UNIVERSITY OF STRASBOURG

MSII GRADUATE SCHOOL: "MATHÉMATIQUES, SCIENCES DE L'INFORMATION ET DE L'INGÉNIEUR"

THESIS PRESENTED BY

CHINMAY SAMANT

ON **DECEMBER** 16th, **2019**

submitted in fulfillment of the requirements for the degree of **Doctor of Philosophy of the University of Strasbourg**

Area/Speciality: **SIAR-Image & Vision (Robotics)**

Ultrasound Laparoscopic Guidance for Minimally Invasive Surgery, Biopsy, and Ablation Procedures

THESIS supervised by:
Adlane HABED, Assistant Professor
Michel DE MATHELIN, Professor

EXTERNAL REPORTERS:
David FOFI, Professor
Pascal VASSEUR, Professor

OTHER MEMBERS OF THE JURY:
John KLEPPER, PhD
Alexandre HOSTETTLER, PhD

Acknowledgements

I thank Siemens Healthcare, France for funding this research (under CIFRE ANRT convention with the ICube - University of Strasbourg, CNRS - Laboratory and IHU, Strasbourg) through ARES project. I thank Mr. John Klepper (Research Scientist, Siemens Healthcare) for realising this project and Mr. Ankur Kapoor (Research Engineer, Siemens Healthcare Princeton) for his contribution in the development of hardware used in this project. I thank Mr. Laurent Goffin for his contributions in developing the software for experimentation. I thank the staff of IHU Strasbourg, in their assistance for using the medical instruments.

I give my sincere thanks to my thesis supervisor, teacher and mentor Mr. Adlane Habed. You showed me how to walk in the world of academics. You supported me in my failures guiding me patiently toward my success. Your passion for achieving excellence in the work is an invaluable treasure for Science and I am fortunate to receive a part of it. Your big heart shielded me in the turbulent times and held me high in the best ones.

I thank my thesis director, Mr. Michel de Mathelin for bringing me in this project. Your decisive and practical advises helped me realise my thesis. With your generous support and encouragement, I could gather the confidence of demonstrating my research to the international platforms. Your leadership and professional efficiency will always be an inspiration for me and ICUBE.

I would like to extend my thanks at my colleagues and members ICUBE laboratory. This endeavour was not possible without your support and encouragement. With the dedication for your work, you keep ICUBE on the forefront of technology and I am fortunate to be a part of it.

I made lifelong friends during my PhD. Each and every moment with you will always bring a smile to my face. Thanks for your help at every step; Lijia, Danda-Pani, Sebastian, Imane, Ana, Maximillian, Arda, Devesh, Jérémy and others.

I thank my parents, Aai and Baba. I thank my second parents, uncle Dadakaka and Smita Kaki. Your support and blessings are the eternal source my success. Ana, I can’t thank you enough for everything you did for me. When I was falling apart, your love held me together. And thank you Mamãe and Papai for believing in me all the time.

Finally, I extend my thanks to the reader, for your time and interest.

Contents

List of Figures

Introduction Générale

Contexte de la thèse

Ce projet est réalisé sous la convention CIFRE-ANRT. Il est financé par Siemens Healthcare (France) en collaboration avec ICube (CNRS), l'IHU et l'Université de Strasbourg à travers le projet ARES. L'objectif de ce projet collaboratif est de développer des technologies liées à la chirurgie mini-invasive et ses instruments.

De nos jours, la chirurgie laparoscopique minimalement invasive guidée par l'image est de plus en plus pratiquée pour le diagnostic et l'intervention. Par rapport à la chirurgie ouverte, une telle procédure ne nécessite que de petites incisions corporelles et conduit à une récupération plus rapide du patient ainsi qu'à un traumatisme post-chirurgical réduit. Dans ce type de chirurgie, le chirurgien s'appuie sur un flux vidéo plutôt que sur un accès visuel direct aux organes. Ceci, combiné avec les progrès technologiques dans les instruments laparoscopiques chirurgicaux et l'imagerie médicale, a conduit les chirurgiens à effectuer la chirurgie de nouvelles façons qui sont uniques à la chirurgie laparoscopique, entraînant, dans certains cas, des procédures plus rapides que celles effectuées dans une chirurgie ouverte et un impact réduit sur l'état de santé du patient.

Bien que la chirurgie laparoscopique devient de plus en plus pratiquée en raison de ses avantages pour les patients, en général, les chirurgiens ont tendance à avoir plus confiance en la chirurgie ouverte. Cela est dû en partie au champ de vision limité, à l'espace confiné dans lequel ils doivent opérer, à la difficile coordination œil-main et au niveau élevé de dextérité manuelle requis en chirurgie laparoscopique. Les chirurgiens peuvent choisir le type de procédure en fonction de leur évaluation des risques. Par exemple, en chirurgie de re-sectionnement du foie, hipatectomie, le chirurgien doit choisir entre une intervention chirurgicale laparoscopique ou une chirurgie ouverte en fonction de la complexité de l'opération. Si les scans pré-chirurgicaux montrent que la propagation de la maladie dans le foie est isolée, la procédure chirurgicale est plus simple et le chirurgien pourrait tout aussi bien utiliser la laparoscopie. Si la maladie est répandue, le chirurgien préférerait une chirurgie ouverte car elle permet de vérifier soigneusement toute la zone et d'éliminer complètement la maladie. La préférence pour la chirurgie ouverte, dans un tel cas, est due au manque de confiance du chirurgien dans les instruments laparoscopiques et l'imagerie.

Dans une chirurgie laparoscopique typique, le chirurgien se limite à une vue 2D de l'anatomie du patient affichée sur un écran. En raison du manque de rétroaction haptique, qui est disponible pour le chirurgien dans une chirurgie ouverte, les chirurgiens s'appuient sur des modalités d'imagerie, telles que les images échographiques (US), en plus de la vue 2D traditionnelle de la caméra du laparoscope sur l'écran. À l'aide de la TDM / IRM préopératoire, des images produites par des appareils de radiologie à arceau et des images laparoscopiques échographiques, les chirurgiens peuvent voir au-delà de la surface des organes avant de les sectionner. Par exemple, dans l'exemple de l'hipatectomie discuté ci-dessus, les chirurgiens s'appuient sur les images médicales pour planifier le chemin de l'incision. Les images préopératoires ainsi que les images peropératoires en direct fournissent des informations visuelles sur l'anatomie du foie, afin que les chirurgiens puissent éviter les vaisseaux majeurs lors de l'ablation de la maladie du foie. Au fur et à mesure que les techniques chirurgicales s'améliorent, la fiabilité et la confiance dans les techniques d'imagerie augmentent. De nombreuses nouvelles techniques de fusion d'images multimodales et de suivi d'instruments laparoscopiques ont été développées pour aider le chirurgien. En conséquence, une vue meilleure, plus complète et plus immersive de la scène chirurgicale augmente la confiance du chirurgien dans les instruments laparoscopiques.

La chirurgie guidée par l'image moderne vise à fournir une vue complète de l'opération au chirurgien. Cette vue peut utiliser différentes modalités d'image fusionnées pour présenter des informations supplémentaires superposées sur la vidéo de la caméra du laparoscope. Par exemple, des images de tomodensitométrie pré-acquises peuvent être superposées sur l'image de la caméra. Les tranches du volume d'image CT-scan sont sélectionnées en temps réel, correspondant aux images de caméra correspondantes du laparoscope. Les étapes de cette technique sont les suivantes:

- Imagerie: les images sont acquises à partir de différents instruments d'imagerie tels que la tomodensitométrie, l'IRM, l'échographie ainsi que la caméra en temps réel avant et pendant la chirurgie.
- Segmentation: la zone de l'opération, telle que la forme de l'organe, doit être séparée du fond des images. Ce processus est appelé segmentation.
- Tracking: il s'agit d'acquérir la pose des instruments laparoscopiques en temps réel par rapport aux objets segmentés de la scène. La pose est acquise à l'aide de capteurs montés sur les instruments.
- Enregistrement: dans cette étape, différentes modalités d'image sont mises en correspondance en identifiant et en faisant correspondre des repères communs. Ce processus peut être très gourmand en ressources CPU dans les applications en temps réel.
- Visualisation et interaction: souvent une partie négligée de l'ensemble du processus, la visualisation des images à l'écran ainsi que l'interaction des chirurgiens avec les instruments et les commandes du système informatique définissent l'utilisabilité, par les chirurgiens, de la technique de guidage.

La machine Siemens S3000, une prouesse technologique pour l'imagerie ultrasonore, intègre la plupart des étapes ci-dessus. Sa technologie logicielle eSieFusion peut effectuer le suivi des sondes à ultrasons pour l'enregistrement des images échographiques en temps réel avec les images de volume de scan CT / IRM préopératoire. La sonde à ultrasons externe est suivie à l'aide de systèmes de capteurs électromagnétiques (EM) tels que NDI Aurora et les informations de pose de ce système sont utilisées pour le récalage. Étant donné que le capteur EM est monté sur la sonde à ultrasons externe, le capteur reste à proximité de l'émetteur EM qui est requis pour le suivi EM. Cependant, dans le cas d'un laparoscope à ultrasons, qui a un réseau d'échographes similaire monté sur sa pointe, le suivi EM traditionnel échoue car la pointe du laparoscope qui doit être suivie pénètre profondément dans le corps. Si le capteur EM est monté sur la pointe, il est soumis à du bruit: des distances plus longues de l'émetteur EM entraînent des erreurs de suivi plus importantes. Dans cette thèse, nous travaillons dans le cadre du suivi des laparoscopes à ultrasons avec le système Siemens eSieFusion. L'objectif est de surmonter les inconvénients du système de suivi EM pour permettre le suivi de la pointe du laparoscope à ultrasons fonctionnant profondément dans le corps du patient.

Portée de la thèse

Dans cette thèse, nous visons à proposer des solutions pour améliorer l'usabilité d'un laparoscope à ultrasons en trouvant sa pose (position et orientation) en temps réel. Cette amélioration serait utilisée pour fournir au chirurgien une information visuelle complète sur l'emplacement de la pointe du laparoscope autour des organes.

Le suivi du laparoscope fait partie d'un pipeline de réalité augmentée dans lequel le chirurgien peut avoir une vue 3D de la procédure chirurgicale sur l'écran. L'objectif est de montrer à l'équipe chirurgicale une image en temps réel du mouvement du laparoscope autour de l'image 3D de l'anatomie du patient. Parmi les nombreuses étapes impliquées dans ce processus, le suivi peut être une limitation pour l'ensemble de la procédure en raison de sa dépendance à l'égard des systèmes de capteurs pour fournir une pose détaillée de l'outil. En suivi, plusieurs capteurs (EM, optiques) sont attachés au laparoscope. En utilisant les informations de ces capteurs, le laparoscope est suivi par rapport à un système de coordonnées de référence. Étant donné que les capteurs sont constamment affectés par l'environnement externe tout en fournissant des données en temps réel, il est difficile d'obtenir une pose en temps réel.

Dans cette thèse, nous abordons les défis du suivi des outils en améliorant le processus d'étalonnage des outils fixes et en proposant un nouveau système de suivi des laparoscopes basé sur des capteurs. Bien que notre objectif soit de suivre la pose d'un laparoscope à ultrasons, la technologie peut être utilisée pour différents instruments laparoscopiques. Par conséquent, nous présentons un aperçu du suivi laparoscopique dans le contexte de la chirurgie guidée par l'image. Tout en présentant les solutions existantes pour le suivi laparoscopique, nous analysons leurs inconvénients et y remédions en proposant notre système de capteurs pour le suivi laparoscopique. Dans le système de suivi proposé, nous avons trouvé deux parties principales du flux de travail essentielles à l'estimation de la pose: le moyennage des données des capteurs et l'étalonnage du corps fixe. Le moyennage des données des capteurs est essentiel pour combiner les données de rotation produisant une valeur unifiée de la composante rotationnelle de la transformation de pose finale. Dans cette thèse, nous présentons différentes méthodes pour résoudre le problème. En outre, l'étalonnage du corps fixe, également connu comme *Hand-Eye calibration*, fournit une transformation fixe entre deux capteurs de pose connectés rigidement à un corps immobile. Notez que, dans toutes les solutions de suivi des laparoscopes à ultrasons existantes, les capteurs de pose sont fixés de manière rigide au corps des instruments laparoscopiques et, par conséquent, le problème de l'étalonnage œil-main se pose. Nous fournissons un historique détaillé de ce problème ainsi que des solutions de pointe. Nous présentons également une nouvelle méthode d'étalonnage main-œil qui est robuste au bruit et aux valeurs aberrantes. Nous testons cette méthode de manière exhaustive par rapport aux méthodes existantes et présentons les résultats en utilisant différents types de données.

Contribution de la thèse

La première contribution de cette thèse est dans le processus d'étalonnage des outils. Les systèmes de suivi optique et EM ont des limitations physiques qui réduisent leurs performances. Par exemple, les systèmes optiques doivent maintenir un champ de vision dégagé et ne peuvent donc pas être utilisés pour suivre la pointe des outils laparoscopiques, car ceux-ci fonctionnent à l'intérieur du corps du patient où la ligne de visée entre le marqueur optique et la caméra ne peut pas être maintenue. Les capteurs de suivi EM sont flexibles et de très petite taille. En conséquence, ils peuvent être utilisés pour suivre la pointe du laparoscope mais peuvent être constamment affectés par des perturbations électromagnétiques. Leurs performances se dégradent également rapidement à mesure que la distance entre l'émetteur EM et le capteur récepteur augmente. Pour compenser les inconvénients de ces deux systèmes, des technologies de capteurs hybrides ont été développées. Ces technologies utilisent conjointement le suivi optique et EM. Un bloc fondamental de ce système est l'étalonnage du système dans lequel un étalonnage fixe entre les capteurs montés sur le laparoscope est obtenu. En robotique, cet étalonnage fixe est appelé «étalonnage main-œil».

Dans cette thèse, nous proposons une nouvelle formulation du problème d'étalonnage main-œil. Notre formulation repose sur la contrainte de la rotation recherchée pour appartenir à ce que l'on appelle l'enveloppe convexe des rotations. Sur la base de cette formulation, nous avons conçu un algorithme robuste déterministe pour résoudre le problème d'étalonnage main-œil. Notre algorithme est robuste aux valeurs aberrantes ainsi qu'au bruit dans les données obtenues à partir des capteurs de pose. En particulier,

- nous proposons une relaxation du problème d'étalonnage main-œil qui peut être résolu en utilisant de la programmation semi-définie, pour laquelle une fonction convexe des matrices de rotation est requise. Nous utilisons le concept d'enveloppe convexe des matrices de rotation pour formuler notre problème sous sa forme convexe;

- afin d'assurer l'appartenance de la solution à l'ensemble des rotations, nous reformulons le problème relaxé en un algorithme d'itération convexe de contrainte de rang. Bien que le problème ne soit pas convexe à ce stade, notre formulation permet d'utiliser un schéma d'optimisation alterné pour converger systématiquement dans la pratique vers la solution recherchée;

- afin de filtrer le bruit élevé des capteurs et les valeurs aberrantes, nous proposons une approche de repondération itérative performante. Avec cet algorithme, nous résolvons systématiquement le problème d'étalonnage main-œil même en présence de niveaux élevés de bruit et de données aberrantes.

La méthode complète est publiée dans la conférence IROS 2019 [Sam+19] y étant acceptée pour une présentation orale. De brefs résultats de cette méthode sont présentés dans CNIV 2019 [SHM19a].

Parallèlement à notre méthode robuste pour l'étalonnage main-œil, nous présentons également une approche conventionnelle utilisant RANSAC (*Random Sample Consensus*) pour filtrer les valeurs aberrantes. Nous présentons un cadre RANSAC basé sur un seuillage «géométrique» dans lequel tout algorithme d'étalonnage main-œil peut être utilisé. Nous avons présenté cet algorithme ainsi que les résultats expérimentaux dans l'édition 2019 de la conférence CRAS-SPIGC [SHM19b].

Une autre contribution de cette thèse est un nouveau système de suivi pour les laparoscopes à ultrasons. Un laparoscope à ultrasons se compose d'un ensemble d'émetteurs d'ultrasons à l'extrémité mobile du laparoscope. Cet équipement fournit une image en temps réel au-delà de la surface des organes, aidant le chirurgien à planifier et à suivre la chirurgie avant de pratiquer une incision. En raison de la nature même de l'image échographique, l'estimation de la pose d'un laparoscope à ultrasons via cette image est un défi. Si la pointe du laparoscope est hors du champ de vision de l'endoscope vidéo qui l'accompagne, le chirurgien peut trouver les images échographiques désorientantes. La technologie de suivi EM peut aider cette situation si un capteur EM est attaché à la pointe du laparoscope, mais la précision du suivi EM est grandement affectée lorsque le laparoscope est à l'intérieur du corps du patient. Nous proposons une nouvelle solution à ce défi en utilisant des capteurs optiques et des centrales inertielles (IMU). Les IMU sont des capteurs électroniques autonomes qui peuvent fournir des informations de rotation fiables. Nous proposons d'utiliser une série d'IMU montés sur le corps et la pointe du laparoscope qui sont accompagnés d'un suivi optique. Le suivi optique fournit les informations de position et d'orientation de la poignée du laparoscope. Combinés ensemble, les IMU et le suivi optique peuvent fournir une pose à 6 degrés de liberté (DOF). Les avantages de ce système de suivi sont qu'il n'est pas limité par la portée comme le système de suivi EM et n'est limité que par la contrainte de ligne de visée du système de suivi optique.

Organisation de la thèse

Le deuxième chapitre de la thèse présente une brève histoire de la chirurgie laparoscopique et des outils. Il décrit également une procédure générale en chirurgie laparoscopique pour fournir un contexte à l'application de la thèse. Il présente ensuite brièvement les étapes d'une chirurgie guidée par l'image. Ensuite, nous présentons un aperçu complet des technologies de suivi des laparoscopes ainsi qu'une description de deux technologies de suivi hybrides. Au final, nous présentons une construction générale des laparoscopes à ultrasons ainsi que notre proposition pour le suivi hybride optique-IMU.

Le troisième chapitre de cette thèse présente la conception proposée du système de suivi hybride basé sur un capteur optique-IMU pour le laparoscope à ultrasons. Tout d'abord, les composants de configuration matérielle du laparoscope sont expliqués en détail. Ensuite, nous présentons la chaîne cinématique et l'étalonnage du laparoscope. Enfin, nous présentons les sources d'erreurs possibles et la solution à celles-ci en utilisant le moyennage des rotations.

Le quatrième chapitre de cette thèse présente l'état de l'art du moyennage des rotations et de l'étalonnage main-œil. Puisque nous utilisons plusieurs IMU sur le laparoscope le long du même axe, ils fournissent les mêmes informations de rotation. Nous utilisons cette redondance pour améliorer l'estimation de la rotation en utilisant des algorithmes de moyennage des rotations. L'étalonnage main-œil est un sous-domaine du moyennage des rotations. L'état de l'art sur l'étalonnage main-œil dans ce chapitre sert de précurseur au chapitre suivant.

Le cinquième chapitre de cette thèse présente notre robuste méthode d'étalonnage main-œil. Dans ce chapitre, le problème d'étalonnage main-œil est réintroduit du point de vue de la robotique. Nous présentons également un bref historique de la programmation semi-définie, de la programmation semi-définie à contrainte de rang ainsi que de l'enveloppe convexe de rotations. Notre méthode est présentée avec l'algorithme proposé. Enfin, des essais sur des configurations synthétiques et différentes configurations de données réelles sont présentés avec les résultats de tests approfondis de cette méthode sur plusieurs critères.

Le sixième et dernier chapitre conclut cette thèse. Il réitère également les apports de cette thèse et leur importance dans ce projet. Au final, nous présentons les travaux restants qui n'ont pas été réalisés pendant la durée du projet. Dernièrement, nous indiquons l'impact positif du laparoscope à ultrasons sur l'imagerie médicale et, en général, sur la chirurgie une fois les futurs travaux réalisés.

Chapter 1

Introduction

1.1 Context of the thesis

This project is realised under CIFRE-ANRT convention. It is funded by Siemens Healthcare (France) in collaboration with ICube (CNRS), IHU and University of Strasbourg through ARES project. The goal of this collaborative project is to develop technologies related to minimally invasive surgery and its instruments.

Nowadays, minimally invasive image-guided laparoscopic surgery is increasingly performed for diagnosis and intervention. In comparison to open surgery, such procedure requires only small body incisions and leads to a faster recovery of the patient as well as a reduced post-surgical trauma. In this type of surgery, the surgeon relies on a video stream rather than a direct visual access to the organs. This, combined with the technological advances in the surgical laparoscopic instruments and medical imaging, has led the surgeons to perform surgery in new ways that are unique to laparoscopic surgery, resulting, in some instances, in faster procedures than those performed in an open surgery and a reduced impact on patient's health.

Although laparoscopic surgery is gaining ground because of its benefits to the patients, in general, surgeons tend to have more confidence in open surgery. This is partly due to the limited field of view, to the confined space they ought to operate in, to the difficult hand-eye coordination, and to the high level of manual dexterity required in laparoscopic surgery. Surgeons may choose the type of procedure based on their risk assessment of its success. For example, in liver re-sectioning surgery, Hipatectomy, the surgeon has to choose between laparoscopic surgical procedure or open surgery depending upon the complexity of the operation. If the pre-surgical scans show that the spread of the disease in liver is isolated then the surgical re-sectioning is simpler and the surgeon might as well use laparoscopy. If the disease is wide-spread, then the surgeon would prefer open surgery as it allows one to thoroughly check the entire area and remove the disease completely. The preference to open surgery, in such a case, is due to the surgeon's lack of confidence in the laparoscopic instruments and imaging.

In a typical laparoscopic surgery, the surgeon is limited to a 2D view of the patient's anatomy displayed on a screen. Due to the lack of haptic feedback, which is available to the surgeon in an open surgery, surgeons rely on imaging modalities, such as ultrasound (US) images, in addition to the traditional 2D view from the laparoscope's camera on the screen. With the help of pre-surgical CT/MRI scan, intra-operative C-arm images and ultrasound laparoscopic images, surgeons can see beyond the surface of the organs before intersecting them. For instance, in the Hipatectomy example discussed above, the surgeons rely on the medical images to plan the path of incision. The pre-operative as well as live intra-operative images provide visual information of the anatomy of the liver, so the surgeons can avoid the important arteries during the removal of the disease from the liver. As minimally surgical techniques are improving, the reliance and confidence on the imaging techniques is growing. Many new techniques of multimodality image fusion and laparoscopic instrument tracking have been developed to aid the surgeon. As a result, a better, more comprehensive and immersing view of the surgical scene increases the surgeon's confidence in Laparoscopic instruments.

Modern image guided surgery aims to provide a comprehensive view of the operation to the surgeon. This view may use different modalities fused together to present additional information overlayed on the video from the laparoscope's camera. For example, images from pre-acquired CT-scans can be overlayed on the camera image. The slices of the CT-scan image volume are selected in real-time, matching the corresponding camera images from the laparoscope. The steps involved in this technique are as follows:

- Imaging: images are acquired from different imaging instruments such as CT-scan, MRI-scan, ultrasound as well as camera in real-time during the surgery and also before the surgery.
- Segmentation: the area of the operation, such as the shape of organ, needs to be separated from the background of the images. This process is known as segmentation.
- Tracking: it involves acquiring the pose of the laparoscopic instruments in real-time with respect to the segmented objects in the scene. The pose is acquired using sensors that are mounted on the instruments.
- Registration: in this step, different image modalities are matched together by identifying and matching common landmarks in them. This process can be very CPU-resource intensive in real-time applications.
- Visualization and interaction: often an overlooked part of the entire process, the visualization of the images on the screen as well as surgeons interaction with the instruments and the controls of the computer system define the usability of the image guidance technique by the surgeons.

Siemens S3000 machine, a technological prowess for ultrasound imaging, incorporates most of the steps above. Its eSieFusion software technology can perform tracking of ultrasound probes for registration of real-time ultrasound images with the preoperative CT/MRI scan volume images. The external ultrasound probe is tracked using electromagnetic (EM) sensor systems such as NDI Aurora and the pose information from this system is used for registration. Since the EM sensor is mounted on the external ultrasound probe, the sensor remains within the vicinity of the EM transmitter unit that is required for EM tracking. However, in the case of an ultrasound laparoscope, which has a similar ultrasound array mounted on its tip, the traditional EM tracking fails as the tip of the laparoscope that needs to be tracked goes deep inside the body. Should the EM sensor be mounted on the tip, it is subjected to noise: longer distances from the EM transmitter lead to larger tracking errors. In this thesis, we work in the context of ultrasound laparoscope tracking with the Siemens eSieFusion system. The goal is to overcome the drawbacks of the EM tracking system to allow the tracking of the US laparoscope tip functioning deep in the body of the patient.

1.2 Scope of the thesis

In this thesis, we aim to propose solutions for improving the usability of an ultrasound laparoscope by finding its pose (position and orientation) in real-time. This improvement would be used to provide a comprehensive visual information to the surgeon about the location of the laparoscope's tip around the organs.

Laparoscope tracking is a part of an augmented-reality pipeline in which the surgeon can see a 3D view of the surgical procedure on the screen. The goal is to show the surgical team a real-time picture of the movement of the laparoscope around the 3D image of the anatomy of the patient. Amongst the many steps involved in this process, tracking can be a bottleneck to the entire procedure due to its reliance on sensor systems to provide a detailed pose of the tool. In tracking, several sensors (EM, optical) are attached to the laparoscope. By using the information from these sensors, the laparoscope is tracked with respect to some reference coordinate system. Since the sensors are affected by the external environment constantly while providing real-time data, there are many challenges in obtaining a real-time pose.

In this thesis, we address the challenges in tool tracking by improving the fixed tool calibration process as well as proposing a new sensor-based laparoscope tracking system. Though our aim is to track the pose of an ultrasound laparoscope, the technology can be used for different laparoscopic instruments. Hence, we present an overview of laparoscopic tracking in the context of image-guided surgery. While presenting the existing solutions for laparoscopic tracking, we analyse their drawbacks and address them by proposing our sensor system for laparoscopic tracking. In the proposed tracking system, we found two main parts of the workflow critical to the pose estimation: sensor data averaging and fixed-body calibration. Sensor data averaging is essential to combine the rotation data to produce a unified value of the rotational component of the final pose transformation. In this thesis, we overview different methods to solve the problem. Furthermore, fixed body calibration, also known as Hand-Eye calibration, provides a fixed transformation between two pose sensors connected rigidly to a non-moving body. Note that, in all existing ultrasound laparoscope tracking solutions, the pose sensors are rigidly attached to the body of the laparoscopic instruments and hence the problem of Hand-Eye Calibration arises. We provide a detailed background on this problem as well as the state-of-the-art solutions. We also present a novel Hand-Eye calibration method which is robust to noise and outliers. We test this method comprehensively against existing methods and present the results using different types of data.

1.3 Contribution of the thesis

The first contribution of this thesis is in the tool calibration process. Both optical and EM tracking systems have some physical limitations reducing their performance. For example, optical systems need to maintain a clear line of sight and hence cannot be used to track the tip of the laparoscopic tools since these operate inside the patient's body where the line of sight between the optical marker and the camera cannot be maintained. EM tracking sensors are flexible and very small in size. As a result, they can be used to track the laparoscope tip but may constantly be affected by electromagnetic disturbances. Their performance also degrades rapidly as the distance between the EM transmitter and the receiver sensor increases. To compensate the drawbacks of these two systems, hybrid sensor technologies have been developed. These technologies use optical and EM tracking together. A fundamental block in this system is the calibration of the system in which fixed calibration between the sensors mounted on the laparoscope is obtained. In robotics, this fixed calibration is known as 'Hand-Eye calibration'.

In this thesis, we propose a novel formulation of the Hand-Eye calibration problem. Our formulation relies on constraining the sought rotation to belong to the so-called convex-hull of rotations. Based on this formulation, we devised a deterministic robust algorithm to solve the Hand-Eye calibration problem. Our algorithm is robust to outliers as well as to noise in the data obtained from the pose sensors. In particular,

- we propose a relaxed version of the Hand-Eye calibration problem that can be solved using semi-definite programming, for which a convex function of rotation matrices is required. We use the concept of convex hull of rotation matrices to formulate our problem in its convex form;
- in order to ensure the solution's membership to the set of rotations, we re-formulate the relaxed problem into a rank-constraint convex iteration algorithm. Although the problem is non-convex at this stage, our formulation allows to use an alternating optimization scheme to consistently converge in practice to the sought solution;

- in order to filter out high sensor noise and outliers, we propose an efficient iterative re-weighting approach. With this algorithm, we consistently solve the Hand-Eye calibration problem even in the presence of high levels of noise and outlier data.

The full method is published in IROS 2019 conference [Sam+19] in which it has been accepted to oral presentation. Brief results of this method are presented in CNIV 2019 [SHM19a].

Alongside our robust method for Hand-Eye calibration, we also present a conventional approach using RANSAC (Random Sample Consensus) to filter out outliers. We present a RANSAC framework based on 'geometric' thresholding in which any Hand-Eye calibration algorithm can be used. We presented this algorithm along with the experimental results in the 2019 edition of CRAS-SPIGC conference [SHM19b].

Another contribution of this thesis is a new tracking system for ultrasound laparoscopes. An ultrasound laparoscope consists of an ultrasound array at the moving tip of the laparoscope. The ultrasound modality provides a real-time image beyond the surface of the organs helping the surgeon to plan and follow the surgery before making an incision. Due to the nature of ultrasound image itself, estimating the pose of ultrasound laparoscope via this image is a challenge. If the laparoscope tip is out of the view of the accompanying video endoscope, the surgeon can find the ultrasound images disorienting. EM tracking technology can help this situation by attaching an EM sensor to the tip of the laparoscope but the accuracy of the EM tracking is greatly affected when the laparoscope is inside the patient's body. We propose a new solution to this challenge by using optical and Inertial Measurement Unit (IMU) based sensors. IMUs are self-contained electronic sensors that can provide reliable rotation information. We propose to use a series of IMUs mounted on the body and the tip of the laparoscope that are accompanied by optical tracking. Optical tracking provides the position and orientation information of the laparoscope handle. Combined together, IMUs and optical tracking can provide pose in 6 degrees of freedom (dof). The advantages of this tracking system is that it is not limited by the range like the EM tracking system and is only limited by the line of sight constraint of the optical tracking system.

1.4 Organization of the thesis

The second chapter of the thesis presents a brief history of laparoscopic surgery and tools. It also describes a general procedure in laparoscopic surgery to provide a context to the application of the thesis. Then it shortly presents the steps involved in an image-guided surgery. Following this, we present an extensive overview of laparoscope tracking technologies along with a description of two hybrid tracking technologies. In the end, we present a general construction of ultrasound laparoscopes along with our proposal for the optical-IMU hybrid tracking.

The third chapter of this thesis presents the proposed design of the Optical-IMU sensor-based hybrid tracking system for the ultrasound laparoscope. First, the hardware setup components of the laparoscope are explained in detail. Then, we present the kinematic chain and calibration of the laparoscope. Finally, we present the sources of possible errors and the solution to those by using rotation averaging.

The fourth chapter of this thesis presents the state-of-the-art of Rotation Averaging and Hand-Eye calibration. Since we use multiple IMUs on the laparoscope along the same axis, they provide the same rotation information. We use this redundancy to improve the rotation estimation using rotation averaging algorithms. Hand-Eye calibration is a sub-field of rotation averaging. The state-of-the-art on Hand-Eye calibration in this chapter serves as a precursor to the next chapter.

The fifth chapter of this thesis presents our robust Hand-Eye calibration method. In this chapter, the Hand-Eye calibration problem is re-introduced from a robotics perspective. We also present a brief background on Semi-Definite programming, Rank-Constrained Semi-Definite Programming as well as the Convex Hull of Rotations. Our method is presented along with the proposed algorithm. Finally, the experiments on synthetic and different real data setups are presented with results of extensive testing of this method on several criteria.

The sixth and final chapter concludes this thesis. It also re-iterates the contributions of this thesis and their importance in this project. In the end, we present the remaining work that was not realised during the period of this project. Finally, we state the positive impact of the ultrasound laparoscope on medical imaging and, in general, on surgery once the future work comes to a realisation.

Chapter 2

Instrument Registration In Laparoscopic Surgery

In this chapter, we present a brief history of the development of laparoscopes that covers the different ways of using laparoscopes throughout the 19th and 20th century. Current standard laparoscopic surgery, in which the a video is presented to the surgeon on the screen, was developed in the late 20th century. This chapter concisely describes laparoscopic surgery to provide a context in which the laparoscopes are used. The innovation in medical imaging along with the camera improvements has led the development of entirely new field in surgery known as image-guided surgery. The technology involved in this field allows the surgeons to access visual information from multiple modalities allowing them to plan and execute the surgeries efficiently with laparoscopic instruments. In this chapter, we describe the steps involved in image-guided surgery. We also focus on the laparoscopic instrument tracking technologies since the tracking is an essential part of image-guided surgery. After discussing different types of laparoscope tracking, we present and discuss the state-of-the-art tracking solutions for ultrasound laparoscope tracking that are the focus of this thesis.

2.1 History of Laparoscopic Surgery

Modern medicine and its sub-branches allowed to greatly improve the quality and the length of life at the turn of 19th century. Along with the advances in medicine, the field of surgery also evolved thanks to the industrial precision tools as well as a greater understanding of human anatomy and chemical compositions. A surgery involved the following steps: identification of the problem (diagnosis), the actual surgical act, which included access to the problem inside human body as well as rectification of the problem, and in the end, post-surgical treatment of the patient. Despite their undeniable advantages, the surgical procedures cause immense trauma to human body and mind. In particular, when the surgeon has to cut open the top layers of skin and muscles to get access to an internal organ, the wound caused by the surgery can be fatal in some cases. Sometimes the diagnosis of a health problem is not just possible by observing the patient from outside and the surgeon requires an internal view of the body. Before the invention of X-ray machines, the only option to the surgeon in this case was to take an incision on the skin to take a peek inside. The invention of laparoscopy lies in these surgical procedures when the surgeon desires to inflict the least damage to the patient's body and perform the diagnosis as well as the surgery. Due to the industrial revolution in the 19th century, precise surgical instruments were available in the form of laparoscopes. These instruments allowed the surgeons to access the insides of human body with minimal wounds on the skin and muscles, which allowed the surgeons a better diagnosis while improving the recovery time of the patients.

The first tools that allowed ancient surgeons to peek inside human orifices such as rectum, ear, nose, vagina etc. were developed over two millennia ago. Ancient Greeks, Romans, Egyptians, Indians used several instruments that provided the ability to look inside the human body using the natural light. The first internally lit device used to inspect the interior of human body was constructed by Philipp Bozzini of Mainz, Germany in 1806. He called this device 'Lichtleiter' or light conductor. It was constructed of a tube, with various attachments, to be inserted into the human body cavity (FIGURE: 2.1).

FIGURE 2.1: Bozzini's Lichtleiter. (Image from www.facs.org)

In 1953, Antonin Jean Desormeaux, a French physician, invented a device with significant improvements to the early versions of laparoscopes and coined the term endoscope (FIGURE: 2.2). His device is called a Cycstoscope today. He also used his device to operate living patients for the first time in history. His device contained an open tube system with mirrors and lenses to examine urinary tract and bladder. This endoscope was manufactured in large quantities proving the popularity of the design. It used a flammable mixture of turpentine and alcohol to produce light.

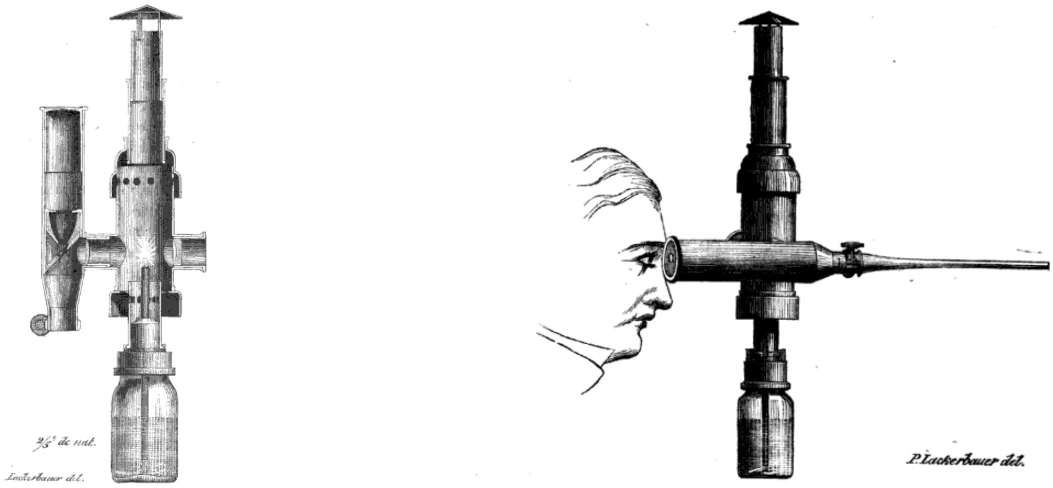

FIGURE 2.2: Left: Internal schematic of endoscope of Antonin Jean Desormeaux by Pierre Lackerbauer showing the mirror system to reflect light. Right: The usage of the endoscope.

In [Feu08], the author provides a short review on the history of the developement of the laparoscopes. The first laparoscope with electrical lighting was invented in 1877 by a German doctor, Maximilian Nitze. He published the designs of a Urethroscope and a Cystoscope to observe Urethra and Bladder respectively (FIGURE: 2.3). In 1879, together with Josef Leiter, he presented an improved version of the cystoscope.

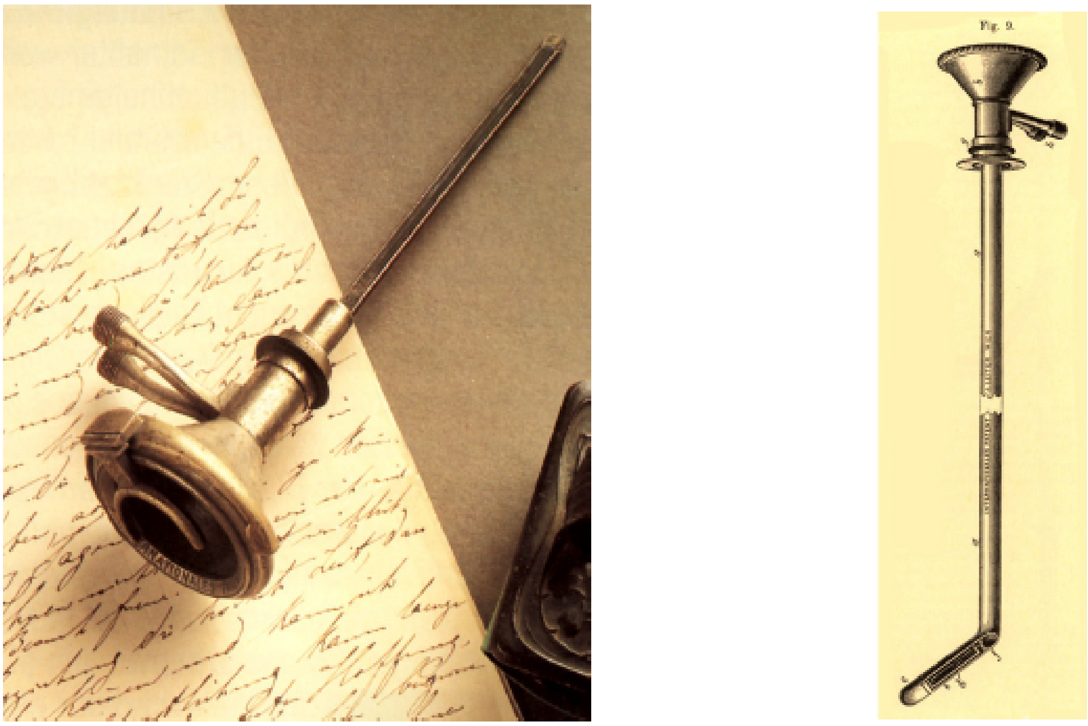

FIGURE 2.3: Nitze's Telescopic Laparoscope instruments. (Images from Nitze-Leiter Museum of Endoscopy)

In 1901, first diagnostic laparoscopic examination was performed by the German surgeon Georg Kelling on a dog using Nitze's Cystoscope. The term 'Laparoscopy' was coined by a Swedish surgeon Hans Christial Jacobaeus as 'Laparothorakoskopie' for the examination of the human peritoneal, thoracic and pericardial cavities. The word comes from the Greek words 'Lapara': the soft part of the body between the ribs, hips and loin i.e. abdomen, and 'Skopein': to survey.

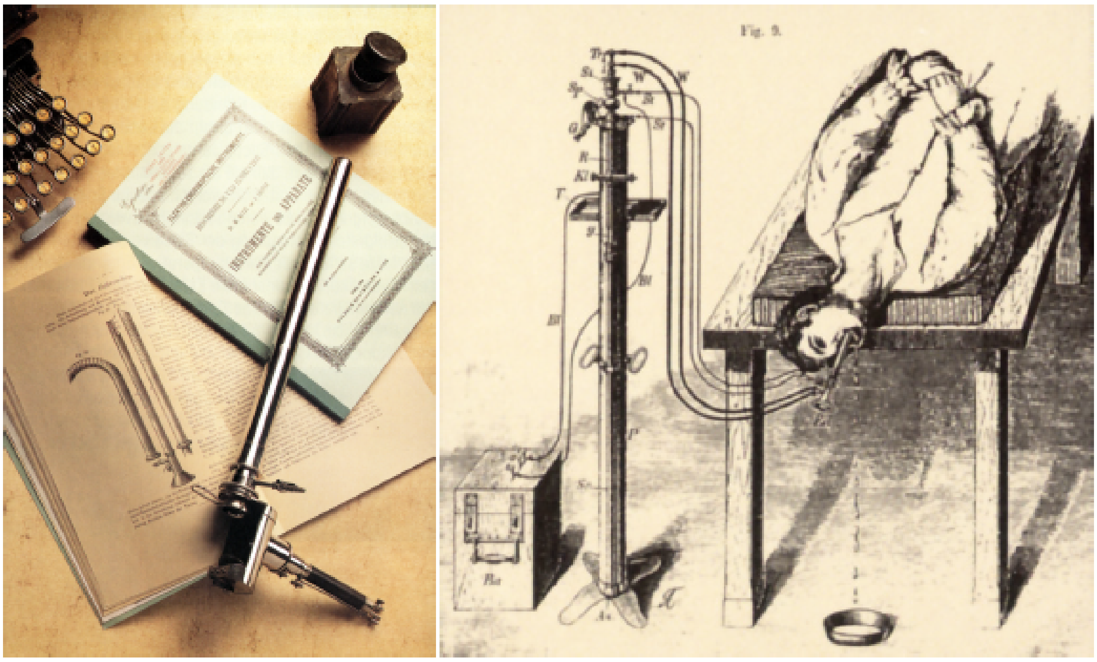

FIGURE 2.4: Left: A Panelectroscope from 1907. Right: Patient undergoing a gastroscopy. (Images from Nitze-Leiter Museum of Endoscopy)

Endoscopy was used mainly for diagnosis until video-based systems were invented in 1980. This allowed the image to be shown on the display and all the members of the surgical team could observe the process. This access meant that new laparoscopic surgical techniques could be developed involving modern practices such surgeons using the laparoscopic tools with their two hands and their assistants holding the endoscopic camera. After a successful cholecystectomy, which involves removing the gallbladder, by O.D. Lukichev in 1983 the laparoscopic surgery started to evolve rapidly and video endoscopy was successfully introduced into other surgical disciplines. Comprehensive reviews on the history of endoscopy and laparopscopy can be found in [BF00], [LLL97] and [Lit99].

2.2 A Standard Laparoscopic Procedure

Laparoscopic surgery is generally performed for the partial re-sectioning of an organ with a disease. Trocars (hollow pipes serving as an access for the laparoscopes) are inserted using few small incisions on the abdomen. Usually two to four trocars of diameter from 10 to 15mm are inserted under general anaesthesia. At least one of the trocar opening is reserved for the laparoscopic camera that allows an internal view of the organs and the instruments. Optimal selection of trocar placement is important for the ease of access during the entire procedure. Surgeon selects the trocar entries by palpation and marks the places with a surgical ink marker before insertion of the trocar. Planning of the instrument placement is crucial for a laparoscopic surgery.

Before inserting the instruments and the laparoscopic camera through trocars, carbon dioxide is inserted in the the body through the trocar opening. This inflates the area of the surgery providing additional space for the surgeon to move the instruments. The laparoscopic camera also relies on this additional space to observe the organs. The laparoscopic camera has a wide angle lens to provide a wider perspective of the organs. It is also enabled with a lighting system. The source of the lighting system is usually situated outside of the patient's body to avoid heat. The light is brought inside by using fibre optics. After the laparoscopic camera is inserted through trocar, the surgeon proceeds with inserting the laparoscopic instrument while observing the video on a screen. (FIGURE: 2.5).

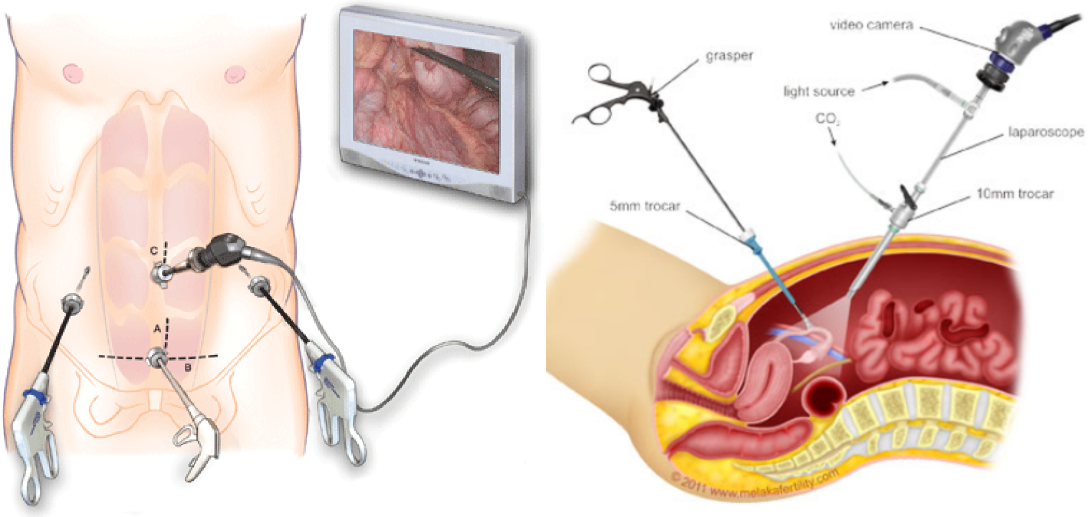

FIGURE 2.5: Left: Overview of a laparoscopic surgery scene (Image from www.columbiasurgery.org). Right: Inside the patient's body during a laparoscopic surgery. (Image from www.longislandsurgery.org)

In Laparoscopic Cholecystectomy, an entire Gallbladder is removed. Contrary to this, laparoscopic liver surgery removes only parts of a liver. Hence before sectioning a liver, the surgeon must have some knowledge about segments of the liver with disease and then proceeds with the surgery. This is often achieved by pre-operative CT or MRI scans that can provide entire volume information of the liver. Based on this information, the surgeon can plan the surgery. Additional to the pre-operative information, surgeons also benefit from live tomographic information from C-Arm. It is a C-shaped X-ray device that can provide live images during the surgery (FIGURE: 2.6).

FIGURE 2.6: Left: C-arm imaging machine (Image from www.kiranxray.com). Right: C-arm machine during a surgery (Image from www.siemens-healthineers.com)

Although C-arm allows a higher resolution image, the X-rays are harmful to the patient as well as the surgical staff. Ultrasound laparoscopes, which contain an ultrasound imaging array at the tip of the laparoscope, can safely provide cross section imaging of an organ such as liver. Using these techniques as well as pre-operative imaging, surgeons can plan the surgery efficiently.

Once surgeons decide how to access an organ, such as liver, they mark the area on the liver using electro-cauterization (burning the tissue with electricity). The same tool can be used for ceiling any open arteries and veins to stop bleeding. Apart from standard tools such as scalpels and forceps, several advanced tools (ultrasound scalpels, ultrasound surgical aspirator, surgical plastic clips, etc.) are used during a surgery. The part with the disease, such as tumour, is retraced through one of the trocars.

2.2.1 Advantages and Limitations

If operated using minimally invasive surgery, the patient gets smaller scars, lesser pain and discomfort during the healing process as well as less physical and emotional trauma. It results into shorter hospital stays and faster healing. Some laparoscopic procedures are much faster compared to an open surgery for the same purpose.

Due to the cost of instruments and the required additional training of the surgical staff, laparoscopic procedures can be very expensive. In the case of cancer related operations, when the surgeon wants to remove as much the malignant parts of the organs as possible, the laparoscopy can restrict the view of the surgeon. Due to the restricted 2D view of the surgery, surgeons are limited with their senses through the laparoscopic instruments. Direct palpation of the organs, vessels and tumours is not available to the surgeon in a laparoscopic surgery. This lack of tactile feedback along with the the restricted 2D view and surgical space can result in longer surgery times and, in some unfortunate cases, post-surgery complications. To compensate these obvious drawbacks of laparoscopic surgery, the surgeons rely upon the their training, advanced imaging as well as proper planning. For example, improper trocar placement relative to the patient's anatomy can increase the surgery time and can cause additional pain to the patient, extending the recovery time. Advanced imaging techniques can provide a hybrid 3D-like view to the surgeons in which they can fuse the pre-operative CT/MRI images with live ultrasound laparoscope or C-arm images. Therefore, the image guidance techniques are useful for the improvement of minimally laparoscopic surgery [Mår+05].

2.3 Image-guided surgery

Image-guided/assisted laparoscopic surgery involves the use of augmented-reality techniques to assist the surgeons during surgery. This type of visualisation provides an advanced view to the surgeons enabling them to plan and execute the surgical procedures more accurately and safely while being faster at the same time. Frequent use of augmented reality in laparoscopic surgery and, in general, in the medical field may result in new efficient techniques to treat the patients. In [Feu08], the author provides a detailed survey on augmented-reality in laparoscopic surgery illustrating all the steps involved in it.

In an open-surgery procedure, surgeons have a direct access to the patient's organ giving them a direct visual-haptic feedback of the surgical situation. Since this feedback is not available in laparoscopic surgery, augmented-reality aims to replace it with visual signals on the display overlayed on the existing laparoscopic video. The video fusion can consist of many different image modalities. The images from these modalities may come from pre-operative images or in real-time. Such systems have been developed and are summarized in [YC06]. This survey focuses on the applications where the tumour region is marked in pre-operative images and, based on its position, the trocar placement is planned. During surgery, the surgeon is assisted by a video in which the tool's path is displayed with respect to the pre-planned path in real-time. This is achieved by registration, a process in which the coordinate system of the pre-operative medical images and the real-time video are matched. In this section, we present the technical and practical steps involved in an augmented-reality based laparoscopic surgery.

2.3.1 Imaging

In modern surgical practices, imaging is involved from the beginning to the end. It may help with the diagnosis of the disease. It may assist during the surgery and also it may show the assessment of the success in terms of recovery during the healing of the patient. Several non-invasive surgical techniques that are used illustrate different anatomical aspects of the patient's body. Each imaging modality presents a unique view of the anatomy of the patient that can add to the existing knowledge of the patient. The modalities used in modern surgery are X-ray, CT (computed tomography)-scan, MRI (magnetic resonance tomography)-scan, PET (positron emission tomography) scan, SPECT(single photon emission computed tomography)-scan, photo-acoustic imaging and ultrasound-based imaging such as bi-planar, Doppler, Elastography, etc.

Along with non-invasive imaging, surgeons also use invasive techniques for diagnosis using laparoscopy. The results of the imaging (pre and intra-operative) are combined together to form a complete picture required to plan and execute a surgery. The information coming from all these modalities can be in 2-D, 3-D or 4-D (3-D image with time) images. Certain imaging techniques such as Doppler ultrasound can also provide the direction of fluid flow through the vessels. All this information can be used in construction of the augmented reality scene for the surgical team. An extensive survey on the imaging techniques can be found in [HR03].

2.3.2 Segmentation

Once the image is analysed by the surgeons, they identify the problem in the organs and mark the boundaries of the organs in the images. The process of marking these areas on an image, associating the image pixels (or voxels) to the target region is called segmentation. Due to the inherent limits of physics, not every imaging modality can visualize the separation of each layer inside the human body. For example, X-rays are blocked by thicker bones but pass easily through a fatty tissue. Hence, they are an ideal choice to observe the bone structure but not tumours. Radiologists use different modalities to construct the big picture of a problem and segments the target areas in one or multiple modalities in a software. The software can extrapolate/connect these segmented areas to provide a 3D visual of the anatomy. Based on the voxel size and other parameters, the software can also present additional details such as tissue volume, rigidity, blood-flow etc. With this information, along with the images, surgeons can plan surgeries efficiently. On the down-side, segmentation is a time consuming process that involves repeated interactions with the physician. To be effective, the process must be tailored to the specific case of a patient. Hence, the prospects of automated segmentation using modern computer vision/machine learning techniques are currently limited in use.

2.3.3 Tracking

Tracking is a process in which the pose (orientation and position) of the surgical instruments involved in a laparoscopic surgery is estimated in real-time. It is achieved by using sensor systems mounted on the instruments as well as in the surroundings. Tracking is required for real-time estimation of the tool's pose which makes it an essential part of the image-guided surgery. Pose of the tool is usually tracked with respect to a fixed frame of reference present in the operation room. In a self contained commercial tracking system product, the tracking is carried out with respect to the coordinate frame of the tracking hardware (e.g transmitter box of EM tracking system). The tracking data usually has 6 degrees of freedom (3 for rotation and as many for translation). This data is fed to the augmented-reality computed along with the calibration information of the reference point with respect to the image acquired by the endoscope (or pre-acquired CT/MRI image data). Using all these transformations together, a real-time view of the tool is overlayed on the video shown to the surgeon.

Tracking is the major bottleneck of image guided surgery. Technological limitation such as electromagnetic interference on EM tracking sensors as well as practical limitations such as occlusions in camera-type sensors can drastically affect the performance of tracking. One of the contributions of this thesis is to present a novel hybrid tracking system that may address some of the problems in tracking. There are several commercial as well as experimental systems available in the market. All of them are either based on optical, mechanical or electromagnetic type of sensors. The commercially available tracking systems use one of the technologies in their products. But given that each tracking technology comes with its inherent limitations, some experimental tracking systems try to overcome these limitations by combining multiple tracking technologies. In [Mar+03], the authors compare an optical and mechanical tracking system and find that both technologies perform with a sub-millimeter accuracy. Due to the bulkiness, installation cost and maintenance of a mechanical tracking systems, they become impractical when compared to other types of systems. In [Kha+00], the authors compare two commercially available tracking systems on an experimental testbed. In 5 different experimental settings, they compare the NDI Polaris system against FlashPoint (Boulder Innovation Group). They state that, both systems are sensitive to the angle between the camera and the marker mounted on the surgical tool. The distance between the camera and the markers also affects the performance introducing jitters in tracking. Optical tracking systems also suffer from inherent limitation such bad lighting, reflections as well as occlusions. In [Sch+05], the authors test the performance of electromagnetic (EM) tracking systems. They conclude that EM tracking systems do not perform well compared to the optical tracking system in terms of stability and accuracy. EM tracking systems are easily affected by electromagnetic (because of the surrounding instruments) and well as ferromagnetic (because of the metal surgical tools) interference. Also, the EM sensors have to be fabricated in a small size to fit them inside the body of laparoscopes, which further reduces the functional area of the system. Given that the laparoscopes are used inside the body of patients, the accuracy decreases critically when the laparoscope is in use. Even then, EM tracking systems remain the only valid option when the tip of the tool is to be tracked inside patient's body due to the fact that optical and mechanical systems can only track the outer parts of the rigid tools such as the handles of the endoscopes. Flexible tools, such as an ultrasound laparoscope, have tips that go inside the patient's body and only miniaturised EM sensors can be mounted on them. Detailed information of different tracking techniques can be found in [Bir00].

2.3.4 Registration

The pre-operative and intra-operative imaging data are expressed in their own respective coordinate systems. To fuse and display all the data together in a useful way, all the data must be brought together in a common frame of reference. Since tracking data is real-time task that governs the functioning of the visual aspect of image-guided surgery, the coordinate frame of tracking is generally considered to be the common frame of reference for all the multi-modal images. Various framing tools and markers are used to obtain the reference frame of each imaging modality such as stereotactic frames, fiducial markers, anatomical landmarks inside the body (e.g. points on bones) and also artificial landmarks created by surgeons (e.g. surgical ink to identify points on the skin or electrocauterized markings on the liver), etc. All of these markers around the patient's anatomy are tracked using certain devices, such as laser/optical cameras, that align their coordinate system to the tracking system.

Pre-operative imaging data obtained from CT, MRI, PET scans can be fused together with the real-time video data obtained from endoscopes using image-to-image registration techniques. There are several recently developed 2D to 2D, 2D to 3D image registration techniques [Mar+12]. Multimodal image registration techniques, such as fusion of a real-time ultrasound image with pre-operative CT/MRI images, are available commercially in ultrasound machines (e.g. S3000 series from Siemens Healthcare using eSieFusion technology).

The abdominal space is often deformable due to activities such as breathing. Organs inside this space are also deformed during a surgery. Registration of these organs with their pre-operative images becomes a complected and hard problem to solve, especially in real-time. New computer vision techniques such as Shape-From-Template [Jim+18] can address the problem of deformable shapes but the commercial solutions are yet to be seen. A comprehensive review of registration methods can be found in [OT14].

2.3.5 Visualization and Interaction

FIGURE 2.7: Left: A view of the surgical room from technicians area. Right: Surgeon's communication with the technician.

FIGURE 2.7 shows an MRI-based real-time image-guided surgery. The surgeon is presented with the MRI image which is constantly acquired from the MRI machine in real-time. The MRI machine is controlled by the technician in the control room and has a view of the operation area through a window. The surgeon directs him to control the

MRI image view using sign directions. The sign interaction is specifically used because the MRI machine makes a loud noise during its operation. Contrary to this, in a typical laparoscopic surgery, the endoscope is controlled by the lead surgeon's assistant who can also observe the screen and can be instructed verbally by the surgeon to adjust the viewpoint. Though this scenario is communication-wise less complicated than that with the MRI machine imaging, a comprehensive interaction between the surgeon and his technician is always required. The surgical team acquires these abilities by periodic rigorous training. Augmented-reality techniques aim to ease this interaction, which in turn, will reduce training hours for the surgical team.

In the case of 3D volumetric data, three standard methods are used: slice-based, surface-based and direct volume rendering. In the most commonly used slice-based method, orthogonal slice images of the patient (pre or intra operative) are displayed. These are called Axial, Sagittal and Coronal plane slices. If a series of consecutive slices is present, a 3D image can be constructed. With the tracking data acquired from the sensors mounted on the laparoscopic tools, the path of the tool can be rendered on the screen with the 3D volume image of patient's anatomy. In practice, surgeons choose to view 3D renderings in pre-operative surgical planning. During surgery, they prefer to see the different orthogonal slices on their screens in real-time. This practice also avoids the computational overhead of 3D real-time rendering and provides a clear 2D view of the necessary slices to the surgeon. Authors of [Tyn+15] present a visualization-based needle insertion surgery. FIGURE 2.8 shows the image on the display of the surgeon while the tool (needle) is observed in each frame during surgery. In this figure, the tool is not tracked by any sensors but only observed by the surgeon on the screen. Tracking the tool using sensors can provide additional control on the visualization such as automatically adjusting the viewpoint with respect to the tracked tool.

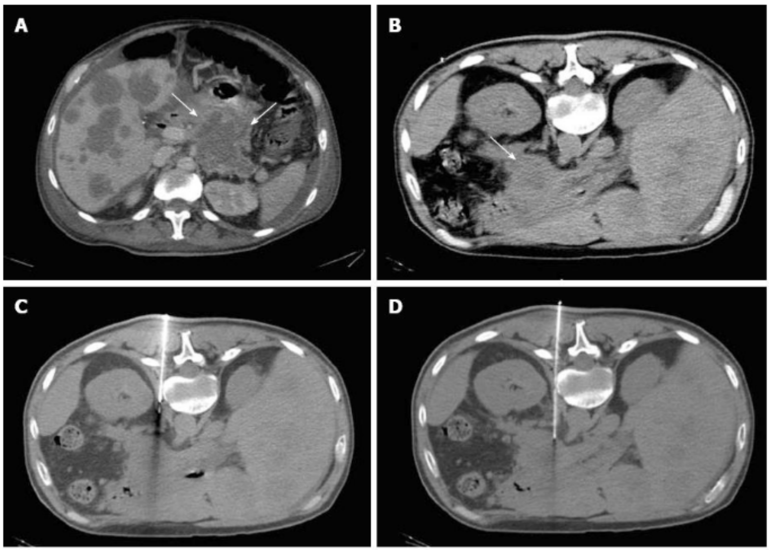

FIGURE 2.8: A: Possible path of needle insertion. Part of the pre-planning. B: The path chosen before the needle insertion. C and D: Real-time progress of needle insertion.(Images from [Tyn+15])

Interaction and user experience are the most overlooked parts of image-guided surgery and often the least discussed in the medical research community. A direct communication is not always efficient between the surgical staff during surgery. Sometimes, the surgeons themselves prefer to control the imaging. In these cases, sterlizable user control is required. Reviews of many of such interaction systems can be found in [YC06]. These include input devices like tool-embedded switches, tracked visual keyboards and also speech/gesture-based systems. The traditional visualization methods use multiple big LCD displays mounted in front of the surgeon. In recent times, some alternative visualization techniques have been used such as mini-LCD screens closer to the surgeons eyes, stereo microscopes, head mounted displays etc [YC06].

2.4 Laparoscope Tracking

In recent years, optical tracking and electromagnetic (EM) tracking solutions are widely adopted in the medical engineering field due to their acceptable accuracy and ease of use. In [Rei13], the author presents these tracking technologies in detail while presenting a hybrid solution of his own. Examples of the commercial products for optical tracking systems are Polaris and Optotrak from NDI(Northern Digital, Waterloo, Canada) and ARTtrack (Germany). In EM tracking, Aurora (NDI) and 3D Guidance medSAFE systems (Ascension Technology, USA) produce commercially available systems having miniature flexible sensors that can be easily be mounted on laparoscopes. Apart from these popular systems, there are other similar systems available that use optical or EM tracking (or sometimes in combination) such as StealthStation (Medtronic, USA), PercuNAV (Philips, Germany), eNLite(Stryker, Germany) and VectorVision(Brainlab, Germany). Though these products are application-specific and not universally adaptable like the ones mentioned previously.

In this section, we discuss the tracking technologies in detail along with a survey of the commercially available solutions. After presenting the technologies based upon optical and EM sensors, we present a recently proposed hybrid solution by [Feu+07] that combines optical and EM tracking together. We also present a novel technique presented in [Ogu+14], purely based on optical tracking, that can track the tool tip of the laparoscope using an endoscopic camera. In the end, we propose our solution to hybrid tracking based on optical and IMU (Inertial Measurement Units) sensors.

2.4.1 Optical Tracking System

An optical tracking system consists of two main sub-systems. The first part contains one or more cameras, fixed on a mounting assembly. This assembly is usually mounted on the ceiling of the operation room to maintain maximum visibility of the markers. The other part is the marker/fiducial subsystem. There are several types of markers that can be classified based on their illumination: visible light pattern, passive/reflective light, active infra-red light pattern.

The visible light patterns are the cheapest and easiest to use in terms of hardware. The markers are LEDs of certain colours observed by a video camera that can record the surgery while detecting the LEDs. The real-time LED detection algorithm is based upon contrast, colour, edges and/or flashing patterns of the LEDs. The flashing can be tuned to a high frequency if it is used with a high frame camera that allows for easier detection of markers. In this case, the markers can be distinguished from the surrounding light. A near-infrared light frequency can be used in combination with the high frequency flashing in order to avoid user discomfort. MicronTracker provides such a system. Though cheap and easy to build and setup, this technique is affected by the intensity of the surrounding light. This is especially true in the surgical room where there are strong sources of light causing multiple reflections.

Passive reflective light markers are fiducials made of a specific shape and are illuminated by either natural light or a nearby artificial light source. These are generally spherical in shape since the 2D captured image always remains a circle irrespective of the camera angle. The size of each fiducial is slightly different than the other in order to differentiate each of them. For all practical purposes, an infrared light source is built inside the camera assembly. The infrared light is reflected over the marker that maybe coated with a hyper-reflective chemical coating such as Scotchlite (3M, Germany). The infrared camera detects the reflected image and segments the circles with subpixel accuracy. The challenge is to detect at least 3 circles to determine the pose of the marker. This task is handled by a computer attached to the camera. This method of optical tracking is popular since the marker assembly need not contain any electronics and hence it is easier to clean. FIGURE 2.9 shows the Polaris products in use. FIGURE 2.9 shows two different products from Polaris and their use in surgical rooms. Polaris systems can work with both passive reflective as well as infrared active tracking.

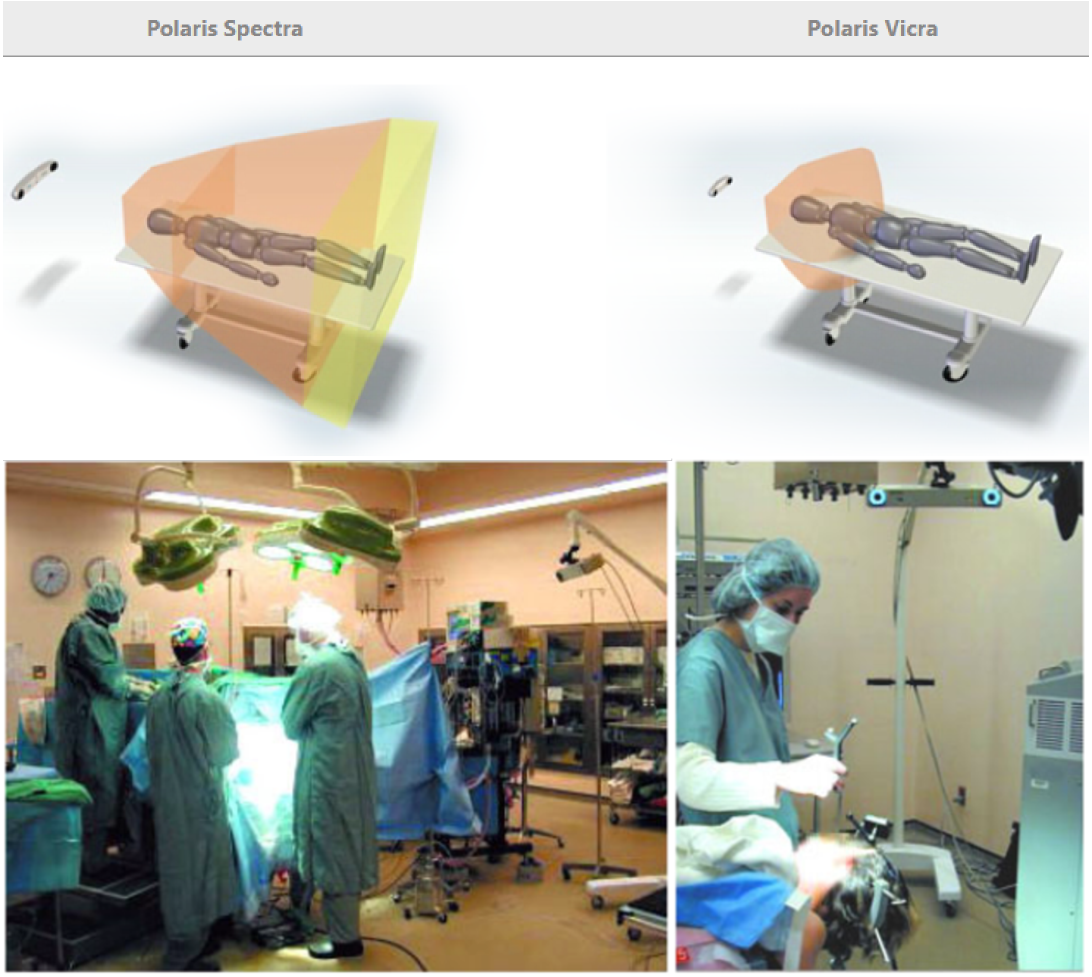

FIGURE 2.9: Top: Polaris products with the schematic of functioning area. Bottom: Placement of Polaris camera in a surgical room (Images from www.ndigital.com)

Active infra-red tracking uses infra-red light LEDs in the marker assembly. The electronics for the LEDs is either mounted on the marker assembly, in the case of a battery operated wireless system (Polaris), or an additional wire is attached to the tool up to its handle. Because of the active LEDs, no additional illumination is needed and a known pattern of LED flashing can be generated to detect and segment the LEDs efficiently.

FIGURE 2.10 shows passive markers mounted on a needle used in ablation procedures. Detection of minimum three points in a non-symmetric configuration is required to obtain a pose of the tool. This arrangement provides a unique relationship between the target pose and the obtained camera images. More than three markers introduce additional redundancy to the tracking system and this can be helpful in case of partial occlusion of the marker.

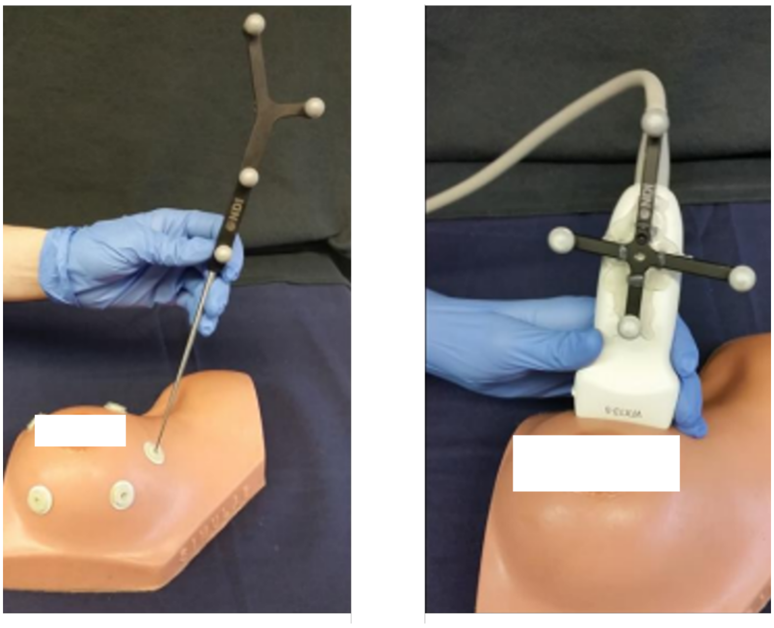

FIGURE 2.10: Mounting of LED markers on a needle tool (Images from www.vanderbilt.edu)

Two 2D cameras mounted on a rig are required to obtain the 3D pose of the marker. One image of the marker can be obtained from each camera and, using triangulation, the pose is estimated. Accuracy of the system increases if the fiducials in the markers are placed apart from each other as much as possible. Similarly the distance between two cameras should be kept as big as possible. In practice, manufacturers pay attention to the ergonomics of the system (by making it compact in size) with a slight trade-off in accuracy and range of the system.

Instead of using two 2D cameras, three 1D cameras can be used for tracking. These systems increase the accuracy up to 0.1mm and the frame rate up to 4600 Hz at the cost of being very bulky. Examples of such systems are Optotrak Certus (NDI) (FIGURE 2.11) and Flashpoint systems (Stryker, Germany). Laser-based tracking systems from laserBIRD2 (Ascension Technology) claim to be accurate up to 0.7mm.

FIGURE 2.11: Optotrack 1D tracking system. Markers for the system shown in top right. (Images from publish.illinois.edu and www.ndigital.com)

2.4.2 Electromagnetic Tracking

Electromagnetic tracking systems are widely popular in endoscope tracking applications([Hay+09], [CMS10]). Due to the small size and flexible shape of the EM sensors, they can be attached to any type of device with a little effort. EM tracking system is the only currently available solution for tracking an instrument like a catheter, which is flexible and goes inside the body. FIGURE 2.12, shows the schematic of Aurora (NDI) EM tracking system. It contains three main parts: an EM field generator, a sensor unit and a system control unit.

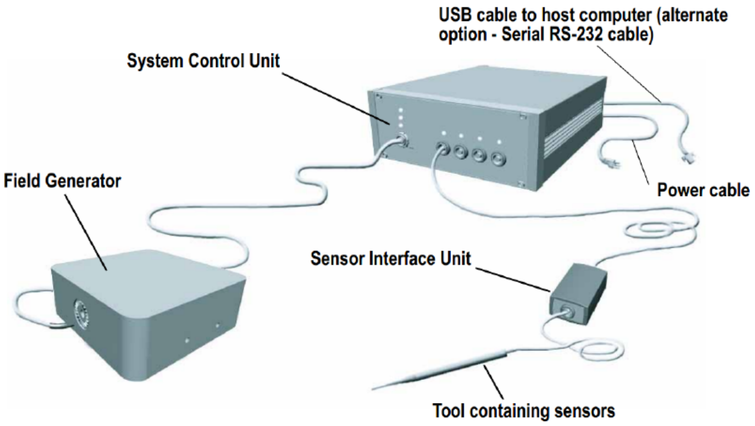

FIGURE 2.12: Schematic of electromagnetic tracking system. (Image from www.ndigital.com)

The EM field generator, also known as Transmitter, generates an electromagnetic field. The EM sensors measure the EM field strength or magnetic flux and, from multiple measurements, the position and orientation of each sensor relative to the transmitter is computed. The early designs of the transmitter had three inductive coils. Each coil was turned-on alternatively in three time phases to create an electromagnetic field which was detected by the coils in the sensors. The EM sensor measured the field strength one dimension at a time. In the early designs, There were three coils in total inside an EM sensor. Hence 3 × 3 measurements provided six degrees of freedom information in total ([Raa+79], [Kui80]). In recent years, six to nine coils are used in tetrahedral configuration in the electric field generator [Sei+00]. With this coil configuration, only one coil is required in the EM sensor to detect 5 degrees of freedom with only one reading (6 × 1 or 9 × 1). Although, the rotation around the sensor's axis cannot be determined in this configuration. Hence two non-parallel coils are combined in the EM sensor to obtain all 6 degrees of freedom. This configuration allows the sensors to be much smaller in size.

There are other configurations available in EM tracking sensors. ENsite NavX catheter tracking system (Endocardinal Solutions, USA) uses six patches with coils arranged closely to the patient's skin. Each pair of two opposite patches defines one of three approximately orthogonal axes. Information from all the pairs is then combined to obtain full pose of the sensor attached to the catheter.

Commercially available EM sensors can be found as small as 0.3mm (5-dof) and 0.9mm (6-dof) in size (FIGURE 2.13). This miniaturization allows the sensors to be integrated into a variety of probes and tools such as endoscopes, needles, catheters, ultrasound probes etc. With a single transmitter, multiple sensors can be tracked without interference from each other. Usually from 4 to 8 sensors can be used simultaneously. EM tracking does not require a line of sight like optical tracking. This allows EM tracking to be widely used in invasive procedures.

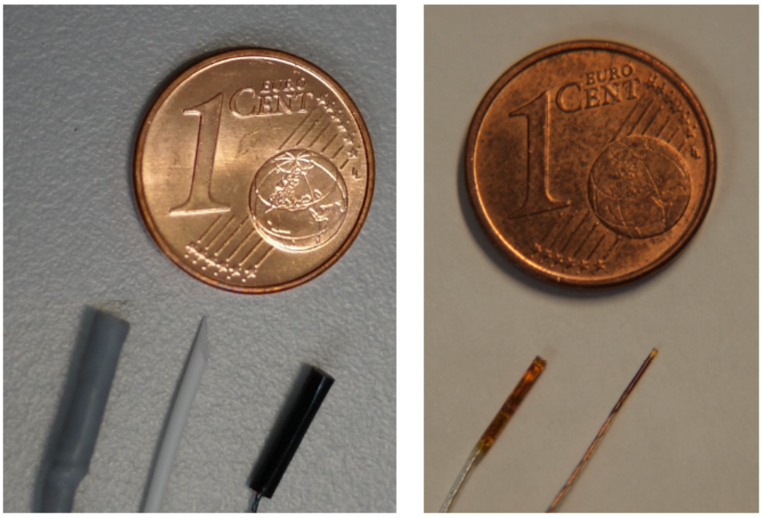

FIGURE 2.13: Left: Sensors from Northern Digital. Right: Sensors from Ascension Technology (Image from [Rei13])

Accuracy of EM tracking is lower in general compared to optical tracking. Typically EM tracking system provides an accuracy of 1.5 mm for position and 0.4° in orientation [Fra+03]. The range of EM tracking is also limited compared to optical tracking. The working volume of an EM system is generally not greater than $50 \times 50 \times 50$cm. (FIGURE 2.14).

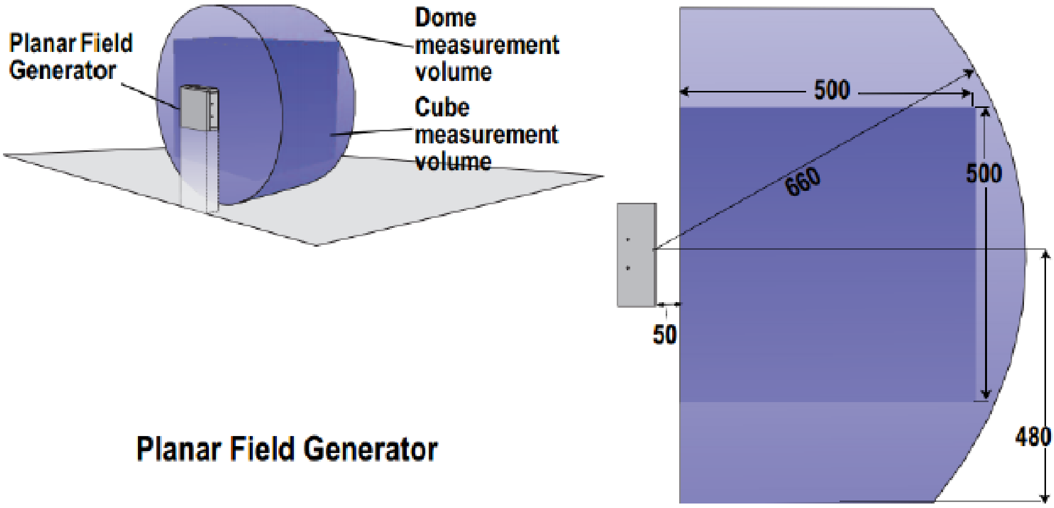

FIGURE 2.14: Active electromagnetic field of an electromagnetic transmitter in which the sensor can be detected. (Image from www.ndigital.com)

Additionally, the detection of the EM sensor in this volume is not uniform. The software associated with the EM tracking provides a quality number for the measurement, which is an indicator of the consistency of the EM sensor detection in the EM field. Based on this "quality" number, we can assess that the EM field strength is lower at the outer boundaries of the working volume. Furthermore, the field is quite non-uniform when the sensor is very close to the transmitter. 5 dof sensors are also susceptible to measurement errors around their rotation axis that is pointing directly towards the transmitter. Due to the range and accuracy limitations, EM sensors are best used when the transmitter is placed close to the body. There are different shapes of transmitters offered from the manufacturers based upon the needs of surgeries. Manufacturers also provide multiple EM field distribution options through the associated software. For example, NDI Aurora system can switch between fields tuned for either bigger working volumes or for better accuracy.

FIGURE 2.15: Different shapes of the EM transmitters. Table top transmitter lies on the surgical table while the box transmitter is usually mounted alongside the surgical table (Image from www.ndigital.com)

Another difficulty in the use of EM tracking is that the system is very sensitive to the distortions of the EM transmitter's EM field. Electrical machines as well as ferromagnetic/metallic objects can disturb the EM tracking field. Imaging devices like, C-arm, ultrasound machines, metallic surgical tools, operation table etc. can affect EM tracking with errors up to 7mm ([PA02], [Yan+09]). However, clinical studies show that the metal object interference was the least concerning factor during the actual use of the EM sensors [Wag+02]. In [Raa+79], the author presents a hypothesis about the effect of a metallic object on the EM tracking field. If the metallic object is about twice the distance (from EM transmitter) than the distance between the transmitter and sensor, then the effect of the object on tracking is less than one percent. In [Nix+98], authors present a comprehensive analysis of metal effects on EM tracking systems.

FIGURE 2.16 (left), shows the distortion of the EM tracking field caused by a Ferromagnetic material. Eddy currents are induced in a metallic object when it is placed in an alternating (AC) EM field. With this effect, the EM energy is partly converted into heat reducing the strength of the EM field. Metals with high electrical conductivity and low permeability affect the tracking system more than their 300 series stainless steel or titanium counterparts [LDM97]. Eddy currents only affect AC EM field systems like NDI Aurora. DC (Direct current) EM field systems like 3D Guidance medSAFE (Ascension Technology) employ a quasi-static EM field and compute steady state values to avoid the Eddy current effect. Although this system brings robustness against EM field distortions, it leads to more static interference by ferromagnetic materials.

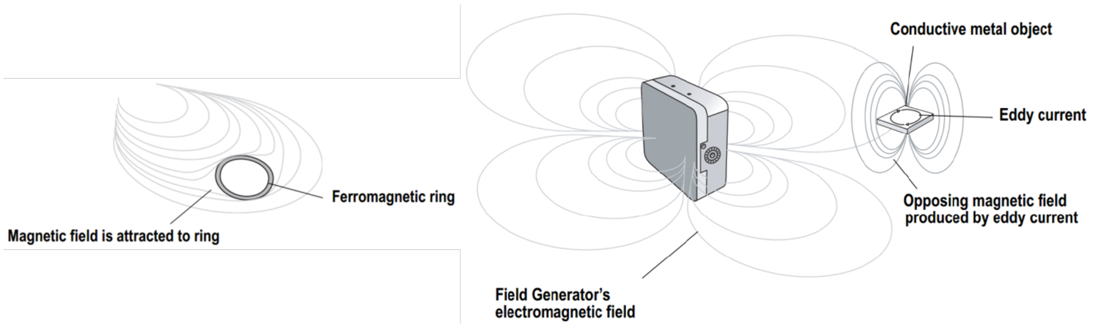

FIGURE 2.16: Left: Ferromagnetic material disturbance to the EM tracking field. Right: Eddy current induction effect in the metallic object in the EM field (Image from www.ndigital.com)

In recent years, many new techniques and designs have helped to improve the robustness of electromagnetic systems [KSB06]. Table top EM transmitters (FIGURE 2.15) are enabled with EM shielding made with high magnetic permeability metals along with aluminium plating [Rei13]. This strategy largely eliminates the effects of surrounding external EM fields while also keeping the effects of ferromagnetic materials low on the EM tracking field.

2.4.3 Mechanical Tracking

Mechanical systems were used in laparoscopic tracking before the contactless tracking technologies such as EM and optical tracking were available. Systems based on joint encoders and potentiometers such as FARO Gage (FARO, Germany) can track the tools with an unsurpassed accuracy of $25\mu m$. The tracking is achieved by detecting the pose via forward kinematics. The weight of this type of system is a big hurdle of these systems. FARO systems weigh up to 9.1 Kg. and can track only one object at a time. Also the mechanical arm connecting the tool to the base can be an obstruction to the surgical staff. Cleaning and sterilization of these complicated systems is also a challenge. Robotic surgical systems such as 'da Vinci' (Intuitive Surgical, USA) are the state-of-the-art of modern robotic surgical techniques. With these systems, remote surgical procedures are possible in which the surgeon operates the patient over a distance. Mechanical tool tracking technology in incorporated in the surgical robots to track the robotic tools.

2.4.4 Hybrid Tracking

Hybrid tracking technologies use two or more tracking sensor systems together. By combining the information from these sensors together, they aim to produce a robust pose estimation of the instrument. The idea behind the hybrid tracking systems is to compensate the drawbacks of one tracking systems with the others. There are no commercially available hybrid tracking systems in the medical instrumentation industry. Although in robotics and computer vision applications, hybrid tracking is readily used. In robotics, SLAM (simultaneous localization and mapping) is used to track a robot's pose in an unknown environment. The robot creates a map of its surrounding at the same time localizing itself inside it. The robot generally uses multiple cameras combined with odometry sensors for localization. Recently in the virtual reality gaming industry, products such as Oculus Rift S and HTC Valve Index use combinations of mounted cameras and IMU (inertial measurement unit) sensors mounted on the virtual reality headset to track them in an indoor environment.

2.4.5 Ultrasound Laparoscope Tracking

Ultrasound (US) laparoscopes have some unique problems associated with tracking of their tips. The tip of an ultrasound laparoscope has an ultrasound array instead of a camera. Hence one cannot use the image from the laparoscope to track the tip using computer vision tracking techniques. The tip of the laparoscope can be moved using the control levers on the handle of the laparoscope. Compared to Endoscopes, in which the entire instrument is rigid and the tip's pose can be estimated by tracking the handle only, ultrasound laparoscopes require an additional sensor mounted on the tip to obtain the pose. Hence two parts of the US laparoscopes need to be tracked separately, the handle and the tip, while also considering the structure of the laparoscope that connects the handle and the tip. There are two notable techniques for tracking the tip of US laparoscopes. The first one from Oguma et al. [Ogu+14] uses only optical cameras while the second from Feuerstein et al. [Feu+07] uses camera-EM sensor hybrid tracking.

Ultrasound laparoscope tracking using endoscope images

This technique, presented by Oguma et al. [Ogu+14], uses and external optical camera system to track the handle of the laparoscope. The authors make assumptions that while using the US laparoscope, an endoscope will be used in the surgery at the same time and tip of the laparoscope will always be visible in the images from the endoscope. FIGURE 2.17 shows an overview of the system.

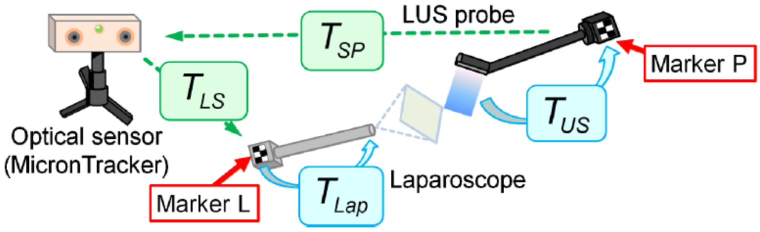

FIGURE 2.17: Overview of the endoscope image based optical tracking system (Image from [Ogu+14])

The endoscope as well as laparoscope handles are detected by the optical ’MicronTracker’ system. Detecting optical markers L and P placed on the handles provides transformations T_{LS} and T_{SP}. T_{Lap} and T_{US} are the fixed body transformations of the endoscope and laparoscope respectively. T_{Lap} is a fixed calibration and it is obtained by placing an optical marker at the tip of the laparoscope as well as on the handle and taking few images from the camera. T_{US} has two fixed components connected with the changing angle of the tip of the laparopscope. The transformation from the handle to the beginning of the tip can be obtained similarly to T_{Lap}. The US image to the tip transformation is obtained by using a cross-wire phantom technique in which several known cross sections points of wires are detected in the ultrasound image providing the transformation between the US array and the origin of the image. After knowing these two fixed transformations, the authors fix the US laparoscope on a test bench and bend the tip of the laparoscope 1000 times while recording marker positions on the tip and on the handle. This set of reading acts as a look-up table for the tip angle.

The US laparoscope tip with a ’yellow’ marker is detected in the endoscope camera providing transformation T_I. A distinct pattern on the yellow marker also allows to detect an angle of the tip rotation in the endoscope image. Transformations T_I, T_{Lap}, T_{LS}, T_{SP} and T_{US} complete the chain of transformations based on which the tip of the US laparoscope can be detected. Authors also test their device in an animal experiment. FIGURE 2.18 shows the US image overlay on the endoscopic image.

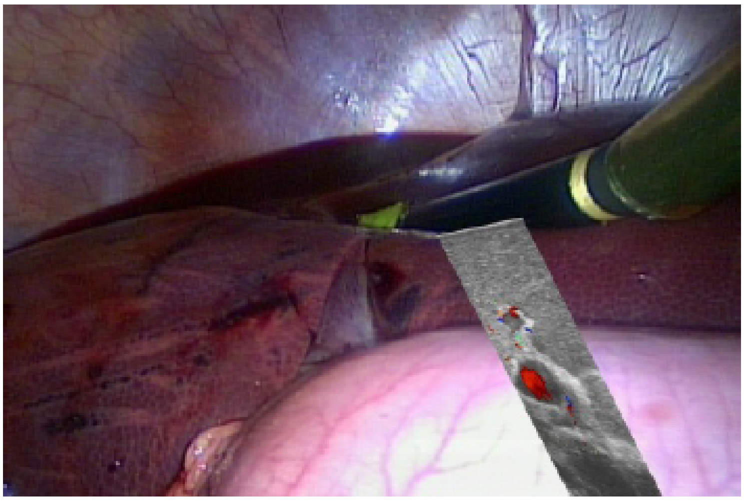

FIGURE 2.18: Ultrasound image overlay on the endoscope optical image (Image from [Ogu+14])

This system heavily relies upon the quality of the endoscope images to detect and track the tip of the laparoscope. The marker used for US laparoscope is very basic and small in size. Detecting the pattern on this marker to compute the roll of the tip in an endoscope image is a hard task given that the image is constantly obstructed by body fluids, smoke and bright lighting. The tip of the US laparoscope is likely to be hidden around the organs making the constant detection of the marker impossible.

Optical-EM sensor-based ultrasound laparoscope tracking

Feuerstein et al. [Feu+07] propose a hybrid tracking system that uses EM sensors as well as optical camera sensors. This system is not dependant upon the endoscope images as the previous system. FIGURE 2.19 shows the schematic of this hybrid tracking system.

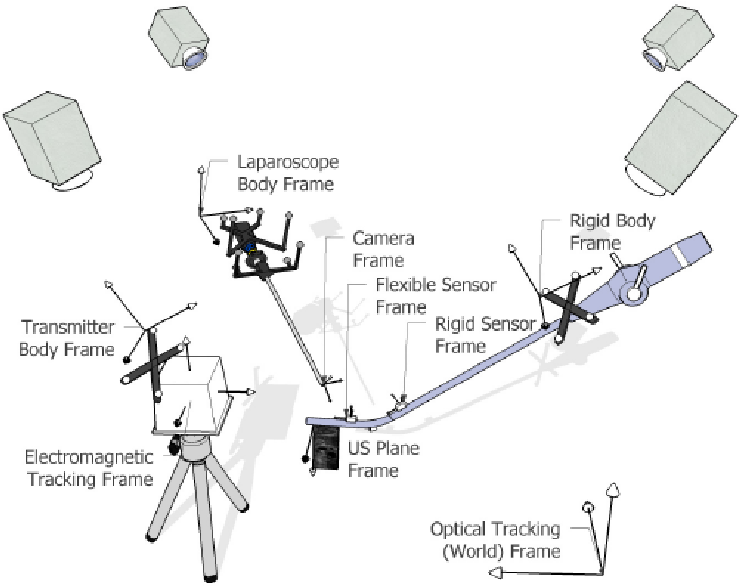

FIGURE 2.19: Electromagnetic sensors and camera based hybrid tracking system for US laparoscopes (Image from [Feu+07])

In this design, NDI Polaris tracking system is used to track different instruments in the scene. The passive markers for the tracking are attached to the handles of the laparoscopes present in the scene. Apart from the tracked instruments, the transformation of the EM transmitter is also tracked with respect to the optical system. Several miniaturised 6 dof EM sensors are mounted on the body of the US laparoscope. NDI Aurora system's transmitter tracks these sensors in real-time. The sensors provide the transformations of the laparoscope's structure that is going to be hidden inside the patient's body during the operation and hence it cannot be tracked using the optical tracking system. The transformation between the EM sensors and the optical marker mounted on the US laparoscope handle is evaluated using the 'Hand-Eye Calibration' technique that computes the fixed body transformation between two rigidly attached pose sensors. In this system, we know all the necessary transformations to complete the kinematic chain with respect to the optical tracking system. As discussed earlier, the EM tracking system suffers from errors when the sensors are small and used at a farther distance from the transmitter.

Although this hybrid tracking system aims to compensate this drawback using the optical system, the accuracy of the tip tracking entirely depends upon the EM tracking sensor mounted on the tip. This means that the tip tracking accuracy is still affected my EM disturbances mentioned earlier. A plastic mount is used to accommodate the EM sensor on the tip of the laparoscope. It increases the width of the laparoscope tip, forcing one to use larger trocar for the US laparoscope.

2.5 Summary and discussion

In this chapter, we presented a brief history of development of image-guided surgery. We provided a surgeon's point of view towards the laparoscopic surgery that gives a context to the laparoscope tracking techniques. We presented different laparoscope tracking techniques using commercially available sensor systems. Given the fact that each sensor system has its advantages and limitations, one can think of an ideal laparoscope tracking system that can eliminate all the drawbacks of presently available ultrasound laparoscope tracking systems. An ideal system for laparoscope tracking must use sensors that are small in size so that they can be fabricated inside the laparoscope tip. The sensors should not depend upon external transmitters to provide the transformations. The sensors should be free from the effects of external disturbances such as occlusion, EM interference, lighting etc. The sensor system should be easy to clean.

Optical tracking remains ideal for tracking the external parts of laparoscopes. The recent advances in image processing and computer vision can reduce the limitations of these systems. IMU (Inertial Measurement Units) can replace the EM tracking sensors since they do not depend upon any external transmitter system to compute the transformation. The IMUs can be fabricated in very small sizes so that they can be incorporated inside the laparoscope body. In the next chapter 3, we propose a solution which is based upon optical sensors and IMUs to track a US laparoscope. Although no system is perfect, we try to answer some of the limitations of the current tracking system with this novel approach.

Chapter 3

Laparoscope positioning based on IMU and PSD Camera

In this chapter, we present a design that allows tracking an ultrasound image in real-time from an ultrasound laparoscope. We track the tip of the laparoscope by using IMUs (Inertial Measurement Units) as well as PSD (Position Sensitive Detectors) camera sensors. First, we explain the hardware of the laparoscope along with a brief review of the tracking systems employing such hardware. Then, we present the design of the laparoscope including the schematics and kinematic chain required to connect all the sensors together. In the end, we provide a general strategy to remove noise from the IMU tracking system using rotation averaging.

3.1 Overview of the Ultrasound Laparoscope

Before we present the details of the US laparoscope tracking system, we first define the main objectives of this system. The considered ultrasound laparoscope has an ultrasound array at its tip. The tip is movable and the surgeon operating the laparoscope can change the direction of the tip using the control levers. These levers are mounted on the tip of the laparoscope. Usually there are two levers to control the movement of the laparoscope tip. Using them, the laparoscope tip can be moved in a hemispherical space in the patient's body (FIGURE 3.1).

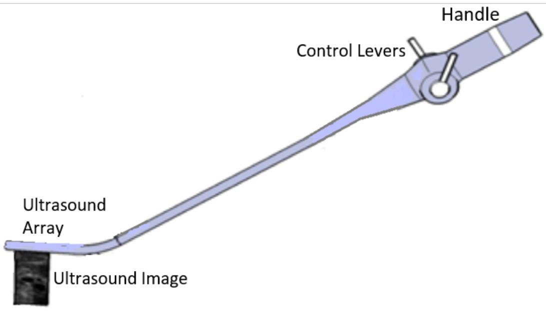

FIGURE 3.1: A standard design of an ultrasound laparoscope without any sensors mounted

Along with the rotation of the tip, the surgeon can also move the entire laparoscope by itself to change the angle of the tip. The ultrasound array mounted on the tip emits an ultrasound signal in a direction that is perpendicular to the non-rotating axis of the tip. This signal is emitted in a plane, usually referred to as ultrasound fan. The ultrasound waves travel in the tissue of the patient and reflect back to the tip. The ultrasound machine computes the time of flight of the reflected signal and an image is produced after signal processing. The image shown on the screen is that of a slice made by the ultrasound plane onto the tissue. The surgeon assumes that the laparoscope tip is at the top of the image. Since the laparoscope is inside the body of the patient, the surgeon has no visual confirmation of the location of the tip with respect to the patient's body (anatomy). Hence the objective is to provide this visual confirmation to the surgeon in which we have to display the tip (and the resulting ultrasound image) on the screen moving with respect to a stationary point in the operation room.

To achieve this objective, we need to track the tip of the ultrasound with some sensors. As we presented in the second chapter, the commercially available solutions for ultrasound probe tracking use electromagnetic (EM) sensors. EM tracking systems come with the drawback of interference and limited range. To counter these problems, we propose to use IMU sensors to track the motions of the laparoscope. The IMU sensors are to be mounted on the body of the laparoscope. Since the circuitry of the IMU sensors is wired through the laparoscope, there is no range limitation issue. IMUs can provide a good quality information about the rotation of the laparoscope. To detect the 3D position of the laparoscope, we use a PSD camera-based system developed by Siemens Healthcare. The PSD sensors track an infrared (IR) LED ring mounted on top of the laparoscope. Combining the rotation information from IMUs and position information from PSDs, we can obtain a full 6-degree-of-freedom (dof) pose of the tip of the laparoscope. There are certain limitations that may affect the performance of this system. The IMUs are affected by noise and drift. The PSDs could be affected by the lighting effects and occlusions. In this chapter, we use rotation averaging to deal with the IMU noise and drift. Since the IR-LEDs emit infrared light and PSDs are tuned to detect the specific light frequency of the infrared light, the lighting problems are mostly reduced. Occlusion problems can be dealt with several computer vision tracking algorithms (this is not covered in this thesis).

3.2 Hardware Setup

In this section we provide a brief background on the sensors systems involved in the US laparoscope tracking.

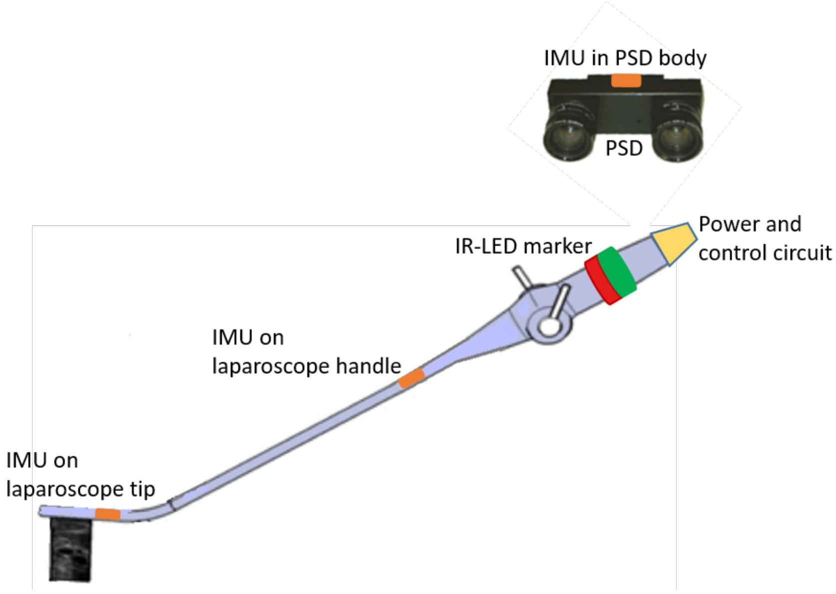

FIGURE 3.2: Schematic of the US laparoscope tracking setup

In this section, we briefly present the hardware modules for the laparoscope tracking. These hardware modules are designed and created by Siemens Healthcare. FIGURE 3.2 shows the schematic of the tracking sensor setup for a US laparoscope. We mount miniaturised IMUs on the body of the laparoscope. We use 4 IMUs in total. Although only two IMUs are required to track the body of the laparoscope, additional two are for redundancy which helps in the case of a failure. All of the IMU circuits are wired together to a control circuit mounted in the handle of the laparoscope. The handle of the laparoscope also hosts the LED ring. The LED system as well as the IMU circuits are powered by a separate power supply connected to the laparoscope.

The PSD camera system consists of a two-PSD rig. The two PSDs are separated by some distance between them. We assume that the angle and distance between the PSDs are known by the manufacturing design. The PSD camera system also hosts an IMU. FIGURE 3.6 shows such a system in operation. The PSD frame capture and the IMU data capture are synchronised by a clock circuit. The synchronised data is then fed to a computer or to a Siemens S3000 ultrasound machine. The machine has a software, named eSie Fusion, that can receive the 6 dof pose in an '.xml' file format and then incorporates it to the captured ultrasound images to show a real-time tracking view of the ultrasound images with respect to pre-acquired CT/MRI images. The semi-automatic calibration between the CT and ultrasound images is also performed in the eSie Fusion software.

3.2.1 Inertial Measurement Units

Initially, the use of IMUs began in the aviation industry for aircraft navigation [ZW11]. Earlier IMU systems were mechanical and the use was restricted to heavy applications. In recent years, micro-electromechanical systems (MEMS) have enabled the production of miniature, inexpensive IMUs that are compact and consume little power.

Several manufacturers produce IMUs such as Bosch, Honeywell, X-Sens etc. IMUs are widely used to determine the movements in terms of acceleration, angular velocity and rotation [ZZ04].

IMUs usually contain three types of sensor units: an accelerometer, a gyroscope and a magnetometer. The accelerometer measures the inertial acceleration. The gyroscope measures the angular rotation. The magnetometer measures the bearing magnetic direction with respect to the Earth's magnetic field. The magnetometer data is usually combined with the gyroscope data in order to improve the accuracy of the output. The gyroscope suffers from drift errors that are known to accumulate over time producing larger and larger errors as time passes by. Magnetometers can fix this problem by correcting the gyroscope data periodically. Although magnetometers prevent the drift, they are affected by electromagnetic noise. Hence proper shielding of IMU circuits is required [ZZ04]. FIGURE 3.3 shows the possible outputs of an IMU sensor system.

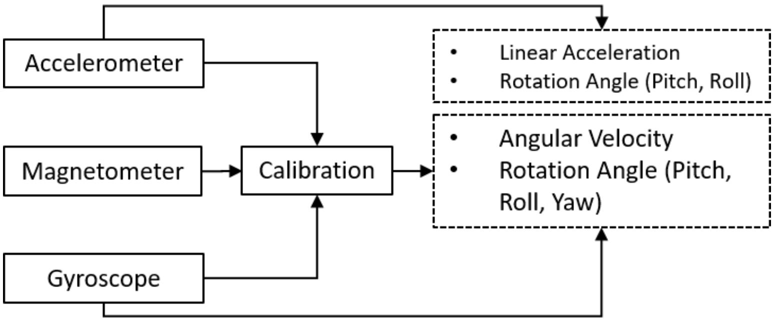

FIGURE 3.3: Outputs of an IMU sensor system

An IMU sensor can be chosen based on following criteria:

- Degrees of freedom: degrees of freedom (dof) determine the number of independent parameters in a system. The dof of an IMU depends upon the type of sensors included in the IMU and the number of axes those sensors will measure in. Most recent IMUs can track from 6 to 9 dof. The accuracy and also the cost of an IMU increases with the increase in dof.

- Data Accuracy: accelerometer accuracy is affected when there is a very slow or extremely sudden acceleration [ZW11]. As mentioned before, gyroscopes and magnetometers are affected by drift and EM noise respectively.

- Rate of response: modern IMU sensors can acquire data at a rate between 50 and 200 Hz range. The faster the data is acquired, the more the resolution we can obtain over a motion. Faster data acquisition also leads to acquiring more data frames. Together, these can be used to obtain more accurate measurements.

- Package size: recent developments in MEMS technology allowed reducing the size of an IMU chip to few millimeters in dimension. To mount an IMU on the laparoscope ultrasound array, we must use a very small IMU chip. At the same time, the circuit around the IMU chip, that ensures communication as well as provides power, must be miniaturized to a smallest possible extent.

For this application, Siemens Healthcare has provided BMI series IMUs from Bosch. The IMU communicates with the computer using I2C communication protocol. The data received in the computer is in the form of 4-vector unit quaternion. The communication and power circuits for the IMUs are designed by Siemens Healthcare.

3.2.2 Position Sensitive Detectors

A popular method for detecting a position of an object is to use a video camera that captures several images of an object. Using image processing techniques, we can detect and segment the object of interest from the background. And then, using 3D view geometry, we can find the position of the object with respect to the camera or in some fixed reference frame. Contrarily to a regular optical camera, position sensitive detectors detect the location of the light source in a plane. They can be tuned to detect a particular frequency of a light source. Compared to a standard camera, the detection of the light is faster and more accurate due to minimal image processing requirements. The cost of a PSD system can be considerably lower than professional grade video camera tracking systems such as NDI Polaris Spectra.

PSD is a precision semiconductor optical sensor that produces a current output related to the center of mass of the light incident to the surface of the device. A two-axis PSD (used in our application), contains a large area silicon p-i-n junction photodiode. It provides continuous position measurements of the incident light spot in 2D. FIGURE 3.4 shows the design of a PSD.

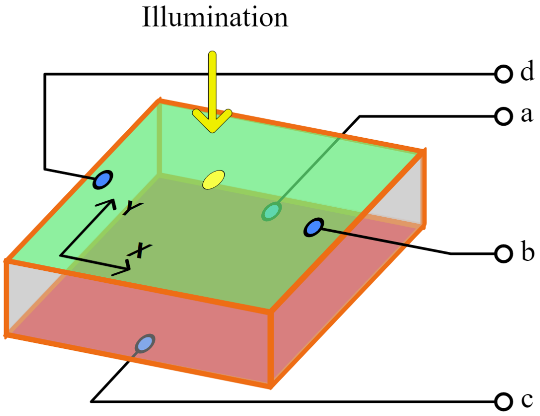

FIGURE 3.4: Design of PSD using a PIN diode. (Image from Georg Wiora)

The PIN diode is a laminar semiconductor and its operation is based on the lateral photoeffect. This phenomenon was first described by [Wal57b]. Later on, several papers discussed the property of non-uniform irradiation of the PIN diode ([Luc60], [Emm67]). In general, when the diode junction is exposed to radiation energy such as light, which is distributed non-uniformly, current is generated and it diffuses across the illuminated surface. The current diffuses through the bulk of the resistance of the material producing a lateral electric field that is measured by the four electrodes attached to the device. From the currents I_a, I_b, I_c and I_d, the location of the light is computed using the following equations

$$\begin{aligned} x &= k_x.\frac{I_b - I_d}{I_b + I_d} \\ y &= k_y.\frac{I_a - I_c}{I_a + I_c}, \end{aligned} \tag{3.1}$$

where k_x and k_y are scaling factors [Wal57a]. Contrary to a traditional video camera, which provides discontinuous measurement of light in the form of picture frames, PSD can provide a continuous measurement of a light spot with measuring rates up to 100 kHz. When the light spot lies at the center of a PSD, the relationship between the measured position and the actual position is approximately linear. When the light spot lies away from the PSD center, the relationship becomes non-linear. There is a new set of formulae proposed by [CS10] that help reducing the non-linearity in 2D PSD. FIGURE 3.5 shows the position response of a 2D PSD after computing the position by using the formulae from [CS10].

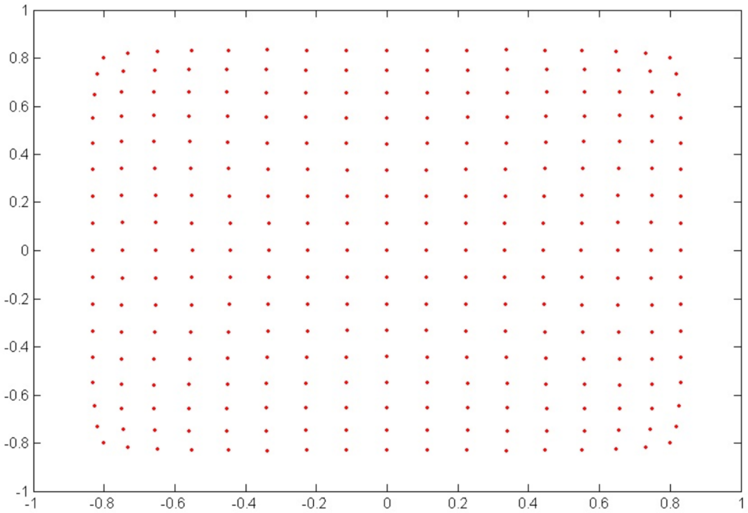

FIGURE 3.5: Position response of a 2D PSD sensor. (Image from [CS10])

In this figure, testing of the PSD is performed by moving a light spot in x and y axes of the plane in equidistant steps. A regular perfect grid pattern should be obtained if the position estimate is linear in all the parts of the sensor. Although the performance is much improved by using the formulae from Cui and Soh [CS10], some non-linear distortion remains.

In our application, we use two PSDs mounted on a camera rig separated by a fixed distance. The PSD system is designed and calibrated by Siemens Healthcare. Each of these PSDs are separately calibrated using a specially made LED grid that uses the same LEDs mounted on the LED ring assembly of the laparoscope. These are very bright infrared LEDs and each LED consumes up to 1W, what is a significant amount of power (in the prototype setup). Given the strong illumination created by each LED, a bright spot is created on the PSDs improving the strength of the current signal and hence the accuracy. The LED grid for the calibration is made by arranging several LEDs in a rectangular 2D grid pattern in which the LEDs are equidistant from each other. The readings from the PSDs are recorded on a computer where the design of the LED grid is also present. A software on the computer compares the ideal grid design to the PSD measurements and corrects the non-linear distortion of the PSDs.

3.2.3 Triangulation using PSDs

Triangulation techniques are used to estimate the 3D position of points with respect to the center of the PSD camera assembly. The LED ring assembly mounted on the laparoscope handle shows only one LED to the PSDs. Other LEDs on the ring are also visible but their brightness is far less compared to the LED pointing right at the PSD camera assembly. Since our PSDs are calibrated, we assume that the projection of the LED light on the PSD sensor is uniform in all the areas of diode. Once the PSDs detect the same 3D point in two PSDs, we can obtain x and y coordinates from both PSDs, corresponding to a single 3D point on the LED ring. Since we know the relative transformation between two PSDs by design, we can project two 3D lines from each 2D point on the two PSDs and find their intersection. In the absence of noise and distortion, this method should work. A suggested approach for this purpose is the 'mid-point' method [BZM94] that handles the case of noise when the lines may not intersect at all. The mid-point method estimates the intersection of the lines at the mid-point of the shortest segment connecting the two lines in the 3D space. Given the accuracy of our PSD devices in the given operational area, this algorithm is sufficient for the 3D point estimation of the LED on the ring assembly. However, we suggest to apply better algorithms for triangulation presented in [HS97]. This paper provides a non-iterative global minimum solution by formulating the triangulation problem as a least-squares minimization problem. The paper assumes a Gaussian noise model for the perturbation of the image coordinates.

3.3 New Laparoscope Design

In this section, we present the design of the new laparoscope with the IMUs and PSDs. FIGURE 3.6 shows the PSD camera in real use.

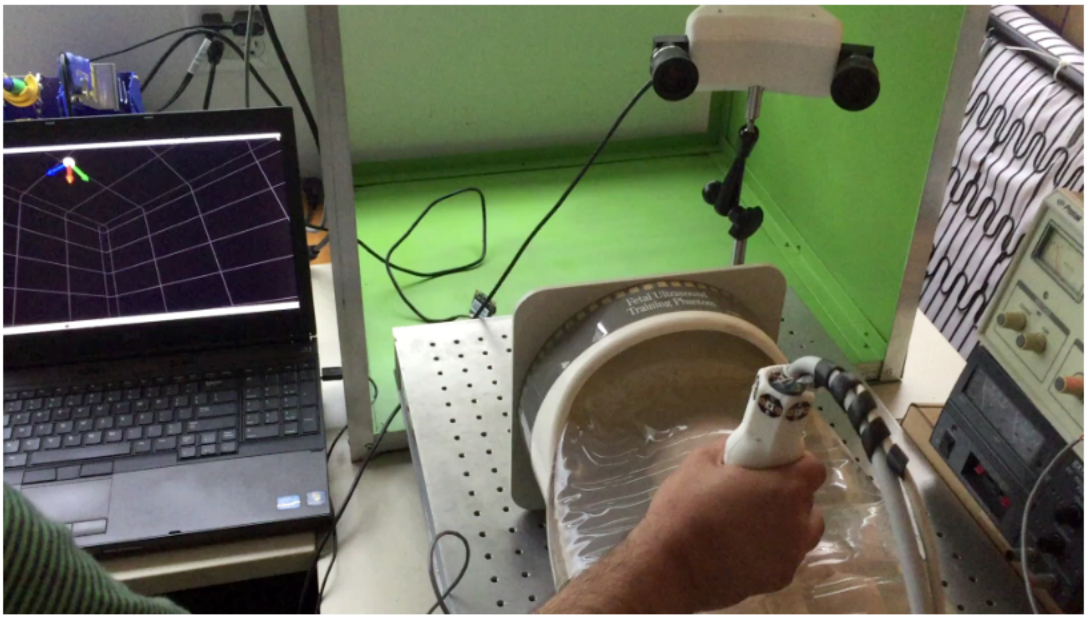

FIGURE 3.6: PSD camera pointing at the LED ring mounted on a US probe handle. (Image from Siemens Healthcare)

The PSD camera rig is mounted nearby on a stand and it is pointing at the infrared LED ring that is attached to the handle of and US probe. The PC display in front of the user shows the tracking of the LED ring mounted on the US probe handle.

In a real operation room, the PSD camera is to be mounted in an ideal place such that the camera has minimal occlusion to the scene. From our experience after attending surgeries, we noticed that, during a laparoscopic surgery, the surgical display monitors are always in front of the surgeon. Hence we can mount the PSD camera on top of such display before surgery. Since the display does not move during a surgical procedure, we can fix the reference coordinate origin of the laparoscope tracking system at the PSD camera. FIGURE 3.7 shows the schematic of this arrangement.

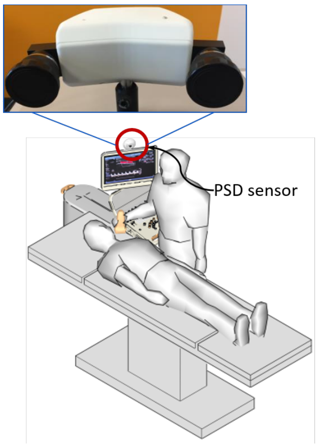

FIGURE 3.7: Ideal position for mounting the PSD camera rig

The PSD camera tracks an infrared LED marker mounted on the handle of the laparoscope. The infrared LEDs are invisible to the human eye and hence they can deliver a very bright LED flash without disturbing the surgeon. The LEDs are mounted on a specially designed plastic ring assembly so that, irrespective of the rotation of the laparoscope, the LEDs are always visible to the PSD camera. FIGURE 3.8 shows the design of the LED ring assembly. In FIGURE 3.6, a similar prototype assembly is mounted on the US probe handle.

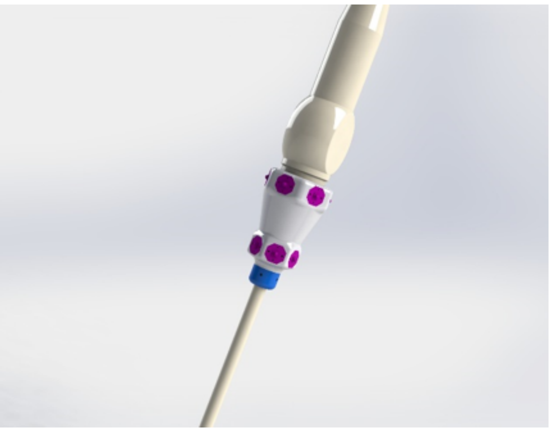

FIGURE 3.8: Design of the LED ring assembly mounted on the laparoscope handle

3.3.1 Kinematic Chain of Sensor Transformations

FIGURE 3.9 shows the schematic of PSD camera and laparoscope system.

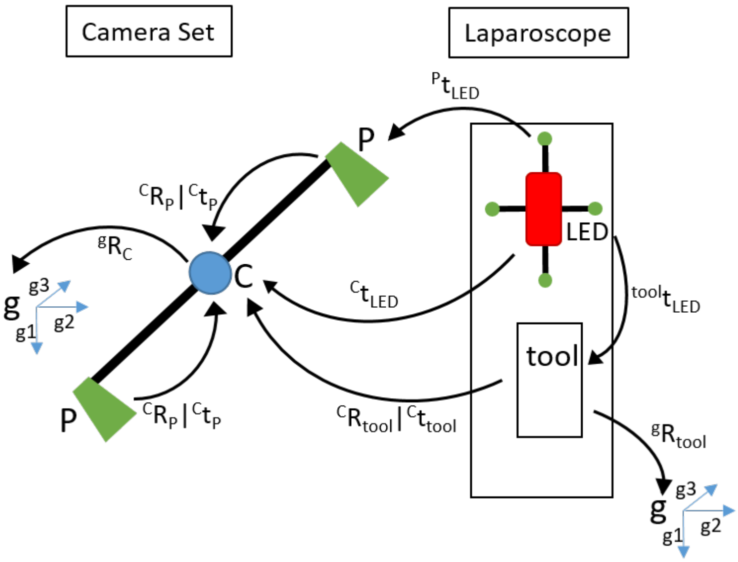

FIGURE 3.9: Schematic of the laparoscope handle and the PSD camera

The laparoscope contains a LED setup mounted in the handle. The IMU is preferably mounted closer to the LED ring assembly. In the PSD camera rig, there are two PSDs mounted on a platform. The center of the platform contains an IMU. Though the IMUs can provide a rough position information by integrating the acceleration data, the accuracy of the computation is not enough for our application. Therefore, we only use rotation information from the IMU setup. Similarly, PSDs only provide the position information and not the orientation. Hence, in our tracking setup, we completely rely upon the PSDs for the position (translation) information and upon IMUs for the rotation information. As seen in FIGURE 3.9, ${}^{g}R_{C}$ and ${}^{g}R_{tool}$ are the orientations of the PSD-camera-rig IMU and of laparoscope handle IMU (*tool*) with respect to gravity $g = [g1, g2, g3]$. ${}^{C}R_{p}|{}^{C}t_{p}$ is the transformation from P (the PSD) to C (center of the PSD camera rig). The IMU on the rig and PSDs are housed by the same sensor assembly of the rig and rigidly fixed. As a consequence, the fixed transformation between them is known by the fabrication design. ${}^{P}t_{LED}$ is the translation from the LED ring assembly center (in the figure's context, *LED*) to each of the PSDs. Similarly, ${}^{C}t_{LED}$ is the translation from the *LED* to *C*. ${}^{tool}t_{LED}$ is the translation from *LED* to the *tool*. We aim to find ${}^{C}R_{tool}|{}^{C}t_{tool}$ by solving the following system of equations. We have

$$ {}^{C}t_{LED} = {}^{C}R_{P}\,{}^{P}t_{LED} + {}^{C}t_{P}, \tag{3.2}$$

where ${}^{C}R_{P}$ is known from the design of the PSD camera rig. It is the angle between the two PSDs. ${}^{C}t_{P}$ is the distance from P to C. ${}^{P}t_{LED}$ is estimated via triangulation. Then we have

$$ {}^{C}R_{tool} = {}^{g}R_{C}^{-1}\,{}^{g}R_{tool}. \tag{3.3}$$

While ${}^{g}R_{C}^{-1}$ is fixed, ${}^{g}R_{tool}$ constantly changes with the movement of the laparoscope. Hence, we can compute

$$ {}^{C}t_{LED} = {}^{C}R_{tool}\,{}^{tool}t_{LED} + {}^{C}t_{tool}. \tag{3.4}$$

Using (3.2) and (3.4), we get

$$\begin{aligned} {}^{C}t_{tool} &= {}^{C}t_{LED} - {}^{C}R_{tool}\,{}^{tool}t_{LED} \\ {}^{C}t_{tool} &= {}^{C}R_{P}\,{}^{P}t_{LED} + {}^{C}t_{P} - {}^{C}R_{tool}\,{}^{tool}t_{LED}. \end{aligned} \tag{3.5}$$

Equations (3.3) and (3.5) together provide ${}^{C}R_{tool}|{}^{C}t_{tool}$. In (3.5), we can obtain a good estimate over the value of ${}^{C}R_{P}$ and ${}^{C}t_{P}$ since we get the value from the design itself. Similarly value of ${}^{tool}t_{LED}$ is estimated from the structure design of laparoscope. ${}^{P}t_{LED}$ is also mostly stable but marginally influenced by non-linear pixel noise especially if it is at the edges of the PSD sensor. This can be fixed using a better design of LED ring assembly in which more than one LEDs are visible to the PSDs at any moment (as shown in FIGURE 3.8). Orientation ${}^{C}R_{tool}$ depends upon the rotation data from the IMU. Since IMU is affected by drift and noise, single rotation averaging technique can provide a better estimate over this value.

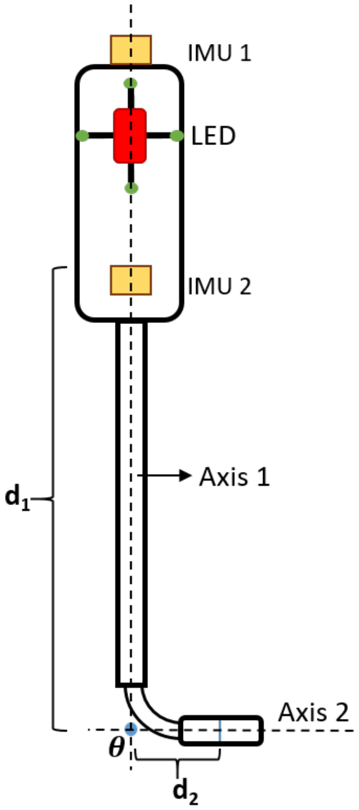

FIGURE 3.10: Schematic of the laparoscope

FIGURE 3.10 shows the schematic of the entire laparoscope. There are two IMUs mounted in the handle of the laparoscope. We only require one IMU to acquire the rotation of Axis 1. The second IMU is for redundancy purposes. If one IMU fails to work, the second IMU keeps providing the data. At the same time, we use the data from IMU 1 and IMU 2 by taking the average of them, which helps to reduce the effect of noise. In (3.5), we see that the ${}^{C}R_{tool}$ is the component related to the rotation of the laparoscope. Since all the IMUs provide rotation with respect to the magnetic rotation reference frame of the Earth (ground), we can only use the IMU at the tip of the laparoscope to obtain the final rotation of the tip instead of IMU 2 (see FIGURE 3.11).

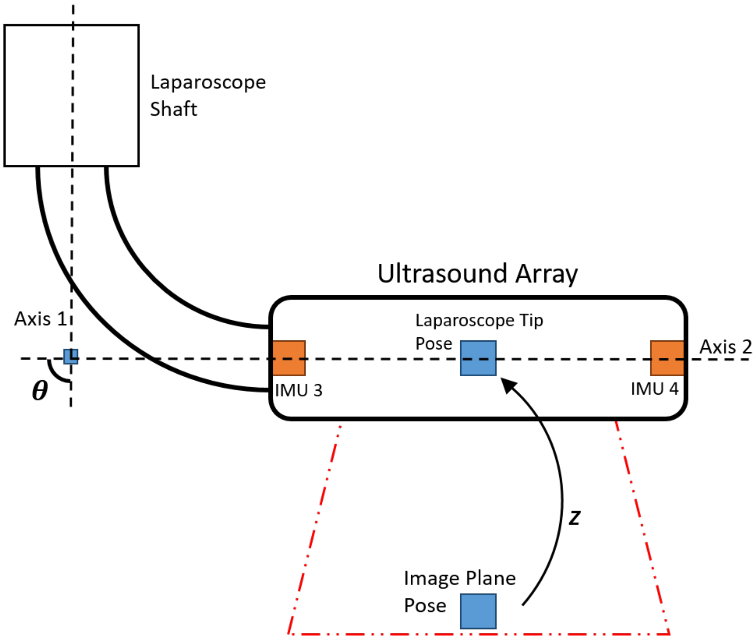

FIGURE 3.11: Schematic of the laparoscope tip

In other words, ${}^{C}R_{tool}$ can be obtained from IMU 3. In contrast, to obtain ${}^{tool}t_{LED}$, we must take into account the geometric structure of the laparoscope. IMU 2 and IMU 3 are connected via two axes, Axis 1 and Axis 2. Both axes intersect in an angle θ. Hence, we obtain the following transformation that estimates IMU 3 in the reference frame of IMU 2. We have a translation with distance $d1$ in the direction of Axis 1 and $d2$ in the direction of Axis 2. Angle R_θ represents a relative rotation between IMU 2 and IMU 3. We have the following kinematic chain between the handle and the tip of the laparoscope (FIGURE 3.10).

$$\begin{bmatrix} I & \begin{bmatrix} 0 \\ d_1 \\ 0 \end{bmatrix} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} R_\theta & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} I & \begin{bmatrix} d_2 \\ 0 \\ 0 \end{bmatrix} \\ 0 & 1 \end{bmatrix} . \tag{3.6}$$

If $R_\theta = R_{IMU2}^{-1} R_{IMU3}$ then

$${}^{tool}t_{LED} = R_\theta \begin{bmatrix} d_2 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ d1 \\ 0 \end{bmatrix} . \tag{3.7}$$

It must be noted that Axis 1 and Axis 2 can occupy any position in the respective vectors (in section 3.3.1) according to the construction of the laparoscope.

In FIGURE 3.11, we see that there are two IMUs, IMU 3 and IMU 4, present. Similarly to IMU 1 and IMU 2, these are for redundancy and rotation averaging purposes. Apart from this, there is one additional future use for these IMUs. The ultrasound array of the tip is barely flexible, thus we assume that the flex between these two IMUs is negligible and at least one of the axes of IMU 3 and IMU 4 is perfectly aligned. If we decide to consider the minor flex of the ultrasound array (perhaps in a slimmer version of the current array), we can measure the relative transformation between these two to evaluate the bending angle occurring at the center of the laparoscope tip.

3.3.2 Ultrasound Image to Laparoscope Tip Calibration

The ultrasound laparoscope array projects the ultrasound waves in one direction in a shape of a fan. The image produced by this fan has an image origin that is considered as the pose of the image plane. There exists a fixed transformation X between the image origin and the center of the ultrasound array. This transformation is computed by using a cross-wire phantom. The process involves using a water bath as shown in FIGURE 3.12.

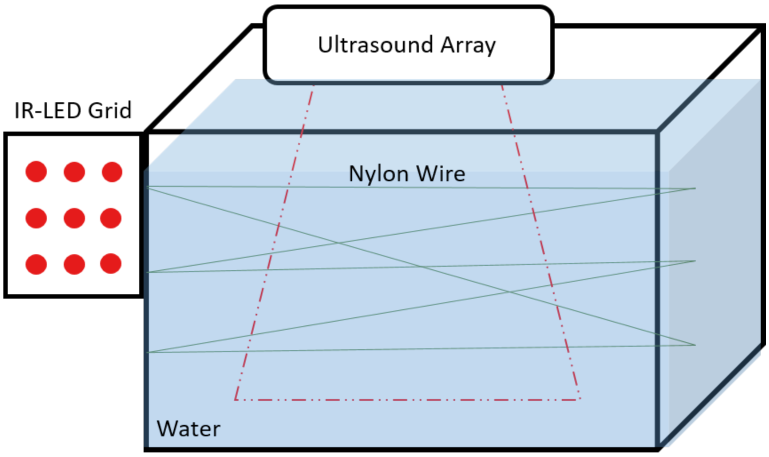

FIGURE 3.12: Schematic of modified wire phantom for ultrasound probe calibration. The IR-LED Grid is detected by the PSD sensors to provide the transformation of the wire phantom

In a traditional cross-wire phantom (water bath) used for this type of calibration in EM sensor based tracking, the probe and cross wire phantom are attached with an EM sensor each. Both EM sensors provide the transformations with respect to the EM transmitter. The water bath container has some crisscross nylon wires attached to its walls. The intersections of these wires create a unique plane inside the phantom. The ultrasound probe is used in a way that it touches the surface of the water in order to propagate the ultrasound waves through the water. The ultrasound waves produce an image of the cross-sections of the wires. The probe is adjusted in a unique way that all the cross-section points appear in the ultrasound image at the same time (FIGURE 3.13). The image with all the wire intersections appearing at the same time confirms that the plane of the ultrasound image is superimposed over the plane of the intersections.


Serial number: 212915

System date/time: Thursday, July 19, 2018 / 12:10

This is a prototype. Not intended for clinical diagnosis and treatment

La sélection de points de repère est active.

FIGURE 3.13: Schematic of the wire phantom for ultrasound probe calibration

Siemens S3000 machine (developer version) has a software to calibrate the compatible ultrasound probes. The dimensions of the water bath and the cross wires as well as the placement of EM sensor are pre-loaded in the software. With these pre-loaded values, the machine can compute the transformation of the ultrasound image with respect to the EM transmitter once the image is aligned with the phantom wire intersection plane. In the calibration mode of the software, the user has to select all the visible points in the ultrasound image. Once all the points are visible and selected in the image, the software confirms that the image is aligned and the software automatically calibrates the ultrasound probe calibration X. Although the dimensions of the cross-wire phantom are proprietary to Siemens Healthcare, the process is well known in the literature. There are several different novel methods for this process mentioned in [She+19], [Che+17]. A review of ultrasound calibration methods is available in [Mer04].

In our case, tracking with EM sensors is replaced with the IMU-PSD system. Hence, the image-tip calibration process is also replaced with these sensors. At this point, we assume that we have the transformation of the laparoscope tip with respect to the PSD rig. To get the transformation of the cross-wire phantom, we attach a IR-LED marker to the phantom as if it is replacing the standard EM sensor attachment. When in direct view of the PSD rig, it can directly obtain the transformation of the water bath with respect to the PSD rig. Using the transformations of the tip and cross-wire phantom as well as the dimensions of the cross-wires, we can obtain X.

Nevertheless, the PSD images have non-linear distortions at the edges. It is a difficult task to fit the entire phantom along with the laparoscope (held in a unique position to match the image plane) in the central part of the PSD image. There are other practical issues with this setup such as requirement of additional power supply to the

IR-LED marker on the cross-wire phantom. To address these issues, we propose replacing the IR-LED marker on the phantom with a commercially available EM-sensor. Now the frame of reference for the phantom (and consequently the ultrasound image) is the EM transmitter while the laparoscope tip has the frame of reference of the PSD camera. In this case, we can use Hand-Eye calibration technique to compute X. This technique provides the fixed transformation between two rigidly attached bodies having different frames of reference for their poses. In the next chapters, we present the Hand-Eye calibration problem in detail as well as our own robust solution to the problem.

3.4 Sensor Noise Reduction

From the kinematic chain relating to our schematic (FIGURE 3.9), we can see that the only source of rotation information are the IMUs mounted on the laparoscope. Since all the IMUs provide the orientation with respect to the gravity vector g, the orientations we receive as data are all in the same coordinate system. Since we know the geometry of the laparoscope fabrication and exactly how the IMUs are mounted on the body of the laparoscope, we can essentially treat all the IMUs on a rigid part as one unit. This means that the two IMUs mounted rigidly on the handle/shaft of the laparoscope can be treated as one unit. With this hypothesis, now we can average the data from these two IMUs to produce a rotation data in which effect of noise and drift is largely reduced. For this, we propose to use the solution given in section 4.1.2.

There are two ways in which we can use rotation averaging. The first one is to take the average of the time-frame matched pairs from the two IMUs. This would allow us to get a reliable output at each instance even when the laparoscope is moving. It will also reduce the accumulation of drift over time. The second and more important averaging is based upon the assumption that the laparoscope does not move sharply in extreme angles as well as it is stationary in one position most of the time. This assumption appears to be true in the case of laparoscopic surgery from our direct observation by attending surgeries during the course of this project. A surgeon moves the laparoscope slowly while in use. In fact, compared to the frame-rate of the IMU data acquisition, the movement of the laparoscope handle is almost stand-still. In other words, the difference in the rotation data at each frame is very small while the laparoscope handle is moved by the surgeon. Another observation we could make was that the laparoscope is generally moved in angles less than $\pi/4\,\text{deg}$. With this assumption on the laparoscope's movement, we can use the proposed solution in section 4.1.2 to take the average of rotations. Hence, whenever the laparoscope is moving slowly or stand-still, we can average the data coming from the same IMU in multiple frames. Given that the IMUs measure acceleration, we can detect the event when laparoscope has started moving based upon a threshold on acceleration value (available from the IMU raw data). When below the threshold, we can state that the laparoscope is not moving and readily start rotation averaging on the data from single IMU (on each IMU separately). When above the threshold, we can stop this process and switch to the averaging with two IMUs together. A combination of these averaging strategies allows us to reduce the effect of noise and drift to a great extent. The following workflow summarizes the algorithm:

1. Define the threshold value to decide between 'moving' or 'not-moving' averaging strategy based upon acceleration value from the IMU3.
2. If below threshold: laparoscope 'not-moving'. Proceed with averaging using the data frames from individual IMUs for a predefined number of frames.
3. Else above threshold: laparoscope 'moving'. Proceed with averaging between the data from IMU1 - IMU2 and IMU3 - IMU4.

3.5 Summary

In this chapter, we presented a novel technique for tracking the pose of a laparoscope tip and consequently the image-pose produced by the ultrasound array mounted on the tip. We also covered some background on the hardware of the sensor systems used in the laparoscope tracking. The IMU-PSD-based tracking has several advantages over a traditional EM sensor-based tracking. But implementation of this tracking system becomes difficult due to the limitations of the IMU dataset that can only provide the rotation and not the translation information reliably. To tackle this issue, we developed a kinematic chain that can use the IMU sensor's rotation data and PSD sensor's position data and use them together to obtain the pose of the laparoscope tip. We overviewed the cross-wire phantom technique that provides the fixed transformation between the laparoscope tip and the ultrasound image origin. Using separate hardware modules provided by Siemens Healthcare for PSD and IMUs, we managed to design a new type of ultrasound laparoscope enabled with a tracking system. To address the noise and drift on the IMU data, we applied a rotation averaging based on a closed-form solution. In chapter 4, we present the problem statements as well as state-of-the-art solutions to the rotation averaging and Hand-Eye calibration techniques used in this chapter.

Chapter 4

Rotation Averaging and Hand-Eye Calibration

In this chapter, we discuss two techniques involved in laparoscope tracking. The first technique of rotation averaging is used in our ultrasound laparoscope tracking proposal to mitigate the effect of noise and drift on the IMU data. The second technique of Hand-Eye calibration is required for fixed body calibrations involved in Hybrid Tracking methods for ultrasound laparoscope tracking.

First, we present the problem of rotation averaging. We discuss the classification of the rotation averaging methods while presenting the problem formulations under different metrics used for solving this problem. We also present a globally optimal closed form solution for the rotation averaging problem under certain constraints that we have proposed to the IMU averaging application in the previous chapter.

Next, we present the Hand-Eye calibration technique along with its problem statement. Hand-Eye calibration is a special case of rotation averaging that is also known as conjugate rotation averaging. In this section, we present a detailed overview and classification of the problem along with the most important state-of-the-art solutions for it. We also briefly present the problem of Robot-World calibration.

4.1 Rotation Averaging

In their article, Hartley et al. [Har+13] have covered the standard rotation averaging techniques in detail. In this section, we present their definitions and problem statements of rotation averaging while classifying them into three categories.

Single Rotation Averaging

Single rotation averaging is used when several estimates of a rotation are obtained and the average of those estimates is the best possible estimation. In other words, rotation data is recorded from a sensor for multiple times while each recording having a minor difference (possibly because of noise) than others. After a certain time interval, we can obtain a mean of all those recordings to obtain a good estimate of the rotation matrix. This is exactly our case in which the IMU provides multiple frames of rotation information per second and we can take an average over these readings. Single Rotation

Averaging in $SO(3)$ is stated as follows: given n rotations R_i, the problem is to find the rotation R that minimizes the cost function

$$C(\mathrm{R}) = \sum_{i=1}^{n} d\left(\mathrm{R}_i, \mathrm{R}\right)^p, \tag{4.1}$$

where d is a distance metric and $p = 1$ or 2.

There are several other applications of single rotation averaging. For instance, several noisy measurements are recorded in a calibrated camera network. We can take a mean of these measurements to reduce the effect of noise.

Multiple Rotation Averaging

Multiple rotation averaging is used when multiple relative rotations R_{ij} are present between different coordinate frame indices by i and j. Only some of these rotations are provided to solve the problem marked by their index pairs i, j in a set $\mathcal{N}$. The goal is to find compatible n absolute rotations R_i so that $R_{ij} \approx R_j R_i^{-1}$. The minimization problem is given by

$$\underset{\mathrm{R}_1,\ldots,\mathrm{R}_n}{\operatorname{argmin}} \sum_{(i,j)\in\mathcal{N}} d^p\left(\mathrm{R}_{ij}, \mathrm{R}_j\mathrm{R}_i^{-1}\right), \tag{4.2}$$

where p =1 or 2 and $i = 1, 2, \ldots, n$. This problem is a complex multi-variable nonlinear optimization problem. This type of rotation averaging is not used in our application related to IMU averaging. Nevertheless, multiple rotation averaging has wide applications in problems such as structure-from-motion(SfM) in which multiple external camera poses are linked together via common image pixels present in those images. This problem has been explored in several papers ([KH08], [HS04], [SH06a], [MP07]). In this problem, it is often assumed that the rotations of the cameras are initially known. These rotations may be estimated separately using multiple rotation averaging. In Govindu [Gov04] and Govindu [Gov06], the authors have used multiple rotation averaging approaches to solve the SfM problem.

Conjugate Rotation Averaging

In this problem, $n \geq 1$ rotation pairs L_i, R_i, left and right rotations, are given and we need to find rotation S such that $R_i = S^i L_i S$ for all i. the rotations L_i and R_i are recorded in different coordinate frames and the fixed coordinate frame S between them is to be determined. The problem statement is as follows,

$$\underset{S}{\operatorname{argmin}} \sum_{i=1}^{n} d\left(\mathrm{R}_i, \mathrm{S}^{-1}\mathrm{L}_i\mathrm{S}\right)^p, \tag{4.3}$$

where p = 1 or 2. Conjugate rotation averaging problem is identical to Hand-Eye calibration problem (presented in the next section) when the rotation part of the transformation is solved separately before the translation part. Solving rotation and translation parts separately have some inherent issues where noise is propagated from the first computation to the next. Therefore, the robotics community prefers to state this problem in a slightly different way to solve for the translation and rotation components simultaneously. We present, in detail, the Hand-Eye calibration problem in the next section.

4.1.1 Distance Metrics

The distance between two rotations is defined by a segment connecting them over in the SO(3). It is analogous to a distance of a line segment connecting two points in Euclidean space. Three metrics are popularly used in the literature.

Angle Metric

Also interchangeably known as Geodesic metric, it is equal to the angle between two rotations. Given two rotations R and S, their product RS^{-1} is also a rotation about some axis by an angle θ in the range $0 \leq \theta \leq \pi$. The angle metric is defined as $d_{\angle}(R,S) = \theta$. With this definition, we can express the angular distance between two rotations R and S to be the angle of the rotation SR^T ($(.)^T$ being the transpose of a matrix) which lies between 0 and π. Hence, $d_{\angle}(S,R) = d_{\angle}(SR^T, I) = \|log(SR^T)\|_2$, where the norm is the Euclidean norm in $\mathbb{R}^3$.

Chordal Metric

The Chordal Metric relates to a specific embedding of a manifold in the Euclidean space $\mathbb{R}^N$. The distance between two points in this manifold is defined as the Euclidean distance in $\mathbb{R}^N$ between the embedded points. Since $R \in SO(3)$ is a 3×3 orthogonal matrix with unit determinant, the natural embedding of a rotation in R is in $\mathbb{R}^9$. Given two rotations R and S, the chordal distance between them is defined as $d_{\text{chord}}(R,S) = \|R - S\|_{\text{frob}}$, where $\|.\|_{\text{frob}}$ is the Frobenius norm of the matrix. $d_{chord}(R,S) = 2\sqrt{2}\sin(\theta/2)$, where $\theta = d_{\angle}(R,S)$.

Quaternion Metric

Another popular distance measure is based upon the Euclidean distance between two quaternions in the embedding space $\mathbb{R}^4$. The distance between two quaternions s and r (representing rotations S and R respectively) is defined by the following formula

$$d_{quat}(S,R) = min(\|s-r\|_2, \|s+r\|_2), \tag{4.4}$$

where the norm is the usual Euclidean norm in $\mathbb{R}^4$. This formula is necessary to compute the quaternion distance between s and r because r and $-r$ represent the same rotation R. If $\theta = d_\angle(S, R) = d_\angle(ST^T, I)$, then the relationship between the quaternion metric to the angle metric is as follows,

$$d_{quat}(S, R) = 2sin(\theta/4).$$

4.1.2 Global Closed-Form Solution for Single Rotation Averaging

The authors of [Har+13] present multiple algorithms for globally and locally optimal solutions for single rotation averaging problem. These solutions are also categorised according to the norms used (L_1 and L_2) as well as using the distance metrics mentioned above. Solving a problem involving rotation matrices can only be solved globally under certain restrictions. The authors of [Har+13] also provide multiple theorems to prove these restrictions. Only one of these solutions is actually a closed form solution. This formula, based upon Chordal L_2-Mean, is only valid if the difference in the angle is less than $\pi/4$ radians. This suits our application with IMUs in which we are using rotation averaging to cancel out the random noise that appears on the IMU readings. To acquire a certain amount of rotation data (to be averaged), the laparoscope must be without any motion. Given that the IMUs provide hundreds of frames of data per second, and the laparoscope is usually moved very slowly and carefully, we obtain a dataset containing multiple rotations for every movement of the laparoscope (more details in chapter 3). Below, we present the borrowed notation of the method from [Har+13] along with the necessary theorems. The theorem is as follows,

Let R_i be rotations satisfying $d_\angle(R_i, S) < \pi/4$ for some rotation S and for all i, then

$$C(\mathrm{R}) = \sum_{i=1}^{n} d_{\text{chord}} \left(\mathrm{R}_i, \mathrm{R}\right)^2 \tag{4.5}$$

is strictly convex on $B(S, \pi/4)$, and hence has a single isolated minimum on that set. A corollary to this theorem states that if R_i rotations lying in a convex set B of radius less than $pi/4$, then the unique global chordal L_2-Mean lies in B and moreover the cost function (4.5) is strictly convex on some ball $B(S, \pi/4) \supset B$. Based upon this theorem, authors of [Har+13] present a closed form solution that can be useful for real-time rotation averaging on IMU data containing rotations with a small difference in angle.

Closed-Form Solution using Quaternions

From [Har+13], we use the following problem statement and solution for single rotation averaging to compute average R_{tool} rotation (Chapter 3, section 3.3.1). Let r_i be the quaternion representations of rotations R_i, where R_i is the data acquired from IMU measurements. And let matrix $A = \sum_{i=1}^{n} r_i r_i^T$ be a 4×4 symmetric matrix. Note that the choice of sign of r does not matter in this case. Let s^* be the eigenvector of A relating to its largest eigenvalue. The following formulation proves that s^* is the quaternion representation for the minimum of the cost function in (4.5) and hence the average of rotation data in quaternions.

Let s be a quaternion. If α is the angle between two unit vectors in $\mathbb{R}^4$, then the inner product between the two vectors is $\cos\alpha$ and angle between the two quaternions (representing two rotations) is $\theta/2$ such as $\theta_i = d_{\angle}(R_i, S)$. Then,

$$s^T A s = \sum_{i=1}^{n} \cos^2(\alpha_i) = \sum_{i=1}^{n} \cos^2(\theta_i/2). \tag{4.6}$$

Vector s^* is obtained by maximizing the product $s^T A s$. Maximizing the cos^2 term is equivalent to minimizing a sin^2 term. Given that $d_{\text{chord}}\left(\mathrm{R}_i, \mathrm{S}\right)^2 = 8\sin^2(\theta/2)$, minimizing $d_{\text{chord}}\left(\mathrm{R}_i, \mathrm{S}\right)^2$ also obtains s^*. Hence, we have,

$$s^* = \underset{\mathbf{s}}{\operatorname{argmin}} \sum_{i=1}^{n} d_{\text{chord}}\left(\mathrm{R}_i, \mathrm{S}\right)^2. \tag{4.7}$$

We obtain the Chordal Mean by using the quaternions (not to be confused with the quaternion mean). Using this algorithm, a unique solution is not possible only when the matrix A has repeated maximum eigenvalues. We can readily apply this problem statement directly taken from [Har+13] to our problem of rotation averaging of IMU data.

Closed-Form Solution using Rotation Matrix

Authors of [Har+13] provide the same solution in terms of matrix representations of rotations as follows. Let $C_e = \sum_{i=1}^{n} R_i \in \mathbb{R}^{3\times 3}$, and let $\langle ., . \rangle$ represent the Frobenius inner product (sum of the elementwise product of two matrices). If R_i and S are rotations, then

$$\begin{aligned}
\sum_{i=1}^{n} d_{chord}(R_i, S)^2 &= \sum_{i=1}^{n} \|R_i - S\|_F^2 \\
&= \sum_{i=1}^{n} \langle R_i - S, R_i - S\rangle \\
&= \sum_{i=1}^{n} (\langle R_i, R_i\rangle - 2\langle R_i, S\rangle + \langle S, S\rangle) \\
&= K - 2\langle C_e, S\rangle,
\end{aligned} \tag{4.8}$$

where K is a constant and is independent of S. Therefore,

$$\begin{aligned}
\underset{\mathbf{s}\in\mathbf{SO(3)}}{\operatorname{argmax}} \sum_{i=1}^{n} d_{\text{chord}}\left(\mathrm{R}_i, \mathrm{S}\right)^2 &= \underset{\mathbf{s}\in\mathbf{SO(3)}}{\operatorname{argmin}} \langle C_e, S\rangle \\
&= \underset{\mathbf{s}\in\mathbf{SO(3)}}{\operatorname{argmin}} \|C_e - S\|_{\text{frob}},
\end{aligned} \tag{4.9}$$

where $\|.\|_{\text{frob}}$ represents Frobenius norm of a matrix. Similar to the previous problem statement using quaternions, R_i and hence C_e, is constructed using IMU measurements and solution of averaging is found in S.

Minimization of the L_2 chordal cost function is equivalent to finding the closest matrix S to C_e under the Frobenius norm. Matrix S is obtained using Singular Value Decomposition. Let the decomposition $C_e = UDV^T$ be such that the diagonal elements of D are arranged in the descending order. If $\det(UV^T) \geq 0$, then set $S = UV^T$. Otherwise set $S = U\,\text{diag}(1, 1, -1)V^T$. Matrix S obtained is the closest rotation to C_e and hence the required rotation minimizing (4.5).

4.2 Hand Eye Calibration

A standard form of Hand-Eye calibration or Conjugate Rotation Averaging problem occurs in an industrial application where a robotic hand manipulator moves an object and a camera observes the robot's hand gripper. In this situation, the camera is steady and the object is held by the hand and is moving with it. In another situation (as shown in FIGURE 4.1), the camera can be mounted on the top of the gripper observing a stationary object. The hand estimates rotations by known geometry of the robot (angle joints etc.) and the camera estimates rotations by observing the stationary object. Though the rotations provided by the camera and hand are in different coordinate frames, we can link them together by solving the Hand-Eye calibration (or Conjugate Averaging) problem.

The Hand-Eye calibration problem is equivalent to Conjugate Rotation Averaging problem as far as the rotational component of a transformation is considered. In Hand-Eye calibration, the translation part is also computed after obtaining the rotation. Computing rotation and translation separately has a drawback of propagating errors from the first rotation computation to the latter. Hence some hand-eye calibration methods solve for the entire transformation at the same time. On this basis, hand-eye calibration methods can be classified into three categories. Decomposed closed-form solutions, simultaneous closed-form solutions and simultaneous iterative solutions. In this section, first we present the basics of the hand-eye calibration problem. Then, we present the state-of-the-art of hand-eye calibration in detail. We briefly explain the Robot-World calibration problem that is solved in the same was as hand-eye calibration. And in the end, we present some of the applications of hand eye calibration.

As discussed in chapter 2, obtaining fix calibration of the body of a laparoscope is fundamental to laparoscope tracking. In the case of a camera-EM sensor tracking setup (presented in [Feu+07]), the fixed calibration between the camera-tracker mounted on the laparoscope and the EM sensor must be obtained. Since this problem is fundamental to laparoscope tracking, in this thesis, we focus on the hand-eye calibration problem in detail. The major contribution of this thesis (presented in chapter 5) falls into the simultaneous iterative solution category of hand-eye calibration problem.

4.2.1 Introduction to Hand-Eye Calibration

Hand-Eye Calibration or Robot-Sensor calibration is a problem in robotics where we estimate the fixed homogeneous transformation between the hand of a robot and the camera mounted on top of it (FIGURE 4.1). In other words, the solution to this problem provides the Euclidean transformation between two rigidly attached motion sensors.

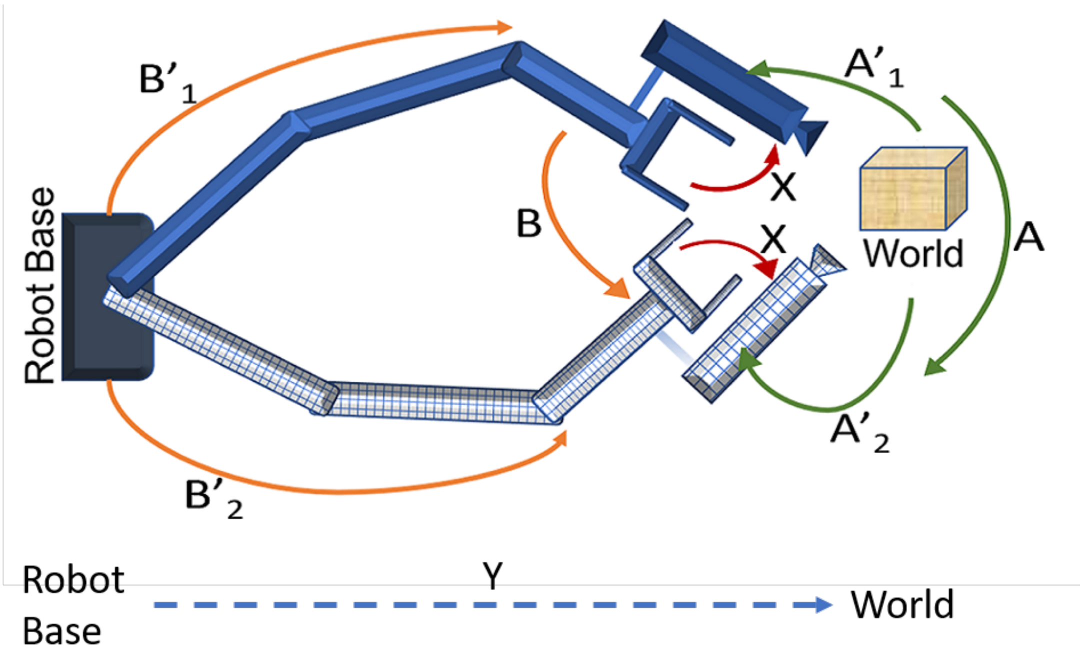

FIGURE 4.1: Hand-Eye calibration setup with a robot hand manipulator and a camera.

In FIGURE 4.1, A'_1 and A'_2 are the transformations provided by the camera by observing a fixed object in the world reference frame. B'_1 and B'_2 are the fixed transformations provided by the kinematic chain of the robot hand with respect to the base of the robot. This figure only shows two positions of the robot. In practice though, the robot hand is moved through multiple positions recording A'_n and B'_n $(n = 1, 2, \dots, n)$ transformations. Transformation A and B correspond to relative motions obtained by the following formula

$$\begin{aligned} A &= A'_2 A'^{-1}_1 \\ B &= B'_2 B'^{-1}_1 , \end{aligned} \tag{4.10}$$

where X is the fixed homogeneous transformation between the robot hand gripper and the camera mounted in top of it. Similarly, Y is also a fixed transformation between Robot Base and the World coordinate frames. All these transformations form two similar systems of equations as follows,

$$\begin{aligned} AX &= XB \\ AX &= YB\,. \end{aligned} \tag{4.11}$$

Since we have multiple motions of the robot hand, we can build an over-determined system of equations using (4.11) and solve for X or Y. Matrices A, B, X and Y are homogeneous transformations in $SE(3)$ and of the form $\begin{pmatrix} R & t \\ 0^T & 1 \end{pmatrix}$ where $R \in SO(3)$. Hence we have

$$\begin{gathered} AX = XB \\ \begin{pmatrix} R_A & t_A \\ 0^T & 1 \end{pmatrix} \begin{pmatrix} R_X & t_X \\ 0^T & 1 \end{pmatrix} = \begin{pmatrix} R_X & t_X \\ 0^T & 1 \end{pmatrix} \begin{pmatrix} R_B & t_B \\ 0^T & 1 \end{pmatrix} \\ \begin{pmatrix} R_A R_X & R_A t_X + t_A \\ 0^T & 1 \end{pmatrix} = \begin{pmatrix} R_X R_B & R_X t_B + t_X \\ 0^T & 1 \end{pmatrix}\,. \end{gathered} \tag{4.12}$$

Thus we have the rotational component of $AX = XB$ as

$$R_A R_X = R_X R_B\,, \tag{4.13}$$

using which we can obtain the translational component as

$$R_A t_X + t_A = R_X t_B + t_X. \tag{4.14}$$

Similarly for $AX = YB$ we can obtain

$$\begin{aligned} R_A R_X &= R_Y R_B \\ R_A t_X + t_A &= R_Y t_B + t_Y\,. \end{aligned} \tag{4.15}$$

Based on these equations, the methods to solve Hand-Eye Calibration problem can be classified into three categories.

- **Decomposed Closed-Form Solutions**: in this category, the rotational component of the hand-eye calibration equations is first solved separately. The translation part is then obtained using the estimated rotation.
- **Simultaneous Closed-Form Solutions**: the rotation and translation parts are obtained simultaneously.
- **Iterative Simultaneous Solutions**: both rotation and translation parts are computed iteratively using optimization techniques. Our method on hand-eye calibration presented in the next chapter falls into this category.

4.2.2 State-of-the-Art of Hand-Eye Calibration

In this section, we present the state of art of hand-eye calibration methods. We also present various problem formulations of some of the known methods in hand-eye calibration. Our survey is partly based upon the paper on Hand-Eye calibration survey from authors of [SEH12].

Decomposed Closed Form Solutions

The hand-eye calibration problem was first introduced in Shiu and Ahmad [SA89] where it is solved separating the rotation and translations parts as follows,

$$\begin{aligned} R_A R_X &= R_X R_B \\ R_A t_X + t_A &= R_X t_B + t_X \,. \end{aligned} \tag{4.16}$$

They use angle-axis parameterization of rotations $R = \mathrm{Rot}(k_R, \theta)$, where k_R is the axis of rotation R and θ is the angle. Hence the proposed solution is as follows,

$$R_X = \mathrm{Rot}(k_{A_i}, \beta_i) R_{X_{P_i}} \,, \tag{4.17}$$

where, $R_{X_{P_i}} = \mathrm{Rot}(v, \omega)$, $v = k_{B_i} \times k_{A_i}$ and $\omega = \mathrm{atan}\,2(|k_{B_i} \times k_{A_i}|, k_{B_i}.k_{A_i})$. β_i is computed by solving a $9 \times 2n$ linear system of equations and the robot motions are $n \geq 2$. In the same paper, the authors provide an important proof that demonstrates that at least two non-parallel rotation axes of rotation R_{A_i} and R_{B_i} are required to solve the equation. Once R_X is computed, we can solve the translation by using

$$(R_{A_m} - I) t_x = R_X t_{B_m} - t_{A_m} \,, \tag{4.18}$$

where $m = 1, 2, \ldots, n$. This is a general way of solving for translation component in decomposed solution solving strategy. Hence, we only present the solutions for the rotation part in the rest of the methods. A drawback of the method from Shiu and Ahmad [SA89] is that the size of the linear system doubles each time a new dataset from a motion is added. Tsai and Lenz [TL89] propose a method that solves hand-eye calibration using a fixed size linear system making it computationally more efficient. They also use angle axis parameterization of rotation $R = \mathrm{Rot}(k_R, \theta)$ using the skew-symmetric matrix $[.]_\times$. They find the rotation axis k_{R_X} as

$$\begin{aligned} [k_{R_{A_i}} k_{R_{B_i}}]_\times k'_{R_X} &= k_{R_{A_i}} - k_{R_{B_i}} \\ k_{R_X} &= 2k'_{R_X} \left(\sqrt{1 + |k'_{R_X}|^2}\right)^{-1} . \end{aligned} \tag{4.19}$$

With this, the angle of rotation is given by

$$\theta = 2\,\mathrm{atan}\,|k'_{R_X}| \,. \tag{4.20}$$

Wang [Wan92] also uses angle-axis representation proposing three methods that correspond to the solution by Tsai and Lenz [TL89].

Park and Martin [PM94] provided a solution for R_X using Lie group theory. For a given rotation R

$$\log R = \frac{\theta}{2\sin\theta}(R - R^T) = [r]_\times \,, \tag{4.21}$$

where $r = \theta k_R$ and θ is the the angle of rotation and k_R is the axis of rotation R. If $\log(A_i)$ and $\log(B_i) = a_i$ and b_i respectively, then $R_{A_i} R_X = R_X R_{B_i}$ is equivalent to $R_X a_i = b_i$. If the noise is present on the motions then according to Park and Martin [PM94] the following minimization problem should be solved,

$$\min_{R_X} \sum_{i=1}^{n} \|R_X a_i - b_i\|^2 \,, \tag{4.22}$$

to which a closed form solution is given as

$$R_X = UV^{-1/2}U^{-1}M^T \,, \tag{4.23}$$

where $M = \sum_{i=1}^{n} b_i a_i^T$ and the eigen-decomposition of $M^T M = UVU^{-1}$.

Chou and Kamel [CK91] use quaternion representation to solve R_X. $R_A R_X = R_X R_B$ is equivalent to $q_A * q_X = q_X * q_B$. q_A, q_B and q_X are quaternion representations of R_A, R_B and R_X respectively. Using the quaternion multiplication properties, we have

$$q_A * q_X - q_X * q_B = (q_A - \overline{q_B}) * q_X = 0 \,. \tag{4.24}$$

They solve this system using singular value decomposition with an additional constraint $\|q_X\| = 1$.

Horaud and Dornaika [HD95] stated that the quaternion representation of R_X can be found as the eigenvector associated with the smallest positive eigenvalue of

$$G = \sum_{i=1}^{n} G_i^T G_i \,, \tag{4.25}$$

where $G_i = \begin{pmatrix} 0 & -a_x^{(i)} + b_x^{(i)} & -a_y^{(i)} + b_y^{(i)} & -a_z^{(i)} + b_z^{(i)} \\ a_x^{(i)} - b_x^{(i)} & 0 & -a_z^{(i)} - b_z^{(i)} & a_y^{(i)} + b_y^{(i)} \\ a_y^{(i)} - b_y^{(i)} & a_z^{(i)} + b_z^{(i)} & 0 & -a_x^{(i)} - b_x^{(i)} \\ a_z^{(i)} - b_z^{(i)} & -a_y^{(i)} - b_y^{(i)} & a_x^{(i)} + b_x^{(i)} & 0 \end{pmatrix}$,

and $a^{(i)} = (a_x^{(i)}, a_y^{(i)}, a_z^{(i)})^T$, $b^{(i)} = (b_x^{(i)}, b_y^{(i)}, b_z^{(i)})^T$ are the rotational axis of R_{A_i} and R_{B_i} respectively. Zhuang and Shiu [ZS92] also use quaternions to solve hand-eye calibration and provide a closed-form solution to the problem similar to Tsai and Lenz [TL89]. Liang and Mao [LM08] use Kronecker product to solve for R_X. With this formulation, they present a linear system as follows,

$$\begin{pmatrix} R_{A_1} \otimes I - I \otimes R_{B_1}^T \\ \vdots \\ R_{A_n} \otimes I - I \otimes R_{B_n}^T \end{pmatrix} \text{vec}\left(R_x\right) = 0\,. \tag{4.26}$$

The Kronecker product between A and B is

$$A \otimes B = \begin{pmatrix} a_{1,1}B & \cdots & a_{1,n}B \\ \vdots & \ddots & \vdots \\ a_{m,1}B & \cdots & a_{m,n}B \end{pmatrix}\,, \tag{4.27}$$

where $a_{i,j}$ is the (i,j)-th element of A, and $\text{vec}(A)$ vectorizes (column-wise) matrix A. The system in (4.26) is solved using SVD and vector X is reconstructed into 3×3 matrix as $R_X = \text{vec}^{-1}(R_X)$. Since there are no additional constraints on R_X, the solution R_X is not generally a rotation matrix. The authors obtain $R_X \in SO(3)$ using U and V^T part of SVD and determining the sign while enforcing it via the D part of the SVD.

Simultaneous Closed-Form Solutions

Chen [Che91] used screw motion theory to analyse the necessary and sufficient conditions for the solution of $AX = XB$. According to these conditions, a unique solution can be obtained if the screws of the screw axes of two robot motions are either in skew or in intersection. Furthermore, even if the screw is ambiguously defined, a partial or full solution can be recovered. Geometrically a screw motion couples rotation to translation (or vice-versa) around an axis. Change in one, causes change in the other. The authors state the fact that rotation and translation of a transformation should not be decoupled before solving the problem. By doing so, one may propagate the errors from the rotational part computation to the translational part, while affecting the generality and efficacy of the resulting algorithm.

Dual-quaternions are the algebraic counterparts of the screws. Daniilidis [Dan99] as well as Daniilidis and Bayro-Corrochano [DBC96] used dual-quaternion parameterization of a transformation in $SE(3)$ to represent the screws and formulated $AX = XB$ as a linear system

$$T \begin{pmatrix} q \\ q' \end{pmatrix} = 0\,, \tag{4.28}$$

where $(q, q')^T \in \mathbb{H}$ is the dual quaternion representing the hand-eye transformation X and matrix T is a $6n \times 8$ matrix

$$T = (S_1^T S_2^T \ldots S_n^T)^T\,. \tag{4.29}$$

If a_i and b_i are the quaternions of rotation R_{A_i} and R_{B_i} respectively, then S_i can be computed as

$$\mathrm{S}_i = \begin{pmatrix} \bar{a}_i - \bar{b}_i & \left[\bar{a}_i + \bar{b}_i\right]_\times & 0_3 & 0_{3\times 3} \\ \bar{a}'_i - \bar{b}'_i & \left[\bar{a}'_i + \bar{b}'_i\right]_\times & \bar{a}_i - \bar{b}_i & \left[\bar{a}_i + \bar{b}\right]_\times \end{pmatrix}, \tag{4.30}$$

where the barred vectors are the imaginary parts of the quaternions. Solution to $(q, q')^T$ is found as an intersection of the null space of (4.28). Daniilidis [Dan99] compared their method against the methods of Tsai and Lenz [TL89] and of Horaud and Dornaika [HD95] only to conclude that this approach is superior to the decomposed solutions. Zhao and Liu [ZL06] also used a similar approach based upon screw-motion theory. Lu and Chou [LC95] used a quaternion representation via an eight step method to solve the hand-eye calibration simultaneously. They used Gaussian elimination and Schur decomposition of a single linear system formed with quaternions.

Andreff, Horaud, and Espiau [AHE99] use Kronecker product to simultaneously solve hand-eye calibration. They form the following linear system,

$$\begin{pmatrix} I - R_{B_i} \otimes R_{A_i} & 0 \\ t_{B_i}^T \otimes I & I - R_{A_i} \end{pmatrix} \begin{pmatrix} \mathrm{vec}\,(R_X) \\ t_X \end{pmatrix} = \begin{pmatrix} 0 \\ t_{A_i} \end{pmatrix}. \tag{4.31}$$

In the same paper, they prove that at least two independent relative motions with non-parallel axes are needed to obtain a solution to hand-eye calibration problem. This method requires orthogonalization of the rotational component since this part can be affected by noise. However, the corresponding translation part is not recomputed along with the orthogonalization step and hence causes errors in the final solution.

Iterative Simultaneous Solutions

Zhuang and Shiu [ZS92] proposed an iterative algorithm that minimizes a non-linear function

$$\sum_{i=1}^{n} \|A_i X - X B_i\|^2. \tag{4.32}$$

Horaud and Dornaika [HD95] also provide an iterative algorithm in which they use Levenberg-Marquardt non-linear optimization algorithm [Mar63] to simultaneously estimate rotation and translation of X. Both methods require an initial estimate to obtain the solution. The algorithm may or may not converge to a global minimum based on the accuracy of the initial estimate. Fassi and Legnani [FL05] also proposed a similar algorithm while providing a geometric interpretation of the hand-eye calibration problem. Wei, Arbter, and Hirzinger [WAH98] propose a method based on the sparse structure of the corresponding normal equations. Mao, Huang, and Jiang [MHJ10] use Kronecker product in their iterative method. Zhao [Zha11] proposed two iterative methods based upon second order cone programming (SOCP). Their first formulation is based upon the method by Andreff, Horaud, and Espiau [AHE99] in which, continuing from (4.31), we have

$$S_i = \begin{pmatrix} I - R_{B_i} \otimes R_{A_i} & 0 \\ t_{B_i}^T \otimes I & I - R_{A_i} \end{pmatrix}$$
$$r_X = \mathrm{vec}(R_X)\,, \tag{4.33}$$

The residual error of the i-th relative motion is defined as

$$e_i = \left\| S_i \begin{pmatrix} r_X \\ r_X \end{pmatrix} - \begin{pmatrix} 0_9 \\ t_{A_i} \end{pmatrix} \right\| . \tag{4.34}$$

The hand-eye calibration is solved as L$_\infty$-norm minimization problem of vector $e = (e_1, e_2, \ldots, e_n)^T$

$$(r_X^*, t_X^*) = \min_{r_X, t_X} \|e\|_\infty \,. \tag{4.35}$$

This method provides a global minimum but requires orthogonalization of r_X. In the same paper, the authors suggest a second method based on dual-quaternion parameterization from Daniilidis and Bayro-Corrochano [DBC96]. With this, the residual e of the i-th motion becomes

$$e_i = \left\| S_i \begin{pmatrix} q \\ q' \end{pmatrix} \right\| , \tag{4.36}$$

where $(q, q')^T$ is dual quaternion representation of X. This method requires an additional constraint to avoid the trivial solution.

Strobl and Hirzinger [SH06b] proposed an iterative method using a parameterization based upon a stochastic model. This method uses a weighting scheme to optimize the accuracy of the solution. Kim et al. [Kim+10] propose a method based on a similar concept while using Minimum Variance.

In their paper, Heller, Henrion, and Pajdla [HHP] present a method to solve hand-eye calibration that is not dependent upon initial estimates and provide a globally optimal solution based upon L-2 norm. In their method, rotation and translation is solved simultaneously by formulating hand-eye calibration problem as multivariate polynomial optimization problem over semi-algebraic sets. They solve the problem using the method of convex linear matrix inequality (LMI) relaxations [Las01]. Multivariate polynomial optimization problem is typically a non-convex problem. In [Las01], the author demonstrated that one can construct a hierarchy of convex relaxations of this problem that produce a monotonically non-decreasing sequence of lower bounds allowing a globally converging solution to the problem. This hierarchy of relaxations is also known as Lasserre's hierarchy. Heller, Henrion, and Pajdla [HHP] use these relaxations to formulate semi-definite programs (SDP) of hand-eye calibration problem under three different parameterizations of rotation matrix. They formulate the first two parameterizations based on the following hand-eye calibration problem

$$\min_{\mathbf{X} \in SE(3)} \sum_{i=1}^{n} \|A_i X - X B_i\|^2 \,, \tag{4.37}$$

which is equivalent to minimization of the sum of the Frobenius norm of $\|A_i X - XB_i\|$. Using this formulation, the problem with orthonormal parameterization of rotation R is stated as

$$\begin{aligned} \min \quad & f\left(u_X, v_X, t_X\right) = \sum_{i=1}^{n} \left\|A_i X\left(u_X, v_X, t_X\right) - X\left(u_X, v_X, t_X\right) B_i\right\|^2 \\ \text{s. t.} \quad & u_X^\top u_X = 1, v_X^\top v_X = 1, u_X^\top v_X = 0\,, \end{aligned} \tag{4.38}$$

where $R = (u, v, w)$ and $u, v, w \in \mathbb{R}^3$ are the orthonormal basis of matrix R. And $X = \begin{bmatrix} R(u,v) & t \\ 0^T & 1 \end{bmatrix}$, where $R(u,v) = (u, v, u \times v)$. Similarly the the problem in quaternion parameterization of rotation is as follows,

$$\begin{aligned} \min \quad & f\left(q_X, t_X\right) = \sum_{i=1}^{n} \left\|A_i X\left(q_X, t_X\right) - X\left(q_X, t_X\right) B_i\right\|^2 \\ \text{s. t.} \quad & q_X^\top q_X = 1, q_{X_1} \geq 0\,, \end{aligned} \tag{4.39}$$

where $X(q,t) = \begin{bmatrix} R(q) & t \\ 0^T & 1 \end{bmatrix}$ and $R(q)$ is the quaternion parameterization of the rotation matrix. Since the quaternions are a double cover of $SO(3)$ (making two global minima for the problem), the second constraint $q_{X_1} \geq 0$ helps the SDP solver to eliminate one of the global minima. The third and final problem formulation is based upon dual quaternions $\hat{a}_i, \hat{b}_i$ and $\hat{q}_X$ that are the dual quaternion representations of rotations A_i, B_i and X while $\otimes$ is the dual quaternion multiplication. The problem is illustrated as follows,

$$\begin{aligned} \min \quad & f\left(\hat{q}_X\right) = \sum_{i=1}^{n} \left\|\hat{a}_i \otimes \hat{q}_X - \hat{q}_X \otimes \hat{b}_i\right\|^2 \\ \text{s. t.} \quad & q_X^\top q_X = 1, \\ & q_{X_1} q_{X_5} + q_{X_2} q_{X_6} + q_{X_3} q_{X_7} + q_{X_4} q_{X_8} = 0, \\ & q_{X_1} \geq 0\,. \end{aligned} \tag{4.40}$$

First two constraints ensure the unity of the dual quaternion using its quaternion parts. The last constraint allows to eliminate one of the two global minima as seen previously.

Heller, Havlena, and Pajdla [HHP16] propose a globally optimal solution to the Hand-Eye calibration problem using a Branch and Bound algorithm. Their algorithm uses camera measurements directly without the prior knowledge of the external camera calibrations. The branch and bound approach minimizes an objective based on the epipolar constraint.

The benefit of the iterative methods is that they can eliminate propagation of the error from the rotation computation to the translation part. However, these type of methods can take significantly longer time to compute a solution. Generally, when a large number of n motions are present, the closed form solutions perform as efficiently as the iterative solutions.

4.2.3 State-of-Art of Robot-World Calibration

In FIGURE 4.1, apart from fixed transformation X, one encounters an additional fixed transformation Y that is the calibration matrix (in $SE(3)$) between the robot base and the stationary object the camera is observing. The methods for solving this transformation are similar to the methods for $AX = XB$. Hence we categorise them similarly in three groups: Decomposed solutions, simultaneous solutions and iterative solutions.

Wang [Wan92] proposed the problem of robot-world calibration for the first time. Their solution assumes the fact that one of the unknowns is given (preferably X). Zhuang, Roth, and Sudhakar [ZRS94] proposed a decomposed closed-form solution based upon the quaternion parameterization of the rotations. Dornaika and Horaud [DH98], extended this formulation in their paper based on quaternions by providing a more accurate solution. Shah [Sha13] provided a solution based on Kronecker product.

Li, Wang, and Wu [LWW10] used dual-quaternions from the hand-eye formulation of Daniilidis and Bayro-Corrochano [DBC96] and Kronecker product from Andreff, Horaud, and Espiau [AHE99] to provide a simultaneous closed form solution to the problem $AX = YB$. In the case of dual quaternion-based solutions, one must take care of the sign of the quaternions of the rotational component while converting the entire transformation to dual quaternions.

Rémy et al. [Rém+97] first introduced the iterative solution for problem $AX = YB$ in which they define the problem in terms of non-linear optimization and use the Levenberg-Marquardt method to solve it. Hirsh, DeSouza, and Kak [HDK01] proposed a method that iteratively solves for rotation and translation separately. The iterative method from Strobl and Hirzinger [SH06b] solves both rotation and translation simultaneously using a stochastic model. Similarly, Kim et al. [Kim+10] solves the problem using Minimum Variance method. Heller, Henrion, and Pajdla [HHP] also provide a globally optimal solution to this problem extending their equations (4.2.2), (4.2.2) and (4.2.2) with additional variables for the robot-world fixed calibration.

4.3 Summary

In this chapter, we presented two techniques: rotation averaging and Hand-Eye calibration. We presented the problem statement of rotation averaging along with its formulations based upon different metrics. We provided a solution to this problem that can be applied to the task of noise reduction on the data from IMUs. In the second part, we presented the problem of Hand-Eye calibration and provided an extensive background on it that includes the state-of-the-art solutions to this problem. In the next chapter 5, we present our own solution to the problem of Hand-Eye calibration which is robust to the noise and outliers.

Chapter 5

Robust Hand-Eye Calibration Method

In this chapter, we revisit the Hand-Eye calibration problem and propose a new deterministic method robust based on convex optimization within an alternating minimization scheme. We first introduce the problem, we present the convex optimization background that is essential to our method. Then, we present the algorithm of our method. We also describe and discuss the extensive experiments we have conducted to test our method against the state-of-the-art methods.

Motivation

The bulk of the literature on Hand-Eye calibration considers a robot-link and a camera, both generally providing satisfactorily accurate measurements. As a consequence, the problem of outlier infested data has known very little attention. Many applications involving a camera coupled with either EM sensors or IMUs are particularly prone to spurious measurement readings and/or large amounts of noise. This may be due to disturbances caused by various physical phenomena such as electromagnetic disturbances, edging or exiting the generated EM field, error accumulation, sensor synchronization issues, etc. Such problems affect predominantly medical instruments mounted with EM sensor where the EM transmitter field is constantly disturbed by EM noise. The excessive use of these instruments may also displace the sensors rendering the need for on-the-fly calibration. To account for outliers, the authors of [Fur+18] use a Random Sample Consensus (RANSAC) framework with the dual quaternion formulation (from [Dan99]) of the transformations. However, RANSAC is non-deterministic and depends on the number of iterations and thresholds chosen within. As a consequence, the RANSAC algorithm can take an unknown amount of time before it finds an acceptable answer. Deterministic approach of our method guarantees to solve the Hand-Eye calibration problem in a definite amount of time while diminishing the influence of outliers at the same time. In the experiments section of this chapter we show the fact that, unlike RANSAC approach, increasing number of randomly placed outliers in the dataset does not affect the time performance of our method. This results into solving the problem in known time while being resistant to outliers and noise in the dataset. This method along with the results have been published in IROS 2019 [Sam+19]. A brief overview of the results of this method have been published in CNIV 2019 [SHM19a].

Revisiting Hand-Eye Calibration Problem

Hand-Eye calibration consists in estimating a fixed Euclidean transformation relating the reference frames of two rigidly attached sensors providing motion measurements. Many robotic and medical imaging devices rely on the presence of a camera (the Eye) that is paired with other pose sensors such as the traditional robotic link (the Hand), Inertial Measurement Units (IMU), odometry, Electromagnetic (EM) or a combination of these. In FIGURE 5.1, A'_1 and A'_2 are the camera transformations obtained from camera pose estimation techniques while B'_1 and B'_2 are those obtained from the robotic link. One motion of the robot hand results into the relative transformations $A = A'_2 A'^{-1}_1$ and $B = B'_2 B'^{-1}_1$. These transformations, along with the rigid transformation X relating the camera and the robot gripper, satisfy $AX = XB$: a relationship originally proposed in [SA89; TL89]. Matrices A, B and X are homogeneous transformations in $SE(3)$. In particular the unknown matrix X is of the form $\begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix}$ where $R \in SO(3)$. Note that at least two relative motions (3 absolute poses) with non-parallel rotational axes are needed to estimate X [SA89]. In practice, several relative motions are used to create multiple sets of matrices A_i and B_i, $i = 1, \dots, n$, leading to an over-determined system (with possibly noisy data). Hence, X is estimated by solving the problem

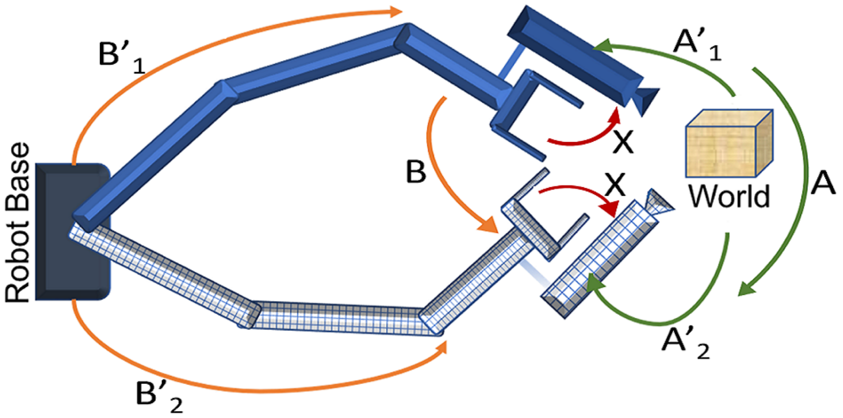

FIGURE 5.1: Traditional Hand-Eye setup in robotics with a camera mounted on a robot gripper attached to the robot base. Transformations are obtained via the kinematic link and camera pose estimation.

$$\begin{aligned} &\min_X \sum_i^n \|(A_i X - X B_i)\| \\ &s.t.\, X \in SE(3)\,, \end{aligned} \tag{5.1}$$

where $\|.\|$ denotes the Frobenius norm.

Based on the Hand-Eye calibration notation mentioned in FIGURE 5.1, we present a deterministic, robust and accurate method for solving the Hand-Eye calibration problem even in the presence of large amounts of outliers and high levels of measurement noise. The proposed method is based on a rank-constrained SDP formulation of the Hand-Eye calibration. This SDP formulation is obtained by considering the membership of the rotational component of X to the convex hull of $SO(3)$. This membership is described by a convex Linear Matrix Inequality (LMI). The full membership to $SO(3)$ is then enforced via a single matrix rank constraint that, alone, encapsulates the nonlinearity of the problem. To enforce this rank constraint, we employ an efficient method alternating between solving two SDP problems: one for enforcing the rank while the other for recovering the calibration matrix. We exploit this formulation of the Hand-Eye calibration problem to robustify the estimation process through an iteratively re-weighted optimization scheme that allows to reduce the influence of outlier motions in the estimation problem.

5.1 Background and Notations

In this section, we present the adopted notations as well as a brief background on Semi-Definite Programming, the notion of convex-hull of rotations in 3-space and the rank-constrained LMI feasibility problem. These are the main ingredients in our rank-constrained SDP reformulation of the Hand-Eye calibration problem (5.1).

5.1.1 Semi-Definite Programming

Semi-Definite Programming (SDP) problems are convex optimization problems with a linear objective function and Linear Matrix Inequality (LMI) constraints, that is,

$$\begin{aligned} &\min \sum_{i=1}^{m} b_i y_i \\ &\text{s.t. } C + \sum_{i=1}^{m} y_i D_i \succcurlyeq 0 \,, \end{aligned} \tag{5.2}$$

where $y_i \in \mathbb{R}$ are the unknown variables while $b_i \in \mathbb{R}$, matrices $C = C^T \in \mathbb{R}^{n \times n}$ and $D_i = D_i^T \in \mathbb{R}^{n \times n}$ are provided. The inequality sign means that the matrix $L = C + \sum_{i=1}^{m} y_i D_i$ is positive semi-definite ($\preccurlyeq$ would refer to negative semi-definiteness) and the inequality constraint is called as a LMI. Matrix L is constrained between its smallest and the largest eigenvalues, respectively, λ_{min} and λ_{max}. For instance, $L - \lambda_{min} I \succcurlyeq 0$ and $L - \lambda_{max} I \preccurlyeq 0$ (matrix I is the identity matrix). There are several methods that can efficiently solve SDPs, notably interior point methods [VB97]. This article [VB97] also provides a generous history and survey of solving SDPs. Though these formulations appear to be very specific form of problems, they are rather widely used in the field of control theory [VB97; Boy+94] and computer vision [Pau+15].

5.1.2 Rank-Constrained LMI Feasibility Problem

Consider a symmetric $m \times m$ matrix M and the rank-constrained semi-definite feasibility problem on the left of (5.3)

$$\begin{array}{ll} \underset{M}{\text{find}} & M \\ \text{s.t.} & M \in \mathcal{C} \\ & M \succcurlyeq 0 \\ & \operatorname{rank} M \leqslant n \end{array} \quad \equiv \quad \begin{array}{ll} \underset{M}{\min} & \operatorname{trace}(MZ^\star) \\ \text{s.t.} & M \in \mathcal{C} \\ & M \succcurlyeq 0 \end{array} \quad , \tag{5.3}$$

where $\mathcal{C}$ is a convex set containing positive semi-definite matrices of rank $n \leqslant m$ or less. If $M^\star$ is a solution to this problem, then there exists a $m \times m$ symmetric positive semi-definite matrix $Z^\star$, of rank at least $m - n$, such that $\operatorname{trace} M^\star Z^\star = 0$. With this $Z^\star$, the solution to the problem on the right-hand side of (5.3) is also $M^\star$.

As a consequence, alternatively, one may solve the rank constrained problem in (5.3) by alternating the solutions of the following two SDP problems

$$\begin{array}{ll} \underset{Z}{\min} & \operatorname{trace}(\mathrm{M}^\star \mathrm{Z}) \\ \text{s.t.} & 0 \preccurlyeq Z \preccurlyeq I \\ & \operatorname{trace}(Z) \geqslant m - n \end{array} \quad \rightleftharpoons \quad \begin{array}{ll} \underset{M}{\min} & \operatorname{trace}(MZ^\star) \\ \text{s.t.} & M \in \mathcal{C} \\ & M \succcurlyeq 0 \end{array} \quad . \tag{5.4}$$

Matrix Z, estimated through solving the problem to the left in (5.4), acts as a direction matrix when estimating M in the problem to the right in (5.4). This process is initialized with some $Z^\star$ and repeated until convergence. The constraint $0 \preccurlyeq Z \preccurlyeq I$ enforces the semi-definiteness of Z as well as bounds its largest eigenvalue by unity. Along with $\operatorname{trace}(Z) \geqslant m - n$, these two constraints are used as surrogates to the rank of Z. The interested reader may refer to [Dat10], chapter 4.4.1, for further details on this approach.

5.1.3 Convex-Hull of Rotations

Consider a 3×3 matrix R and the function $\mathtt{C}$ mapping $\mathbb{R}^{3\times3}$ to $\mathbb{R}^{4\times4}$, as given in (5.8), where R_{ij} denotes R's entry at the i^{th} row and j^{th} column. The set

$$\operatorname{conv} SO(3) = \{R \in \mathbb{R}^{3\times3} : \mathtt{C}(R) \succcurlyeq 0\} \tag{5.5}$$

represents the tightest convex hull of the set of 3-space rotation matrices. This result was first reported in [SSS11] while characterizing orbitopes, then in [SPW15] as part of the characterization of the convex hull to rotations. Indeed, should R be a rotation matrix, the conversion between a unit quaternion $z = a + \boldsymbol{i}b + \boldsymbol{j}c + \boldsymbol{k}d$ (where $|z| = 1$) and matrix R is given by

$$R = \begin{bmatrix} a^2+b^2-c^2-d^2 & 2bc-2ad & 2bd+2ac \\ 2bc+2ad & a^2-b^2+c^2-d^2 & 2cd-2ab \\ 2bd-2ac & 2cd+2ab & a^2-b^2-c^2+d^2 \end{bmatrix}. \tag{5.6}$$

It can be observed that the entries of R are constructed from the terms of a rank-1 4×4 symmetric matrix U

$$U = \frac{1}{a^2+b^2+c^2+d^2}\begin{bmatrix} a^2 & ab & ac & ad \\ ab & b^2 & bc & bd \\ ac & bc & c^2 & cd \\ ad & bd & cd & d^2 \end{bmatrix}, \tag{5.7}$$

the entries of which can be obtained by inverting the relations $R_{11} = a^2 + b^2 - c^2 - d^2$, $R_{12} = 2bc - 2ad, \ldots$ etc. This inverse mapping characterizes matrix $\mathtt{C}(R)$ (as given by (5.8)). Matrix $\mathtt{C}(R)$ being positive semi-definite ensures R's membership to conv $SO(3)$. If in addition $\mathtt{C}(R)$'s rank is 1, then R is a guaranteed to be a rotation matrix.

$$\mathtt{C}(R) = \begin{bmatrix} 1+R_{11}+R_{22}+R_{33} & R_{32}-R_{23} & R_{13}-R_{31} & R_{21}-R_{12} \\ R_{32}-R_{23} & 1+R_{11}-R_{22}-R_{33} & R_{21}+R_{12} & R_{13}+R_{31} \\ R_{13}-R_{31} & R_{21}+R_{12} & 1-R_{11}+R_{22}-R_{33} & R_{32}+R_{23} \\ R_{21}-R_{12} & R_{13}+R_{31} & R_{32}+R_{23} & 1-R_{11}-R_{22}+R_{33} \end{bmatrix}. \tag{5.8}$$

5.2 Deterministic Robust Hand-Eye Calibration

In this section, we present the main contribution of the paper, that is: our SDP formulation of the Hand-Eye calibration problem and the underlying robust algorithm.

5.2.1 Semi-Definite Problem Formulation

We first formulate the Hand-Eye calibration problem as a rank-constrained SDP and show how this can be solved using alternating optimization of two convex problems. The formulation herein does not explicitly address the problem of robustness. This is addressed in the next section which also provides our full algorithm.

Assuming $n \geqslant 2$ motions available, an alternative formulation to the Hand-Eye calibration problem (5.1) is the following:

$$\begin{aligned} &\min_X \sum_i \sigma_{max}(A_i X - X B_i) \\ &\text{s.t. } X \in SE(3), \end{aligned} \tag{5.9}$$

where $\sigma_{max}(.)$ denotes the largest singular of its matrix argument. Note that this problem is not equivalent to the one stated in (5.1). The latter minimizes the L_2-norm, which is generally preferred over other norms. However, relying on the cost function in (5.9) carries a number of advantages. Firstly, minimizing the largest singular value of a matrix is a SDP problem. As a result, such minimization may be subjected to additional LMI constraints such as, for the problem at hand, the membership of the rotational component of X to the convex hull of rotations conv $SO(3)$ (5.5) rather than

$SO(3)$. Then, the largest singular value alone is an excellent indicator for the relative quality of motion measurements employed to estimate X. Indeed, with the correct estimate of X, motions with small largest singular values of $A_iX - XB_i$ indicate that A_i and B_i are good measurements and bad otherwise. Such singular values can hence be trusted to provide weights to the various motions involved in X's estimation.

To see that minimizing the largest singular value of a matrix is a SDP problem, observe that the singular values of a matrix, say $(AX - XB)$, are the square root of the eigenvalues of $(AX - XB)^T(AX - XB)$. Therefore, as discussed in section 5.1.1, the latter matrix and its largest eigenvalue σ_{max}^2 satisfy $(AX - XB)^T(AX - XB) - \sigma_{max}^2 \leqslant 0$. Equivalently, and because singular values are always nonnegative, we can write

$$\frac{1}{\sigma_{max}}(AX - XB)^T(AX - XB) - \sigma_{max} I \leqslant 0. \tag{5.10}$$

Note that the matrix on the left-hand side of (5.10) is the Schur complement of the matrix on the left-hand side of

$$\begin{bmatrix} \sigma_{max} I & (AX - XB) \\ (AX - XB)^T & \sigma_{max} I \end{bmatrix} \geqslant 0. \tag{5.11}$$

Furthermore, by Schur's complement lemma (a proof of which can be found in [VB00]), the two inequalities (5.10) and (5.11) are equivalent. As a result, considering X of the form $\begin{bmatrix} R_X & t_X \\ 0^T & 1 \end{bmatrix}$, problem (5.9) may be equivalently restated as:

$$\begin{aligned} &\min_{R_X, t_X, \sigma} \sum_i^n \sigma_i \\ &\quad \text{s.t.} \\ &\qquad \begin{bmatrix} \sigma_i I & (A_iX - XB_i) \\ (A_iX - XB_i)^T & \sigma_i I \end{bmatrix} \geqslant 0, \\ &\qquad \mathtt{C}(R_X) \geqslant 0, \\ &\qquad \operatorname{rank} \mathtt{C}(R_X) = 1. \end{aligned} \tag{5.12}$$

Note that $\mathtt{C}(R_X)$ ensures R_X's membership to $\operatorname{conv} SO(3)$ (5.5). Together with the rank condition, these two constraints ensure the resulting R_X to be in $SO(3)$, hence $X \in SE(3)$. Problem (5.12) carries the advantage of confining the nonlinearity of the problem to a single rank constraint. A relaxed convex problem can simply be solved by dropping this constraint. However, the resulting solution would not guarantee the estimated R_X to be in $SO(3)$. We show, in the following, that a better solution satisfying $\mathtt{C}(R_X)$'s rank condition in practice can be obtained. The proposed approach is based on the alternating minimization procedure described in Section 5.1.2. Given some initial 4×4 matrix $Z^\star$ of rank 3, the idea is to alternate between solving

$$\begin{aligned} &\min_{R_X, t_X, \sigma} \sum_i^n \sigma_i + \mu \, \mathrm{trace}(\mathtt{C}(R_X) Z^\star) \\ &\text{s.t.} \\ &\qquad \begin{bmatrix} \sigma_i I & (A_i X - X B_i) \\ (A_i X - X B_i)^T & \sigma_i I \end{bmatrix} \succcurlyeq 0, \\ &\qquad \mathtt{C}(R_X) \succcurlyeq 0, \end{aligned} \tag{5.13}$$

which involves a regularization term controlled by a fixed nonnegative parameter μ, and the problem

$$\begin{aligned} &\min_Z \ \mathrm{trace}(\mathtt{C}(R_X^\star) Z) \\ &\text{s.t.} \ \mathrm{trace}(Z) \geqslant 3, \ 0 \preccurlyeq Z \preccurlyeq I. \end{aligned} \tag{5.14}$$

Here, $R_X^\star$ is obtained by solving (5.13) while the direction matrix $Z^\star$ is the minimizer obtained by solving (5.14). Note that both (5.13) and (5.14) are SDP, hence convex, problems. The algorithm iterates until there is no change in X. Note that the choice and influence of the regularization parameter μ are discussed in the experiments section. The choice of the regularization parameter μ and the initial $Z^\star$ is delayed to the next section, following the presentation our full robust algorithm.

5.2.2 Robust Hand-Eye Calibration

The convexity of the alternating problems (5.13) and (5.14), along with the estimates of the largest singular values, provide a very suitable ground for an iteratively reweigthted procedure. Such procedures are known to handle efficiently both noise and outliers. To do so, consider a set of singular values $\sigma_i^{(k-1)}$ and $X^{(k-1)}$ all estimated at iteration $k-1$ of the algorithm, $k \geqslant 1$. With $X^{(k-1)}$, one can estimate the direction matrix $Z^{(k)}$ to be used to estimate $X^{(k)}$ at the next iteration k by solving (5.14). In turn, the weights $w_i^{(k)}$ are necessary to estimate $X^{(k)}$. They are inversely proportional to $\sigma_i^{(k-1)}$ and can be first computed as

$$w_i^{(k)} = \frac{1}{\sqrt{\sigma_i^{(k-1)}} + \epsilon}, \tag{5.15}$$

where ϵ is a very small number, then normalized as described in Algorithm 1. Given the weights and direction matrix, $X^{(k)}$ is estimated by solving

$$\begin{aligned} &\min_{R_X, t_X, \sigma} \sum_{i}^{n} w_i^{(k)} \sigma_i + \mu \operatorname{trace}(\mathtt{C}(R_X) Z^{(k)}) \\ &\text{s.t.} \\ &\quad w_i^{(k)} \begin{bmatrix} \sigma_i I & (A_i X - X B_i) \\ (A_i X - X B_i)^T & \sigma_i I \end{bmatrix} \succcurlyeq 0, \\ &\quad \mathtt{C}(R_X) \succcurlyeq 0. \end{aligned} \tag{5.16}$$

Such weighting scheme efficiently reduces the influence of corrupt measurements. The full algorithm is given by Algorithm 1. In this algorithm, $\gamma = 1e-7$ is a threshold value to determine that X's estimation is unchanged compared to its estimate at the previous iteration (in the Frobenius norm sense). This is the stopping criterion. The initial matrix $Z^{(0)}$ must be rank-3. In practice one may start from any diagonal matrix with all ones on the diagonal except for one zero-entry. We suggest to run the algorithm with all 4 such possible values of matrix $Z^{(0)}$ and retain the one with the minimum cost of the SDP. Furthermore, we found that $\mu = 1e-6$ was a good choice in all our experiments. Its influence on the algorithm is discussed in the experiments section. All weights can be initialized to 1.

Algorithm 1: Robust Hand-Eye calibration algorithm

```
Input: $A_{i=1...n}, B_{i=1...n}, \gamma, Z^{(0)}$
Output: $X$
$w^{(0)}_{i=1...n} = 1, \zeta = +\infty, k = 0$
Estimate $X^{(0)}$ by solving (5.16)
while $\zeta > \gamma$ do
    $k = k + 1$
    Estimate $Z^{(k)}$ by solving (5.14)
    $w^{(k)}_{i=1...n} = \frac{1}{\sqrt{\sigma_i^{(k-1)}}} + \epsilon$
    $w^{(k)}_{i=1...n} = \frac{w^{(k)}_{i=1...n}}{\sum_{j=1}^{n} w_j^{(k)}}$
    Estimate $X^{(k)}$ by solving (5.16)
    $\zeta = \|X^{(k-1)} - X^{(k)}\|$
return $X = X^{(k)}$
```

5.3 Experiments and results

We have conducted experiments using 3 types of datasets: synthetic dataset, real motion dataset of hand-eye robot setup and real motion dataset from an EM sensor-camera setup. In all the experiments, our robust Algorithm 1 was used with parameter

$\mu = 1e - 6$, threshold $\gamma = 1e - 7$ (as discussed given in section 5.2.2). Also, our initial $Z^{(0)}$ was arbitrarily chosen as the diagonal matrix with all ones but the first entry.

5.3.1 Synthetic Data Experiments

In the synthetic experiments, we simulated a scene of 50 points confined to the surface of a 1 meter radius sphere. Random views of the scene were generated by moving the virtual camera around the sphere surface while constraining all the points to be visible in all the images. The camera intrinsic parameters were adjusted accordingly. The camera was placed at a 2.5 meters mean distance from the center of the sphere and 0.7 meters standard deviation. We provided a Hand-Eye transformation X which, along with the camera pose, was used to create ground truth absolute hand transformations. These were used to create relative motions for both camera and hand. For each level of changing noise or percentage of outliers, we ran 100 independent trials over which the error measurements were averaged.

To quantify the results, we used the RMS errors in rotation unit quaternion $||q - \hat{q}||$ and RMS of the relative errors in translation $||t - \hat{t}|| / ||t||$. These are customary metrics used in [TL89; HD95; Dan99]. Though the metrics for rotation and translation are given independently, note that the success of a method depends upon the performance in both metrics of rotations and translations combined.

In the synthetic experiment results, methods *qhec*, *uvhec* and *dqhec* are from [HHP], *tsai* refers to [TL89], *Inria* is from [HD95] and *navy* refers to [PM94]. For these methods, we used the Matlab implementations from [Wen16]. *traceZ* refers to the Matlab implementation of Algorithm 1 using YALMIP [Löf04] as a parser and SeDuMi [Stu99] as a SDP solver.

Experiments with varying noise

In the first experiment, we used 30 motions of the virtual Hand-Eye setup throughout 100 trials for increasing noise levels. An absolute Gaussian noise was added on the quaternion rotation of hand transformations with increasing standard deviation (0 to 0.2 quaternions). Relative Gaussian noise was added on the translation vector of both hand transformations (0 to 20%). For the camera transformations, a Gaussian pixel noise was added on the 2D pixels from the camera views with an increasing intensity (0 to 2 pixels). The results are illustrated in FIGURE 5.2. Note that the best performing methods are *qhec* (as well as *uvhec*) [HHP] and ours with a better translation estimation with our method.

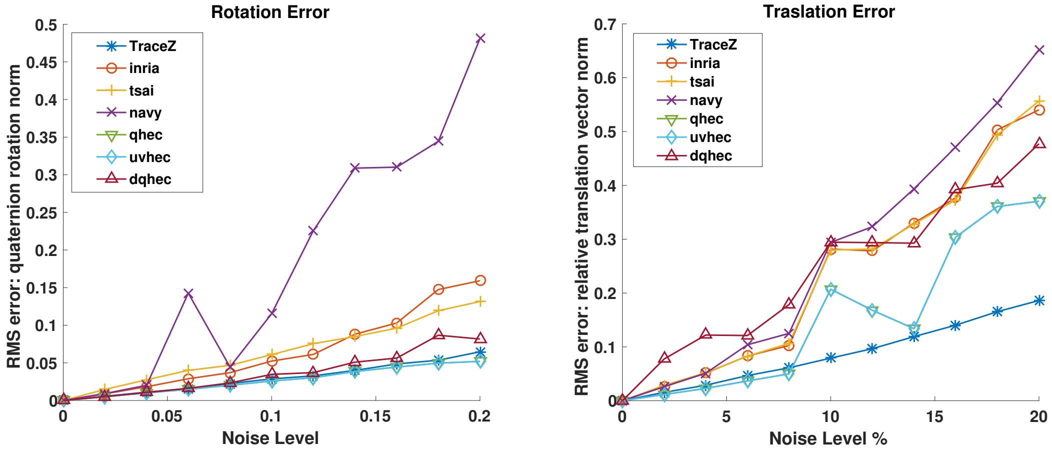

FIGURE 5.2: Gaussian noise on hand and camera motions. RMS of quaternion rotation errors (left), RMS of relative translation vector errors (right).

In the second experiment, we used the same noise adding scheme as above except that the Gaussian noise on the hand transformations is replaced by uniform noise. It is generally hard to model the noise from IMUs/EM sensors and it is usually non-Gaussian. Pixel noise remained Gaussian as in the previous experiment. Our results are illustrated in FIGURE 5.3 in which we note a similar performance as the previous experiment.

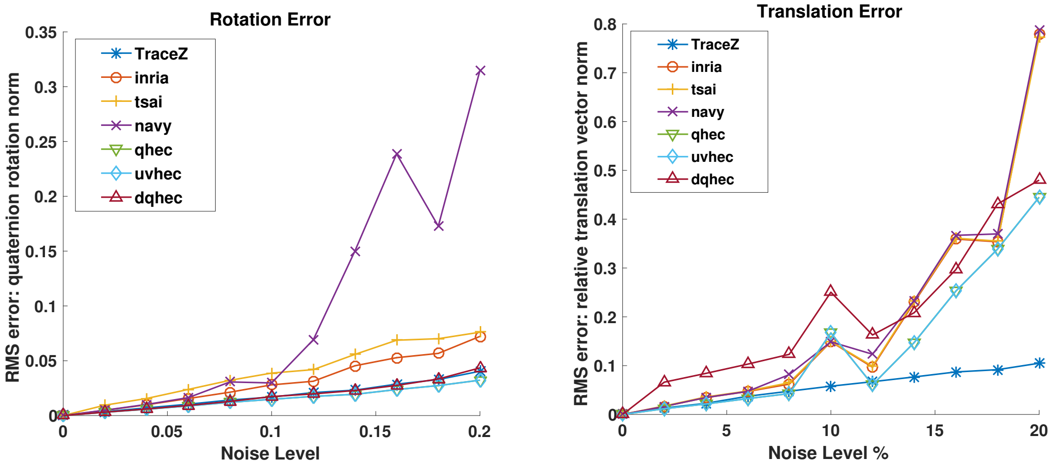

FIGURE 5.3: Uniform noise on hand and camera motions. RMS of quaternion rotation errors (left), RMS of relative translation vector errors (right).

Experiments with Outliers

In this experiment, 30 motions with a marginal noise were used throughout the experiment. We added a constant amount of 0.05 standard deviation absolute Gaussian noise on quaternion of hand transformations, 5% relative Gaussian noise on translation vectors of the hand and 1 pixel Gaussian noise on cameras. To create the outlier data, we added an increasing percentage of outliers (totally random data) in the total number of motions. Only hand transformations (occurring from IMU, EM sensors, odometry or the combinations of these sensor measurements) were affected by outliers. The camera transformations were not affected by outliers since cameras have a direct view of the target. Also, there are various robust algorithms already present on the camera hardware/software platforms allowing one to obtain good camera pose estimation. The results reported in FIGURE 5.4 show that the outliers are efficiently rejected up to 70% (only 9 out of 30 good relative motions). After 70%, our method starts to fail, albeit not by a big margin.

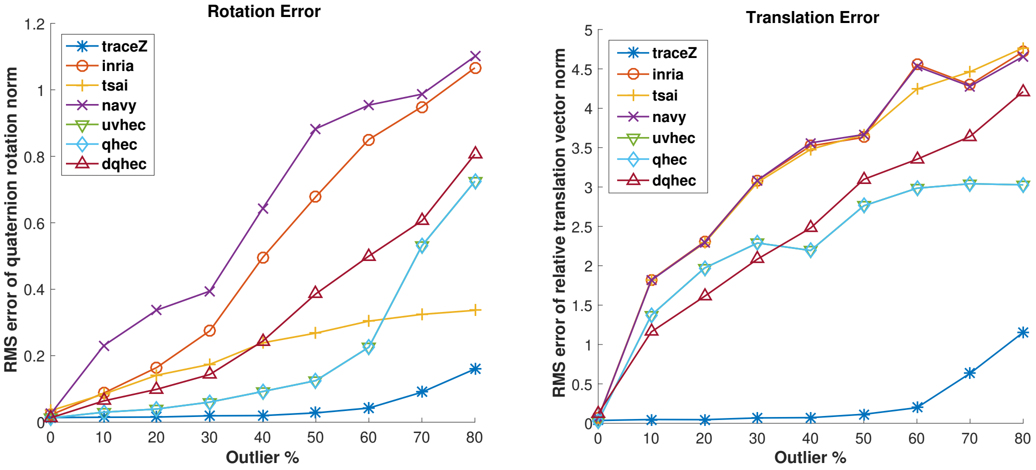

FIGURE 5.4: Outliers replacing increasing percentage of motions. RMS of quaternion rotation errors (left), RMS of relative translation vector errors (right).

Performance at Minimal Case

In this experiment, we keep the noise at a constant level for both hand and camera transformations throughout the experiment. We vary the number of relative motions from 2, 3, 4, 6, 8, 10, 12, 14, 16, 18, 22, 26 and 30. The constant noise level throughout the experiment is 0.05 standard deviation absolute Gaussian noise on the quaternion rotation of hand transformations, 5% relative Gaussian noise on the translation vector of the hand transformations and 1 pixel Gaussian noise on the camera transformations. Results are in FIGURE 5.5. The minimal case for solving the equation $AX = XB$ is 2 relative motions. However, due to the presence of noise on the data, an acceptable solution can be obtained with minimum of 3 relative motions. The graphs for *qhec*, *uvhec*, *dqhec* and *traceZ* are plotted from 2 motions since they could find a numerical solution in this case. Other methods failed, hence their graphs start from 3 motions onwards. This experiment intends to show the performance of the compared methods under the minimal relative motions criterion.

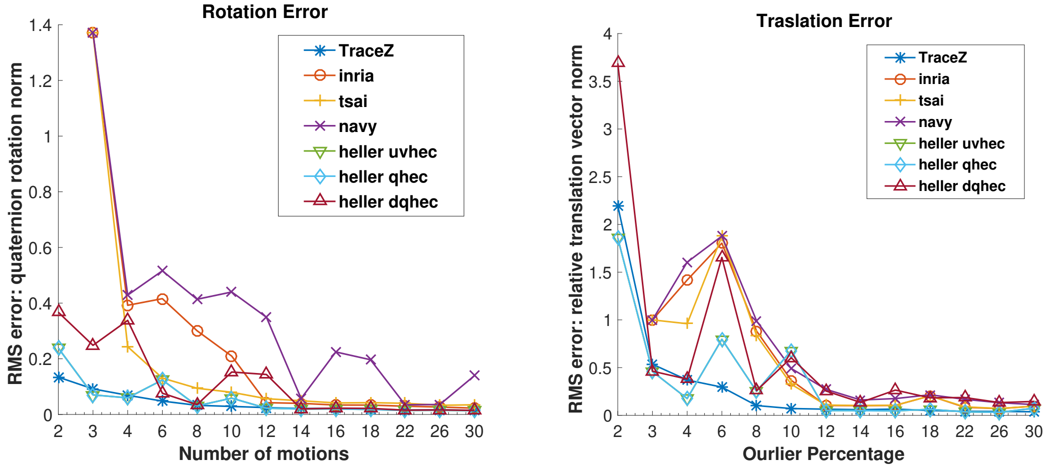

FIGURE 5.5: Increasing number of motions. RMS of quaternion rotation norm (left), RMS of relative translation vector norm (right).

Our method with nonlinear refinement

In this experiment, we demonstrate the importance of the weights for removing the outliers. We add an additional step of nonlinear refinement at the end of our algorithm. The final output of our method is provided as an initial estimate to the refinement algorithm. With this, we create two versions of our method using nonlinear refinement: the first one with using weights and the second, without using them. We compare these two versions with our original method, *traceZ* (without any refinement), as well as the method from [HHP] with quaternion representation. The nonlinear problem with fixed weights w_i is as follows:

$$\begin{aligned} &\min_X \sum_i w_i \|(A_i X - X B_i)\|^2 \\ &\text{s.t. } X \in SE(3) \end{aligned} \tag{5.17}$$

where $\|.\|$ represents Frobenius norm. The rotation can be parametrized to quaternions to constrain it to SO(3). In Figure 5.6, we can see that *traceZ* provides the best results without any refinement when compared to the others. The version of *traceZ* with nonlinear refinement without weights performs almost as same as the *qhec* method from [HHP].

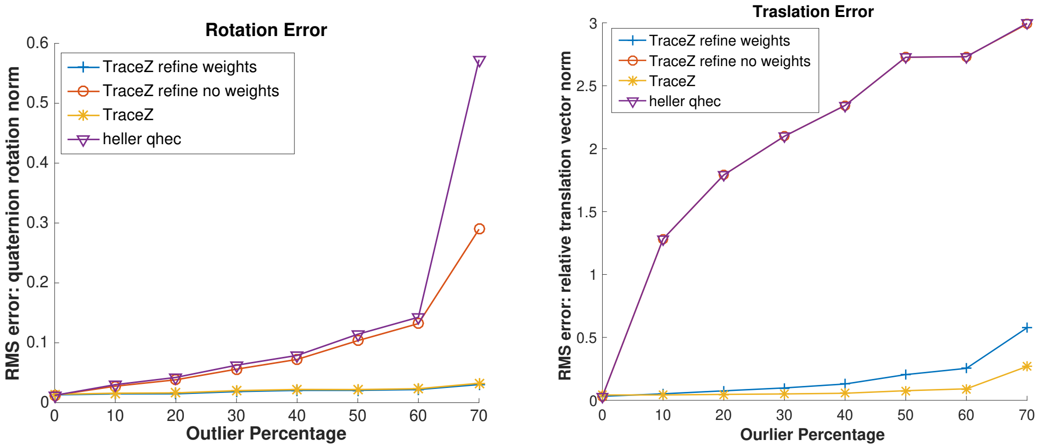

FIGURE 5.6: Outliers replacing increasing percentage of motions. RMS of quaternion rotation norm (left), RMS of relative translation vector norm (right).

Time performance of our method

In this experiment, the noise levels and outliers were used as in the outlier experiments. Here we present two types of graphs (FIGURE 5.7). The line graph presents the average time taken by one iteration of the weighting of the motions (the 'while loop' in Algorithm 1). The bar graph presents the number of iterations taken by the 'while loop' before termination. The average of this time was computed for 50 trials for each outlier (completely random data) percentage. The bar part of the graph shows that the number of iterations hardly increases with increasing number of outliers. Similarly, the line graph also fluctuates only in milliseconds highlighting the apparent disconnection between the ratio of outliers and computation time.

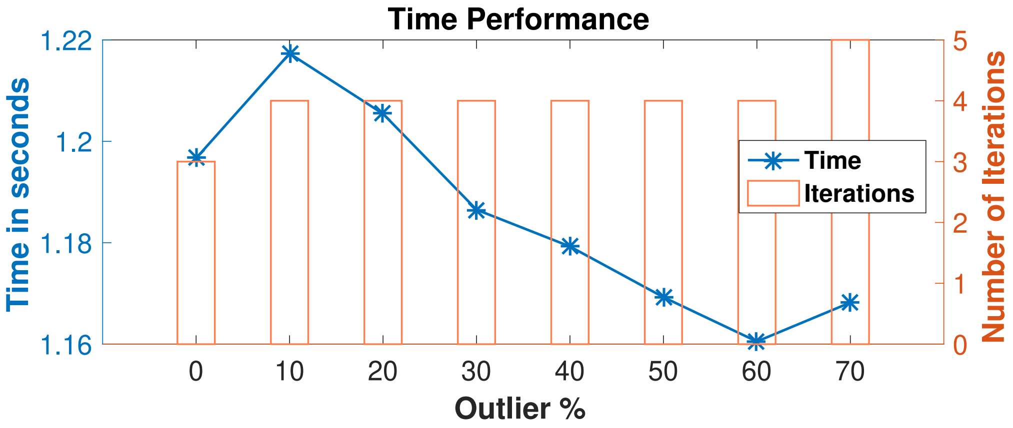

FIGURE 5.7: Average time performance of each iteration. Left Y-axis: Average time taken by each iteration in 50 trials. Right Y-axis: Total number of iterations.

Experiments with regularization parameter

We have used μ as a regularization parameter in our problem formulation (5.13). Changing the value of μ affects the time of convergence of our method. In this experiment, we imposed a nominal absolute noise of 0.05 standard deviation (in quaternions) to hand motion rotations and 5% relative Gaussian noise on hand translations. 1 pixel noise was added to cameras. FIGURE 5.8 shows the time per iteration and number of iterations taken for each μ. We observe that choosing a correct μ value is critical for the time. For instance, with $\mu = 1e-4$, the program takes very long (approx. 16 minutes). With $\mu = 1e-5$, it only takes a 4 to 5 seconds to terminate.

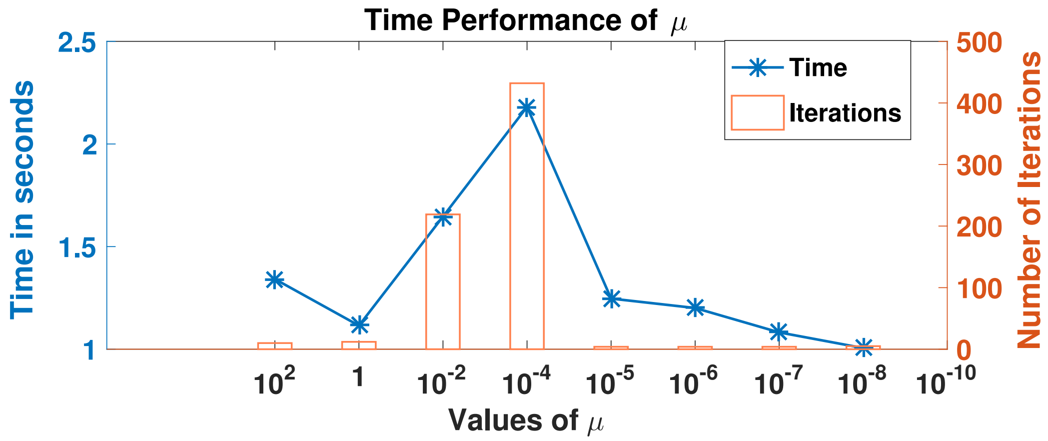

FIGURE 5.8: Average time performance while changing the regularization parameter. Left Y-axis: Average time taken by each iteration in 50 trials. Right Y-axis: Total number of iterations.

5.3.2 Real Data Experiments with Robot-Camera Setup

The authors of [Fur+18] published an open source dataset for hand-eye calibration recorded on a real robot: a UR-10 arm equipped with a RealSense SR300 RGB-D sensor mounted rigidly close to the end effector. They also provide an algorithm to time-align the motions according to the time-stamps and also to pre-filter the motions and reject those with very similar information. Only those with the maximum information are retained. Using this algorithm, we could obtain 1686 ’time-aligned’ motion pairs to test our method. Their ’RANSAC Scalar (RS) based inlier check’ algorithm uses 30 pre-selected motions out of 1686 motions to perform Hand-Eye calibration. The RS method uses RANSAC framework with Dual-Quaternion parameterization [Dan99] to eliminate the outliers. RS method is a RANSAC framework and different Hand-Eye calibration methods can be used within this framework. The goal of this experiment was not to demonstrate the performance of the Hand-Eye calibration methods but to compare the performance of the non-deterministic RANSAC outlier removal approach against our deterministic method. We used the following chain to recompute camera motions $\hat{A}_j$. Subscript $j = 1 \ldots n$ relates to all the motions corresponding to the A_i where $i = 1$.

$$\hat{A}_i = X^{-1} B_i B_j^{-1} X A_j. \tag{5.18}$$

Then, we compared $\hat{A}_{i=1}$ against the camera pose A_1 given by camera motions from the dataset. We evaluated the difference between the two rotations in angle metric and the relative translation using Euclidean distance. These are the same metrics used in [Fur+18]. The final rotation and translation error is the RMS error of all $\hat{A}_1$s computed with all corresponding A_js. We devise two different experiments with this dataset.

In the first experiment, we randomly drew 100 sets of 30 motions out of 1686 time-aligned motions *without* using the pre-selection algorithm from [Fur+18]. Using each sample of these 30 motions, we computed X. We compared the results using our kinematic chain (5.18). FIGURE 5.9 summarizes the results and shows that the overall performance is superior at every sample even without using the pre-selection. Note that, in FIGURE 5.9, *ransac* line is entirely flat because RS method selects the best 30 motions to produce one Hand-Eye calibration output against which we compare our results from 100 samples. In this comparison, especially in FIGURE 5.9 (left), our worst result obtained with some randomly obtained 30 motions is better than the result from RS method with pre-selected motions.

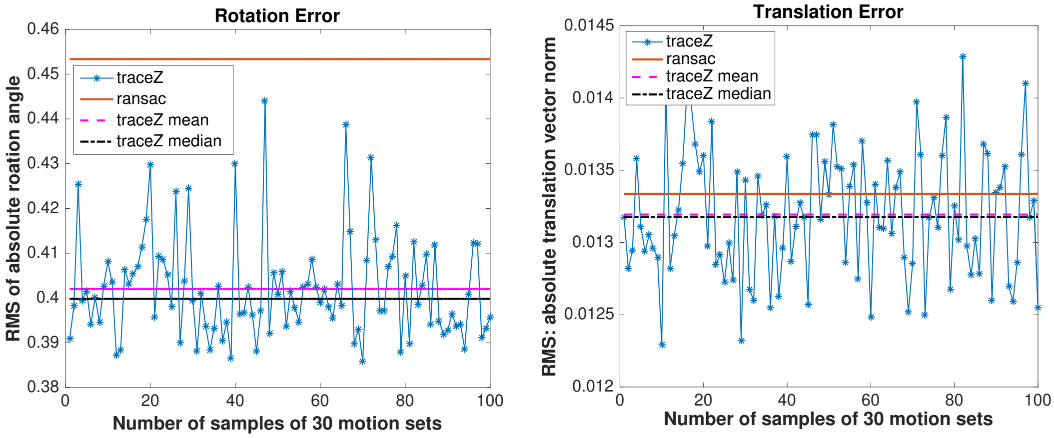

FIGURE 5.9: Real data from robot hand. Comparison against *ransac* method [Fur+18] using 30 pre-selected motions.

In the second experiment, we use 30 pre-selected motions to compute X, and we purposely add a percentage of outliers replacing some motions from the 30 motions. These outliers are generated by selecting random hand motions out of time-aligned motions and then pairing them up with camera motions with a big time shift (e.g hand-motion with time stamp 1 is paired with camera-motion with a time-stamp of 100). At 0% outliers, our method is using exactly the same data as its competitor. We compute the X with increasing number of outliers upto 80%. In FIGURE 5.10 we can see that out *traceZ* method manages to perform as well as the compared algorithm from [Fur+18] upto 70% of the outlier level. This experiment is meaningful in real-life situation in which the incoming data from the sensors is often time-wise misaligned. Authors in [Fur+18] have devised their method to align all the data with the provided time stamp before feeding it to their algorithm. This experiment shows that, if our method is provided with a very small amount of data correctly aligned with time-stamps (which can be achieved with hardware synchronization after a definite clock cycle), our method can overcome the rest of the time-misalignment to produce ideal results.

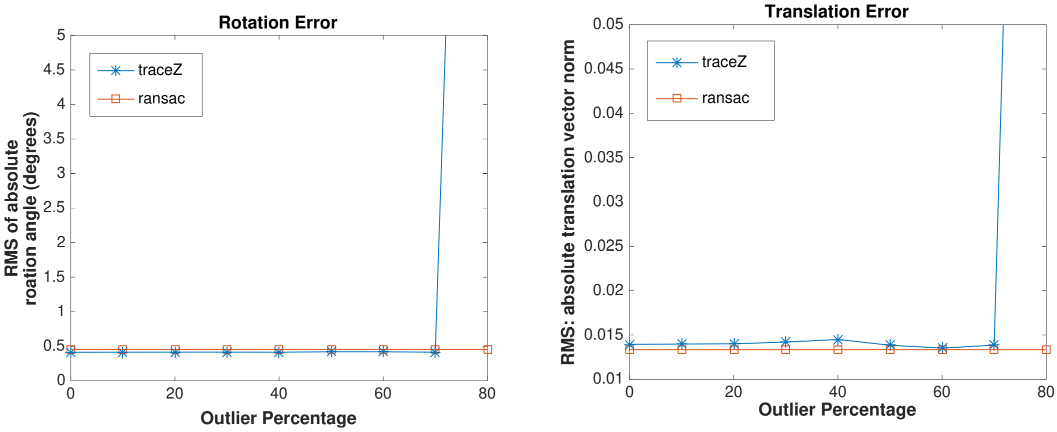

FIGURE 5.10: Real data from robot hand. A certain percentage of time-shifted outliers are added to replace good motions from the 30 pre-filtered motions.

5.3.3 Real Data Experiments with EM-Camera Setup

FIGURE 5.11 shows our own setup for Hand-Eye calibration experiment. In this setup, we used a checkerboard calibration pattern in the view of a calibrated camera providing us the camera transformations. We used NDI Aurora electromagnetic tracking system to generate hand transformations. This system consists of an EM transmitter, an EM receiver and a power unit. The EM receiver was mounted on the calibration pattern. Hence, the origin of the calibration pattern and the EM sensor (receiver) are related via a rigid transformation X. The transformation between the camera and EM transmitter was also fixed. We recorded a sequence of motions by moving the calibration pattern in front of the camera with the receiver being in the EM field of the transmitter at the same time. We avoided the cases of optical occlusions as well as rapid movements of checkerboard pattern to acquire time-synchronized pairs of camera-EM transformations. Our software can synchronize the image and EM-data acquisition while recording the data. We added high electromagnetic interference to the electromagnetic receiver for some of the EM transformations. This can be easily achieved by holding a varying electromagnetic field device such as a cell-phone (while calling) close to the EM receiver. The noise is extreme enough to distort the EM field near the EM receiver randomly. The Windows-based driver/software which comes with NDI Aurora system can display the sensor accuracy values in real-time.

We can observe (in real-time) the high disturbance being caused by cell-phone EM field, which helps us to record an 'outlier' value during data acquisition. As we only needed a randomly caused high EM disturbance, we can accept this dataset without accurate specifications of cell-phone EM field.

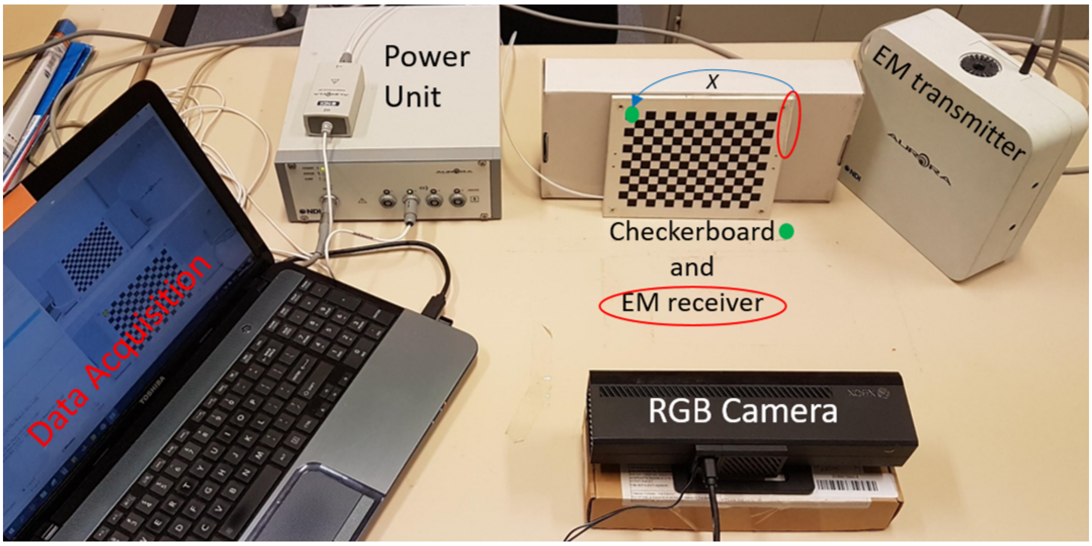

FIGURE 5.11: Our camera-EM data acquisition setup. The camera and EM transmitter remain fixed. With each motion, the checkerboard pattern moves in front of camera along with the attached EM receiver.

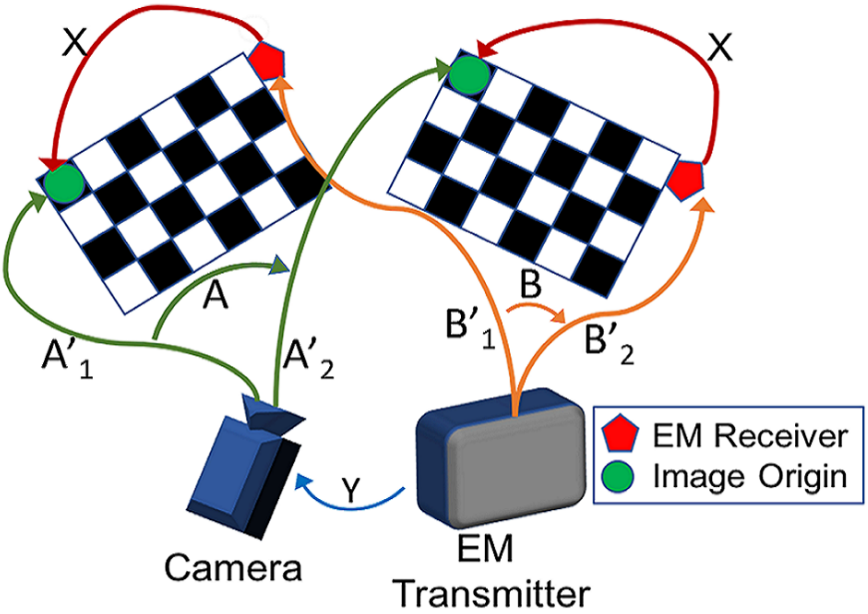

FIGURE 5.12: Schematic of EM sensor - camera setup to demonstrate Hand-Eye calibration. Contrary to the original Hand-Eye calibration setup, the camera is fixed and the calibration pattern moves: the chain of transformation remains equivalent.

FIGURE 5.12 shows the schematic of our EM-camera setup. There exists a Y fixed transformation between the camera and EM transmitter. This is equivalent to World-Robot Base calibration in the robot hand-camera setup. Estimation of this transformation is not necessary to compute Hand-Eye calibration of the checkerboard as long as the EM transmitter and camera are not moved during the experiment. Similarly to the experiments in section 5.3.2, we used the chain of transformations given in (5.18) with a small modification to compute the camera transformations. Instead of only using A_1, we used all the N motions A_is ($i = 1 \dots N$) for this experiment. Then we re-projected the computed pixels onto to the reference image using 2D homography equation $\hat{x} = Hx$, where H is 3×3 non-singular matrix given by the camera intrinsics and x are the coordinates of 2D points. FIGURE 5.13 shows the average pixel error (Euclidean distance) corresponding to all $\hat{A}_i$s computed from all A_js (5.18). In this experiment, we used 20 (for *Easy*) and 10 (for *Hard*) relative motions of the EM sensor(Hand)-checkerboard(Eye) setup. In FIGURE 5.13, *Easy* dataset is without outliers and in *Hard* dataset, 3 out of 10 of the motions are affected by EM interference outliers. We can see that in the *Easy* case our method performs fairly in comparison to the other methods (similarly seen in the Gaussian noise simulation experiment in FIGURE 5.2) and excels in eliminating the outliers in the *Hard* case. We compared our results against *Inria* [HD95], *Tsai* [TL89], *Navy* [PM94] and *Heller-dqhec* [HHP]. In comparison to synthetic experiment section 5.3.1, we left out a few methods because of their poor performance with this dataset.

This experimental setup entirely represents our proposed laparoscope tip-image calibration setup presented in section 3.3.2. The optical camera is same as the PSD-IMU setup providing the transformations of the laparoscope tip. Both setups have the EM sensor that provides the transformations of the laparoscope image. The obtained Hand-Eye calibration X provides the fixed calibration between the laparoscope tip and the image.

5.3.4 Additional Experiments with RANSAC

In the previous experiment section, we used a dataset in which the relative motions were provided. It is practical and also advised to pre-compute the relative motions using the absolute motion before feeding them as input to the Hand-Eye calibration algorithm. Experts are unanimous to approve this strategy. If we are provided with a dataset with absolute motions, apart from using them to compute the relative motions, we can also use them to form a pixel error based geometric threshold for a RANSAC algorithm. Current RANSAC algorithms for Hand-Eye calibration methods (such as in [Fur+18]) use algebraic thresholds. The algebraic distance measurement is a representation of the distance between two transformation matrices. In a transformation matrix, there are two main parts: translation and rotation. The distance between two translation vectors can be Euclidean. But distance between two rotations can vary depending upon the parameterization of the rotations used. At the same time, the distance computation can be susceptible to the presence of noise on rotation matrices. Along with these problems, we also have to devise a strategy to combine the distances

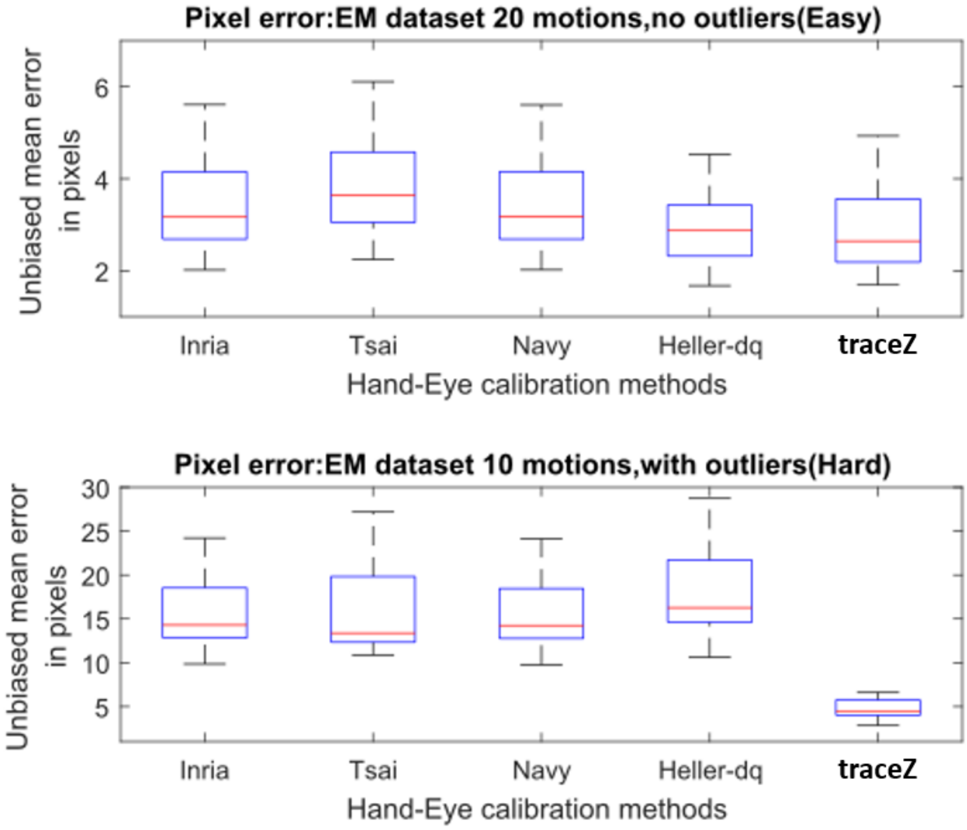

FIGURE 5.13: Dataset from EM-sensor setup. The graphs show the distribution of the pixel errors. In the box plot, the red central line indicates the median pixel error. The top and bottom blue lines of the box indicate 75th and 25th percentile respectively. The whiskers extend to the most extreme data points.

from the rotation part and the translation part. In their RANSAC algorithm, Furrer et al. [Fur+18] present a dual quaternion based distance measure to combine rotation and translation distances together. They use the dot product of the screw axis of the transformations and compare them with a threshold. This threshold and in general the algebraic threshold is very sensitive to the dataset, meaning that for each new experimental setup, the threshold needs to be adjusted.

To counter this issue, we propose a new type of threshold for the RANSAC-based Hand-Eye calibration algorithms results, which we published in [SHM19b]. Our 'geometric' threshold uses the average Euclidean distance between the pixels of the reprojected camera image. Another benefit of our algorithm is that instead of pre-computed relative motions we can use absolute robot-hand motions to compute the Hand-Eye calibration. If a given absolute robot hand motion is an oulier, pre-computing the relative motion using the outlier absolute motion and it's adjacent good absolute motion produces an outlier relative motion. If we filter out the oulier absolute motion in the first place using RANSAC algorithm, then we can obtain a dataset of outlier-free abolute motions using which we can make a dataset of pre-computed relative motions available to any Hand-Eye calibration algorithm.

Our RANSAC algorithm

Following are the steps involved in our RANSAC algorithm.

1. Choose a set of n random samples from the dataset of hand and camera motions. In our case, we select 4.
2. Using the sample of 4 absolute motions, create 3 relative motions.
3. Compute Hand-Eye calibration using the Hand-Eye calibration methods for this dataset (in our experiments methods from Dornaika and Horaud [DH98] and Heller, Henrion, and Pajdla [HHP]). Obtain the value of X.
4. Using X and absolute hand motions, re-project all the calibrated camera motions into each camera motion one-by-one. Compute the average pixel error between the obtained camera motion and the reprojected camera motion. Based on this 'geometric' threshold, determine the consistent inlier pairs eliminating the outliers.
5. Retain the largest inlier set of the absolute camera and hand motions.
6. After all the inliers are found, compute relative camera and hand motions using the inliers.
7. Re-compute the Hand-Eye calibration using the largest inlier dataset of the relative motions.

Geometric threshold

In step 3 of our RANSAC algorithm, we obtain Hand-Eye calibration X. Using X and absolute hand and camera motions, we can re-project all the pixels of each camera motion in using the following equation

$$\hat{A}_i = X^{-1} B_i B_j^{-1} X A_j \,. \tag{5.19}$$

In this equation, $\hat{A}$ is the camera image computed after re-projection. Index $i = 1, 2, \ldots, n$ is for selecting image for re-projection, where $n =$ total number of images. It is re-projected $j = 1, 2, \ldots, n$ times. With this strategy, we obtain i number of $\hat{A}$ versions of ground truth A_i. We compute root-mean-square (rms) 2D pixel error on each pixel in all the images. The resulting matrix containing the rms error values of all the i re-projections of j camera images clearly shows a large error for the images which are outliers. In our experiment, we obtained rms values below 20 for good images, and for outliers the values were greater than 500 for our dataset. With this great margin, we can comfortably set a 'geometric' threshold that is suited for a wide range of datasets.

Results

We test our algorithm with synthetic data experiments. The dataset is similar to the one in section 5.3.1. 31 random absolute camera motions and hand motions are generated. We introduce a nominal noise on the pixels of the camera motions and directly on the quaternions as well translation vectors of the hand motions. A certain percentage of the hand motions are replaced with outliers at each step (from 10% to 50%). The probability value of RANSAC is set to 0.99%. With this probability value, in the hardest case, RANSAC algorithm will keep running infinitely looking for a solution. Hence, we truncate the experiment's run-time to 90 seconds. Within this time, our RANSAC algorithm finds a solution using upto 50% of outlier infested dataset. We use Hand-Eye calibration methods from Heller, Henrion, and Pajdla [HHP] (*heller*) and Dornaika and Horaud [DH98] (*inria*) in this experiment and compare the results of these methods with and without using the RANSAC algorithm. FIGURE 5.14 shows the results of this experiment.

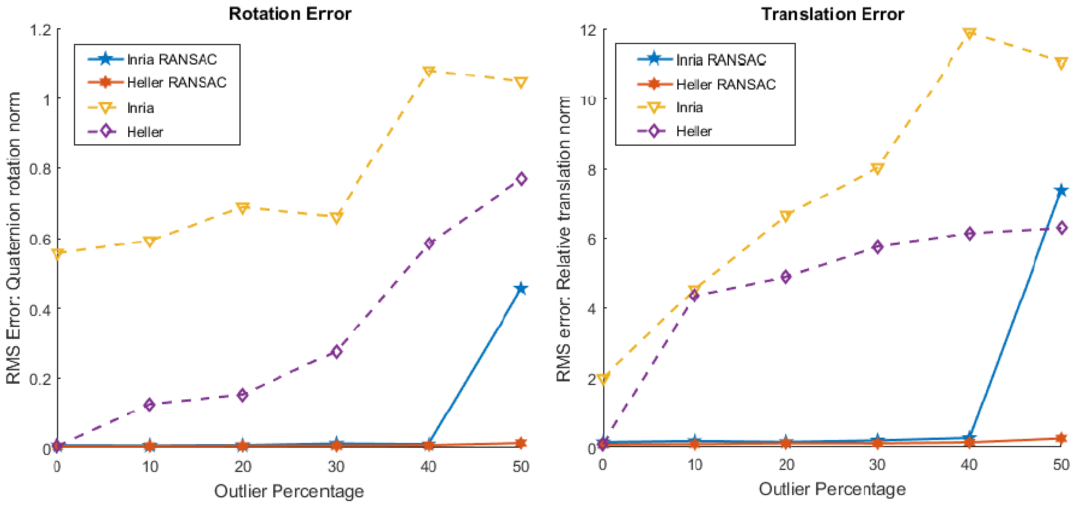

FIGURE 5.14: On the Y-axis, an increasing percentage of outliers is replacing the motions. Left: RMS of quaternion rotation errors. Right: RMS of relative translation vector errors.

FIGURE 5.15 shows the time performance of the RANSAC algorithm. We can see that upto 40% outliers, the algorithm computes results below 50 seconds. After this, the time increases exponentially.

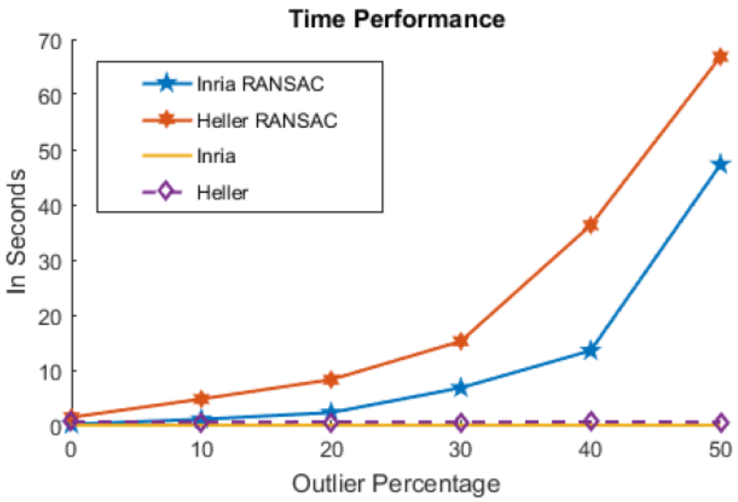

FIGURE 5.15: Average time taken for 50 iterations at each step of outlier percentage.

5.4 Summary

In this chapter, we presented a method to solve the Hand-Eye calibration problem. The data measurements from sensors such as EM or IMUs can be corrupted due to interference, time misalignment and/or communication errors. We have reformulated the problem using a SDP formulation in which the nonlinearity of the problem is confined to a single rank constraint. Based on this, we proposed an alternating procedure to cope with the nonlinearity and an iterative re-weighting scheme against the outliers. This is a deterministic and robust method that has provided excellent results in comparison to the existing ones. We tested this method extensively against other state-of-the-art methods using different types of experiments. At the end, we presented a real-data experiment setup which can be applied to the laparoscope tip-image calibration in a PSD-IMU sensor based ultrasound laparoscope tracking setup.

Chapter 6

Conclusion

In the recent years, laparoscopic surgery has been increasingly used to benefit from its numerous obvious advantages, particularly leading to less post-surgical trauma to the patient and shorter recovery times. Minimally invasive surgery procedures, such as laparoscopic surgery, entirely rely upon the video feed from an Endoscope that the surgeon sees on a screen. The improved quality of optical imaging involved in these surgeries have immensely contributed to the success of laparoscopic surgery. Parallelly, we have also witnessed significant progress in other types of imaging technologies such as ultrasound imaging, CT/MRI scans etc. These imaging techniques bring valuable additional image information to the surgeons based upon which they can plan their procedures more efficiently. These advances in the imaging techniques, along with the improvements in surgical tools, have paved a way to novel techniques that are unique to laparoscopic surgery. To display multimodal images overlayed on each other in real-time, one must know the exact pose of the surgical instruments at all times during their use. The focus of this thesis was to track the ultrasound laparoscope while addressing some of the core problems involved in tracking this instrument.

In the introduction of this thesis, we presented a general context of image-guided laparoscopic surgery. We explained the limits of a laparoscopic surgery from a surgeon's point of view and how the image guidance technology can address these limitations. We also presented the steps involved in image-guided surgery while identifying the fact that the tracking is the most critical part of the surgical process due to its reliance on sensors. Since the tracking sensors have their physical limitations, the tracking suffers from noise affecting the overall performance of image-guided surgery. We also introduced an ultrasound laparoscopic instrument which is used as a probe to the Siemens S3000 ultrasound machine. Tracking the ultrasound laparoscope allows eSieFusion software present on the Siemens S3000 machine to perform image fusion between pre-operative CT/MRI images with the real-time tracked ultrasound images.

We also presented a brief history of the development of laparoscopic instruments beginning from the nascent time of laparoscopic instruments, when were used for only for diagnosis, to the current era of image-guided surgery. Adding to it, we presented a general laparoscopic procedure to help understand the context in which a laparoscope is used. This effort was crucial to understand the surgeon's expectations from the image-guided surgery. Following to this, we also discussed the steps involved in a laparoscopic surgery in moderate details while focusing on the tracking step in the whole process. We presented the available tracking sensor technologies while discussing their advantages and limitations. Finally in this chapter, we presented the state-of-the-art hybrid technologies for laparoscope tracking that can utilise different sensor technologies combined to mitigate the limitations of one another.

After understanding the limitations of the available tracking systems, we proposed a novel setup for hybrid tracking based on Position Sensitive Detectors (PSDs) for optical tracking and Inertial Measurement Units (IMUs) as orientation sensors. We presented an overview of the schematics and functioning of the sensors while discussing the important steps involved in connecting all the sensors together to form a kinematic chain. We presented a technique to obtain the pose of the ultrasound image in the reference frame of the optical PSD setup considering it to be a stationary point in operation room. In this technique, we also identified the transformations that are affected by noise and suggested rotation averaging as a strategy to counter the effect of noise and drift on the dataset acquired from the IMUs.

We also presented the problem of rotation averaging along with the classification of rotation averaging methods. We identified that our problem of IMU data averaging falls under single rotation averaging case and provided a pre-existing globally optimal closed-form solution for the problem found in [Har+13]. Focusing onto the second important part involved in tracking, we presented the concept of Hand-Eye calibration relating it to rotation averaging. We underlined the importance of Hand-Eye calibration in the case of hybrid tracking in which the pose sensors are attached onto the fixed body of a laparoscope. Given that Hand-Eye calibration is used to compute the fixed body transformation of two rigidly attached pose sensors, it becomes an essential part of hybrid tracking. We presented a detailed overview of the hand-eye calibration literature along with the problem formulations of the important methods.

Building upon the Hand-Eye calibration literature, we present one of the main contributions of this thesis, that is, a novel method for Hand-Eye calibration. Our method is robust to the outliers occurring in the medical pose sensors such as electromagnetic sensors that are constantly affected by their surroundings producing completely erroneous readings in a dataset. After introducing the problem statement, we provided some background on Semi-Definite Programming, Rank-Constrained LMI feasibility problem and Convex-Hull of rotations: the concepts used in our method. We restated the Hand-Eye calibration problem based upon Semi-Definite programming formulation using LMIs. To ensure the rotation's membership to $SO(3)$, we solved the problem using an alternating SDP formulations. For robustness, we added weights to the problem formulation as to diminish the influence of outliers on the data and noise. We observed during rigorous trials that our method consistently converges to the sought solution for most practical cases. Through experiments, we tested our method against other Hand-Eye calibration methods using different type of datasets. In the end, we presented a RANSAC framework for Hand-Eye calibration methods that utilises a geometric threshold.

This thesis focused on two important problems related to the ultrasound laparoscope tracking. The problem of Hand-Eye calibration is inherent to the current state-of-the-art hybrid tracking solutions. We solved this non-convex problem providing a solution that is robust to noise and outliers. The semi-definite programming approach of solving this problem also confines the non-linearity of the problem to a single matrix rank constraint. Our solution addresses this problem in a unique way of alternating SDP approach making it the highlight of this thesis. In the future, we would like to explore the applicability of this technique to other problems in computer vision and Robotics. We would also like to extend our robust hand-eye calibration algorithm to the rotation averaging problem and test it against the existing rotation averaging methods.

The second contribution of this thesis elaborates the mechanics of pose sensors for the laparoscope. We proposed a sensor setup for the ultrasound laparoscope tracking that combines the IMU data to optical pose. We would like to realise this project to the point of real-world testing of the laparoscope. Due to the unavailability of the actual laparoscope, we did not manage to construct the setup for the laparoscope tracking. We would like to integrate this laparoscope in the Siemens S3000 machine system to perform image fusion between CT/MRI scan data and the real-time tracked ultrasound images. We would like to test different rotation averaging algorithms for the IMU data averaging. This laparoscope tracking technology is not unique to the ultrasound laparoscope. It can also be implemented on different types of laparoscopic instruments possessing a moving tip/end. The pose of a laparoscope tip obtained from tracking can be used for applications such as needle tracking for ablation procedure, endoscope image overlay etc.

The main contribution of this thesis, that is the Hand-Eye calibration method, received a good response from the international academic community. Likewise, I would like to receive a good feedback from the users of this new ultrasound laparoscope with sensors. This project exposed me to the world of academic research as well as gave me technical experience inside surgical rooms. While working with the surgeons and medical staff, I understood the value of medical instrumentation in saving human life. Therefore, I would like to see the entirety of the technology involved in designing this ultrasound laparoscope to be used by surgeons to save human lives. It indeed would be the pinnacle of this project.

Bibliography

[AHE99] Nicolas Andreff, Radu Horaud, and Bernard Espiau. "On-line hand-eye calibration". In: *3-D Digital Imaging and Modeling, 1999. Proceedings. Second International Conference on*. IEEE. 1999, pp. 430–436.

[BF00] G Berci and KA Forde. "History of endoscopy". In: *Surgical endoscopy* 14.1 (2000), pp. 5–15.

[Bir00] Wolfgang Birkfellner. *Tracking systems in surgical navigation*. na, 2000.

[Boy+94] Stephen Boyd et al. *Linear matrix inequalities in system and control theory*. Vol. 15. Siam, 1994.

[BZM94] Paul A Beardsley, Andrew Zisserman, and David W Murray. "Navigation using affine structure from motion". In: *European Conference on Computer Vision*. Springer. 1994, pp. 85–96.

[Che+17] Alexis Cheng et al. "Active phantoms: a paradigm for ultrasound calibration using phantom feedback". In: *Journal of Medical Imaging* 4.3 (2017), p. 035001.

[Che91] Homer H Chen. "A screw motion approach to uniqueness analysis of head-eye geometry". In: *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*. IEEE. 1991, pp. 145–151.

[CK91] Jack CK Chou and M Kamel. "Finding the position and orientation of a sensor on a robot manipulator using quaternions". In: *The international journal of robotics research* 10.3 (1991), pp. 240–254.

[CMS10] Lee Yik Ching, Knut Möller, and Jackrit Suthakorn. "Non-radiological colonoscope tracking image guided colonoscopy using commercially available electromagnetic tracking system". In: *2010 IEEE Conference on Robotics, Automation and Mechatronics*. IEEE. 2010, pp. 62–67.

[CS10] Song Cui and Yeng Chai Soh. "Linearity indices and linearity improvement of 2-D tetralateral position-sensitive detector". In: *IEEE Transactions on Electron Devices* 57.9 (2010), pp. 2310–2316.

[Dan99] Konstantinos Daniilidis. "Hand-eye calibration using dual quaternions". In: *The International Journal of Robotics Research* 18.3 (1999), pp. 286–298.

[Dat10] Jon Dattorro. *Convex optimization & Euclidean distance geometry*. Lulu. com, 2010.

[DBC96] Konstantinos Daniilidis and Eduardo Bayro-Corrochano. "The dual quaternion approach to hand-eye calibration". In: *proceedings of 13th International Conference on Pattern Recognition*. Vol. 1. IEEE. 1996, pp. 318–322.

[DH98] Fadi Dornaika and Radu Horaud. "Simultaneous robot-world and hand-eye calibration". In: *IEEE transactions on Robotics and Automation* 14.4 (1998), pp. 617–622.

[Emm67] RB Emmons. "The lateral photoeffect". In: *Solid State Electronics* 10 (1967), pp. 505–506.

[Feu+07] Marco Feuerstein et al. "Magneto-optic tracking of a flexible laparoscopic ultrasound transducer for laparoscope augmentation". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007*. Springer, 2007, pp. 458–466.

[Feu08] Marco Feuerstein. *Augmented Reality in Laparoscopic Surgery – New Concepts and Methods for Intraoperative Multimodal Imaging and Hybrid Tracking in Computer Aided Surgery*. Vdm Verlag Dr. Müller, 2008. ISBN: 978-3-8364-7783-3.

[FL05] Irene Fassi and Giovanni Legnani. "Hand to sensor calibration: A geometrical interpretation of the matrix equation AX= XB". In: *Journal of Robotic Systems* 22.9 (2005), pp. 497–506.

[Fra+03] Don D Frantz et al. "Accuracy assessment protocols for electromagnetic tracking systems". In: *Physics in Medicine & Biology* 48.14 (2003), p. 2241.

[Fur+18] Fadri Furrer et al. "Evaluation of Combined Time-Offset Estimation and Hand-Eye Calibration on Robotic Datasets". In: *Field and Service Robotics*. Springer. 2018, pp. 145–159.

[Gov04] Venu Madhav Govindu. "Lie-algebraic averaging for globally consistent motion estimation". In: *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.* Vol. 1. IEEE. 2004, pp. I–I.

[Gov06] Venu Madhav Govindu. "Robustness in motion averaging". In: *Asian Conference on Computer Vision*. Springer. 2006, pp. 457–466.

[Har+13] Richard Hartley et al. "Rotation averaging". In: *International journal of computer vision* 103.3 (2013), pp. 267–305.

[Hay+09] Caroline Hayhurst et al. "Application of electromagnetic technology to neuronavigation: a revolution in image-guided neurosurgery". In: *Journal of neurosurgery* 111.6 (2009), pp. 1179–1184.

[HD95] Radu Horaud and Fadi Dornaika. "Hand-eye calibration". In: *The international journal of robotics research* 14.3 (1995), pp. 195–210.

[HDK01] Robert L Hirsh, Guilherme N DeSouza, and Avinash C Kak. “An iterative approach to the hand-eye and base-world calibration problem”. In: *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No. 01CH37164)*. Vol. 3. IEEE. 2001, pp. 2171–2176.

[HHP] Jan Heller, Didier Henrion, and Tomas Pajdla. “Hand-eye and robot-world calibration by global polynomial optimization”. In: *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. IEEE, pp. 3157–3164.

[HHP16] Jan Heller, Michal Havlena, and Tomas Pajdla. “Globally optimal hand-eye calibration using branch-and-bound”. In: *IEEE transactions on pattern analysis and machine intelligence* 38.5 (2016), pp. 1027–1033.

[HR03] William R Hendee and E Russell Ritenour. *Medical imaging physics*. John Wiley & Sons, 2003.

[HS04] Richard Hartley and Frederik Schaffalitzky. “L/sub/spl infin//minimization in geometric reconstruction problems”. In: *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.* Vol. 1. IEEE. 2004, pp. I–I.

[HS97] Richard I Hartley and Peter Sturm. “Triangulation”. In: *Computer vision and image understanding* 68.2 (1997), pp. 146–157.

[Jim+18] David Fuentes Jiménez et al. “Deep Shape-from-Template: Wide-Baseline, Dense and Fast Registration and Deformable Reconstruction from a Single Image”. In: *arXiv preprint arXiv:1811.07791* (2018).

[KH08] Fredrik Kahl and Richard Hartley. “Multiple-View Geometry Under the L-∞ Norm”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30.9 (2008), pp. 1603–1617.

[Kha+00] Rasool Khadem et al. “Comparative tracking error analysis of five different optical tracking systems”. In: *Computer Aided Surgery* 5.2 (2000), pp. 98–107.

[Kim+10] Sin-Jung Kim et al. “Robot head-eye calibration using the minimum variance method”. In: *2010 IEEE International Conference on Robotics and Biomimetics*. IEEE. 2010, pp. 1446–1451.

[KSB06] Stefan R Kirsch, Christian Schilling, and Georg Brunner. “Assesment of metallic distortions of an electromagnetic tracking system”. In: *Medical Imaging 2006: Visualization, Image-Guided Procedures, and Display*. Vol. 6141. International Society for Optics and Photonics. 2006, 61410J.

[Kui80] Jack B Kuipers. “SPASYN-an electromagnetic relative position and orientation tracking system”. In: *IEEE Transactions on Instrumentation and Measurement* 29.4 (1980), pp. 462–466.

[Las01] Jean B Lasserre. “Global optimization with polynomials and the problem of moments”. In: *SIAM Journal on optimization* 11.3 (2001), pp. 796–817.

[LC95] Ying-Cherng Lu and Jack CK Chou. “Eight-space quaternion approach for robotic hand-eye calibration”. In: *1995 IEEE International Conference on Systems, Man and Cybernetics. Intelligent Systems for the 21st Century*. Vol. 4. IEEE. 1995, pp. 3316–3321.

[LDM97] Daniel F Leotta, Paul R Detmer, and Roy W Martin. “Performance of a miniature magnetic position sensor for three-dimensional ultrasound imaging”. In: *Ultrasound in Medicine and Biology* 23.4 (1997), pp. 597–609.

[Lit99] Grzegorz S Litynski. “Endoscopic surgery: the history, the pioneers”. In: *World journal of surgery* 23.8 (1999), pp. 745–753.

[LLL97] WY Lau, CK Leow, and Arthur KC Li. “History of endoscopic and laparoscopic surgery”. In: *World journal of surgery* 21.4 (1997), pp. 444–453.

[LM08] Rong-hua Liang and Jian-fei Mao. “Hand-eye calibration with a new linear decomposition algorithm”. In: *Journal of Zhejiang University-SCIENCE A* 9.10 (2008), pp. 1363–1368.

[Löf04] J. Löfberg. “YALMIP : A Toolbox for Modeling and Optimization in MATLAB”. In: *In Proceedings of the CACSD Conference*. Taipei, Taiwan, 2004.

[Luc60] Gerald Lucovsky. “Photoeffects in nonuniformly irradiated p-n junctions”. In: *Journal of Applied Physics* 31.6 (1960), pp. 1088–1095.

[LWW10] Aiguo Li, Lin Wang, and Defeng Wu. “Simultaneous robot-world and hand-eye calibration using dual-quaternions and Kronecker product”. In: *International Journal of Physical Sciences* 5.10 (2010), pp. 1530–1536.

[Mar+03] Sandra Martelli et al. “Comparison of an optical and a mechanical navigation system”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2003, pp. 303–310.

[Mår+05] R Mårvik et al. “Image-guided laparoscopic surgery. Review and current status.” In: *Minerva chirurgica* 60.5 (2005), pp. 305–325.

[Mar+12] Primoz Markelj et al. “A review of 3D/2D registration methods for image-guided interventions”. In: *Medical image analysis* 16.3 (2012), pp. 642–661.

[Mar63] Donald W Marquardt. “An algorithm for least-squares estimation of nonlinear parameters”. In: *Journal of the society for Industrial and Applied Mathematics* 11.2 (1963), pp. 431–441.

[Mer04] Laurence Mercier. “Review of Ultrasound Probe Calibration Techniques for 3D Ultrasound”. PhD thesis. McGill University, 2004.

[MHJ10] Jianfei Mao, Xianping Huang, and Li Jiang. “A flexible solution to AX= XB for robot hand-eye calibration”. In: *Proceedings of the 10th WSEAS international conference on Robotics, control and manufacturing technology. World Scientific and Engineering Academy and Society (WSEAS)*. 2010, pp. 118–122.

[MP07] Daniel Martinec and Tomas Pajdla. “Robust rotation and translation estimation in multiview reconstruction”. In: *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 2007, pp. 1–8.

[Nix+98] Mark A Nixon et al. “The effects of metals and interfering fields on electromagnetic trackers”. In: *Presence* 7.2 (1998), pp. 204–218.

[Ogu+14] Ryo Oguma et al. “Ultrasound image overlay onto endoscopic image by fusing 2D-3D tracking of laparoscopic ultrasound probe”. In: *Augmented Environments for Computer-Assisted Interventions*. Springer, 2014, pp. 14–22.

[OT14] Francisco PM Oliveira and Joao Manuel RS Tavares. “Medical image registration: a review”. In: *Computer methods in biomechanics and biomedical engineering* 17.2 (2014), pp. 73–93.

[PA02] François Poulin and L-P Amiot. “Interference during the use of an electromagnetic tracking system under OR conditions”. In: *Journal of biomechanics* 35.6 (2002), pp. 733–737.

[Pau+15] Danda Pani Paudel et al. “LMI-based 2D-3D registration: From uncalibrated images to Euclidean scene.” In: *CVPR*. 2015, pp. 4494–4502.

[PM94] Frank C Park and Bryan J Martin. “Robot sensor calibration: solving AX= XB on the Euclidean group”. In: *IEEE Transactions on Robotics and Automation* 10.5 (1994), pp. 717–721.

[Raa+79] Frederick H Raab et al. “Magnetic position and orientation tracking system”. In: *IEEE Transactions on Aerospace and Electronic systems* 5 (1979), pp. 709–718.

[Rei13] Tobias Reichl. “Advanced Hybrid Tracking and Navigation for Computer-Assisted Interventions”. PhD thesis. Technische Universität München, 2013.

[Rém+97] Sandrine Rémy et al. “Hand-eye calibration”. In: *Proceedings of the 1997 IEEE/RSJ International Conference on Intelligent Robot and Systems. Innovative Robotics for Real-World Applications. IROS'97*. Vol. 2. IEEE. 1997, pp. 1057–1065.

[SA89] Yiu Cheung Shiu and Shaheen Ahmad. “Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form AX= XB”. In: *ieee Transactions on Robotics and Automation* 5.1 (1989), pp. 16–29.

[Sam+19] Chinmay Samant et al. “Robust Hand-Eye Calibration via Iteratively Re-weighted Rank-Constrained Semi-Definite Programming”. In: *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2019, pp. 4482–4489.

[Sch+05] Kurt Schicho et al. “Stability of miniature electromagnetic tracking systems”. In: *Physics in Medicine & Biology* 50.9 (2005), p. 2089.

[SEH12] Mili Shah, Roger D Eastman, and Tsai Hong. "An overview of robot-sensor calibration methods for evaluation of perception systems". In: *Proceedings of the Workshop on Performance Metrics for Intelligent Systems*. ACM. 2012, pp. 15–20.

[Sei+00] PG Seiler et al. "A novel tracking technique for the continuous precise measurement of tumour positions in conformal radiotherapy". In: *Physics in Medicine & Biology* 45.9 (2000), N103.

[SH06a] Kristy Sim and Richard Hartley. "Recovering Camera Motion Using L\infty Minimization". In: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*. Vol. 1. IEEE. 2006, pp. 1230–1237.

[SH06b] Klaus H Strobl and Gerd Hirzinger. "Optimal hand-eye calibration". In: *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*. IEEE. 2006, pp. 4647–4653.

[Sha13] Mili Shah. "Solving the robot-world/hand-eye calibration problem using the Kronecker product". In: *Journal of Mechanisms and Robotics* 5.3 (2013), p. 031007.

[She+19] Chunxu Shen et al. "A method for ultrasound probe calibration based on arbitrary wire phantom". In: *Cogent Engineering* 6.1 (2019), p. 1592739.

[SHM19a] C. Samant, A. Habed, and M. de Mathelin. "Hand-Eye Calibration for Surgeries using Ultrasound Laparoscope". In: *Congres National d'Imagerie du Vivant - CNIV*. 2019.

[SHM19b] C. Samant, A. Habed, and M. de Mathelin. "Robust laparoscope body calibration using hand-eye calibration methods". In: *9th Joint Workshop on New Technologies for Computer/Robot Assisted Surgery, Genoa, Italy*. 2019.

[SPW15] James Saunderson, Pablo A Parrilo, and Alan S Willsky. "Semidefinite descriptions of the convex hull of rotation matrices". In: *SIAM Journal on Optimization* 25.3 (2015), pp. 1314–1343.

[SSS11] Raman Sanyal, Frank Sottile, and Bernd Sturmfels. "Orbitopes". In: *Mathematika* 57.2 (2011), pp. 275–314.

[Stu99] J.F. Sturm. "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones". In: *Optimization Methods and Software* 11–12 (1999). Version 1.05 available from `http://fewcal.kub.nl/sturm`, pp. 625–653.

[TL89] Roger Y Tsai and Reimar K Lenz. "A new technique for fully autonomous and efficient 3D robotics hand/eye calibration". In: *IEEE Transactions on robotics and automation* 5.3 (1989), pp. 345–358.

[Tyn+15] Chiang J Tyng et al. "Computed tomography-guided percutaneous core needle biopsy in pancreatic tumor diagnosis". In: *World Journal of Gastroenterology: WJG* 21.12 (2015), p. 3579.

[VB00] Jeremy G. VanAntwerp and Richard D. Braatz. "A tutorial on linear and bilinear matrix inequalities". In: *Journal of Process Control* 10.4 (2000), pp. 363 –385. ISSN: 0959-1524.

[VB97] Lieven Vandenberghe and Venkataramanan Balakrishnan. "Algorithms and software for LMI problems in control". In: *IEEE Control Systems* 17.5 (1997), pp. 89–95.

[Wag+02] A Wagner et al. "Quantitative analysis of factors affecting intraoperative precision and stability of optoelectronic and electromagnetic tracking systems". In: *Medical physics* 29.5 (2002), pp. 905–912.

[WAH98] Guo-Qing Wei, Klaus Arbter, and Gerd Hirzinger. "Active self-calibration of robotic eyes and hand-eye relationships with model identification". In: *ieee Transactions on Robotics and Automation* 14.1 (1998), pp. 158–166.

[Wal57a] J Torkel Wallmark. "A new semiconductor photocell using lateral photoeffect". In: *Proceedings of the IRE* 45.4 (1957), pp. 474–483.

[Wal57b] J Torkel Wallmark. "Photocell measures light direction". In: *Electronics* 1165 (1957).

[Wan92] C-C Wang. "Extrinsic calibration of a vision sensor mounted on a robot". In: *ieee Transactions on Robotics and Automation* 8.2 (1992), pp. 161–175.

[Wen16] Christian Wengert. *Hand-Eye calibration implementation*. `https://github.com/christianwengert/calib_toolbox_addon`. 2016. URL: `https://github.com/christianwengert/calib_toolbox_addon`.

[Yan+09] Ziv Yaniv et al. "Electromagnetic tracking in the clinical environment". In: *Medical physics* 36.3 (2009), pp. 876–892.

[YC06] Ziv Yaniv and Kevin Cleary. "Image-guided procedures: A review". In: *Computer Aided Interventions and Medical Robotics* 3 (2006), pp. 1–63.

[Zha11] Zijian Zhao. "Hand-eye calibration using convex optimization". In: *IEEE International Conference on Robotics and Automation (ICRA), 2011*. IEEE. 2011, pp. 2947–2952.

[ZL06] Zijian Zhao and Yuncai Liu. "Hand-eye calibration based on screw motions". In: *18th International Conference on Pattern Recognition (ICPR'06)*. Vol. 3. IEEE. 2006, pp. 1022–1026.

[ZRS94] Hanqi Zhuang, Zvi S Roth, and Raghavan Sudhakar. "Simultaneous robot/world and tool/flange calibration by solving homogeneous transformation equations of the form AX= YB". In: *IEEE Transactions on Robotics and Automation* 10.4 (1994), pp. 549–554.

[ZS92] Hanqi Zhuang and Yiu Cheung Shiu. “A noise tolerant algorithm for wrist-mounted robotic sensor calibration with or without sensor orientation measurement”. In: *Intelligent Robots and Systems, 1992., Proceedings of the 1992 lEEE/RSJ International Conference on*. Vol. 2. IEEE. 1992, pp. 1095–1100.

[ZW11] He Zhao and Zheyao Wang. “Motion measurement using inertial sensors, ultrasonic sensors, and magnetometers with extended kalman filter for data fusion”. In: *IEEE Sensors Journal* 12.5 (2011), pp. 943–953.

[ZZ04] Rong Zhu and Zhaoying Zhou. “A real-time articulated human motion tracking using tri-axis inertial/magnetic sensors package”. In: *IEEE Transactions on Neural systems and rehabilitation engineering* 12.2 (2004), pp. 295–302.

Université de Strasbourg

msii ÉCOLE DOCTORALE

Chinmay SAMANT

Ultrasound Laparoscopic Guidance for Minimally Invasive Surgery, Biopsy, and Ablation Procedures

Résumé

La chirurgie laparoscopique minimalement invasive guidée par l'image permet la réduction de la durée des séjours à l'hôpital pour le patient, réduisant ainsi les traumatismes postopératoires et accélérant le temps de guérison. Avec les progrès récents des techniques d'imagerie, les chirurgiens peuvent planifier une chirurgie de manière efficace et en toute confiance en utilisant différentes modalités d'image telles que la tomodensitométrie / IRM, les images échographiques, etc. Les techniques de fusion d'images en temps réel permettent la superposition de différents types d'images pour fournir une vue complète au chirurgien. Un aspect important de la fusion en temps réel est que l'instrument laparoscopique est suivi en temps réel à l'aide de capteurs. Dans cette thèse, nous présentons une analyse détaillée de ces technologies de suivi tout en fournissant une nouvelle configuration de capteurs pour le suivi d'images par laparoscope à ultrasons. Nous présentons une chaîne cinématique pour la configuration des capteurs et nous fournissons une solution pour la réduction du bruit présent dans les données des capteurs en utilisant la technique de moyennage des rotations. Le Hand-Eye calibration (étalonnage main-œil) est également un élément fondamental des systèmes de suivi hybrides. Nous présentons une révision détaillée de cette technique. Nous présentons également une méthode déterministe, robuste et précise pour résoudre le problème d'étalonnage main-œil, même pour de grandes quantités de valeurs aberrantes et des niveaux élevés de bruit de mesure. La méthode proposée est basée sur une reformulation d'un problème de programmation semi-définie à contraintes de rang où la robustesse est renforcée via une approche d'optimisation pondérée de façon itérative.

Mots-clés : Hand-Eye Calibration, Robotics, Hybrid-Tracking, Laparoscopes.

Abstract

Minimally invasive image-guided laparoscopic surgery allows shorter hospital stays for the patient reducing post-operative trauma and faster healing time. With the recent advances in imaging techniques, surgeons can efficiently and confidently plan a surgery by using different image modalities such as CT/MRI scans, ultrasound images etc. Real-time image fusion techniques can overlay the images from different modalities together to provide a comprehensive view to the surgeon. An important aspect of real-time fusion is that the laparoscopic instrument is tracked in real-time using sensors. In this thesis, we present a detailed analysis of such tracking technologies while providing a novel sensor setup for ultrasound laparoscope image tracking. We present a kinematic chain for the sensor setup and provide a solution for noise reduction in the sensor data using rotation averaging technique. Hand-Eye calibration is also a fundamental part of hybrid tracking systems. We present a detailed review of this technique. We also present a deterministic, robust and accurate method for solving Hand-Eye calibration problem even for large amounts of outliers and high levels of measurement noise. The proposed method is based on a reformulation of a rank-constrained semi-definite programming problem allowing for robustness to be enforced via an iteratively re-weighted optimization approach.

Keywords: Hand-Eye Calibration, Robotics, Hybrid-Tracking, Laparoscopes.