Inter-operative biopsy site relocalization in gastroscopy : application to oesophagus

Anant Suraj Vemuri

UNIVERSITY OF NICE SOPHIA ANTIPOLIS
DOCTORAL SCHOOL STIC
(INFORMATION AND COMMUNICATION TECHNOLOGIES AND SCIENCES)

T H E S I S

to obtain the title of

PhD of Science

of the University of Nice - Sophia Antipolis
Specialty : AUTOMATION, SIGNAL AND IMAGE PROCESSING

Prepared by
Anant Suraj VEMURI

Inter-Operative Biopsy Site Relocalization in Gastroscopy: Application to Oesophagus

Thesis Advisors: Nicholas AYACHE and Luc SOLER
prepared at IHU, Strasbourg and INRIA, Sophia Antipolis, ASCLEPIOS Team
defended on April 26th, 2016

Jury :

Reviewers :	Lena MAIER-HEIN	-	Deutsches Krebsforschungszentrum (DKFZ)
	Tom VERCAUTEREN	-	Univeristy College London
President :	Nassir NAVAB	-	Technische Universität München
Advisor :	Nicholas AYACHE	-	INRIA (Asclepios)
Co-Advisor :	Luc SOLER	-	IRCAD (Strasbourg)
Examiners :	Stéphane NICOLAU	-	IRCAD (Strasbourg)
Invited :	Michel DELVAUX	-	CHRU (Strasbourg)
	Lee L. SWANSTROM	-	The Oregon Clinic & IHU Strasbourg

UNIVERSITÉ DE NICE SOPHIA ANTIPOLIS

ÉCOLE DOCTORALE STIC

(SCIENCES ET TECHNOLOGIES DE L'INFORMATION ET DE LA COMMUNICATION)

T H È S E

pour obtenir le titre de

Docteur en Sciences

de l'université de Nice - Sophia Antipolis

Mention : AUTOMATIQUE, TRAITEMENT DU SIGNAL ET DES IMAGES

Preparé par

Anant Suraj VEMURI

Relocalisation de site de biopsie en gastroscopie: application à l'œsophage

Thèse dirigée par Nicholas AYACHE et Luc SOLER

préparée à IHU, Strasbourg et INRIA, Sophia Antipolis, projet ASCLEPIOS

soutenue le 26 Avril, 2016

Jury :

Rapporteurs :	Lena MAIER-HEIN	-	Deutsches Krebsforschungszentrum (DKFZ)
	Tom VERCAUTEREN	-	Univeristy College London
Président :	Nassir NAVAB	-	Technische Universität München
Directeur :	Nicholas AYACHE	-	INRIA (Asclepios)
Co-Directeur :	Luc SOLER	-	IRCAD (Strasbourg)
Examinateurs :	Stéphane NICOLAU	-	IRCAD (Strasbourg)
Invités :	Michel DELVAUX	-	CHRU (Strasbourg)
	Lee L. SWANSTROM	-	The Oregon Clinic & IHU Strasbourg

To Mom, Dad and Baba.

Acknowledgments

Foremost, I would like to thank my advisor, Prof. Luc Soler for providing the opportunity to work with IHU/IRCAD and establishing the environment for conducting my work. Luc, you energy levels never seized to amaze me. Despite an intense schedule you always took time to follow my work. Your guidance, advice and motivation were invaluable. Your vision of the future and ability to present those ideas, inspire me tremendously.

Then, I would like to express my gratitude to Dr. Stéphane Nicolau, my immediate supervisor and my dear friend. Thank you for taking such an large amount of time to guide my work. You channelled my energies in the right direction, whenever I exploded with ideas. It was a pleasure working with you.

To my advisor, Prof. Nicholas Ayache, it was an honour working with you. Thank you for accepting me into the asclepios team. Although much of the project related work, you advanced to Luc and Stéphane; from time to time you provided the necessary environment to stimulate my thinking. Whether it was through organizing visits to INRIA to meet the team or the wonderful Auron retreats, where I met the most brilliant scientific minds every year.

To Dr. Adrien Sportes, my partner in crime from the clinical front, your contribution in my work cannot be understated. Thank you for those long and fruitful discussions on gastro-intestinal procedures, I gained a lot of insight into the clinical aspect through you. You are the right clinical partner an engineer could have wished for.

The reviewers of my thesis, Dr. Lena Maier-Hein and Dr. Tom Vercauteren; your insights and comments were invaluable. Thank you for taking the time to make yourselves available for the review process. To Dr. Michel Delvaux and Dr. Lee Swanstrom for their clinical assessment of my work. In medical technology, the opinions of key medical experts, such as yourselves, who are the end users of our work, is most important. Thank you all for accepting to be part of my thesis Jury.

To the IRCAD/IHU development team (current and past); Julien Waechter, Emilie Harquel, Pascal Monnier, Nicolas Philipps, Jessica Gromer, Johan Moreau, Frédéric Champ, Arnaud Charnoz, Marc Schweitzer and Flavien Bridault. I would like to extend my deepest gratitude for working with me to transfer the approach presented in this thesis to clinical application and building a professional looking software that could be used in the operating room. I would like to also thank Jessica especially, for taking the time to participate in the clinical experiments. Your efforts were very much appreciated. Pascal, you are a very good friend and I will miss our morning tea and coffee sessions. To Dr. Oscar Garcia, I would like to say thank you. Your presence helped me understand the mathematics behind the Lie algebra and our discussions on calibration, error computation and propagation were very insightful. Your critiques made me delve deeper into the subject. Jalal Chaabane, who worked as an intern for an year and is an amazing programmer. You made my job of interfacing optical tracker over the network with my desktop easier through your application. Without it, the quantitative evaluation of the system would have been impossible. Thank you so much. Dr. Vincent Agnus, my neighbour at Virtual Surg for 3 years, I cannot understate your contribution through discussions, suggestions and guidance. You spent nearly 8 hours over a span of 10 days to help formulate my slides for the defence, even when you were

busy with things. Thank you very much Vince, I will miss your company.

To Dr. Xavier Pennec, a tireless mathematician. Your ability to model the most intricate parts of the surgical application was amazing. You helped me formulate a simplified trajectory based registration approach and understand the concept of error propagation in presence of isotropic noise.

Pamela Lhote, you are an amazing person and were super helpful throughout. From your arranging meetings with potential rentals when I first arrived in Strasbourg, to coordinating with everyone until my final defence, your role in this thesis completion, though behind the scenes, did not go unnoticed. Clélia Kinderstuth, you are another one of those behind the scenes person who organized so many things during my employment with IHU. Thank you!

From the IHU radiology team Rodrigo Cararo, Mourad Bouhadjar and Gael Fourré, and , who organized the experiment room professionally and facilitated Adrien and I, to perform experiments with ease, I appreciate the work you guys do, thank you! To the staff at the endoscopy unit at NHC, Strasbourg for cooperating with us to conduct our data collection. It would have been quite challenging without your support.

Then to Prof. Jacques Marescaux and Jean-Luc Dimarcq for supporting my research through IRCAD and IHU foundation.

I would like to thank the patients who consented to us for collecting data using the electromagnetic tracker. To the 5 pigs on whom I perform the animal trials, I apologize for the pain you may have been put through. I extend my sincere gratitude for your sacrifice. To the printer, the trees (for the paper) and to Google scholar for supporting my work silently.

To the members of the asclepios team, thank you for making me feel home every time I visited. Dr. Chloé Audigier, thank you for being so helpful in clarifying questions about UNice from time to time and just being such a nice person. Isabelle Strobant, you were amazing in coordinating things with UNice and INRIA for me. I really appreciate your help and support. To Regine Saelens from EDSTIC, thank you for being patient with me and replying to my “english” emails with care. My dear friend, Dr. Debarshi Dey, thank you for the last minute clarification for the calculation of statistics for my presentation.

To my family, for their never ending support, through the toughest of times. My Mom, Sister, Uncles, Aunts, Grandma and brother-in-law who have stood by me and given me the confidence to continue striving. My Dad and Grandpa, who passed on before they could see me complete, you have both been pillars in my life and your presence will be sorely missed.

Last and the most important acknowledgement of all, is the contribution and sacrifice of my wife, Archana Vemuri, who gave up so much to be with me and help me complete my work. Without your support, this would not have been possible. Thank you from the bottom of my heart.

Abstract

Minimally invasive surgery in gastrointestinal (GI) endoscopy has evolved from being a diagnostic tool to a therapeutic solution. It is common that GI procedures involve periodic monitoring or surveillance of the internal anatomy. Specifically in oesophageal procedures (the target of this thesis), surveillance interventions involve obtaining biopsies at different regions along the oesophagus. The tracking and relocalization of these biopsy sites "inter-operatively" poses a significant challenge for providing targeted treatments.

This thesis, clarifies the concept of relocalization, and analyses the need for a platform to aide GI endoscopy. Then, based on the understanding of the clinical context in oesophageal procedures; a novel framework to use electromagnetic tracking system is proposed, which is used to perform a "recording" of an intervention. This framework and the recording is then used to provide a guided navigation to the GI expert, during a follow-up surveillance endoscopy; for accurate re-positioning of the endoscope at previously targeted sites. This is achieved using inter-operative video synchronization, and the various steps involved in achieving this are described in this thesis.

Following the description of system design and the methodology, a careful analysis of noise affecting the system is performed. Using a Gaussian noise model, quantitative evaluation of the system is performed with synthetic and real (porcine) data. The results indicate that the relocalization was achieved with an uncertainty in depth inside the oesophagus of $\pm$10mm, which was considered acceptable for the GI expert. Additionally, qualitative experiments were performed using data from pigs, to simulate the real task of biopsy site relocalization, which was evaluated by 10 GI experts. The results of these experiments showed an improvement in biopsy site relocalization rate from 47.5% to 94%, thus clearly demonstrating the benefits of the proposed system towards assisted guidance. Furthermore, an incremental improvement in inter-operative video synchronization is proposed, that uses additional information obtained during the course of the intervention. Synthetic experiments indicated, inclusion of this additional information reduced the error in video synchronization by $\sim 50\%$.

This framework, was then extended by proposing a constrained inter-operative image matching, for further improvement in quality of video synchronization. Within this context, the effect of, the choice of feature descriptors and colour-space, filtering of uninformative frames and the endoscopic modality in use, are investigated and shown that further improvement is achieved in image synchronization to $[92\%, 87\%]$ from $[73\%, 76\%]$ for both narrow band imaging and white-light endoscopic modalities.
This research work has been implemented in a software (at IHU and IRCAD, Strasbourg), allowing us to validate our results in clinical conditions. This work was supported by IHU Strasbourg, through a grant# ANR-01AHU-02.

Keywords: Gastro-Intestinal Endoscopy, Inter-operative relocalization, Electro-magnetic tracking

Résumé:

La chirurgie mini-invasive en endoscopie gastro-intestinale a évolué d'un outil de diagnostic à une solution thérapeutique. En règle générale, les procédures gastro-intestinales impliquent un contrôle ou une surveillance périodique de l'anatomie interne. Dans le contexte des interventions de l'œsophage, les exploration de de surveillance implique la réalisation de multiples biopsies régulièrement le long de l'œsophage. Ainsi, les défis les plus importants auxquels il faut faire face au cours de ces procédures sont le suivi et la relocalisation inter-opératoires de ces sites de biopsies (pour un même patient opéré plusieurs fois). L'objectif de cette thèse est de proposer une solution informatisée afin de guider le gastroentérologue pendant de telles procédures.

Cette thèse précise tout d'abord le concept de relocalisation et analyse la nécessité d'une plate-forme pour aider l'endoscopie gastro-intestinale. Ensuite, après une analyse des procédures de l'œsophage et de leurs besoins, nous proposons un cadre novateur utilisant un système de suivi électromagnétique pour réaliser des enregistrements d'intervention de l'œsophage, couplant la vidéo à la profondeur de l'endoscope inséré. Ces enregistrements sont utilisés pour fournir une navigation aux gastroentérologues pendant des procédures de surveillances afin de repositionner l'endoscope de façon précise sur des sites de biopsie préalablement ciblés. Cette navigation consiste en une synchronisation vidéo entre la vue endoscopique courante et celles des surveillances endoscopiques précédentes enregistrées.

Une première version de notre système est présentée, associée à une analyse rigoureuse des bruits affectant le système. Cette évaluation incrémentale est réalisée sur des données d'abord synthétiques puis réelles recueillies sur des cochons. Les résultats montrent que la relocalisation est obtenue avec une précision de l'ordre de ±10mm, considérée comme largement acceptable par les experts. En outre, une expérience qualitative a été conçue à partir de données provenant de cochons pour simuler une tâche réelle de relocalisation de site de biopsie, et a été évaluée par 10 gastroentérologues. Celle-ci a clairement démontré l'avantage du système de guidage assisté en améliorant le taux de prélèvement de sites de biopsie de 47,5% à 94%. Une seconde version est alors proposée, utilisant la trajectoire 3D complète de l'œsophage acquise pendant l'intervention pour synchroniser les vidéos. Cette version permet d'éviter les erreurs importantes dues à des facteurs humains et permet de réduire l'erreur globale du système qui est améliorée d'environ 50%.

Ce cadre est finalement étendue afin d'améliorer encore la précision de la relocalisation à partir d'une sélection optimale de l'image vidéo pré-enregistrée dont le point de vue est le plus proche de celui de l'image endoscopique courante. Dans ce contexte d'appariement d'images, l'influence du choix des descripteurs, de l'espace de couleurs, de la présence d'une étape de filtrage d'images peu informatives et de la modalité d'images (lumière blanche ou NBI) sont examinés et démontrent qu'une amélioration encore plus significative peut être obtenue pour la synchronisation des images.

Contents

Chapter 1

Introduction

Contents

1.1 A Historical Perspective

Since its inception in early 1800's minimally invasive surgery (MIS) has become the de-facto standard today. It reduces the operative trauma for the patient and has significantly improved postoperative recovery. The earliest descriptions of endoscopic examinations from the era of *Hippocrates* described the use of rigid tubes supported by natural lighting to examine the insides of the patient to perform diagnosis. In 1805 Phillipe Bozzini developed an instrument for inspecting the bladder and rectum with candle light reflected by mirrors which kick-started a new era in endoscopic diagnosis. The introduction of rigid telescopic instruments and improvement in artificial lighting using an incandescent light bulb (developed by Edison), revolutionized endoscopic diagnosis. But, it was Boisseau du Rocher, in 1889, who introduced a separate channel in a telescopic instrument, that established the potential for modern endoscopy and endoscopic surgery was realised. The journey to flexible instrumentation began in 1881 when Johann Von Mickulicz, designed an instrument that could be angled at 30 degrees towards its lower third section. However it was not until 1936 that Wolf and Schindler developed the first semi-flexible gastroscope which ultimately initiated the field of flexible endoscopy.

Until the 1950's, a key area in endoscopic technology, the choice of an ideal light source was lacking. Endoscopic illumination was provided by a small tungsten filament lamp positioned at the tip of the viewing instrument which was subsequently augmented by the use of telescopic lenses. This arrangement, however, was less than satisfactory as it provided poor illumination and introduced significant colour distortion. Heinrich Lamm had demonstrated in 1930 that fine threads of glass fibres could be bundled together to act as a conduit for a light source, and that the bundle could be flexed or bent without losing its transmission capabilities. However, it remains a mystery why this idea languished for

nearly 25 years, until 1954. Thus, changes in the light source from the distal electric bulb to the external light unit and Heinrich's light-conducting fibreglass technology eliminated these problems. The first gastrocamera was envisioned by Lange and Meltzing in 1898. 62 years later, the first prototype of the modern day gastroscope was developed at the University of Michigan, School of Medicine by Basil Hirschowitz, Wilbur Peters, and Lawrence Curtis. This initiated the modern era of endoscopy which has evolved from looking through a rigid tube to viewing a high definition image of the anatomy using a flexible scope on a digital screen. With the invention of flexible instrumentation, the access to the internal organs with minimal or no external incisions, while negotiating the natural curves of the human anatomy, paved the way for modern diagnostic and therapeutic endoscopy; thereby establishing the field of Gastrointestinal (GI) endoscopy. We refer the reader to [Sliker 2014, Menciassi 2014] which provides a review of flexible and other instrumentation currently in use or under clinical evaluation for screening, diagnosis and treatment in the GI tract.

1.2 Endoluminal Endoscopy

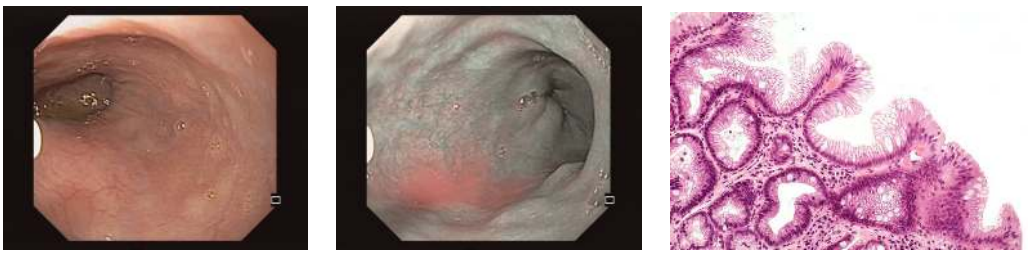
The development of flexible endoscopes with fibre-optics has allowed therapeutic procedures to be performed throughout the GI tract. GI endoscopy is a non-invasive procedure that allows a endoscopist to look at the lining of the oesophagus, stomach, biliary system, pancreas, small and large intestine, rectum and anus, using a thin flexible viewing tool called an endoscope shown in Figure 1.1.

Figure 1.1: Karl Storz Gastroscope.

In GI endoscopy, the tip of the scope is inserted through the mouth or the anus to view internal structure. Over the years, GI endoscopy has evolved from being a purely diagnostic tool for endoscopists to a minimally (or non-) invasive surgical tool. The advancements in high-definition imaging in laparoscopy has been extended to endoluminal gastric surgery. The word endoluminal literally means "*within the lumen*" and is synonymous with incisionless, transluminal and natural orifice transluminal endoscopic surgery. Operations performed within the lumen of the GI tract using an endoscope, include simple procedures such as foreign body removal, dilation of strictures, and excision of polyps, first performed through

rigid endoscopes in the early 20th century. Endoluminal procedures have since combined the techniques of flexible GI endoscopy with MIS to provide therapeutic treatment of diseases such as gastro-oesophageal reflux disease (GORD), morbid obesity, ablation of pre-malignant tissue etc. Endoluminal approaches are also being used in conjunction with laparoscopy to drain pseudocytes and necrosis of the pancreas and to excise stromal tumours. Likewise, a new generation endoluminal surgical techniques such as, transanal endoscopic microsurgery, transgastric endoscopic surgery etc., are being investigated. The use of endoscopic ultrasound has enabled assessment of nature and depth of penetration of lesions in the GI tract. With needle biopsy and other techniques its utility has been extended to areas outside the GI tract, bridging the gap between laparoscopic and endoscopic techniques.

In the upper GI tract, oesophageal varices are routinely treated with banding, injection therapy, or both, in most cases obviating the need for emergency surgery. Diagnosis and treatment of oesophagitis, gastritis, chronic inflammation, GORD, and Barrett's oesophagus are being routinely performed using endoluminal procedures. Treatments for Barrett's, a pre-malignant condition that can lead to oesophageal cancer is being closely studied under different imaging methods (Appendix A). Figure 1.2 illustrates the oesophagus under three different endoscopic modalities. For therapy of Barrett's mucosa, ablation and excision of the suspicious tissue are routinely performed. However, initial evidence ([Wang 2008a, Fitzgerald 2014]) suggests that *regular screening and biopsies have improved the median survival rate. For colonic endoscopy as well, aggressive resection of polyps and regular surveillance has decreased the need for surgery [Karlen 1998]. Thus, the need for* routine surveillance using different imaging modalities has become integrated into most healthcare systems.



(a) (b) (c)

Figure 1.2: (a) Endoscopic view under standard white-light imaging. (b) Endoscopic view under Narrow band imaging. (c) Histopathology of an extracted tissue sample *Source*: `http://pathology2.jhu.edu/beweb/fig1.htm`. All images are from the human oesophagus.

1.3 Gastrointestinal cancer

GI cancer refers to malignant conditions of the GI tract and accessory organs of digestion, including the oesophagus, stomach, biliary system, pancreas, small and large intestine, rectum and anus. The symptoms relate to the affected organ and can include obstruction (leading to difficulty in swallowing or defecating), abnormal bleeding or other associated

problems. The diagnosis usually requires endoscopy to allow, biopsy of the suspicious tissue. The treatment depends on the location of the tumour, as well as the type of cancer cell and whether it has invaded other tissues or spread elsewhere, which then determines the prognosis. Such procedures often lead to regular surveillance to be performed on patients, which motivate the need for additional navigational aides. Surveillance of pre-malignant conditions, refers to endoscopic follow-up of individuals who are at an increased risk for malignancy or in whom a neoplastic lesion has been identified and removed [Hirota 2006, Laine 2015].

1.4 Guidance in Endoluminal Interventions

Flexible endoscopic interventions are often impeded by the difficulty to orient in the endoscopic view. This is due to the small field of view, the inhomogeneous illumination and the deformation of organs in the presence of complex movements. This presents difficulties for the orientation and localization of target structures. The rationale for a guidance platform is to be able to provide the endoscopist with adequate visual feedback about the location and orientation of the endoscope in order to improve instrument navigation and facilitate in the recognition of anatomical structures. It has been proven to have statistically significant benefits in enabling smooth navigation and increased confidence during the procedure [Córdova 2013, Fernández-Esparrach 2010, Azagury 2012].

Several methods for navigation in laparoscopic procedures have been presented in the literature; Baumhauer *et al.* [Baumhauer 2008] reviews some of these methods. A more recent review of methodologies in urological procedures is presented in [Rassweiler 2014]. Most of these techniques use registration between an alternate intra-operative imaging such as intra-operative ultrasound or CT in combination with an existing pre-operative model. The most significant challenges for navigation in endoluminal endoscopy are; (a) Pre-operative imaging is seldom available; and is usually not sufficiently discriminative to identify the suspicious tissue structure that the GI experts require. (b) In cases where pre-operative imaging is performed, the deformation in the GI tract during the surgery, relative to the condition at the stage of pre-operative imaging, makes localization quite challenging. (c) Tracking of flexible instrumentation, which is necessary to combine with any pre-operative data, if present. (d) Furthermore, the use of a flexible scope prevents application of traditionally used optical tracking techniques. There are several commercially available navigation systems in laparoscopic surgery, for example, BrainLab for neurosurgery, CAScination for hepatic surgery, PercuNav for interventional radiology and HipSextant™ for orthopaedic, to name a few.

For flexible endoscopy, alternate tracking platform have been employed; in [Mori 2007, Leong 2012, Grand 2011] the authors have used registration with pre-operative CT using an electromagnetic tracker for guidance in bronchoscopy. It has also been developed into a commercial product SPiNView® by Veran technologies. Olympus has pioneered the inclusion of electromagnetic sensors in the colonoscope called ScopeGuide®. However, their device is limited to providing guidance to avoid loops during the procedure and it's use has been restricted to clinical studies thus far. But, there exists no commercially available

navigation platform for GI procedures.

Owing to the fact that many endoluminal interventions follow dedicated guidelines, calling for systematic surveillance and biopsy; navigation in this context can be differentiated as, "Intra-operative" (during a single surgical intervention) and "Inter-operative" (between two surveillance interventions). Besides the traditional navigational guidance, two very important paradigms exist in inter-operative methods; (1) *Biopsy site re-localization*: to reposition the endoscope at a previously biopsied site, during a follow-up endoscopy; (2) *Temporal differential surveillance*: to perform a comparative assessment of tissue evolution between two surveillance endoscopies for an informed diagnosis. A detailed description is provided in Section 2.3. The challenge in the inter-operative methodologies require the tackling of both the above mentioned navigational issues in the presence of important clinical hurdles (Section 2.5), whereas, intra-operative methods deal only with biopsy site re-localization. These two paradigms form the chief aspects of navigational guidance for endoluminal procedures. Providing a solution in this context is one of the motivations for this thesis.

1.5 Objectives of the thesis

This section identifies the main goals proposed to be achieved in this thesis. The next section then provides the connection between these objectives and the organization of the thesis by chapter.

1. To provide an initial proposition for a navigation system for endoluminal surgery, without using a pre-operative model. The proposed solution uses an electromagnetic tracking system. The characteristics of the approach are its ease of set up and integration with the clinical environment. Additionally, the approach must be invariant to the endoscopic modality employed during the procedure.

2. Development of an interface to suit the GI clinical workflow and enable a clear understanding for its users. This is significant, because, in order to fit into the clinical workflow, the information provided must utilize the existing tools in GI procedures.

3. Performing a complete quantitative and qualitative analysis of the system, to provide a conservative margin of error in live cases that is quantifiable for the user.

4. Exploring the extent to which the image based relocalization approaches presented in the literature, are applicable for endoluminal procedures and to justify the need for an EM based solution. Then, extending the initial proposition from item (1) in order to perform a constrained image based navigation.

1.6 Manuscript Organization

- Chapter 2 begins with a study of the clinical conditions in upper and lower GI procedures, that use surveillance based methodologies for long-term treatments. By analysing the various scenarios, the rationale for relocalization is identified. Further

review of the literature, revealed two solutions to the relocalization problem. Studying the specifics of these solution methodologies, determined the various challenges encountered and the shortcomings of the existing methods for solving them. This motivated the need for proposing an alternate approach to relocalization based on the electromagnetic tracking system (EMTS), which is presented in Chapter 3. Furthermore, discussions with clinicians and a careful analysis of the problem domain, identified, two important paradigms of relocalization from a technical standpoint. These, in turn provided the guidelines for development of the desired software interface to be presented to the clinicians.

- Chapter 3 realises the conceptual model presented in 2; describing the various elements of the system design and set up for using an EMTS. It introduces the notion of "*recording an intervention*", to be used in the proposed framework and builds the required mathematical notation for the rest of the thesis. Broadly speaking, the proposed methodology, aims to provide inter-operative video synchronization between two surveillance procedures performed at different times on the patient. In this context, the application of inter-operative registration is explained and a approach using three landmarks is presented. Using the two paradigms of relocalization introduced in Chapter 2, this chapter, then describes the design of, a first version of the software interface to be presented to the clinician during an intervention.

- Chapter 4 performs the evaluation of the proposed system described in Chapter 3. The purpose of these evaluations is to assess the performance of the system both quantitatively and qualitatively. This chapter begins with a careful analysis of the various sources of uncertainty in the system, followed by a series of experiments to empirically measure them. These measurements were used to generate synthetic data-sets for the first set of quantitative evaluations. A second set of experimental evaluations were performed on interventions using pigs. Both these experiments compute the error in average depth during relocalization inside the oesophagus, as the reported metric to measure the system's performance. Then, a set of experiments to quantify the subjective assessment of the clinical experts using the proposed conceptual and software framework is presented. The experiments in this chapter demonstrated benefit to the end user.

- Chapter 4 computed the error in relocalization caused by uncertainties in the system. In extreme cases, due to human factors or otherwise, large error values were observed. This motivated the need for an improvement in the inter-operative registration approach, which is the focus of Chapter 5. The approach in this chapter extends the 3-point inter-operative registration, discussed in Chapter 3 to use the complete oesophagus trajectory, and then comparing the results to demonstrate an improvement in performance, even in extreme cases.

- Chapter 6 extends the methodology presented in until Chapter 5, in which the relocalization was performed by video synchronization; only used the information obtained from the EMTS. This chapter highlights that such a video synchronization, may not

always provide a qualitative visual match from a GI expert's point of view. This could be due to a difference in view-point of the synchronized image or due to an image being uninformative. To alleviate this, the available image information must be utilized. Hence, this chapter reviews the application of computer vision concepts in the GI endoscopy literature for scene understanding and matching; in order to identify a few key methods that would be compared in 7 for an image based video synchronization.

- Chapter 7 follows the discussion in Chapter 6 and introduces the concept of "view-point" selection, using inter-operative image matching. Broadly, a constrained scene matching framework, extending the work-flow presented in Chapter 3 is proposed. It draws on the techniques selected in Chapter 6 to firstly, propose an alternate approach to detection of uninformative frames and secondly, provide a view-point match for inter-operative relocalization. Experiments conducted using images from interventions recorded on 7 patients indicated that, the constrained image-based match provided a significant improvement in video synchronization's qualitative visual match score.

- Chapter 8 summarizes the thesis, followed by an analysis of perspectives on the proposed framework and a discussion on the possible extensions for future research.

CHAPTER 2

Analysis and Problem Statement

Contents

The previous chapter indicated the lack of a commercially available guidance platform for endoluminal endoscopy and highlighted the importance of regular screening in GI endoscopy. This chapter further describes the clinical context for a navigational system to aide in such surveillance procedures. Specifically, the focus here is on two pathological conditions: (1) Oesophageal adenocarcinoma and; (2) Colorectal cancer. The context for surveillance endoscopy is explained; describing the need for a navigational platform that can provide inter-operative relocalization.

Sections 2.1 and 2.2 present the various pathological conditions in GI endoscopy that demand periodic surveillance. Section 2.3 discusses the rationale for relocalization in such procedures. Section 2.4 reviews the state of the art and Section 2.5 presents the primary challenges in the context of "inter"-operative relocalization. Then, in Section 2.6, there is a presentation on the use of electromagnetic tracking in clinical applications and finally there is a brief overview of the proposed approach, based on the problem analysis.

2.1 Oesophageal Adenocarcinoma

There are two primary types of oesophageal cancers: squamous cell cancer and oesophageal adenocarcinoma (OAC). Squamous cell cancer occurs most commonly in people who smoke cigarettes and drink alcohol excessively. Whereas, OAC occurs most commonly in people with gastro-oesophageal reflux disease (GORD). The latter condition has seen an increase in

frequency in the last two decades. GORD is a benign complication caused when the stomach acid escapes into the lower part of the oesophagus. When this disorder becomes a chronic condition, it can lead to changes in the oesophageal lining, causing the tissue to resemble the intestinal lining. This pathological condition is termed as Barrett's oesophagus (BO). The British society of gastroenterology provides a working definition of BO [Fitzgerald 2014]:

> *BO is defined as an oesophagus in which any portion of the normal distal squamous epithelial lining has been replaced by metaplastic columnar epithelium, which is clearly visible endoscopically (≥ 1cm) above the gastro-oesophageal junction and confirmed histopathologically from oesophageal biopsies.*

Several studies have indicated a direct link of BO with OAC. OAC appears to arise from the Barrett's mucosa through progressive degrees of dysplasia [Conteduca 2012, Evans 2012] observed in the cells of the lower oesophagus as shown in Figure 2.1. The possibility of being able to perform staging of the precancerous tissue, provides room for early diagnosis and targeted treatments, avoiding emergency surgical interventions such as oesophagectomy.

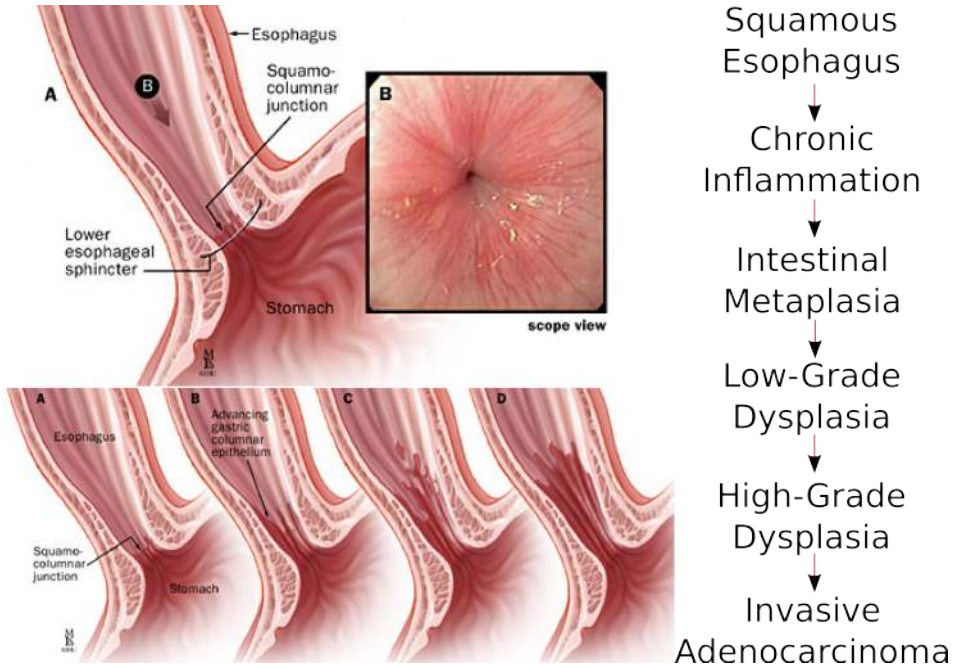


Figure 2.1: Progression of dysplasia in the oesophagus. The observation of many stages provides room for early detection and treatment.

The presence of BO is associated with increased risk of developing OAC. Adenocarcinoma in BO develops in a sequence of changes, from non-dysplastic (metaplastic) columnar epithelium, through intermediate-grade, low-grade and then high-grade dysplasia (precancerous change detected microscopically) and finally into OAC. This makes early

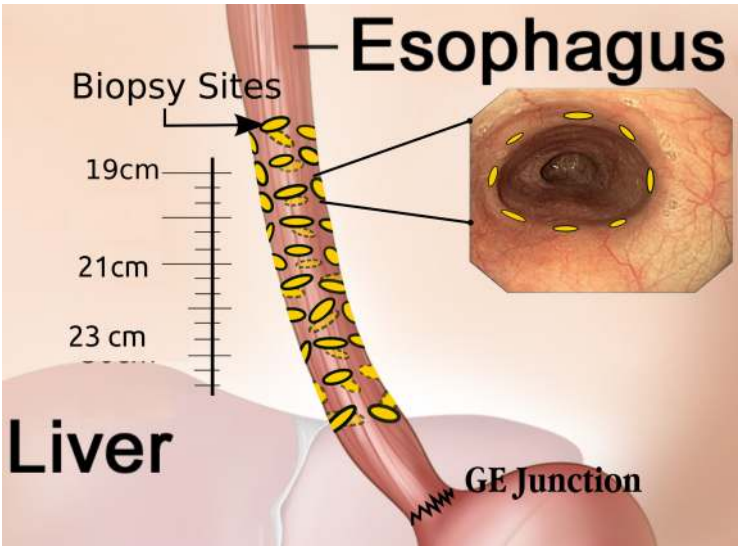


Figure 2.2: Seattle protocol based quadrant biopsies.

detection and treatment a possibility through surveillance. The guidelines [Wang 2008b] prescribe different levels of surveillance intervals depending on the degree of dysplasia, with a minimum of two surveillance endoscopies with biopsy per year. According to the Seattle protocol [Levine 1993, Levine 2000] a typical surveillance procedure involves taking four quadrant biopsies every 2 cms towards the distal end of the oesophagus and in suspicious regions. The biopsied tissue is sent to the pathology for evaluation. For most treatment procedures involving high-grade dysplasia, a 3-month follow-up for 1 year and an yearly follow-up thereafter is recommended. With the introduction of new imaging modalities (Appendix A), in-*vivo* evaluation of the tissue is now possible.

2.2 Colorectal Cancer

All colorectal cancers (CRC) develop from dysplastic precursor lesions. This is true either in the presence of a predisposing factor such as in inflammatory bowel diseases (IBD) or lack thereof, with lesions occurring sporadically. The fact that there is a pre-malignant phase in CRC, allows a window of opportunity for early detection and cure through planned surveillance. Macroscopically the shape of lesions observed in the colon have been classified as follows [Laine 2015, Inoue 2003];

1. *Visible dysplasia*: Dysplasia identified on targeted biopsies from a lesion visualized during colonoscopy.
    (a) *Polypoid*: Lesion protruding from the mucosa into the lumen ≥2.5 mm Figure (a) and (b).
        i. *Pedunculated*: Lesion attached to the mucosa by a stalk
        ii. *Sessile*: Lesion not attached to the mucosa by a stalk: entire base is contiguous with the mucosa

(b) *Non-polypoid*: Lesion with little (<2.5 mm) or no protrusion above the mucosa Figure (c)-(f).

i. *Superficial elevated*: Lesion with protrusion but <2.5 mm above the lumen (less than the height of the closed cup of a biopsy forceps)

ii. *Flat*: Lesion without protrusion above the mucosa.

iii. *Depressed*: Lesion with at least a portion depressed below the level of the mucosa.

(c) General descriptors

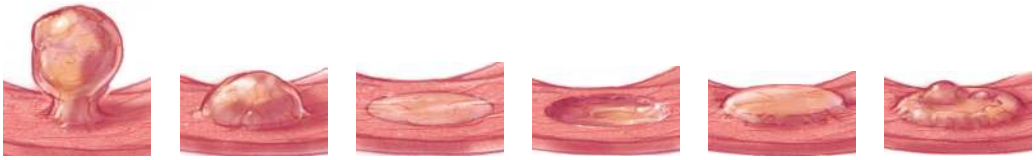
i. *Ulcerated*: Ulceration (fibrinous-appearing base with depth) within the lesion.

ii. Borders

A. *Distinct border*: Lesion's border is discrete and can be distinguished from surrounding mucosa.

B. *Indistinct border*: Lesion's border is not discrete and cannot be distinguished from surrounding mucosa.

2. *Invisible dysplasia*: Dysplasia identified on random (non-targeted) biopsies of colon mucosa without a visible lesion.



(a) (b) (c) (d) (e) (f)

Figure 2.3: Macroscopic classification of lesions in the colon: (a) Polypoid pedunculated, (b) Polypoid sessile, (c) Non-polypoid flat, (d) Non-polypoid depressed, (e) and (f) Non-polypoid Elevated. *Source*: [Soetikno 2008].

In some CRC where the dysplastic precursor lesion is of polypoid type; polypectomy is performed and the CRC risk is localized. In contrast, for patients with long standing IBD, dysplasia can be polypoid or flat, localized or multi-focal and, once found, marks the colon at high risk for CRC [Rembacken 2000, Soetikno 2008]. In addition, these types of lesions are significantly more difficult to identify, which makes the surveillance procedures, tissue characterization and disease evolution in IBD more challenging. The typical surveillance protocol involves endoscopic mucosal resection (EMR) of the suspicious regions and biopsies in the surrounding mucosa along with biopsies every 10 cm. The guidelines [Lieberman 2012, Lutgens 2008] prescribe repetition of surveillance endoscopy at least once every 1-3 years in moderate to high risk patients.

2.3 Rationale for Relocalization

Broadly, the relocalization in endoscopic procedures can be classified as; (1) *"Intra"-operative* (IAO) relocalization methods focus on tracking and mapping the biopsy sites, as they move in and out of the field-of-view of the endoscope frame, during a single intervention. This is to enable the GI expert a global view of all the biopsy locations during a single procedure. (2) *"Inter"-operative* (IRO) methods, on the other hand, aim to provide guided navigation between two surveillance endoscopies. Relocalization in surveillance endoscopies can be considered to serve two main purposes; which are discussed in the following subsections.

2.3.1 Biopsy Site Relocalization

This is the process of finding the biopsy sites either during the same intervention or those from an earlier diagnostic endoscopy, which were identified as suspicious during pathological analysis. Although the use of advanced imaging techniques has improved detection of suspicious regions; during a follow-up inspection, the GI specialist may be required to locate the previously biopsied or surveyed locations precisely with limited prior knowledge. Typically, the GI specialist uses the scale markings on the endoscope, to return the scope to previously surveyed region. For oesophageal procedures this distance is measured from the mouth (in trans-oral) and nasal opening (for transnasal endoscopy). In colorectal procedures the distance is marked from the anus. This information can be highly unreliable, especially when relocating previous biopsy sites and while using microscopic imaging devices such as CLE or OCT. Due to the lack of deterministic tools for providing such inter-operative relocalization, the GI specialist has to survey or biopsy a large mucosal region in the absence of clear markers on the tissue, which prevents targeted treatments. In both upper and lower GI procedures, relocalization is performed based on the clinician's experience, prior anatomic description and identification of tissue scaring. In colorectal procedures an additional marker using tattoo ink injection can be performed [Cho 2007]. However, this approach has its disadvantages; (1) Ink fades over time after several months; (2) Tattooing is an invasive procedure and damages the tissue; and (3) It is used mainly post-EMR and not a recommended guideline for all surveillance procedures.

2.3.2 Temporal Differential Surveillance

Visualizing the temporal changes in the tissue structure of the oesophagus or the colon is a significant challenge. Even in the presence of sufficient anatomical information (or tattoos), allowing localization during follow-up surveillance procedures, the clinician does not have access to an approach to visually compare the temporal changes in the tissue structure. This limits the clinician's ability to effectively track the disease evolution, which is an important goal for surveillance endoscopies. It is especially significant since, there is a considerable misdiagnosis owing to the lack of clear understanding of diseased tissue appearance [Wang 2013b, Lieberman 2012]. In addition, the anatomic location of the diseased tissue may be significantly correlated with risk of cancer [Triantafillidis 2009];

making it important to augment the current endoscopic procedures with knowledge from previously conducted surveillance endoscopies.

From the medical application view-point, this thesis aims to introduce a new framework for "inter"-operative relocalization for flexible endoscopy. From the discussions in Sections 2.1 and 2.2 it is clear that many GI protocols involve regular surveillance. Thus, there is a need to provide additional tools for navigational guidance in these procedures. In addition, inter-operative relocalization for endoscopic procedures should aide the endoscopist not only in biopsy site relocalization (BSR), but also in providing temporal differential surveillance (TDS) defined earlier in this section. This manuscript focuses on the relocalization in oesophageal interventions. The oesophagus was chosen as the target organ since; (1) It is relatively fixed on both ends and hence, does not exhibit strong deformation unlike the colon; (2) and the external anatomical landmarks can be used for providing inter-operative registration which will be discussed in Section 2.6. (3) In addition, the clinical protocols for oesophageal surveillance are fairly well laid out as opposed to colonoscopic procedures, which are subject to change based on the disease condition and experience of the GI specialist.

2.4 Current solutions

Following the identification of the two important paradigms of relocalization in surveillance procedures this section presents the state-of-the-art on the techniques published in literature for two types of tasks depending on when the relocalization is actually performed.

2.4.1 Intra-Operative relocalization

The IAO based approaches as defined earlier in Section 2.3 focus on detecting, tracking and localizing biopsy sites during a single procedure. Primarily, these approaches focus on BSR. One of the first methods in IAO relocalization, was published by Allain *et al.* [Allain 2009, Allain 2010]. In their approach, the authors proposed to compute feature points in scale-space around the biopsy location and then extracted descriptors for these points using scale invariant feature transform (SIFT) for the two endoscopic views to be matched. Then employing the epipolar constraint, a fundamental matrix was computed between the two views, that mapped the biopsy site to facilitate re-targeting. In [Allain 2012] a framework for characterizing and propagation of the uncertainty in the localization of the biopsy points was presented. Mountney *et al.* [Mountney 2007] performed a review of various feature descriptors applied to deformable tissue tracking and in [Mountney 2006] proposed an Extended Kalman filter (EKF) framework for simultaneous localization and mapping (SLAM) based method for feature tracking in deformable scene, such as in laparoscopic surgery. This EKF framework was then extended in [Mountney 2009] for maintaining a global map of biopsy sites for endoluminal procedures, intra-operatively. The authors presented an evaluation of the EKF-SLAM on phantom models of stomach and oesophagus. Giannarou *et al.* [Giannarou 2009] presented an affine-invariant anisotropic region detector robust to soft tissue deformations. This was used by [Atasoy 2009] along with SIFT descriptors. The feature matching problem was then modelled as a global optimization of

an Markov Random Field (MRF) labelling. Recently, Ye *et al.* [Ye 2013, Ye 2014, Ye 2016] addressed the biopsy site re-targeting in three stages. First using the Tracking-Learning-Detection (TLD) method proposed by Kalal *et al.* [Kalal 2012]. TLD was used for tracking multiple regions around the selected biopsy site. Under the assumption that the regional tissue deformations can be approximated using local affine transformations, a local homography between matched region centres was estimated. In this way multiple regions around the biopsy sites are tracked, which were then used for homography estimation and mapping the biopsy sites. Wang *et al.* [Wang 2014a] proposed to learn a graph (atlas) from a sequence of images from several gastroscopic interventions. Considering that the stomach's deformation is not large between similar frames, the nodes of the learnt graph atlas were connected by an estimated rigid transformation. Thus, the mapping of the biopsy sites from a single (reference) frame to subsequent frames for any given intervention was reduced to a graph search problem. First, for the reference frame and the moving frame their corresponding matching nodes in the graph were computed. Using Dijkstra's algorithm, the shortest path between these matched nodes was obtained. Hence, the transformation between the reference frame and moving frame was obtained as the associated combination of rigid transforms along the shortest path between the corresponding matched nodes of the graph.

2.4.2 Inter-Operative relocalization

In contrast, the IRO methods attempt to provide localization between interventions. In [Atasoy 2011, Atasoy 2012b], Atasoy *et al.* proposed to formulate the relocalization as a image-manifold learning process. The method involved, building an adjacency graph between the images of a surveillance intervention. Normalized cross-correlation was used as the similarity measure between image frames to compute the adjacency graph. Then using laplacian eigenmaps decomposition that was initially proposed in [He 2005], a linear projection matrix was computed. This approximation for projection on to the manifold was used to compute the low-dimensional representation for all the images in the intervention. Then, two separate methodologies for performing inter-operative relocalization was proposed using scene association. In [Atasoy 2012b], the scene association is performed by computing the nearest neighbour directly over the low-dimensional representation from an earlier surveillance endoscopy. However, in [Atasoy 2011] a two-run surveillance endoscopy was suggested, in which a dummy surveillance is initially performed, that is then used for scene association with the actual surveillance. The authors claimed that the modified approach in [Atasoy 2011] allowed for scene association in presence of significant structural changes in the tissue.

[Liu 2014] proposed the use of electromagnetic tracking system (EMTS) for localizing the biopsy sites in the stomach. They construct a 3D model of the stomach using SLAM and map the biopsy points tracked using the EMTS on to the 3D model. The inter-operative registration was performed by selecting five reference points manually, during each intervention.

For colonoscopic procedures, the need to provide navigational assistance is substantial. One of the earliest approaches involved a combination of 3D reconstruction from pre-

operative CT with endoscopic video known as virtual colonoscopy. The chief aspect of this involved the computation of optical flow to estimate the ego-motion of the colonoscope. Ego-motion or visual odometry involves first, extracting features from the image and computing optical flow fields. Then, using the flow fields, the camera motion would be estimated. In [Puerto-Souza 2014] the authors presented a comparison of two ego-motion estimation schemes, supervised and unsupervised. Supervised methods, as shown in [Bell 2013] require training data to be available in the form of optical-flow measurements and corresponding camera motion data. Unsupervised approaches, however, used image correspondences between video frames and multiple-view geometry to estimate endoscope motion, as was shown in [Liu 2008].

Pre-operative CT scans are uncommon in routine GI endoscopies, as they are not cost effective. Intra-operative X-rays are sometimes used for guidance, to areas in the GI tract that are difficult to access, but they do not provide sufficient information for 3D reconstruction. Theoretically, the first endoscopy can be used to compute a 3D model of the target lumen, which can be used in the follow-up surveillance procedures. But, the video based 3D reconstruction in GI procedures is still an open are for research.

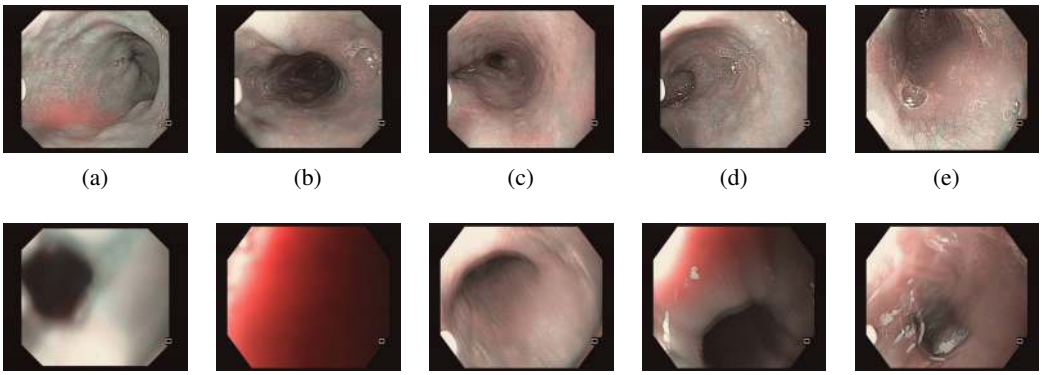
2.5 Challenges

This section presents some of the problems associated with the existing methods in the literature and the various challenges encountered in clinical settings of endoluminal surgeries which ultimately motivated the direction of the proposed methodology of this thesis. There are several hurdles involved in both of the relocalization schemes discussed in Section 2.4. An exhaustive list of these challenges that were identified during the course of this research is provided below.

Uninformative (UI) Frames: A typical endoscopic exploration lasts for several minutes and may suffer from large number of uninformative frames, as shown in Figure 2.4. Relocalization method requires the robustness to UI frames during the procedure. It should be noted that seldom are the images completely informative or UI. The presence of fluid in part of the image or contraction, in only a section of the tissue can give us a frame that is usable but that which is not completely UI. This is addressed in detail in Chapter 7. It should be noted that the presence of UI frames affects both IAO and IRO relocalization.

Endoscope and tissue motion: An obvious challenge in IAO relocalization is to accounting for tissue deformation while tracking and mapping the biopsy sites. Scene recognition in endoscopic videos is a challenging task, not only due to the tissue movement but also due to the motion of the endoscope, causing a change in view-point.

Evolution of tissue structure: Surveillance procedures in BO are performed at an interval of at least 3-6 months. One of the most important constraints as a result of this is the evolution of the tissue structure due to progression of the disease. There has been no study conducted thus far, which unequivocally demonstrates the deterministic nature of the image-based



(a) (b) (c) (d) (e)

(f) Blurred (g) Contact (h) Motion-Blur (i) Contraction (j) Fluid

Figure 2.4: (a)-(e): Informative frames. (f)-(j) Uninformative frames.

methods described in Section 2.4. However, an IRO relocalization must be robust to these changes in scene understanding. Chapter 6 presents a complete survey of image analysis methods applied to detection, classification and tracking in GI endoscopy.

Localization of the endoscope: For manual IRO relocalization, the endoscopist uses markings on the outer sheath of the endoscope to re-position. This can be affected by the patient position on the table (decubitus lateral or supine) and also by the position and orientation of the neck, which introduces an additional source of uncertainty.

Large search area: With the introduction of new imaging devices (Appendix A), *in-vivo* visualization of the tissue has improved. With devices such as the confocal microscope and the optical coherence tomography, it is now possible to make an initial diagnosis before sending tissue to the pathologist. However, the area of exploration using these devices is in the micro metric range and the area of a typical biopsy is $\approx 0.25 \sim 4mm^2$; whereas, the search area extends over a region of $\approx 450mm^2$. In presence of uncertainty in the localization of the endoscope, this area can be larger. This is closely linked with the view-point problem mentioned earlier. However, owing to a good IRO relocalization, IAO methodologies can be applied to map the regions of interest to reduce this search area.

Orientation difference: An aspect which has not been clearly identified in IRO literature has been to account for the orientation difference during relocalization. When IRO relocalization is performed using video synchronization, the synchronized frames need not be at the same orientation as the endoscope inside the lumen. Clinically, the visual disparity does not provide an ideal view for comparison to the GI expert and is an important point that needs to be clearly addressed. From a technical standpoint, the approaches discussed in Section 2.4.2, rely heavily on, either the development of orientation invariant global image description or the estimation of this difference through matched features; which would be hindered by the challenges discussed previously.

2.5.1 EMTS in Clinical Applications

The approach presented in this thesis is inspired by the need to be independent of the assumptions about, the evolution of tissue structure, rotation invariance of descriptor etc. in image-based methodologies; that were described in the previous section. Moreover, with the development of different endoscopic modalities, it is essential that the IRO framework integrates with all of them, without the need for significant modifications. To this effect, an EMTS based approach for tracking the endoscope in the lumen is proposed here. An EMTS based method offer a distinct advantage over the optical tracking methods in situations where line-of-sight is an issue and where flexible instruments are in play such as in the GI endoscopy. Additionally, multiple EM sensors can be integrated together without significant increase in the endoscope dimensions. There are several commercial EMTS available for medical applications: Ascension trakSTAR/driveBAY®, Northern Digital Inc (NBI) Aurora®, Polhemus Fastrak® and Calypso (Figure 2.5); are a few. EMTS is the only tracker that enables real-time tracking of objects without the line-of-sight restrictions. This has permitted its extensive use in MIS applications to provide navigation to reach deep-seated anatomy.

Broadly, the application of EMTS in clinical procedures can be classified as, (a) *Interventional radiology*: Typically, such procedures are percutaneous. They involve either obtaining biopsies for diagnosis or ablative procedures for therapeutic treatment. Such as in, [Grand 2011], the authors performed a clinical evaluation of the EMTS guided lung biopsies on 60 patients under CT imaging and in [Franz 2012], where a modified ultrasound probe fitted with a compact EMFE for navigation and guidance is proposed, also for percutaneous procedures. (b) *Catheter interventions*: These are procedures that use the human circulatory system for guiding the instrument. Tracking of electrophysiological catheters in cardiological interventions allows the creation of high-resolution spatial maps of the electrical activity of the heart [Linte 2012, Manstad-Hulaas 2012]. (c) *Endoscopic*: In pulmonary and GI procedures, researchers have considered tracking the tip of the scope using an EMTS to provide accurate localization. This approach has been applied to bronchoscopy by [Mori 2007, Leong 2012, Grand 2011]. For GI endoscopy, Olympus has designed a colonoscope (ScopeGuide®) with multiple embedded EM sensors that would track the position of the tip, and additionally provide the shape of a proximal section of the scope. (d) *Laparoscopic*: EMTS has been used in tracking of absolute and relative poses of laparoscopic ultrasound and video for procedures involving kidney, liver, gall bladder, lymph nodes etc. In [Jolles 2004] the authors employed EM tracking to perform computer assisted total hip arthroplasty to improve accuracy of cup placement. EMTS has been used for navigation in neuroendoscopy [Suess 2001, Peng 2002] as well. (e) *Miscellaneous*: There are several other clinical applications where EMTS has been tested for guidance, such as in the placement of feeding tubes [October 2009, Krenitsky 2011].

Franz *et al.* [Franz 2014] have provided a comprehensive overview of the application of EMTS in clinical scenarios. They provide a detailed review of the commercially available devices and cite various clinical evidence presented in literature, validating the methodologies. Further research is under-way to design new devices such as the one presented by [Lucarini 2013] that use wireless capsule endoscope with an EMTS for localization.

Figure 2.5: Commercial Electromagnetic trackers: (a) Northern Digital Inc Aurora® [NDI], (b) Calypso EM tracker [Varian], (c) Cortrack [Corpack-Medical], (d) Fastrack by Polhemus [Polhemus], (e) Ascension TrackSTAR and driveBAY [Ascension-tech].

This thesis does not study the GI endoscopic workspace for assessment of adaptability of EMTS in GI endoscopy. However, it relies on several such studies already conducted on phantom models and *in-vivo* in clinical environment. A few of the most significant studies have been listed here; [Wilson 2006, Frantz 2003, Fischer 2005, Wang 2013a, Lugez 2015, Cleary 2005, Peters 2008, Yaniv 2009, Krücker 2007, Franz 2014]. It should be noted that, apart from environment, the complex work-flow of the surgical procedure introduces additional sources of noise in the usage of EMTS. Since the use of EMTS in the clinical environment varies depending on the kind of intervention, no fixed evaluation methodology can be employed. Although some of the above mentioned methods can be employed for distortion correction, it is out of scope of this work to incorporate these models.

During endoscopic procedures it was observed that the following devices were employed routinely,

(1) *Endoscope*: Plastic or Non-ferromagnetic outer sheath with some electronic components and very low current devices.

(2) *Biopsy catheter*: also typically with a plastic sheath with a small gauge wire connecting to biopsy forceps. These are used in between the procedure and will not stay in place to continuously affect tracking.

(3) *Radio frequency ablation probes and Coagulation forceps*: These are high current devices and can affect the tracking. However typically when using these device the

tracking would generally not be employed and in case it is required to use tracking, the positioning would be completed using the navigation before using them.

Hence, for the purposes of this application, since the workspace for oesophageal procedures is limited to about $25 \times 10 \times 10\ cm^3$. This is well within the prescribed bounds of the EMFE specifications, and the distortions introduced during the procedure, can be neglected without significant loss of quality. Thus no distortion correction measures have been employed here, however, a complete assessment of it's effect with other instrumentation in the operating room must be definitely studied in the future.

2.6 Problem Analysis and Proposed Solution

Let us consider that the intervention performed a few months prior was termed as the "diagnostic endoscopy" (DE). During the DE procedure several biopsies are taken and samples sent for histopathological analysis. Depending on the outcome of their analysis, the patient is asked to return for a repeat intervention. This follow-up is termed as the "surveillance endoscopy" (SE). Broadly speaking, the proposed approach aims to achieve video synchronization between the SE, with image frames recorded during the DE or between two successive SE procedures. This is conceptually inspired, in part, by the method presented in [Atasoy 2012a]. However, their approach is constrained by the implicit assumptions made for image-based methodologies, which were discussed in Section 2.5.

To perform video synchronization of two interventions, a correspondence between the images of these interventions must be made. Over the length of the lumen in the oesophagus, the tissue texture, is quite uniform and visually does not exhibit strong variation. Thus, using the image matches to synchronize frames leads to a strong uncertainty and low confidence in obtaining the right match. To overcome this, the proposed framework includes an EM sensor that is inserted in the endoscopic channel to track its position in the oesophagus. A synchronized capture of the pose of the embedded EM sensor and the endoscopic image frame is performed, to generate a database. So, each captured image has a corresponding 3D pose associated with it.

Considering an ideal scenario, if the DE and SE were performed in succession, without disturbing the patient and the EMTS set up in the operating room, a 3D position correspondence from the EM sensor trajectory, would directly lead to video synchronization of SE with DE. However, since typical surveillance interventions are performed 3-6 months apart, to establish such a correspondence, registration of the EMTS reference frames for the two interventions must be performed. One approach that was presented in [Vemuri 2013], involved using the gastro-oesophageal junction (GOJ) as an anatomical landmark, along with the trajectory of EM sensor in the oesophagus to perform registration.

However, this approach poses difficulties in clinical settings. It requires the expert to necessarily reach the GOJ first before receiving any guidance. This however is not always possible; in case of oesophagus stenosis due to a tumour, for instance. It is also a constraint that since the clinician may come across lesions before reaching the GOJ, and may prefer to treat them immediately. Additionally, the combined variation in head and neck orientation introduces an uncertainty that is difficult to model. In fact, the patient is not always laying in

a supine position and can be set in the lateral decubitus position, depending on the surgical conditions making the orientation compensation for estimating the 6-DOF registration from the 3D endoscope trajectory alone unreliable. Hence, an alternative method using two additional sensors attached to the sternum of the patient is considered, with details provided in Chapter 3. It will be explained how, the three sensor set up is used to perform IRO registration.

Two important paradigms of relocalization, were identified in Section 2.3. The importance of this distinction between BSR and TDS, when providing a software framework for relocalization, cannot be understated; and this has not been clearly delineated in the literature. For BSR, the guidance platform requires identification and tagging of the biopsy sites during DE, thus providing relocalization to specific points in the oesophagus during SE interventions. Whereas, for TDS, the GI expert must be provided with the synchronization of the complete endoscopic intervention from DE (that may or may not contain any specific tags for biopsy sites). It is important to note that the interface difference though subtle adds substantial value to the expert under varying clinical conditions and will also be discussed further in the following chapter.

CHAPTER 3

System Description and Methodology

Contents

This chapter is based on work published in [Vemuri 2013, Vemuri 2015b].

This chapter describes the system framework and methodology of the proposed approach to provide Inter-operative (IRO) video synchronization, using an electromagnetic tracking system (EMTS). The goal is to provide a seamless integration of the hardware into the endoscopy suite with minimal additional time for set up. As was briefly described in Section 2.6, the endoscopic intervention is recorded, it is thus important, to clearly identify the various elements of such a recording, to provide a consistent mathematical structure to the problem. For the interventions to be synchronized; diagnostic endoscopy (DE) and surveillance endoscopy (SE), are performed several weeks or months apart, to provide video synchronization, a global registration between the EMTS reference frames for SE and DE, must be performed. Additionally, the synchronized frames could have been captured at different orientations of the endoscope camera *w.r.t* a reference plane, such as the patient table. This can happen due to two factors; (a) orientation of the patient (decubitus lateral or supine); (b) orientation of the endoscope (sensor) in the oesophagus at the synchronized frames. . This can lead to ambiguity in visual correlation of the scenes for the GI expert. Thus, an approach to reliably estimate this orientation difference must also be provided.

Section 3.1 describes the components of the system set up and highlights the need for, the two supplementary sensors that were initially discussed in Section 2.6. Section 3.2 outlines the description of the various elements recorded during an intervention and lists the notation that would be used in the rest of the thesis. Section 3.3 depicts the inter-operative registration

performed for registering the data collected during SE and DE, followed by the approach to perform video synchronization between these interventions. Section 3.4, discusses the various approaches to estimate the orientation difference between the synchronized frames. It also presents the clinical work-flow constraints that were considered before proposing an approach. Finally, Section 3.5 combines these methodologies into a software interface and elaborates the subtle differences in the presentation of this information to the GI expert from the standpoint of BSR and TDS.

3.1 System Setup

The system consists of a Northern Digital Inc (NDI) Aurora® EMTS with three 6-DOF EM sensors, a titanium arm and a dual channel Karl Storz flexible endoscope as shown in Figure 3.1(g), while Figure 3.1(e) shows the video interface system. The NDI Aurora EMTS consists of an electromagnetic field emitter (EMFE) Figure 3.1(b), EM sensors in Figure 3.1(a) and tracker control interface (EMCI) shown in Figure 3.1(c). The Aurora EMFE is available in two different forms Figure 2.5 (a). The first is a box type emitter, that can be mounted onto the non-ferromagnetic arm, as illustrated in Figure 3.1(d). The second type is a flat-bed device which can be placed on top of the operating table such that the patient lies on top of it. The proposed approach is independent of the type of the device, however the former device was employed; since it presented the ease of initial integration into the operating room. The range of the box type EMFE is a cubic region of volume 50 x 50 x 50 cm^3. An EM sensor is a copper wire shielded in a rubber sheath of 2mm (diameter). The EMCI is connected to the EMFE and EM sensors. The EMTS reference frame as established by the manufacturer, is located on the surface EMFE that generates the EM field. When an EM sensor is placed in the working volume of EMFE, the EMCI computes its complete pose (6-DOF) with respect to EMFE. The EMCI can simultaneously track up to 4 (6-DOF) sensors placed in the working volume of the EMFE. The 6-DOF information provided by the EMCI is of the form of a 7D vector, comprising of the three position values and four elements of a quaternion for the orientation in EMTS reference. The EMFE has a blind spot $\sim$ 5cm from its surface, hence the operating region is suitably adjusted.

Of the three EM sensors, one is inserted into the channel of the endoscope and fixed at the tip as shown in Figure 3.1(f). The two remaining sensors are used as external anatomical landmarks on the patient. There are several possible locations for placement of the sensors. However, two important constraints must be imposed to suit the clinical requirements. Firstly, that the sensors exhibit minimal movement during the breathing cycle. Secondly, that the identification of their locations is easily reproducible by the GI expert between two interventions. Based on these criteria, the jugular notch (or Suprasternal notch) was used as the primary landmark. Anatomically it is located at the superior border of the manubrium of the sternum and is between the clavicular heads. Externally it is visible as a large dip at the base of the neck. It is a stable landmark that stays stationary during a breathing cycle or when the patient position is changed. Most importantly, it is unaffected by the movement of the neck during the procedure. The third sensor is placed on the sternum $\sim$ 10 cm below the jugular notch. As will be described later in Section 3.3, only the direction which this sensor

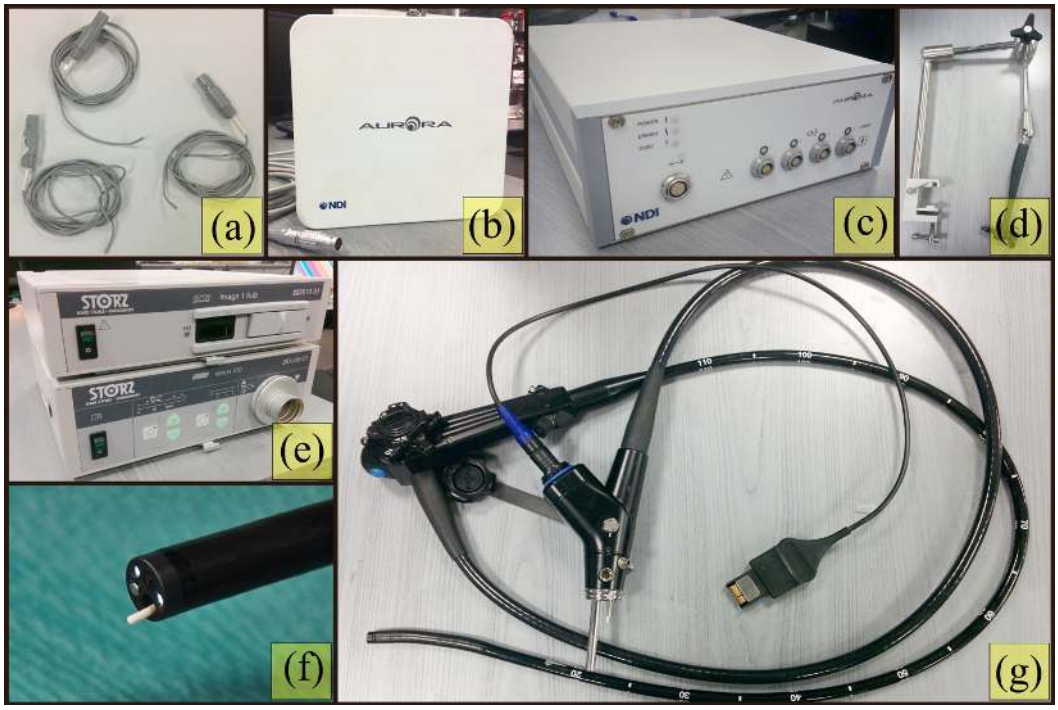


Figure 3.1: (a) Titanium arm to mount the EMFE; (b) Storz video acquisition interface; (c) NBI Aurora® EMFE; (d) Storz dual-channel gastroscope; (e) 3 Aurora 6DOF cable tools, the EM sensors; (f) the EMCI for interfacing the EMTS with the computer.

is placed is important; its exact position from the jugular notch sensor does not affect the registration.

3.2 Data Acquisition

To perform a recording of an intervention, a simultaneous capture of the video frame from the endoscope and the corresponding position tracked using the EM sensor in the channel was performed (shown in Figure 3.3). Practically, this was accomplished by maintaining a global time-line. The frame-position pair that were closest in this time-line were stored in the database. It should be noted that the accuracy of the EMTS falls when tracking fast moving objects. Hence, any data captured from the EMTS above a certain velocity threshold (10mm/sec, as was suggested in [Frantz 2003]) was discarded.

3.2.1 Notation

Defining $\mathcal{I}$ as the recording of an intervention, which is a collection of the following four items;

1. The trajectory of the endoscopic sensor in the oesophagus $\mathcal{T} = \{P_i^E \in \mathbb{R}^3; R_i^E \in SO(3),\ i \in [1, N]\}$ for N points recorded on the oesophagus trajectory, where; P_i^E = position and R_i^E = orientation as generated by the EM sensor.

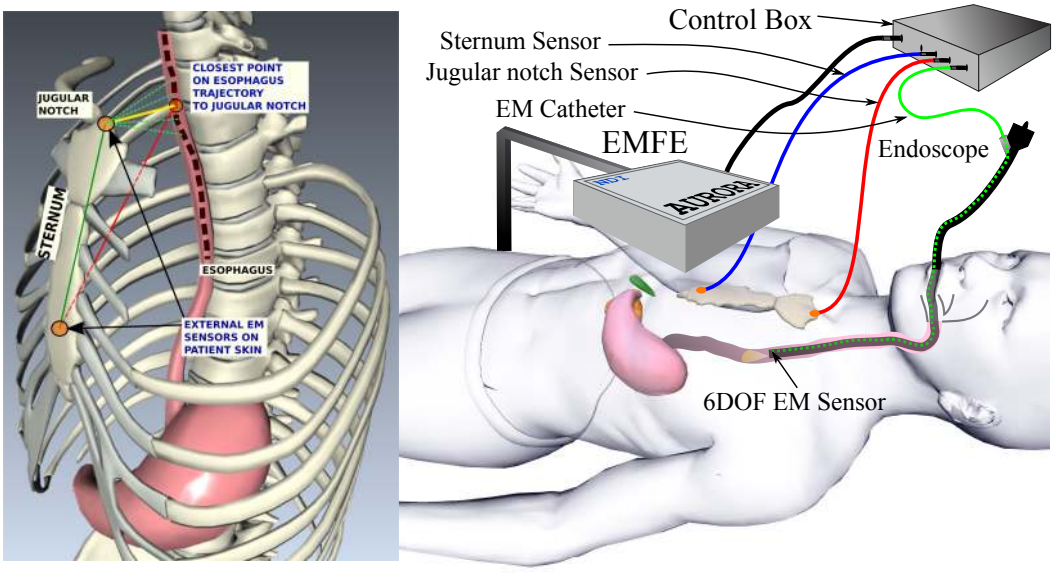


Figure 3.2: System setup: Placement of sensors, patient position and other device setup in the operating room.

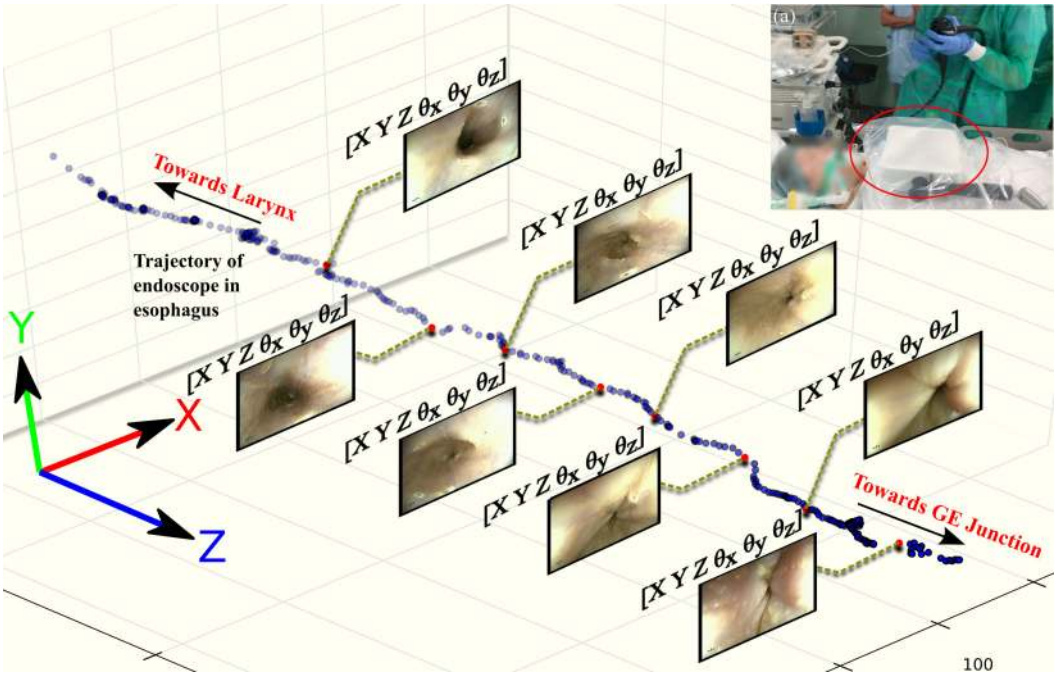


Figure 3.3: Synchronized capture of data. For each captured endoscopic frame a 6-*dof* pose is recorded for the EM sensor placed in the channel of the dual-channel endoscope.

2. A collection of sensor positions for the external landmarks on the Jugular notch ($\mathcal{L}_A$) and the sternum ($\mathcal{L}_B$) for each corresponding position (P_i^E) of the endoscopic sensor; $\mathcal{L}_{A,B} = \{P_i^A, P_i^B \in \mathbb{R}^3; R_i^A, R_i^B \in SO(3),\ i \in [1, N]\}$.

3. Set of images recorded $\boldsymbol{I} = [I_1, ...I_N]$ in the oesophagus at each position of the endoscopic sensor.

4. Tag for the entry, $k \in [1, N]$ in $\mathcal{T}$ and $\mathcal{L}_{A,B}$, which corresponds to the shortest distance of the segment $P_i^E P_i^A$ for $i \in [1, N]$. As we move along the oesophagus, the distance of the point P_i^E from P_i^A first decreases monotonically and then increases. The index k marks the position at which this distance is a minimum. k identifies a reference plane ($\mathcal{D}$) formed by the three sensors $\mathcal{D} = \{P_k^E, P_k^A, P_k^B; \in \mathbb{R}^3\}$ as show in Figure 3.2. The entry k is computed at the beginning of the procedure. For notational simplicity the suffix k can be dropped and the closest point is labelled as C. Thus, the reference plane is represented as $\mathcal{D} = \{\mathbf{P}^C, \mathbf{P}^A, \mathbf{P}^B; \in \mathbb{R}^3\}$. $\mathcal{D}$ is used for inter-operative registration as will be explained in the next section.

3.3 Inter-Operative Registration: Basic Model Using 3-points

Considering $\mathcal{I}_1$ as the SE and $\mathcal{I}_2$ as the DE; the goal is to synchronize the video frames recorded during $\mathcal{I}_2$ with the live intervention $\mathcal{I}_1$. In order to achieve this, the reference frames of the EMTS of $\mathcal{I}_1$ and $\mathcal{I}_2$ must be registered. This is accomplished by performing a rigid registration between $\mathcal{D}_1$ and $\mathcal{D}_2$. However, there is an uncertainty in the positioning of $\mathbf{P}_l^B$ and computation of $\mathbf{P}_l^C$ for the formation of the plane $\mathcal{D}_l$, $l = [1, 2]$. Hence instead of performing a direct point-to-point registration, a patient specific reference frame is computed using the plane $\mathcal{D}_l$ at the reference point ($\mathbf{P}_l^A$) as defined below;

$$T_l \equiv [\mathbf{P}_l^A; \mathbf{n}_1^l = \widehat{\mathbf{P}_l^A \mathbf{P}_l^B}; \mathbf{n}_2^l = \mathbf{n}_1^l \times \widehat{\mathbf{P}_l^A \mathbf{P}_l^C}] \tag{3.1}$$

This representation is used to perform a rigid registration ($\mathcal{I}_1$ with $\mathcal{I}_2$; $T \in SE(3)$). The registration is performed on-line during the procedure and there are no additional steps to initialize the video synchronization. T is computed only once, at the beginning of the procedure. Any significant movement of the patient can be noted from change in $\mathcal{D}_l$; in such a case k is recomputed and registration would be performed again.

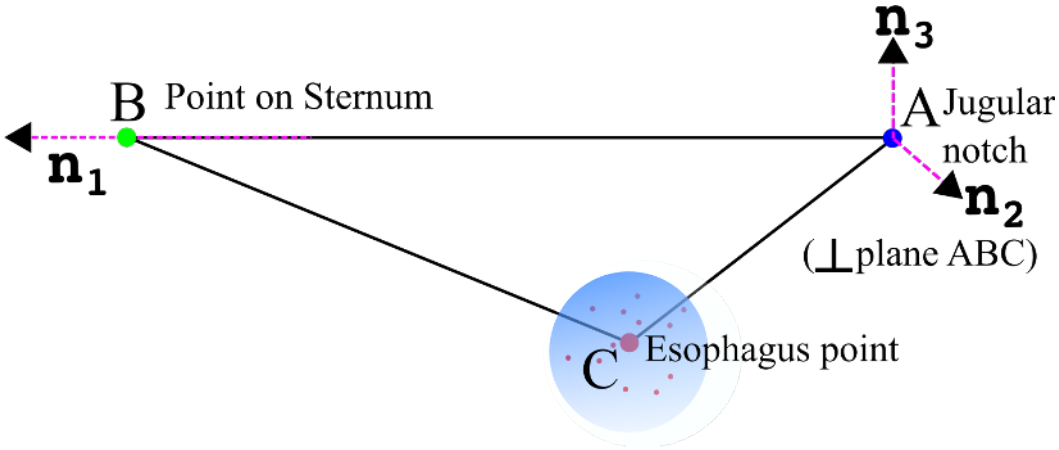


Figure 3.4: Formation of the reference frame for an intervention $\mathcal{I}_l$.

3.3.1 Video Synchronization

After the registration of the reference frames of the patient in $\mathcal{I}_1$ with $\mathcal{I}_2$, a spatial correspondence between the sensor position in live from $\mathcal{T}_1 \in \mathcal{I}_1$ is made with $\mathcal{T}_2 \in \mathcal{I}_2$, by a search for the closest neighbour (in Euclidean space). Let the function for finding the closest neighbour be denoted by f. Due to small movement of the endoscope tip (yaw and pitch motion) along the cross-section of the oesophagus, the trajectories are not completely smooth. If only the Euclidean distance was used in computing the nearest neighbour, it would lead to false matches along the oesophagus due to the situation observed in Figure 3.6. Globally, the trajectory of the endoscope in the oesophagus is sufficiently linear along $\mathbf{n}_1^l$, computed in Eq. (3.1). Hence to alleviate the false matches during synchronization, the search space is constrained along $\mathbf{n}_1^l$ (= $4mm$ chosen empirically) using $\mathcal{T}_l^{projected}$ as in Equation 3.2 and as shown in Figure 3.6.

$$\mathcal{T}_l^{projected} = \langle \mathcal{T}_l, \mathbf{n}_1^l \rangle \quad l \in [1, 2] \tag{3.2}$$

Figure 3.5 presents an example result of the synchronization performed on pigs.

3.4 Orientation Difference Estimation

Following the synchronization of frames from (pre-recorded) DE with (live) SE, the image orientation or the rotation of the camera view-up, about the image centre *w.r.t* a fixed plane, such as the operating table could be different for the synchronized frames. The two factors that could lead to such a situation were explained earlier. Presence of an orientation difference between the synchronized frames makes it difficult for the GI expert to clearly identify differences (or similarities) in them. Typically this can be corrected by defining the ground plane, such as the operating table. It can be computed at the beginning of an intervention by sweeping, for instance, an EM sensor on the table plane before the procedure starts. However, there are two clinical work-flow issues that were encountered,

1. This would add an additional step before the procedure which was not acceptable to our clinicians because, in a typical GI procedure, the patient is wheeled in on the operating table after which he/she gets into the position before being anaesthetized, following which the EMFE device is mounted (which means the calibration should be performed again). Such pre-calibration is viable only if the EMFE is integrated in the OP table, which is feasible, but still limited by the following point.

2. If a pre-calibration step was performed, it would be dependent on the position of the patient w.r.t to this reference plane. If that changes (which can happen during a procedure), the calibration would have to performed again.

Following the IRO registration and synchronization; considering a match for the u^{th} point $\{P_{u,1}^E R_{u,1}^E\} \in \mathcal{I}_1$ and v^{th} point $\{P_{v,2}^E R_{v,2}^E\} \in \mathcal{I}_2$, was obtained. An estimation of the correction in orientation of image $I_v^2 \in \boldsymbol{I}_2$ *w.r.t* image $I_u^1 \in \boldsymbol{I}_1$, must be made. Now, using

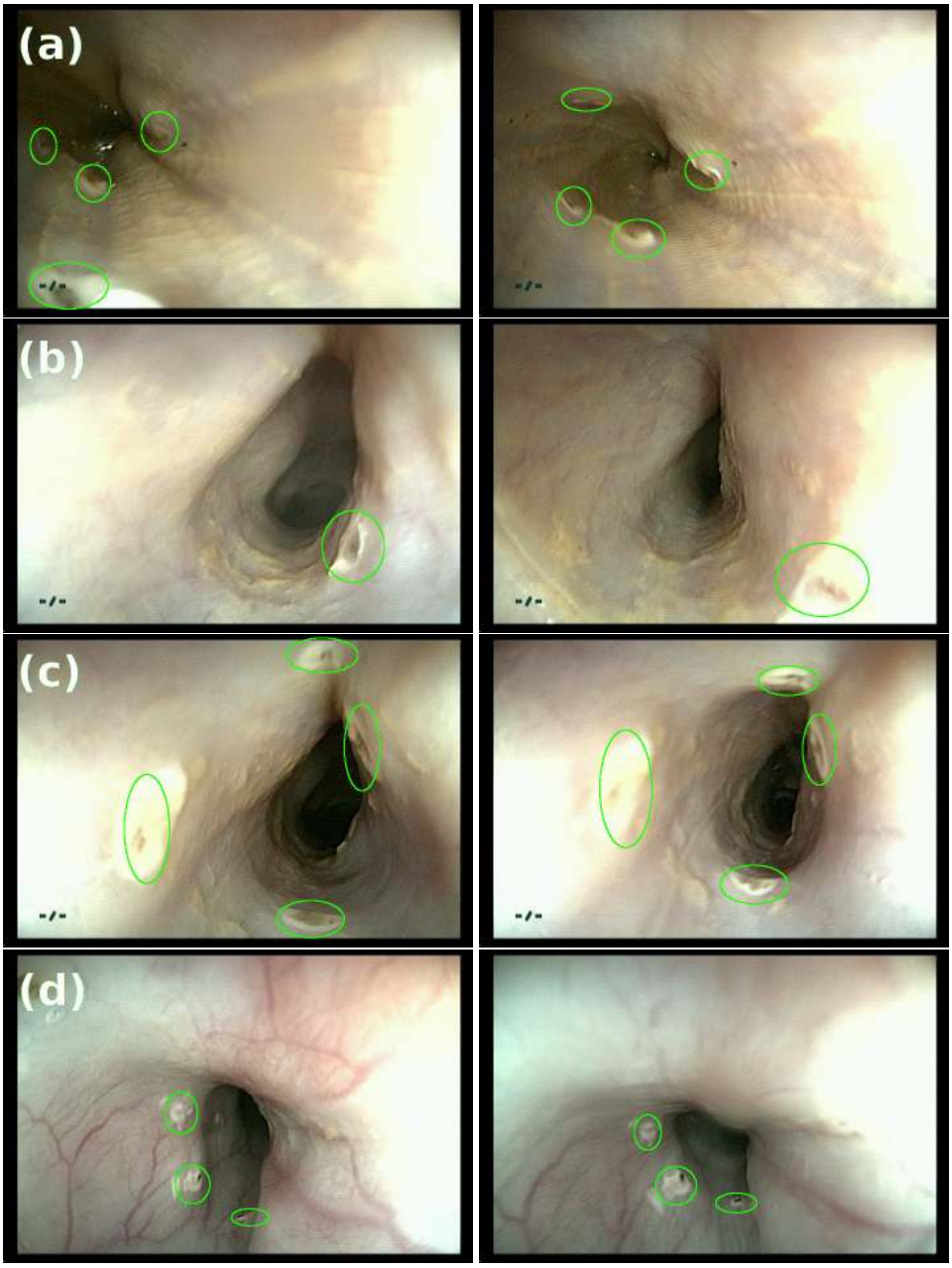


Figure 3.5: The figure shows results from interventions performed on pigs. The 1^{st} column shows the frames from the video frames from a SE, the 2^{nd} column shows the closest match found from a DE obtained using the video synchronization presented in Section 3.3.1. The marking made on these images are to highlight the locations that were used by the GI expert for comparing the matches.

$R^E_{u,1}$ and $\tilde{R}^E_{v,2}$ (corrected with T from Section 3.3), the residual orientation can be computed as,

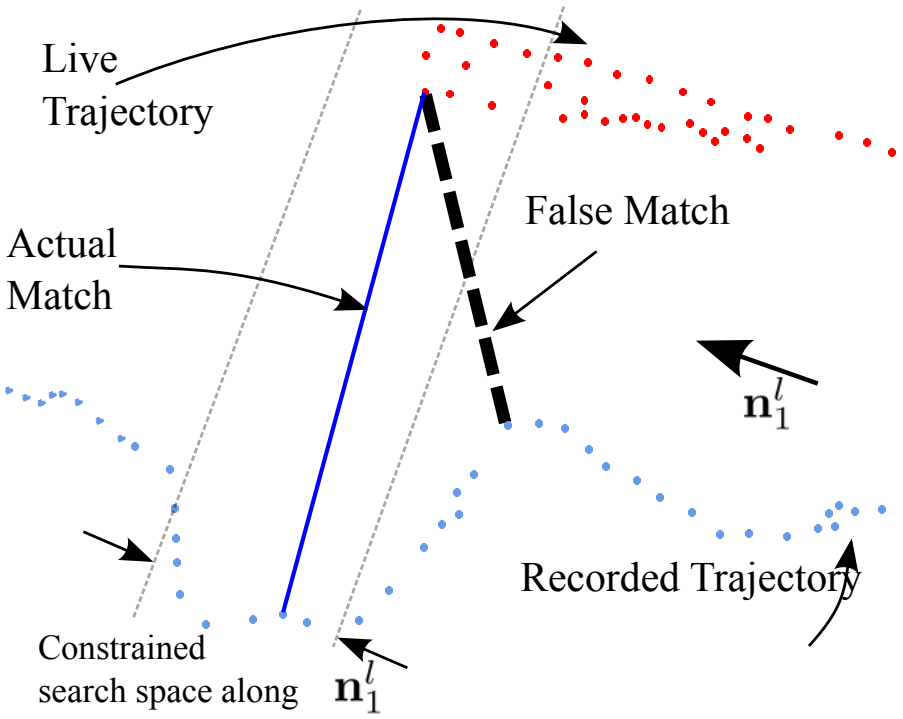


Figure 3.6: False match using Euclidean distance. The bold green dotted line indicates the closest matched point using Euclidean distance. However, the actual match is the bold blue line, which is much further away.

$$R = (R_{u,1}^{E\ -1})\ \tilde{R}_{v,2}^{E} \tag{3.3}$$

The image orientation correction can then be estimated from the roll angle (θ_c) by decomposing R into Euler angle formulation. However, this decomposition is not stable because the estimation of θ_c depends on the complete decomposition into yaw-pitch-roll. During an intervention, the endoscope can perform yaw and pitch motion, which would affect the estimated roll angle, in case of a direct estimation from R. Ideally the effect of yaw-pitch combination must be dissociated from the estimation of the roll angle (θ_c). To achieve this the following approach described below was proposed.

Considering the component vectors of the orientation, $R_{u,1}^{E} = (\mathbf{n}_x^u\ ,\ \mathbf{n}_y^u\ ,\ \mathbf{n}_z^u)$ and $\tilde{R}_{v,2}^{E} = (\mathbf{n}_x^v\ ,\ \mathbf{n}_y^v\ ,\ \mathbf{n}_z^v)$, to effectively estimate the rotation about $\mathbf{n}_z^u$ by reducing the effects due to angle between $\mathbf{n}_z^u$ and $\mathbf{n}_z^v$, the projections $\mathbf{n}_x^{v,proj}$ and $\mathbf{n}_y^{v,proj}$ on the $\mathbf{n}_x^u\mathbf{n}_y^u$ plane were computed, as shown in Figure 3.7. Then the roll angle was computed as, $\theta_c = (\alpha + \beta)/2$, about $\mathbf{n}_z^u$, as the angles required to align vectors $(\mathbf{n}_x^{v,proj},\ \mathbf{n}_y^{v,proj})$ with $(\mathbf{n}_x^u,\ \mathbf{n}_y^u)$; where, α and β are the two independent estimates of θ_c for the x and y component vectors respectively. This estimation provides the orientation difference between the endoscope position in live $\mathcal{I}_1$ (SE) and the corresponding match computed in the previously recorded $\mathcal{I}_2$ (DE).

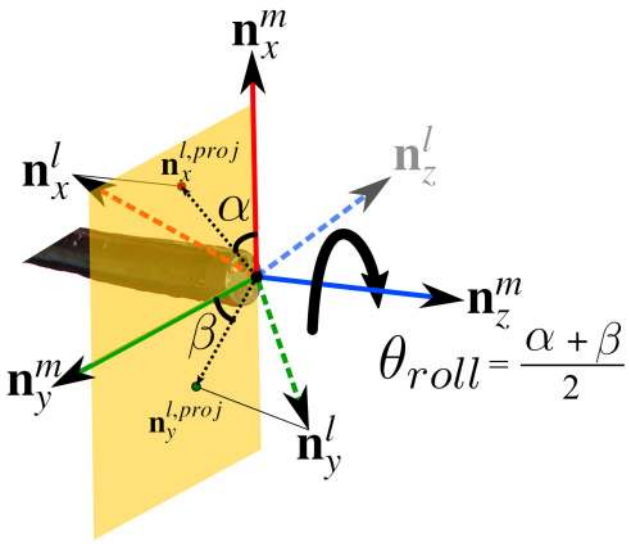


Figure 3.7: Figure showing the approach to obtain roll angle about the endoscope axis by negating the effects of yaw and pitch motions. We use this for orientation correction between the two interventions. Note that the orientation difference between $\mathbf{n}_z^m$ and $\mathbf{n}_z^l$ have been exaggerated to explain the point.

3.5 Application Interface

Section 2.3 presented two approaches in which the IRO relocalization can be applicable in a clinical settings. The subtle differences in the requirements between BSR and TDS, motivated two separate interfaces that are explained in the following sections, in a form that fits the GI procedure work-flow. We recall that $\mathcal{I}_1$ as the SE and $\mathcal{I}_2$ as the DE.

3.5.1 BSR Interface

Biopsy site relocalization or BSR, involves repositioning the endoscope during SE, at the location of biopsy performed in DE. To achieve BSR, the location of biopsy site must be tagged in the DE recording. This tag includes, the frame number of the recorded image, its corresponding 6-DOF EM sensor information and the pixel position of biopsy in the frame. Tagging, could be performed manually, during the procedure or off-line while reviewing the recording. This thesis does not present methods to perform automatic tagging of biopsy sites or identification of abnormalities. [Ye 2013, Mountney 2009, Allain 2012] have presented methods for automatic detection of optical biopsies, which can be employed in this scenario. It should be noted that to achieve BSR, only the frames which contain the biopsy would be necessary.

Figure 3.8 presents a view of the application available to the clinician during the procedure. There are four parts to this interface;

(a) *3D-View*: It depicts the current 3D position of the EM sensors during the live intervention $\mathcal{I}_1$ and the trajectory recorded during DE, $\mathcal{I}_2$. The recording for DE intervention

contains biopsy sites, which are shown as white spheres in the Figure 3.8(a).

(b) *Live View*: This is shown in Figure 3.8(b), it shows the live endoscopic image $\{I_m \in \boldsymbol{I}_1 : \boldsymbol{I}_1 \in \mathcal{I}_1\}$ that is seen by the GI expert during the SE.

(c) *Synchronized View*: This frame shows the image $\{I_l \in \boldsymbol{I}_2 : \boldsymbol{I}_2 \in \mathcal{I}_2\}$ that is synchronized with the live view. For BSR, since only the images containing biopsy are utilized, the video synchronization provides the closest corresponding image on the EM trajectory, that contains the biopsy tag. This view also consists of, a 3D arrow overlaid on the matching image that indicates the direction in which to move the endoscope to reach this biopsy site. '↑' indicates the biopsy site is further inside the oesophagus, '↓' that the site can be reached by retracting the endoscope. In addition the distance in millimetres, to this biopsy site, is also shown at the bottom left of this frame. A virtual marker is drawn on the matched image to indicate the position of biopsy site. This information is presented in Figure 3.8(c).

(d) *Biopsy Map View*: A map of biopsies that were recorded during $\mathcal{I}_2$ is provided in the interface as illustrated in Figure 3.8(d). This map is based on a standard format of reporting used by GI experts to document the biopsies taken during a procedure. It is an important aspect of the interface since it facilitates plotting the biopsy points in a context clearly understood by a clinician. The map tells the GI expert, the approximate depth of the biopsy in the oesophagus and its orientation *w.r.t* a reference on the patient. This map also shows the live pose of the endoscope during $\mathcal{I}_1$. The following section describes in details the construction of this map and the approach to incorporating the orientation information in it.

3.5.1.1 Biopsy Map Construction

The biopsy map consists of concentric circles, with each annuli depicting regions of depth of 5cms inside the oesophagus. The circle in the centre is the z-line (gastro-oesophageal junction or GOJ). During a DE, $\mathcal{I}_2$, each biopsy point is first mapped onto the corresponding annulus. During the course of procedure, when the GI expert navigates to the GOJ, this information is used to compute the annulus position. The orientation of the biopsy point in the annulus, denoted by θ^l_{map}, is estimated as the roll angle between the patient's reference frame $\mathcal{D}_2$ and the reference frame of the EM sensor at the biopsy site. The approach similar to one presented earlier in Section 3.4 is employed, in which, the EM sensor reference frame is projected onto the patient's reference frame at $\mathcal{D}_2$. Figure 3.8(d) illustrates the final outcome. Observe that the smaller black circles (marked 1-4) indicate biopsy sites performed during $\mathcal{I}_2$.

Now during SE, since the two interventions have been registered; to map the live position of the endoscope, the location of GOJ from $\mathcal{I}_2$ is used. If a biopsy site exists in this region that is closest to the live endoscope position (within a radius of 25mm), it is highlighted by a change in colour. Also, since, $\mathcal{D}_1$ and $\mathcal{D}_2$ are perfectly aligned, the live orientation θ^m_{map} is estimated in the same way using $\mathcal{D}_1$, as was described earlier. Assuming that the closest biopsy site is the required target to be reached, the orientation difference between the

synchronized frames becomes, $\theta_c = \left|\theta^l_{map} - \theta^m_{map}\right|$. As the GI expert rotates the endoscope inside the lumen, $\theta_c \rightarrow 0$ and the live view I^m_1 is aligned to fit the matched view I^l_2.

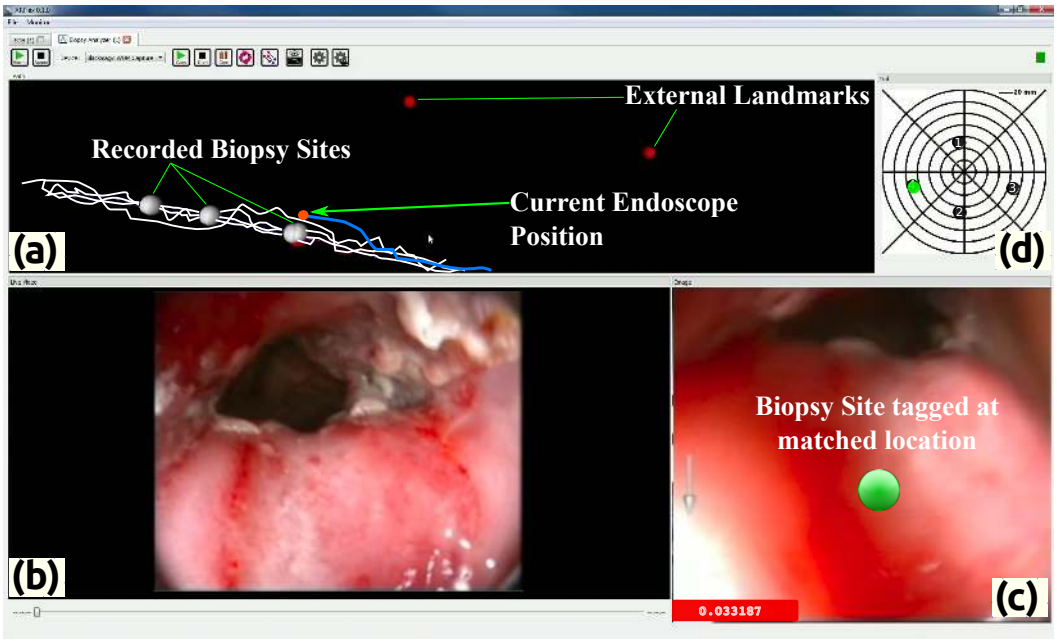


Figure 3.8: The Biopsy site relocalization application view presented to the GI expert. (a) Frame indicating the 3D view containing the trajectory of the sensor in the oesophagus. (b) Live video frame from the endoscope. (c) The corresponding matched image from a previous intervention. (d) Displays the map of biopsy sites from intervention $\mathcal{I}_2$ in dark black circles and the current position and orientation of the endoscope in the oesophagus by bright green circle.

3.5.2 TDS Interface

Temporal differential surveillance or TDS, on the other hand, involves performing a comparative assessment of tissue evolution between DE and SE for an informed diagnosis. The goal is to provide the GI expert with the option to do complete review of the anatomy, to observe temporal changes in the tissue all along the oesophagus (in this particular case). In such situation, no special tagging is necessary. The decision of determining the tissue structural changes is deferred to the GI expert's judgement.

The recording for TDS involves capturing all the frames during the intervention. Video synchronization is performed following the approach presented in Section 3.3, with every frame in $\mathcal{I}_1$ matched to its closest neighbour in 3D oesophagus trajectory to frames in $\mathcal{I}_2$. The same four views are provided for TDS that were identified for BSR, with the following differences; (a) the *3D-View* does not include any tagged biopsy sites; (b) the *Synchronized View*, does not have any arrow or distance measure indicators to guide to a specific location in the oesophagus; (c) the *Biopsy Map View* also does not contain the locations of the biopsy sites. For TDS, instead of matching the orientation to a biopsy site, the view-up indicated

in the map points to the orientation of the synchronized frame and the green dot shows the orientation of the live-view *w.r.t* the synchronized frame. This was calculated in the Section 3.5.1.1 and denoted by θ_c.

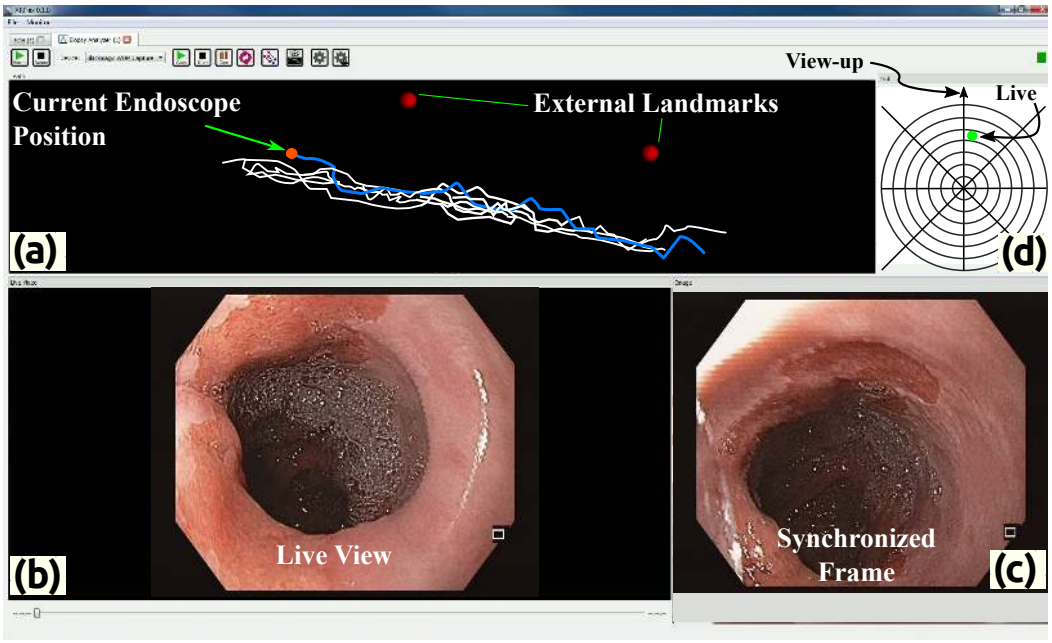


Figure 3.9: The temporal differential surveillance application view presented to the GI expert. (a) Frame indicating the 3D view containing the trajectory of the sensor in the oesophagus. (b) Live video frame from the endoscope. (c) The corresponding synchronized image from a previous intervention. (d) Displays indicates the orientation difference calculated between the live image and the synchronized frame from $\mathcal{I}_2$ using a bright green marker which rotates as the value of θ_c changes.

3.6 Conclusion

This chapter has provided a complete overview of the proposed framework to provide IRO relocalization using EMTS as part of the solution. The method discussed is independent of the image information captured during the intervention and can be applied to any of the imaging modalities reported in Appendix A. The proposed system provides a generic guidance platform that is usable in any flexible endoscopic procedures, that require relocalization. The system set-up is quite straightforward, it works in real-time (trajectory matching $\sim$ 100Hz), can be used with minimal change to the operating room protocol. Moreover, because the system set up and methodology allows the data to be reusable, the recordings can be shared between GI specialists without loss of information. Since, the proposed method does not rely on the quality of images, it is robust to the typical endoscopic image artefacts for IRO comparison. The IRO registration framework presented here uses a basic model using a reference frame and a point. Figure 3.10 summarizes the work-flow presented

in this chapter by indicating the various components discussed.

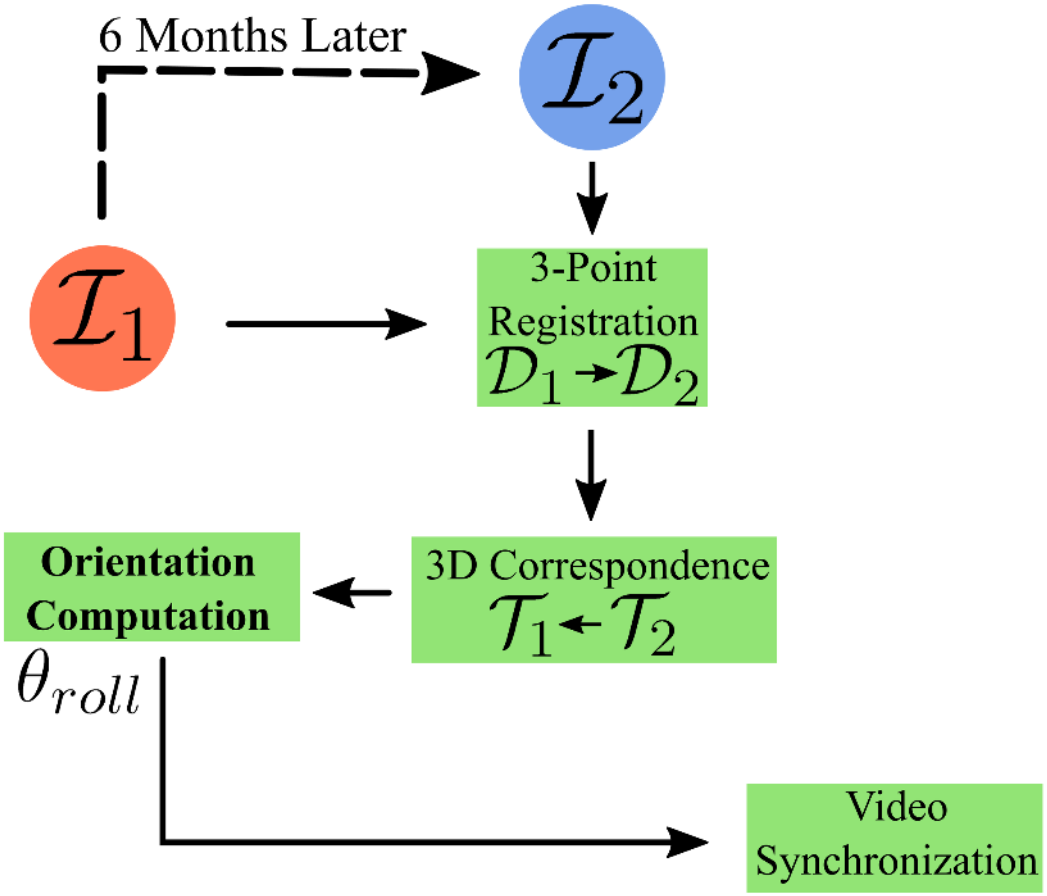


Figure 3.10: This figure summarizes the approach presented in this chapter as a flowchart.

The following chapter will present a complete evaluation of this system framework.

Chapter 4

System Evaluation

Contents

This chapter is based on work published in [Vemuri 2015b].

Chapter 3 presented a new approach to inter-operative relocalization problem. The characteristic of the proposed method over existing ones is the integration of an EM sensor at the flexible endoscope tip, so that the endoscopic camera depth inside the oesophagus is obtained in real-time during surveillance endoscopy (SE), allowing one to retrieve and display an image from a previous diagnostic endoscopy (DE) or another SE, at the same depth. The application of this concept, to two relocalization scenarios of BSR and TDS was presented.

To be clinically applicable, it is important to study and evaluate the performance of the proposed system. Hence, in this chapter a set of quantitative and qualitative experiments are proposed. To begin with, the system is carefully analysed to identify the various sources of error. Using a Gaussian model, synthetic data is generated to analyse the maximum bounds for the synchronization error. Then using real-data from interventions on pigs and an optical tracking system to provide ground-truth, further evaluation is conducted. An error metric is chosen in a way that is easily comprehensible for the GI expert under clinical setting. The goal of quantitative evaluation is to correlate the effect of the identified sources of uncertainty, to the computed error metric. This would establish for the GI expert, the crucial aspects of the system set up. Finally, to quantify the subjective assessment of clinicians, a third set of experiments were designed using data from pigs to simulate a real task of biopsy

site relocalization (BSR). It evaluates from a clinical standpoint if the proposed system provides improvement for BSR.

This chapter is organized as follows; Section 4.1 analyses and presents an exhaustive list of sources of uncertainty in the system. Section 4.2 proposes two sets of quantitative evaluations to assess the system performance. It also quantifies the error metrics understandable by a GI expert. Then in Section 4.3 a subjective assessment of the system with application to BSR is presented. The experimental results are quantified which indicate a substantial improvement to the relocalization rate.

4.1 Analysis of Error Sources

To generate realistic synthetic data, it is essential to identify the various sources of error in the application domain. Careful analysis revealed five main sources of uncertainty in the proposed system;

(a) In the placement of landmarks (external and internal), which corresponds to the uncertainty in the estimation of $\mathcal{D} = \{\mathbf{P}^C, \mathbf{P}^A, \mathbf{P}^B \in \mathbb{R}^3\}$

(b) Deformation in the patient during the intervention and between two interventions. This can occur due to change in the posture in which the patient is laying on the table (decubitus lateral or supine) and also due to the change in the positioning of the neck.

(c) Deformation due to breathing. Typically under general anaesthesia, the breathing is steady. However, under rare conditions, heavy breathing there can cause extension of the oesophagus up to 1-2 cms.

(d) Tracking errors due to EMTS, could be as a result of error in calibration of EMFE or due to distortion caused by instrumentation in the working volume.

(e) Since the endoscope is smaller in diameter than the oesophagus lumen, there is a positional uncertainty of the endoscope tip in the cross section of the oesophagus.

After analysing the sources of error, this section presents two experiments to empirically measure these uncertainties which are modelled as Gaussian. The modelling of errors (a) and (e), are discussed in Section 4.1.1 and Section 4.1.2 respectively. The effect of (b) would not be significant since the oesophagus does not exhibit longitudinal elongation. However a more significant effect would be due to the change in the position of the neck. This however, would not influence the proposed approach, since it does not rely on the complete oesophagus trajectory. The effect due to breathing (c), is non-linear along the oesophagus length and is maximum closest to the GI junction. However, this error was not modelled in the present work; but could be easily incorporated as will be discussed in Section 4.4. The error (d) due to EMTS tracking has been explored extensively in [Frantz 2003, Fischer 2005, Yaniv 2009, Wang 2013a] to characterize the distribution of error in the working volume. However, the effect on the tracking by EMTS is dependent on the environment and varies non-linearly in the working volume. Here, the volume of interest of the EMTS is small (25 x 10 x 10 cm^3) compared to its maximum tracking

volume. Therefore, the effect of distortions in EM field would be minimal and hence, has been neglected in the current study. For the synthetic evaluations, errors described by (*b*), (*c*) and (*d*), have not been modelled. These however, have been implicitly considered in the real-data and thus, their effect has been studied in combination with (*a*) during *in-vivo* evaluations. Due to the way the ground-truth was defined in Section 4.2.1, the effect of uncertainty (*e*) has not be included in the final error computation. However, it would not significantly impact the analysis, since it only affects the view-point in the oesophagus.

4.1.1 Measurement of Uncertainty in $\mathcal{D}$

The registration of two interventions $\mathcal{I}_1$ (SE) and $\mathcal{I}_2$ (DE) relies on the registration of their corresponding reference planes $\mathcal{D}_1$ and $\mathcal{D}_2$. The uncertainty in $\mathcal{D}$ is modelled by measuring the uncertainty in the positioning of the landmarks at, *a*) Jugular notch ($\mathbf{P}^A$) *b*) landmark on the sternum ($\mathbf{P}^B$) and *c*) the closest point on oesophageal trajectory to $\mathbf{P}^A$ ($\mathbf{P}^C$). Since the external landmark are placed independent of each other, they can be modelled independently as a normal distribution $\mathcal{N}(0, \Sigma)$. However, the position of the landmark, $\mathbf{P}^C$ depends on the sensor at $\mathcal{L}_A$ and must be verified if its uncertainty is uncorrelated with that of $\mathbf{P}^A$. The experiments for measuring this distribution were sub-divided in two independent steps.

First, for the external landmarks an empirical study was conducted on 18 test subjects. Each subject was positioned on the surgical table once, in the decubitus lateral position and the external sensors on the jugular notch and sternum were repositioned 12-15 times by a GI expert. To average the measurements for all subjects, a third fixed sensor was attached relative to the sensor at the jugular notch so as to create a platform (as shown in Figure 4.1), that was perpendicular to the subject's chest, thus, forming a frame of reference at the jugular notch for the patient. All the sensor positions were then transformed to this reference frame and the positional uncertainty ($\Sigma_{\mathbf{P}^A}, \Sigma_{\mathbf{P}^B}$) on the landmarks was measured at this reference frame, for each of the subjects.

In the second step, experiments were conducted on a phantom model of the oesophagus. After placing the external landmarks, the closest point was estimated on the oesophagus trajectory in 108 trials. The points were transformed to the reference frame at the jugular notch so that the computed covariance matrix ($\Sigma_{\mathbf{P}^C}$) was independent of the position of the EMFE *w.r.t* the patient. The process was repeated by repositioning the sensor at the jugular notch and the sternum sensor. A similar uncertainty matrix was observed in each case, which led to the conclusion that the uncertainty in the point $\mathbf{P}^C$ was independent of uncertainty in $\mathbf{P}^A$ and $\mathbf{P}^B$. The empirically estimated uncertainties are shown below in Equation 4.1;

$$\Sigma_{\mathbf{P}^A} = \begin{bmatrix} 8.5 & -1.4 & -2.43 \\ -1.4 & 7 & -1.3 \\ -2.43 & -1.3 & 12.5 \end{bmatrix} \qquad \Sigma_{\mathbf{P}^B} = \begin{bmatrix} 14 & -6.7 & -4.23 \\ -6.7 & 13.759 & 5.8 \\ -4.23 & 5.8 & 94.5 \end{bmatrix}$$
$$\Sigma_{\mathbf{P}^C} = \begin{bmatrix} 26 & -1.7 & -1.68 \\ -1.7 & 25 & -1.19 \\ -1.68 & -1.19 & 109 \end{bmatrix} \tag{4.1}$$

The values of the matrices presented above are in millimetres.

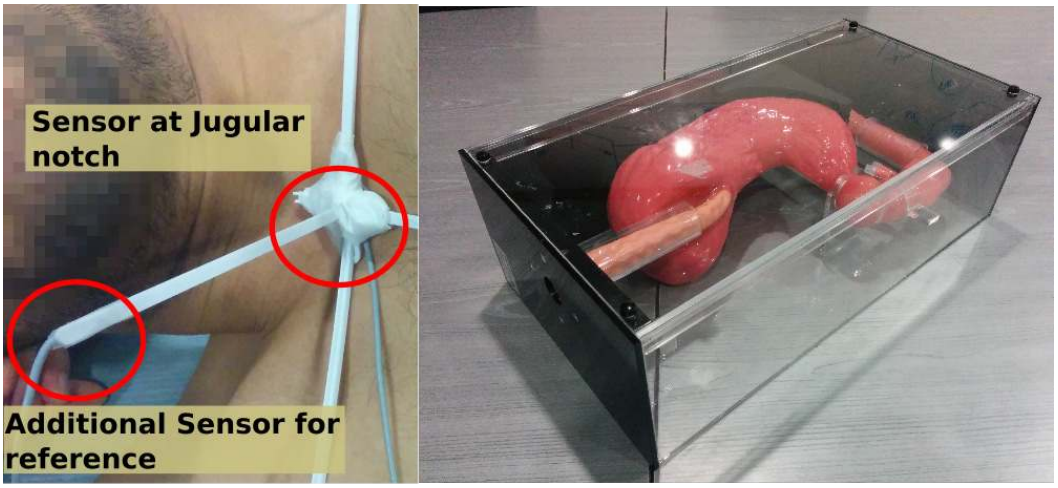


Figure 4.1: (Left) Using the second sensor at the jugular notch for providing a reference frame when measuring uncertainty in $\mathcal{D}$. (Right) oesophagus and stomach phantom, used for measuring uncertainty in the oesophagus cross-section.

4.1.2 Measuring Uncertainty in the Endoscope Tip

An additional source of uncertainty identified in Section 4.1 was the position of the tip of the endoscope inside the cross-section of the oesophagus. 100 back and forth trajectories were recorded and 10 regions each of thickness 2 mm, were selected to compute the uncertainty of points in the chosen cross-sections. The uncertainty of points in this region was then modelled as a normal distribution $\mathcal{N}(0, \Sigma_{CS})$. The experiment was conducted on an oesophagus phantom shown in Figure 4.1.

The estimated uncertainty matrix is; $\Sigma_{CS} = \begin{bmatrix} 25.7 & -1.52 & 0 \\ -1.52 & 22.3 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

The values of the matrix presented above are in millimetres. The above matrix is of rank-2, since we are interested in measuring the uncertainty in positioning the endoscope in a cross-section of the oesophagus in a plane that is perpendicular to the direction of $\mathbf{n}_1^l$ (the direction of increasing oesophageal depth).

4.2 Quantitative Evaluation

The quantitative evaluation of the system is performed in two phases; synthetic data evaluation and analysis on data collected on pigs *in-vivo*. Firstly, the empirically estimated uncertainties in the previous section are used to generate synthetic data as shown in Section 4.2.1. Section 4.2.2 presents the approach to compute the error that is meaningful in the GI endoscopic context. This error definition is used in evaluations discussed in the rest of this and the following chapter. Section 4.2.4 describes the steps involved in the quantitative evaluation on interventions on pigs. In this section it is explained how an optical tracking system (OTS) is employed to establish the necessary ground-truth.

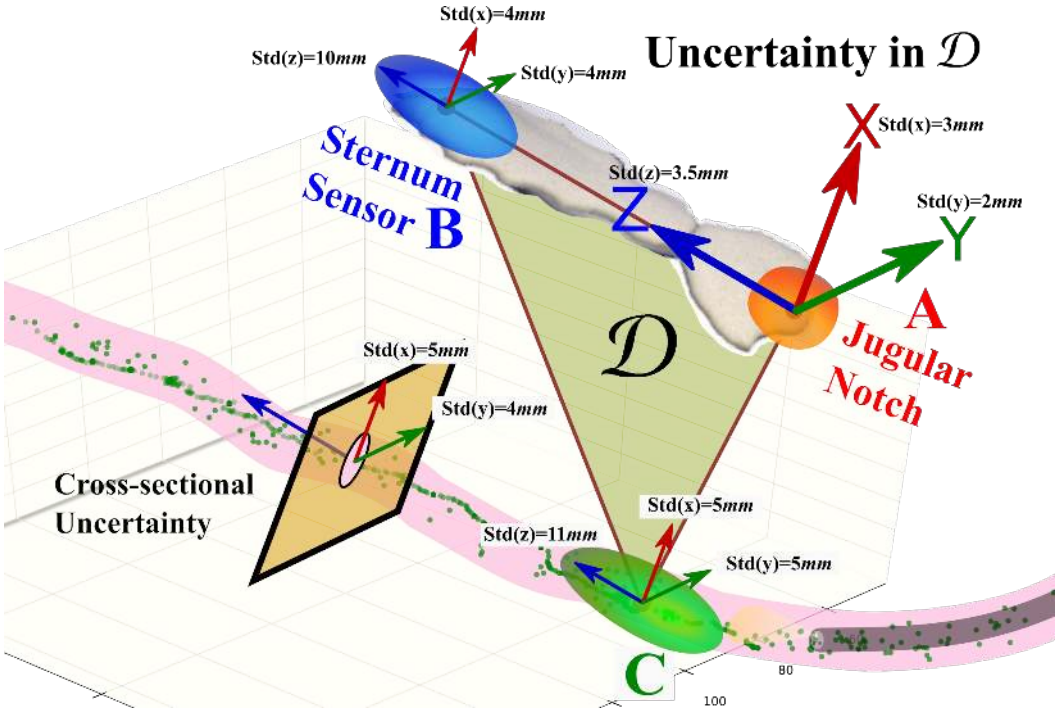


Figure 4.2: This figure summarizes the uncertainties computed empirically for $\mathcal{D}$ in Section 4.1.1 and cross-sectional uncertainty in Section 4.1.2.

4.2.1 Generation of Synthetic Data

Using the phantom model shown in Figure 4.1, a recording $\mathcal{I}_1$ was performed, as explained earlier in Section 3.2. To generate synthetic data ($\mathcal{I}_2$) at a new position, a known transform T_{GT} was applied; that was generated using a random rotation vector and translation ($\pm 1000mm$ limits) in x, y and z direction to $\mathcal{I}_1$. Then, using the estimated uncertainties $\mathcal{N}(0, \Sigma_{\mathbf{P}^A})$, $\mathcal{N}(0, \Sigma_{\mathbf{P}^B})$ and $\mathcal{N}(0, \Sigma_{CS})$, a noisy version $\tilde{\mathcal{I}}_2$ was created. Let,

$$
\begin{aligned}
[\ \tilde{\mathbf{P}}_2^A &= T_{GT}\mathbf{P}_2^A + \mathcal{N}(0, \Sigma_{\mathbf{P}^A}),\\
\tilde{\mathbf{P}}_2^B &= T_{GT}\mathbf{P}_2^B + \mathcal{N}(0, \Sigma_{\mathbf{P}^B}),\\
\tilde{\mathcal{T}}_2 &= T_{GT}\mathcal{T}_1 + \mathcal{N}(0, \Sigma_{CS})\ ] \in \tilde{\mathcal{I}}_2
\end{aligned}
\tag{4.2}
$$

be the noisy versions of the external landmark positions and the oesophageal trajectory. Now using $\tilde{\mathbf{P}}_2^A$, the new closest point ($\hat{\mathbf{P}}_2^C$) on the oesophageal trajectory $\tilde{\mathcal{T}}_2$ was computed. Then, $\tilde{\mathbf{P}}_2^C = \hat{\mathbf{P}}_2^C + \mathcal{N}(0, \Sigma_{\mathbf{P}^C})$, a noisy version of $\mathbf{P}_2^C$ was generated. Finally, for $\tilde{\mathcal{I}}_2$ the new reference plane $\tilde{\mathcal{D}}_2 = \{\tilde{\mathbf{P}}_2^C, \tilde{\mathbf{P}}_2^A, \tilde{\mathbf{P}}_2^B; \in \mathbb{R}^3\}$ was obtained. To eliminate any bias in the registration using $\mathcal{D}_1$ and $\tilde{\mathcal{D}}_2$; Gaussian noise was applied to $\mathcal{D}_1$ in the same way as described above to obtain $\tilde{\mathcal{D}}_1$. Using this synthetic data, $f : \mathcal{T}_1 \rightarrow \tilde{\mathcal{T}}_2$ is bijective (f defined in Section 3.3.1), which provides the necessary ground-truth to estimate the error in synchronization. Seven $\mathcal{I}_1$, interventions were recorded, which were transformed using 30 different T_{GT}. For each $\mathcal{I}_2$ generated by transforming $\mathcal{I}_1$, 1500 $\tilde{\mathcal{I}}_2$ samples were generated. Thus, for each base trajectory $\mathcal{I}_1$, 45000 synthetic *trajectory matches* were computed.

4.2.2 Error Measurement

To measure the error, the value that is important from a clinician's viewpoint is the discrepancy in the final synchronization of images. Computing a metric to measure similarity between matched frames from $\mathcal{I}_1$ and $\tilde{\mathcal{I}}_2$ would not yield much information since in real cases the tissue in the oesophagus may undergo a change due to natural oesophagus muscular contractions and tissue structural changes due to disease propagation. Error is computed as the mismatch along the depth of the oesophagus trajectory. As shown in the Figure 4.3, since $\mathcal{T}$ exhibits linearity along $\mathbf{n}_1$ (that was computed in Equation 3.1), the 3D Euclidean position error was projected on to $\mathbf{n}_1$, as in Equation 3.2, to obtain the depth error along the oesophagus, by normalizing the position in the oesophagus in the range [0=Beginning of the oesophagus, 1=Close to the gastro-oesophageal junction], to obtain uniformity in different oesophageal interventions.

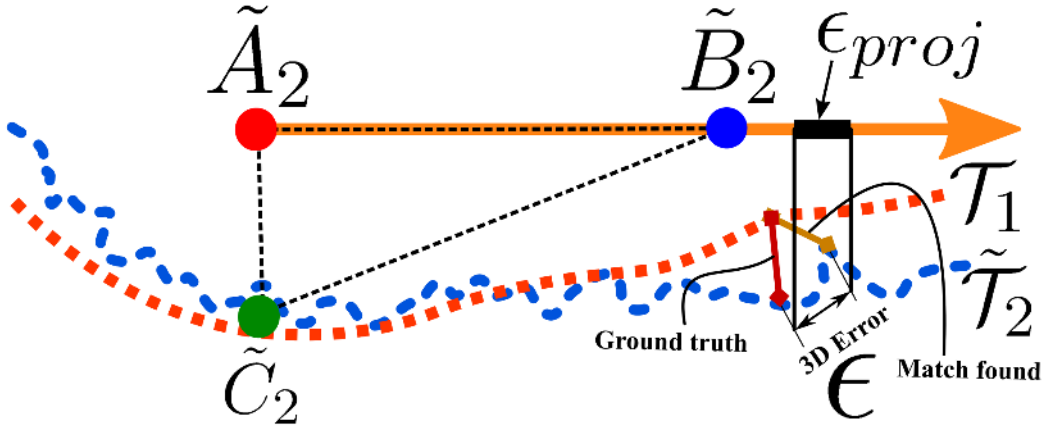


Figure 4.3: This figure shows the approach to error estimation when the reference planes are aligned with the world reference frame. The error in-effect becomes the error in depth inside the oesophagus.

4.2.3 Results

Each $\mathcal{I}_1$ is matched with the $\tilde{\mathcal{I}}_2$ synthetic samples generated. From the error measurement illustrated in Figure 4.4, a consistent bias in the trajectory due to a mismatch along $\mathbf{n}_1$ can be observed. This bias remains fairly constant along the trajectory and was observed to be on an average in the range (5-6 mm). Practically, this is due to the translational part of the registration error. The rotational error is small and has a negligible influence on the depth estimation, whereas translational error along Z has a direct influence on depth estimation, and mainly depends on the (simulated) error in positioning the landmarks at the jugular notch $\mathbf{P}^A$, the variance of which, along $\mathbf{n}_1^l$ was computed as $12.5mm^2$ in Equation 4.1. When adding variance on both $\tilde{\mathbf{P}}_1^A$ and $\tilde{\mathbf{P}}_2^A$, a standard deviation of $5mm$ is obtained, which corresponds to the reported registration error. Hence, a single point could be studied to understand the trend of registration error along the trajectory, for the 315000 synthetic samples generated, for understanding the distribution of error. It can be observed from Figure 4.5 that the error

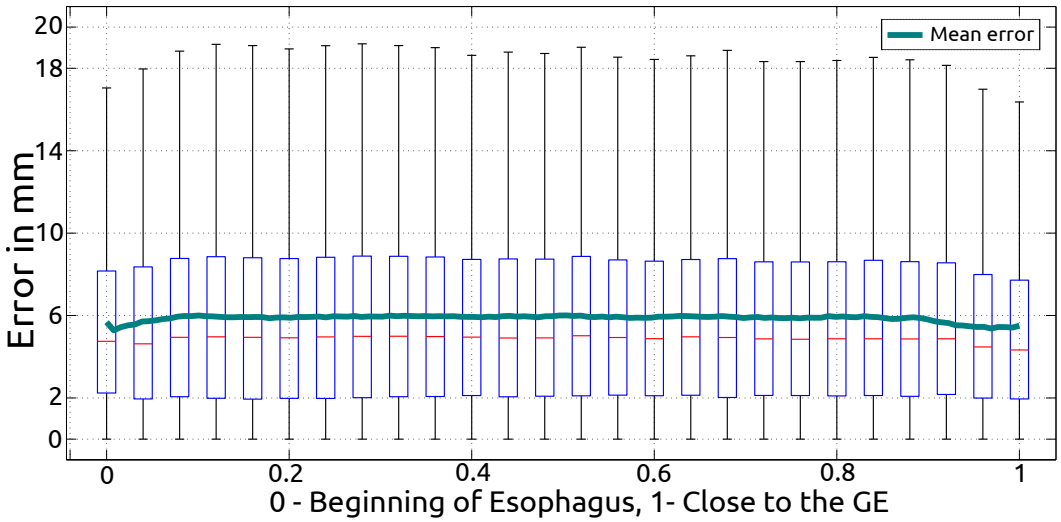


Figure 4.4: Absolute value of error along the oesophagus length for each of the 7 $\mathcal{I}_1$ trajectories, matched with their corresponding 45000 synthetic $\tilde{\mathcal{I}}_2$ trajectories. A constant offset was observed which is a result of the error in the translational component of registration.

is within 16 mm in 96 percentile of the cases. The synchronization error observed is within acceptable limits since in effect it finally corresponds to a manageable variation in image matches inside the oesophagus Figure 4.6. A $\pm 2cm$ length in the oesophagus is visible in a single frame of the endoscope. In the worst case, GI expert would be at the centre of this region and may have passed the real matched location; which can be easily retrieved by going back and forth around during the live SE. In Figure 4.4 we observed a decrease on either ends of the error curve. This is due to the fact that towards the end of the trajectory, the error due to mismatch decreases, as can be observed from ε_a, ε_b and ε_c in Figure 4.3. The maximum error observed is about 4.9 cms, corresponds to the largest observed error in 315000 matches. The error condition illustrated in Figure 4.7 is an unlikely situation to occur in reality. As it can be seen, the error in $\tilde{\mathbf{P}}_2^C$ and $\tilde{\mathbf{P}}_2^B$ are unrealistically large. This may happen only in situations where the patient is breathing heavily. However, in most procedures the patients are placed under general anaesthesia and the breathing is very steady. Hence, this effect would not be encountered in a real scenario.

4.2.4 Evaluation on Real Data

The second phase of quantitative evaluation experiments were performed on data collected on pigs. To the original set up for data acquisition, discussed in Section 3.1, an optical tracking system (OTS) was included to provide the necessary ground-truth for recording the pose of the EMFE. An optical marker M_1 was fixed on the EMFE, as shown in Figure 4.8.

This section is organised as follows; firstly the co-calibration of OTS and EMTS is performed to ultimately estimate the matrix $T_{EM}^{M_1}$, which is referred to as the hand-eye matrix in robotics literature. The calibration is divided into independent estimation of rotation and translation. Secondly, using the additional information from the OTS, a description

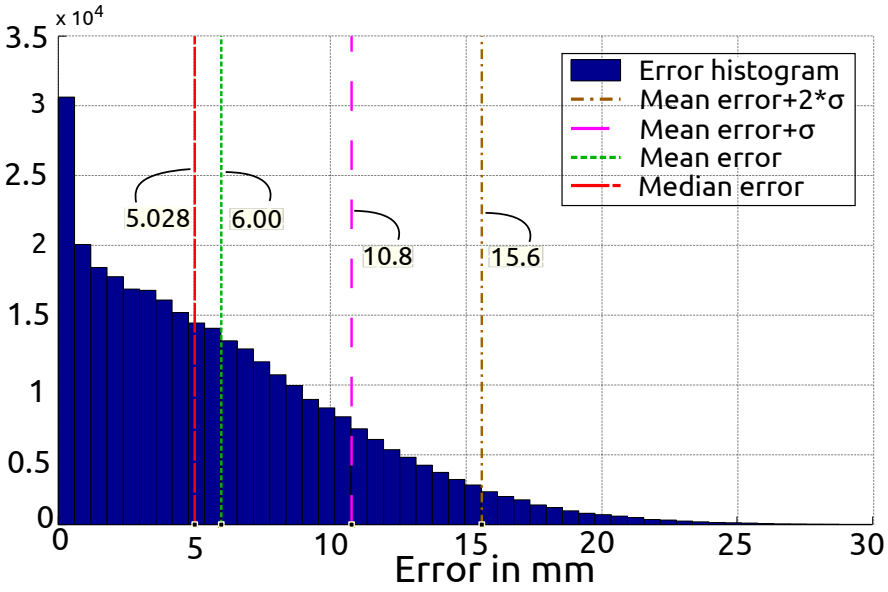


Figure 4.5: Error at a single point on the oesophagus length for 315000 synthetic trajectories.

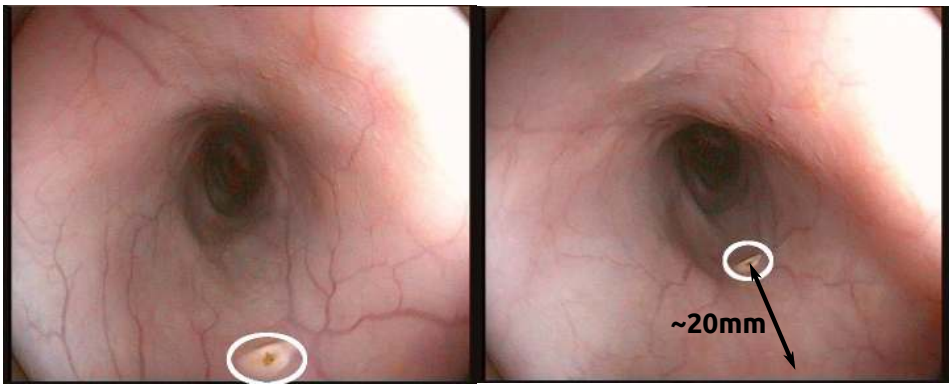


Figure 4.6: Approximate distance measure inside the oesophagus translated to displacement in image space.

of the methodology to establish the ground-truth for the collected data is presented. This information is then used to compute the error as described in Section 4.2.2.

4.2.4.1 Hand-Eye Calibration

To perform the hand-eye calibration an optical marker M_1 is fixed on the EMFE for which the hand-eye transform T_{EM}^{M1} is to be estimated. Mathematically the hand-eye calibration involves solving the equation $AT_{EM}^{M_1} = T_{EM}^{M_1}B$ [Shiu 1989]. Where $A \in SE(3)$ is the relative motion of the EMFE and $B \in SE(3)$ represents the relative motion of the marker M_1 attached to EMFE. Several approaches have been presented in literature to solve this [Tsai 1988, Chou 1991, Daniilidis 1999, Barreto 2009]

As explained in [Malti 2010], a simultaneous estimation (of rotation and translation) is

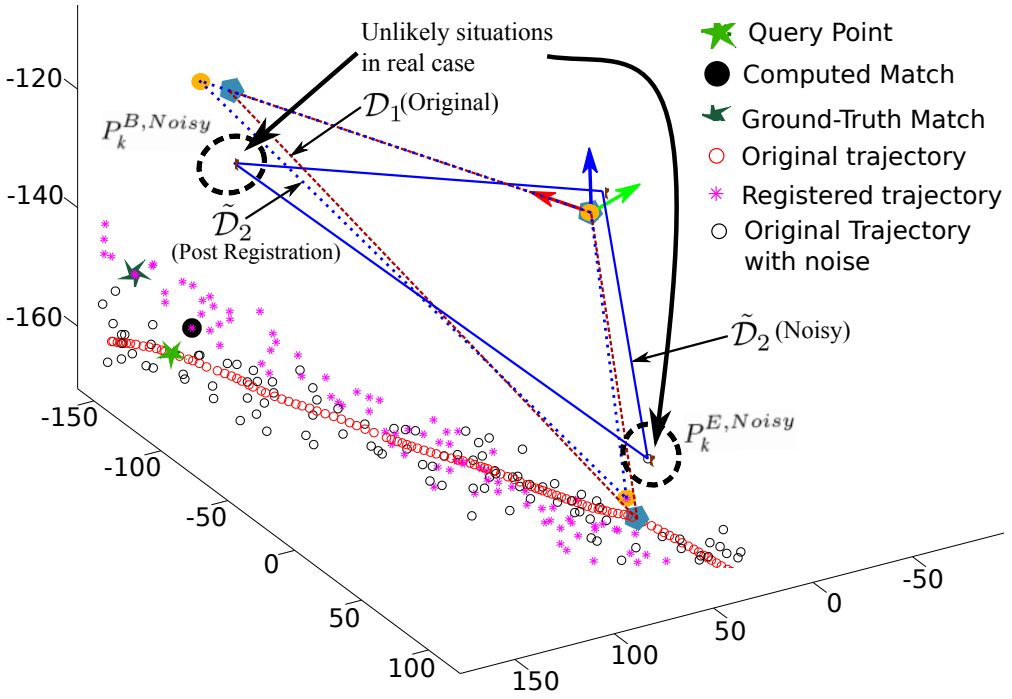


Figure 4.7: Case of maximum error of ≈ 4.9 cm. It can be seen that $\tilde{\mathbf{P}}_2^C$ and $\tilde{\mathbf{P}}_2^B$ is larger than that encountered in a real scenario. In this particular simulated case, the error in $\tilde{\mathbf{P}}_2^C$ causes it to be tending into the patient body and the noise in $\tilde{\mathbf{P}}_2^B$ causing it to be further away from the oesophageal trajectory. A registration using these points leads to an error in rotation along an axis perpendicular to the plane $\mathcal{D}_1$ through $\mathbf{P}^A$. The final registration causes the two trajectories to be crossed, leading to larger errors as we move into the oesophagus.

usually error prone. Motivated by their approach, an alternative is proposed by subdividing the estimation of the $T_{EM}^{M_1} \in SE(3)$ transform in two separate steps. First, for estimating the $R \in SO(3)$ and the second data set for computing the translation $t \in \mathbb{R}^3$. Figure 4.8 shows the set up for performing this calibration. A non-ferromagnetic flat metal strip was used to fix an EM sensor on one end and an optical marker on the other end.

Estimation of Rotation: For estimation of rotation, the metal strip was translated along a straight line and, the motion of the EM sensor and the optical marker were captured in their respective reference frames. From the recorded point clouds, a linear least-squares fit for lines is computed, that is used to obtain the unit vectors $\hat{v}^{EM}, \hat{v}^{OT} \in \mathbb{R}^3$. Then using the approach presented in [Somani 1987], where each point represents the coordinates of a unit vector in $\mathbb{R}^3$; an estimation $\hat{R}$ to align the two sets of vectors was performed. This can be computed by solving the criterion Equation 4.3 using the approach presented in Appendix B.

$$\hat{R} = \underset{R}{\operatorname{argmin}} \sum_{i=1}^{N} \|\hat{v}_i^{OT} \times (R\hat{v}_i^{EM})\| \tag{4.3}$$

Estimation of Translation: The platform with the sensor and marker, was mounted onto

a turn-table as shown in Figure 4.8. This allows both the EM sensor and the optical marker to rotate about a fixed axis and point in space. Using the arc formed by the EM sensor and optical marker, a circle $\in \mathbb{R}^3$ is estimated by first, least squares fit for the point cloud onto a 2D plane and solving the general equation of the circle in this plane, to obtain the centres (C_{EM}, C_{OT}) and the corresponding radii of the two circles. Using these, the translation was then estimated as, $t = C_{OT} - RC_{EM}$. This estimation provided the co-calibration matrix T^{OT}_{EM}. The hand-eye matrix $T^{M_1}_{EM}$ was then estimated using $T^{M_1}_{EM} = T^{OT}_{EM}(T^{OT}_{M1})^{-1}$. Calibration was performed using 11 data-sets and averaged to obtain a final hand-eye calibration matrix $T^{M_1}_{EM_{avg}}$.

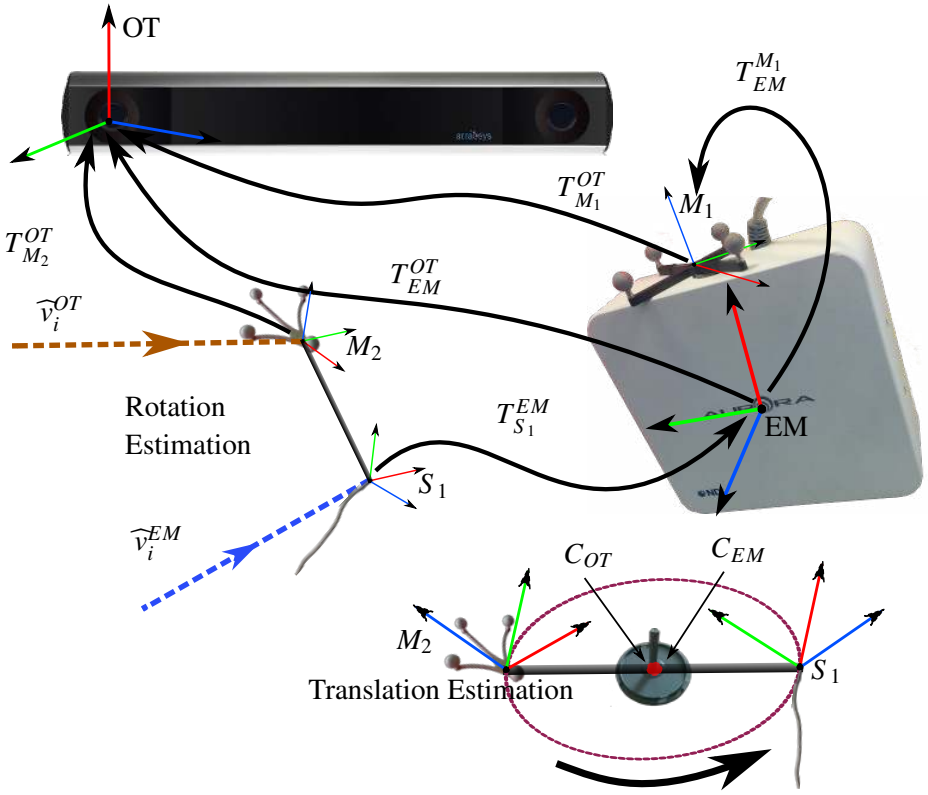


Figure 4.8: Calibration set up, M1 and M2 are optical markers. S1 is an electromagnetic sensor. The calibration is performed in two steps; line segments for estimating rotation and circle for estimation of translation.

4.2.4.2 Establishing Ground-Truth for Real Data

Now the next step to the quantitative evaluation on real data is to establish the ground-truth by following the steps outlined below;

Step-1 Record trajectory $\mathcal{I}_1$ as described in Section 3.2 for the first position of EMFE at EM_1.

Step-2 Move the EMFE to a new position EM_2. Remove and reposition the sensors at landmarks $\mathcal{L}_A$ and $\mathcal{L}_B$.

Step-3 At EM_2 perform a new recording $\mathcal{I}_2^{tmp}$, which is used to compute the closest point on the oesophageal trajectory to estimate the reference plane, $\mathcal{D}_2$. For this recording $\mathcal{I}_2^{tmp}$, only the reference plane $\mathcal{D}_2$ is retained and the rest of the recording discarded.

Step-4 Compute the ground-truth transform from the optical tracker as, $X = T_{EM}^{M_1} T_{M_{12}}^{M_{11}} T_{M_1}^{EM}$.

Step-5 Using the registration approach presented in Section 3.3 obtain $\hat{X}$ from registration $\mathcal{D}_1 \rightarrow \mathcal{D}_2$.

Step-6 Compute $\mathcal{I}_2 = X\mathcal{I}_1$, which represents the ground-truth.

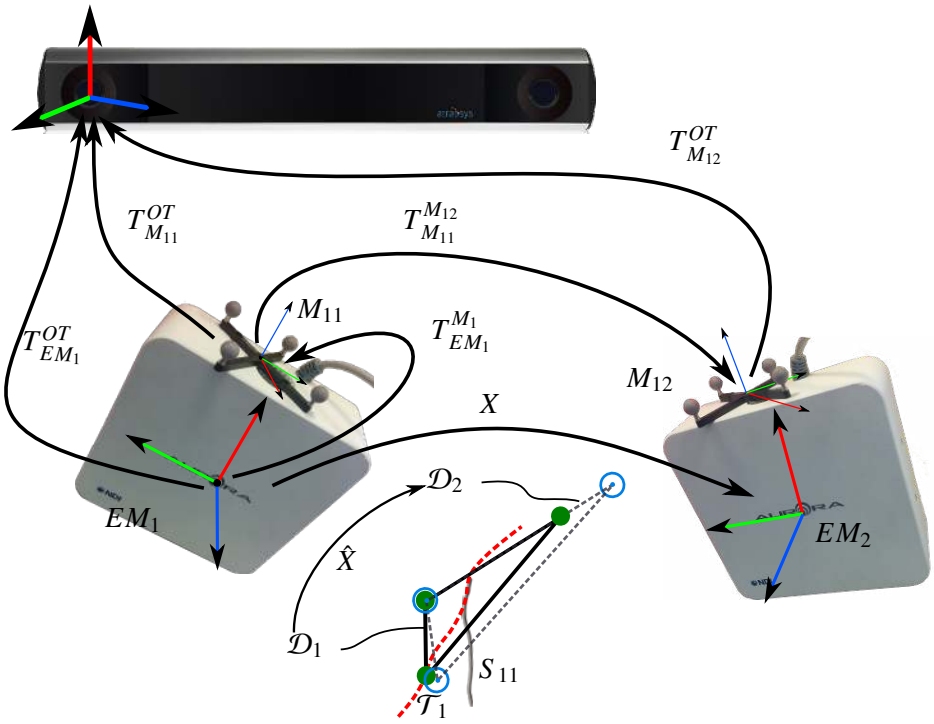


Figure 4.9: This figure corresponds to discussion presented in reference-to-ground-truth-real-section. The transformation is applied to the trajectories, $\mathcal{T}_2 = X\mathcal{T}_1$.

To compute the matching error both the trajectories $\mathcal{I}_2$ and $\hat{\mathcal{I}}_2 = \hat{X}\mathcal{I}_1$ were used and the matching, $f : \mathcal{T}_2 \rightarrow \hat{\mathcal{T}}_2$ is again bijective. The synchronization error projected along the oesophagus trajectory was obtained as in the synthetic case.

4.2.4.3 Results

In the experimental set up, an Atracsys InfiniTrack™ OTS, was employed. The EMTS and OTS were both configured on a Linux machine (i7 3.2 GHz octa-core processor) for synchronized capture using a Blackmagic Decklink Intensity pro acquisition card. 8

recordings were performed (with an average of 200 points in $\mathcal{T}$) for separate positions of the EMFE and the pig. Each recorded trajectory was registered and matched with the rest 7 recorded trajectories as described in Section 4.2.4.2. The final matching error along the oesophagus trajectory is shown in Figures 4.10 and 4.11. The figures indicate that the error results are comparable with those observed in the synthetic case. It thus indicates that our model is a reasonable choice and the empirical values of the uncertainties are relevant. A similar error phenomenon at the beginning and end of the trajectory was observed, as was previously explained in Section 4.2.2. In 96 percentile of the cases the error computed was in the range [10.9mm, 13.47mm], which is within reasonable limits as explained in Section 4.2.3.

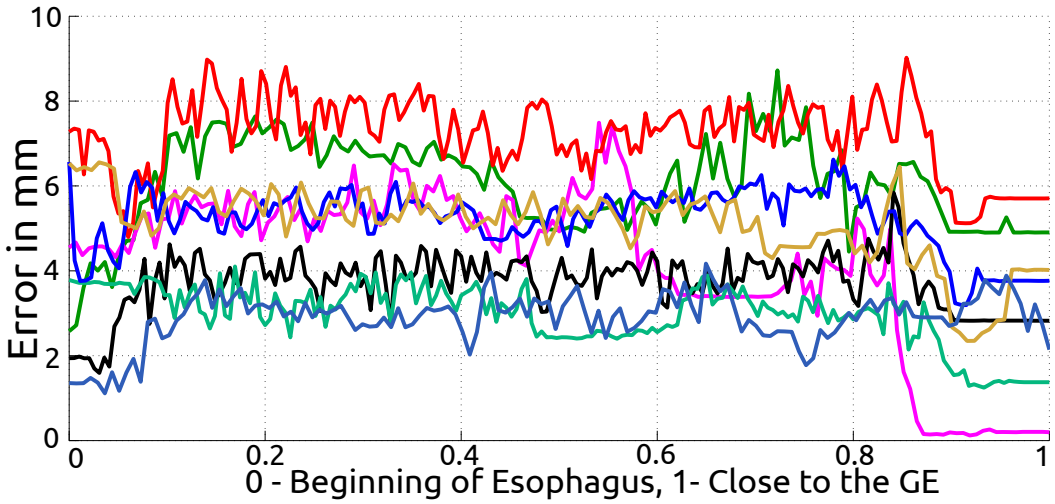


Figure 4.10: Each curve represents the mean error along the oesophagus for a trajectory, when synchronized with 7 other recorded trajectories. The above figure shows the result for 8 such trajectory synchronizations.

4.3 Qualitative Evaluation

This section describes the final set of experiments proposed for qualitative analysis of the system when applied to BSR relocalization. The evaluation was conducted to compare the performance of the proposed system against the classical approach currently being clinically followed. In the classical approach, a GI expert prints a picture of the locations where the biopsies were performed and notes the corresponding approximate length of the endoscope inside the oesophagus. For a follow-up procedure, the biopsy site are approximately relocalized using the endoscope length.

Here, a comparison is drawn between the performance of the proposed system against the classical approach using 10 gastroenterologists. The evaluation was conducted a posteriori on data collected from pigs. In order to simulate real conditions, two sets of recordings were performed in the pig's oesophagus. The first set of recordings ($\mathcal{S}_1$) were performed with no coagulation markings on the oesophagus tissue. Then, the clinician used a coagulation

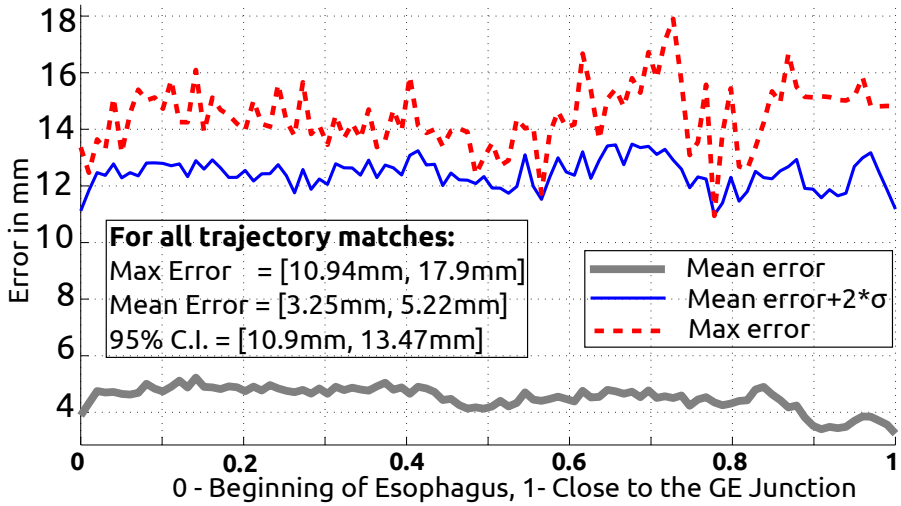


Figure 4.11: Error distribution along the oesophagus trajectory for all 56 trajectory matches for interventions on pig. This figure consolidates the results of the errors in Figure. 4.10

device to make 12 markings (tags) along the oesophagus lining. The markings corresponded to a typical four quadrant biopsy performed in any real diagnostic intervention. A second set of recordings ($\mathcal{S}_2$) were performed on the oesophagus tagged with coagulation markings. Between each recording of $\mathcal{S}_1$ and $\mathcal{S}_2$ all the sensors were removed and replaced and the EMFE was repositioned at a new location. For the actual experimental evaluation, each of the recordings from the first set $\mathcal{S}_1$ were used as simulated live interventions, which would be matched with each of the recordings from $\mathcal{S}_2$. Using a test environment built specifically for qualitative assessment, evaluation data was presented in two phases. The test environment presented the evaluator with a video sequence from an intervention in $\mathcal{S}_1$, which was navigated using a keyboard interface, to go back and forth on the recorded images.

In the first phase, the evaluator was provided with a printed picture of all the biopsy locations taken from $\mathcal{S}_2$ with their approximate length inside the oesophagus. As the sequence was being navigated the approximate distance inside the oesophagus was also presented on the screen. It corresponded to the length of the endoscope inside the oesophagus, which was shown as the approximated depth recorded using the EMTS. In the second evaluation phase, the proposed EMTS based system was used to provide a guided navigation. As the evaluator navigated the sequence of images from an intervention in $\mathcal{S}_1$, a corresponding matched image was presented from an intervention in $\mathcal{S}_2$ as shown in Figure 4.12.

In each phase, the evaluator was asked to select an ROI in the images from $\mathcal{S}_1$ where he/she felt confident about the possible location of the biopsy sites simulated by coagulation markings, by reviewing the images from $\mathcal{S}_2$. To establish the ground-truth, the expert who performed the original procedures, tagged the ROI in all the image sequences from $\mathcal{S}_1$; where the coagulation markings were made within few minutes after the completion of the original data collection, by closely reviewing the images from the sequences in $\mathcal{S}_2$. By

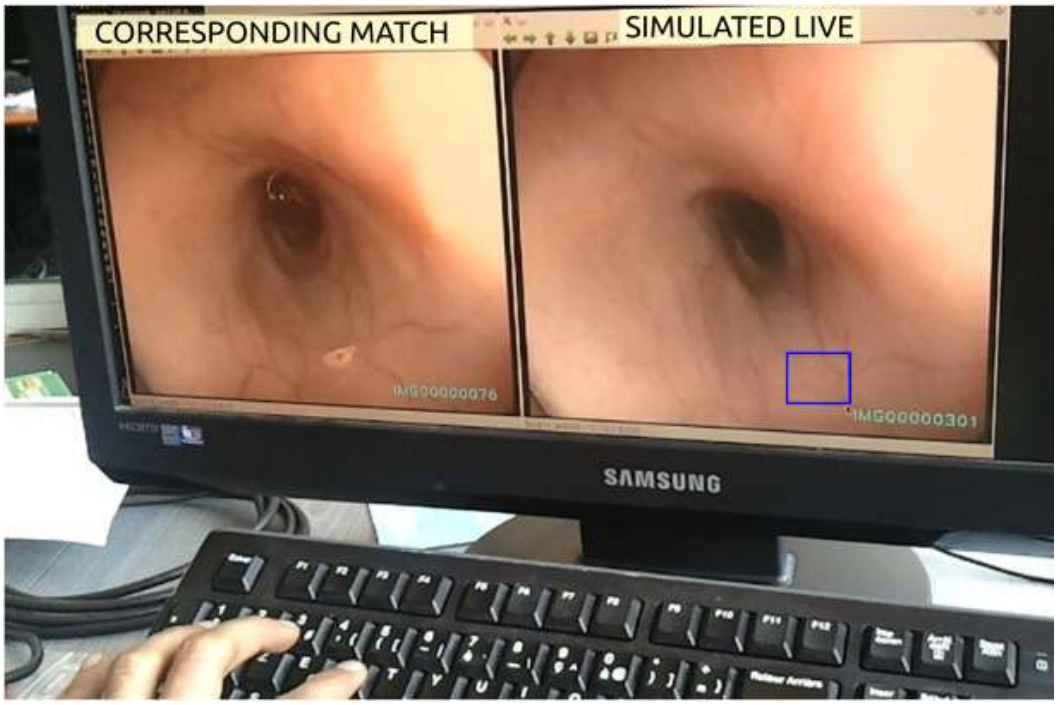


Figure 4.12: Interface for qualitative evaluation. Right image corresponds to the simulated live view. Left image corresponds to the image matched using the EM position. The blue rectangle is the marking made by the evaluator on the estimated location of the biopsy point in the simulated live view. The navigation between each frame is done using a keyboard. The evaluator has to mark on only a single image where he/she expects the coagulation markings were made in the matched view.

comparing the tags made by the evaluators against ground-truth, the relocalization accuracy was computed as the number of tags successfully identified by the clinician in 3 separate trajectory matches between sets $\mathcal{S}_1$ and $\mathcal{S}_2$. To avoid any bias, both phases of the evaluation were conducted on different days.

4.3.1 Results

Table 4.1 provides the results of qualitative evaluation of the proposed navigation system by 10 GI experts. As can be seen from the last two rows of the table, the rate of relocalization accuracy improved for each of the individual experts. It is evident that the proposed system and approach clearly improves the localized targeting in the oesophagus over the classical unguided methodology. The GI experts were instructed to mark all the regions which they could not find with sufficient confidence as “Not Found". Clinically, a “Not Found" marking would lead to sampling a larger portion of the oesophagus during the SE. It is typically better if the GI expert marks a region as “Not Found" than localize at an incorrect position in the oesophagus. Two important goals under clinical settings need to be achieved, to increase the confidence of the GI expert for IRO relocalization. First, by reducing the number of regions

that are marked as “Not Found”. Second, to reduce the number of incorrectly relocalized regions. This poses a greater challenge, because, ideally this number should be zero. It can be seen from Table 4.1, that, using the classical approach the number of region marked as “Not Found” were 34 which reduced to 4 using the guided navigation. The number of incorrectly matched locations decreased from 55 to 3.

In this experiment *True positive* was identified as a region that was marked as a biopsy site and was observed as one by the ground-truth. A *False positive* is, in this case a non-biopsy site that was marked as a biopsy site. The other conditions in the confusion matrix are not applicable here, since only the simulated biopsy sites were asked to be identified, as in any realistic situation. Based on this evaluation criteria, the overall accuracy for relocalization improved from 47.5% in the classical scenario to 94% using the EMTS based relocalization and the rate of recall improved from 63.3% to 97%.

4.4 Discussion

Thus far this chapter identified various errors affecting the system performance of the proposed framework in Chapter 3. The uncertainty was modelled as a Gaussian distribution. Three separate experiments were conducted to test the performance of the proposed system. The first experiment, identified the sources of uncertainty in the system; and determined the extent of noise empirically from the data recorded on 18 test subjects. This was then used to generate synthetic (but realistic) data. To report the performance, error was characterized to be the difference in depth (in millimetres) in the oesophagus between a matched image from DE and simulated live SE.

In the second experiment, data was collected from surveillance interventions on pigs. Ground-truth was established by using an OTS to record the position of the EMFE between two surveillance interventions. In the final experiment, an actual procedure was replicated offline and presented to the GI expert. For doing this, intervention data on pigs was used, but with biopsy sites simulated by using coagulation forceps. The experts were presented with the recording of an image from $\mathcal{I}_1$ (SE) and $\mathcal{I}_2$ (DE). This experiment demonstrated the clinical relevance of the proposed framework for IRO relocalization. The subjective assessment was quantified into a confusion matrix, which indicated improvement of the relocalization performance by reduction in the number of incorrectly localized regions.

Overall, the evaluation experiments provided satisfactory results for GI experts. The error of magnitude of up to 49 mm was observed in the synthetic experiments. However, such large errors correspond to the Gaussian tail, which is not expected in practical scenarios (Figure 4.5 and 4.7). Indeed, it is highly unlikely that the GI expert would position both the external EM sensors on the patient’s skin with an error above 10 mm. Certainly, the Gaussian should be truncated to reflect more precisely the true noise on the EM sensor positioning. It is thus expected that the maximum error that will be encountered should remain less than 15 mm. This expectation is well confirmed by the second set of experiments involving *in-vivo* data on pigs and realistic sensor positioning error. It was observed that maximum error average on 7 data-sets was less than 14 mm (Figure 4.11), which corresponds to an acceptable visual discrepancy by the GI experts.

	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10
	Classical Approach *(error in depth "cms")*									
B1	≤ 1	≤ 1	≤ 1	≤ 1	NF	≤ 1	≤ 1	≤ 1	≤ 1	≤ 1
B2	≤ 1	≤ 1	≤ 1	3.5	NF	≤ 1	≤ 1	≤ 1	≤ 1	≤ 1
B3	≤ 1	≤ 1	NF	3	NF	≤ 1	≤ 1	≤ 1	≤ 1	≤ 1
B4	5	≤ 1	≤ 1	≤ 1	NF	≤ 1	≤ 1	4.2	≤ 1	4.2
B5	5.9	≤ 1	NF	≤ 1	4.5	NF	≤ 1	5.3	≤ 1	6.1
B6	NF	≤ 1	NF	≤ 1	NF	NF	≤ 1	4.1	≤ 1	NF
B7	≤ 1	≤ 1	NF	≤ 1	NF	NF	≤ 1	5.1	≤ 1	NF
B8	≤ 1	≤ 1	NF	4.3	NF	NF	≤ 1	5.5	≤ 1	NF
B9	≤ 1	≤ 1	NF	5.8	6.3	NF	NF	NF	≤ 1	≤ 1
B10	5.4	5.1	NF	≤ 1	3.2	≤ 1	3.5	NF	3.9	NF
B11	4.3	3.8	NF	NF	4.3	≤ 1	3.9	NF	4.1	≤ 1
B12	4.1	6.1	NF	NF	4.4	≤ 1	4.1	NF	≤ 1	NF
# Err ≤ 1cm	**6**	**9**	**3**	**6**	**0**	**7**	**8**	**3**	**10**	**5**
Accuracy	**50%**	**75%**	**25%**	**50%**	**0%**	**58%**	**67%**	**25%**	**83%**	**42%**
	Synchronization Approach *(error in depth "cms")*									
B6	≤ 1	≤ 1	≤ 1	≤ 1	≤ 1	2.8	≤ 1	≤ 1	≤ 1	NF
B9	≤ 1	≤ 1	≤ 1	≤ 1	≤ 1	3	≤ 1	≤ 1	NF	≤ 1
B10	NF	≤ 1	≤ 1	≤ 1	≤ 1	≤ 1	≤ 1	≤ 1	≤ 1	NF
B11	3.3	≤ 1	≤ 1	≤ 1	≤ 1	≤ 1	≤ 1	≤ 1	≤ 1	≤ 1
# Err ≤ 1cm	**10**	**12**	**12**	**12**	**12**	**10**	**12**	**12**	**11**	**10**
Accuracy	**83%**	**100%**	**100%**	**100%**	**100%**	**83%**	**100%**	**100%**	**92%**	**83%**

Table 4.1: The above table provides the details of the biopsy sites as marked by 10 clinical experts (columns). The rows marked B1-B12 indicate the biopsy sites. Each cell (in rows B1-B12) provides the error in depth (in cms) for identifying a biopsy site by the expert for the classical approach. The label NF implies the biopsy site was "Not Found" by the expert. For the synchronization approach only the biopsy sites incorrectly identified or marked NF have been presented here. The % values indicate the accuracy.

The third set of experiments clearly indicated the benefit to the medical community. They corresponded closely to the clinical work-flow, and measured the extent to which the system could help in the task of relocalization of biopsy sites. This experiment also demonstrated that an approximate depth information (like the markings on the endoscope shaft) is clearly not sufficient for relocalization. This was observed from the results in Table. 4.1, that the experts could not retrieve simulated biopsies with the same accuracy, in comparison to guided navigation discussed in Chapter 3.

Synthetic experiments were organized to be as close as possible to the realistic scenario, in terms of noise affecting the process. However, the possible deformation due to breathing or due to patient repositioning on the table was not modelled. We believe the latter should have little influence on the system accuracy, since an analysis of the thoracic CT of the same patient indicated no variation in the oesophagus's relative shape and position. However, this should be rigorously assessed; but a first comparison on several patients showed this trend.

The case of a patient taking long and deep breaths, which would result in a large shift (1 cm or more [Yaremko 2008]) in the relative position of the gastro-oesophageal junction should not be neglected. Practically however, in GI endoscopy, the patient is always under general anaesthesia and the breathing is regulated. Thus, we believe that the elongation of the oesophagus at its distal end would be negligible. Further experiments on patients are nevertheless necessary, to assess this phenomenon. However, since the proposed approach uses external sensors on the sternum; by studying the relative displacement of the sensor at B ($\mathcal{L}_b$), a dynamic model of the oesophagus length, as a function of the relative position of the sensor P_i^B can be proposed, as in [Hostettler 2007].

The accuracy of EM apparatus can degrade in presence of distortion due to ferromagnetic equipment. However, in the current application, the equipment is compatible with the EMTS and no EM tracking errors were observed during a procedure. Additionally, due to the limited working volume, the extremities of the EMFE workspace are never reached, hence the EMFE accuracy should remain within 1 mm, as specified by the manufacturer. However, in some cases when the patient's size is large, it would be necessary to provide a robust approach to correcting errors in the EM tracker [Luo 2014, Reichl 2013].

Although the video synchronization based navigational system is already very useful for the GI experts; a few issues still need to be addressed to reach a complete automated navigational help.

(a) *Oesophagus model*: A typical surveillance procedure includes a few densely captured regions (corresponding to biopsy sites) and many sparsely populated regions recorded in $\mathcal{T}$. A patient specific model, that would include a smooth trajectory representing the central line in the oesophagus and corresponding frames with clear lumen visibility, is necessary to eliminate redundant data and provide seamless video synchronization. This could be further used to generate patient specific virtual models for training using simulators. The generation of the oesophagus model will be discussed in Section 5.1.

(b) *Improving registration*: In addition to using the three landmarks for registration as discussed so far in this chapter; incorporating a patient specific oesophagus model could aide as a secondary approach to registration in the presence of errors due to placement of sensor at $\mathcal{L}_A$ and $\mathcal{L}_B$. This is also necessary for patients who are bulky, when placement of the secondary sensor accurately can be difficult. Additionally, the sensor $\mathcal{L}_B$ can be eliminated from the set up when using the oesophagus trajectory for registration. However, this would cause an additional delay in the registration, due to the need for acquiring more points on oesophagus trajectory. Along with an approach to form an oesophagus model, a more generalized registration using the complete trajectory will be presented in Chapter 5.

(c) *Scene matching*: The navigational guidance discussed thus far presents a matched image, only in terms of the depth in the oesophagus. However, such an image may not contain the right view point corresponding to the live-view. Additionally the matched image could be uninformative. Choosing the image from $\mathbf{I}_2 \in \mathcal{I}_2$ in the vicinity of the match obtained using the EM tracker, such that, the image best matches the live view in terms of its visual content; is an important aspect that needs to be tackled. A complete understanding of the visual information affecting the GI procedures will be presented in Chapter 7.

CHAPTER 5

Inter-operative Registration: Using Complete Oesophagus Trajectory

Contents

It was suggested in the previous chapter that the 3-point approach can introduce error in registration due to the approximate placement of external sensors at $\mathcal{L}_A$ and $\mathcal{L}_B$. The error in $\mathcal{L}_A$ reflects in the translational component along the oesophagus depth and in case of $\mathcal{L}_B$, it takes effect as an error in rotation affecting the alignment of oesophagus trajectories, $\mathcal{T}$ between DE and SE recordings as was shown in Figure 4.7. Such an error is unlikely to occur normally during a procedure, however, it can be attributed to error in placement of sensors due to human factors or caused by the displacement of the sensors between the time of their placement and the patient settling into the right position before intervention. Such a situation must be detected and appropriate correction must be presented to avoid critical misguidance during the procedure. In this chapter, an extension to the 3-point registration approach introduced in Chapter 3 is presented, which uses the complete oesophagus trajectory. A typical GI endoscopic intervention, contains regions of continuous motion in the lumen and those containing the highly dense regions of biopsy as shown in Figure 5.1. As was noted in Section 4.4, generation of a patient specific oesophagus model would aide in providing a smooth trajectory inside the oesophagus. Using a patient specific model, jumps in video after synchronization can be avoided and a jitter-free video synchronization can be provided.

This chapter is organized as follows. Section 5.1 presents an algorithm for the formulation of an oesophagus model that is based on point cloud decimation. Section 5.2 reviews the literature on point cloud registration which is used in Section 5.3 to present the methods selected for comparative evaluation against the 3-Point registration approach (from Section 3.3). The experiments are described in Section 5.4, indicating that using the complete oesophagus trajectory improves the overall accuracy of the system by almost 50% and handles the worst case error scenarios that were encountered during the evaluation in Chapter 4.

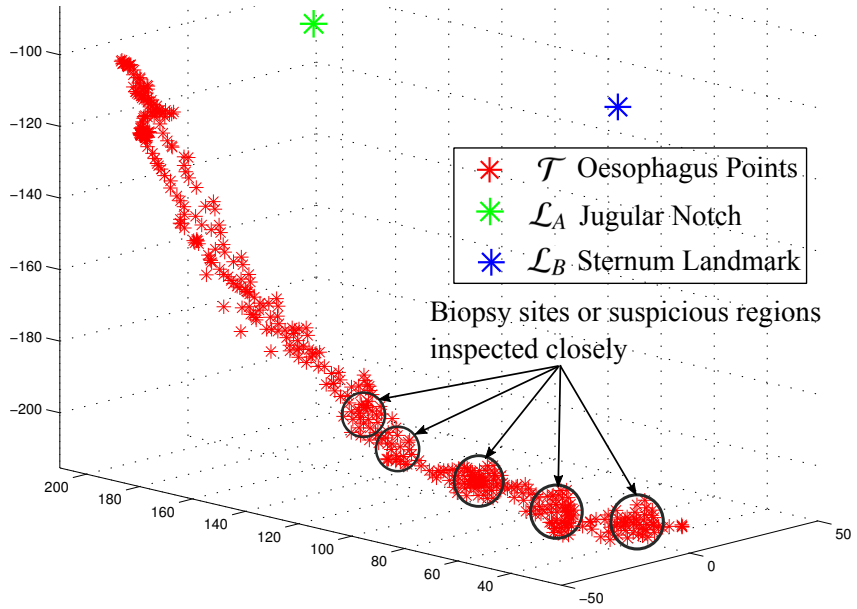


Figure 5.1: Depicting regions of continuous motion and clustered point clouds representing locations of biopsy sites.

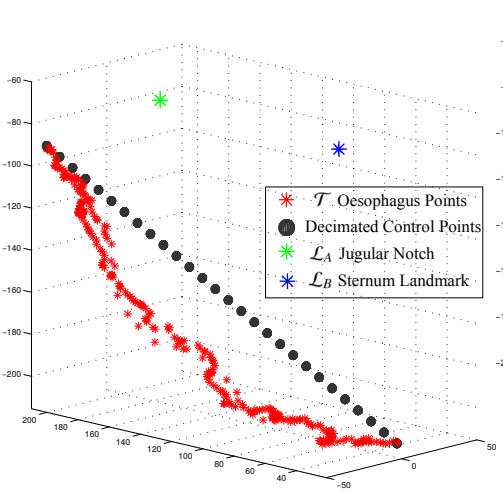
5.1 Formulation of Model of Oesophagus Trajectory

This section presents a simple approach to formulate an oesophagus model. An important criteria while formulating this model, is to subsample the points such that, they represent uniform non-overlapping subsets of the original trajectory. This is necessary to provide consistent subsampled versions from each new intervention performed on the same patient. Although, the oesophagus is relatively rigid and fixed at both ends; in noisy versions of the original trajectory, back and forth movements during endoscopy and contractions in the oesophageal muscles, introduce deviations which should be smoothed before registration.

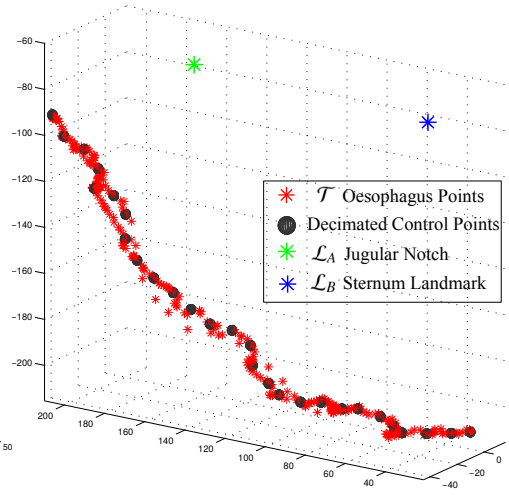
Algorithm-1 describes the approach to sub-sample the trajectory. An appropriate value for r (radius of point decimation) must be selected based on the knowledge that between swallows (the automatic muscle contractions to smoothly pass the food), the oesophagus is collapsed, but the lumen can distend to approximately 20 mm in the anterior-posterior dimension and up to 30 mm laterally to accommodate a swallowed bolus [Kuo 2006]. In case of a diseased oesophagus, for example, in case of oesophageal stenosis, the diameter can shrink to less than 13 mm, at which point swallowing solid food becomes almost impossible. Since the gastroscope has a longitudinal slender shape and there is intermittent insufflation during endoscopy, an appropriate choice of lumen diameter would be between 18-20 mm.

With the knowledge that the oesophageal trajectory is convex, the line segment joining the endpoints of the trajectory, can be decimated to obtain a first down-sampled version. Then using the chosen value of r, firstly a stable centroid of points within the radius r is established as an initialization of a control point. With each new control point initialized, its position is optimized under the constraint that its distance from the previous point is $2 \times r$.

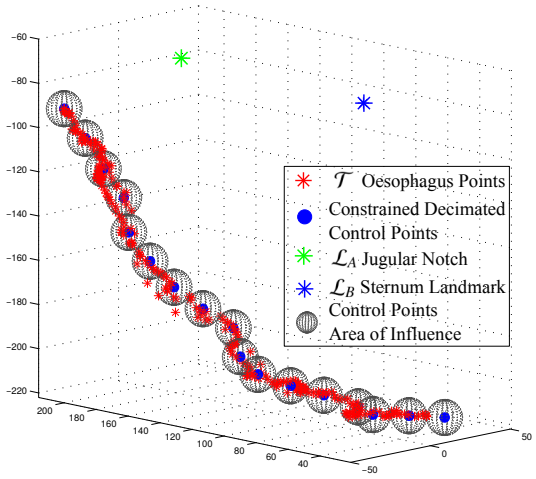
Figure 5.2 shows the results of the intermediate steps of the oesophagus model generation.



(a) Initialized model control points for oesophagus trajectory using the knowledge of convexity of oesophagus. The end points of the trajectory are connected using a line segment and decimated.



(b) The decimated points on the line segment are then mapped on to the oesophagus by considering the nearest neighbour point on the recorded trajectory. The mapped points represent an initial down-sampled version of the oesophagus trajectory.



(c) The down-sampled points are then optimized to obtain control points that are separated by $2 \times r$ from each other. The spheres indicate the region of influence of the control points.

Figure 5.2: (a) (b) (c) Intermediate steps for the methodology presented in Algorithm 1.

5.2 Registration of Point Clouds

Registering 3D point clouds is an important task in a variety of applications such as, computer vision, biometrics [Zhang 2009], 3D model construction from multi-directional

Algorithm 1 Modelling Oesophagus Trajectory

Require: $\mathcal{T}$ (Oesophagus trajectory point cloud). r = Radius for final constrained decimation.

function FitCurve($\mathcal{T}$, r)
- $C \leftarrow \emptyset$ ▷ Initialize control points
- $n + 1$ =number of anchor points.
- Compute c_0 and c_n anchor points as extremities of $\mathcal{T}$.
- Generate $c_k = (n - 2)$ equispaced linear points between c_0 and c_n.
- Assign $C = [c_0, ..., c_n]$.
- $C \leftarrow$ nearest neighbour in $\mathcal{T}$ $\forall$ C.
- $C \leftarrow$ mean of associated nearest points in $\mathcal{T}$ for each C.
- Eliminate anchor points C not linked to their mean.
- Redistribute points using linear interpolation.
- $\nu \leftarrow \emptyset$ and $i = 0$ ▷ Initialize constrained control points.
- **while** *size_of*(C)>0 **do**
  - Append $C(0)$ to C
  - **while** $\varepsilon > 1e - 3$ **do** ▷ Loop to initialize control point.
    - $\kappa \leftarrow C(i)$
    - $\gamma \leftarrow ||C(i) - C||^2$
    - $\Omega = C(\gamma <= r * r)$
    - $C(i) \leftarrow median(\Omega)$
    - $\varepsilon = ||C(i) - \kappa||^2$
  - $x_0 \leftarrow C(i)$ ▷ Initial condition for optimization.
  - **if** $i > 0$ **then** ▷ Optimize current point to be placed at distance 2*r.
    - $\Omega = C(\gamma <= 4 * r * r)$.
    - $C = C - (C \cap \Omega)$
    - Solve:

$$\arg\min_{x} \sum_{i=1}^{N} ||x - \Omega||^2$$
$$subject\ to: \quad ||x - C(i-1)||^2 = 4 \times r \times r$$

    - $\nu(i) = x$.
  - $i \leftarrow i + 1$
- **return** ν.

range-based images [Armesto 2010], computer-assisted medical applications [Cash 2007], etc. The goal of registration of point clouds is to be able to provide an assignment of correspondences between two sets of points and to recover the transformation that maps one point cloud to the other. For the registration of oesophagus trajectories modelled using the approach presented in the previous section, we are interested in a rigid registration framework. Several methods have been proposed in the literature to tackle this problem. Starting from simple deterministic methods such as Principal Component Analysis (PCA) and Singular Value Decomposition (SVD), to more recent methods such as the Iterative Closest Point (ICP) and probabilistic approaches such as the Coherent Point Drift (CPD).

Rigid registration of two point clouds, can be formulated by defining a cost function, that represents the matching error between corresponding point clouds. This cost function is then minimized using various optimization techniques. If the Euclidean distance between corresponding points in each point cloud is minimized, this can then be simplified to a linear least-squares minimization problem by representing each point using homogeneous coordinates as in Equation B.1. The closed form solution can be obtained using PCA and SVD based methods as described in Appendix B. Whereas, the PCA and SVD based approaches directly solve the least-square problem, assuming perfect data; Besl and Mc. Kay [Besl 1992] and Chen and Madioni [Chen 1991] introduced a method that iteratively discards outliers in order to improve upon the estimate of rotation and translation at each step (as illustrated in the flowchart in Figure 5.3). Several variants of this method have been proposed in literature but [Besl 1992] came to be termed as the standard ICP algorithm. Rusinkiewicz and Levoy [Rusinkiewicz 2001] classified the various ICP algorithms based on the following sub-tasks;

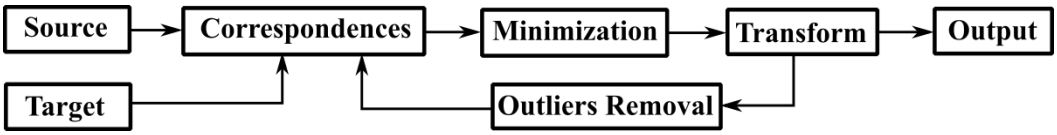


Figure 5.3: Standard ICP scheme.

1. *Selection* of points from the two point clouds to be registered, is usually performed to speed-up the algorithm and/or to eliminate outliers.

2. *Matching* of the selected subset for the two point-sets. The original ICP algorithm by Besl and McKay [Besl 1992], employed the euclidean distance metric for establishing point correspondences. Several other variants have been proposed, for example those using mahalanobis distance [Granger 2006, Maier-Hein 2012].

3. *Weighting* the corresponding matches appropriately. Several authors have also investigated assigning weights to point pairs, in order to improve the robustness of the algorithm. The proposed methods range from assigning weights based on the interpoint distance [Godin 1994] or weighting based on the noise characteristics as in [Rusinkiewicz 2001]. However, the weighting is generally performed using scalar values or along certain directions of a common coordinate system [Kaneko 2003].

4. *Outliers* have a large effect on least-squares minimization based methods. Several approaches have been proposed to detect and reject pairs iteratively that could have a negative effect on the registration. Among others, rejection of the worst percent of matched pairs [Pulli 1999, Chetverikov 2005], excluding pairs with distance above a certain threshold [Rusinkiewicz 2001], have been proposed.

5. *Error metric and its minimization*: Commonly, the sum of squared distances between correspondences is minimized, using one of the few closed form solutions for determining the rigid transformation, that minimizes this error; for example as described in [Horn 1987]. Alternatively, point-to-point metrics or point-to-plane metrics have been proposed [Rusu 2010, Low 2004, Chen 1991]; as shown in Figure 5.4. In [Granger 2006] authors proposed to formulate ICP as maximum likelihood problem and use expectation-maximization to solve for optimal transform. Several other variants have also been presented in literature; an overview of these can be found in [Maier-Hein 2012, Granger 2006, Bellekens 2014].

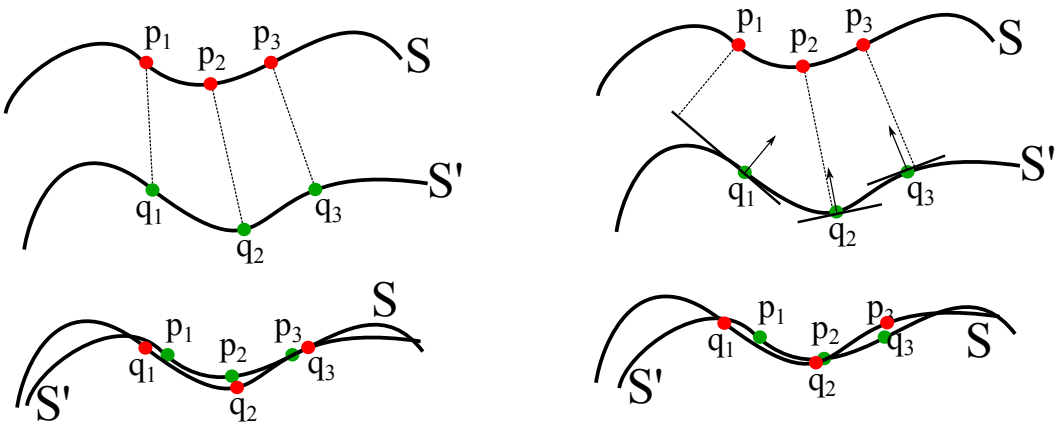


Figure 5.4: (*Left*): ICP alignment point to point, (*right*): ICP alignment point to plane.

5.3 Registration Methods Considered for Comparison

Based on the criteria described in the previous section, the selection of points from the point clouds recorded during gastroscopy, is performed using the method described in Section 5.1. The euclidean distance metric is selected here for ICP based registration. It should be noted that, with a model of uncertainty on each control point on the trajectory, Mahalanobis distance can also be employed. However, this has not been considered here, but can be easily extended using statistical information obtained from several interventions recorded on the same patient under varied control scenarios.

In the following section, an evaluation of the performance of rigid registration, using the standard ICP and three other methods; against the 3-point approach presented in Section 3.3, is performed. The first of these approaches is a RANSAC based standard ICP, that attempts to iteratively filter outliers from the decimated oesophagus trajectory, during registration.

The second approach in consideration here (was proposed in [Stewart 2003]); is a variant of ICP, that formulates the standard ICP objective function as an iterative re-weighted least squares problem, which is very similar to the weighting of point pairs; however, the weights are estimated using a robust M-estimation criterion (termed as IRL-ICP). Consider the points p_i $i \in [1, N]$ are to be registered with model points X, the least squares formulation can be modified as follows;

$$\min_{\mathbf{R},\mathbf{t}} \sum_{i=1}^{N} \varrho(d(\mathbf{R}p_i + \mathbf{t}, X)) \tag{5.1}$$

Where, d is the distance operator; in the standard ICP it is the $\mathbf{L_2}$ norm and ϱ is the robust criterion function. If $\psi(r) = \varrho'(r)$ is the derivative of ϱ then, a robust criterion function must satisfy the following properties;

1. $\varrho(r)$ is an even function and $\boldsymbol{C}^1$ continuous on $\mathbb{R}$, and $\varrho(r) = 0$,
2. $\varrho(r)$ is monotonically increasing on $[0, \infty)$, and
3. $\psi(r)/r$ is monotonically decreasing and bounded above on $(0, \infty)$.

The weight function for the given $\varrho(r)$ can be defined as,

$$w(r) = \begin{cases} \dfrac{\psi(r)}{r} & \text{if} \quad r \neq 0, \\ \lim\limits_{r \to 0} \dfrac{\psi(r)}{r} = \psi'(0) & \text{if} \quad r = 0. \end{cases} \tag{5.2}$$

In [Bergström 2014] the authors compare three robust criterion functions namely, Huber's, Cauchy's and Beaton-Turkey's bi-weighted functions. Turkey's and Cauchy's methods exhibits the strongest protection against influence of outliers, however, have the tendency to create local minima. Here, the Huber's criterion function was considered for the evaluation, because it does not suffer from the same situation. However in this particular case, with the use of an oesophagus model, the number of sample points in consideration, are very few and the possibility of reaching a local minima is very low.

The third is a probabilistic approach, introduced by Myronenko and Song [Myronenko 2010], to register two point clouds and was termed as the Coherent Point Drift (CPD) algorithm. For both rigid and non-rigid point sets, they formulated the alignment of two point clouds as a probability density estimation problem. They fit the Gaussian mixture model (GMM) centroids (representing the first point cloud) to the data (the second point cloud) by maximizing the likelihood and forcing the GMM centroids to move coherently as a group to preserve the topological structure of the point sets.

In summary, the following section, evaluates of the performance for rigid registration of the complete oesophagus trajectory, with standard ICP and ICP with RANSAC to filter outliers, robust ICP proposed by [Bergström 2014] and CPD; against the 3-point approach.

5.4 Comparative Evaluation of Registration

The approach presented in Section 4.2.1 was followed for generating synthetic sample trajectories. Figure 5.5 presents a few sample synthetic trajectories used for evaluating the registration. Figure 5.6 presents the modelling of these synthetic trajectories using the proposed decimation algorithm. Four trajectories recorded on two pigs and 1800 synthetic datasets were generated from each. Inter-operative registration was then performed using the four additional methods described in the previous section. The results of the registration were compared by computing the video synchronization error that was described in Section 4.2.2. This was then compared with the error statistics computed using the 3-point registration.

Figures 5.7, 5.8, 5.9, and 5.10 show the plot of the mean and standard deviation of error along the trajectory. The results of synchronization error have been presented in Table 5.1, and highlights the important results for comparison. In general, using the oesophagus trajectory in addition to the 3-point registration does improve the accuracy of video synchronization. Specially, the RANSAC-ICP, IRL-ICP and CPD algorithms provide the best registration accuracies, by eliminating outliers effectively. The results from Dataset-4 however, shows only a marginal decrease in error from the 3-point approach, which requires further investigation.

Specifically, the methods reduces the synchronization error due to placement of sensor at the jugular notch by half. Additionally, using the complete oesophagus information reduces the error in the 99.7 percentile cases, which is where errors as large as 4.9 cms were observed (Section 4.2.3).

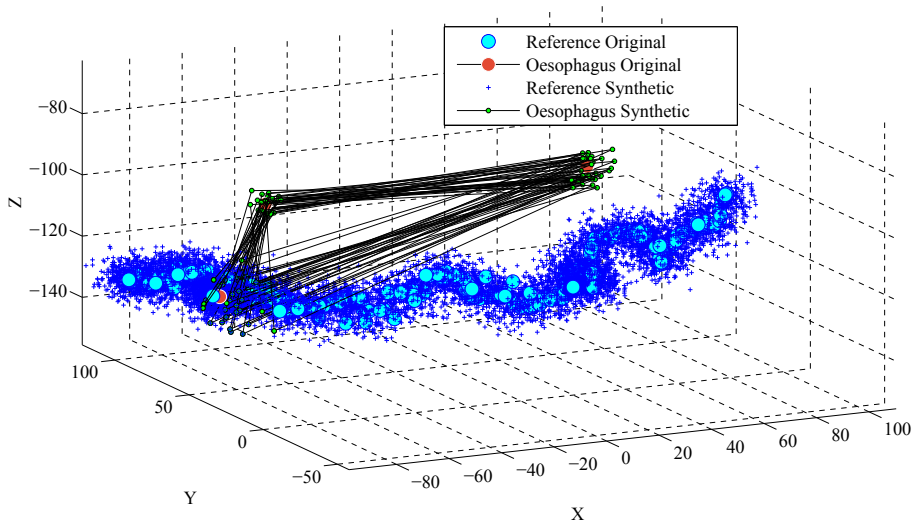


Figure 5.5: 15 Noisy sample trajectories generated using the Gaussian noise model presented earlier in the chapter are plotted along with the original recorded trajectory.

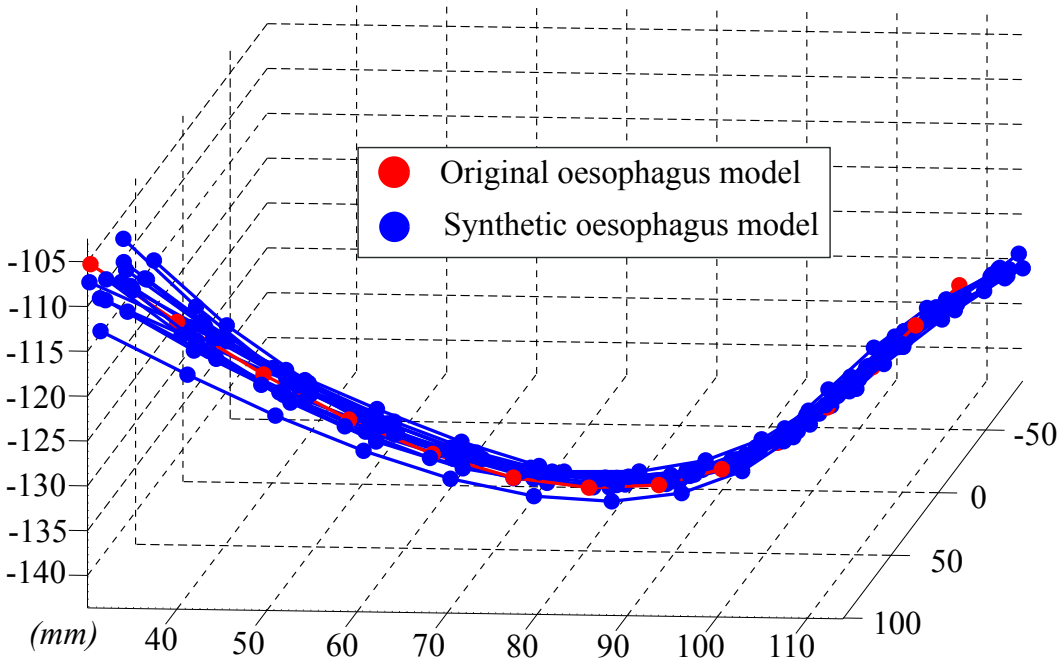


Figure 5.6: Original and 15 noisy sample trajectories modelled using method presented in Algorithm-1 are shown here. We observe that the noisy trajectory is smoothed out by the oesophagus modelling.

5.5 Conclusion

The results obtained in this chapter are significant, since they provide an incremental improvement to the methodology presented in Chapter 3. It addresses some of the problems that were discussed in Chapter 5 during the evaluation of the 3-point registration methodology. The presented approach, still fits the clinical work-flow without the need for any additional steps. All these corrections are necessary for qualitative certification of the system for introduction into the clinical routine. However, as was discussed in Section 5.3 the use of uncertainty based distance metric such as the Mahalanobis should also be verified. However, from a practical standpoint we believe that this approach would not further reduce the synchronization error.

This initial result obtained in this analysis, warrants further investigation using intervention data obtained on human subjects to verify this observation. From the qualitative experiments performed in Chapter 4, the 3-point registration methodology is a reasonable choice for the oesophageal procedures. It allows the process to be simplified without additional overhead in the work flow of the clinical routine. This should be verified for the complete oesophagus trajectory based registration as well.

Figure 5.11 updates the work-flow presented in Chapter 3 with the registration refinement using oesophageal trajectory.

		Mean	95.45 percentile	99.7 percentile
Dataset-1	3-Point	[4.15, 5.12]	[10.65, 13.22]	[13.90, 17.27]
	ICP	[2.30, 3.26]	[6.40, 9.36]	[8.45, 12.41]
	RANSAC-ICP	[2.10, 3.05]	[6.87, 7.36]	[8.78, 9.99]
	IRLS-ICP	[1.95, 2.85]	[5.59, 8.55]	[7.41, 11.40]
	CPD	[2.01, 2.69]	[5.63, 8.19]	[7.44, 10.94]
Dataset-2	3-Point	[4.23, 5.18]	[10.45, 13.32]	[13.56, 17.39]
	ICP	[2.61, 3.45]	[6.01, 9.47]	[7.71, 12.48]
	RANSAC-ICP	[2.39, 3.11]	[5.99, 8.41]	[7.79, 11.06]
	IRLS-ICP	[2.35, 2.91]	[5.85, 9.11]	[7.60, 12.21]
	CPD	[2.40, 2.61]	[5.70, 7.01]	[7.35, 9.21]
Dataset-3	3-Point	[3.20, 4.70]	[7.82, 12.20]	[10.13, 15.95]
	ICP	[2.41, 3.74]	[5.83, 10.08]	[7.54, 13.25]
	RANSAC-ICP	[2.35, 3.15]	[5.35, 8.53]	[6.85, 11.22]
	IRLS-ICP	[2.25, 2.83]	[6.25, 8.93]	[8.25, 11.98]
	CPD	[2.31, 2.55]	[5.89, 6.97]	[7.68, 9.18]
Dataset-4	3-Point	[2.35, 2.65]	[5.37, 6.95]	[6.88, 9.10]
	ICP	[2.90, 4.35]	[5.80, 9.95]	[7.25, 12.75]
	RANSAC-ICP	[2.25, 3.15]	[5.27, 7.99]	[6.78, 10.41]
	IRLS-ICP	[2.40, 2.80]	[5.70, 8.10]	[7.35, 10.75]
	CPD	[2.49, 2.60]	[5.39, 6.36]	[6.84, 8.24]

Table 5.1: The table shows the error statistics of quantitative evaluation for four datasets recorded on two pigs. All values are in *millimetres*. The analysis was conducted by generating synthetic data using the noise model presented in Section 4.2.1. Column-1 shows the mean error, columns-2 and 3 show the 95 and 99 percentile errors along the trajectory. Each cell in the table includes a [minimum,maximum] value of the respective errors along the trajectory.

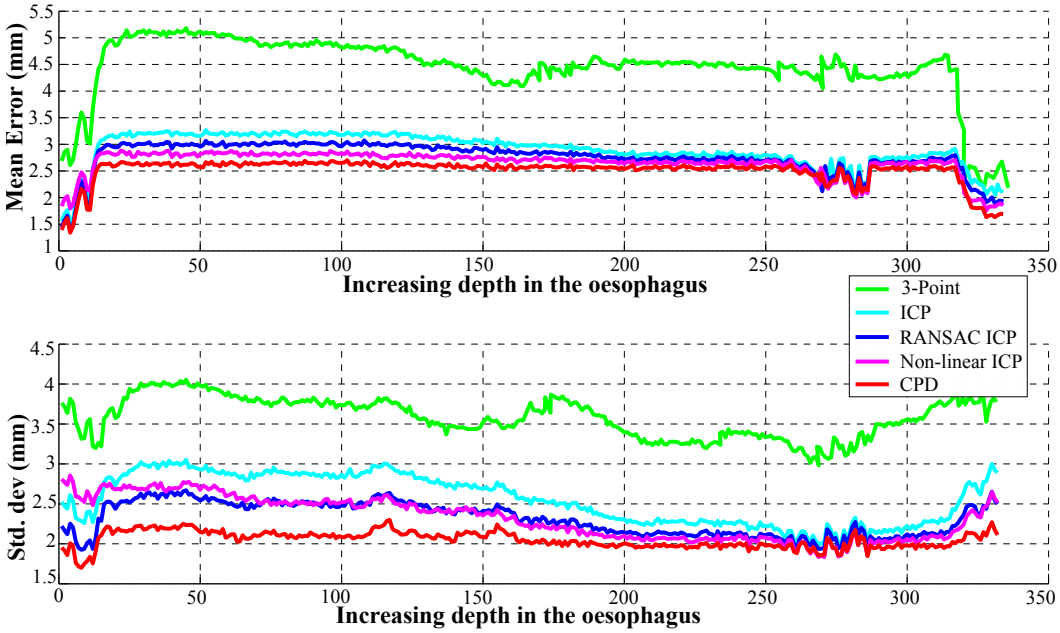


Figure 5.7: Figure shows the mean (top) and standard deviation (bottom) for analysis on Dataset-1 using the four registration methods.

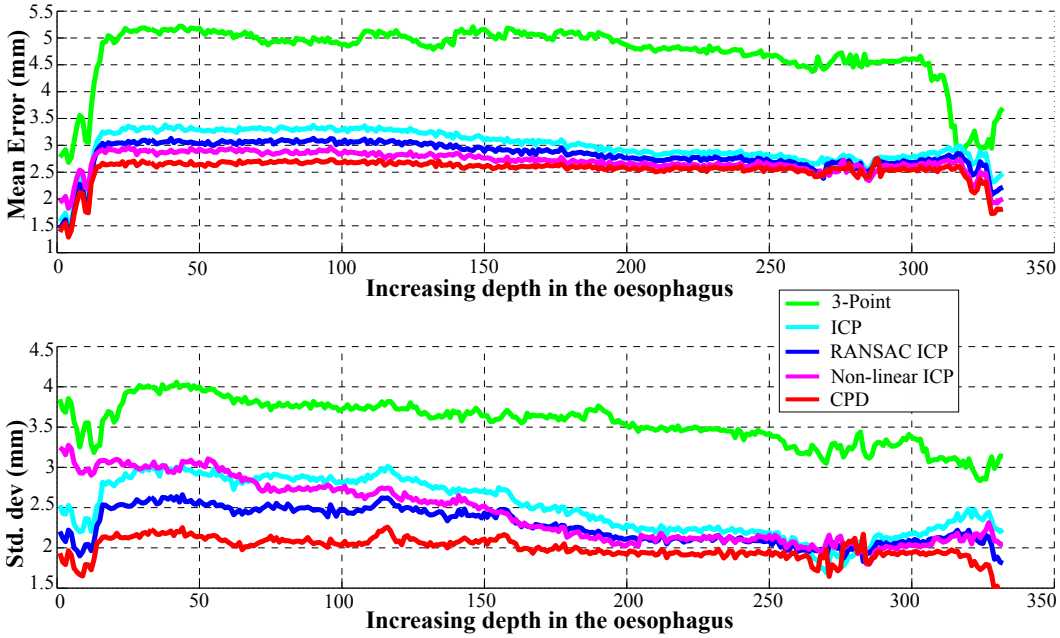


Figure 5.8: Figure shows the mean (top) and standard deviation (bottom) for analysis on Dataset-2 using the four registration methods.

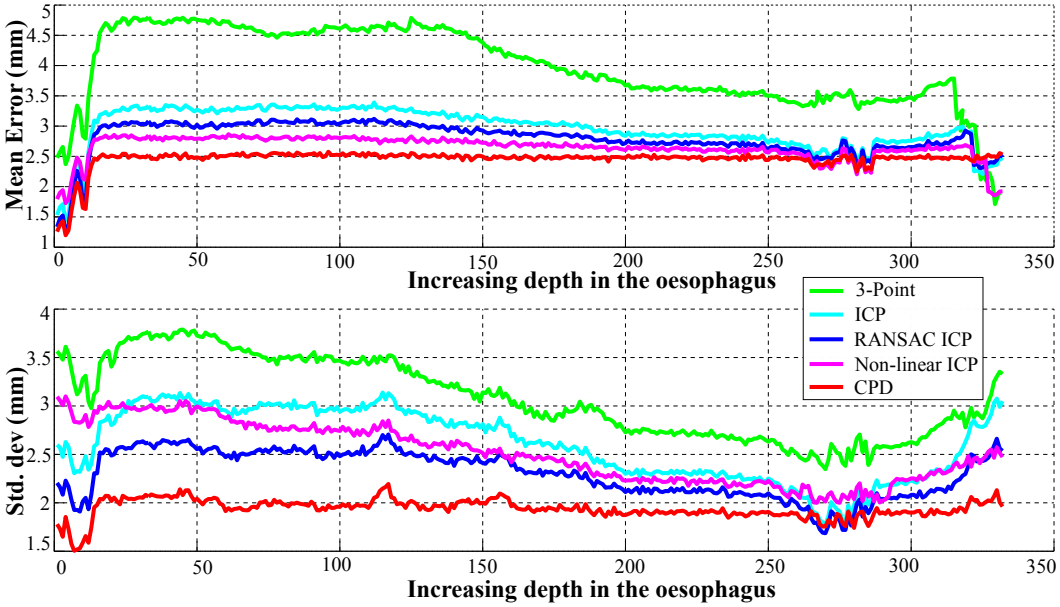


Figure 5.9: Figure shows the mean (top) and standard deviation (bottom) for analysis on Dataset-3 using the four registration methods.

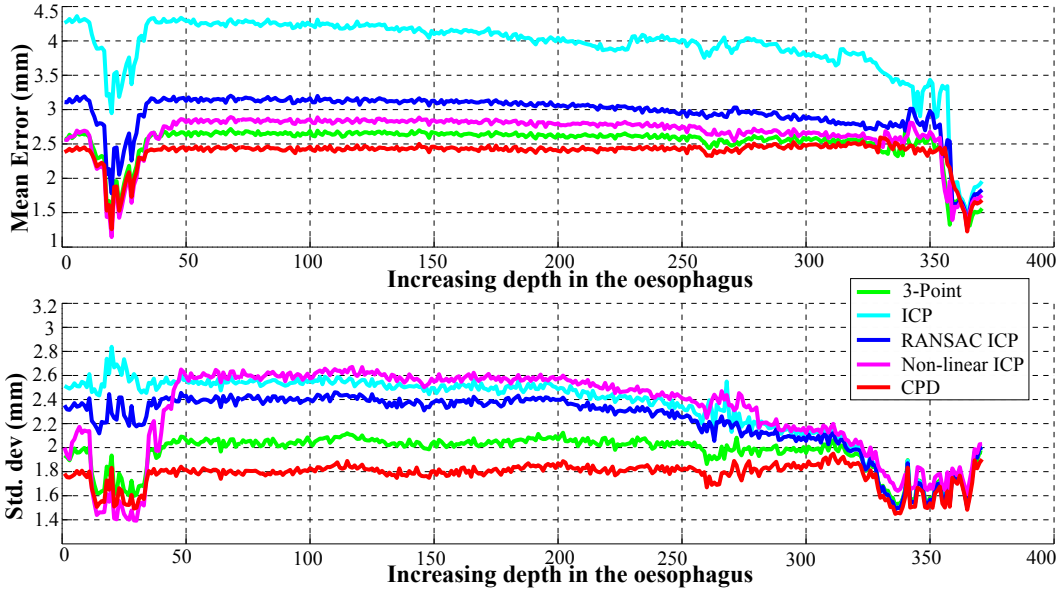


Figure 5.10: Figure shows the mean (top) and standard deviation (bottom) for analysis on Dataset-4 using the four registration methods.

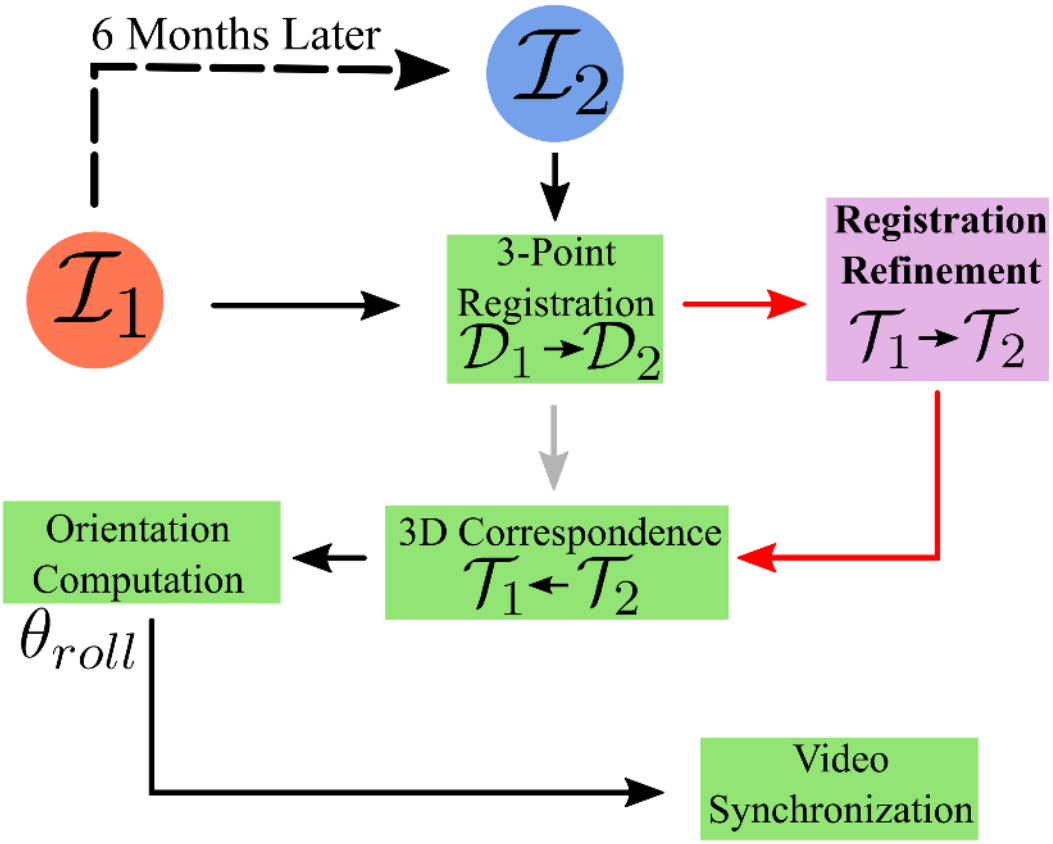


Figure 5.11: This figure summarizes the registration refinement step presented in this chapter as a flowchart.

CHAPTER 6

Endoscopic Image Analysis

Contents

This chapter is based on work under review in a journal.

Chapter 3 presented a first approach to inter-operative video synchronization using only an EMTS with EM markers, for GI endoscopy. The oesophagus was chosen as the first target clinical application since, relocalization is a clinically relevant here. This chapter also described how the 3D position obtained from the EM sensor could be used to determine the nearest neighbour (EMNN) and obtain video synchronization. Between, the pre-recorded diagnostic endoscopy (DE) and the live Surveillance endoscopy (SE), this selected image provides a localized view in the oesophagus during DE, corresponding to the current location of the endoscope during the SE. However, complete reliance on an only EM based video synchronization has some drawbacks, which can be classified as follows;

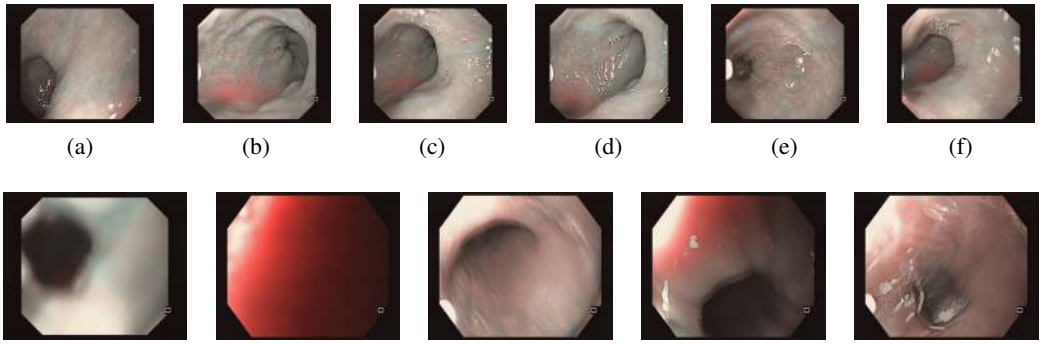
1. *Depth error*: In Chapter 4 it was established that the influence of uncertainty in the placement external markers effectively resulted in a final depth matching error along the oesophagus, of $\pm 10mm$ for a 95% confidence interval. The alternate registration presented in Chapter 5 showed that this error can be further reduced. However, there could still be a depth error that may be corrected, using the endoscope image information.

2. *Lateral error*: Within the localized region using EMTS the synchronized image obtained from EMNN does not necessarily provide the best view-point from GI expert's perspective as shown in the Figures 6.1(b)-6.1(f). This could be attributed to two factors; (a) orientation of endoscope in the lumen and (b) position of endoscope in the lumen cross-section.

3. *Uninformative (UI) frames*: Additionally, the EMNN could be an UI frame, as shown in Figure 6.1(g)-6.1(k).

It is thus important to provide an intelligent selection of the best matching view using the images from the neighbourhood of EMNN. To accomplish this, GI endoscope scene analysis and matching must be performed.

Video-endoscopic image analysis is an important approach for extracting relevant information from the surgical view. From a technical standpoint, this problem description is similar to scene analysis problems, which is an important and challenging task in various application domains, including computer vision, biomedical imaging, biometry, vehicle and robotic navigation, industrial visual inspection, and remote sensing [Lazebnik 2006, Zhou 2003, Datta 2008, Filliat 2007, Grasa 2014]. It provides a stepping stone towards reasoning with the world around us. This chapter thus, provides a review of literature, towards the application of scene understanding, classification and matching in GI endoscopy. The techniques presented in this chapter provide the necessary background to motivate the choice of algorithms, that are eventually employed in Chapter 7.



(a) (b) (c) (d) (e) (f)

(g) Blurred (h) Contact (i) Motion-Blur (j) Contraction (k) Fluid

Figure 6.1: (a): Query, (b)-(f): Five best EMNN frames matches with scores: $[0, 1, 0, 2, 1]$ as assigned by the expert section. 7.3. (g)-(k) present the sample UI frames.

Typically, analysing a scene involves extraction of low-level salient information. There are two parts to describing a scene; first, is recognizing a feature and the second is its description. There is no exact definition for a feature and it depends on the type of application. Generally a feature can be thought of as an "interesting" part of an image. This could be in the form of edges, corners and interest points or regions-of-interest such as blobs; which will be described in Section 6.1. Feature description, on the other hand, encompasses encoding the information from a feature, so as to have a representation that is distinctive and yet, repetitive enough to be applied to similar scenes. A descriptor could be in the form of, for example, image pixels, colour, texture, gradient or any information extracted by applying various mathematical operations to the selected image region. This will be detailed in Section 6.2. Finally, using the scene description, machine learning models are utilized to perform high-level abstraction and decision making; which are discussed in Section 6.3.

Each section in this chapter is concluded by a brief discussion on their application in GI endoscopic images.

6.1 Feature Detectors

Corners are among the first low-level features used for image analysis and in particular, tracking [Gauglitz 2011, Moravec 1980]. Moravec's work was further developed by Harris and Stephens [Harris 1988] which became known as the Harris Corner Detector. This formed the basis for the detectors proposed by Förstner [Förstner 1994] and Shi and Tomasi [Shi 1994]. The framework was extended by Mikolajczyk and Schmid in [Mikolajczyk 2001] to incorporate scale invariance to Harris detector. [Schmid 2000, Gauglitz 2011] present evaluations of many feature detectors, specifically based on intensity and contour-based approaches.

Blobs provide a complementary description of image structures in terms of regions. Blob detectors may contain an interest point or corners which could be local extrema of responses of certain filters or a centre of gravity. Thus, blob detectors may also be regarded as interest point operators detecting areas in an image which are too smooth to be detected by a corner detector. In particular, approaches aim to approximate the response to filters such as the Laplacian of a Gaussian which, given an appropriate normalization, are scale invariant when applied at multiple image scales as shown by Lindeberg [Lindeberg 2013]. Lowe extended this framework in [Lowe 1999, Lowe 2004] by approximating the local extrema detection using image filtered with differences of Gaussian, which are separable and hence faster to compute, than the Laplacian.

Endoscopic Applications: [Castaneda 2009] proposed to use SIFT features using normalized SSD based monoSLAM for 3D reconstruction of the oesophagus. A combination of, histograms of SIFT Flow Directions to describe the flow course; SIFT descriptors to represent image intestine structure and; SIFT flow magnitude to quantify intestinal deformation, was proposed in [Drozdzal 2010]. [Fan 2011, Fan 2010] proposed to use, affine invariant version of SIFT detector and descriptor to estimate the epipolar geometry and recover the 3D. [Spyrou 2015] provides a comparative assessment of various feature extraction methods for determining the position and orientation of a wireless capsule endoscopy. [Ciuti 2012] employed Shi-Tomasi features and used them in shape from shading framework for reconstruction.

Furthermore, detectors invariant to affine changes in the image have been proposed [Mikolajczyk 2002, Schaffalitzky 2002]; and exhibit better repeatability for large distortions [Lowe 2004, Mikolajczyk 2002]. However, they are typically expensive to compute as evaluated in [Mikolajczyk 2005, Moreels 2007]. The methods presented in Section 2.4.1 for intra-operative relocalization have employed feature detectors, such as, SIFT or the anisotropic region detector [Giannarou 2009]. It will be discussed in Chapter 7, that these methods are applicable for inter-frame mapping on synchronized frames, to facilitate, matching the regions of interest (ROI), such as those containing biopsy sites. Image preprocessing has also been employed for segmenting the necessary ROI based on texture similarity or using filtering techniques as discussed in [Arnold 2011, Bejakovic 2009,

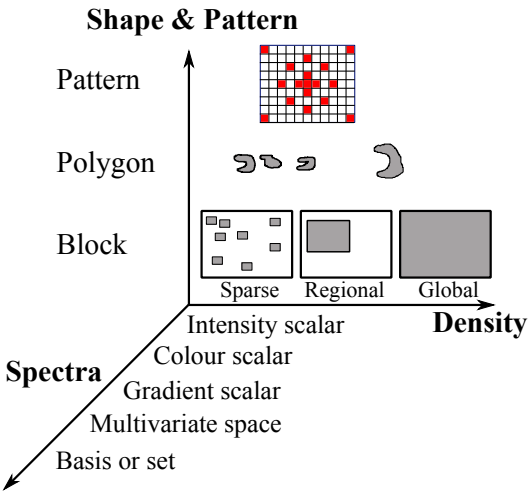


Figure 6.2: Taxonomy for feature descriptor dimensions; (1) feature density as global, regional, and sparse local; (2) shape and pattern of pixels used to compute the descriptor, these may include rectangles, circles, and sparse sampling patterns; (3) spectra, which includes the spectrum of information contained in the feature itself. This figure is reproduced from [Krig 2014] (*page* 192).

Bernal 2014b]. In [Arnold 2011], specular highlights is addressed using a segmentation method based on non-linear filtering and colour image thresholding followed by a fast in painting method. [Bejakovic 2009] provides the framework for lesion segmentation with application to Crohn's disease. [Bernal 2014b] used specular highlights and blood vessels segmentation for polyp detection and classification.

Feature detectors, although form an important pre-processing step, it is not always possible to find suitable points or regions depending on the target image. Especially in endoscopic scenes, where uniform texture can be found, salient points or regions are not repeatable inter-operatively. Thus the use of feature detection is such clinical application is limited.

6.2 Feature Descriptors

The step following feature detection is its description. An ideal descriptor is required, to not only be distinctive, but also robust to geometric and photometric transformations. Such invariance properties of the image representation are especially significant for endoscopic interventions, due to the presence of curved structures which lead to specularities, fluids or blood in the tract and due to the deformable nature of the tissue.

The simplest form of a descriptor is a vector of pixel values. Several literature reviews have been published, providing classification of various feature descriptors based on the target application. However, the review and classification presented in [Krig 2014], provided the most apt approach for the discussion in this section to relate their application to endoscopic image data. Krig classifies the feature descriptor taxonomy using three important characteristics, as shown in Figure 6.2; (1) *Shape and pattern*: How pixels are taken from the image. (2) *Density*: The extent of image used to form the descriptor, by differentiating them into local, regional and global descriptors. (3) *Spectra*: The scalar and vector quantities used for the metrics, and a breakdown of the algorithms and computations used to formulate them. Based on these criteria, descriptors can be classified into the following four families.

6.2.1 Shape Descriptors

Some of the earliest descriptors presented in the literature have been oriented towards abstracting the shape of objects in the scene as measured by their statistical metrics, such as area, perimeter, centroid etc. Typically, shapes are extracted using morphological operations on binary or grey scale images. Shape moments and metrics [Zhang 2004, Yang 2008], image moments [Jan 2009] have been used for describing shapes. Polygon shape methods are commonly used in medical and industrial applications, such as automated pathology for cellular biology, and also for industrial inspection of parts.

Endoscopic applications: Shape features are useful for describing lesions, polyps and dysplasia observed by the localized tissue texture changes. [Angermann 2015] proposed using edges, followed by a Hough transform of the image before using GLCM for texture features detection. [Eskandari 2012] used region-based Active Contour Method (ACM) and geometric feature for automatic detection of polyps. In [Häfner 2015] a new texture analysis framework using local fractal dimensions is employed by extending it to include shape and gradient information. [Hämmerle-Uhl 2012] studied in the context of celiac disease classification that, using distortion correction does not lead to significantly different classification performance when using edge- and shape-related features. [Wang 2010] employed geometric shape, saturation and intensity changes along the norm direction (cross-section) of an edge to provide localization during wireless capsule endoscopy.

6.2.2 Spectra Descriptors

Spectra, simply put is a quantity that can be measured or computed, for example, light intensity, colour, surface normals, local area gradients or statistical features, moments and sorted information such as 2D or 3D histograms of any spectral type, such as histograms of local gradient direction. There is no practical limit to the spectra that could be used with this particular class of descriptors. Some of the most popular descriptors in this category were evaluated by [Mikolajczyk 2005]. Perhaps the most famous of these is the Scale-Invariant Feature Transform (SIFT) developed by Lowe in [Lowe 1999, Lowe 2004]. The method achieves descriptor invariance to changes in scale and rotation, by operating in a local reference frame relative to a dominant scale, and rotation that is computed from the image. After the detection of a stable feature point, the descriptor is computed on a square grid sampled around the key-point. Several variants of SIFT have been proposed in literature such PCA-SIFT [Ke 2004], opponent-SIFT, C-SIFT and rg-SIFT [Brown 2011].

Endoscopic applications: Table 6.1 provides a (non-exhaustive) list of spectra descriptors encountered in GI endoscopy for varied applications [Georgieva 2015, Drozdzal 2014, Bell 2014]. In [Bernal 2012], the authors proposed a novel region descriptor built on the concept of depth of valleys (DoV) image, which combines valley localization given by a valleys detector with the intensity information provided by morphological gradient. A Valley Orientation-Depth of Valleys Accumulation descriptor (VO-DOVA) is presented, which consists of accumulating, by using a series of radial sectors, the maxima of the DoV image. This descriptor were incorporated into a polyp detection scheme in order to classify segmented regions into polyp or non-polypoid. In [Manivannan 2014], a variation called

Root-SIFT and a multi-resolution local patterns descriptors were extracted from image patches, for each colour channel for colon image classification.

6.2.3 Local Binary Descriptors

The family of descriptors in this category are features represented as binary bit vectors. To compute the descriptors, image pixel point-pairs are compared in a pre-defined pattern and the results are stored as binary values in a vector. Binary descriptors are very efficient to compute, store, and match. Generally, local binary pattern methods have been shown to achieve very good accuracy and robustness compared to other methods. Based on the sampling patterns employed, different binary descriptors have been proposed. However, since local binary patterns (LBP) form the basis for many variants in this category, a short description is given here. The LBP in its original form was presented in [Ojala 1994] and [Ojala 1996]. LBP has been found to be a powerful feature for texture classification. Below, the variant presented in [Ojala 2002] is described, which involves the following steps;

Step-1 The region of interest is divided into cells as shown in Figure 6.3.

Step-2 Each pixel in a cell is compared with its 8 neighbours (ie. its left-top, left-middle, left-bottom, right-top, etc.). Following a clockwise or counter-clockwise direction.

Step-3 A value "1" is output where the center pixel's value is greater than its neighbour's value, else, a "0". Finally giving an 8-digit binary number, which is converted to decimal.

Step-4 Over each cell a histogram is obtained with the frequency of occurrence of numbers computed in the previous step.

Step-5 The histogram is normalized. The histograms from each of the cells is concatenated, giving a feature vector for the region of interest. This feature vector can now be processed by a machine learning algorithm, such as the support vector machines (SVM) for classification [Shan 2009].

Endoscopic applications: Several references have employed LBP and its variants for disease detection and classification in GI tract, [Gadermayr 2015, Alexandre 2008, Amelling 2009, Engelhardt 2010, Charisis 2013, Constantinescu 2015, Meng 2010].

6.2.4 Basis Space Descriptors

The fourth and final category of feature descriptors involves detection of alternative basis space to describe image features. A set of vectors in a vector space V is called a basis, or a set of basis vectors, if the vectors are *linearly independent* and every vector in the vector space is a linear combination of this set [Halmos 1958]. Generally, a basis is a linearly independent spanning set, and it is useful to transform a dataset from one basis space to another to gain further insight, or to just process the data. Four important basis space methodologies have been encountered in the reviewed literature on GI endoscopic image analysis.

Descriptor	Comments
SIFT	Scale Invariant feature transform (details in text), has been employed for varied applications such as 3D reconstruction, biopsy point description, motility descrition etc. [Castaneda 2009, Drozdzal 2010, Manivannan 2014, Spyrou 2015].
SURF	Speeded up Robust Features, presented by Bay et al. [Bay 2008] uses the similar concept as SIFT however, for finding the dominant feature orientation, they use HAAR-like feature responses around the detected interest point. [Spyrou 2012, Iakovidis 2013] have employed it for capsule endoscope tracking.
MPEG-7	The MPEG-7 is a visual Standard under development specifies content-based descriptors that allow users to measure similarity in images or video based on visual criteria. Specifically, MPEG-7 describes colour, texture, object shape, global motion, or object motion features. It has been used in [Bulat 2007, Coimbra 2006, Duda 2007] for various applications in GI endoscopy.
HOG	Histogram of oriented gradients [Dalal 2005] relies on computing local region gradients over a dense grid of overlapping blocks, rather than at interest points. [Iwahori 2015, Wang 2014c] have used HOG for polyp detection.
PHOG	Pyramid HOG [Bosch 2007] was designed for global or regional image classification. PHOG combines regional HOG features with whole image area features using spatial relationships between features spread across the entire image in an octave grid region subdivision. This method has been employed in [Bae 2015] for polyp detection.
DSIFT	Dense SIFT as proposed in [Lazebnik 2006], uses an approximation over gradients before computing the histogram to substantially improve the speed with little or no loss of performance in applications. [Miyaki 2013] applied it for identification of mucosal gastric cancer under magnifying endoscopy.
GLCM	Gray level co-occurrence matrix is a histogram of co-occurring grayscale values at a given offset over an image. This is used to extract various texture features such as entropy, contrast, correlation, energy, homogeneity etc. as presented in [Haralick 1973]

Table 6.1: Table presents the most commonly used spectra descriptors presented in literature.

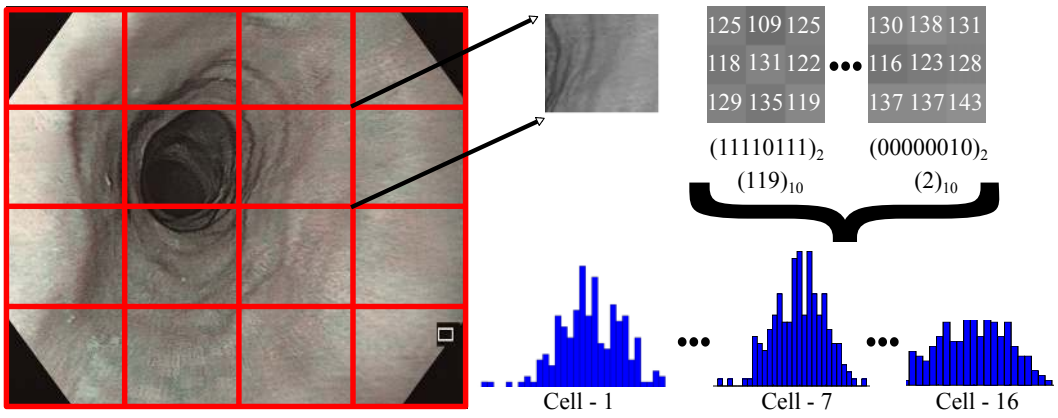


Figure 6.3: Steps involved in the computation of LBP feature vector.

1. ***Spatial Frequency Based Descriptors***: Basis transforms, such as the Fast Fourier Transform (FFT), Discrete Wavelet Transforms (DWT), Dual-tree Complex Wavelet Transform (DT-CWT), Gabor Filters, Contourlet Transform, Contourlet Transforms etc.; decompose an image into a set of basis vectors from which the image can be synthesized or reconstructed. Viewing the set of basis vectors as a spectrum is a valuable method for understanding image texture and for creating a signature. Two broad approaches for texture retrieval in the spatial-frequency domain are generally discussed in literature [Baaziz 2010]; energy-based approach and statistical modelling-based approach.

   Endoscopic Applications: In GI endoscopy the former has been widely used for diseased tissue classification and on-line decision support. Variants of DT-CWT are explored for automatic classification of endoscopic images using the Marsh classification, in [Häfner 2010, Uhl 2011]. The feature vector was composed of mean and standard deviations of the sub-bands from DT-CWT variant or Weibull parameter of the sub-bands. Enhanced scale invariance was obtained by applying DFT or DCT across the scale dimension of the feature vector. [Liu 2012] used, color wavelet covariance as features for ulcer and blood detection. For polyp detection, [Iakovidis 2005] employed colour wavelet covariance. [Kalpathy-Cramer 2009], proposes using GLCM, colour histogram, GIST and Gabor, wavelet, Maximum response (MR8), Leung-Malik (LM) filter bank, and the Schmid filter banks responses as feature descriptors for image retrieval using a naive Bayes Nearest neighbour classifier. In [Karkanis 2003], colour wavelet covariance based texture features are extracted for abnormality detection in endoscopic images.

2. ***Feature Learning***: This class includes a set of techniques that learn a feature: as a transformation of raw image data or other features described earlier in this section. Such an approach obviates the need for traditional hand-crafted features and allowing a machine to learn them. Feature learning can be divided into two categories: (a) *Supervised Feature Learning*: Attempts to learn features from labelled data. Supervised dictionary learning and neural networks fall in this category. However, most applica-

tions encountered in this survey belonged to the second category; (b) *Unsupervised Feature Learning*: These approaches attempt to learn features from unlabeled data. The goal is often to discover low-dimensional features that captures some structure underlying the high-dimensional input data. When the feature learning is performed in an unsupervised way, it enables a form of semisupervised learning where first, features are learned from an unlabeled dataset, which are then employed to improve performance in a supervised setting with labeled data. Approaches that are categorized in this are , visual vocabularies, bag of words (or features), textons etc.

Manifold learning approaches, can also be categorized here. Methods such as, Principal component analysis (PCA), ISOMAP, Locally linear embeddings (LLE), Laplacian eigenmaps (LE) etc., statistically estimate the topological space that is locally Euclidean for the training data. In contrast to linear vector spaces, on a manifold the Euclidean metric properties do not hold globally. Therefore, when measuring the distance between data points lying on a manifold, the global structure of the underlying manifold needs to be taken into account [Atasoy 2012a]. Non-linear approaches attempt to reduce the dimensionality, while preserving the local structure of the data.

Endoscopic Applications: [Tamaki 2013], explores local features (extracted by using sampling schemes such as Difference-of-Gaussian and grid sampling), BoW, and provides extensive experiments on a variety of technical aspects for feature description to use with content based image retrieval systems. [Yuan 2015] presents an improved bag of feature method to assist classification of polyps in WCE images. Several different descriptors are combined into the bag-of-words framework. Various parameters are studied in the context of classification of polyps. [Francisco 2015] proposed using image patches in the BoW model for classification. In [Manivannan 2013], authors present two schemes. The first, working on the full-resolution image, the second on a multi-scale pyramid space. With this framework any feature descriptor could be employed; but a multi-resolution LBP was tested. [Gallo 2010, Nawarathna 2013, Riaz 2011] are examples of methods using textons as feature descriptors. Manifold based approaches on the other hand have been extensively used for feature compression and identifying the most discriminating features for classification, such as in [Tjoa 2003, Berens 2005, Biswas 2014, Duda 2007, Atasoy 2012b]. They provide an elegant representation of the extracted features to extract the underlying structure as in [Atasoy 2012b].

Last, but largely unexplored approach in GI endoscopy literature is using auto-encoding to learn an over-complete representation of input data. An autoencoder is a simple 3-layer neural network. Unsupervised learning using a single layer algorithm could extract salient features, but because of the limited capacity of that layer, the features can be seen as low-level features. It is conceivable that by stacking two autoencoders, such that the second layer takes output of the first layer as input, could extract slightly higher-level features. In this way, one could imagine that high-level abstractions characterizing the input could emerge with each new stacked layer, which could be more discriminating that the original feature-set. Autoencoders, have

sometimes been considered as simplified models of Deep belief networks. Such methods have been employed in pre-trained models, such as CaffeNet [Jia 2014], OverFeat [Sermanet 2013], DenseNet [Iandola 2014], Alexnet [Krizhevsky 2012], Clarifai [Zeiler 2014] etc. for natural image scene classification. Deep learning strategies should be ideally suited for obtaining application specific features in GI endoscopy as they do away the need to build hand-crafted features.

6.3 Matching and Classification

The next step in the process is matching and/or classification using the feature representations described in the previous section. The first and the most direct approach is by assigning a meaning to an observed feature set, by comparing a pair of observations by a similarity function. While searching for a correspondence for query image, represented by feature **P**, with an image described by feature set **Q**, from among the elements of the database of images; knowledge of the domain is expressed by formulating a similarity measure $S_{Q,P}$. Various similarity (distance) measures have been proposed in literature [Cha 2007, Krig 2014, Yang 2006]; and their usage depends on the type of features. Alternatively, for classification, the feature descriptors can be used to learn a structure from the data, to determine to which category an incoming feature belongs.

1. ***Nearest Neighbour (NN) Classifier***: Classification of unlabelled images is performed based on the closest training samples in the feature space. Considering the descriptor-label pairs, $(\mathbf{X}, \mathbf{Y}) = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$, such that, $X \in \mathbb{R}^d$ and $Y \in \{1, 2\}$. The nearest neighbour over the distance metric defined by $\|\cdot\|$ on $\mathbb{R}^d$ is used to defined the 1-NN or the k-NN classifiers. The nearest neighbour is the simplest classifier and has been largely used in methods such as in [Kwitt 2008, Vecsei 2008, Hu 2015, Mathew 2015, Rajan 2009].

2. ***Clustering***: It is an unsupervised classification of patterns into clusters [Jain 1999]. Although a "cluster" cannot be precisely defined, a common denominator among different models is the "grouping" of data objects. Clustering algorithms use knowledge about the data, either implicitly or explicitly. Implicit knowledge is used in (a) selecting a pattern representation scheme (e.g., from prior user experience to select and encode features); (b) selection of the similarity measures; (c) choosing the grouping scheme (i.e. specifying the k-means algorithm when the clusters are known to be hyperspherical). Explicit knowledge, constraints or guides the clustering process. [Abouelenien 2013, An 2005, Asari 2000, Chen 2015, Drozdzal 2015, Francisco 2015] are some examples in GI endoscopy.

3. ***Kernel Methods***: Kernel methods derive their name from their usage of kernel (or similarity) functions. These functions operate in a high-dimensional, implicit feature space, without ever performing the actual mapping of the data in that space, but rather by simply computing the inner products between the images of all pairs of data in the feature space. This is often computationally cheaper to perform than

the explicit computation of the coordinates, and is referred to as the "kernel trick". The most popular example of kernel machines is the support vector machine (SVM) [Boser 1992]. The SVM is a learning machine for binary classification problems and are the most commonly used classifiers in GI literature; [Abe 2015, Abouelenien 2013, Zhou 2014, Rajan 2009, Mackiewicz 2006, Drozdzal 2015].

4. ***Decision Tree Classifiers***: Decision trees are another example of supervised learning approaches. Decision tree learning uses a data structure called as the decision tree for building a predictive model that maps observations about an item to conclusions about the item's target value. Several variations of decision trees have been employed in GI endoscopy literature, discussed in [Breiman 2001, Rokach 2005, Khan 1996, Hai 2009, de Sousa 2008, Serpa-andrade 2014, Gallo 2012, Iwahori 2015, Li 2014, Wang 2014b].

5. ***Artificial Neural Networks (ANN)***: They are an extension of probabilistic classifiers, that are able to predict, given a sample input, a probability distribution over a set of classes. ANN is a learning paradigm, inspired by the way biological nervous systems process information. ANN are the second most commonly used approaches in GI literature after SVMs; [Barbosa 2012, Boulougoura 2004, Bourbakis 2005, Bulat 2007, Fu 2011, Karargyris 2009, Khun 2009, Li 2009].

6.4 Discussion and Conclusion

A wide range of methods have been employed in GI endoscopic literature for scene understanding and classification. In this thesis, we are specifically interested in understanding the applications to oesophageal tissue classification. Although any of the methods described so far in this chapter can be employed, however, the most important criteria in inter-operative scene matching is invariant global image description under varying contrast and scene lighting along with real-time speed of computation. In addition, the information extracted must also include some geometric data to capture any specific shape features. Apart from being globally invariant, the description must also include local information to allow partial scene matching.

From the review, we have selected the following the evaluation described in the next chapter.

- The review in the previous sections indicated that LBP and its variants are the best performing descriptors for tissue texture description.
- We are interested in understanding if using the lumen shape as an additional feature would aid endoscopic image matching hence an LBP descriptor combined with the HOG is also considered for the comparison.
- Local ternary pattern descriptor (LTP), which is a variant of LBP, has been shown to perform well under difficult lighting conditions. In endoscopic scenes where illumination variation is a significant challenge, LTP was chosen as an alternate approach.

- SIFT is one of the best performing descriptors for natural scene classification. An alternate approximation of the original SIFT, proposed by [Lazebnik 2006] where descriptors are densely extracted using a flat rather than a Gaussian window on a regular grid at certain step size has shown to perform well. It is relatively faster than the original SIFT computation and thus considered to be an ideal choice in the comparison.

- Inspired by the dense SIFT approach, a dense version of LBP is proposed in which the cells are obtained as overlapping blocks, in contrast to the original approach by [Ojala 2002].

- A Local Invariant Order Pattern (LIOP) descriptor proposed in [Wang 2011] is considered that is a local image descriptor based on the concept of local order pattern. Order pattern is the order obtained by sorting selected image samples by increasing intensity. Considering in particular a pixel $\mathbf{x}$ and n neighbours $x_1, x_2, \ldots, x_n$. The local order pattern at x is the permutation σ that sorts the neighbours by increasing intensity $I(x_{\sigma(1)}) \leq I(x_{\sigma(2)}) \leq \ldots \leq I(x_{\sigma(n)})$. Such a representation is invariant to monotonic changes of the image intensity. However, this only describes a small portion of an image patch and is not sufficiently distinctive. LIOP, on the other hand collects local order patterns computed over the complete image to formulate a representation, which is distinctive. Once the order patterns are obtained, they can be pooled together into a histogram to form a descriptor.

Two aspects in the descriptor computation must be highlighted. First, using different image scales to formulate the descriptors. This is useful since, at varying scales of the image, the local information contained in the image decreases and only the global texture will be encoded. Second, the choice of appropriate colour-space. It should be noted that, in alternate imaging modalities, such as for SPIES, NBI or FICE (Appendix A), converting images to Gray-Scale (GS) is not an ideal choice, since it does not provide a true mapping of the wavelengths. In such cases, it is important to study the appropriate colour-space that is stable for extraction of features. The choice of suitable colour-spaces is also important to achieve invariance to illumination, contrast changes and specular highlights, which are strongly observed in GI endoscopic images. Thus, the descriptors presented above, have been computed in RGB, HSV, GS, normalized RGB (norm), log and opponent colour-spaces (chosen appropriately for each descriptor).

Chapter 7

Image classification and Fine Positioning

Contents

This chapter is based on work published in [Vemuri 2015a].

Thus far an inter-operative (IRO) video synchronization method using only the information obtained from an EMTS was presented. In this framework, an EM sensor was used inside the endoscope channel to track its position in the oesophagus, and two external sensors provide the anatomical landmarks on the patient for IRO registration and video synchronization for the (live) surveillance endoscopy (SE), with the (pre-recorded) diagnostic endoscopy (DE). It was explained in Chapter 6 that, complete reliance on the EM based match has certain drawbacks. Since we have a localized region within the oesophagus; the image information can be used more reliably to perform IRO scene matching.

Before discussing the specifics of this approach, we revisit the task of IRO endoscopic relocalization. On a broad level this can be subdivided into three stages:

1. *Gross-localization* (GL): In the discussion thus far, the EM based approach can be considered as GL. Computing the nearest neighbour from the 3D position obtained from EMTS, an approximate location of the endoscope in the oesophagus is obtained. It is termed as GL, since it could be error prone due the reasons explained earlier in Chapter 6. As shown in Figure 7.1, this is a biased match and a method to remove this bias must be provided.

2. *Fine-positioning* (FP): Considering the neighbourhood of the gross-localized region, an intelligent selection of the best matched position can be made. This could be based on using the localized image search (Figure 7.2(a)) or using the EM sensor information (Figure 7.2(b)) to determine the ideal direction of view in the oesophagus. This is a refinement step and this chapter proposes a solution using the image information available, that has been left unused until this step in the framework.

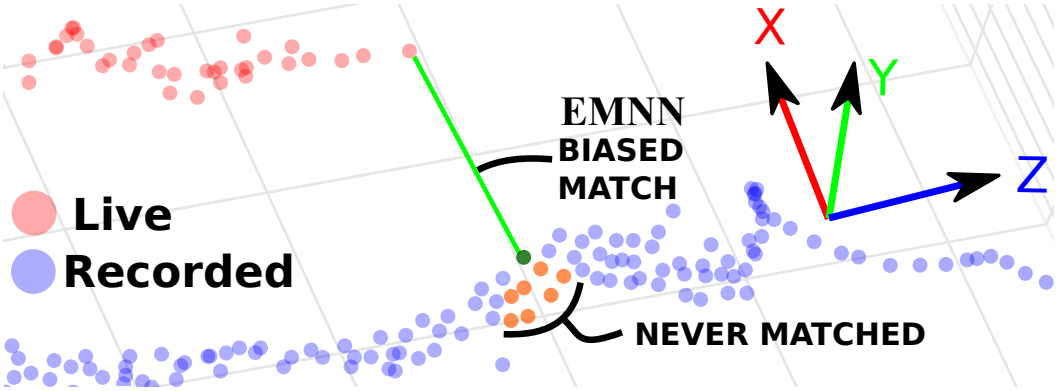


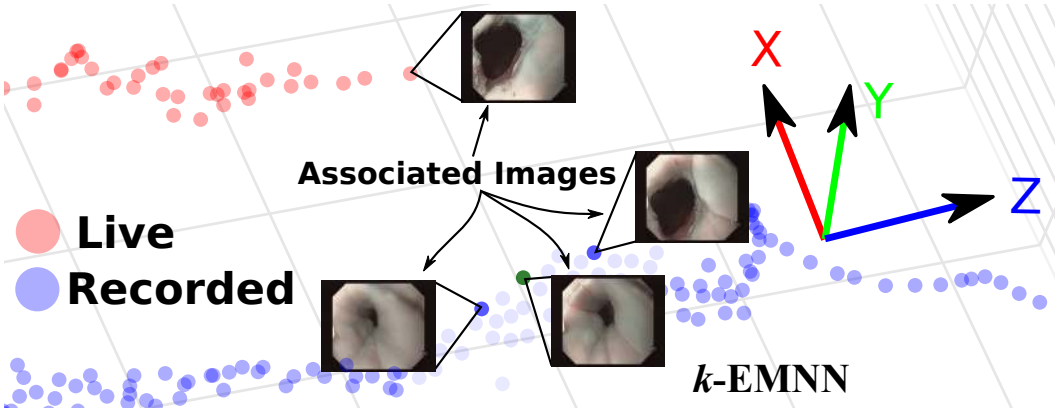
Figure 7.1: The EMTS based match (ENNN) is a biased match since it does not consider the vicinity of the EMNN which is at the same depth in the oesophagus, however depending on the shape of the trajectory and the registration, the points shown around the EMNN (in orange colour) do not get considered during gross-localization.

3. *Inter-frame mapping* (IFM): Finally, from the refined image match obtained from fine-positioning, regions of interest (and tagged biopsy sites) can be mapped from the matched image to the live frame for re-targeting. This thesis does not delve deeper on this subject, however, it forms a part of prospective research direction that will be discussed in Chapter 8.

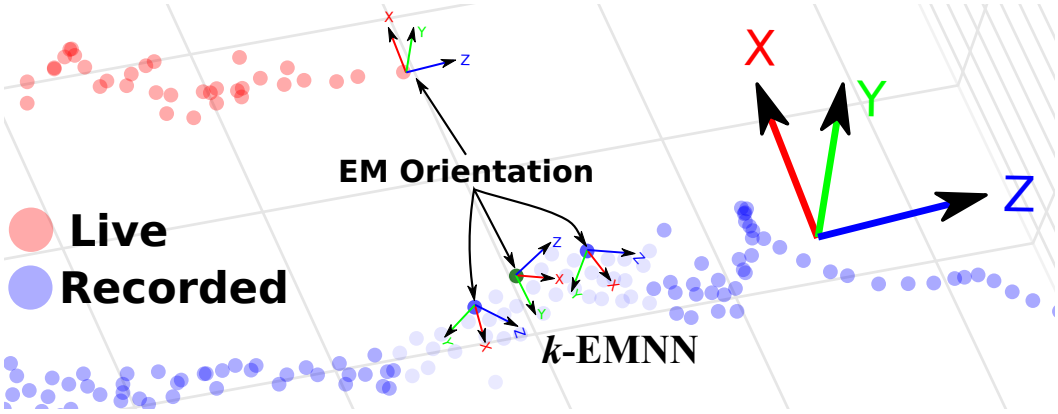
This classification is essential in the context of IRO relocalization, since, it allows a deeper understanding of the problem, to make the choice for appropriate computer vision and machine learning concepts discussed in the previous chapter. The current state-of-the-art methods in providing IRO relocalization attempt to solve the GL and FP problems jointly. Such as, in [Atasoy 2012b] had proposed training locality preserving projections, for low-dimensional embedding of images, from a single intervention for localization in subsequent procedures. [Wang 2014a] proposed to learn a graph (atlas) from the sequence of images from several gastroscopic interventions. However, among the challenges discussed in Section 2.5, the evolution of tissue structure between procedures, poses the most significant difficulty in IRO GI endoscope image analysis. Thus, it is clear that at each of the stages presented above, specific issues need to be tackled.

This chapter, focuses on the FP problem which is closely related to scene understanding, classification and recognition that has received considerable attention in computer vision [Van De Sande 2010] and GI endoscopy as presented in the previous chapter. In the scope of this work, however, the method proposed here uses the outcome of Chapter 3 to provide further refinement in video synchronization.

In this chapter the following points are explored. Firstly, we establish the need for GL using EMTS. As was identified in the previous chapter that presence of uninformative (UI) frames can also be disruptive in the video synchronization. Hence, an approach to filtering the UI frames is proposed. Finally, an evaluation of 6 descriptors in various colour-spaces, for narrow band imaging (NBI) and white light (WL) endoscopic modalities (which are commonly used in clinical practice) is performed, to prove the validity of the approach



(a)



(b)

Figure 7.2: (a) The images recorded during diagnostic endoscopy, around the localized EMNN point. (b) The EM sensor orientations around the localized EMNN point.

and demonstrate on multiple patient data that a relevant choice of parameters allows to considerably improvement in the inter-operative scene matching.

Section 7.1 presents an outline of the proposed methodology. Then, extending the discussion presented in Section 6.4; it adds further insight into the clinical conditions encountered and presents a brief summary of the proposed descriptors employed for evaluation. Section 7.2 presents the approach for the removal of UI frames and studies the effect of various parameters in the classification performance. Section 7.3 describes the process of data collection and assignment of ground-truth for both scene recognition and UI frame classification with their results. Section. 7.4 lists the important observations in this study.

7.1 Constrained Scene Matching

For the refinement of match using the nearest neighbour of the 3D EM sensor position (EMNN), it is proposed here to consider the k-EMNN matches within a chosen search radius as shown in Figure 7.3. This localized region should contain frames that are visually closest

to the endoscopic live view, thus constraining the search space for localization. Within this constrained space, the methods discussed in Chapter 6 can be employed. Here, the task of FP can be identified as "*view-point selection*".

Since the recorded video can be analysed in advance, a database containing the various scene descriptors can be maintained, to compute the best matching viewpoint for the live view. It should be noted that the value of k would vary depending on the number of points within the chosen search radius from the closest EMNN match; since, it is related to the density of points in $\mathcal{T}$, in the vicinity of the EMNN match. Typically the biopsy sites or location of particular interest have higher density of points as was shown in Figure 5.1. It is important to note that, to obtain view-point selection, a corresponding matching frame containing a similar view-point to the live view, may not necessarily be available. However, it is important to be able to provide the "best available" matching view.

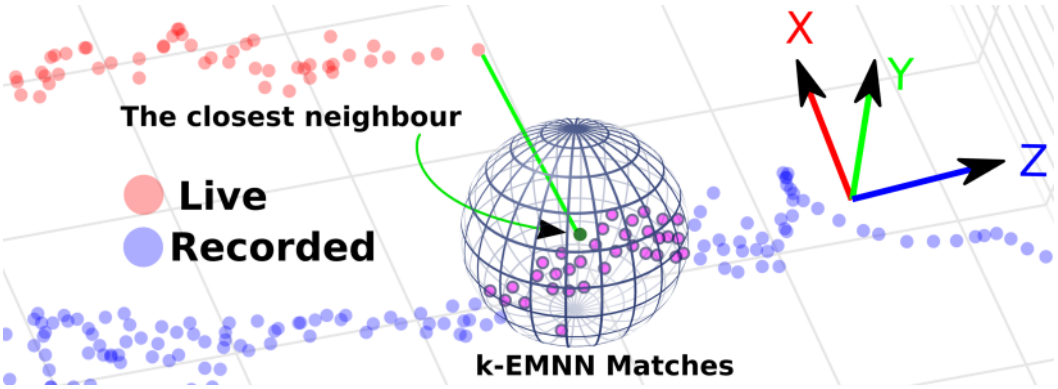


Figure 7.3: k-Nearest neighbour electromagnetic tracker matches. Firstly, the EMNN is obtained using the 3D position match. Then within a radius r around this match, all the points on the trajectory are considered as the k-EMNN matches.

Scene description for endoscopic frames involves several important considerations. The review presented in the previous chapter highlights the different approaches and salient features that have been used for endoscopic image classification in the literature. As with most scene recognition problems, we are interested in performing a global match between competing frames, the most important features in this context include; the shape of lumen, descriptors capturing information from the blood vessels, invariant (local or global) texture, colour and spectra description of the tissue and detection of abnormalities, such as lesions and polyps, to localize based on regions of interest.

When the disease spreads over a period of time or between treatments (such as radio-frequency ablation), reliance on image descriptors for capturing the tissue texture in various forms is not feasible. In such cases, the lumen shape is an important characteristic. During GI endoscopy, the insufflation is usually artificially monitored and regulated for patient safety. Thus, assuming that the insufflation is similar between two procedures, regions localized using the EMTS would have similar characteristics of the oesophagus muscle contractions. This observation, pointed out by the GI expert, was qualitatively observed in preliminary review on 4 patients.

Some of the most robust descriptors for performing scene matching, have been presented in Chapter 6. Following the trend observed in that review, Section 6.4 identified a set of 6 descriptors summarized in Table 7.1, that would be considered for a comparative evaluation for view-point localization. Since, the GI expert is presented with an approach to match the orientation (method described in Section 3.4) in the live view with a recording from the previous intervention; rotation invariance in descriptors is not a necessity.

Table 7.1, briefly presents the various descriptors that were discussed in Section 6.4. As was also analysed in Section 6.4, these descriptors have been computed in RGB, HSV, Gray-Scale (GS), normalized RGB (norm), log and opponent colour-spaces (chosen appropriately for each descriptor), to evaluate their performance. The Chi-squared distance was used as the similarity measure, since the features are histogram based and the vector of scalars are independent variables.

ID	*Descriptor*
mLBP[Mäenpää 2003]	Multi-scale Local Binary Patterns. At each level of the pyramid, the image is divided into non-overlapping cells. An LBP feature vector is computed for each cell which was concatenated into a large feature vector representing the image.
mHOG[Newell 2011]	Multi-scale Histogram of Oriented Gradients. Approach similar to mLBP but with an HOG descriptor for each cell.
sw-mLBP	Sliding window mLBP. Similar to mLBP, with each cell is a region within a sliding window over the image.
mLTP[Tan 2010]	Multi-scale Local Ternary Patterns.
mLBP+mHOG	A combined mLBP and mHOG descriptor.
dSIFT[Lazebnik 2006]	Dense scale invariant feature transform. A fast variant computed on non-overlapping cells for each image.
mLIOP[Wang 2011]	Multi-scale Local Intensity Order Pattern.

Table 7.1: Summary of feature descriptors evaluated in this study. For multi-scale approaches a scale space image pyramid was constructed.

7.2 Uninformative Frame Removal

Several approaches have been proposed for removal of UI frames [Bashar 2010, Alizadeh 2015, Bernal 2014a]. However, most have focused on images from either colonoscopic videos or those recorded using capsule endoscopy (CE). Typically, the UI frame observed in colonoscopic videos is quite different, owing to the presence of faecal matter, turbidity due to water jet spray and presence of bubbles. For CE, in addition to these, there is no insufflation during the procedure, thus, the contractions of colon muscles are more prominent, along with an unclean tract. Thus the methods presented for colonoscopy or CE cannot be directly applied here. In [Atasoy 2012b], the authors proposed using locality preserving projections using k-means clustering to identify UI segments with the recorded

intervention. Thus, they do not propose labelling every image in the intervention. This is however not the case in the current scenario.

In Chapter 6, LBP was identified as the most widely used and best performing descriptor for tissue texture classification. Hence, here an approach using mLBP has been proposed. Since, the alternate colour-spaces discussed earlier store redundant information about the scene, here, only the GS version of the image has been considered. Since, the goal is to capture the global texture variation. Furthermore, to further eliminate the descriptor information redundancy, PCA was applied. PCA is an ideal choice to statistically decorrelate and extract the most representative information from the descriptors; which can be performed in an unsupervised manner. To achieve this, a randomly selected balanced subset of the available data is used. From this subset, PCA basis vectors were computed and used them for obtaining low-dimensional representation (PCA-LBP) for the complete descriptor set. Figure 7.4 shows an example case of the first three components of the PCA-LBP descriptors, where the basis vectors were computed using 10% (stratified sub-sampling) of the data. Qualitatively, two clusters can be observed to emerge from these projections, leading us to pursue employing these as an input to a classifier. The trade off for the percentage of data needed to obtain the PCA basis vectors against the classifier performance, is studied later in this chapter.

An RBF-kernel SVM classifier was trained on the PCA-LBP descriptors. SVM was considered because, it can be trained on small data-sets, it is scalable on new data and an RBF kernel function is well established method in capturing the non-linearities in data efficiently. Their downside, however is that the training time is much longer as it is much more computationally intensive. In the current application scenario, since the training of the model (with the inclusion of new data) would be performed between the two procedures; real-time performance is not an important criteria. The *LibSVM* package developed by [Chang 2011] is the most well tested tool available. The training was performed in two steps. Firstly, a leave-one-out (LOO) cross validation was performed for the best parameter selection. In this context, for LOO cross-validation, data from k-1 interventions were used for training, and the k^{th} intervention for testing. There are two important parameters to be analysed here,

(1) The percentage of data to use for computing the principal components (ν).

(2) The number of principal components (κ) to consider.

Their influence *w.r.t* the performance of the classifier is analysed in the next section. Figures 7.5 and 7.6 illustrate the computed decision boundary for the sample case that was described earlier, considered here for $\kappa = 2$, for NBI and WL respectively.

7.3 Experiments and Results

Data from 7 human subjects was collected, with two surveillance procedures per subject. An Olympus gastroscope was employed, that is capable of both WL and NBI modalities, for recording the interventions. Seven diagnostic and surveillance interventions each, were recorded, first using WL followed by NBI imaging. From each recorded diagnostic

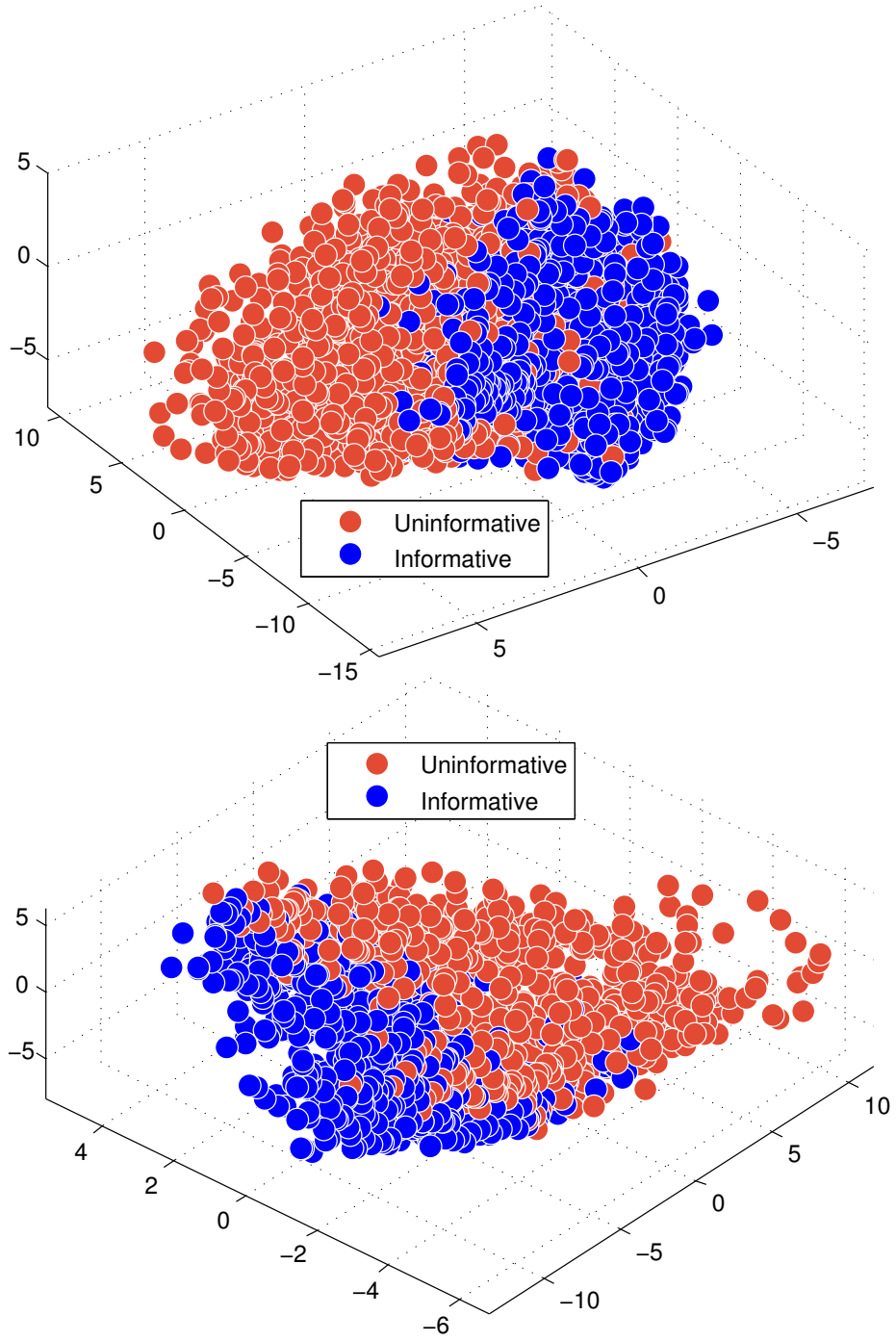


Figure 7.4: The first three dimensions of the projected feature vectors, NBI (top) and WL (bottom), on to the trained basis vectors from PCA.

intervention trajectory, 9 equally spaced query locations were selected to cover about 25 cm along the oesophagus length. For each selected query locations, k-EMNN matches were obtained from the surveillance interventions, with increasing search radii (10 mm - 70 mm), as shown in the Figure 7.3. Hence, for 7 pairs of diagnostic endoscopies, a total of 63 query locations (each for NBI and WL) were selected. The GI expert reviewed the k-EMNN matches obtained for each of the search radii and scored the matched images, 2 - best match, 1 - partial match and 0 - incorrect match. Although these are subjective scores, they help quantify the ideology and the approach to feature based matching in the choice of a good view-point. For each query frame and k-EMNN frames, descriptors were computed and matched using chi-squared distance metric. In this chapter, 36 descriptor-colour-space combinations were compared. Variations of algorithms presented in the *Vlfeat* ([Vedaldi 2008]) have been used.

For training the classifier of UI frames, the GI expert reviewed the images from 10 NBI and 8 WL surveillance interventions. A score of 2 for informative frame (IF), 1 for partially informative (PIF) and 0 for UI frame were assigned. It was observed that the PCA-LBP descriptor, was not sufficiently discriminative between IF and PIF, hence, the IF and PIF classes were grouped into a single class. A total of 4236 NBI and 2643 WL frames were tagged. The illustration below presents the various elements of the confusion matrix,

		True Condition	
		UI	***I***
Predicted Condition	***UI***	True Positive (TP)	False Positive (FP)
	I	False Negative (FN)	True Negative (TN)

The following metrics were then computed to analyse the performance of classifier;

$$F - Measure = \frac{2TP}{2TP + FP + FN} \quad Precision = \frac{TP}{TP + FP} \quad Recall = \frac{TP}{TP + FN}$$

$$Positive\ Likelihood\ Ratio\ (PLR) = \frac{TPR}{FPR} = \left(\frac{TP}{TP + FN}\right) / \left(\frac{FP}{FP + TN}\right)$$

$$Negative\ Likelihood\ Rate\ (NLR) = \frac{FNR}{TNR} = \left(\frac{FN}{TP + FN}\right) / \left(\frac{TN}{FP + TN}\right)$$

The PLR and NLR, are complementary indicators of the classifier performance combining sensitivity and specificity. Ideally, a large PLR value indicates a strong TP prediction rate. Whereas, NLR in an ideal situation, would be small ($0 < NLR \ll 1$). Figure 7.7(a), illustrates the variation of the classifier performance *w.r.t* change in ν and κ. It should be noted that the values plotted in these graphs are obtained from the LOO cross-validation described in the previous section. For each graph different values of $\nu = [0.05, 0.10, 0.35, 0.55]$ were used. Thus, the maximum number of estimated principal components (κ), is limited by the data collected from the experiments for the specific modality (4236 for NBI and 2643 for WL).

The following observations were made;

1. For NBI and WL, using $\kappa = 5$, PLR obtained was $[14, 15]$, however, NLR is also high $[0.6, 0.5]$. Which implies, that the FPR is small, but FNR is large indicating, low confidence in identifying an IF. This observation is further confirmed from the low F-Measure scores for $\kappa = 5$. Also the variance of the other metrics, is large. Clearly, low values for κ, will not be able to sufficiently capture the required informativeness of an image.

2. For $\kappa = 50, \cdots, 1000$, the F-Measure=$[0.89, 0.87]$, is relatively constant for NBI and WL respectively. This indicates that including more information in the form of κ, does not capture any additional pattern in the data.

3. Figure 7.8, illustrates the receiver operating curves (ROC) and the area under the ROC curve. It also indicates that the proposed approach shows high probability of identifying an UI frame randomly presented.

4. Larger values of ν also does not improve the information captured in the principal components (κ). Thus, using only a small subset of data, the required set of κ basis vectors can be trained.

For a sample case, with $\nu = 0.1$ for obtaining the PCA basis vectors and $\kappa = 100$ principal components were considered. Using these, the low-dimensional PCA-LBP descriptors was computed to report the, precision and recall obtained for the test data as; NBI = $[95\%, 89\%]$ and WL = $[96\%, 88\%]$. The average scores for the EM based match for NBI and WL improved from $[0.97, 0.82]$ to $[1.2, 1.2]$, after filtering the UI frames. Following the evaluation of RBF-SVM UI frame classifier, a comparison of the scene matching was performed using the various descriptor-colour space combinations presented earlier. For the sake of completeness, Tables 7.2 and 7.3 reports the results for all the combinations for NBI and WL respectively. It can be observed that the best matching score using the constrained image based match improved to $[1.7, 1.46]$ for NBI and WL. Figures 7.13 and 7.14 shows the best matches from EM and imaged based approach for NBI and WL endoscope modalities, respectively. For these sample cases, the Table 7.4 describes why the image-based match is better than the closest EMNN.

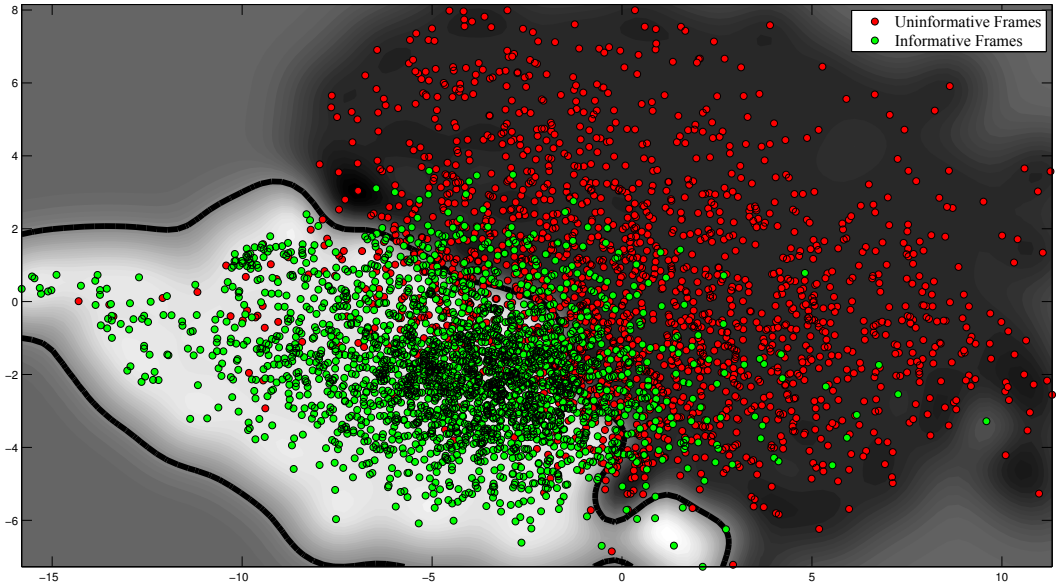


Figure 7.5: 2D Decision Boundary for NBI, using $\kappa = 2$ principal components.

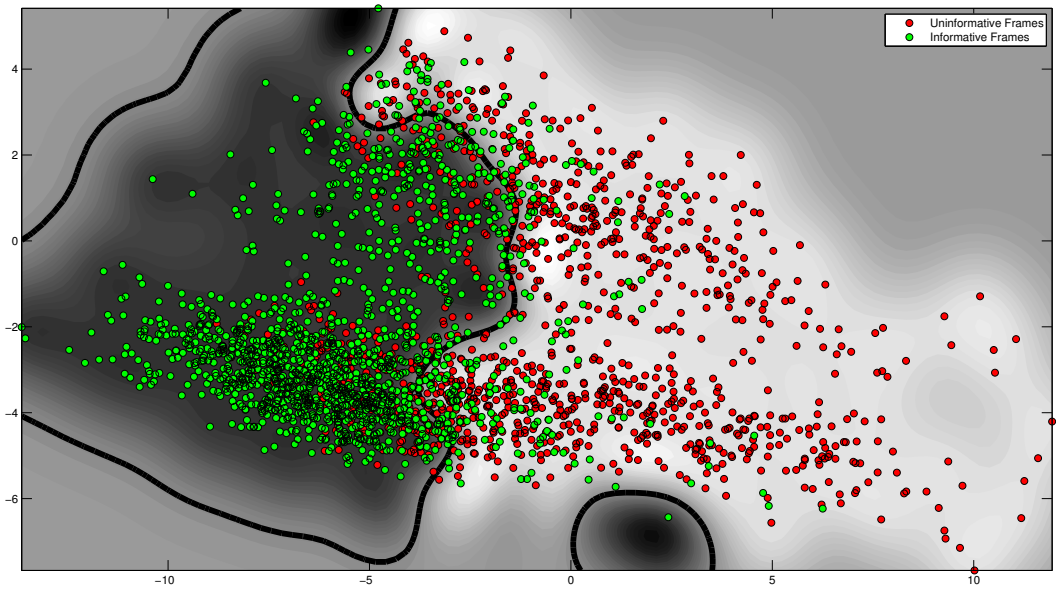


Figure 7.6: 2D Decision Boundary WL, using $\kappa = 2$ principal components.

7.4 Discussion and Conclusion

This chapter extends the EM based IRO synchronization, presented in Chapter 3, to include view-point localization. This is an important concept that has not previously been explored in the literature. In this chapter, the relocalization task has been reclassified into three stages. Various descriptors that were discussed in the literature were used for texture and scene classification, and have been compared; then an alternative approach to UI frame removal was proposed. Most importantly, the presented framework allows for a quantitative

evaluation of IRO view-point matching, which is an important aspect of temporal differential surveillance and by extension for SE. The following observations were made;

(i) The performance is better on NBI than on WL, because of the higher texture observable in NBI modality.

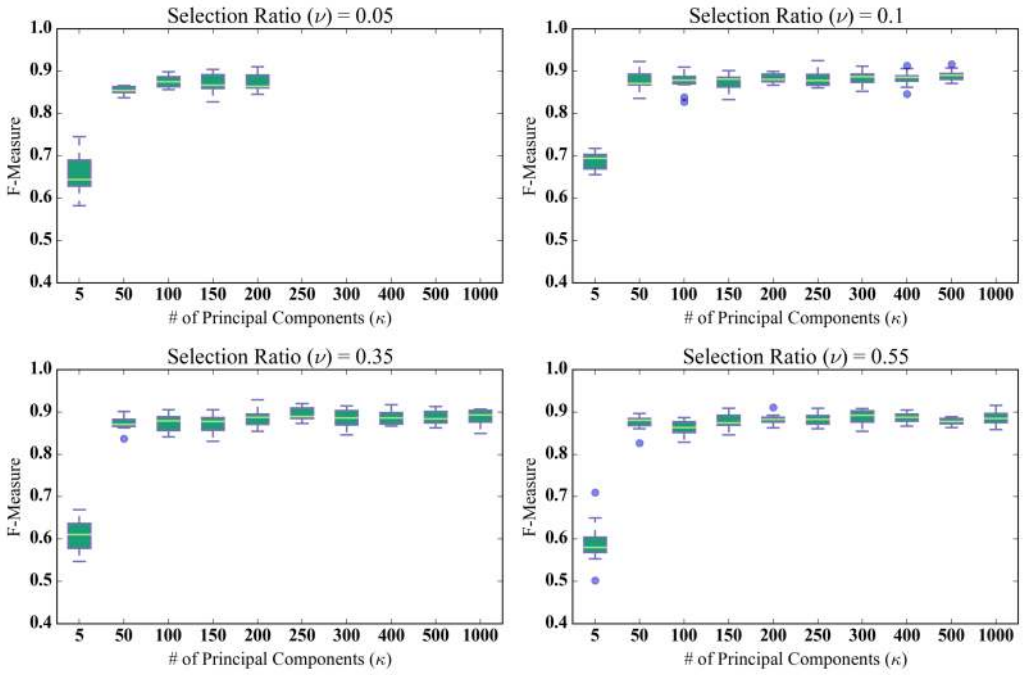
(ii) Figure 7.9 depicts the general trend of decreasing average score with increasing search radius, indicating the need for constraining the search space using GL, whereas in [Atasoy 2012b] only temporal localization of frames was considered.

(iii) From Figures 7.10 and 7.11, it is clearly observable that filtering the UI frames reduces the matches with worst score when using global descriptors. This is however more observable in WL than in NBI. Additionally, it must be noted that the matches filtered by UI frames are highly correlated with the accuracy of relocalization. If there is high confidence in EM based localization, then theoretically, using matches obtained by filtering of UI images (using a perfect classifier) would give the best possible viewpoint match.

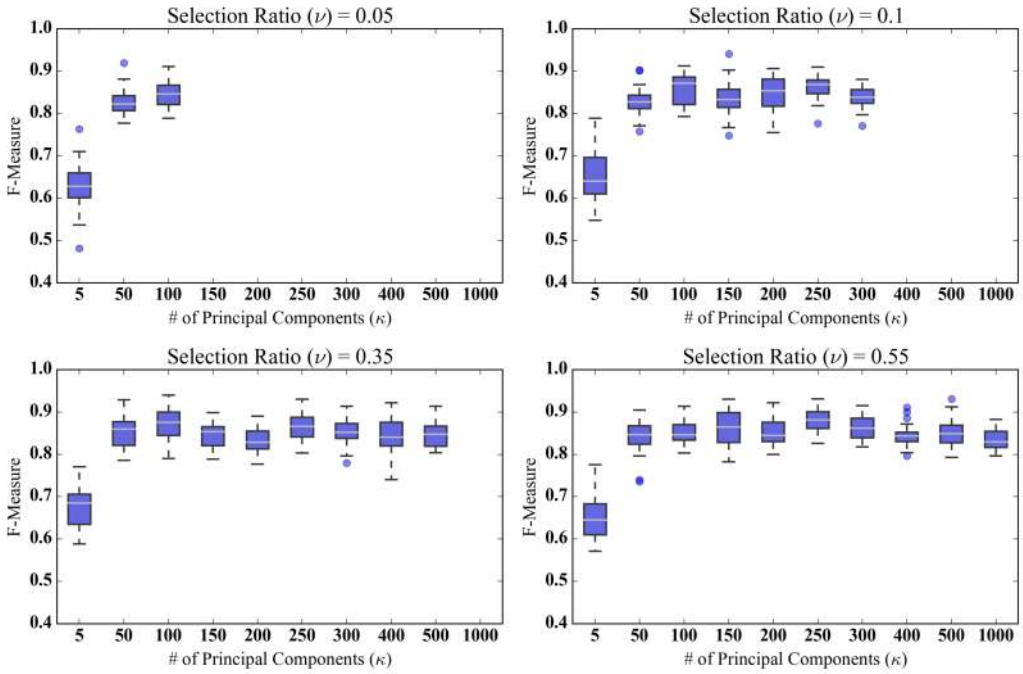
(iv) Using GS for NBI images is not meaningful as observed from Table 7.2, because it is not a true mapping from RGB space. Thus, for alternate modalities, appropriate colour spaces must be chosen, for robustness of descriptors.

(v) Tables 7.2 and 7.3 also shows that texture based descriptors such as LBP, LIOP and LTP are much better suited in this scenario. Figure 7.12 shows the plot of the 96 percentile confidence interval for the average scores obtained using EM tracker, EM tracker match filtered of UI frames and for global image match using LBP for NBI and LIOP for WL. We observe that an image based match is significantly better than an EM based match. Using UI frame filtering is, as explained earlier, correlated with the EM based match and hence its performance is only conservatively better. If we can strive for a perfect UI classifier, the EM based score can be definitely improved.

Along with the choice of illumination invariant colour-spaces such as hsv, norm and log; variations of these must be explored further. The descriptor performance in scene matching in GI endoscopy for SEs performed after many months, need to be studied, to understand the effect of tissue structure changes on them: for example, the Barrett's evolution or post radio-frequency ablation, where oesophageal scene would change completely. In such scenarios however, it is recommended to use the EM based matching, filtered of the UI frames. The method presented in this chapter discussed about providing the "best available" matching view. However such a view may not always exist in a DE. In Chapter 8, an alternative proposal is envisioned for synthetically generating a view-point using the available informative images. Figure 7.15 provides the summary work-flow of the complete approach updated by the constrained image-based refinement and off-line UI filtering step.



(a)



(b)

Figure 7.7: The box plot for the variation of F-Measure for $\nu = [0.05, 0.1, 0.35, 0.55]$ and κ principal components. (a) for NBI and (b) for WL imaging modalities.

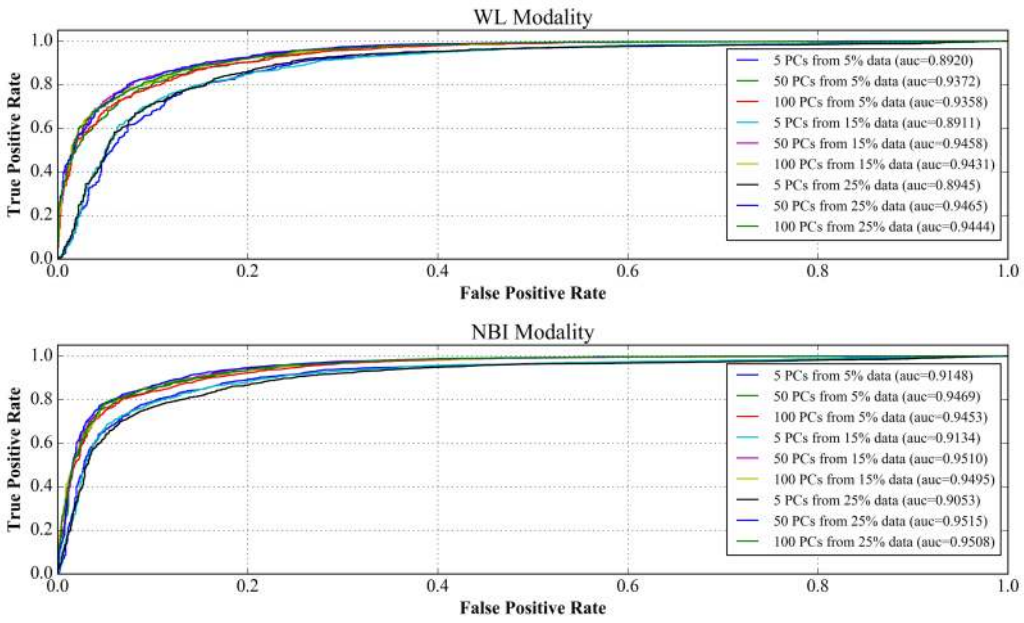
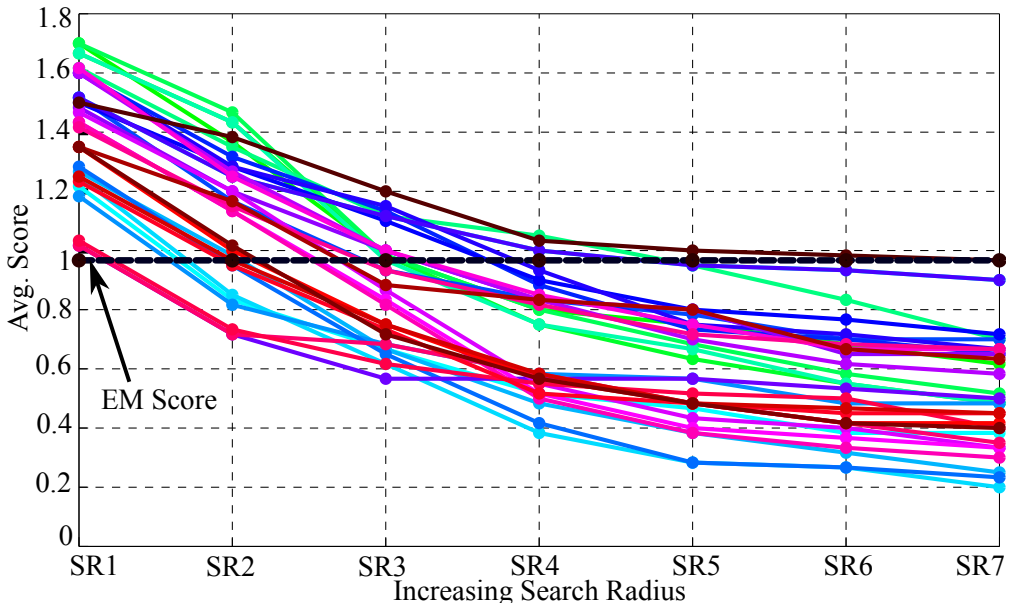
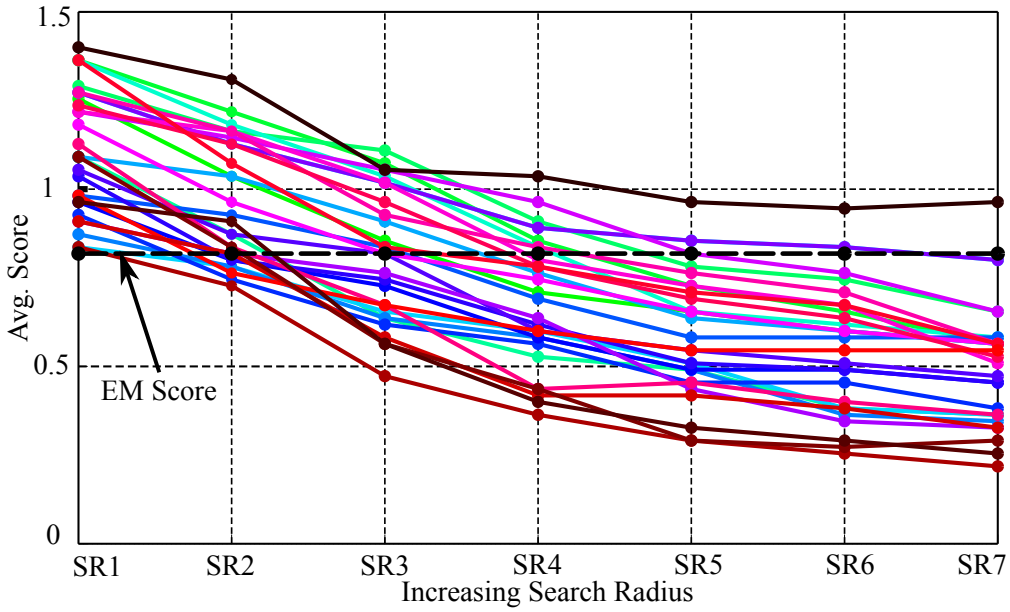


Figure 7.8: ROC curves for $\nu = [0.05, 0.1, 0.15]$ and $\kappa = [5, 50, 100]$ with area under the curve (AUC) shown in the legend. We observe here that for $\kappa = 5$, the AUC is lower than for the rest of the values. This phenomenon is observed for WL and NBI.



(a)



(b)

Figure 7.9: Average scores of descriptor and colour-space combinations that perform better than the EMTS based matching. The graph plots over an increasing search radii (10mm to 70mm). NBI (a) and WL (b).

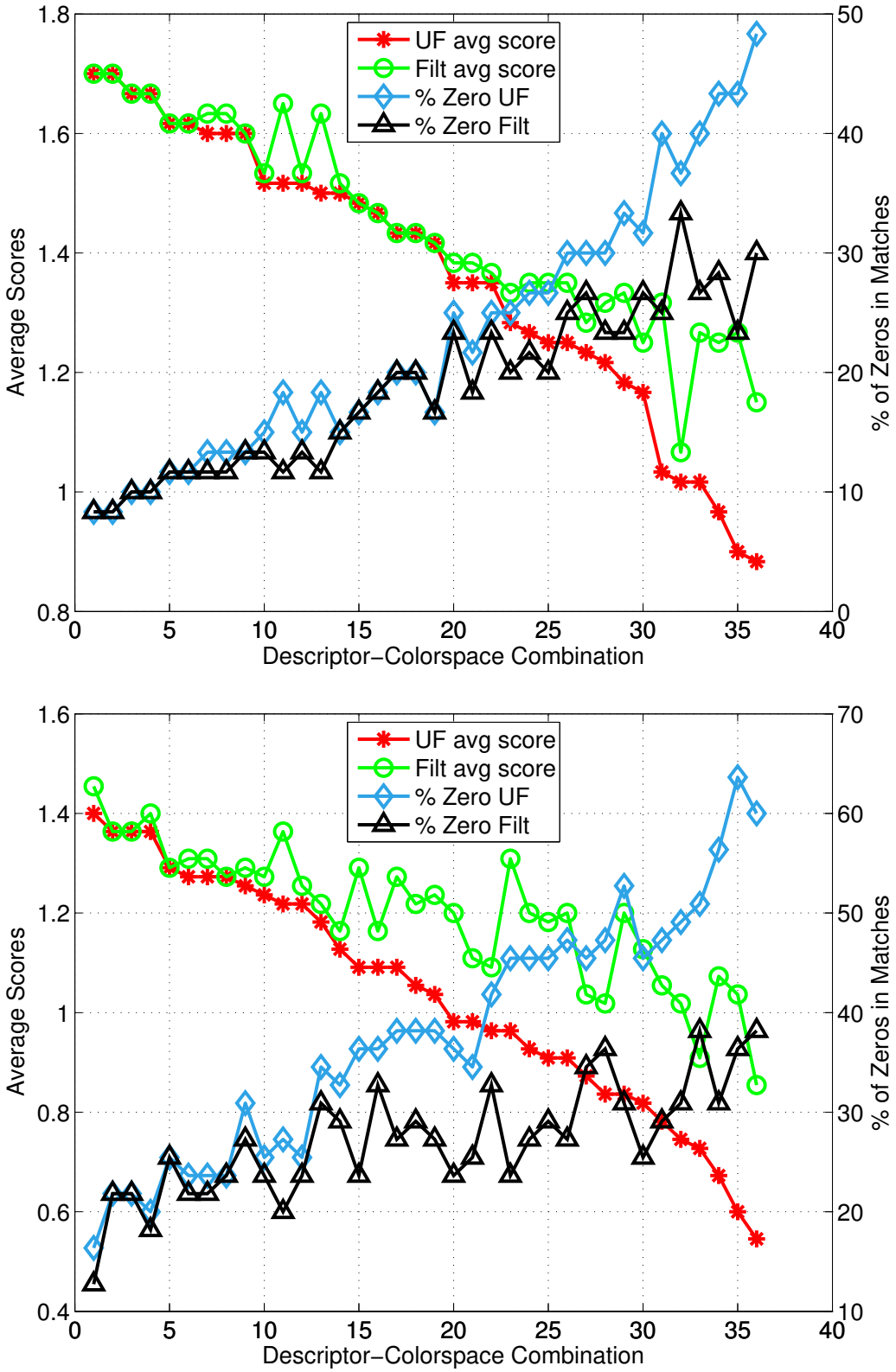


Figure 7.10: Avg. scores on the left axis and % of images matched with score zero on the right axis for all 36 of the descriptor-colorspace combinations. The x-axis is sorted by the best avg. score. NBI (*top*) and WL (*bottom*).

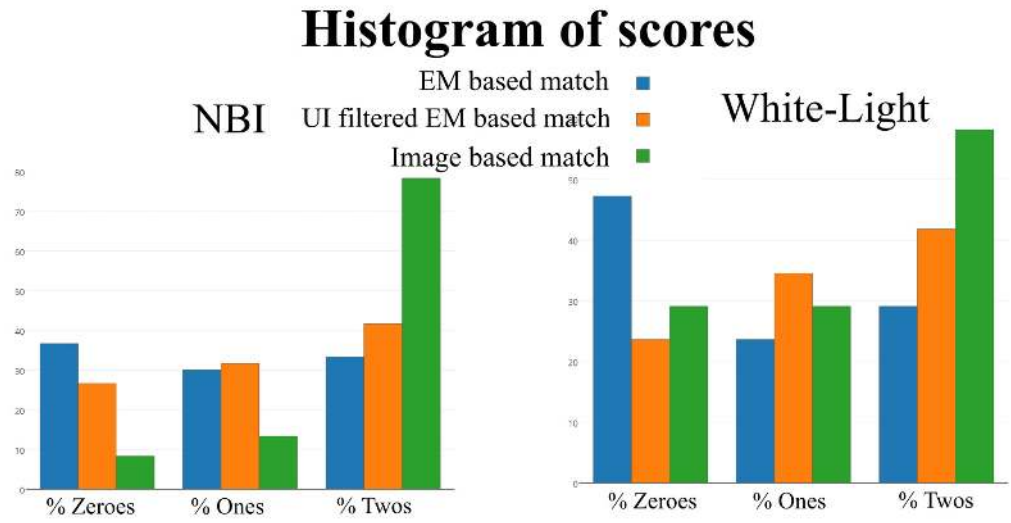


Figure 7.11: This figure shows the comparison of percentage of synchronized frames that were assigned a score=0,1 or 2 by the clinician that were detected by the three approaches; EM based match, UI filtered EM match and the best performing descriptor (mLBP, mLIOP) for constrained image match for NBI (left) and WL (right).

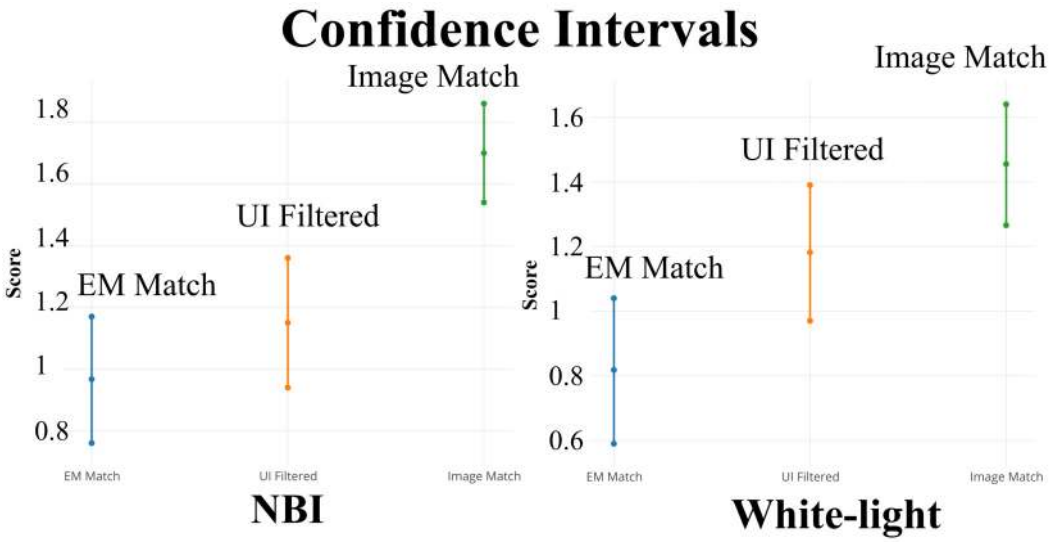


Figure 7.12: This figure shows the comparison between confidence intervals for the average scores for matches obtained using EM based approach, UI filtered EM match and the best performing descriptor (mLBP, mLIOP) for constrained image match.

NBI						
Descriptor	*ColorSpace*	*MeanScore*	*StdDev*	*Percent Zeroes*	*Percent Ones*	*Percent Twos*
mLBP	hsv	1.7	0.619	8.33	13.33	78.33
mLBP	rgb	1.7	0.619	8.33	13.33	78.33
mLBP	norm	1.667	0.655	10	13.33	76.67
mLBP	log	1.667	0.655	10	13.33	76.67
mLBP+mHOG	hsv	1.65	0.685	11.67	11.67	76.67
mLBP+mHOG	norm	1.633	0.688	11.67	13.33	75
mLBP+mHOG	rgb	1.633	0.688	11.67	13.33	75
mLBP+mHOG	log	1.633	0.688	11.67	13.33	75
mLBP	OPP	1.617	0.691	11.67	15	73.33
LBPMod	OPP	1.617	0.691	11.67	15	73.33
LBPMod	hsv	1.6	0.718	13.33	13.33	73.33
mLBP	gray	1.533	0.724	13.33	20	66.67
mLBP+mHOG	gray	1.533	0.724	13.33	20	66.67
LIOP	gray	1.517	0.748	15	18.33	66.67
LTP	gray	1.483	0.77	16.67	18.33	65
LBPMod	norm	1.467	0.791	18.33	16.67	65
LBPMod	rgb	1.433	0.81	20	16.67	63.33
LBPMod	log	1.433	0.81	20	16.67	63.33
LBPMod	gray	1.417	0.766	16.67	25	58.33
LIOP	norm	1.383	0.846	23.33	15	61.67
LIOP	OPP	1.383	0.783	18.33	25	56.67
LIOP	log	1.367	0.843	23.33	16.67	60
LIOP	rgb	1.35	0.86	25	15	60
mHOG	rgb	1.35	0.82	21.67	21.67	56.67
mHOG	norm	1.35	0.799	20	25	55
mHOG	OPP	1.333	0.837	23.33	20	56.67
mHOG	log	1.333	0.795	20	26.67	53.33
dSIFT	gray	1.317	0.854	25	18.33	56.67
mHOG	hsv	1.317	0.833	23.33	21.67	55
LIOP	hsv	1.283	0.865	26.67	18.33	55
dSIFT	rgb	1.267	0.861	26.67	20	53.33
dSIFT	hsv	1.267	0.821	23.33	26.67	50
dSIFT	opponent	1.25	0.876	28.33	18.33	53.33
mHOG	gray	1.25	0.856	26.67	21.67	51.67
mLBP+mHOG	OPP	1.15	0.86	30	25	45
EM-Based	N.A.	1.15	0.82	26.67	31.67	41.67
LTP	rgb	1.067	0.861	33.33	26.67	40

Table 7.2: Average scores of descriptors colour-space combinations for NBI (filtered of UI frames), sorted in the descending order of column 3 (*mean score*). Columns 5, 6 and 7 are the % of cases where the matched image has a score of zero, one and two respectively.

WL						
Descriptor	*ColorSpace*	*MeanScore*	*StdDev*	*Percent Zeroes*	*Percent Ones*	*Percent Twos*
LIOP	gray	1.455	0.715	12.73	29.09	58.18
LBPMod	gray	1.4	0.784	18.18	23.64	58.18
mLBP	norm	1.364	0.825	21.82	20	58.18
mLBP	log	1.364	0.825	21.82	20	58.18
LTP	gray	1.364	0.802	20	23.64	56.36
LIOP	log	1.309	0.836	23.64	21.82	54.55
mLBP	gray	1.309	0.814	21.82	25.45	52.73
mLBP+mHOG	gray	1.309	0.814	21.82	25.45	52.73
mLBP	hsv	1.291	0.875	27.27	16.36	56.36
mLBP	rgb	1.291	0.854	25.45	20	54.55
mLBP	OPP	1.291	0.832	23.64	23.64	52.73
LIOP	rgb	1.273	0.87	27.27	18.18	54.55
LBPMod	rgb	1.273	0.827	23.64	25.45	50.91
LBPMod	log	1.273	0.827	23.64	25.45	50.91
LBPMod	norm	1.255	0.821	23.64	27.27	49.09
mLBP+mHOG	rgb	1.236	0.86	27.27	21.82	50.91
LBPMod	hsv	1.218	0.896	30.91	16.36	52.73
mLBP+mHOG	log	1.218	0.875	29.09	20	50.91
LIOP	norm	1.2	0.89	30.91	18.18	50.91
mLBP+mHOG	hsv	1.2	0.848	27.27	25.45	47.27
LIOP	hsv	1.2	0.848	27.27	25.45	47.27
mHOG	gray	1.2	0.803	23.64	32.73	43.64
LTP	rgb	1.182	0.863	29.09	23.64	47.27
EM-Based	N.A.	1.182	0.796	23.64	34.55	41.82
mHOG	rgb	1.164	0.898	32.73	18.18	49.09
LBPMod	OPP	1.164	0.856	29.09	25.45	45.45
dSIFT	hsv	1.127	0.795	25.45	36.36	38.18
dSIFT	OPP	1.109	0.786	25.45	38.18	36.36
mLBP+mHOG	norm	1.091	0.867	32.73	25.45	41.82
LIOP	OPP	1.073	0.836	30.91	30.91	38.18
dSIFT	rgb	1.055	0.803	29.09	36.36	34.55
mLBP+mHOG	OPP	1.036	0.881	36.36	23.64	40
mHOG	log	1.036	0.86	34.55	27.27	38.18
mHOG	norm	1.018	0.871	36.36	25.45	38.18
dSIFT	gray	1.018	0.805	30.91	36.36	32.73
mHOG	hsv	0.909	0.823	38.18	32.73	29.09
mHOG	OPP	0.855	0.78	38.18	38.18	23.64

Table 7.3: Average scores of descriptors colour-space combinations for WL (filtered of UI frames), sorted in the descending order of column 3 (*mean score*). Columns 5, 6 and 7 are the % of cases where the matched image has a score of zero, one and two respectively.

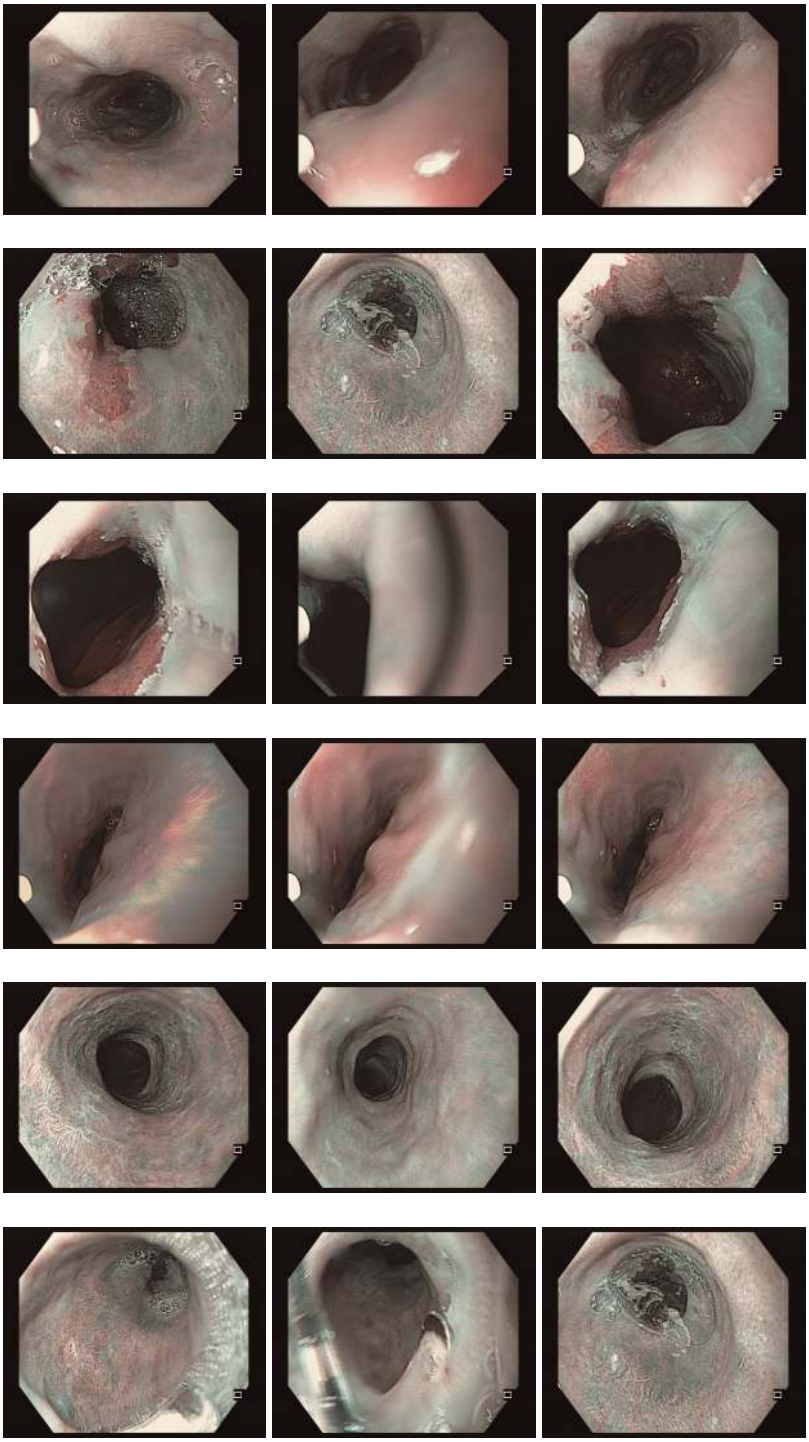


Figure 7.13: View-point match results for NBI Modality. Column-1 displays the query frames. Column-2: present the closest EMNN match and Column-3, the best view-point localized images from the available image set. Scores for matches for each row ordered by their columns [2,3]. Row 1: [0,2], Row 2: [0,2], Row 3: [0,2], Row 4: [0,2], Row 5: [0,2], Row 6: [1,2]

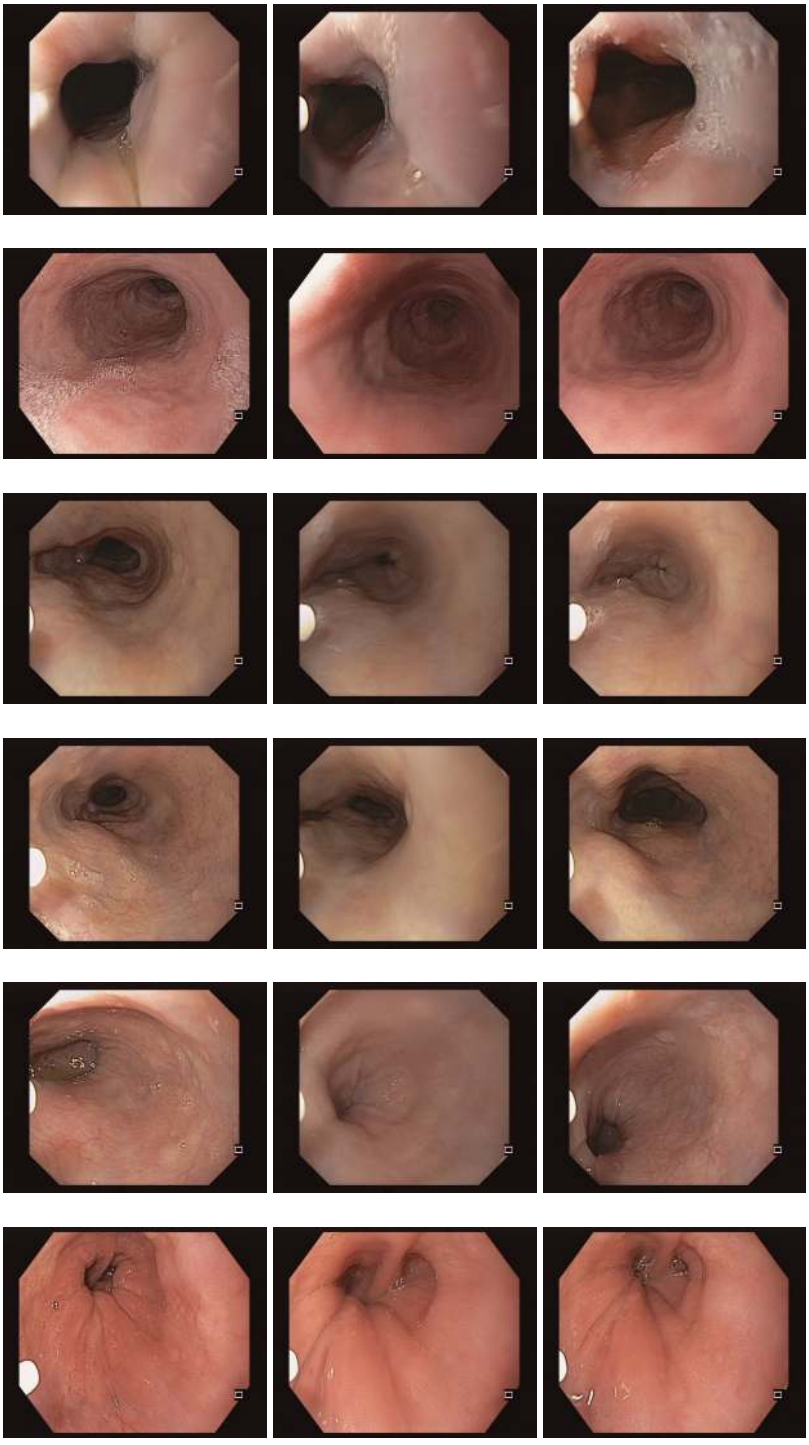


Figure 7.14: View-point match results for WL image Modality. Column-1 displays the query frames. Column-2: present the closest EMNN match and Column-3, the best view-point localized images from the available image set. Scores for matches for each row ordered by their columns [2,3]. Row 1: [0,2], Row 2: [1,2], Row 3: [1,2], Row 4: [0,2], Row 4: [1,2], Row 5: [1,2]

NBI	
Row#	Comment
1	EM-based match is only partially informative and; therefore the IA-based match is better.
2	The query frame shows the gastro-oesophageal junction, which is not clearly visible in the EM-match but is shown in image-based match. This case is also a classic example of the exact view-point not being available from the diagnostic endoscopic recording. And the IA-based match selects the best available view from the recorded data.
3	The EM-match is UI and the image based approach presents a good matching view.
4	The EM-match has slight positioning error, also partially informative, hence, the image-based match is better in this case.
5	The quality of selected image from the point of view of the visible texture, the image-based match is better than the EM-match.
6	We see that there is a slight error in depth and also a large view-point difference. This is a situation resulting due to the muscular contractions, which is somewhat corrected by the image-based approach.
WL	
1	The image-based match selects a more informative view-point from the GI expert's point-of-view.
2	The EM-match is UI which is corrected by the image-based approach.
3	The EM-match is partially informative and has poor contrast. Image-based approach selects a better image in this scenario.
4	The image-based match selects a better view-point as compared to the EM-match.
5	The EM-match is partially informative which is revised by the image-based approach.
6	In this the EM-match and image-based selection, provides almost similar results, but the quality of selected image, in terms of the visible texture is better in image-based approach.

Table 7.4: Table explaining the matches presented in each row of the Figures 7.13 and 7.14.

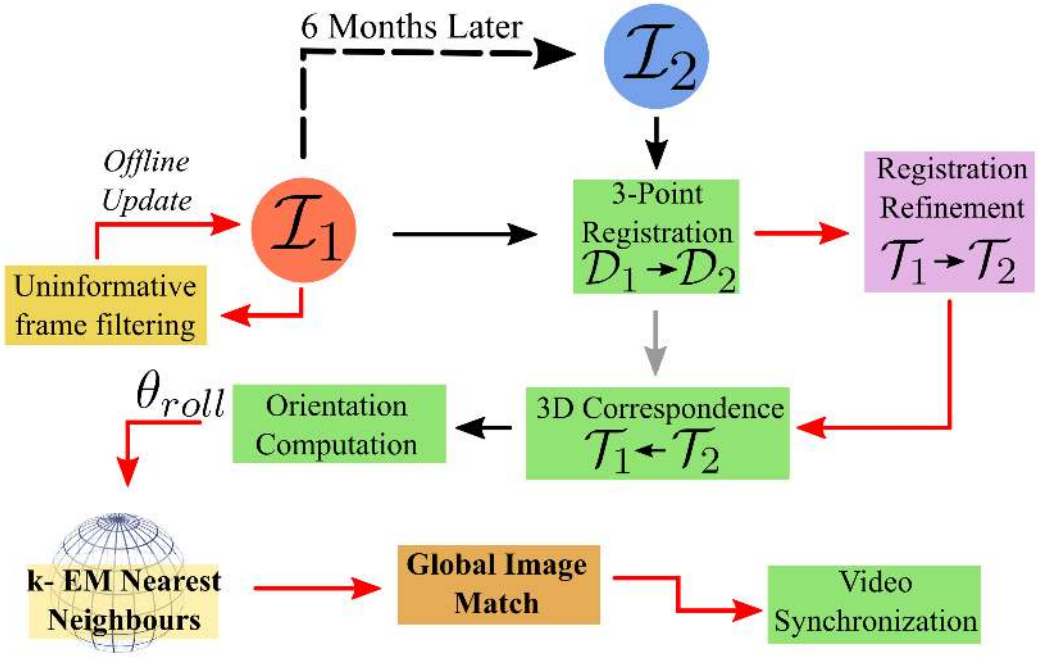


Figure 7.15: This figure summarizes the refinement step presented in this chapter as a flowchart, which includes the off-line filtering of UI frames and the constrained global image matching between synchronized frames and the k-EMNN matches.

CHAPTER 8

Conclusion and Future Perspectives

Contents

8.1 Summary

Theis thesis has presented a complete system for inter-operative relocalization for oesophageal procedures, which can be extended to other GI interventions. During the course of this thesis, the clinical constraints involved during a GI intervention were closely studied, before proposing a solution that could be easily integrated into the routine clinical work-flow. For the proposed system an assessment of noise affecting it was performed and used for generating synthetic data, which was then used for quantitative evaluation of the system. Then, on interventions recorded on pigs, using an optical tracking system as the benchmark for inter-operative registration, *in-vivo* evaluation was performed. Both these experiments provided consistent results and showed that the considered noise model closely followed the real scenario. Then, a qualitative evaluation of the system was conducted by 10 clinical specialists, with varying degrees of expertise. The experiments demonstrated that the proposed navigational guidance system provided an advantage to the clinical experts in GI endoscopy. The 3-point registration framework that was presented, is currently the most efficient in terms of the clinical work-flow. However, an alternate registration approach performed on the oesophagus model was presented and showed that in cases of large error caused due to human error, it improved the video synchronization accuracy. As will be discussed in Section 8.3, this conceptual framework can be extended to other GI interventions.

The quantitative evaluations performed on the system defined thus far, showed a region of uncertainty in relocalization using the EM based match. Additionally, the selection of best image view-point match to the live endoscopic view was also essential to provide better guidance during endoscopy. A constrained image based search near the EM based localized region was proposed. In this context, a review of the various image analysis and machine learning concepts applicable to GI endoscopy was presented. The approach involved, firstly, for detection of uninformative frames in the intervention, an approach using LBP descriptors combined with PCA, was proposed. It was shown to works well with low false positive rates. Secondly, using concepts reviewed in literature, a comparison of different features

was presented, for constrained oesophageal scene matching. Experiments conducted using patient data, indicated a significant improvement in the subjective score provided by our GI expert, for the synchronized frame, compared to using only an EM based match.

8.2 Contributions

I would like to highlight the following contributions made by this thesis.

1. *Problem Analysis*: Chapters 1 and 2, analysed the need for a navigational platform in GI endoscopy. In this chapter, GI interventions were analysed to identify two important paradigms of relocalization in surveillance endoscopy; (a) Biopsy site relocalization (BSR) and (b) Temporal differential surveillance (TDS). This distinction; although subtle, has been left unidentified in the medical literature and depending on the target application, the approach for addressing the problem or the presentation of a solution to the clinician, should vary. An analysis on these concepts in GI endoscopy was introduced in [Vemuri 2015a] and is also under review in the article [Vemuri 2016].

2. *Novel Navigational Platform*: One of the most significant challenges in GI endoscopy has been the lack of a complete navigational platform, that does not modify the existing clinical routines. One of these aspects, that was discussed in Chapter 2, was the lack of pre-operative imaging in routine GI procedures. This limits the choice of solutions available for the relocalization problem. In Chapter 3, a novel approach is presented, that introduces an electromagnetic tracking system (EMTS) in a GI endoscopic intervention to record data during the procedure and reuse in future surveillance interventions for relocalization. This approach, thus effectively pre-empts the need for any additional imaging and breaks away from the existing paradigm of instrument localization using a 3D model, thus establishing an alternate approach for flexible endoscopic navigation. The concept of using EMTS for inter-operative relocalization was first presented in [Vemuri 2013] and later expanded on in [Vemuri 2015b].

3. *Integration with Clinical Work-flow*: To achieve inter-operative relocalization, the recorded 3D EM sensor data must be registered between two surveillance interventions. Chapter 3 presents a 3-point registration methodology, that was motivated by studying the work-flow of oesophageal procedures. Chapter 5, extended this approach to use additional information of the oesophagus trajectory recorded during the intervention to perform registration. Both the proposed methodologies do not interfere with the existing clinical work-flow, thus allowing easier integration. Experimental evaluation of the second approach showed that it effectively identified any large errors due to human factor and corrected them with no significant overhead to the system. Thus, depending on the state of the system, the second method could be initiated without interference with the work-flow. The analysis and solution design was presented earlier in [Vemuri 2013, Vemuri 2015b].

4. *View-point Selection OR Image based Localization*: Chapter 7 further analyses the inter-operative relocalization from a technical standpoint to identify, three intermediate steps; (a) Gross-localization (b) Fine positioning or View-point selection and (c) Inter-frame mapping. This classification is important because it influences the choice of approach to be used at each stage. The gross-localization was tackled using EM-based video synchronization in this thesis. However, subjectively selecting the "best" view-point to match the live image is important for decision support, from clinician's perspective. This aspect of inter-operative relocalization has not been addressed in literature and is presented in Chapter 7.

   Some of the methods proposed in literature have approached the inter-operative relocalization as a content based image retrieval problem. In this thesis, it has been established, through experimental evaluation on patient data that, this in fact is not an ideal approach. It is also experimentally demonstrated, that to perform image-based matching in GI endoscopy, constraints provided by (methods such as,) EM-based localization, would be essential to reduce false matches.

   Using the conceptual model of constrained scene matching, Chapter 7 compares, 5 descriptors under two different endoscopic imaging modalities, NBI and WL. Due to the use of different wavelengths in these modalities, the choice of suitable colour-space has also been studied. The chapter presents results of this comparison, on surveillance interventions performed on 7 patients. The concept of view-point localization and as a extension the constrained image-based relocalization was originally presented in [Vemuri 2015a].

5. *Uninformative Frame Classification*: Chapter 7 also explores an alternate (generic) approach to classification of uninformative frames in endoscopy, using texture based descriptors, such as LBP. It is shown that, removing redundant information from the texture descriptors by performing dimensionality reduction, by for example, Principal component analysis, can be effectively used for this classification. The proposed approach is tested on Narrow band and White light imaging modalities, using image data from 11 interventions of each modality. It is shown that the subjective scores of the EM-based match can be improved, simply by filtering out uninformative images from the sequence. The uninformative frame classification approach discussed in this thesis was presented in [Vemuri 2015a].

6. *System Evaluation Experiments*: A detailed analysis of the various sources of uncertainty in the system was performed in Chapter 4 and a series of experiments were designed to measure them empirically. The values were used for generating synthetic interventions, which were used along with the data obtained from interventions on pigs for performing a complete quantitative evaluation of the proposed EM-based system. In the context of relocalization, an appropriate error metrics to quantify the system performance must be selected, which must not only be meaningful to the scientific community, but also informative for the clinician, to take decisions in the operating room. For oesophageal procedures, the mapping of 3D position error after video synchronization, on to the vector formed by sternum sensors provided such a

metric and corresponded to the depth error in the oesophagus. This metric was also employed to quantify the subjective responses of 10 clinical experts during qualitative evaluation of the system.

In a similar qualitative assessment of view-point selection, a new set of subjective scores (Section 7.3) were assigned to quantify the system response from a GI expert's standpoint. Although, these scores may be biased, they provide a basis for quantifying a GI expert's interpretation of view-point. The complete evaluation experiments were published in [Vemuri 2015b].

7. *Software Interface*: The work presented in this thesis was developed into a clinically inspired software interface, designed with the gastroscopy work-flow in view. This was presented in Section 3.5, with different interfaces for BSR and TDS. The software integrated the 3D trajectory information, to show the GI expert a global view of the operating region, depicting the current endoscopic position and the locations of the biopsy sites from an earlier intervention. The software interface was presented in [Vemuri 2015b].

   Further discussions with the clinicians revealed that biopsy sites during an intervention are plotted on a polar graph with the gastro-oesophageal junction at the centre. Such a graph was then integrated into the software interface to provide a familiar view for the GI experts.

8. *Literature Review*: Finally, while researching the various approaches to scene analysis and matching in GI endoscopy, it became essential to collect all the existing literature to understand from algorithmic perspective, the applicability of various feature detectors, descriptors and machine learning concepts to the GI endoscopic vision community. The discussion presented in Chapter 6 provides the necessary starting point in formulating a strategy for image analysis in GI macroscopic imaging modalities.

8.3 Perspectives

The approach presented in this thesis has provided an initial solution for the problem of inter-operative relocalization in flexible endoscopic procedures. There are many aspects to this problem which still need to be addressed, which are discussed below. Furthermore, extensions to other clinical applications in GI anatomy that can also be envisioned, are presented here.

1. Since, during an endoscopic procedure, the patient is under general anaesthesia; the breathing is stable and smooth. Using the movement of the external sensor at $\mathcal{L}_B$ during the breathing cycle, the elongation of the oesophagus can be modelled as function of this variable. This could be modelled as a deformation which quadratically decreases, when the distance to the stomach increases. In order to use the information from the sensor at $\mathcal{L}_B$, its optimal position on the patient must be firstly studied.

2. Using the endoscopic data to generate a 3D reconstruction is a widely researched topic in GI community. A realistic reconstruction can be used to artificially generate view-point closest to the matched view from DE, addressing the problem discussed in Chapter 7. Additionally, they can be used for building realistic simulators, for performing post-operative review or pre-operative surveillance planning for future interventions. In Visentini *et al.* [Visentini-Scarzanella 2012], the authors present the estimation of perspective SFS models for surfaces lit by a near point light source. A combination of the EM sensor positions with the information from SFS reconstruction can be used to obtain 3D view inside the GI tract. An alternative to SFS reconstruction would be to use visual SLAM by integrating the EM sensor position for deeper navigation in the gastric tract.

3. An important aspect of video synchronization is the need for selecting the best matching frame with the correct view-point. In absence of the desired view-point, choosing images that are lumen centric would be a more realistic option than using global image matching as described in Chapter 7. In this context, an approach to identify an incorrect match must be researched.

4. A combination of the methods described in Chapters 3 and 7 can be extended to colonoscopy. The most important challenge for re-localization in colonoscopic procedures is to counter the elastic nature of the colon. However, in a typical colonoscopic intervention, the clinician inserts the scope all the way to the beginning of the colon and retracts it slowly while cleaning and inspecting the surface. In such a situation, the colon is almost completely extended. The problem can be approximated to that of non-rigid curve matching or rigid registration in presence of uncertainty as described in [Maier-Hein 2012, Granger 2006]. Using known anatomic landmarks (such as the hepatic and splenic flexure) along the colon, gross-localization could be performed, using the EMTS. Fine positioning can be obtained using the image information capturing the colon shape and texture, that was proposed in Chapter 7. This would effectively reduce the region of uncertainty for inter-operative re-localization in colonoscopy. The proposal is illustrated in Figure 8.1. However, firstly, the repeatability in the detection of the above mentioned (or other) anatomical landmarks must be studied on several human subjects.

5. Using the inter-operative localization framework a top-down image database can be built that would include the endoscopic images to microscopic imagery from CLE or OCT along with the histopathological images of the biopsied tissues. Thus, simplifying the storage and filing of patient records. To achieve this, a standardized format (such as DICOM for radiological data), must be established. Then, a suitable interface to navigate the recorded data can be developed for cross-platform access.

6. For colonoscopy, inter-modality relocalization is an important problem to be tackled. Since, the invention of wireless capsule endoscopy (WCE), it has gained in popularity for pre-operative diagnostic inspection. Although the volume of information recorded using a WCE is rather large, several approaches have been proposed to

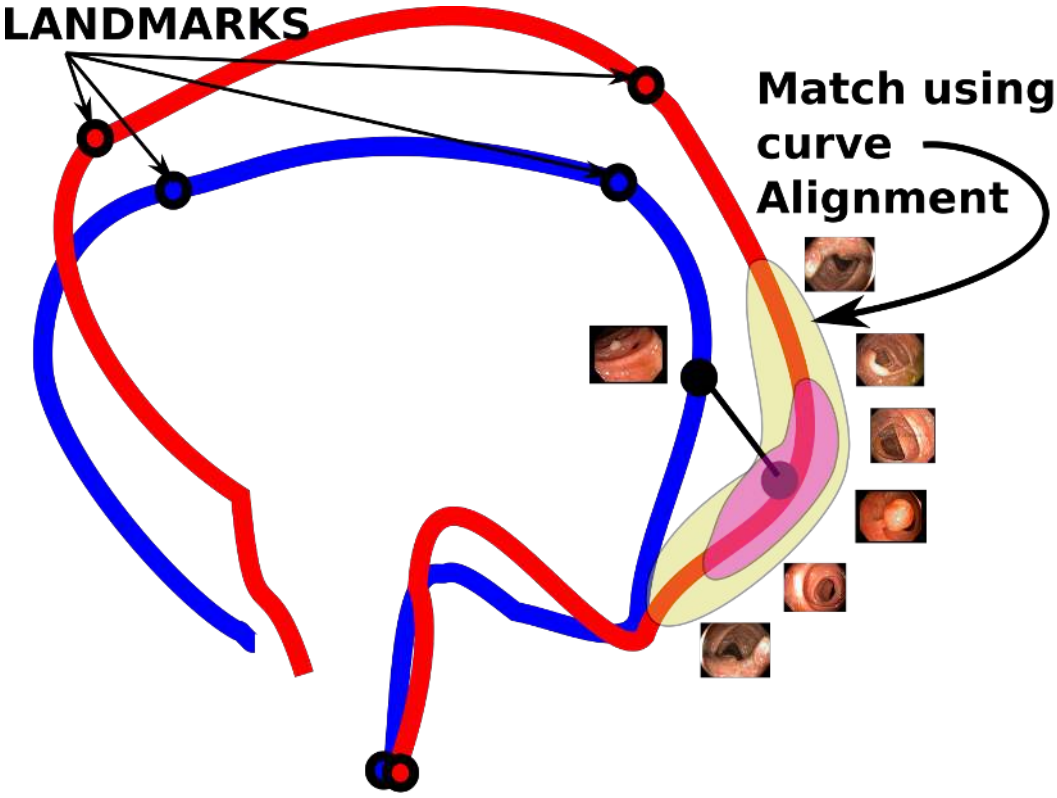


Figure 8.1: The extension of the concept discussed in Chapters 3 and 7 is illustrated here for the colon. The hepatic and splenic flexure are visible in the colonoscopic frames and can be used as landmarks for trajectory registration. But, the variations in the tissue texture along the colon can also be used as landmarks. The 3D EM based curve matching can be used to provide the gross localization. The shaded areas indicate the regions of uncertainty for searching the best match.

summarize them based on anatomy and presence of suspicious regions. Although, this information is available pre-operatively to the GI expert, there is a large uncertainty in the relocalization of any region of interest that was identified during capsule endoscopy inspection. An extension of the system proposed in this thesis to perform relocalization with WCE would be highly beneficial to the GI expert.

7. Although the video synchronization based navigational system is already very useful for the GI experts; displaying in the live view, an augmented reality based view of the previous biopsy positions can be considered as a final goal. To achieve this, an important aspect that needs to be addressed is the inter-frame biopsy (or interest) site mapping. Once the best view-point has been obtained, the next step involves mapping the regions of interest such as, biopsy sites from the matched image to the live view, as in [Atasoy 2009]. However, to obtain such inter-operative mapping an understanding of the variation in tissue texture with time and the reliability of the key-point matching approaches is necessary. The vasculature network along the oesophageal mucosa

is visible only on some patients, and also depends on the imaging modality (NBI™ by Olympus, FICE™ by Fuji, SPIES™ by Karl Storz and i-scan™ by Pentax). The visibility of this vascular structure changes due to contractions of the oesophagus muscles. Additionally, it is yet unknown if the vascular network remains static or changes over a period of time. In case of a Barrett's oesophagus, it is well known that the length of the Barrett's evolves with time. Hence, patient data must be regularly collected on a cohort to analyse the behaviour of the oesophagus mucosa, under different endoscopic modalities, to evaluate the texture dependence on repeatability.

Appendix A

Endoscopic Imaging

Several imaging modalities have been developed over the last two decades to aide in better visualization, characterization and demarcation of abnormal tissue from the normal mucosa [Subramanian 2014, Neumann 2011, Kiesslich 2011].

- Methylene Blue Chromoendoscopy (MBC): It involves spraying a blue dye (in the oesophagus or colon) and is used to improve identification of suspicious areas of the mucosa. The methylene blue temporarily stains the mucosa, helping identify areas of dysplasia and early cancers to help identify the best places to obtain biopsies. Research studies have showed variable success using this technique to identify dysplasia [Sharma 2001, Kiesslich 2003].

- Storz Professional Image Enhancement System (SPIES™): It is a platform developed by Storz to provide varying degrees of improvement to the traditional endoscopic image. Such as the SPIES CLARA™, which provides uniform illumination to every part of the endoscopic system. The SPIES CHROMA™, intensifies image colour contrast while retaining the natural colour perception. The SPIES SPECTRA™, to filter out suitable spectra to clearly differentiate between tissues.

- Narrow band imaging (NBI): A method that uses special filters on the endoscope to narrow the color spectrum of red, green, and blue light and increase the intensity of the blue light. When NBI is combined with high magnification or high resolution endoscopy, the detailed pattern of the mucosa and blood vessels can be seen [Gono 2004, Yoshida 2004]. Some studies have shown small, but significant benefit to adding NBI to high resolution endoscopy for identifying high grade dysplasia and early cancers.

- Autofluorescence imaging (AFI): It uses laser light during endoscopy to stimulate the natural fluorescence of the oesophageal mucosa. The fluorescence is captured and processed by a computer showing the difference between normal and abnormal mucosa that can be seen during the endoscopic procedure. AFI has a may prove to be an excellent method for scanning large areas of mucosa for neoplasia [Matsuda 2008, Haringsma 2001]. The standard endoscopy image and autofluorescence image could potentially allow the endoscopist to identify the most suspicious areas to biopsy.

- Confocol Laser Endosmicroscopy (CLE): A new method for examining mucosal histology in real-time during an endoscopic procedure. The CLE is a standard probe with a microscope built into the tip. It magnifies the mucosa 1000 times normal, so microscopic pictures of cells can be obtained. Normal mucosal cells can be

distinguished from dysplasia and cancers directly [Kiesslich 2006, Polglase 2005, Kiesslich 2004]. In colorectal procedures a combination of MCB with CLE has shown improved performance [Bae 2014].

- Optical Coherence Tomography (OCT): This modality relies on light scattering by tissues [Bouma 2000, Sivak 2000].
- Fuji Intelligent Chromo Endoscopy (FICE): It is a software based spectral estimation to produce spectral images, and select spectral images of given wavelengths, and finally assign the chosen spectral images to the Red, Green, and Blue monitor input channels. FICE can select various combinations of wavelengths from all the light captured by the CCD and display a variety of different images [Chung 2010]. However, fewer studies have been reported on this modality.
- Capsule Endoscopy: Typically used for colonoscopic surveillance. It consists of a camera fitted in a small capsule that is swallowed by the patient. The capsule traverses through the GI tract while capturing images that are later reviewed by the clinician before performing an endoscopy. It aides in identifying suspicious regions such as polyps or dysplasia. It is generally used for pre-surveillance of colon [Iddan 2000]. Although its diagnostic value is variable, however, review of the acquired images provides the GI expert with an idea of what to expect.
- High frequency endoscopic ultrasonography (EUS): has an important role in the evaluation of benign and malignant gastrointestinal diseases. Echoendoscopes operate from 5 to 20 MHz, permitting a spectrum of depth of penetration and image resolution. Higher frequencies provide higher resolution but less penetration, while lower frequencies provide higher penetration but lower resolution.
- HD Zoom Endoscopy.
- PENTAX Medical i-scan: It image processing (Surface, Contrast, and Tone Enhancement) to provide digital image enhanced endoscopy (IEE). i-scan provides real-time virtual chromoendoscopy for a detailed view of mucosal and vascular patterns.

Appendix B

Optimal Rotation and Translation Between Corresponding 3D Points

---

The registration of 3D point sets ($N \geq 3$) can be obtained by solving for R and t using the least-squares formulation;

$$\min_{R,t} \sum_{i=1}^{N} \|P_A^i - (RP_B^i + t)\|^2 \tag{B.1}$$

There are a few ways to finding the optimal solution (REFS). One of the approaches is to use Singular value decomposition (SVD). To find the optimal rotation, as shown in [Huang 1986]; we first re-centre both the point-sets so that both centroids are at the origin, which removes the translation component. Based on the correspondences, we can construct the cross correlation matrix **H** between two centred point clouds as,

$$H = \sum_{i=1}^{N} (P_A^i - centroid_A)\,(P_B^i - centroid_B)^T \tag{B.2}$$

The optimal solution for the least squares problem can be defined by the rotation R as;

$$[U, S, V] = svd(H) \tag{B.3}$$

$$R = VU^T \tag{B.4}$$

The special reflectance case, when SVD will return a reflection of R matrix. This can be addressed by checking $determinant(R) < 0$ and multiplying the third column of R by -1. The translation from the target point cloud to the source can then be defined as;

$$t = centroid_B - R\, centroid_A \tag{B.5}$$

Bibliography

[Abe 2015] Koji Abe, Hidenori Takagi, Masahide Minami and Haiyan Tian. *Computer-aided Diagnosis for Internal Hemorrhoids by Measuring the Congestive Extent in Endoscopic Images*. Journal of Biomedical Engineering and Medical Imaging, vol. 1, no. 6, pages 10–21, Society of science and enducation, UK, Dec 2015. (Cited on page 79.)

[Abouelenien 2013] Mohamed Abouelenien, Xiaohui Yuan, Balathasan Giritharan, Jianguo Liu and Shoujiang Tang. *Cluster-based sampling and ensemble for bleeding detection in capsule endoscopy videos*. American Journal of Science and Engineering, vol. 2, no. 1, pages 24–32, 2013. (Cited on pages 78 and 79.)

[Alexandre 2008] Luís a. Alexandre, Nuno Nobre and João Casteleiro. *Color and position versus texture features for endoscopic polyp detection*. BioMedical Engineering and Informatics: New Development and the Future - Proceedings of the 1st International Conference on BioMedical Engineering and Informatics, BMEI 2008, vol. 2, pages 38–42, 2008. (Cited on page 74.)

[Alizadeh 2015] Mahdi Alizadeh, Kaveh Sharzehi, Alireza Talebpour, Hamid Soltanian-Zadeh, Hoda Eskandari and Omid Haji Maghsoudi. *Detection of uninformative regions in wireless capsule endoscopy images*. In Biomedical Engineering Conference (NEBEC), 2015 41st Annual Northeast, pages 1–2. IEEE, 2015. (Cited on page 85.)

[Allain 2009] Baptiste Allain, Mingxing Hu, Laurence B Lovat, Richard Cook, Sebastien Ourselin and David Hawkes. *Biopsy site re-localisation based on the computation of epipolar lines from two previous endoscopic images*. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), vol. 12, pages 491–8. Springer, Heidelberg, Jan 2009. (Cited on page 14.)

[Allain 2010] Baptiste Allain, Mingxing Hu, Laurence B Lovat, Richard J Cook, Tom Vercauteren, Sebastien Ourselin and David J Hawkes. *A system for biopsy site re-targeting with uncertainty in gastroenterology and oropharyngeal examinations*. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), vol. 13, pages 514–21. Springer, Jan 2010. (Cited on page 14.)

[Allain 2012] Baptiste Allain. *Re-localisation of microscopic lesions in their macroscopic context for surgical instrument guidance*. PhD thesis, UCL (University College London), 2012. (Cited on pages 14 and 31.)

[Ameling 2009] Stefan Ameling, Stephan Wirth, Dietrich Paulus, Gerard Lacey and Fernando Vilarino. *Texture-based polyp detection in colonoscopy*. In Bildverarbeitung für die Medizin 2009, pages 346–350. Springer, 2009. (Cited on page 74.)

[An 2005] Yong Hwan An, Sae Hwang, JungHwan Oh, JeongKyu Lee, Wallapak Tavanapong, Piet C de Groen and Johnny Wong. *Informative-frame filtering in endoscopy videos*. In Medical Imaging, pages 291–302. International Society for Optics and Photonics, 2005. (Cited on page 78.)

[Angermann 2015] Quentin Angermann, Aymeric Histace, Olivier Romain, Xavier Dray, Andrea Pinna and Bertrand Granado. *Smart Videocapsule for Early Diagnosis of Colorectal Cancer: Toward Embedded Image Analysis*. In Computational Intelligence in Digital and Network Designs and Applications, Part 2: Digital, Network Designs and Applications, Chapter 12, page 25. Springer, May 2015. (Cited on page 73.)

[Armesto 2010] Leopoldo Armesto, Javier Minguez and Luis Montesano. *A generalization of the metric-based iterative closest point technique for 3D scan matching*. In Proceedings of IEEE International Conference on Robotics and Automation (ICRA), pages 1367–1372. IEEE, 2010. (Cited on page 59.)

[Arnold 2011] Mirko Arnold, Stefan Ameling, Anarta Ghosh and Gerard Lacey. *Quality Improvement of Endoscopy Videos*. In Proceedings of the 8th IASTED International Conference on Biomedical Engineering, Insbruck, Austria. ACTA Press, 2011. (Cited on page 72.)

[Asari 2000] K Vijayan Asari. *A fast and accurate segmentation technique for the extraction of gastrointestinal lumen from endoscopic images*. Medical engineering & physics, vol. 22, no. 2, pages 89–96, Elsevier, 2000. (Cited on page 78.)

[Ascension-tech] Ascension-tech. *TrakSTAR and driveBAY*. http://www.ascension-tech.com. (Cited on page 19.)

[Atasoy 2009] Selen Atasoy, Ben Glocker, Stamatia Giannarou, Diana Mateus, Alexander Meining, Guang-Zhong Yang and Nassir Navab. *Probabilistic region matching in narrow-band endoscopy for targeted optical biopsy*. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pages 499–506. Springer, 2009. (Cited on pages 14 and 108.)

[Atasoy 2011] Selen Atasoy, Diana Mateus, Alexander Meining, Guang-Zhong Yang and Nassir Navab. *Targeted optical biopsies for surveillance endoscopies*. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pages 83–90. Springer, 2011. (Cited on page 15.)

[Atasoy 2012a] Cana Selen Atasoy. *Endoscopic Video Manifolds for Scene Recognition in Targeted Optical Biopsy*. PhD thesis, Technische Universität München, 2012. (Cited on pages 20 and 77.)

[Atasoy 2012b] Selen Atasoy, Diana Mateus, Alexander Meining, Guang-Zhong Yang and Nassir Navab. *Endoscopic video manifolds for targeted optical biopsy*. Medical Imaging, IEEE Transactions on, vol. 31, no. 3, pages 637–653, IEEE, 2012. (Cited on pages 15, 77, 82, 85 and 91.)

[Azagury 2012] DE Azagury, M Ryou, SN Shaikh, R San José Estépar, BI Lengyel, J Jagadeesan, KG Vosburgh and CC Thompson. *Real-time computed tomography-based augmented reality for natural orifice transluminal endoscopic surgery navigation*. British Journal of Surgery, vol. 99, no. 9, pages 1246–1253, Wiley Online Library, 2012. (Cited on page 4.)

[Baaziz 2010] Nadia Baaziz, Omar Abahmane and Rokia Missaoui. *Texture feature extraction in the spatial-frequency domain for content-based image retrieval*. arXiv preprint arXiv:1012.5208, 2010. (Cited on page 76.)

[Bae 2014] Song I Bae and You Sun Kim. *Colon cancer screening and surveillance in inflammatory bowel disease*. Clinical endoscopy, vol. 47, no. 6, pages 509–515, 2014. (Cited on page 112.)

[Bae 2015] Seung Hwan Bae and Kuk-Jin Yoon. *Polyp Detection via Imbalanced Learning and Discriminative Feature Learning*. Medical Imaging, IEEE Transactions on, pages 2379–2393, IEEE, 2015. (Cited on page 75.)

[Barbosa 2012] Daniel C Barbosa, Dalila B Roupar, Jaime C Ramos, A Tavares and C Lima. *Automatic small bowel tumor diagnosis by using multi-scale wavelet-based analysis in wireless capsule endoscopy images*. Biomedical Engineering Online, vol. 11, no. 3, 2012. (Cited on page 79.)

[Barreto 2009] Joao Barreto, Jose Roquette, Peter Sturm and Fernando Fonseca. *Automatic camera calibration applied to medical endoscopy*. In 20th British Machine Vision Conference (BMVC), pages 1–10. The British Machine Vision Association (BMVA), 2009. (Cited on page 44.)

[Bashar 2010] Md Khayrul Bashar, Takayuki Kitasaka, Yasuhito Suenaga, Yoshito Mekada and Kensaku Mori. *Automatic detection of informative frames from wireless capsule endoscopy images*. Medical Image Analysis, vol. 14, no. 3, pages 449–470, Elsevier, 2010. (Cited on page 85.)

[Baumhauer 2008] Matthias Baumhauer, Marco Feuerstein, Hans-Peter Meinzer and J Rassweiler. *Navigation in endoscopic soft tissue surgery: perspectives and limitations*. Journal of Endourology, vol. 22, no. 4, pages 751–766, Mary Ann Liebert Publishers, 2008. (Cited on page 4.)

[Bay 2008] Herbert Bay, Andreas Ess, Tinne Tuytelaars and Luc Van Gool. *Speeded-up robust features (SURF)*. Computer vision and image understanding, vol. 110, no. 3, pages 346–359, Elsevier, 2008. (Cited on page 75.)

[Bejakovic 2009] Srdan Bejakovic, Rajesh Kumar, Themistocles Dassopoulos, Gerard Mullin and Gregory Hager. *Analysis of Crohn's disease lesions in capsule endoscopy images*. In Proceedings of IEEE International Conference on Robotics and Automation (ICRA), pages 2793–2798. IEEE, 2009. (Cited on page 72.)

[Bell 2013] Charreau S Bell, Keith L Obstein and Pietro Valdastri. *Image partitioning and illumination in image-based pose detection for teleoperated flexible endoscopes*. Artificial intelligence in medicine, vol. 59, no. 3, pages 185–196, Elsevier, 2013. (Cited on page 16.)

[Bell 2014] Charreau S. Bell, Gustavo A. Puerto, Gian-Luca Mariottini and Pietro Valdastri. *Six DOF motion estimation for teleoperated flexible endoscopes using optical flow: A comparative study*. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), pages 5386–5392. IEEE, May 2014. (Cited on page 73.)

[Bellekens 2014] Ben Bellekens, Vincent Spruyt, Rafael Berkvens and Maarten Weyn. *A Survey of Rigid 3D Pointcloud Registration Algorithms*. In Fourth International Conference on Ambient Computing, Applications, Services and Technologies, pages 8–13. IARA, 2014. (Cited on page 60.)

[Berens 2005] Jeff Berens, Michal Mackiewicz and Duncan Bell. *Stomach, intestine, and colon tissue discriminators for wireless capsule endoscopy images*. In Medical Imaging, pages 283–290. International Society for Optics and Photonics, 2005. (Cited on page 77.)

[Bergström 2014] Per Bergström and Ove Edlund. *Robust registration of point sets using iteratively reweighted least squares*. Computational Optimization and Applications, vol. 58, no. 3, pages 543–561, Springer, 2014. (Cited on page 61.)

[Bernal 2012] Jorge Bernal, Javier Sánchez and Fernando Vilarino. *Integration of valley orientation distribution for polyp region identification in colonoscopy*. In Abdominal Imaging. Computational and Clinical Applications, pages 76–83. Springer, 2012. (Cited on page 73.)

[Bernal 2014a] Jorge Bernal, Debora Gil, Carles Sánchez and F Javier Sánchez. *Discarding Non Informative Regions for Efficient Colonoscopy Image Analysis*. In Proceedings of the Computer-Assisted and Robotic Endoscopy Workshop, MICCAI, pages 1–10. Springer, 2014. (Cited on page 85.)

[Bernal 2014b] Jorge Bernal, Joan Manel Núñez, F Javier Sánchez and Fernando Vilariño. *Polyp Segmentation Method in Colonoscopy Videos by Means of MSA-DOVA Energy Maps Calculation*. In Clinical Image-Based Procedures. Translational Research in Medical Imaging, pages 41–49. Springer, 2014. (Cited on page 72.)

[Besl 1992] P.J. Besl and Neil D McKay. *A method for registration of 3-D shapes*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 14, no. 2, pages 239–256, IEEE, Feb 1992. (Cited on page 59.)

[Biswas 2014] Mainak Biswas. Hilbert Huang Transform Based Video Analysis for Detecting Various Colon Diseases using Composite Similarity Measure. Master's thesis, Jadavpur University, 2014. (Cited on page 77.)

[Bosch 2007] Anna Bosch, Andrew Zisserman and Xavier Munoz. *Image classification using random forests and ferns*. In Proceedings of the 11th IEEE International Conference on Computer Vision (ICCV), pages 1–8. IEEE, 2007. (Cited on page 75.)

[Boser 1992] Bernhard E Boser, Isabelle M Guyon and Vladimir N Vapnik. *A training algorithm for optimal margin classifiers*. In Proceedings of the fifth annual workshop on Computational learning theory, pages 144–152. ACM, 1992. (Cited on page 79.)

[Boulougoura 2004] Maria Boulougoura, Edmund Wadge, V S Kodogiannis and Hardial S Chowdrey. *Intelligent systems for computer-assisted clinical endoscopic image analysis*. In Proceedings of the 2nd International Conference on Biomedical Engineering, pages 405–408. ACTA Press, 2004. (Cited on page 79.)

[Bouma 2000] Brett E Bouma, Guillermo J Tearney, Carolyn C Compton and Norman S Nishioka. *High-resolution imaging of the human esophagus and stomach in vivo using optical coherence tomography*. Gastrointestinal Endoscopy, vol. 51, no. 4, pages 467–474, Elsevier, 2000. (Cited on page 112.)

[Bourbakis 2005] N Bourbakis, Sokratis Makrogiannis and Despina Kavraki. *A neural network-based detection of bleeding in sequences of WCE images*. In Proceedings of the 5th IEEE Symposium on Bioinformatics and Bioengineering (BIBE), pages 324–327. IEEE, 2005. (Cited on page 79.)

[Breiman 2001] Leo Breiman. *Random forests*. Machine learning, vol. 45, no. 1, pages 5–32, Springer, 2001. (Cited on page 79.)

[Brown 2011] Matthew Brown and Sabine Süsstrunk. *Multi-spectral SIFT for scene category recognition*. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pages 177–184. IEEE, 2011. (Cited on page 73.)

[Bulat 2007] J Bulat, K Duda, M Duplaga, R Fraczek, A Skalski, M Socha, P Turcza and T P Zielinski. *Data processing tasks in wireless GI endoscopy: image-based capsule localization and navigation and video compression*. In Proceedings of the 29th Annual International Conference of IEEE Engineering in Medicine and Biology Society (EMBS), 3, pages 2815–8. IEEE, 2007. (Cited on pages 75 and 79.)

[Cash 2007] David M Cash, Michael I Miga, Sean C Glasgow, Benoit M Dawant, Logan W Clements, Zhujiang Cao, Robert L Galloway and William C Chapman. *Concepts and preliminary data toward the realization of image-guided liver surgery*. Journal of Gastrointestinal Surgery, vol. 11, no. 7, pages 844–859, Springer, 2007. (Cited on page 59.)

[Castaneda 2009] V. Castaneda, S. Atasoy, D. Mateus, N. Navab and a Meining. *Reconstructing the Esophagus Surface from Endoscopic Image Sequences*. In Proceedings of the 5th Russian-Bavarian Conference on Bio-Medical Engineering, 2009. (Cited on pages 71 and 75.)

[Cha 2007] Sung-Hyuk Cha. *Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions*. International Journal of Mathematical Models and Methods in Applied Sciences, vol. 1, no. 4, pages 300–307, North Atlantic University Union, 2007. (Cited on page 78.)

[Chang 2011] Chih-Chung Chang and Chih-Jen Lin. *LIBSVM: A library for support vector machines*. ACM Transactions on Intelligent Systems and Technology (TIST), vol. 2, no. 3, page 27, ACM, 2011. (Cited on page 86.)

[Charisis 2013] Vasileios S. Charisis, Christina Katsimerou, Leontios J. Hadjileontiadis, Christos N. Liatsos and George D Sergiadis. *Computer-aided capsule endoscopy images evaluation based on color rotation and texture features: An educational tool to physicians*. In Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems (CBMS), pages 203–208. IEEE, 2013. (Cited on page 74.)

[Chen 1991] Yung Chen and Gérard Medioni. *Object modeling by registration of multiple range images*. In Proceedings of IEEE International Conference on Robotics and Automation (ICRA), pages 2724–2729. IEEE, 1991. (Cited on pages 59 and 60.)

[Chen 2015] J Chen, Y Wang and Y X Zou. *An adaptive redundant image elimination for Wireless Capsule Endoscopy review based on temporal correlation and color-texture feature similarity*. In Digital Signal Processing (DSP), 2015 IEEE International Conference on, pages 735–739. IEEE, 2015. (Cited on page 78.)

[Chetverikov 2005] Dmitry Chetverikov, Dmitry Stepanov and Pavel Krsek. *Robust Euclidean alignment of 3D point sets: the trimmed iterative closest point algorithm*. Image and Vision Computing, vol. 23, no. 3, pages 299–309, Elsevier, 2005. (Cited on page 60.)

[Cho 2007] Yong Beom Cho, Woo Yong Lee, Hae Ran Yun, Won Suk Lee, Seong Hyeon Yun and Ho-Kyung Chun. *Tumor localization for laparoscopic colorectal surgery*. World journal of surgery, vol. 31, no. 7, pages 1491–1495, Springer, 2007. (Cited on page 13.)

[Chou 1991] Jack CK Chou and M Kamel. *Finding the position and orientation of a sensor on a robot manipulator using quaternions*. The international journal of robotics research, vol. 10, no. 3, pages 240–254, Sage Publications, 1991. (Cited on page 44.)

[Chung 2010] Su Jin Chung, Donghee Kim, Ji Hyun Song, Min Jung Park, Young Sun Kim, Joo Sung Kim, Hyun Chae Jung and In Sung Song. *Efficacy of computed virtual chromoendoscopy on colorectal cancer screening: a prospective, randomized, back-to-back trial of Fuji Intelligent Color Enhancement versus conventional colonoscopy to compare adenoma miss rates*. Gastrointestinal Endoscopy, vol. 72, no. 1, pages 136–142, Elsevier, 2010. (Cited on page 112.)

[Ciuti 2012] Gastone Ciuti, Marco Visentini-Scarzanella, Alessio Dore, Arianna Menciassi, Paolo Dario and Guang Zhong Yang. *Intra-operative monocular 3D reconstruction*

for image-guided navigation in active locomotion capsule endoscopy. In Proceedings of the IEEE RAS and EMBS International Conference on Biomedical Robotics and Biomechatronics, pages 768–774. IEEE, 2012. (Cited on page 71.)

[Cleary 2005] Kevin Cleary, Hui Zhang, Neil Glossop, Elliot Levy, Bradford Wood and Filip Banovac. *Electromagnetic tracking for image-guided abdominal procedures: Overall system and technical issues*. In Proceedings of the 27th Annual International Conference of IEEE Engineering in Medicine and Biology Society (EMBS), pages 6748–6753. IEEE, 2005. (Cited on page 19.)

[Coimbra 2006] Miguel Tavares Coimbra and João Paulo Cunha Cunha. *MPEG-7 Visual Descriptors-Contributions for Automated Feature Extraction in Capsule Endoscopy*. Circuits and Systems for Video Technology, IEEE Transactions on, vol. 16, no. 5, pages 628–637, IEEE, 2006. (Cited on page 75.)

[Constantinescu 2015] Adriana Florentina Constantinescu, Mihaela Ionescu, Ion Rogoveanu, Marius Eugen Ciurea, Costin Teodor Streba, Vlad Florin Iovanescu, Stefan Alexandru Artene and Cristin Constantin Vere. *Analysis of Wireless Capsule Endoscopy Images using Local Binary Patterns*. Applied Medical Informatics, vol. 36, no. 2, pages 31–42, 2015. (Cited on page 74.)

[Conteduca 2012] V. Conteduca, D. Sansonno, G. Ingravallo, S. Marangi, S. Russi, G. Lauletta and F Dammacco. *Barrett's esophagus and esophageal cancer: an overview*. International journal of oncology, vol. 41, no. 2, pages 414–424, Spandidos Publications, 2012. (Cited on page 10.)

[Córdova 2013] Henry Córdova, Raúl San José Estépar, Antonio Rodríguez-D'Jesús, Graciela Martínez-Pallí, Pedro Arguis, Cristina Rodríguez de Miguel, Ricard Navarro-Ripoll, Juan M Perdomo, Miriam Cuatrecasas, Josep Llach*et al.* *Comparative study of NOTES alone versus NOTES guided by a new image registration system for navigation in the mediastinum: a study in a porcine model*. Gastrointestinal Endoscopy, vol. 77, no. 1, pages 102–107, Elsevier, 2013. (Cited on page 4.)

[Corpack-Medical] Corpack-Medical. *Cortrak*. http://www.corpakmedsystems.com/cortrak/. (Cited on page 19.)

[Dalal 2005] Navneet Dalal and Bill Triggs. *Histograms of oriented gradients for human detection*. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, pages 886–893. IEEE, 2005. (Cited on page 75.)

[Daniilidis 1999] Konstantinos Daniilidis. *Hand-eye calibration using dual quaternions*. The International Journal of Robotics Research, vol. 18, no. 3, pages 286–298, SAGE Publications, 1999. (Cited on page 44.)

[Datta 2008] Ritendra Datta, Dhiraj Joshi, Jia Li and James Z Wang. *Image retrieval: Ideas, influences, and trends of the new age*. ACM Computing Surveys (CSUR), vol. 40, no. 2, page 5, ACM, 2008. (Cited on page 70.)

[de Sousa 2008] André Morais Correia de Sousa. Analysis of colour and texture features of vital magnification-endoscopy images for computer diagnosis of precancerous and cancer lesions. Master's thesis, Universidade do Porto, 2008. (Cited on page 79.)

[Drozdzal 2010] Michal Drozdzal, Laura Igual, Jordi Vitria, Carolina Malagelada, Fernando Azpiroz and Petia Radeva. *Aligning endoluminal scene sequences in wireless capsule endoscopy*. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pages 117–124. IEEE, 2010. (Cited on pages 71 and 75.)

[Drozdzal 2014] Michal Drozdzal. *Sequential image analysis for computer-aided wireless endoscopy*. PhD thesis, Universitat de Barcelona, 2014. (Cited on page 73.)

[Drozdzal 2015] Michal Drozdzal, Santi Seguí, Petia Radeva, Carolina Malagelada, Fernando Azpiroz and Jordi Vitrià. *Motility bar: A new tool for motility analysis of endoluminal videos*. Computers in Biology and Medicine, Elsevier, 2015. (Cited on pages 78 and 79.)

[Duda 2007] K. Duda, T. Zielinski, R. Fraczek, J. Bulat and M Duplaga. *Localization of Endoscopic Capsule in the GI Tract Based on MPEG-7 Visual Descriptors*. The proceedings of the IEEE International Workshop on Imaging Systems and Techniques, vol. 1, pages 1–4, IEEE, 2007. (Cited on pages 75 and 77.)

[Engelhardt 2010] S Engelhardt, S Ameling, S Wirth and D Paulus. *Features for classification of polyps in colonoscopy*. In CEUR Workshop Proceedings, vol. 574, pages 350–354, 2010. (Cited on page 74.)

[Eskandari 2012] H Eskandari, a Talebpour, M Alizadeh and H Soltanian-Zadeh. *Polyp detection in Wireless Capsule Endoscopy images by using region-based active contour model*. In 19th Iranian Conference of Biomedical Engineering, ICBME 2012, pages 305–308. IEEE, Dec 2012. (Cited on page 73.)

[Evans 2012] John A Evans, Dayna S Early, Norio Fukami, Tamir Ben-Menachem, Vinay Chandrasekhara, Krishnavel V Chathadi, G Anton Decker, Robert D Fanelli, Deborah A Fisher, Kimberly Q Foley*et al. The role of endoscopy in Barrett's esophagus and other premalignant conditions of the esophagus*. Gastrointestinal Endoscopy, vol. 76, no. 6, pages 1087–1094, Elsevier, 2012. (Cited on page 10.)

[Fan 2010] Yichen Fan, M.Q.-H. Meng and Baopu Li. *3D reconstruction of wireless capsule endoscopy images*. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), pages 5149–5152. IEEE, 2010. (Cited on page 71.)

[Fan 2011] Yichen Fan and Max Q.-H Meng. *3D reconstruction of the WCE images by affine SIFT method*. In 2011 9th World Congress on Intelligent Control and Automation, pages 943–947. IEEE, June 2011. (Cited on page 71.)

[Fernández-Esparrach 2010] G Fernández-Esparrach, R San José Estépar, C Guarner-Argente, G Martínez-Pallí, R Navarro, C Rodríguez de Miguel, H Córdova, CC Thompson, AM Lacy, L Donoso*et al. The role of a computed tomography-based image registered navigation system for natural orifice transluminal endoscopic surgery: a comparative study in a porcine model*. Endoscopy, vol. 42, no. 12, page 1096, Thieme Medical Publishers, 2010. (Cited on page 4.)

[Filliat 2007] David Filliat. *A visual bag of words method for interactive qualitative localization and mapping*. In Proceedings of IEEE International Conference on Robotics and Automation (ICRA), pages 3921–3926. IEEE, 2007. (Cited on page 70.)

[Fischer 2005] Gregory S. Fischer and Russell H. Taylor. *Electromagnetic tracker measurement error simulation and tool design*. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pages 73–80. Springer, 2005. (Cited on pages 19 and 38.)

[Fitzgerald 2014] Rebecca C Fitzgerald, Massimiliano di Pietro, Krish Ragunath, Yeng Ang, Jin-Yong Kang, Peter Watson, Nigel Trudgill, Praful Patel, Philip V Kaye, Scott Sanders*et al. British Society of Gastroenterology guidelines on the diagnosis and management of Barrett's oesophagus*. Gut, vol. 63, no. 1, pages 7–42, BMJ Publishing Group Ltd and British Society of Gastroenterology, 2014. (Cited on pages 3 and 10.)

[Förstner 1994] Wolfgang Förstner. *A framework for low level feature extraction*. In Proceedings of the European Conference on Computer Vision (ECCV), pages 383–394. Springer, 1994. (Cited on page 71.)

[Francisco 2015] Sara Isabel Moreira Francisco. Recognition of Cancer using Random Forests as a Bag-of-Words Approach for Gastroenterology. Master's thesis, Universidade do Porto, 2015. (Cited on pages 77 and 78.)

[Frantz 2003] Don D. Frantz, A. D. Wiles, S.E. Leis and S. R. Kirsch. *Accuracy assessment protocols for electromagnetic tracking systems*. Physics in medicine and biology, vol. 48, no. 14, pages 2241–2251, IOP Publishing, 2003. (Cited on pages 19, 25 and 38.)

[Franz 2012] Alfred M Franz, Keno März, Johann Hummel, Wolfgang Birkfellner, Rolf Bendl, Stefan Delorme, H-P Schlemmer, H-P Meinzer and Lena Maier-Hein. *Electromagnetic tracking for US-guided interventions: standardized assessment of a new compact field generator*. International Journal of Computer Assisted Radiology and Surgery, vol. 7, no. 6, pages 813–818, Springer, 2012. (Cited on page 18.)

[Franz 2014] Alfred M Franz, Tamas Haidegger, Wolfgang Birkfellner, Kevin Cleary, Terry M Peters and Lena Maier-Hein. *Electromagnetic Tracking in Medicine-A Review of Technology, Validation, and Applications*. Medical Imaging, IEEE Transactions on, vol. 33, no. 8, pages 1702–1725, IEEE, 2014. (Cited on pages 18 and 19.)

[Fu 2011] Yanan Fu, Mrinal Mandal and Gencheng Guo. *Bleeding region detection in WCE images based on color features and neural network*. In Proceedings of the Midwest Symposium on Circuits and Systems, pages 1–4. IEEE, 2011. (Cited on page 79.)

[Gadermayr 2015] Michael Gadermayr and Andreas Uhl. *Making Texture Descriptors Invariant to Blur*. Technical Report 2015-04, Universität Salzburg, 2015. (Cited on page 74.)

[Gallo 2010] Giovanni Gallo and Eliana Granata. *WCE video segmentation using textons*. In Proceedings of SPIE Medical Imaging, pages 76230X–76230X–7. International Society for Optics and Photonics, 2010. (Cited on page 77.)

[Gallo 2012] Giovanni Gallo and Alessandro Torrisi. *Random forests based WCE frames classification*. In Proceedings of the 25th IEEE International Symposium on Computer-Based Medical Systems (CBMS), pages 1–6. IEEE, 2012. (Cited on page 79.)

[Gauglitz 2011] Steffen Gauglitz, Tobias Höllerer and Matthew Turk. *Evaluation of interest point detectors and feature descriptors for visual tracking*. International journal of computer vision, vol. 94, no. 3, pages 335–360, Springer, 2011. (Cited on page 71.)

[Georgieva 2015] Veska Georgieva, Szilvia Nagy, Andras Horvath and Elena Kamenova. *An Approach for Pit Pattern Recognition in Colonoscopy Images*. Egyptian Computer Science Journal, vol. 39, no. 2, 2015. (Cited on page 73.)

[Giannarou 2009] Stamatia Giannarou, Marco Visentini-Scarzanella and Guang-Zhong Yang. *Affine-invariant anisotropic detector for soft tissue tracking in minimally invasive surgery*. In Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI), pages 1059–1062. IEEE, 2009. (Cited on pages 14 and 71.)

[Godin 1994] Guy Godin, Marc Rioux and Rejean Baribeau. *Three-dimensional registration using range and intensity information*. In Photonics for Industrial Applications, pages 279–290. International Society for Optics and Photonics, 1994. (Cited on page 59.)

[Gono 2004] Kazuhiro Gono, Takashi Obi, Masahiro Yamaguchi, Nagaaki Ohyama, Hirohisa Machida, Yasushi Sano, Shigeaki Yoshida, Yasuo Hamamoto and Takao Endo. *Appearance of enhanced tissue features in narrow-band endoscopic imaging*. Journal of Biomedical Optics, vol. 9, no. 3, pages 568–577, International Society for Optics and Photonics, 2004. (Cited on page 111.)

[Grand 2011] David Justin Grand, Michael A Atalay, John J Cronan, William W Mayo-Smith and Damian E Dupuy. *CT-guided percutaneous lung biopsy: comparison of conventional CT fluoroscopy to CT fluoroscopy with electromagnetic navigation system in 60 consecutive patients*. European Journal of Radiology, vol. 79, no. 2, pages e133–e136, Elsevier, 2011. (Cited on pages 4 and 18.)

[Granger 2006] Sébastien Granger and Xavier Pennec. *Multi-scale EM-ICP: A fast and robust approach for surface registration*. In Proceedings of the European Conference on Computer Vision (ECCV), pages 69–73. Springer, 2006. (Cited on pages 59, 60 and 107.)

[Grasa 2014] Oscar G Grasa, Ernesto Bernal, Santiago Casado, Iñigo Gil and JM Montiel. *Visual SLAM for handheld monocular endoscope*. Medical Imaging, IEEE Transactions on, vol. 33, no. 1, pages 135–146, IEEE, 2014. (Cited on page 70.)

[Häfner 2010] Michael Häfner, Andreas Uhl, Andreas Vecsei, Georg Wimmer and Friedrich Wrba. *Complex Wavelet Transform variants and Discrete Cosine Transform for scale invariance in magnification-endoscopy image classification*. In Proceedings of the 10th IEEE International Conference on Information Technology and Applications in Biomedicine, pages 1–5, 2010. (Cited on page 76.)

[Häfner 2015] Michael Häfner, Toru Tamaki, Shinji Tanaka, Andreas Uhl, Georg Wimmer and Shigeto Yoshida. *Local fractal dimension based approaches for colonic polyp classification*. Medical Image Analysis, vol. 26, no. 1, pages 92–107, Elsevier, Dec 2015. (Cited on page 73.)

[Hai 2009] V U Hai, Tomio Echigo, Ryusuke Sagawa, YAGI Keiko, Masatsugu Shiba, Kazuhide Higuchi, Tetsuo Arakawa and YAGI Yasushi. *Controlling the display of capsule endoscopy video for diagnostic assistance*. IEICE transactions on information and systems, vol. 92, no. 3, pages 512–528, The Institute of Electronics, Information and Communication Engineers, 2009. (Cited on page 79.)

[Halmos 1958] Paul R Halmos. *Finite-dimensional vector spaces. The University Series in Undergraduate Mathematics*, 1958. (Cited on page 74.)

[Hämmerle-Uhl 2012] Jutta Hämmerle-Uhl, Yvonne Höller, Andreas Uhl and Andreas Vécsei. *Endoscope Distortion Correction Does Not (Easily) Improve Mucosa-Based Classification of Celiac Disease*. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), vol. 7512, pages 574–581. Springer, 2012. (Cited on page 73.)

[Haralick 1973] Robert M Haralick, Karthikeyan Shanmugam and Its' Hak Dinstein. *Textural features for image classification*. Systems, Man and Cybernetics, IEEE Transactions on, no. 6, pages 610–621, IEEE, 1973. (Cited on page 75.)

[Haringsma 2001] Jelle Haringsma, Guido NJ Tytgat, Hiroyuki Yano, Hiroyasu Iishi, Masaharu Tatsuta, Tatsuo Ogihara, Haruo Watanabe, Nobuhiro Sato, Norman Marcon, Brian C Wilson*et al. Autofluorescence endoscopy: feasibility of detection of GI neoplasms unapparent to white light endoscopy with an evolving technology*. Gastrointestinal Endoscopy, vol. 53, no. 6, pages 642–650, Elsevier, 2001. (Cited on page 111.)

[Harris 1988] Chris Harris and Mike Stephens. *A combined corner and edge detector*. In Alvey vision conference, vol. 15, page 50. Citeseer, 1988. (Cited on page 71.)

[He 2005] Xiaofei He, Shuicheng Yan, Yuxiao Hu, Partha Niyogi and Hong-Jiang Zhang. *Face recognition using Laplacianfaces*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 27, no. 3, pages 328–340, IEEE, 2005. (Cited on page 15.)

[Hirota 2006] William K Hirota, Marc J Zuckerman, Douglas G Adler, Raquel E Davila, James Egan, Jonathan A Leighton, Waqar A Qureshi, Elizabeth Rajan, Robert Fanelli, Jo Wheeler-Harbaugh*et al. ASGE guideline: the role of endoscopy in the surveillance of premalignant conditions of the upper GI tract*. Gastrointestinal Endoscopy, vol. 63, no. 4, pages 570–580, Elsevier, 2006. (Cited on page 4.)

[Horn 1987] Berthold KP Horn. *Closed-form solution of absolute orientation using unit quaternions*. JOSA A, vol. 4, no. 4, pages 629–642, Optical Society of America, 1987. (Cited on page 60.)

[Hostettler 2007] A Hostettler, SA Nicolau, L Soler and Y Remond. *Toward and accurate real-time simulation of internal organ motions during free breathing from skin motion tracking and an a priori knowledge of the diaphragm motion*. In Proceedings of Computer-Assisted Radiology and Surgery, pages S100–S102, 2007. (Cited on page 53.)

[Hu 2015] Yifan Hu, Bowen Song, Perry J Pickhardt and Zhengrong Liang. *Distance weighted 'inside disc' classifier for computer-aided diagnosis of colonic polyps*. In Lubomir M. Hadjiiski and Georgia D Tourassi, editeurs, SPIE Medical Imaging, page 94140R. International Society for Optics and Photonics, Mar 2015. (Cited on page 78.)

[Huang 1986] TS Huang, SD Blostein and EA Margerum. *Least-squares estimation of motion parameters from 3-D point correspondences*. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol. 10, pages 112–115. IEEE, 1986. (Cited on page 113.)

[Iakovidis 2005] Dimitris K Iakovidis, Dimitris E Maroulis, Stavros Karkanis and A Brokos. *A Comparative Study of Texture Features for the Discrimination of Gastric Polyps in Endoscopic Video*. In Proceedings of the 18th IEEE Symposium on Computer-Based Medical Systems (CBMS), pages 575–580. IEEE, 2005. (Cited on page 76.)

[Iakovidis 2013] Dimitris K. Iakovidis, Evaggelos Spyrou and Dimitris Diamantis. *Efficient homography-based video visualization for wireless capsule endoscopy*. In 13th IEEE International Conference on BioInformatics and BioEngineering, pages 1–4. IEEE, 2013. (Cited on page 75.)

[Iandola 2014] Forrest Iandola, Matt Moskewicz, Sergey Karayev, Ross Girshick, Trevor Darrell and Kurt Keutzer. *Densenet: Implementing efficient convnet descriptor pyramids*. ArXiv e-prints, 2014. (Cited on page 78.)

[Iddan 2000] Gavriel Iddan, Gavriel Meron, Arkady Glukhovsky and Paul Swain. *Wireless capsule endoscopy*. Nature, vol. 405, page 417, 2000. (Cited on page 112.)

[Inoue 2003] H Inoue, H Kashida, S Kudo, M Sasako, T Shimoda, H Watanabe, S Yoshida, M Guelrud, CJ Lightdale, K Wang*et al.* *The Paris endoscopic classification of superficial neoplastic lesions: esophagus, stomach, and colon: November 30 to December 1, 2002*. Gastrointestinal Endoscopy, vol. 58, no. 6 Suppl, pages S3–S43, 2003. (Cited on page 11.)

[Iwahori 2015] Yuji Iwahori, Akira Hattori, Yoshinori Adachi, M K Bhuyan, Robert J Woodham and Kunio Kasugai. *Automatic Detection of Polyp Using Hessian Filter and HOG Features*. Procedia Computer Science, vol. 60, pages 730–739, Elsevier, 2015. (Cited on pages 75 and 79.)

[Jain 1999] Anil K Jain, M Narasimha Murty and Patrick J Flynn. *Data clustering: a review*. ACM Computing Surveys, vol. 31, no. 3, pages 264–323, ACM, 1999. (Cited on page 78.)

[Jan 2009] Flusser Jan, Suk Tomáš and Zitová Barbara. *Moments and Moment Invariants in Pattern Recognition*. Chippenham, UK: Wiley & Sons Ltd, 2009. (Cited on page 73.)

[Jia 2014] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama and Trevor Darrell. *Caffe: Convolutional architecture for fast feature embedding*. In Proceedings of the ACM International Conference on Multimedia, pages 675–678. ACM, 2014. (Cited on page 78.)

[Jolles 2004] Brigitte M Jolles, Patrick Genoud and Pierre Hoffmeyer. *Computer-assisted cup placement techniques in total hip arthroplasty improve accuracy of placement*. Clinical orthopaedics and related research, vol. 426, pages 174–179, Wolters Kluwer Health, 2004. (Cited on page 18.)

[Kalal 2012] Zdenek Kalal, Krystian Mikolajczyk and Jiri Matas. *Tracking-learning-detection*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 34, no. 7, pages 1409–1422, IEEE, 2012. (Cited on page 15.)

[Kalpathy-Cramer 2009] Jayashree Kalpathy-Cramer. Classification and retrieval of endoscopic images from the clinical outcomes research initiative (CORI) collection. Master's thesis, Oregon Health and Science University, 2009. (Cited on page 76.)

[Kaneko 2003] Shun'ichi Kaneko, Tomonori Kondo and Atsushi Miyamoto. *Robust matching of 3D contours using iterative closest point algorithm improved by M-estimation*. Pattern Recognition, vol. 36, no. 9, pages 2041–2047, Elsevier, 2003. (Cited on page 59.)

[Karargyris 2009] Alexandros Karargyris and Nikolaos Bourbakis. *Identification of ulcers in wireless capsule endoscopy videos*. In Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI), pages 554–557. IEEE, 2009. (Cited on page 79.)

[Karkanis 2003] S.A. Karkanis, D.K. Iakovidis, D.E. Maroulis, D.A. Karras and M Tzivras. *Computer-aided tumor detection in endoscopic video using color wavelet features*. Information Technology in Biomedicine, IEEE Transactions on, vol. 7, no. 3, pages 141–152, IEEE, 2003. (Cited on page 76.)

[Karlen 1998] P Karlen, D Kornfeld, O Broström, R Löfberg, PG Persson and A Ekbom. *Is colonoscopic surveillance reducing colorectal cancer mortality in ulcerative colitis? A population based case control study*. Gut, vol. 42, no. 5, pages 711–714, BMJ Publishing Group Ltd and British Society of Gastroenterology, 1998. (Cited on page 3.)

[Ke 2004] Yan Ke and Rahul Sukthankar. *PCA-SIFT: A more distinctive representation for local image descriptors*. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2, pages II–506. IEEE, 2004. (Cited on page 73.)

[Khan 1996] Gul N. Khan and Duncan F Gillies. *Vision based navigation system for an endoscope*. Image and Vision Computing, vol. 14, no. 10, pages 763–772, Elsevier, 1996. (Cited on page 79.)

[Khun 2009] Poh Chee Khun, Zhang Zhuo, Liang Zi Yang, Li Liyuan and Liu Jiang. *Feature selection and classification for Wireless Capsule Endoscopic frames*. In Proceedings of the International Conference on Biomedical and Pharmaceutical Engineering, pages 1–6. IEEE, 2009. (Cited on page 79.)

[Kiesslich 2003] Ralf Kiesslich, Johannes Fritsch, Martin Holtmann, Heinz H Koehler, Manfred Stolte, Stephan Kanzler, Bernhard Nafe, Michael Jung, Peter R Galle and Markus F Neurath. *Methylene blue-aided chromoendoscopy for the detection of intraepithelial neoplasia and colon cancer in ulcerative colitis*. Gastroenterology, vol. 124, no. 4, pages 880–888, Elsevier, 2003. (Cited on page 111.)

[Kiesslich 2004] Ralf Kiesslich, Juergen Burg, Michael Vieth, Janina Gnaendiger, Meike Enders, Peter Delaney, Adrian Polglase, Wendy McLaren, Daniela Janell, Steven Thomas*et al. Confocal laser endoscopy for diagnosing intraepithelial neoplasias and colorectal cancer in vivo*. Gastroenterology, vol. 127, no. 3, pages 706–713, Elsevier, 2004. (Cited on page 112.)

[Kiesslich 2006] R Kiesslich, L Gossner, M Goetz, A Dahlmann, M Vieth, M Stolte, A Hoffman, M Jung, B Nafe, PR Galle*et al. In vivo histology of Barrett's esophagus and associated neoplasia by confocal laser endomicroscopy*. Clinical Gastroenterology and Hepatology, vol. 4, no. 8, pages 979–987, Elsevier, 2006. (Cited on page 112.)

[Kiesslich 2011] Ralf Kiesslich, Martin Goetz, Arthur Hoffman and Peter Robert Galle. *New imaging techniques and opportunities in endoscopy*. Nature Reviews Gastroenterology and Hepatology, vol. 8, no. 10, pages 547–553, Nature Publishing Group, 2011. (Cited on page 111.)

[Krenitsky 2011] Joe Krenitsky. *Blind bedside placement of feeding tubes: treatment or threat?* Practical Gastroenterology, page 32, 2011. (Cited on page 18.)

[Krig 2014] Scott Krig. Computer vision metrics: Survey, taxonomy, and analysis. Apress, 2014. (Cited on pages 72 and 78.)

[Krizhevsky 2012] Alex Krizhevsky, Ilya Sutskever and Geoffrey E Hinton. *Imagenet classification with deep convolutional neural networks*. In Advances in neural information processing systems, pages 1097–1105. Curran Associates, Inc., 2012. (Cited on page 78.)

[Krücker 2007] Jochen Krücker, Sheng Xu, Neil Glossop, Anand Viswanathan, Jörn Borgert, Heinrich Schulz and Bradford J Wood. *Electromagnetic tracking for thermal ablation and biopsy guidance: clinical evaluation of spatial accuracy*. Journal of Vascular and Interventional Radiology, vol. 18, no. 9, pages 1141–1150, Soc Intervent Radiol, 2007. (Cited on page 19.)

[Kuo 2006] Braden Kuo and Daniela Urma. *Esophagus-anatomy and development*. GI Motility online, Nature Publishing Group, 2006. (Cited on page 56.)

[Kwitt 2008] Roland Kwitt and Andreas Uhl. *Color wavelet cross co-occurrence matrices for endoscopy image classification*. In Proceedings of the 3rd International Symposium on Communications, Control, and Signal Processing (ISCCSP), pages 715–718. IEEE, 2008. (Cited on page 78.)

[Laine 2015] Loren Laine, Tonya Kaltenbach, Alan Barkun, Kenneth R McQuaid, Venkataraman Subramanian and Roy Soetikno. *SCENIC international consensus statement on surveillance and management of dysplasia in inflammatory bowel disease*. Gastroenterology, vol. 148, no. 3, pages 639–651, WB Saunders, 2015. (Cited on pages 4 and 11.)

[Lazebnik 2006] Svetlana Lazebnik, Cordelia Schmid and Jean Ponce. *Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories*. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2, pages 2169–2178. IEEE, 2006. (Cited on pages 70, 75, 80 and 85.)

[Leong 2012] Steven Leong, Hong Ju, Henry Marshall, Rayleen Bowman, Ian Yang, Ann-Maree Ree, Cathy Saxon and Kwun M Fong. *Electromagnetic navigation bronchoscopy: a descriptive analysis*. Journal of Thoracic Disease, vol. 4, no. 2, page 173, Pioneer Bioscience Publishing Company, 2012. (Cited on pages 4 and 18.)

[Levine 1993] DOUGLAS S Levine, RODGER C Haggitt, PATRICIA L Blount, PETER S Rabinovitch, VALERIE W Rusch and BRIAN J Reid. *An endoscopic biopsy protocol can differentiate high-grade dysplasia from early adenocarcinoma in Barrett's esophagus*. Gastroenterology-BALTIMORE THEN PHILADELPHIA, vol. 105, pages 40–40, WB SAUNDERS CO, 1993. (Cited on page 11.)

[Levine 2000] Douglas S Levine, Patricia L Blount, Rebecca E Rudolph and Brian J Reid. *Safety of a systematic endoscopic biopsy protocol in patients with Barrett's esophagus*. The American Journal of Gastroenterology, vol. 95, no. 5, pages 1152–1157, Nature Publishing Group, 2000. (Cited on page 11.)

[Li 2009] Baopu Li and Max Q H Meng. *Computer-based detection of bleeding and ulcer in wireless capsule endoscopy images by chromaticity moments*. Computers in Biology and Medicine, vol. 39, no. 2, pages 141–147, Elsevier, 2009. (Cited on page 79.)

[Li 2014] Baopu Li, Ran Zhou, Can Yang, Max Q.-H. Meng, Guoqing Xu and Chao Hu. *Capsule endoscopy images classification by random forests and ferns*. In Proceedings of the IEEE International Conference on Information Science and Technology, pages 414–417. IEEE, 2014. (Cited on page 79.)

[Lieberman 2012] David A Lieberman, Douglas K Rex, Sidney J Winawer, Francis M Giardiello, David A Johnson and Theodore R Levin. *Guidelines for colonoscopy surveillance after screening and polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer*. Gastroenterology, vol. 143, no. 3, pages 844–857, WB Saunders, 2012. (Cited on pages 12 and 13.)

[Lindeberg 2013] Tony Lindeberg. Scale-space theory in computer vision, vol. 256. Springer Science & Business Media, 2013. (Cited on page 71.)

[Linte 2012] Cristian A Linte, Pencilla Lang, Maryam E Rettmann, Daniel S Cho, David R Holmes III, Richard A Robb and Terry M Peters. *Accuracy considerations in image-guided cardiac interventions: experience and lessons learned*. International Journal of Computer Assisted Radiology and Surgery, vol. 7, no. 1, pages 13–25, Springer, 2012. (Cited on page 18.)

[Liu 2008] Jianfei Liu, Kalpathi Subramanian, Terry Yoo and Robert Van Uitert. *A stable optic-flow based method for tracking colonoscopy images*. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pages 1–8. IEEE, 2008. (Cited on page 16.)

[Liu 2012] Xiaoying Liu, Jia Gu, Yaoqin Xie, Jun Xiong and Wenjian Qin. *A new approach to detecting ulcer and bleeding in Wireless capsule endoscopy images*. In Proceedings of the IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), pages 737–740. IEEE, 2012. (Cited on page 76.)

[Liu 2014] Jiquan Liu, Bin Wang, Weiling Hu, Yun Zong, Jianmin Si and Huilong Duan. *A non-invasive navigation system for retargeting gastroscopic lesions*. Bio-medical materials and engineering, vol. 24, no. 6, pages 2673–2679, 2014. (Cited on page 15.)

[Low 2004] Kok-Lim Low. *Linear least-squares optimization for point-to-plane icp surface registration*. Chapel Hill, University of North Carolina, 2004. (Cited on page 60.)

[Lowe 1999] David G Lowe. *Object recognition from local scale-invariant features*. In Computer vision, 1999. The proceedings of the seventh IEEE international conference on, vol. 2, pages 1150–1157. IEEE, 1999. (Cited on pages 71 and 73.)

[Lowe 2004] David G Lowe. *Distinctive image features from scale-invariant keypoints*. International journal of computer vision, vol. 60, no. 2, pages 91–110, Springer, 2004. (Cited on pages 71 and 73.)

[Lucarini 2013] Gioia Lucarini, Alessandro Tozzi, Carlo Bruni, Alessio Vallesi, Elena Gaggini, Gastone Ciuti and Arianna Menciassi. *Electromagnetic design for capsule endoscope navigation: a preliminary study*. In 3rd joint workshop on new technologies for computer/robot assisted surgery, Verona, Italy: September, 2013. (Cited on page 18.)

[Lugez 2015] Elodie Lugez, Hossein Sadjadi, David R Pichora, Randy E Ellis, Selim G Akl and Gabor Fichtinger. *Electromagnetic tracking in surgical and interventional environments: usability study*. International Journal of Computer Assisted Radiology and Surgery, vol. 10, no. 3, pages 253–262, Springer, 2015. (Cited on page 19.)

[Luo 2014] Xiongbiao Luo and Kazuo Mori. *Robust endoscope motion estimation via an animated particle filter for electromagnetically navigated endoscopy*. Biomedical Engineering, IEEE Transactions on, vol. 61, no. 1, pages 85–95, IEEE, 2014. (Cited on page 53.)

[Lutgens 2008] Maurice WMD Lutgens, Frank P Vleggaar, Marguerite EI Schipper, Pieter CF Stokkers, C Janneke van der Woude, Daan W Hommes, Dirk J de Jong, Gerard Dijkstra, Ad A van Bodegraven, Bas Oldenburg*et al.* *High frequency of early colorectal cancer in inflammatory bowel disease*. Gut, vol. 57, no. 9, pages 1246–1251, BMJ Publishing Group Ltd and British Society of Gastroenterology, 2008. (Cited on page 12.)

[Mackiewicz 2006] Michal Mackiewicz, Jeff Berens, Mark Fisher and Duncan Bell. *Colour and texture based gastrointestinal tissue discrimination*. In Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on, vol. 2, pages II–II. IEEE, 2006. (Cited on page 79.)

[Mäenpää 2003] Topi Mäenpää and Matti Pietikäinen. *Multi-scale binary patterns for texture analysis*. In Image Analysis, pages 885–892. Springer, 2003. (Cited on page 85.)

[Maier-Hein 2012] Lena Maier-Hein, Alfred M Franz, Thiago R dos Santos, Mirko Schmidt, Markus Fangerau, Hans-Peter Meinzer and J Michael Fitzpatrick. *Convergent iterative closest-point algorithm to accomodate anisotropic and inhomogenous localization error*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 34, no. 8, pages 1520–1532, IEEE, 2012. (Cited on pages 59, 60 and 107.)

[Malti 2010] Abed Malti and Joao P. Barreto. *Robust hand-eye calibration for computer aided medical endoscopy*. In Proceedings of IEEE International Conference on

Robotics and Automation (ICRA), pages 5543–5549. IEEE, May 2010. (Cited on page 44.)

[Manivannan 2013] Siyamalan Manivannan, Ruixuan Wang, Emanuele Trucco and Adrian Hood. *Automatic normal-abnormal video frame classification for colonoscopy*. In Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI), pages 644–647. IEEE, 2013. (Cited on page 77.)

[Manivannan 2014] Siyamalan Manivannan, Ruixuan Wang, Maria P Trujillo, Jesus Arbey Hoyos and Emanuele Trucco. *Video-specific SVMs for colonoscopy image classification*. In Proceedings of the Computer-Assisted and Robotic Endoscopy Workshop, MICCAI, vol. 8899, page 11. Springer, 2014. (Cited on pages 73 and 75.)

[Manstad-Hulaas 2012] Frode Manstad-Hulaas, Geir Arne Tangen, Torbjørn Dahl, Toril AN Hernes and Petter Aadahl. *Three-dimensional electromagnetic navigation vs. fluoroscopy for endovascular aneurysm repair: a prospective feasibility study in patients*. Journal of Endovascular Therapy, vol. 19, no. 1, pages 70–78, International Society of Endovascular Specialists 1928 East Highland Avenue,# F104-605, Phoenix, AZ 85016 USA, 2012. (Cited on page 18.)

[Mathew 2015] Mekha Mathew and Varun P Gopi. *Transform based bleeding detection technique for endoscopic images*. In Electronics and Communication Systems (ICECS), 2015 2nd International Conference on, pages 1730–1734. IEEE, 2015. (Cited on page 78.)

[Matsuda 2008] Takahisa Matsuda, Yutaka Saito, Kuang-I Fu, Toshio Uraoka, Nozomu Kobayashi, Takeshi Nakajima, Hisatomo Ikehara, Yumi Mashimo, Tadakazu Shimoda, Yoshitaka Murakami*et al. Does Autofluorescence Imaging Videoendoscopy System Improve the Colonoscopic Polyp Detection Rate: A Pilot Study*. The American Journal of Gastroenterology, vol. 103, no. 8, pages 1926–1932, Nature Publishing Group, 2008. (Cited on page 111.)

[Menciassi 2014] Arianna Menciassi, Gastone Ciuti and Carmela Cavallotti. *Future Developments of Video Capsule Endoscopy: Hardware*. In Video Capsule Endoscopy, pages 543–556. Springer, 2014. (Cited on page 2.)

[Meng 2010] Max Q.-H Meng. *Detection of lymphangiectasia disease from wireless capsule endoscopy images with adaptive threshold*. In 8th World Congress on Intelligent Control and Automation, pages 3088–3093, 2010. (Cited on page 74.)

[Mikolajczyk 2001] Krystian Mikolajczyk and Cordelia Schmid. *Indexing based on scale invariant interest points*. In Proceedings of the 8th IEEE International Conference on Computer Vision (ICCV), vol. 1, pages 525–531. IEEE, 2001. (Cited on page 71.)

[Mikolajczyk 2002] Krystian Mikolajczyk and Cordelia Schmid. *An affine invariant interest point detector*. In Proceedings of the European Conference on Computer Vision (ECCV), pages 128–142. Springer, 2002. (Cited on page 71.)

[Mikolajczyk 2005] Krystian Mikolajczyk and Cordelia Schmid. *A performance evaluation of local descriptors*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 27, no. 10, pages 1615–1630, IEEE, 2005. (Cited on pages 71 and 73.)

[Miyaki 2013] R Miyaki, S Yoshida, S Tanaka, Y Kominami, Y Sanomura, T Matsuo, S Oka, B Raytchev, T Tamaki, T Koide, K Kaneda, M Yoshihara and K Chayama. *Quantitative identification of mucosal gastric cancer under magnifying endoscopy with flexible spectral imaging color enhancement*. Journal of Gastroenterology & Hepatology, vol. 28, no. 5, pages 841–847, Wiley Online Library, 2013. (Cited on page 75.)

[Moravec 1980] Hans P Moravec. *Obstacle avoidance and navigation in the real world by a seeing robot rover*. Technical report, DTIC Document, 1980. (Cited on page 71.)

[Moreels 2007] Pierre Moreels and Pietro Perona. *Evaluation of features detectors and descriptors based on 3d objects*. International Journal of Computer Vision, vol. 73, no. 3, pages 263–284, Springer, 2007. (Cited on page 71.)

[Mori 2007] Kensaku Mori, Daisuke Deguchi, Kazuyoshi Ishitani, Takayuki Kitasaka, Yasuhito Suenaga, Yosihnori Hasegawa, Kazuyoshi Imaizumi and Hirotsugu Takabatake. *Bronchoscope tracking without fiducial markers using ultra-tiny electromagnetic tracking system and its evaluation in different environments*. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pages 644–651. Springer, 2007. (Cited on pages 4 and 18.)

[Mountney 2006] Peter Mountney, Danail Stoyanov, Andrew Davison and Guang-Zhong Yang. *Simultaneous stereoscope localization and soft-tissue mapping for minimal invasive surgery*. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pages 347–354. Springer, 2006. (Cited on page 14.)

[Mountney 2007] Peter Mountney, Benny Lo, Surapa Thiemjarus, Danail Stoyanov and Guang Zhong-Yang. *A probabilistic framework for tracking deformable soft tissue in minimally invasive surgery*. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pages 34–41. Springer, 2007. (Cited on page 14.)

[Mountney 2009] Peter Mountney, Stamatia Giannarou, Daniel Elson and Guang-Zhong Yang. *Optical biopsy mapping for minimally invasive cancer screening*. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pages 483–490. Springer, 2009. (Cited on pages 14 and 31.)

[Myronenko 2010] Andriy Myronenko and Xubo Song. *Point set registration: Coherent point drift*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 32, no. 12, pages 2262–2275, IEEE, 2010. (Cited on page 61.)

[Nawarathna 2013] Ruwan D Nawarathna. *Detection of temporal events and abnormal images for quality analysis in endoscopy videos*. PhD thesis, University of North Texas, 2013. (Cited on page 77.)

[NDI] NDI. *Aurora electromagnetic tracker*. http://www.ndigital.com/. (Cited on page 19.)

[Neumann 2011] Helmut Neumann, Markus F Neurath and Jonas Mudter. *New endoscopic approaches in IBD*. World Journal of Gastroenterology, vol. 17, no. 1, page 63, WJG Press, 2011. (Cited on page 111.)

[Newell 2011] Andrew J Newell and Lewis D Griffin. *Multiscale histogram of oriented gradient descriptors for robust character recognition*. In Proceedings of the International Conference on Document Analysis and Recognition (ICDAR), pages 1085–1089. IEEE, 2011. (Cited on page 85.)

[October 2009] Tessie W October and George E Hardart. *Successful placement of post-pyloric enteral tubes using electromagnetic guidance in critically ill children**. Pediatric Critical Care Medicine, vol. 10, no. 2, pages 196–200, Wolters Kluwer Health, 2009. (Cited on page 18.)

[Ojala 1994] Timo Ojala, Matti Pietikainen and David Harwood. *Performance evaluation of texture measures with classification based on Kullback discrimination of distributions*. In Proceedings of 12th International Conference on Pattern Recognition Vol. 1-Conference A: Computer Vision & Image Processing, vol. 1, pages 582–585, Oct 1994. (Cited on page 74.)

[Ojala 1996] Timo Ojala, Matti Pietikäinen and David Harwood. *A comparative study of texture measures with classification based on featured distributions*. Pattern recognition, vol. 29, no. 1, pages 51–59, Elsevier, 1996. (Cited on page 74.)

[Ojala 2002] Timo Ojala, Matti Pietikäinen and Topi Mäenpää. *Multiresolution gray-scale and rotation invariant texture classification with local binary patterns*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 24, no. 7, pages 971–987, IEEE, 2002. (Cited on pages 74 and 80.)

[Peng 2002] YP Peng, ST Qi, G Zheng, JL Zhao and BH Qiu. *Application of electromagnetic navigation in surgical treatment of intracranial tumors: analysis of 12 cases*. Academic journal of the first medical college of PLA, vol. 22, no. 7, pages 662–662, 2002. (Cited on page 18.)

[Peters 2008] Terry Peters and Kevin Cleary. Image-guided interventions: technology and applications. Springer Science & Business Media, 2008. (Cited on page 19.)

[Polglase 2005] AL Polglase, WJ McLaren, SA Skinner, R Kiesslich, MF Neurath and PM Delaney. *A fluorescence confocal endomicroscope for in vivo microscopy of the upper-and the lower-GI tract*. Gastrointestinal Endoscopy, vol. 62, no. 5, pages 686–695, Elsevier, 2005. (Cited on page 112.)

[Polhemus] Polhemus. *FASTRACK*. http://polhemus.com/. (Cited on page 19.)

[Puerto-Souza 2014] Gustavo A Puerto-Souza, Aaron N Staranowicz, Charreau S Bell, Pietro Valdastri and Gian-Luca Mariottini. *A Comparative Study of Ego-Motion Estimation Algorithms for Teleoperated Robotic Endoscopes*. In Computer-Assisted and Robotic Endoscopy Workshop at MICCAI, pages 64–76. Springer, 2014. (Cited on page 16.)

[Pulli 1999] Kari Pulli. *Multiview registration for large data sets*. In 3-D Digital Imaging and Modeling, 1999. Proceedings. Second International Conference on, pages 160–168. IEEE, 1999. (Cited on page 60.)

[Rajan 2009] P Rajan, M Canto, E Gorospe, A Almario, A Kage, C Winter, G Hager, T Wittenberg and C Münzenmayer. *Automated Diagnosis of Barrett's Esophagus With Endoscopic Images*. In World Congress on Medical Physics and Biomedical Engineering, pages 2189–2192. Springer, 2009. (Cited on pages 78 and 79.)

[Rassweiler 2014] Jens Rassweiler, Marie-Claire Rassweiler, Michael Müller, Hannes Kenngott, Hans-Peter Meinzer, Dogu Teber*et al.* *Surgical navigation in urology: European perspective*. Current opinion in urology, vol. 24, no. 1, pages 81–97, Wolters Kluwer Health, 2014. (Cited on page 4.)

[Reichl 2013] Tobias Reichl, José Gardiazabal and Nassir Navab. *Electromagnetic servoing - a new tracking paradigm*. Medical Imaging, IEEE Transactions on, vol. 32, no. 8, pages 1526–1535, IEEE, 2013. (Cited on page 53.)

[Rembacken 2000] BJ Rembacken, T Fujii, A Cairns, MF Dixon, S Yoshida, DM Chalmers and ATR Axon. *Flat and depressed colonic neoplasms: a prospective study of 1000 colonoscopies in the UK*. The Lancet, vol. 355, no. 9211, pages 1211–1214, Elsevier, 2000. (Cited on page 12.)

[Riaz 2011] Farhan Riaz, Miguel Areia, F Baldaque Silva, Mário Dinis-Ribeiro, P Pimentel Nunes and M Coimbra. *Gabor textons for classification of gastroenterology images*. In Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI), pages 117–120. IEEE, Mar 2011. (Cited on page 77.)

[Rokach 2005] Lior Rokach and Oded Maimon. *Top-down induction of decision trees classifiers-a survey*. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, vol. 35, no. 4, pages 476–487, IEEE, 2005. (Cited on page 79.)

[Rusinkiewicz 2001] Szymon Rusinkiewicz and Marc Levoy. *Efficient variants of the ICP algorithm*. In Proceedings of the 3rd International Conference on 3-D Digital Imaging and Modeling, pages 145–152. IEEE, 2001. (Cited on pages 59 and 60.)

[Rusu 2010] Radu Bogdan Rusu. *Semantic 3d object maps for everyday manipulation in human living environments*. KI-Künstliche Intelligenz, vol. 24, no. 4, pages 345–348, Springer, 2010. (Cited on page 60.)

[Schaffalitzky 2002] Frederik Schaffalitzky and Andrew Zisserman. *Multi-view matching for unordered image sets, or "How do I organize my holiday snaps?"*. In Proceedings of the European Conference on Computer Vision (ECCV), pages 414–431. Springer, 2002. (Cited on page 71.)

[Schmid 2000] Cordelia Schmid, Roger Mohr and Christian Bauckhage. *Evaluation of interest point detectors*. International Journal of computer vision, vol. 37, no. 2, pages 151–172, Springer, 2000. (Cited on page 71.)

[Sermanet 2013] Pierre Sermanet, David Eigen, Xiang Zhang, Michaël Mathieu, Rob Fergus and Yann LeCun. *Overfeat: Integrated recognition, localization and detection using convolutional networks*. arXiv preprint arXiv:1312.6229, 2013. (Cited on page 78.)

[Serpa-andrade 2014] Luis Serpa-andrade, Vladimir Robles-bykbaev, Eduardo Calle-ortiz and Luis Gonzalez-delgado. *A proposal based on color descriptors and Local Binary Patterns Histogram as support tool in presumptive diagnosis of hiatus hernia*. In Power, Electronics and Computing (ROPEC), 2014 IEEE International Autumn Meeting on, pages 1–5. IEEE, 2014. (Cited on page 79.)

[Shan 2009] Caifeng Shan, Shaogang Gong and Peter W McOwan. *Facial expression recognition based on local binary patterns: A comprehensive study*. Image and Vision Computing, vol. 27, no. 6, pages 803–816, Elsevier, 2009. (Cited on page 74.)

[Sharma 2001] Prateek Sharma, Margarita Topalovski, Matthew S Mayo and Allan P Weston. *Methylene blue chromoendoscopy for detection of short-segment Barrett's esophagus*. Gastrointestinal Endoscopy, vol. 54, no. 3, pages 289–293, Elsevier, 2001. (Cited on page 111.)

[Shi 1994] Jianbo Shi and Carlo Tomasi. *Good features to track*. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pages 593–600. IEEE, 1994. (Cited on page 71.)

[Shiu 1989] Yiu Cheung Shiu and S Ahmad. *Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form AX=XB*. Robotics and Automation, IEEE Transactions on, vol. 5, no. 1, pages 16–29, Feb 1989. (Cited on page 44.)

[Sivak 2000] Michael V Sivak, Kenji Kobayashi, Joseph A Izatt, Andrew M Rollins, R Ung-Runyawee, Amitabh Chak, Richard CK Wong, Gerard A Isenberg and Joseph Willis. *High-resolution endoscopic imaging of the GI tract using optical coherence tomography*. Gastrointestinal Endoscopy, vol. 51, no. 4, pages 474–479, Elsevier, 2000. (Cited on page 112.)

[Sliker 2014] Levin J Sliker and Gastone Ciuti. *Flexible and capsule endoscopy for screening, diagnosis and treatment*. Expert review of medical devices, vol. 11, no. 6, pages 649–666, Taylor & Francis, 2014. (Cited on page 2.)

[Soetikno 2008] Roy M Soetikno, Tonya Kaltenbach, Robert V Rouse, Walter Park, Anamika Maheshwari, Tohru Sato, Suzanne Matsui and Shai Friedland. *Prevalence of nonpolypoid (flat and depressed) colorectal neoplasms in asymptomatic and symptomatic adults*. Jama, vol. 299, no. 9, pages 1027–1035, American Medical Association, 2008. (Cited on page 12.)

[Somani 1987] Arun K. Somani, Thomas S. Huang and Steven D. Blostein. *Least-squares fitting of two 3-D point sets*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, no. 5, pages 698–700, IEEE, 1987. (Cited on page 45.)

[Spyrou 2012] Evaggelos Spyrou and Dimitris K Iakovidis. *Homography-based orientation estimation for capsule endoscope tracking*. In 2012 IEEE International Conference on Imaging Systems and Techniques Proceedings, pages 101–105. IEEE, 2012. (Cited on page 75.)

[Spyrou 2015] Evaggelos Spyrou, Dimitris K Iakovidis, Stavros Niafas and Anastasios Koulaouzidis. *Comparative Assessment of Feature Extraction Methods for Visual Odometry in Wireless Capsule Endoscopy*. Computers in Biology and Medicine, 2015. (Cited on pages 71 and 75.)

[Stewart 2003] Charles V Stewart, Chia-Ling Tsai and Badrinath Roysam. *The dual-bootstrap iterative closest point algorithm with application to retinal image registration*. Medical Imaging, IEEE Transactions on, vol. 22, no. 11, pages 1379–1394, IEEE, 2003. (Cited on page 61.)

[Subramanian 2014] Venkataraman Subramanian and Krish Ragunath. *Advanced endoscopic imaging: a review of commercially available technologies*. Clinical Gastroenterology and Hepatology, vol. 12, no. 3, pages 368–376, Elsevier, 2014. (Cited on page 111.)

[Suess 2001] O Suess, Th Kombos, R Kurth, S Suess, S Mularski, S Hammersen and M Brock. *Intracranial image-guided neurosurgery: experience with a new electromagnetic navigation system*. Acta neurochirurgica, vol. 143, no. 9, pages 927–934, Springer, 2001. (Cited on page 18.)

[Tamaki 2013] Toru Tamaki, Junki Yoshimuta, Misato Kawakami, Bisser Raytchev, Kazufumi Kaneda, Shigeto Yoshida, Yoshito Takemura, Keiichi Onji, Rie Miyaki and Shinji Tanaka. *Computer-aided colorectal tumor classification in NBI endoscopy: Using local features*. Medical Image Analysis, vol. 17, no. 1, pages 78–100, Elsevier, 2013. (Cited on page 77.)

[Tan 2010] Xiaoyang Tan and Bill Triggs. *Enhanced local texture feature sets for face recognition under difficult lighting conditions*. Image Processing, IEEE Transactions on, vol. 19, no. 6, pages 1635–1650, IEEE, 2010. (Cited on page 85.)

[Tjoa 2003] Marta P Tjoa and Shankar M Krishnan. *Feature extraction for the analysis of colon status from the endoscopic images*. BioMedical Engineering Online, vol. 2, no. 1, page 9, 2003. (Cited on page 77.)

[Triantafillidis 2009] John K Triantafillidis, Georgios Nasioulas and Paris A Kosmidis. *Colorectal cancer and inflammatory bowel disease: epidemiology, risk factors, mechanisms of carcinogenesis and prevention strategies*. Anticancer research, vol. 29, no. 7, pages 2727–2737, International Institute of Anticancer Research, 2009. (Cited on page 13.)

[Tsai 1988] Roger Y Tsai and Reimar K Lenz. *Real time versatile robotics hand/eye calibration using 3D machine vision*. In Proceedings of IEEE International Conference on Robotics and Automation (ICRA), pages 554–561. IEEE, 1988. (Cited on page 44.)

[Uhl 2011] Andreas Uhl, Andreas Vécsei and Georg Wimmer. *Complex wavelet transform variants in a scale invariant classification of celiac disease*. In Pattern Recognition and Image Analysis, pages 742–749. Springer, 2011. (Cited on page 76.)

[Van De Sande 2010] Koen EA Van De Sande, Theo Gevers and Cees GM Snoek. *Evaluating color descriptors for object and scene recognition*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 32, no. 9, pages 1582–1596, IEEE, 2010. (Cited on page 82.)

[Varian] Varian. *Calypso Extracranial Tracking*. https://www.varian.com/. (Cited on page 19.)

[Vecsei 2008] Andreas Vecsei, T Fuhrmann and Andreas Uhl. *Towards automated diagnosis of celiac disease by computer-assisted classification of duodenal imagery*. Advances in Medical, Signal and Information Processing, 2008. MEDSIP 2008. 4th IET International Conference on, pages 1–4, IET, 2008. (Cited on page 78.)

[Vedaldi 2008] A. Vedaldi and B. Fulkerson. *VLFeat: An Open and Portable Library of Computer Vision Algorithms*. http://www.vlfeat.org/, 2008. (Cited on page 88.)

[Vemuri 2013] Anant S. Vemuri, Stephane A Nicolau, Nicholas Ayache, Jacques Marescaux and Luc Soler. *Inter-Operative Trajectory Registration for Endoluminal Video Synchronization: Application to Biopsy Site Re-localization*. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pages 372–379. Springer, 2013. (Cited on pages 20, 23 and 104.)

[Vemuri 2015a] Anant S. Vemuri, Stephane A Nicolau, Jacques Marescaux, Luc Soler and Nicholas Ayache. *Automatic View-Point Selection for Inter-Operative Endoscopic Surveillance*. In Workshop on Medical Content-based Retrieval for Clinical Decision Support, MICCAI, pages 1–8. HAL, 2015. (Cited on pages 81, 104 and 105.)

[Vemuri 2015b] Anant S. Vemuri, Stephane A Nicolau, Adrien Sportes, Jacques Marescaux, Luc Soler and Nicholas Ayache. *Inter-Operative Biopsy Site Relocalization in Endoluminal Surgery*. Biomedical Engineering, IEEE Transactions on (In Press), pages 1–12, IEEE, 2015. (Cited on pages 23, 37, 104 and 106.)

[Vemuri 2016] Anant S. Vemuri, Stephane A Nicolau, Jacques Marescaux, Luc Soler and Nicholas Ayache. *Improved registration and View-point localization for Inter-Operative Biopsy Site Relocalization*. Biomedical Engineering, IEEE Transactions on (under review), pages 1–14, 2016. (Cited on page 104.)

[Visentini-Scarzanella 2012] Marco Visentini-Scarzanella, Danail Stoyanov and Guang-Zhong Yang. *Metric depth recovery from monocular images using shape-from-shading and specularities*. In Proceedings of 19th IEEE International Conference on Image Processing (ICIP), pages 25–28. IEEE, 2012. (Cited on page 107.)

[Wang 2008a] Hanzi Wang, Daniel Mirota, Gregory Hager and Masaru Ishii. *Anatomical reconstruction from endoscopic images: Toward quantitative endoscopy*. American journal of rhinology, vol. 22, no. 1, page 47, NIH Public Access, 2008. (Cited on page 3.)

[Wang 2008b] Kenneth K. Wang and Richard E. Sampliner. *Updated guidelines 2008 for the diagnosis, surveillance and therapy of Barrett's esophagus*. The American Journal of Gastroenterology, vol. 103, no. 3, pages 788–97, Mar 2008. (Cited on page 11.)

[Wang 2010] Yi Wang, Wallapak Tavanapong, Johnny S. Wong, Junghwan Oh and Piet C De Groen. *Detection of quality visualization of appendiceal orifices using local edge cross-section profile features and near pause detection*. Biomedical Engineering, IEEE Transactions on, vol. 57, no. 3, pages 685–695, IEEE, 2010. (Cited on page 73.)

[Wang 2011] Z. Wang, B. Fan and F Wu. *Local intensity order pattern for feature description*. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), pages 603–610. IEEE, 2011. (Cited on pages 80 and 85.)

[Wang 2013a] Yancheng Wang, Chris Spangler, Bruce L Tai and Albert J Shih. *Positional accuracy and transmitter orientation of the 3D electromagnetic tracking system*. Measurement Science and Technology, vol. 24, no. 10, pages 105–109, IOP Publishing, 2013. (Cited on pages 19 and 38.)

[Wang 2013b] Yize R Wang, John R Cangemi, Edward V Loftus and Michael F Picco. *Rate of early/missed colorectal cancers after colonoscopy in older patients with or without inflammatory bowel disease in the United States*. The American Journal of Gastroenterology, vol. 108, no. 3, pages 444–449, Nature Publishing Group, 2013. (Cited on page 13.)

[Wang 2014a] Bin Wang, Weiling Hu, Jiquan Liu, Jianmin Si and Huilong Duan. *Gastroscopic Image Graph: Application to Noninvasive Multitarget Tracking under Gastroscopy*. Computational and mathematical methods in medicine, vol. 2014, Hindawi Publishing Corporation, 2014. (Cited on pages 15 and 82.)

[Wang 2014b] Shuai Wang, Dongying Tian, Yang Cong, Yunsheng Yang, Yandong Tang and Huaici Zhao. *Automatic gastroscopy video quality assessment*. In Proceedings

of the IEEE International Conference on Robotics and Biomimetics (ROBIO), pages 2709–2714. IEEE, Dec 2014. (Cited on page 79.)

[Wang 2014c] Yi Wang, Wallapak Tavanapong, Johnny Wong, JungHwan Oh and Piet C de Groen. *Part-based multiderivative edge cross-sectional profiles for polyp detection in colonoscopy*. IEEE Journal of Biomedical and Health Informatics, vol. 18, no. 4, pages 1379–1389, IEEE, 2014. (Cited on page 75.)

[Wilson 2006] Emmanuel Wilson. *Accuracy analysis of electromagnetic tracking within medical environments*. Technical report, tech. rep., Georgetown University, Imaging Science and Information Systems Center, Washington, DC, USA, 2006. (Cited on page 19.)

[Yang 2006] Liu Yang and Rong Jin. *Distance metric learning: A comprehensive survey*. Michigan State Universiy, vol. 2, 2006. (Cited on page 78.)

[Yang 2008] Mingqiang Yang, Kidiyo Kpalma and Joseph Ronsin. *A survey of shape feature extraction techniques*. Pattern recognition, pages 43–90, IN-TECH, 2008. (Cited on page 73.)

[Yaniv 2009] Ziv Yaniv, Emmanuel Wilson, David Lindisch and Kevin Cleary. *Electromagnetic tracking in the clinical environment*. Medical Physics, vol. 36, no. 3, pages 876–892, 2009. (Cited on pages 19 and 38.)

[Yaremko 2008] BP Yaremko, TM Guerrero, MF McAleer, MK Bucci, J Noyola-Martinez, LT Nguyen, PA Balter, R Guerra, R Komaki and Z Liao. *Determination of Respiratory Motion for Distal Esophagus Cancer Using Four-Dimensional Computed Tomography*. International Journal of Radiation Oncology * Biology * Physics, vol. 70, no. 1, pages 145–153, 2008. (Cited on page 53.)

[Ye 2013] Menglong Ye, Stamatia Giannarou, Nisha Patel, Julian Teare and Guang-Zhong Yang. *Pathological site retargeting under tissue deformation using geometrical association and tracking*. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pages 67–74. Springer, 2013. (Cited on pages 15 and 31.)

[Ye 2014] Menglong Ye, Edward Johns, Stamatia Giannarou and Guang-Zhong Yang. *Online Scene Association for Endoscopic Navigation*. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pages 316–323. Springer, 2014. (Cited on page 15.)

[Ye 2016] Menglong Ye, Stamatia Giannarou, Alexander Meining and Guang-Zhong Yang. *Online tracking and retargeting with applications to optical biopsy in gastrointestinal endoscopic examinations*. Medical Image Analysis, vol. 30, pages 144 – 157, 2016. (Cited on page 15.)

[Yoshida 2004] Tatsuya Yoshida, Haruhiro Inoue, Shinsuke Usui, Hitoshi Satodate, Norio Fukami and Shin-ei Kudo. *Narrow-band imaging system with magnifying endoscopy*

for superficial esophageal lesions. Gastrointestinal Endoscopy, vol. 59, no. 2, pages 288–295, Elsevier, 2004. (Cited on page 111.)

[Yuan 2015] Yixuan Yuan, Baopu Li and Max Q Meng. *Improved Bag of Feature for Automatic Polyp Detection in Wireless Capsule Endoscopy Images*. Automation Science and Engineering, IEEE Transactions on, vol. PP, no. 99, pages 1–7, 2015. (Cited on page 77.)

[Zeiler 2014] Matthew D Zeiler and Rob Fergus. *Visualizing and understanding convolutional networks*. In Proceedings of the European Conference on Computer Vision (ECCV), pages 818–833. Springer, 2014. (Cited on page 78.)

[Zhang 2004] Dengsheng Zhang and Guojun Lu. *Review of shape representation and description techniques*. Pattern recognition, vol. 37, no. 1, pages 1–19, Elsevier, 2004. (Cited on page 73.)

[Zhang 2009] David Zhang, Guangming Lu, Wei Li, Lei Zhang and Nan Luo. *Palmprint recognition using 3-D information*. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, vol. 39, no. 5, pages 505–519, IEEE, 2009. (Cited on page 57.)

[Zhou 2003] Xiang Sean Zhou and Thomas S Huang. *Relevance feedback in image retrieval: A comprehensive review*. Multimedia systems, vol. 8, no. 6, pages 536–544, Springer, 2003. (Cited on page 70.)

[Zhou 2014] Shangbo Zhou, Xinying Song, Muhammad Abubakar Siddique, Jie Xu and Ping Zhou. *Bleeding Detection in Wireless Capsule Endoscopy Images Based on Binary Feature Vector*. In Proceedings of the 5th International Conference on Intelligent Control and Information Processing (ICICIP), pages 29–33, 2014. (Cited on page 79.)