



Markerless tracking of a fiber-optical probe by endoscopic vision : Towards multispectral enhanced biopsy

Omar Zenteno

► To cite this version:

Omar Zenteno. Markerless tracking of a fiber-optical probe by endoscopic vision : Towards multispectral enhanced biopsy. Signal and Image Processing. Université d'Orléans, 2020. English. NNT : 2020ORLE3181 . tel-03921701

HAL Id: tel-03921701

<https://theses.hal.science/tel-03921701>

Submitted on 4 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

*ÉCOLE DOCTORALE MATHÉMATIQUES, INFORMATIQUE, PHYSIQUE
THÉORIQUE ET INGÉNIERIE DES SYSTÈMES*

LABORATOIRE PRISME

Thèse présentée par :

Omar ZENTENO

soutenue le : 16 Juin 2020

pour obtenir le grade de : **Docteur de l'Université d'Orléans**

Discipline/ Spécialité : **Sciences et Technologies Industrielles**

**Markerless tracking of a fiber-optical probe
by endoscopic vision: Towards multispectral
enhanced biopsy**

Thèse dirigée par :

Sylvie TREUILLET

Maître de Conférences, Université d'Orléans, Laboratoire PRISME

Yves LUCAS

Maître de Conférences, Université d'Orléans, Laboratoire PRISME

RAPPORTEURS :

Su RUAN

Professeur des universités, Université de Rouen, Laboratoire LITIS

Christophe DOIGNON

Professeur des universités, Université de Strasbourg, Laboratoire ICube

EXAMINATEURS :

Mireille GARREAU

Professeur des universités, Université de Rennes, Laboratoire LTSI

Dominique LAMARQUE

Praticien hospitalier, Hôpital Ambroise-Paré, Service d'Hépto-gastro-entérologie et oncologie digestive

Benjamín CASTANEDA

Professeur des universités, Pontificia Universidad Católica del Perú, Laboratoire LIM

Rachid HARBA

Professeur des universités, Université d'Orléans, Laboratoire PRISME

Acknowledgments

I would like to express my thankful gratitude to all the people involved in this work. Firstly, I would like to acknowledge the incredible support, scientific guidance and inspiration provided by my supervisors Sylvie Treuillet and Yves Lucas. I would especially like to thank them for trusting on me and pushing me during these three years to improve my researching skills. Thank you for remembering me that it's possible to find family figures outside of my blood circle. You are one of a kind and I am so lucky that I can count on you!. Also, thanks to all the PRISME team, I have developed both professionally and personally by taking part of such an awesome research group.

To my friends, I blame all of you. Writing this thesis has been an extensive discovery journey but also an exercise in sustained suffering. The casual reader may, perhaps, exempt from excessive guilt, but for those of you who have played the larger role in prolonging my agonies with your encouragement and support, well you know who you are, and you owe me, you owe me big.

To Rania, my loving companion in adventures and misadventures, thank you for your constant support and love. We made it as a team!

Last but not least, to my family who have always supported me throughout my studies and life. To my mother and father who encourage me to be ambitious and pursue my passions and my sister Fiorella who always provides a funny perspective of things. Thank you for sharing valleys and peaks.

Abstract

This work addresses the problem of bi-modal registration to provide the surgeon with an accurate optical biopsy location in endoscopic frames during image guided examination. A multispectral-augmented endoscopic prototype was developed based on a commercial endoscopic system: A fiberscope is inserted into the operating channel and connected to a snapshot multispectral camera (41 bands in the range of 470 to 975 nm), working as a localized optical biopsy for a complementary medical exploration with a much smaller field of view. The registration between modalities is challenging due to the lack of texture in gastric wall which causes direct image-content registration methods to fail. Therefore, we approach the registration problem as a relative pose estimation between two non-homogeneous cameras (endoscope and fiberscope) not rigidly linked. Both cameras can be separately calibrated using a chessboard pattern so their instantaneous relative pose can be estimated by tracking the moving camera (fiberscope). To do this, we take advantage of the fiberscope presence in the endoscopic images and its geometry. Therefore, the pose of the fiberscope is estimated by minimizing the re-projection error of a virtual cylinder using a predicted pose in four steps: (i) the segmentation of the fiberscope is achieved through CNNs models trained on our endoscopic sequences; (ii) a 3D cylinder axis is initialized using the initial segmentation and Plücker’s coordinates; (iii) the pose estimate is optimized by maximizing Jaccard’s similarity index between an automatic segmentation of the fiberscope in the image and the virtual cylinder projection in the endoscopic reference frame; (iv) we use the pose of both cameras to estimate the transformation through direct homography between virtually projected points achieving a markerless tracking. This approach was validated using a virtual framework to realistically simulate the movements of the fiberscope inside the operating channel (i.e, insertion and precession) and in-vivo sequences acquired with our multimodal prototype in a medical center. The final registered images offer the surgeon real time visualization of the area targeted by the fiberscope at the gastric wall and its corresponding spectral data. In an off-line process, it is also possible to locate the hyperspectral biopsy on a panoramic map constructed by image mosaicing techniques from the endoscopic video acquired during examination. In a general way, the proposed method for locating optical biopsy can be extended to any type of imaging, specially in the presence of poorly textured images or when markerless probe tracking is necessary, and could be valuable for patient monitoring.

Contents

List of figures	v
List of tables	x
Introduction	1
1 Optical biopsy and Multimodal gastro-endoscopy	5
1.1 Endoscopic optical imaging systems	6
1.2 Optical biopsy	7
1.3 Multimodal endoscopic imaging	8
1.3.1 Chromoendoscopy	9
1.3.2 Autofluorescence imaging	10
1.3.3 Optical coherence tomography	11
1.3.4 Confocal endomicroscopy	12
1.3.5 Outlook	12
1.4 Towards hyperspectral gastro-chromoendoscopy	13
1.5 Hyperspectral enhanced chromoendoscopy	15
1.6 Conclusion	17
2 MS-augmented endoscopic prototype	19
2.1 System description	19
2.1.1 Endoscope	19
2.1.2 Fiberscope and lightsource	20
2.1.3 Dual camera image splitter	21
2.1.4 Hyperspectral cameras	22
2.2 Data acquisition setup	23
2.3 System calibration	25
2.3.1 Spatial calibration	26
2.3.2 Spectral calibration	33
2.4 Conclusion	35

3	Prototype modeling and virtual simulation framework	37
3.1	Laparoscopic instrument virtual modeling	37
3.2	Virtual framework modeling	38
3.2.1	Notation	38
3.2.2	Projection matrix	38
3.2.3	Cameras joint kinematics	40
3.2.4	Fiberscope simulation and projection	42
3.3	Illustration of the kinematic system deployment	43
3.4	Conclusion	44
4	Markerless fiberscope tracking based on binary mask similarity	49
4.1	Related work	49
4.1.1	Vision-based markerless tracking	50
4.1.2	Biopsy localization through optical probe tracking	52
4.2	Instrument localization and tracking	53
4.2.1	Contour detection and pose estimation	55
4.2.2	Cylinder axis estimation using Plücker coordinates	56
4.2.3	Fiberscope pose estimation by binary mask similarity maximization	58
4.3	Conclusion	61
5	Validation on synthetic and real images	63
5.1	Propagation of error for axis estimation	63
5.2	Propagation of error for borders estimation	66
5.3	Test on synthetic data	67
5.4	Estimated pose error on chessboard pattern images	72
5.5	Propagation of reprojection error by pattern's distance uncertainty	73
5.6	Conclusion	75
6	Clinical application	79
6.1	Segmentation of the fiber-optic instrument on RGB images	79
6.1.1	Acquisition and database	80
6.1.2	Neural network training	81
6.1.3	Neural network deployment	83
6.2	HS and RGB synergy	85
6.2.1	Enhanced endoscopy by image registration	86
6.3	Spectral bands content analysis from the HS images	88
6.4	Biopsy location on the mosaic	89
6.5	Conclusion	93

Conclusions and perspectives	95
Appendices	101
A Introduction (<i>traduction française</i>)	101
B Résumé français	105
B.1 Introduction	105
B.2 Caractérisation du système	106
B.3 Estimation de la pose du fibroscope	108
B.3.1 Segmentation automatique du fibroscope	109
B.3.2 Estimation de l'axe du fibroscope via les coordonnées Plücker . . .	109
B.3.3 Recalage multimodal	111
B.4 Résultats expérimentaux	111
B.4.1 Test de simulation de suivi	111
B.4.2 Tests de recalage sur une mire	113
B.4.3 Biopsie hyperspectrale	114
B.4.4 Localisation sur la mosaïque	115
B.5 Conclusion	117
Bibliography	119

List of Figures

1	EMMIE project block diagram	3
1.1	Different areas of interest for endoscopy in the upper and lower gastroin- testinal tracks, from [1]	6
1.2	High-grade tumor. (A) Endoscopic view. (B) Confocal laser endomi- croscopy view (Optical biopsy). (C) Histopathological pattern. Images extracted from [2]	8
1.3	Paired sample comparison on total correct answers on images taken using various enhance imaging modalities. Images extracted from [3]	11
1.4	Comparison of trans- and cross- sectional biopsies	13
1.5	Electromagnetic spectrum used for novel methods of optical biopsy result- ing in different depths of tissue penetration through the mucosa (graded layers) and submucosa (1 mm depth), adapted from [4]	16
2.1	Concept diagram of the multispectral prototype	20
2.2	Manufacturer specifications of GIF-H185 [5]	20
2.3	Manufacturer specifications of GIF-1TH190 [6]	21
2.4	a) Principle of fiberoptics b) FoV and DoV illustration	21
2.5	a) Xenon light source b) Internal condenser lens diagram c) microFlex m25-2500 fiberscope	22
2.6	Experimental setup: (a) Exterior of the TwinCam system, (b) Interior of the TwinCam system	22
2.7	(a) Position and peak responses of the filters in the mosaic pattern on the CMV2K SSM VIS _{4x4} sensor, (b) Example filter responses of the SSM VIS _{4x4} sensor in the active range of 470-620nm	24
2.8	(a) Position and peak responses of the filters in the mosaic pattern on the CMV2K SSM NIR _{5x5} sensor, (b) Example filter responses of the SSM NIR _{5x5} sensor in the active range of 600-872nm	24
2.9	a) Pinhole camera model and b) Fisheye camera model	27
2.10	Chessboard images used during endoscopic calibration	29

2.11	Reprojection error during endoscopic camera calibration using pin-hole and fish-eye camera models	30
2.12	Comparison of the reprojection error for the 16 bands of the VIS4x4 camera	33
2.13	Fiberscopic calibration: (a) Chessboard images , (b) Reprojection error, (c) Camera positions relative to the pattern	34
3.1	World coordinates to Pixel coordinates pipeline	39
3.2	Simulated fiberscope movements: a) insertion (z-axis), b and c) precession components: yaw (y-axis) and pitch (x-axis) respectively.	40
3.3	a) Kinematic description of the system including the endoscope and fiberscope camera reference frames b) Fiberscope optical axis direction (\hat{l}) definition.	40
3.4	Spatial calibration of the system: a) Fiberscopic image and b) Endoscopic image c) Virtual modeling of the fiberscopic instrument in the endoscopic reference frame (The colored points represent the initial calibration pattern)	43
3.5	a) Point cloud of the modeled cylinder, b) Point cloud projection in 2D . .	44
3.6	Kinematic deployment of the system a) Precession of the fiberscope through its original position's axis, different positions of the fiberscope tip are presented with their corresponding 2D projection	45
3.7	Kinematic deployment of the system a) Insertion of the fiberscope through its original position axis, different positions of the fiberscope tip are presented with their corresponding 2D projection	46
3.8	a) Point cloud of the modeled cylinder, b) point cloud projection in 2D, c) original endoscopic image	47
3.9	a) Point cloud of the modeled cylinder, b) point cloud projection in 2D, c) original endoscopic image	47
4.1	Block diagram of the minimization process	53
4.2	Graphical pipeline of the binary mask comparison process	54
4.3	Fiberscope's apparent contour lines estimation through RANSAC	55
4.4	a) Plücker coordinates ($l, p \times l$) of a line. Here $p \perp$ is the closest point on the line to the origin O b) Plücker representation of the cylinder's axis and its projection into the endoscopic image (Adapted from [7])	57
4.5	Behaviour of objective function using Jaccard, Dice and F-score similarity measures, according to each parameter deviation: a) x_F , b) y_F and c) z_F . Zero corresponds to the convergence objective (i.e. the perfect superposition of the two masks)	60

4.6	Binary mask similarity comparison. Mask A and B are colored and outlined by red and blue respectively.	61
4.7	Binary mask similarity maximization: a) Initial state b) Final state c) Pose parameters evolution vs the number of iteration.	61
5.1	Parametric representation of a line on terms of θ and ρ	64
5.2	Simulated apparent contours on terms of θ from 0 to 5 in steps of 0.1 degrees	64
5.3	Monte Carlo test for 100 random trials : Mean and STD of the angular deviation between \mathbf{l} and \mathbf{l}_{GT} (in degrees) versus noise on θ	65
5.4	Monte Carlo test for 100 random trials : Mean and STD of E_D between $\hat{\mathbf{C}}_F$ and \mathbf{C}_{GT} (in mm) versus noise on θ	65
5.5	a) Noise corrupted virtual instrument b) Instrument's projection	66
5.6	Monte carlo tests for 100 random trials : Mean and STD of the slope measurement error for the instrument borders a) Upper contour line slope b) Lower contour line slope	67
5.7	Simulated helicoidal trajectory mimicking an endoscopic exploration: comparison of estimated poses and ground truth.	68
5.8	Distribution of error metrics according to Jaccard index greater or less than 0.99 with for the four variants: 3DOF-min (left-top), Plücker+3DOF-min (right-top), Tracking without Plücker (left-bottom) and Tracking with Plücker (right-bottom)	70
5.9	Some singular cases to illustrate the limitations of the method.	71
5.10	a) HS image from a chessboard pattern (green) registered to its corresponding RGB image (red), b) Transformed HS image, c) Original RGB image .	73
5.11	a,b,c,d,e,f) HS image from a chessboard pattern (green) registered to its corresponding RGB image (red). The images correspond to samples 1 to 6	74
5.12	Reprojected points and the original points at 3 different patterns positions: a)-2.5mm b)0mm and c)2.5mm)	75
5.13	Re-projection error as a function of estimated distance from plane of interest in a range of ± 5 mm. The graphs correspond to samples 1 to 6 respectively	76
6.1	Example of a frames segmented using DeepLab.	84
6.2	Example of a frame where segmentation fails due to the fibercope's proximity to the gastric wall.	84
6.3	Interface to simultaneously view RGB and HS images	85
6.4	a) NN segmentation b) maximization of similarity c) virtual model projection	86
6.5	Example of a plane projection over a particular relative positions of cameras	87

6.6	a) Original HS image, b) Transformed HS image, c) Original RGB image and d) Multimodal overlay	88
6.7	a) Original HS image, b) Transformed HS image, c) Original RGB image and d) Multimodal overlay	89
6.8	HS-RGB overlay and its corresponding raw reflectance data	90
6.9	HS-RGB overlay and its corresponding reflectance data	91
6.10	Resulting mosaic image and the original images used	92
6.11	HSI localization into the RGB mosaic	93
6.12	HSI overlaid to the RGB mosaic	93
A.1	EMMIE project block diagram	103
B.1	Prototype d'endoscope augmenté d'une sonde hyperspectrale	107
B.2	Mouvements simulés du fibroscope : a) insertion (axe z), b et c) composantes de précession : lacet (axe y) et tangage (axe x)	107
B.3	Estimation de la pose des instruments	108
B.4	Exemple d'un trames segmentées à l'aide de DeepLab	109
B.5	Estimation des lignes de contour apparentes du fibroscope par RANSAC .	110
B.6	Représentation par Plücker de l'axe du cylindre et sa projection dans l'image endoscopique (Adapté de [7])	110
B.7	Evolution de la pose du fibroscope (C_F) par rapport à la vérité terrain simulée sus une trajectoire de mouvement séquentiel de 100 iterations . . .	112
B.8	Distribution des mesures d'erreur selon l'indice Jaccard supérieure ou inférieure à 0,99 avec pour les quatre variantes : 3DOF-min (en haut à gauche), Plücker+3DOF-min (en haut à droite), Tracking without Plücker (en bas à gauche) et Tracking with Plücker (en bas à droite)	114
B.9	a) Image HS d'une mire d'échiquier (en vert) recalée sur l'image RGB (en rouge), b) Image HS transformée, c) Image RGB originale	116
B.10	a) Image HS originale, b) Image HS transformée, c) Image RGB originale et d) Superposition multimodale	116
B.11	Superposition HS-RGB et les données de réflectance correspondantes . . .	117
B.12	Points de données HS dans la mosaïque 2D	117

List of Tables

1.1	Gastrointestinal imaging modalities.	14
2.1	Camera specification.	23
2.2	Endoscopic camera calibration with pin-hole model	29
2.3	Endoscopic camera calibration with fish-eye model	30
2.4	Reprojection error during endoscopic calibration using pin-hole and fish-eye camera model	30
2.5	Fiberscopic results corresponding to the pin-hole camera model	31
2.6	Fiberscopic results corresponding to the fish eye camera model	32
5.1	Error metrics over the helicoidal trajectory (100 positions)	69
5.2	Estimated pose deviation on real images (mm)	72
5.3	Re-projection error as a function of estimated distance from plane of interest (pixels)	73
6.1	Distribution of the Training, Validation and Test datasets	81
6.2	Comparison between network performance in the Test set using Jaccard similarity measure as metric	83
B.1	Comparaison des performances du réseau par mesure de similarité avec l'indice de jacquard pour métrique	109
B.2	Métriques d'erreur sur la trajectoire hélicoïdale (100 positions)	113
B.3	Estimation de l'écart de pose sur des images réelles (mm)	115
B.4	Écart de rétroprojection en fonction de la distance estimée par rapport au plan d'intérêt (pixels)	115

Introduction

Endoscopy is a nonsurgical procedure used to examine a person's interior organs using an optical instrument called endoscope. An endoscope is a long, thin, flexible tube connected to a light source and camera at one end. Visual exploration and targeted biopsies during endoscopy are the current standard for the evaluation of diseases affecting the upper digestive system. However, this approach has a limited ability to detect diseases in their early stages where no pathognomonic visual features are present or when they are predominantly located beneath the visible surface.

The majority of endoscopes currently used in clinical practice can only produce videos acquired under white light and the recent innovations for these devices are limited to increasing the image resolution or adding dye-less chromoendoscopy (DLC) capabilities. The few systems open to exploitation of DLC like Narrow Band Imaging (NBI; Olympus, Japan), Fujinon intelligent color enhancement (FICE; Fujinon, Japan), and I-Scan (Pentax, Japan) [8] are capable of analysing a limited number of light wavelengths (usually 2 or 3). In addition, while these methods enable analysing suspect zones on enhanced white light; they do not detect flat lesions, such as inflammatory gastritis or atrophic sequelae. Moreover, valuable information contained in other bands of the full spectrum is not used. These limitations justify the development of innovative imaging techniques to meet the need for early detection and characterization of esophageal and stomach diseases.

In this thesis we propose a prototype imaging system which enables simultaneous acquisition of a video composed by hyper-spectral (HS) data in addition to RGB. The HS-enhanced RGB images have the potential to provide new information for tissue characterization while the clinician remains in the context of a conventional endoscopic exploration. Numerous publications demonstrate the real interest of the combination of spatial and spectral information for characterizing gastric tissue on a macro or mesoscopic scale [9, 10, 11].

The ANR EMMIE project

Under the framework of the EMMIE (Endoscopie MultiModale pour les lésions inflammatoires de l'Estomac) project funded by the ANR-15-CE17-0015 and the collaboration of our research partners we introduce an HS-augmented endoscopic prototype. The system is centered around a custom gastro-endoscopic probe comprising a fiber bundle boroscope, an optical splitter, hyperspectral cameras and a commercial medical endoscope.

Figure 1 depicts the task distribution and pipeline of the project. Preliminary analyses in an animal model were performed to characterize the gastric tissue of mice after explantation. The findings were presented to the Ambroise Paré Hospital as *a priori* knowledge of the tissue characteristics to develop an acquisition protocol which involved patients referred to the endoscopy unit of the A.P. Hospital for an upper GI endoscopy under general anaesthesia. The conduct of the routine endoscopy remained unchanged and the results of analyses from the samples taken by the Department of Pathological Anatomy of the A.P. Hospital were stored and labeled to serve as reference diagnostic.

The system setup, calibration and acquisition interface was designed in collaboration with the Le2i laboratory from the Université Bourgogne Franche-Comté. The pre-processing, spatial calibration and registration of both modalities was in charge of PRISME laboratory while the spectral calibration and hyperspectral analysis was entrusted to Le2i laboratory. Finally the 2D and 3D mosaicing was carried by the CRAN laboratory using optical flow and structure from motion (SfM) techniques, we (PRISME) complemented these results by locating the registered HS-image in the endoscopic panoramic view.

The simultaneous multimodal acquisition requires geometric alignment in order to correctly exploit the additional information obtained which are of different size and resolution. This is not a trivial task since, due to the aseptic practices, the fiberscope must be sterilized between exams, forcing the physician to remove and reintroduce it into the instrument channel on each examination. In addition the gastric wall texture is full of specular patches and gastric fluid which make difficult the task of finding common features for registration.

After processing the data could be analyzed in two stages to obtain different levels of insight from the physician:

- On-line, providing a diagnostic aid during endoscopy procedures by alerting the physician when the analysis of spectral images detects potentially inflammatory areas.
- Off-line, a more detailed analysis of the spectral reflectance, recognition and reproducibility of endoscopic images through the mapping obtained by mosaicing of images.

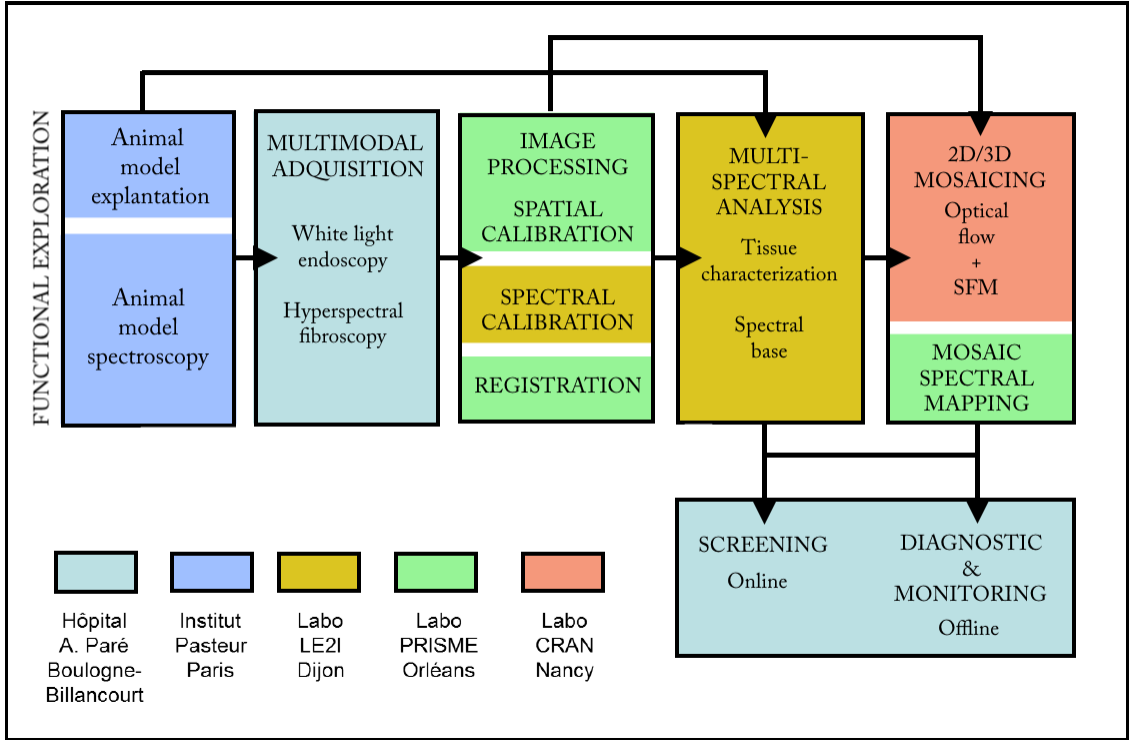


Figure 1: EMMIE project block diagram

Contribution of the work

The three main contributions of this work are: (i) A novel markerless vision tracking approach, (ii) the localization of spectral data during the endoscopic analysis and (iii) the use of the snapshot HS cameras to perform gastric screening.

We used the presence of the fiberoptic probe on the RGB images as a feature for tracking. We tracked its movement based on the similarity measure between a projection of a virtual modeling of the system and an automatic segmentation of the instrument in the endoscopic image. This method is robust against conditions that hinder the efficiency of other methods found in the literature. For example, low FPS acquisition or the impossibility to use markers in the probe.

The project also develops a mosaicing method for mapping the targeted areas, which is an added diagnostic functionality aid. With our method we can localize the spectral data in any RGB frame of the video and enhance the exploration providing the clinicians with the ability to analyse the spectral information of localized areas and deferred time processing. This aspect responds to a legitimate demand from the gastroenterologist and anatomo-pathologist community.

The final prototype designed by the EMMIE team during the project, makes possible to insert a fibroscope, connected to snapshot HS cameras that are sensitive to both visible

and NIR light into this instrument channel, working as a localized HS probe. Something which has never been implemented before and therefore opens a window of opportunities for spectral exploration of gastric tissue.

Organization of manuscript

Each chapter in this thesis contains a literature review which explores prior work and details the rationale behind the specific experiments presented. Then, the presentation of the experimental methods and results is followed by a discussion on the significance of the results and the identification of areas of further work. The main findings of each section are subsequently summarised. The following is an overview of the chapters of the thesis:

Chapter 1, introduces the optical biopsy and multimodal gastro-endoscopy system relation with the current imaging methods present in the literature. Chapter 2, presents the different devices which comprise the MS-augmented endoscope and the system calibration. Chapter 3, details the prototype modeling and the simulation framework developed based on the camera parameters of both boroscopes. Chapter 4, presents the tracking algorithm for the fiberoptic instrument. Chapter 5, assess the validation of the system in synthetic and real images and Chapter 6, the application of our method in clinical data. The document ends with a final discussion and analysis of the further work in the conclusion.

Chapter 1

Optical biopsy and multimodal gastro-endoscopy

Gastrointestinal imaging technologies have been traditionally divided into two categories: classical optical imaging and spectroscopic imaging. Classical optical imaging methods are aimed at recording an image of an area of the sample of interest at one (grayscale) or three specific wavelengths (RGB). White Light Endoscopy (WLE) is currently the primary classic optical method used for wide-area imaging in the gastrointestinal tract; particularly the tissues that line the outer surfaces of organs and blood vessels throughout the body and the inner surfaces of cavities in many internal organs called epithelium. Conventional WLE represents only the first of many possible methods to analyze tissue in-vivo.

The current practice of WLE often requires incisional biopsies with ex-vivo histological analysis and interpretation to complement the findings and make clinical decisions. This method of evaluation creates a significant delay in diagnosis, introduces the possibility of sampling error, and adds to the risk and cost of the procedure. Furthermore, the lack of quantitative parameters during endoscopic inspection often results in significant inter- and intra-observer variability.

In contrast, spectroscopic diagnostic techniques are generally used to obtain an entire spectrum of the tissue under analysis within a wavelength region of interest. Recent advances in optical fibers, light sources, detectors, and molecular biology have stimulated the development of these imaging technologies that promise to significantly improve our ability to visualize and evaluate human epithelium in vivo. These methods, collectively termed “optical biopsy” perform nondestructive inspection of mucosal histopathologic states using tissue’s response to light which providing instantaneous tissue assessment to enhance RGB endoscopy as an alternative to physical biopsy.

In the following sections we will introduce the standard endoscopy optical system, present some imaging modalities commonly used to enhance gastric exploration, their

limitations and how they inspired us to conceptualize the proposed solution.

1.1 Endoscopic optical imaging systems

Endoscopes play an important role in modern medicine. In medical practice different cavities within the body exist which are regularly inspected with an endoscope (e.g., bronchoscopy, cystoscopy, gynoscopy, etc). However, there also exist procedures which are performed through small incisions to reach cavities which are normally closed, such as the abdominal or pelvic cavity (laparoscopy) or organs of the chest (thorascopy).

In this document we will focus on the first group of interventions, particularly in the inspection of the gastrointestinal (GI) tract. When passed through the mouth, an endoscope can be used to examine the esophagus, the stomach, and some of the small intestine (upper GI tract). When passed through the anus, an endoscope can be used to examine the rectum, anus, and the entire large intestine (lower GI tract). Different endoscopes are used for different procedures, and the tubes vary in length and size (Figure 1.1). Some examples of conditions which are known to be premalignant or to increase the risk of cancer in the GI tract are adenomas, Barrett’s esophagus, Crohn’s disease, celiac disease, and a *Helicobacter pylori* infection. The detection of GI bleeding, being a sign of malignancy, is also considered an important finding in gastrointestinal endoscopy.

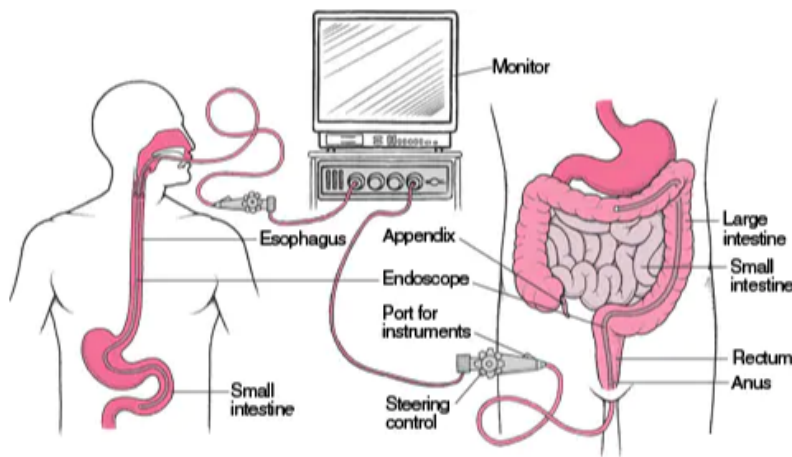


Figure 1.1: Different areas of interest for endoscopy in the upper and lower gastrointestinal tracks, from [1]

The quality of visualization in this modality is a function of resolution and magnification which are determined by the quality of the objective lens and the pixel density of the charge-coupled device (CCD). The CCD chips used in HD endoscopes produce signal images containing 850,000 pixels to >1 million pixels, in contrast to images produced by

standard-definition endoscopes that contain 400,000 pixels. The utility of high-pixel-density CCD chips is enhanced when these systems are used together with magnification technology, because the selected region can be visualized in detail and the increased pixel density of HD CCDs enables magnification without compromising resolution [12].

HD-WLE is widely available today and along with biopsy have become the standard of care in clinical practice worldwide. Yet this approach is troubled by issues related to cumbersome biopsy sampling, errors and cost. Therefore to overcome such adversity, there needs to be evolutionary advancement in terms of diagnosis, which should address these concerns and ideally enhance risk stratification in order to provide timely management in real time.

1.2 Optical biopsy

In this context, optical biopsy presents itself as an attractive alternative to current state of the art endoscopic techniques. The primary goals of using optical biopsy are targeting premalignant mucosa, grading and staging pathological progression and reducing the risk of physical biopsy. Gastric surveillance is challenging because the surface areas of disease involvement, such as that for Barrett's esophagus and ulcerative colitis, are large, ranging from 10 to 100 cm^2 . In addition, deep-tissue penetration of light is needed to detect lesions below the mucosal surface, for example, the presence of abnormal glands beneath the squamous epithelium of the esophagus and to evaluate submucosal tumor invasion to evaluate an endoscopic mucosal resection [13, 14, 4].

Incisional/excisional biopsy and cytology present several disadvantages when compared to optical biopsy. Their main downside is that both require physical extraction and laboratory processing of tissue specimens which incurs in a relatively high cost because of specimen handling; but more importantly, diagnostic information is not available in real time. Moreover, in the context of detecting early neoplastic changes, both can have unacceptable false negative rates, often arising from sampling errors [15, 16]. In contrast, optical biopsy allows earlier detection of pre-malignant lesions and minimize number of target biopsies and frequency of surveillance. Moreover, it can differentiate malignant and benign ulcers and structures during the exploration while quantitatively evaluating tissue areas based on numerical data, limiting inter-observer variability and providing information in the plane perpendicular to the mucosal surface (vertical cross-section). Also, it reduces the risk on exploration in patients with bleeding diathesis

Some examples of common clinical endpoints [17, 18, 19] for optical biopsy in the gastrointestinal exploration could be listed as :

- Detect and localize high-grade dysplasia in Barrett's esophagus

- Identify neoplasia beneath neosquamous esophageal mucosa
- Assess depth of tumor invasion for possible mucosal resection
- Preoperative identification of tumor margins
- Detect and localize dysplasia in the setting of ulcerative colitis
- Distinguish adenomatous from hyperplastic polyps
- Surveillance of polypectomy and mucosal resection site

An illustrative example of new endoscopic imaging technologies are the real-time confocal laser endomicroscopy (CLE) images (Figure 1.2) provided by the Cellvizio™ [2] system (Mauna Kea, Technologies, Paris, France) which allows dynamic microscopic evaluation of cellular architecture and morphology. In this system a confocal microprobe is inserted through the auxiliary channel allowing the physician to examine the cellular structure from areas of interest during the exploration. This is only one of the several multi-modal-imaging techniques which intend to surpass WLE's limitations and enhance the screening procedure.

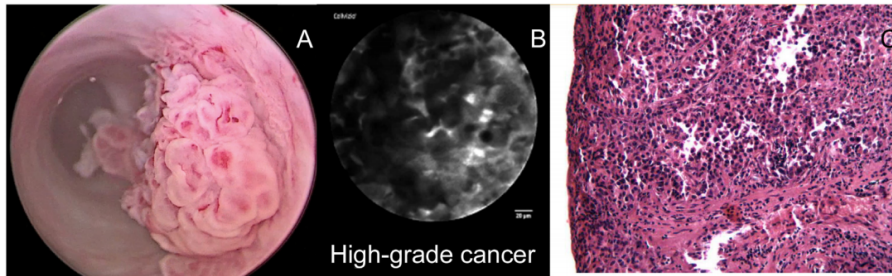


Figure 1.2: High-grade tumor. (A) Endoscopic view. (B) Confocal laser endomicroscopy view (Optical biopsy). (C) Histopathological pattern. Images extracted from [2]

1.3 Multimodal endoscopic imaging

WLE's sensitivity and specificity are operator-dependent and especially poor for small and flat lesions. A multimodal approach is effective because it adds information which may help to detect the presence of pre-malignant and malignant lesions during early stages of their development. Therefore, physicians are able to establish a better diagnosis during the endoscopic procedure, and identify subtle mucosal and structural changes that are related to inflammatory and precancerous lesions [20]. Endoscopic image enhancement technologies are becoming increasingly versatile and complex. A multimodal approach

provides opportunities to visualize normal and abnormal tissues in the gastrointestinal tract, in addition to complementing information derived from conventional WLE. Virtual chromoendoscopy and other advances such as autofluorescence, confocal endomicroscopy and optical coherence tomography (OCT) are becoming increasingly important and will determine the future of optical biopsy diagnostic to an even greater extent [21]. The following subsections make a review of the aforementioned techniques.

1.3.1 Chromoendoscopy

Chromoendoscopy or intravital staining is a dye-spraying technique that accentuates architectural details of the mucosa. Contrast (e.g., indigo carmine) or absorptive (e.g., methylene blue) dyes can be sprayed over mucosal surfaces to highlight abnormal areas and to unmask flat and subtle-appearing lesions.

Virtual Chromoendoscopy can be activated at the press of a button, is essentially comparable to intravital staining (chromoendoscopy) but the light source produces light that is transmitted to the distal end of the endoscope to illuminate the area under inspection and evaluate its reaction. Both procedures highlight specific details on the surface of the mucosa, although intravital staining is more specific because it reacts with various epithelial entities and is still regarded as the gold standard [20].

Virtual Chromoendoscopy includes Narrow Band Imaging (NBI; Olympus, Japan), Fujinon intelligent color enhancement (FICE; Fujinon, Japan), and I-Scan (Pentax, Japan) [8]. NBI applies two optical filters with narrow band centered at 415 nm (blue) and 540 nm (green). Because the maximum light absorption of hemoglobin occurs at these particular wavelengths, blood vessels will appear very dark, providing enhanced visibility and identification of other structures found in the surface. FICE and I-Scan are based on similar physical principle as NBI, but they depend on some computed spectral estimation technology, other than the presence of optical filters [22]. FICE converts images into spectral images with individual wavelengths and reconstructs them to generate fine high contrast images. Then it reconstruct virtual images in real time, by increasing blue component and decreasing red and green components resulting on improved contrast of capillary patterns. I-Scan consists of three types of algorithms: surface enhancement (SE), contrast enhancement (CE), and tone enhancement (TE). SE enhances light-dark contrast by obtaining luminance intensity data for each pixel to allow detailed observation of a mucosal surface structure. CE digitally adds blue color in relatively dark areas to allow detailed observation of subtle irregularities around the surface. TE dissects and analyzes the individual RGB components of a normal image. I-scan is effective to predict neoplasia as precisely as DB chromoendoscopy [8]. However, recent studies shows that FICE could not improve the delineation of diseases, such as ulcers and erosions in Crohn's

disease [23].

Also, several studies have compared NBI and FICE for polyp detection, all of which have shown no difference. The largest of these studies was a prospective, randomized, controlled tandem colonoscopy trial with 1650 subjects that compared HD-WLE with NBI and FICE. In this study, neither NBI nor FICE increased the mean number of adenomas detected per patient compared with WLE (WLE, 0.37 vs NBI, 0.35 and FICE, 0.36; $P=0.591$). The percentage of missed adenomas also did not differ between the 3 groups (20.8% by WLE vs 22.9% by NBI and 26.0% by FICE, $P = .3$) [24]

In a single-center, open, prospective cohort study of 142 consecutive patients undergoing screening or surveillance colonoscopy, NBI and i-SCAN had a significantly improved sensitivity and accuracy compared with white light endoscopy (WLE), for the prediction of adenomas (sensitivity, 88.8% for NBI vs 68.8% for HD-WLE, $P= 0.002$). For i-SCAN, the sensitivity was 94.6% versus 74.3% for HD-WLE; $P=0.001$). There were no significant differences between the NBI and i-SCAN (sensitivity, 88.8% vs 94.6%; specificity, 86.8% vs 86.4%; accuracy, 87.8% vs 90.7%, respectively; $P=0.05$). Additionally, there was good intra- and inter- observer agreement between the NBI and i-SCAN ($k =0.7$) [24].



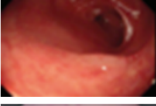








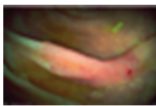



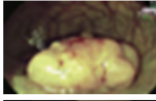

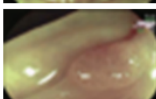
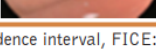
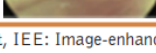
No study has yet directly compared the merits of all endoscopic system in the same patient distribution and it is the users individual preference to use one or the other system. Thus, virtual chromoendoscopy differs in several ways from traditional chromoendoscopy: the normal image can be transformed into an enhanced image, and back again to the white-light image, with a button press; the process dispenses with the laborious and frequently nonuniform application of contrast agent. An example of different modalities and their influence in medical diagnostic is presented on Figure 1.3

1.3.2 Autofluorescence imaging

Autofluorescence imaging (AFI) detects subtle changes in tissue activated by specific short wavelengths of light, leading to the emission of fluorescent light of a longer wavelength. The most abundant target of these approaches is collagen. The large majority of changes identified in the gastrointestinal tract with fluorescence systems rely on the loss of collagen in dysplastic tissue, resulting in the reduction of green fluorescence and the enhancement of red fluorescence. However, the technique has some key limitations, including the inferior resolution of images produced in the AFI mode and a high false-positive rate. The specificity of AFI can be enhanced by combining it with WLE and virtual chromoendoscopy (such as NBI), which jointly form the so-called endoscopic tri-modal imaging procedure [25].

Tri-modal imaging permits the investigator to utilize the advantages of AFI by ‘red-flagging’ fluorescent reddish appearing lesions when screening the gastrointestinal tract.

1.3. MULTIMODAL ENDOSCOPIC IMAGING

IEE	White light	Correct identification	CI 95%	Enhanced imaging	Correct identification	CI 95%	P
NBI		45%	33%-57%		42%	30% to 54%	0.36
		28%	17%-38%		56%	44%-68%	<0.01
FICE		73%	62%-83%		66%	54%-77%	0.08
		51%	38%-62%		61%	49%-72%	0.94
		67%	54%-79%		67%	54%-79%	0.50
		37%	26%-48%		39%	26%-50%	0.58
iScan		84%	76%-93%		77%	67%-87%	0.10
		61%	49%-73%		70%	58%-81%	0.90
		53%	41%-64%		56%	44%-67%	0.65
							

CI: Confidence interval, FICE: Flexible spectral imaging color enhancement, IEE: Image-enhanced endoscopy, NBI: Narrow band imaging

Figure 1.3: Paired sample comparison on total correct answers on images taken using various enhance imaging modalities. Images extracted from [3]

The lesions can then be characterized further by means of high-resolution endoscopy and virtual chromoendoscopy.

1.3.3 Optical coherence tomography

OCT permits high-resolution, cross-sectional imaging of the internal microstructure of biological tissue by measuring back-scattered or back-reflected light using an interferometer with a short coherence length light source [26]. Optical interference between the reference light and the one reflected by the sample occurs only when the optical delays match the coherence length of the light. One arm of the interferometer contains a scanning system while the second arm has a scanning reference path delay to produce a 2D cross sectional image corresponding to the histological layer. Moreover, swept source domain detection is able to provide a very large number of axial samples, as determined by the

speed of the data acquisition systems. The high resolution is able to demonstrate tissue microstructure, but this aspect is achieved at the expense of a decreased depth of imaging (1–2 mm). Thus, OCT permits easy and accurate differentiation of the normal epithelium from abnormal tissue. However, the detection and grading of dysplasia and early cancer remain challenging. OCT is currently not widely spread in clinical use but its potential remain dormant while being a point of interest on gastric research.

1.3.4 Confocal endomicroscopy

A confocal microscope produces images of very shallow depth-of-field. By varying the focal plane of the lens at different depth levels in the sample, it is possible to acquire a series of optical slices corresponding to the focal plane from which a 3D representation of the object can be obtained.

Confocal endomicroscopy allows the histological analysis *in vivo* when performing endoscopy of the mucosal layer at cellular and even subcellular resolution [27]. The histological findings of confocal laser endomicroscopy (CLE) exceed those of conventional histology because cellular interaction can be observed in real time. Two devices are available at the present time: the endoscope-based system (eCLE) and the probe-based system (pCLE). The eCLE works efficiently when integrated into the conventional endoscope (Pentax, Japan). The pCLE is incorporated into the working channels of various standard endoscopes (Cellvizio, Mauna Kea Technologies, France). Notably, the Pentax system is no longer commercially available (probably because of the difficulties to explore and analyse samples at the same time) but is worthy of mention because a number of crucial studies were performed with this system and it is still used.

Confocal imaging can be coupled with the use of intravenous or topical fluorescent dyes to enhance image contrast and resolution. Intravenous fluorescent dye penetrates the epithelium and stains the extracellular matrix while topical acriflavine provides superficial staining, with a predilection for cell nuclei. In the future, the use of fluorescence-labelled probes targeting disease-specific biomarkers will probably enhance the value of CLE in combination with molecular imaging.

1.3.5 Outlook

A large body of research has emphasized the merits and drawbacks of novel imaging modalities, most of these techniques are required to meet set standards before they can be recommended for general use. A summary of the aforementioned imaging modalities is presented in Table 1.1

These technologies are mainly divided in two groups: tran- and cross-sectional imaging

(Figure 1.4). The advantage of trans-sectional imaging over the latter relies in the possibility to directly correlate the acquired information with the endoscopic image. Therefore, tran-sectional modalities may permit the detection of abnormalities at the cellular level without the need for tissue sampling enhancing the already available endoscopic image.

Currently chromoendoscopy is the most used and extended modality over tran-sectional techniques. This is not only due to its commercial availability and clinical insertion but also for its high level of accuracy while screening gastric tissue. However, the analysis is usually limited to a small number of bands and the information should be interpreted by the physician. This does not only lead to inter-operator variability but also leaves a wide range of bands which are not explored. On the other hand, Hyperspectral Imaging (HSI) is an enhanced mode of chromoendoscopy which is able to deliver images containing bio-marker information, and provide assessments of tissue pathophysiology based on a higher range of characteristics from the full spectrum [28]. HSI is increasingly being used for medical diagnosis and image-guided surgery. For instance, HSI has been applied to the diagnosis of hemorrhagic shocks [29], the detection of laryngeal disorders [30], the early detection of dental caries [31], the fast characterization of kidney stone types [32], the assessment of peripheral artery disease [33], and so on. We believe that due to its tran-sectional nature and its superiority when compared to the current chromoendoscopic technics it can effectively enhance the diagnostic if it is properly overlaid to an endoscopic image.

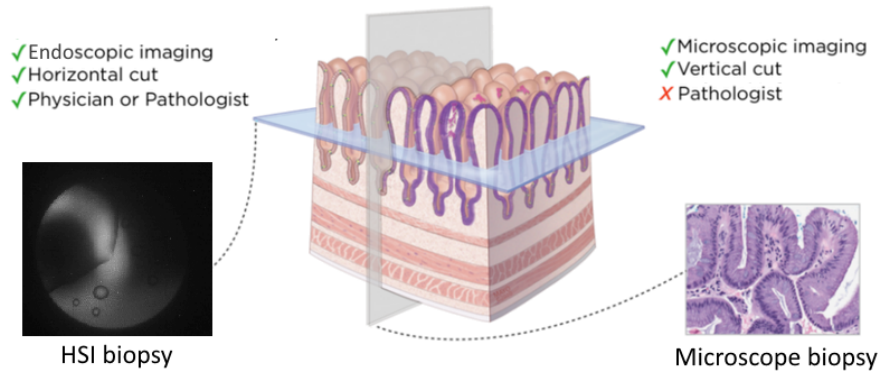


Figure 1.4: Comparison of trans- and cross- sectional biopsies

1.4 Towards hyperspectral gastro-chromoendoscopy

In the gastrointestinal field, most multispectral (MS) and hyperspectral (HS) imaging studies have involved ex-vivo biopsies, resected tumor tissues, or organs such as the skin, tongue, or larynx. Clancy et al. [34] developed a laparoscopic HS system based on a liquid-

Table 1.1: Gastrointestinal imaging modalities.

Imaging modality	Principle	Key features	Resolution and interactions
HD-WLE	An objective lens, linked through a fiber-optic cable carrying light back out of the patient's body to a CCD	External xenon arc for illumination CCD sensor and endoscopic lense	Up to 1M pixels Can be used with optical and electronic magnification
Virtual chromo-endoscopy	Light penetrates tissue to variable depths depending on wavelenght	The use of optical filters or post-processing for particular wavelength enhancement. Diagnosis through improved vessel and mucosal pattern characterization	Normally combined with HD endoscopy Can be toggled to switch from normal to chromo-enhanced imaging
Autofluorescence	Tissue can fluorescence differently when exposed by specific wavelengths of light.	The image is a combination of green and purple colours Dysplastic tissue and inflammation is located in low wavelengths	Lower image resolution High false-positive rate due to the inability to distinguish inflammation and dysplasia
Optical coherence tomography	Uses an interferometer to obtain subsurface and cross-sectional images. (Similar to optical US)	Separates the broad and light source into two different beams. Light is reflected differently from each layer of tissue	Cross-sectional imaging technique for the visualization of mucosal and submucosal structures in microscopic detail Similar resolution to HF-US
Confocal endomicroscopy	A low power laser is focused at a single point in a defined microscopic field, then the same lens is used as the condenser	In vivo histology during ongoing endoscopy Intravenous or topical fluorescent dyes are needed to improve image contrast and resolution	Resolution is similar to that of conventional fluorecence Can be enfoscope- or probe-based

crystal tunable filter (LCTF). In Martin et al. [35] and [30] an HS system with fluorescence for imaging of the larynx was developed. Also Dohi et al. [36], used a micro Fabry-Perot interference filter placed at the tip of a flexible endoscope to create a wavelength-adjustable spectral endoscope. Nevertheless, this has not been used clinically yet. Galeano et al. [37] and Kiyotoki et al. [38] reported certain differences observed between healthy and pre-cancerous ex-vivo colon tissues. However, the color of the resected sampling tissues differed from what is normally observed in vivo, which suggests that the spectral properties of tissue may change after the resection process. In a recent study Martinez et al. [39] assessed the difference in the in-vivo spectral response of malignant colorectal tumors and normal mucosa. Nevertheless, the acquisition systems used a color filter wheel, which makes temporal registration in different wavelengths a non-trivial task.

Conventional endoscopy observes reflected visible light (400–700 nm) from the mucosal surface, leaving a wide range of bands unattended. Indeed, the light spectrum extends to shorter wavelengths in the ultraviolet (UV) and longer wavelengths in the near-infrared (NIR) that also can be used for endoscopic imaging (Figure 1.5). UV and blue light are absorbed by bio-molecules to produce fluorescence. The visible band is dominated by hemoglobin absorption and thus has relatively short penetration depths, typically $<100\text{ }\mu\text{m}$, useful for imaging. NIR light is much less sensitive to tissue scattering and hemoglobin absorption and thus can usually penetrate $<1000\text{ }\mu\text{m}$ through the mucosa [14, 4]. In addition to conventional imaging methods, point detection techniques (i.e., light scattering and Raman’s spectroscopy) can be used to collect molecular information during endoscopy with optical fiber probes inserted through the working channel of a commercial endoscope. These methods have the potential to be extended to imaging, although geometric localization is required in order to correctly exploit the additional sources of information obtained from these two channels, which are of different size and resolution.

1.5 Hyperspectral enhanced chromoendoscopy

In [40], [41], [42] externally tracked cameras were used to augment the surgeon’s view by fusing preoperative data with the actual endoscopic view. As mentioned before, none of them enhance the endoscopic image with HS information. In the same way that virtual chromoendoscopy (VC) surpasses the limitations of normal chromoendoscopy, we propose to extend the capabilities of VC over a wider range of bands in the NIR and VIS range. This can be achieved due to the majority of endoscopic systems are built-in with an instrument channel into which different tools (i.e., biopsy sampler, clamp tool, etc.) can be inserted. This makes possible to insert a fiberscope, connected to HS cameras using

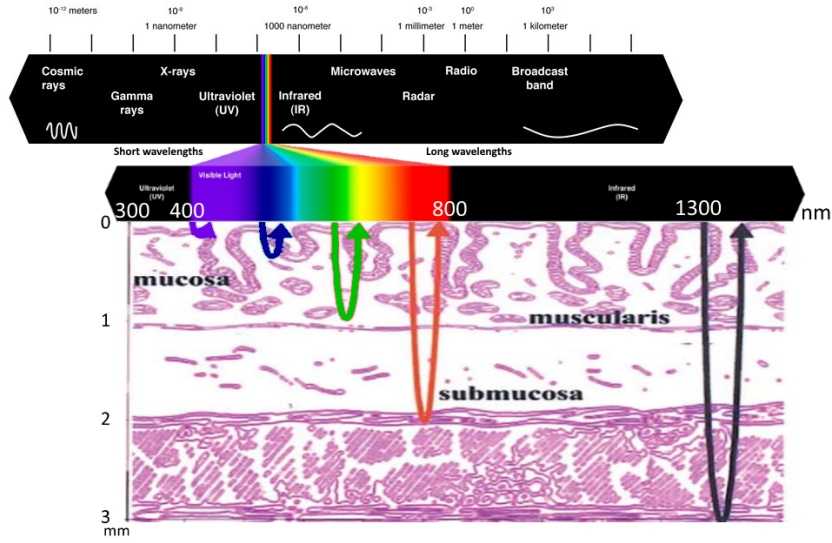


Figure 1.5: Electromagnetic spectrum used for novel methods of optical biopsy resulting in different depths of tissue penetration through the mucosa (graded layers) and submucosa (1 mm depth), adapted from [4]

charge-coupled device (CCD) detectors that are sensitive to both visible and NIR light into this instrument channel, working as a localized HS probe.

However, both modalities must be registered to provide medically relevant information. This is not a trivial task, since due to the aseptic practices during medical and surgical procedures, the fiberscope must be sterilized between exams, forcing the physician to remove and reintroduce it into the instrument channel on each examination. This procedure leads to different relative positions between the fiberscope and endoscope for each new video. In addition, it is impossible to estimate the relation between coordinate systems using calibration patterns before the insertion of the fiberscope because, during the medical examination, the physician introduces first the endoscope into the patient for exploration and then the fiberscope into the instrument channel for localized screening. Thus, registration cannot be performed with conventional approaches (e.g., by matching singular points). Fortunately the end-tip of the fiberscope is visible throughout the endoscopic exploration, then it can then be used as a feature for probe tracking if it is properly detected in a frame by frame basis. By maintaining the relation between the two modalities, the prototype provides physicians with hyperspectral information in small regions of interest overlaid to endoscopic images.

1.6 Conclusion

High-definition white-light endoscopy (WLE) is valuable for early detection of cancer because it provides detailed information about the lesion. The results obtained from WLE are usually complemented by other modalities which allow the physician to surpass its limitations.

We have reviewed the benefits of these complementary modalities in clinical practice, chromoendoscopy and virtual chromoendoscopy allow the detection and characterization of dysplasia in the gastrointestinal tract and identification of changes in the vascular system. Autofluorescence imaging detects subtle changes in tissue after it has been activated by specific wavelengths of light. Confocal endomicroscopy is used for the in vivo diagnosis of pre-malignant lesions and early gastric cancer and Molecular imaging renders pathological changes visible at the cellular level.

With the rapid development of various novel technologies, it seems that the ideal endoscopy system could well be on the horizon. Ideally, such a system would incorporate a ‘red flag’ modality with a lower rate of false-positives based on the collection of optical biopsies. We inspired our self in the NBI and cellvizio systems to propose a prototype including a fiberscope connected to an HSI camera through the working channel of an endoscope which could fill this niche in medical practice .

However, still numerous issues need to be overcome, including standardization of the various classification systems and incorporation of all these devices into a single, easily managed prototype. In the next chapter an overview of the prototype and the different elements it comprises will be presented. Also a detailed description of the devices it comprises and the experimental setup we used to accomplish the realization of a hyperspectral enhanced chromoendoscopy system.

Chapter 2

MS-augmented endoscopic prototype

In this chapter, the multispectral(MS)-augmented imaging prototype developed during EMMIE project and this thesis is presented. The prototype is designed to allow the simultaneous acquisition of two video streams from a targeted area of the stomach (i.e., white light and spectral images). The first modality is captured by an endoscopic device and the latter by a fiberscope introduced through the auxiliary channel of the endoscope. We also present the system's experimental setup and its calibration procedures. Spatial calibration is needed for image distortion correction and to model both cameras in a virtual framework for pose calculation, while spectral calibration is essential for extracting reflectance spectra.

2.1 System description

To obtain an hyperspectral visualization of biological tissues in visible and NIR bands the fiberscope is attached to an optical system composed of two cameras (i.e., VIS and NIR) and a lighting source rigidly fixed with a dual emission image splitter. The system comprises six units: a mercury (Xenon) light source, an endoscope imaging unit, a visible (VIS) range MS camera, a Near Infrared (NIR) range MS camera, a fiberscope and a twin-cam camera splitter. An overview of how the multispectral gastro-intestinal prototype interacts is used in clinical practice is depicted in Figure 2.1.

2.1.1 Endoscope

We used an Olympus endoscopic system (EVIS EXERA III) with two different versions of gastrointestinal videoscope. Both endoscopes (GIF-H185 and GIF-1TH190) combine HDTV and Narrow Band Imaging in a slim-diameter design (less than 10mm). NBI provides greater contrast for interpretation of mucosal morphology and blood vessel, however

2.1. SYSTEM DESCRIPTION

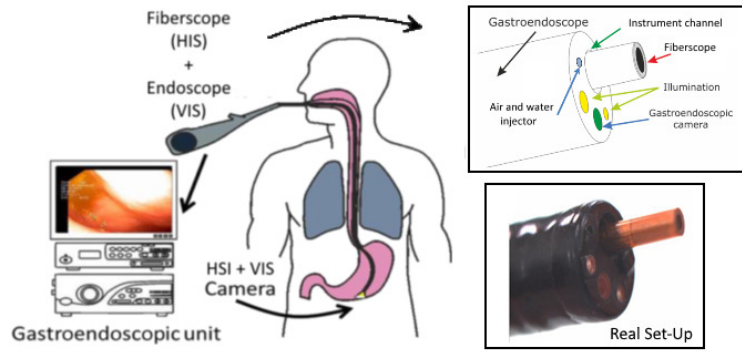


Figure 2.1: Concept diagram of the multispectral prototype

Specifications

Optical system	Field of view	140°
	Direction of view	Forward viewing
	Depth of field	2–100 mm
Insertion section	Distal end outer diameter	9.2 mm
	Distal end enlarged	
Instrument channel	Insertion tube outer diameter	9.2 mm
	Working length	1030 mm
	Channel inner diameter	2.8 mm
	Minimum visible distance	3.0 mm from the distal end
Direction from which endotherapy accessories enter and exit the endoscopic image		



Bending section	Angulation range	Up 210°
		Down 90°
		Right 100°
		Left 100°
Total length	1350 mm	
Compatible	Video system center OLYMPUS CV-190	
EVIS EXERA system	Xenon light source OLYMPUS CLV-190	

Figure 2.2: Manufacturer specifications of GIF-H185 [5]

is limited to only 3 bands of analysis. The endoscopes also have a large field of view (140°), straight direction of view (0°) and close focus which enables close observation of the mucosal surface (2 mm) without the need for electronic magnification. The flexible tip can be controlled by maneuvering wheels in two axes. Further detail for each endoscope are presented in Figure 2.2 and Figure 2.3

2.1.2 Fiberscope and lightsource

We used a microFlex m25-2500 fiberscope from IT Concepts GmBh (Lahnau, Germany). Fiberscopes use optical glass fibers to relay the image. Resolution depends on the number of fibers and their diameter. Each fiber forms a pixel in the final image. The selected optical image conductor is designed originally for industrial applications but has been adapted to our system due to the absence of medical specific fiberscopes in the market. The microFlex contains an optical image guide with an outer diameter of 2.5 mm. made

2.1. SYSTEM DESCRIPTION

Specifications

Optical system	Field of view	140°
	Direction of view	Forward viewing
	Depth of field	2.0-100 mm
Insertion section	Distal end outer diameter	10.0 mm
	Distal end enlarged	
	Air/water nozzle	Up
	Light-guide lens	Left
	Objective lens	Right
	Instrument channel outlet	Down
	Auxiliary water channel	Up
	Insertion tube outer diameter	10.9 mm
	Working length	1,030 mm
Instrument channel	Channel inner diameter	3.7 mm
	Minimum visible distance	3.0 mm from the distal end
	Direction from which EndoTherapy accessories enter and exit the endoscopic image	



Bending section	Angulation range	Up 210°
		Down 90°
		Right 100°
		Left 100°
Total length	1,350 mm	
Compatible	Video system center OLYMPUS CV-190	
EVIS EXERA system	Xenon light source OLYMPUS CLV-190	

Figure 2.3: Manufacturer specifications of GIF-1TH190 [6]

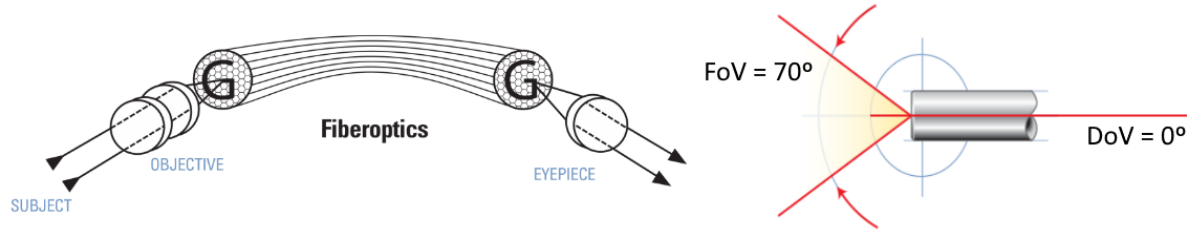


Figure 2.4: a) Principle of fiberoptics b) FoV and DoV illustration

of 2500 highly flexible, thin and parallel single fibres connected directly into the optical system providing a direction of view and field of view of 0° and 70° respectively (Figure 2.4).

A turnable ring found at the main body of the fiberscope allows to adjust the objective's focal point. The light source can be connected to an auxiliary light post found in the the optic system and the eye cup uses a 32mm standard DIN 58105 connector. We used a mercury (Xenon) light source unit from Oriel Instruments® (California, USA), these high power (1000 W) light sources use long lifetime Xe arc lamps to provide high intensity ultraviolet (UV) to near-infrared (NIR) light output with minimum light ripple. An image of the light source system and the fiberscope body is presented in Figure 2.5

2.1.3 Dual camera image splitter

The Cairn TwinCam LS Image Splitter (presented in Figure 2.6) has been designed as a solution for pixel alignment of simultaneous two-channel imaging coupled to a single optical microscopic port without sacrificing field of view for sensors up to 22 mm. It allows to simultaneously acquire the images from the VIS and NIR cameras with an equal

2.1. SYSTEM DESCRIPTION

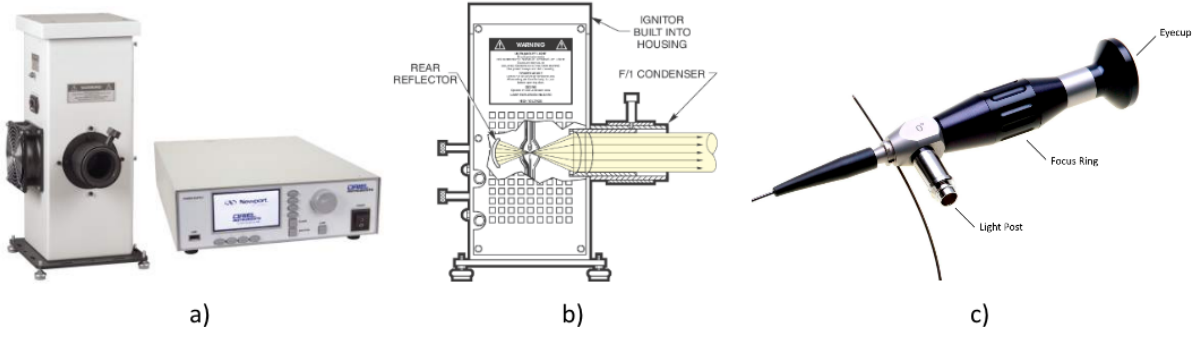


Figure 2.5: a) Xenon light source b) Internal condenser lens diagram c) microFlex m25-2500 fiberscope

distribution of light.

Each cube has fine mechanical X-Y adjustment for pixel overlay or deliberate image offset. The input of the instrument has a variable rectangular aperture enabling the use of cropped sensor mode on both detectors.

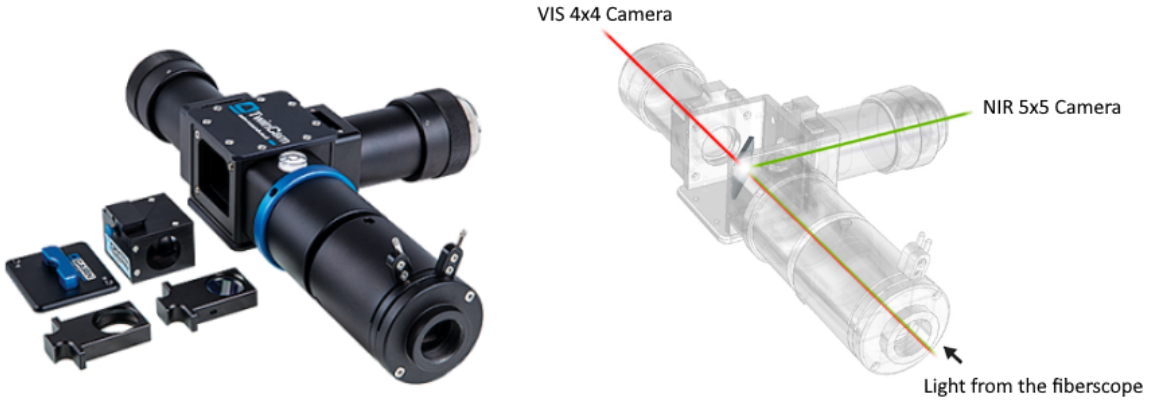


Figure 2.6: Experimental setup: (a) Exterior of the TwinCam system, (b) Interior of the TwinCam system

2.1.4 Hyperspectral cameras

The Hyperspectral imaging (HSI) setup consists of two hyperspectral snapshot cameras from XIMEA (Münster Germany). A snapshot camera has the advantage of acquiring the complete hyperspectral dataset in a single shot, at the cost of spatial resolution due to the larger filter array.

The sensor of both cameras have a resolution of 2048×1088 pixels, at which the filter area is 2048×1024 pixels for the 4x4 HSI camera and 2045×1080 pixels for the 5x5 HSI camera. Both cameras hold a 75 mm f/2.8 lens and are aligned through the TwinCam

splitter system. We use two cameras to increase the number of spectral channels over the complete visual and near-IR range (465 nm to 975 nm) while keeping a reasonable resolution.

The first camera (SM4X4-VIS) holds a 4x4-filter array, resulting in 16 HSI-bands between 460 nm to 630 nm (Figure 2.7). The second camera (SM5X5-NIR) holds a 5x5-filter array, sensitive from 600 nm to 975 nm. Since the filters of the camera show two main peak responses in some pixels of the 5x5-filter array, a 675 nm long-pass filter is placed in front of the camera, resulting in 25 HSI-bands between 675 nm to 975 nm (Figure 2.8). Both cameras record the spectral irradiance $I(x,y,n)$ of a scene through a multispectral image, i.e. a 3-D dataset typically called a datacube or hypercube where n is the number of bands. The nominal capture rate is 170 datacubes per second. However, this value is constrained by the exposure time in practice. Each band has a specific response pattern containing primary and eventually secondary peaks in the sensitive interval of the sensor. Therefore, the band response is not a correct spectral signature as the signal is a combination of the sensor level response and the system level components. To measure a correct spectral signature, it is required to calibrate the signal through spectral correction.

Furthermore, the exposure times for both hyperspectral cameras will differ due to different sensor response characteristics. For the 4x4 HSI camera, the peak response intensity for each band is about 20%, except for band 3 accounting for 10% only. For the 5x5 HSI camera, the peak response intensity for each band is lower by approximately 10%. Hence, the exposure time has to be longer for the 5x5 HSI camera to achieve a decent sensor response.

We mounted the SSM NIR_{5x5} and a SSM VIS_{4x4} Ximea camera to the TwinCam system as slave and master respectively. The detailed specifications of both cameras are presented in Table 2.1.

Table 2.1: Camera specification.

Camera	Resolution	Bands
XIMEA SM5X5-NIR	409x216	25
XIMEA SM4X4-VIS	512x256	16

2.2 Data acquisition setup

The cameras are connected to the PC on two USB 3.0 ports for control and data transfer and one USB 2 port for powering the trigger circuit. We configured one of the cameras as “master” and the other camera as “slave”. The master runs at a configurable frame rate

2.2. DATA ACQUISITION SETUP

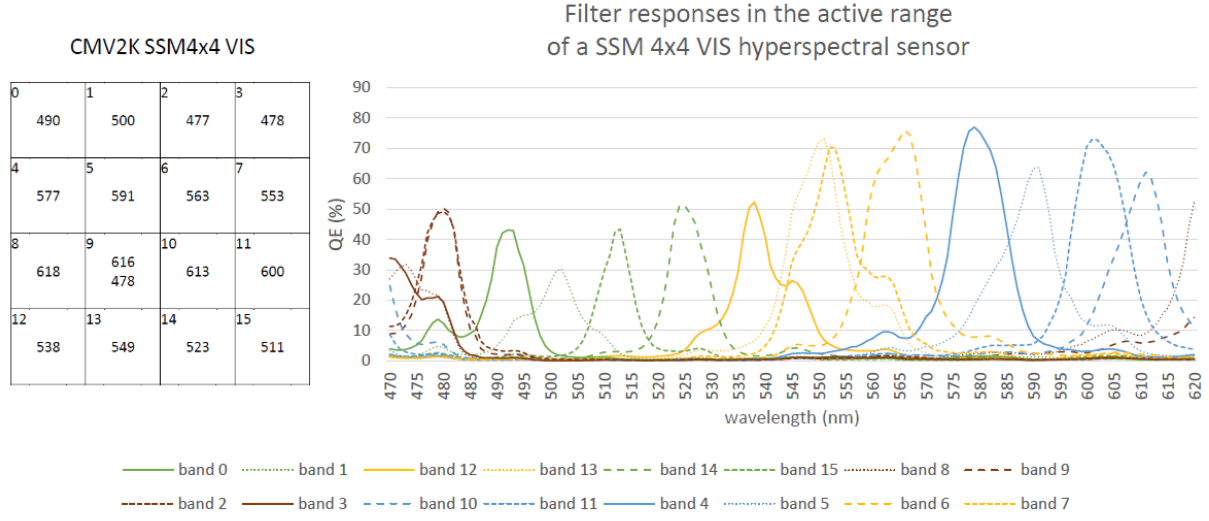


Figure 2.7: (a) Position and peak responses of the filters in the mosaic pattern on the CMV2K SSM VIS_{4x4} sensor, (b) Example filter responses of the SSM VIS_{4x4} sensor in the active range of 470-620nm

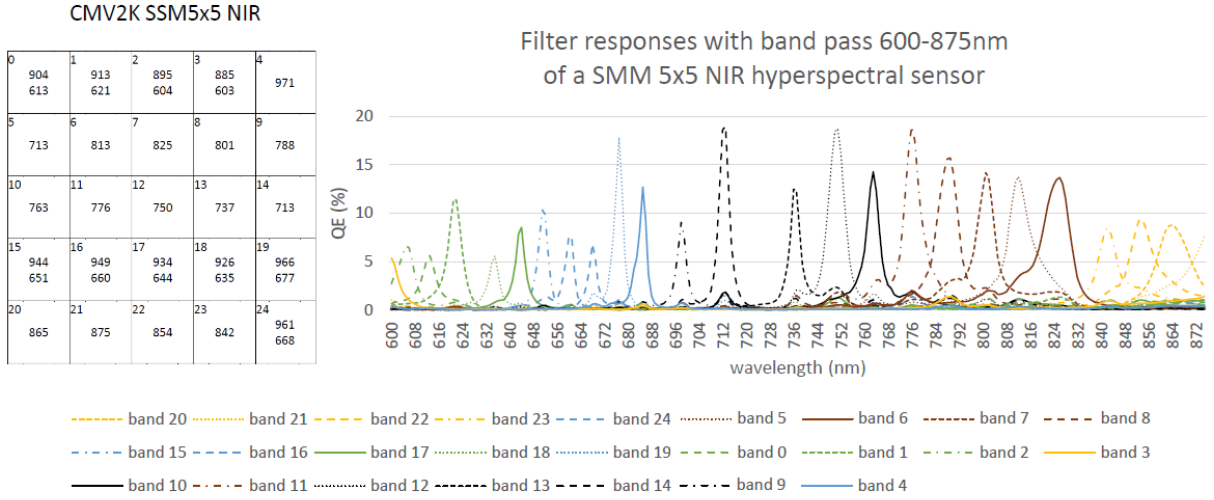


Figure 2.8: (a) Position and peak responses of the filters in the mosaic pattern on the CMV2K SSM NIR_{5x5} sensor, (b) Example filter responses of the SSM NIR_{5x5} sensor in the active range of 600-872nm

and send out a pulse on acquisition start and the slave runs in a triggered mode and starts his acquisition on receiving the pulse from the master. We optically aligned the images of the cameras and acquisition was activated using a custom interface. The fiberscope is connected in the input port of the TwinCam system using an Edmund Optics adaptor and is then inserted in the endoscopic auxiliary channel.

A custom user interface for data acquisition was developed using C# programming

language. The interface allows the user to capture raw MS images and gastroendoscopic video stream simultaneously while the three cameras are connected to the computer. The endoscope is connected through a firewire interface and the MS cameras are connected via USB 3.0 interface. The interface captures up to two images per second due to the high exposure time required for the MS system. All frames are timestamped to be able to associate each MS images to its closest endoscopic image. The association of both modalities allows the combination of textural information in endoscopic images and the spectral signature in MS images. The complete processing pipeline of the images is currently done off-line (i.e., image matching, filtering, spectral analysis).

2.3 System calibration

After setup the optical sensors of the system needs to be calibrated. Due to the particular characteristics of our systems two steps are needed, namely spatial and spectral calibration.

Strong radial distortions are a typical problem in endoscopy. They cause a non-linear geometric distortion of the image, proportional to the radius from the center [43]. They severely affect the perception of relative size and depth, and it is difficult to be able to tell that the imaged scene is planar [44]. The distortion-correction issue has merited the attention of the manufacturers and they have started to focus more efforts in reducing the endoscopic image deformation. Some new proposals diminish the distortion by combining a new optical design with cropping of the image periphery where radial distortions are more noticeable. Unfortunately, the images are far from being geometrically correct perspective; the main disadvantage of this is that cropping decreases the effective field of view.

Spatial calibration provides the necessary distortion coefficients to compensate this effect. In addition, spatial calibration can gives the pose of each camera in the 3D space as long as we have a common reference point. In this way, it is possible to obtain a realistic relative pose between the camera of the fiberscope and that of the endoscope, and thus to obtain the registration relationship between the two images. This operation is necessary for the virtual framework modeling and the pose estimation of the fiberscope as well. Further details of these procedure will be presented in Chapter 3.

On the other hand, the measured raw hyperspectral data cube contains pixel values within each spectral band, which do not necessarily represent the real reflectance of the scanned area. Artifacts of the optics and electronic components influence the measured signals and have to be corrected. Further, as described in Aasen and Bolten [45], differences in the illumination settings have an influence on the calibration process and the

data processing. Finally, the measured signal has to be transformed to a physical system-independent value, e.g. reflectance, to make the data comparable among different settings and useful for signal processing. These aspects make a robust spectral calibration process necessary for potential clinical practice [46].

2.3.1 Spatial calibration

Spatial calibration consists in estimating a computational model for a camera to describe how 3D points are projected onto the image including geometrical distortions.

Various mathematical models for different cameras have been proposed in the literature. Clarke and Fryer [47] gave an excellent overview of the historical development of calibration methods and models up to the end of the 1990s, while more recent extensions are presented by Remondino and Fraser [48] and Boland [49]. A comprehensive nomenclature of cameras computational models is proposed in the books of Sturm [50].

The first thing to consider is whether or not to choose a central projection model. Most models have a single optical center through which all camera rays pass. Non-central camera models do not possess a single optical center (in the case of multi-lens or multi-mirror system). The back-projection is simplified for the central projection, since it only gives the direction of the camera's beam. A second property to take into account when choosing a camera model is the global or local character of a model, according to the nomenclature proposed in [50]. A global model have a small set of parameters each of which influences the projection function over the entire field of view. Conversely, a local model affects the projection function only over a subset of the field of view (portion of the image). These models have usually more parameters than global ones, but have a higher descriptive power. Some models, named discrete, can go as far as defining a projection function for each individual pixel. Such a discrete camera model are sometimes referred to as non-parametric models, which is inappropriate since they have many parameters and needs to be associated with an interpolation scheme. Finally, a third property to consider is the ease with which forward or backward projection can be performed depending on different tasks: forward projection for distortion correction of images and bundle adjustment, back-projection for minimal methods for various structure-from-motion tasks, such as pose and motion estimation.

Considering the characteristics of our snapshot cameras and optical devices, a global central projection model seems appropriate. We chose the global models because they have a lower number of intrinsic parameters and can be used for endoscopy: the classical pinhole model and fish-eye model are in this category. For endoscopy, the pinhole model needs to be enhanced with radial distortions. The classical radial distortion model is written using polynomials, neglecting tangential distortion. Even with only one distortion

coefficient, the classical polynomial distortion model are cubic, making the back-projection cumbersome to write. Such models were applied for endoscope calibration using planar grids [44, 51, 52]. Some fully automatic calibration processes for the endoscopic camera have also been developed specifically for use in the operating room [53, 54]. The main contribution of these studies is the possibility to quickly perform the calibration in the intervention room with a limited number of images. In [54], the proposed calibration process consists of replacing the classic pre-printed checkerboard with a checkerboard displayed on a computer screen in front of which the user can present the endoscopic camera in different orientations. In our particular application, the calibration is performed off-line only once to characterize the entire imaging device (endoscopic and fiber optic channels). In the following sections, we briefly introduce the two most commonly used camera models (i.e., pin-hole and fish-eye) before evaluating them with our prototype.

Camera models

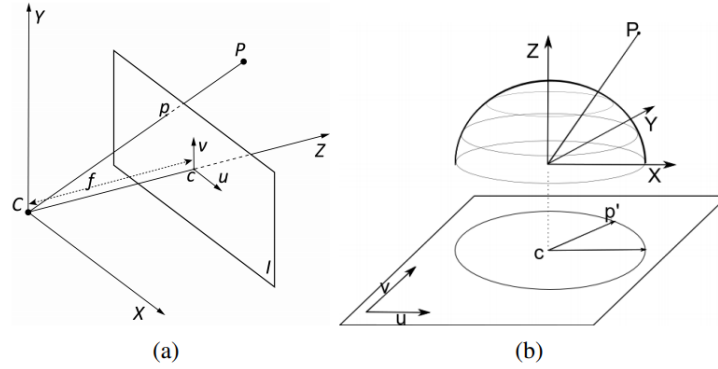


Figure 2.9: a) Pinhole camera model and b) Fisheye camera model

The camera-centered coordinates of a 3D point P are projected into the image plane differently depending on the adopted camera projection model (Figure 2.9).

Pin-hole camera model

The pinhole model, or perspective projection, assumes that all camera rays pass through a single point, called optical center, and that there is a linear relationship between image point position and the direction of the associated camera ray. Therefore, the relationship between the camera-centered coordinates $(X, Y, Z)^T$ of a 3D point P and the image coordinates $(u, v)^T$ of its projection can be expressed via five intrinsic parameters (see Equation 2.1).

$$\begin{pmatrix} u \\ v \end{pmatrix} = \frac{f}{Z} \begin{pmatrix} k_u \cdot X \\ k_v \cdot Y \end{pmatrix} + \begin{pmatrix} u_0 \\ v_0 \end{pmatrix} \quad (2.1)$$

where f is the focal length, k_u and k_v the scale factors for pixels units conversion and $(u_0, v_0)^T$ the image coordinates of the principal point c at the intersection of the optical axis and the image plane to adjust the $(0,0)$ location of the image plane.

To accurately represent a real camera, especially with a short focal length, a polynomial radial distortions model is usually added to this central projection by equations 2.2 and 2.3, generally considering tangential distortions as negligible.

$$u_{distorted} = u(1 + k_1 * r^2 + k_2 * r^4 + k_3 * r^6) \quad (2.2)$$

$$v_{distorted} = v(1 + k_1 * r^2 + k_2 * r^4 + k_3 * r^6) \quad (2.3)$$

where $r^2 = u^2 + v^2$ and k_i the polynomial coefficients.

Typically, one or two coefficients are sufficient for radial distortions of most of the lenses. For severe distortion, such as in a wide-angle lens, we can select 3 coefficients.

Fish eye camera model

The omnidirectional camera model proposed by Scaramuzza [55] is suitable for cameras with field of view (FOV) up to 150 degrees. This projection model uses equations 2.4 and 2.5 to relate a 2D image projection $(u, v)^T$ to a 3D point $(X, Y, Z)^T$.

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \lambda \begin{pmatrix} u \\ v \\ a_0 + a_2 r^2 + a_3 r^3 + a_4 r^4 \end{pmatrix} \quad (2.4)$$

$$\begin{pmatrix} u'' \\ v'' \end{pmatrix} = D \begin{pmatrix} u \\ v \end{pmatrix} + \begin{pmatrix} u_0 \\ v_0 \end{pmatrix}, \quad \text{with} \quad D = \begin{pmatrix} c & d \\ e & 1 \end{pmatrix} \quad (2.5)$$

where λ is a scalar factor, $r^2 = u^2 + v^2$ and a_i the polynomial coefficients described by Scaramuzza ($a_1 = 0$). A stretch matrix D is also introduced to relate the real distorted coordinates (u'', v'') to the ideal distorted coordinates (u, v) and compensate the sensor-to-lens misalignment and the distortion center c .

Endoscopic channel

For the endoscopic camera calibration we used a 15x17 chessboard pattern with a grid resolution of 2 mm. The acquisition of the pattern was performed from different points of view and we selected the best focused images trying to fill all the 3D space described

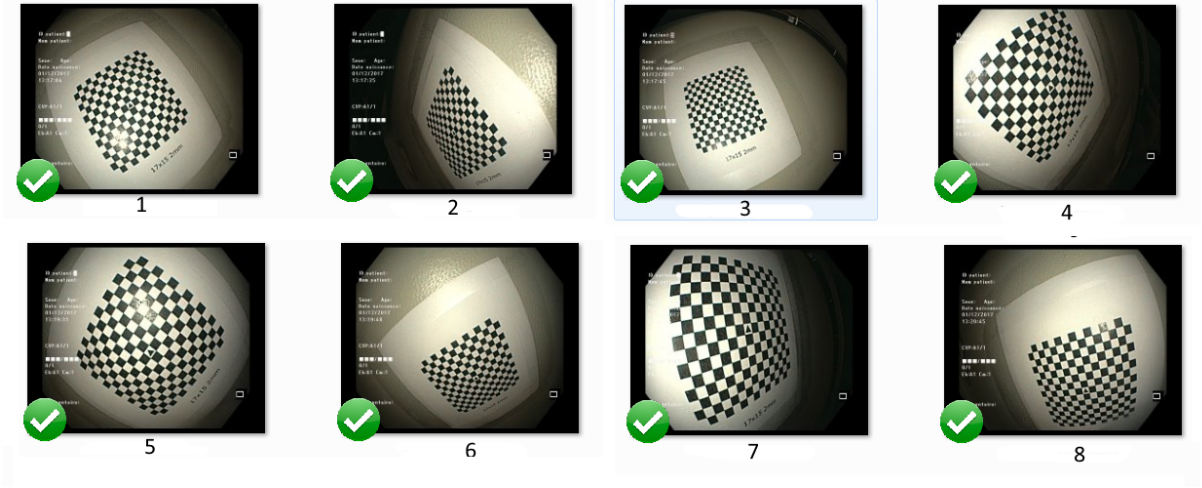


Figure 2.10: Chessboard images used during endoscopic calibration

Table 2.2: Endoscopic camera calibration with pin-hole model

	X	Y
Focal length	362.97	397.30
Principal point	357.08	278.86

	k_1	k_2	k_3
Radial distortion	-0.4213	0.1930	-0.0461

by the camera during a clinical exam. The intrinsic parameters estimates are presented in Table 2.2 and 2.3 for the pin-hole and fish-eye model respectively. The set of images used for the calibration are presented in Figure 2.10.

For the endoscopic calibration we used a chessboard pattern grid of 15x17 squares of 2mm each. The capture of the pattern was performed from different points of view. We selected the best focused images trying to fill all the 3D space described by the camera during a clinical exam. The results for the camera calibration using the pin-hole and fish-eye models are presented in Table 2.2 and Table 2.3 respectively. Table 2.4 presents the comparison between the reprojection error obtained with both models. The different images used for the calibration and their corresponding reprojection errors are presented in Figure 2.10 and Figure 2.11

Fiberscopic channel

For the fiberscopic bundle calibration we used a pattern grid of 10x9 squares of 1mm each. We used the same experimental setup and calibration method than while calibrating the

2.3. SYSTEM CALIBRATION

Table 2.3: Endoscopic camera calibration with fish-eye model

	X	Y		
Distortion center (pixels)	357.3976	281.0434		
	a_0	a_2	a_4	a_6
Mapping coef (pixels)	397.6535	-0.0012	8.78E-07	-4.88E-09
	c	d	e	
Stretch matrix	0.9140	0.0341	-0.02970	

Table 2.4: Reprojection error during endoscopic calibration using pin-hole and fish-eye camera model

	Mean	STD	Max
Pin-hole	0.434	0.111	0.606
Fisheye	0.420	0.119	0.603

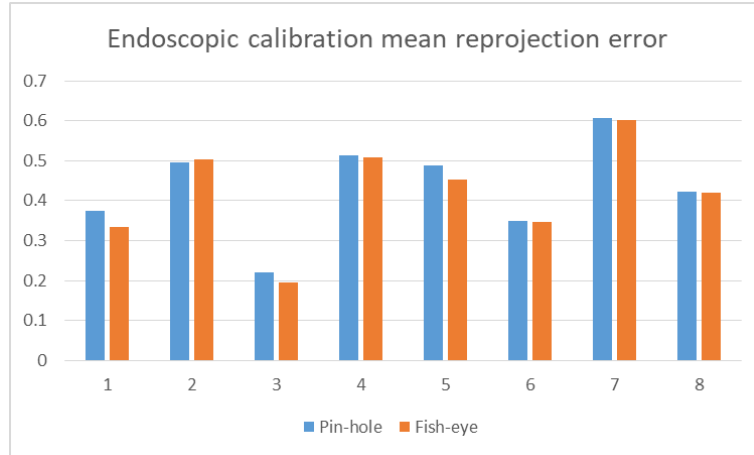


Figure 2.11: Reprojection error during endoscopic camera calibration using pin-hole and fish-eye camera models

endoscope but this time the analysis was performed independently for each band of the hyperspectral image cube. In our prototype setup we make use of the TwinCam-fiberscope bundle. This requires that we register the $NIR_{5 \times 5}$ images to the $VIS_{4 \times 4}$ which will act as the master. Therefore the calibration can be performed using only the $VIS_{4 \times 4}$ images and be used to undistort the $NIR_{5 \times 5}$ images once registered. The results and details for each band using the pin-hole and fish-eye camera model are presented in Table 2.5

2.3. SYSTEM CALIBRATION

Table 2.5: Fiberscopic results corresponding to the pin-hole camera model

BAND	Repro. error	Focal length (in pixels)		Principal point (in pixels)		Polynomial coefficients			Images
		horizontal	vertical	horizontal	vertical	k ₁	k ₂	k ₃	
1	0.35	151.59	150.02	241.67	121.20	-0.7957	1.3355	-4.2033	4
2	0.35	146.53	145.56	244.50	124.00	-0.6176	0.9034	-0.7888	6
3	0.38	151.20	150.49	240.66	120.79	-0.6595	1.2262	-1.6491	6
4	0.39	145.43	144.85	238.45	123.67	-0.6360	1.1100	-1.0939	5
5	0.38	150.42	150.12	239.54	125.47	-0.6219	0.7213	-0.8217	7
6	0.28	153.19	153.23	236.80	125.10	-0.6206	0.7858	-0.6913	5
7	0.33	150.68	150.88	239.25	126.43	-0.6161	0.8743	-0.9528	8
8	0.35	150.94	150.53	238.53	124.89	-0.7306	1.8537	-3.2575	7
9	0.31	151.03	151.15	238.03	126.11	-0.6278	0.9095	-0.9544	6
10	0.33	151.53	151.00	240.37	123.76	-0.6372	1.0008	-1.2120	8
11	0.34	149.22	148.37	242.21	124.41	-0.6347	0.8552	-0.9241	5
12	0.32	151.36	151.35	239.16	125.42	-0.6647	1.0888	-1.2914	5
13	0.41	149.75	149.53	240.49	122.84	-0.7039	1.1800	-3.3858	5
14	0.42	149.49	148.61	240.40	123.23	-0.6644	0.7068	-0.9122	5
15	0.40	148.18	147.64	240.78	123.54	-0.6980	1.2755	-2.7987	7
16	0.38	142.85	141.99	241.56	122.39	-0.6499	1.1140	-2.1618	6

MEAN	0.36	150.04	149.56	240.06	124.06	-0.6619	1.0564	-1.6625	5.94
STD	0.04	2.04	2.23	1.88	1.63	0.0509	0.2988	1.1480	1.18
CV	11%	1%	1%	1%	1%	8%	28%	69%	20%

MAX	0.42	153.19	153.23	244.50	126.43	-0.6161	1.8537	-0.6913	8
MIN	0.28	142.85	141.99	236.80	120.79	-0.7957	0.7068	-4.2033	4

2.3. SYSTEM CALIBRATION

Table 2.6: Fiberscopic results corresponding to the fish eye camera model

BAND	Repro. error	Distortion center (in pixels)		Polynomial coefficients				Images
		horizontal	vertical	a ₀	a ₂	a ₃	a ₄	
1	0.35	239.20	125.18	150.78	-0.011545	0.00021548	-1.6067E-06	4
2	0.35	240.10	125.90	149.51	-0.005786	4.5151E-05	-2.5183E-07	6
3	0.39	239.08	124.38	151.83	-0.004446	1.4569E-05	-6.7871E-08	6
4	0.40	236.73	126.58	145.97	-0.007248	8.853E-05	-5.5571E-07	5
5	0.30	239.02	127.38	150.43	-0.005619	5.9097E-05	-5.0469E-07	7
6	0.28	238.64	127.30	151.32	-0.003707	-4.1848E-06	3.0191E-08	5
7	0.33	239.72	126.50	150.32	-0.004294	7.4181E-06	-1.5325E-08	8
8	0.35	237.42	126.57	152.17	-0.008462	0.00012972	-9.8263E-07	7
9	0.31	238.76	127.55	150.80	-0.005690	4.7518E-05	-3.2643E-07	6
10	0.33	239.51	126.09	151.56	-0.005303	3.9532E-05	-2.927E-07	8
11	0.34	239.48	126.96	150.46	-0.005814	5.3879E-05	-4.0264E-07	5
12	0.32	238.69	126.59	151.95	-0.006631	6.9894E-05	-4.9074E-07	5
13	0.41	239.52	124.91	152.69	-0.009669	0.00015751	-1.1585E-06	5
14	0.42	238.00	125.65	150.38	-0.005706	6.6255E-05	-5.8282E-07	5
15	0.40	238.80	125.22	151.27	-0.009707	0.00016804	-1.282E-06	7
16	0.38	238.46	125.08	151.17	-0.008077	0.00011587	-8.543E-07	6
MEAN	0.35	238.82	126.12	150.79	-0.006732	7.9643E-05	-5.8405E-07	5.94
STD	0.04	0.87	0.97	1.52	0.002208	6.2128E-05	4.7416E-07	1.18
CV	0.12	0.00	0.01	0.01	33%	78%	81%	20%
MAX	0.42	240.10	127.55	152.69	-0.003707	0.00021548	3.0191E-08	8
MIN	0.28	236.73	124.38	145.97	-0.011545	-4.1848E-06	-1.6067E-06	4

and Table 2.6 respectively. The analysis reveals a mean reprojection error of 0.36 pixels over all the bands (Figure 2.12) but more importantly the variability between the focal length and the principal point for the different bands was lower than 1% in all the cases. Moreover, the first distortion coefficient variability was very stable and is less than 8% with a mean value of -0.662 ± 0.05 . The second order coefficient is also lower than 28% with a mean value of 1.056 ± 0.29 . This lead us to believe that we can use the average values to undistort the different bands. The different images used for the calibration and an example of the corresponding reprojection errors are presented in Figure 2.13

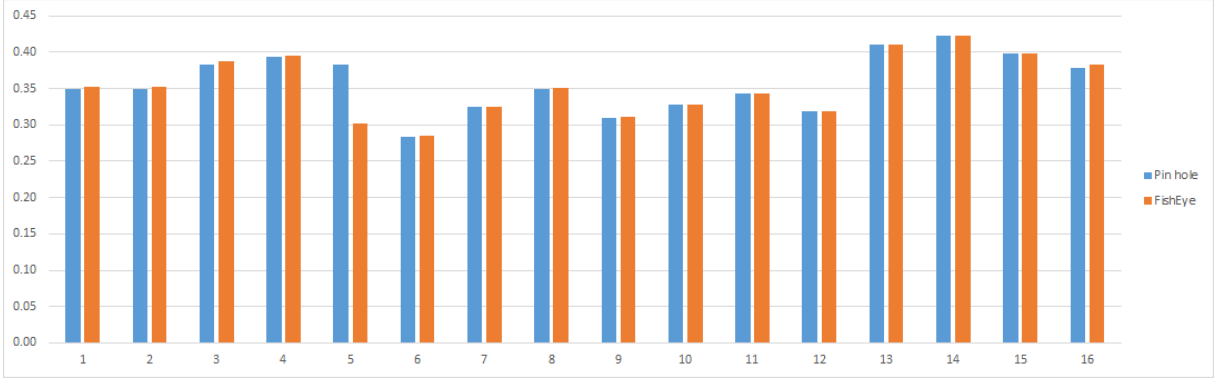


Figure 2.12: Comparison of the reprojection error for the 16 bands of the VIS4x4 camera

2.3.2 Spectral calibration

The multispectral cameras do not directly give a reflectance spectra, they measure the radiance of the observed object. Thus a correction must be applied to mitigate effects introduced by the sensor, optics and illumination (i.e., quantum efficiency, transmission efficiency). The spectral calibration of both cameras was performed by out partners of the Le2i laboratory from the Université Bourgogne Franche-Comté.

The spectral correction is divided in two parts. The first step is to subtract the reference image b and divide by the exposure time τ as shown in Equation 2.6. Where b represents what the camera captures while completely in the dark during the same exposure time τ as the measure. Secondly, the reflectance is obtained by a linear combination of the camera's bands. More formally, let x be a raw pixel of a MS image, represented as a row matrix of 41 values (16 in the visible and 25 in the NIR) and y the corresponding reflectance row matrix obtained through an spectrometer. Then, there exist a transformation matrix W such that

$$\frac{(x - b)}{\tau} W = x' W = y. \quad (2.6)$$

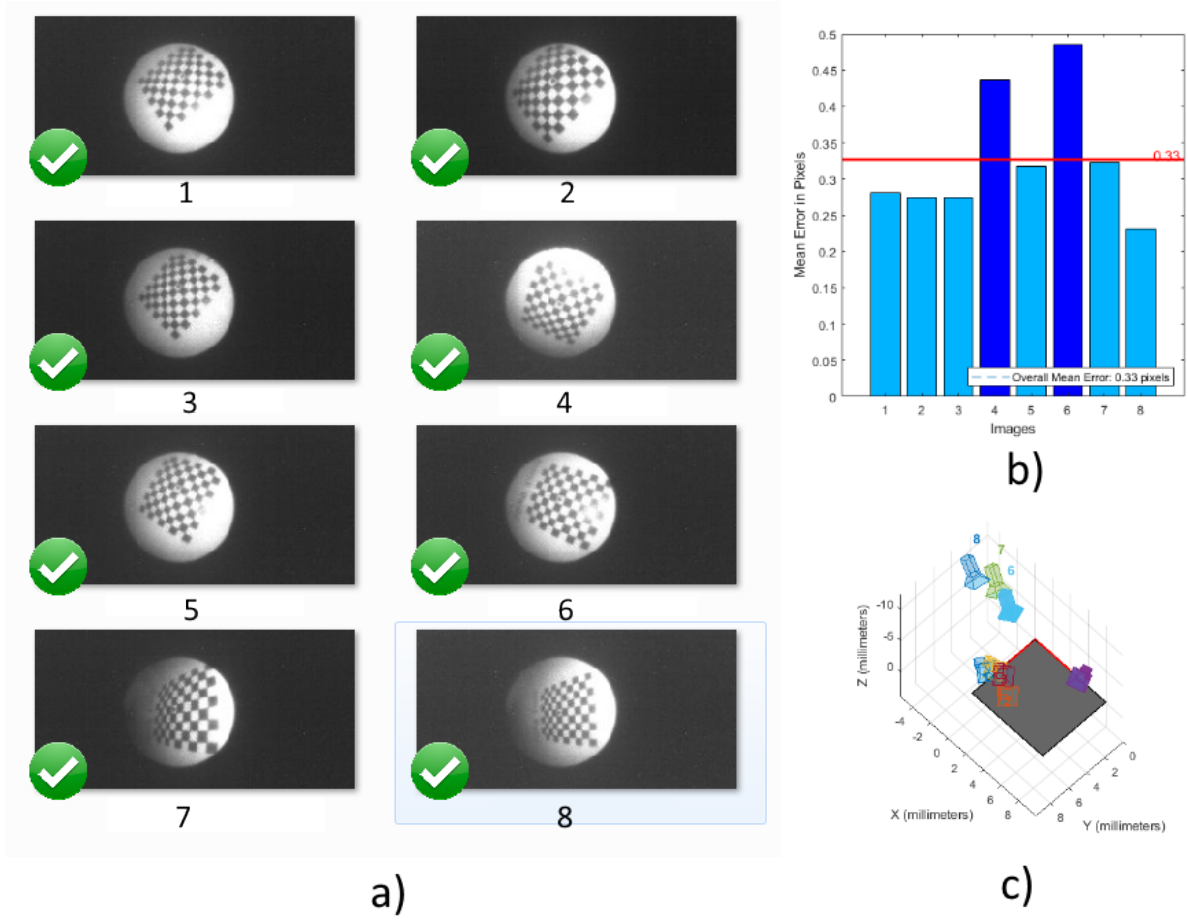


Figure 2.13: Fiberscopic calibration: (a) Chessboard images , (b) Reprojection error, (c) Camera positions relative to the pattern

Finding W can be done by imaging patches from a Digital ColorChecker® SG with MS cameras (x vectors). In practice, 96 patches were used for calibration and corrected to get x' vectors. The true reflectance y is acquired with a AvaSpec-ULS2048XL-EVO spectrometer from Avantes (Apeldoorn, Netherlands) and is able to acquire spectra between 200 nm and 1160 nm with non-constant spectral resolution. To provide a diffuse illumination on patches, an integrating sphere combined with an halogen light source AvaSphere-50-LS-HAL-12V, both from the same manufacturer as the spectrometer, are used. W can then be obtained by classical least square method by minimizing

$$W^* = \operatorname{argmin}_i \sum (x'_i W - y_i)^t (x'_i W - y_i). \quad (2.7)$$

To asses the performance of the spectral calibration, the aforementioned technique is used to simulate NBI and white light (WL) modes of an endoscope and compare them to the actual RGB endoscopic data. As for equation (2.6), there exist a transformation

matrix W_{NBI} and a transformation matrix W_{WL} such that:

$$x'W_{NBI} = y_{NBI}. \quad (2.8)$$

and

$$x'W_{WL} = y_{WL}. \quad (2.9)$$

where y_{NBI} is a vector of three components corresponding to the color of a patch under NBI light and y_{WL} is a vector of three components corresponding to the R,G,B channels of a patch taken under white light.

2.4 Conclusion

In this chapter we presented the MS-augmented endoscopic prototype and the elements it comprises. Also the experimental setup used for acquisition and the calibration of the optical system as a single multimodal image acquisition system. The system uses commercially available devices to assemble a complete solution. The spectral calibration of the HSI system is performed to obtain reflectance information from the raw luminance data captured from the sensors. Thanks to the procedure to overlay the imaged and an hyperspectral equivalent image of 512x256 resolution of 41 bands was recomposed.

While the field of view of the fiberscope is low, the endoscope has a higher field of view (140°) and therefore present sever radial distortion. The spatial calibration results show that the pin-hole model can effectively describe the endoscopic channel and the fiberscopic channel as well without losing precision when compared to the fish-eye camera model. Therefore, we decided to use the simpler but effective camera model to homogenize the model for both cameras. Also, being the values relatively similar between bands, we decided to use the average distortion coefficient for the 41 bands. We also obtain the intrinsic parameters of the endoscopic system which will allow us to create a virtual framework for simulation in the following chapter.

Chapter 3

Prototype modeling and virtual simulation framework

The MS-augmented endoscopic prototype presented in the previous chapter opens the possibility to explore the gastric wall using the two imaging systems simultaneously. The multispectral probe works as a localized optical biopsy with a much smaller field of view. As common approaches by matching images content fail to correctly register the two modalities, we choose to dynamically estimate the pose changes between the two cameras (fiberscope vs. endoscope). Then deduce the homography between the two images. In this chapter, we present a complete modeling of the system as a pair of non-rigidly fixed cameras whose different intrinsic characteristics has been previously calibrated separately using a chessboard pattern. A virtual framework software has been developed based on this modeling to realistically simulate the movements of the fiberscope within the operating channel. This simulation allows to generate a representation of the fiberscope's tip as observed in the endoscopic image with a controlled joint kinematics. It will be used to assess the accuracy of the single view 2D/3D pose estimation algorithm presented in the following chapter.

This chapter begins with a literature revision of instrument virtualization in laparoscopic systems. We continue with the formal presentation of the virtual framework, including projection matrix, camera joint kinematics and fiberscope modeling in 3.2. Finally, the simulated backprojection from 3D to 2D is illustrated in Section 3.3.

3.1 Laparoscopic instrument virtual modeling

Most of the literature in virtualization of instruments is centered in laparoscopic systems. Contrary to traditional surgical operations, laparoscopy does not require the opening of the abdominal cavity. Surgeons handle the instruments based on visual feedback from a

monitor that shows the inside of the cavity [56]. However, due to the restricted vision, reduced degree of freedom (DoF) and difficulty in hand-eye coordination, could easily cause injuries to patient [57]. Therefore, it is necessary to train surgeons and improve their skill. At present, laparoscopic training is performed on animal tissue, cadavers and boxes. These approaches have clear drawbacks: animal tissue is different from humans, cadaveric models are expensive and boxes can only train surgeons for simple skills [58].

Laparoscopic simulation based on virtual reality (VR) can replicate conditions that are accurate in terms of human anatomy. Organs and tissues in the virtual environment exhibit physical and physiologic characteristics of real human organs and tissues through special algorithm. Trainees can practice different levels of operation repeatedly in the virtual system. This can lower the cost and improves the efficiency of training. So far, there have been many studies that prove the suitability of virtual-reality simulator for training [59, 60, 61, 62]. Some studies targeted the simulation of a special operation process. Basdogan et al. [63] made simulated training of laparoscopic surgery on the common bile duct. Heng et al. [57] focused on the development of a training system for knee arthroscopic surgery. Pan et al. [58] studied a virtual laparoscopic rectum surgery.

We observe that the virtualization of laparoscopic instruments have been limited historically to training of clinicians through visual feedback, mostly because of the absence of dual camera systems in the applications in which this type of interventions are common. In the present work, the virtual simulation framework is focused on testing algorithms for pose estimation and image registration. The framework definition is achieved through the simultaneous calibration of both cameras.

3.2 Virtual framework modeling

3.2.1 Notation

Vectors are indicated by bold symbols since scalars are represented by plain letters. 2D and 3D points are expressed in homogeneous coordinates. The operator \sim means that equality is only defined to within a scalar factor (which is case when using homogeneous coordinates).

3.2.2 Projection matrix

The camera reference coordinates system is defined by the origin at optical center C , and the three axes are referred to as C_X , C_Y and C_Z , where C_Z represents the optical axis pointing in the viewing direction of the camera. The plan spanned by axes X and Y is the image plane. The mathematical relationship between the coordinates of a 3D point

in and its projection on the image plane is based on a pinhole model with additional radial distortions. This choice is justified by the calibration experiments presented in the previous chapter.

In computer vision, this 3D/2D relationship for undistorted images is given by a 3x4 projection matrix P using homogeneous coordinates :

$$\mathbf{x} \sim P\mathbf{X} \quad (3.1)$$

where $\mathbf{x} = (u, v, 1)^T$ and $\mathbf{X} = (x, y, z, 1)^T$

The camera projection matrix P is the combination of the extrinsic parameters, which expressed the position and orientation of a camera with 6 degrees of freedom (rotation R and translation \mathbf{t}), and the intrinsic parameters, which represent the characteristics of the camera and are grouped in an upper triangular matrix K .

$$P \sim KR(I - \mathbf{t}) \quad K = \begin{bmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.2)$$

With α_u and α_v are horizontal and vertical focal length (in pixels) considered as equal for cameras with squared pixels; u_0 and v_0 are pixels coordinates of the principal point (i.e. intersection between image plan and optical axis).

World points are transformed to camera coordinates using the extrinsic parameters and the camera coordinates are mapped into the image plane using the intrinsic parameters (Figure 3.1).

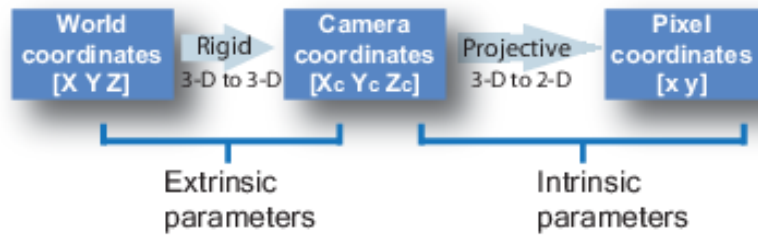


Figure 3.1: World coordinates to Pixel coordinates pipeline

Therefore, the image projection of a 3D point is computed by :

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \sim \begin{pmatrix} \alpha_u & 0 & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} R & -R\mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (3.3)$$

Note that, for a 3D point represented in a camera centered coordinate system, the coordinates of its projection in image are directly computed using identity matrix for rotation and vector null for translation.

The intrinsic parameters of the camera are invariant to its position and are estimated once by calibration. This model is used for both endoscopic camera and fiberscopic one as well, with the appropriate K-matrix for each.

3.2.3 Cameras joint kinematics

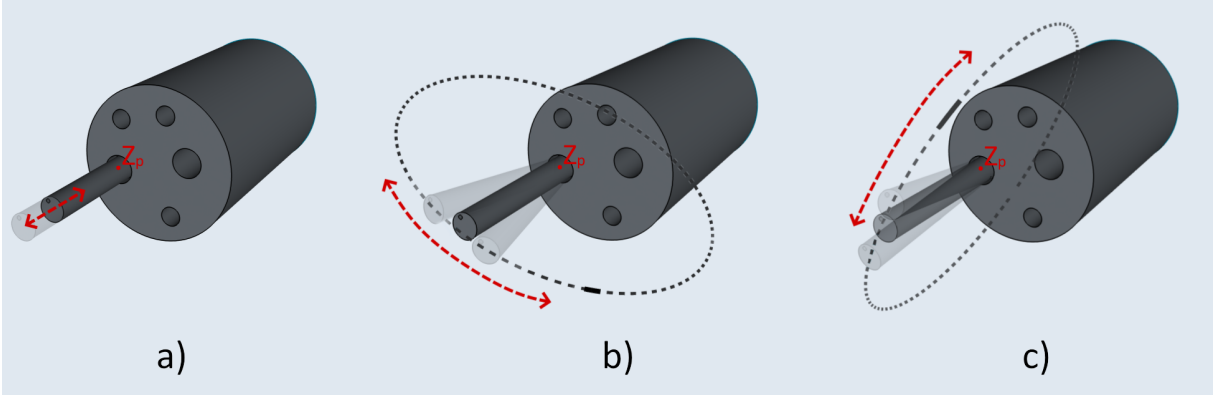


Figure 3.2: Simulated fiberscope movements: a) insertion (z-axis), b and c) precession components: yaw (y-axis) and pitch (x-axis) respectively.

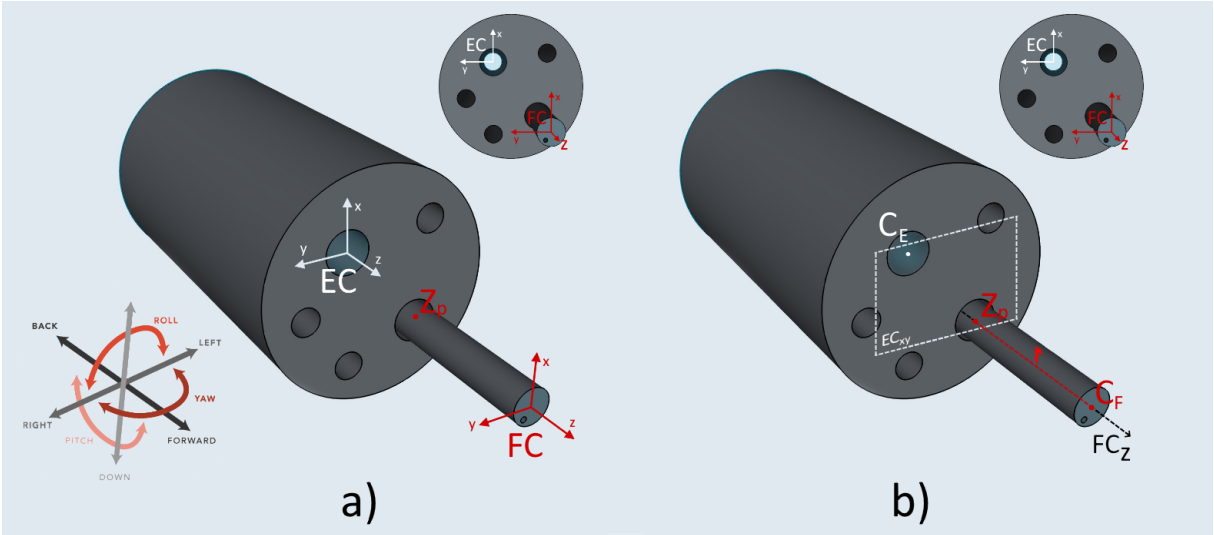


Figure 3.3: a) Kinematic description of the system including the endoscope and fiberscope camera reference frames b) Fiberscope optical axis direction (\hat{l}) definition.

Our multimodal prototype can be compared to a pair of non-rigidly coupled cameras as the fiberscope is re-inserted through the operating canal of the endoscope for each

new patient and it can slide into this channel during the examination. The relative pose between the two cameras generally considers a six degrees-of-freedom motion. However, due to our specific setup, the kinematic model between the two cameras can be simplified under two assumptions:

- As the fiberscope is inserted through the endoscope instrument channel, it can easily slide longitudinally but its lateral displacement is limited due to the similarity between the channel internal diameter (2.8 mm and 3.7 mm depending of the endoscope) and the fiberscopic instrument external diameter (2.5 mm). Therefore the segment of the fiberscope visible to the endoscope camera can be considered coupled to the instrument channel opening which we define as the pivot point \mathbf{Z}_p . The kinematic model relative to the pivot point \mathbf{Z}_p and the extremity of the fiberscope implies two explicit and easy to interpret movements: Insertion and precession. The first corresponds to the longitudinal translation along the instrument axis and the latter to an angular rotation between the two optical axes comprising two components: pitch and yaw on the x- and y-axis respectively. (see Fig 3.2).
- The fiberscope has a perfect axial symmetry and the rotation around its axis (roll) cannot be estimated without the use of markers on the tip. Moreover, the rotation of the fiberscope over its own axis is limited because it is fixed to the optical splitter (Twin-Cam) at the opposite end. We have therefore chosen to ignore this motion parameter. This does not affect the value of the captured data because the biopsy is centered on a small field of view and we can perform averaged measurements or higher order statistics in the spectral space to characterize the texture while respecting this symmetry.

Based on these assumptions, we adopt a simplified kinematic model with three degrees-of-freedom to describe the relative motion between the two cameras. First, we define the coordinate reference frames attached to the endoscope and the fiberscope camera as EC and FC respectively (Fig 3.3a). Then, the optical center of the fiberscope camera is located at the fiberscopic instrument extremity $\mathbf{C}_F = [x_F, y_F, z_F]^T$ while the optical center of the endoscope camera is located at \mathbf{C}_E . Finally, the pivot point \mathbf{Z}_p is defined at the intersection of the fiberscope optical axis \mathbf{FC}_z and the endoscope camera image plane \mathbf{EC}_{xy} using the following equations:

$$\mathbf{p} = \mathbf{C}_F + (s \times \mathbf{d}_{Fz}) \quad (3.4)$$

$$(\mathbf{p} - \mathbf{C}_E) \cdot \mathbf{d}_{Ez} = 0 \quad (3.5)$$

The locus for any point \mathbf{p} in the fiberscope optical axis $\mathbf{F}\mathbf{C}_z$ is given by Eq.3.4, where \mathbf{d}_F is the unity vector of the optical axis and s a scalar in the real number domain. In the same manner, the locus for any point \mathbf{p} in the endoscope image plane is given by Eq.3.5, where \mathbf{d}_E corresponds to the normal vector of the image plane.

Then, by replacing Eq. 3.4 into Eq. 3.5 we can isolate s :

$$s = \frac{(\mathbf{C}_E - \mathbf{C}_F) \cdot \mathbf{d}_{Ez}}{\mathbf{d}_{Fz} \cdot \mathbf{d}_{Ez}} \quad (3.6)$$

and if $\mathbf{d}_{Fz} \cdot \mathbf{d}_{Ez} \neq 0$, there is a single intersection point \mathbf{Z}_p :

$$\mathbf{Z}_p = \mathbf{C}_F + \frac{(\mathbf{C}_E - \mathbf{C}_F) \cdot \mathbf{d}_E}{\mathbf{d}_F \cdot \mathbf{d}_E} \cdot \mathbf{d}_F \quad (3.7)$$

Therefore, the fiberscope optical axis direction is defined as $\overrightarrow{\mathbf{Z}_p\mathbf{C}_F} = \mathbf{C}_F - \mathbf{Z}_p$ and its insertion $depth_F$ by the norm $\|\overrightarrow{\mathbf{Z}_p\mathbf{C}_F}\|$ (Fig 3.3b) and the unity director vector of the cylinder axis

$$\mathbf{l} = \frac{\overrightarrow{\mathbf{Z}_p\mathbf{C}_F}}{depth_F} \quad (3.8)$$

3.2.4 Fiberscope simulation and projection

Our simulation framework is centered on reproducing the observation from the endoscopic camera of the fiberscope movement. Therefore EC is chosen as the system reference frame. The initial definition of $\mathbf{C}_F, \mathbf{C}_E, \mathbf{d}_{Ez}$ and \mathbf{d}_{Fz} in EC is performed using a set of images simultaneously acquired from a chessboard pattern (see Fig 3.4).

The fiberoscopic instrument geometry can be described as a flexible tube of 2.5mm diameter. Its distal tip is rigid and can be modeled as a straight cylinder while its observed by the endoscope. To simulate the fiberscope distal tip observation from EC , a virtual cylinder is computed and projected into EC_{xy} using the matrix P as presented in Sec. 3.2.2.

For 3D rendering, we adopted the numerical model of a cylindrical surface consisting of a point cloud defined by a sequence of n circles of the same radius stacked on top of one another. Each circle is generated by 360 points (i.e., one for each degree) while the number of circles is fixed to 100 per mm of insertion. This allows to achieve a sufficient density of projected points in the endoscopic view. Given the pivot point \mathbf{Z}_p and the pose of the fiberscope tip \mathbf{C}_F , the sequence of circle centers \mathbf{c}_n complies with Eq. 3.9.

$$\mathbf{c}_n = \mathbf{Z}_p + \frac{n}{100} \overrightarrow{\mathbf{Z}_p\mathbf{C}_F} \quad (3.9)$$

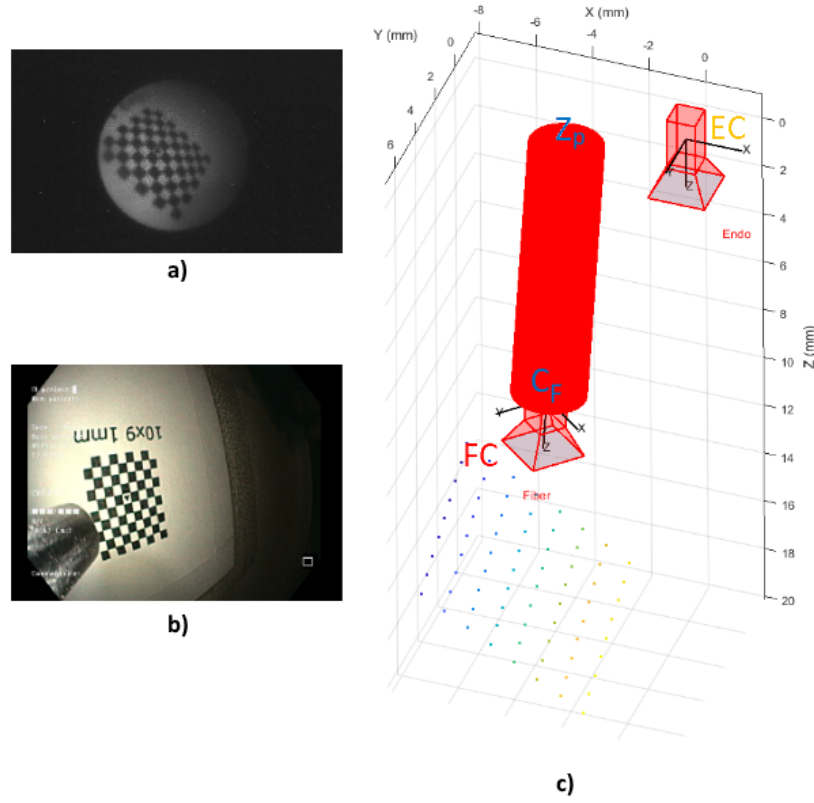


Figure 3.4: Spatial calibration of the system: a) Fiberscopic image and b) Endoscopic image c) Virtual modeling of the fiberscopic instrument in the endoscopic reference frame (The colored points represent the initial calibration pattern)

An example of a generated point cloud and its projection in EC_{xy} is shown in Figure 3.5, the cloud was intentionally created with a lower point density for illustrative purposes.

3.3 Illustration of the kinematic system deployment

The aforementioned system allows not only to simulate the movement of the fiberscope as seen from the endoscope camera but also the instantaneous pose of both cameras. Therefore, we can define \mathbf{C}_F and \mathbf{d}_{Fz} in all the space of the achievable range of movement to obtain on-demand visual feedback from a specific camera pose.

Fig 3.6 and 3.7 depict different configurations of kinematic parameters for relative pose with their corresponding fiberscope projection in the endoscopic view.

Fig 3.8 and 3.9 depict the simulated system and their corresponding projection on EC based on information extracted from images acquired during gastroendoscopic exploration with our prototype. The accuracy of the virtual model projection is backed up by a very high Jaccard index ($J > 0.98$) and the visual feedback illustrates the adequacy of our

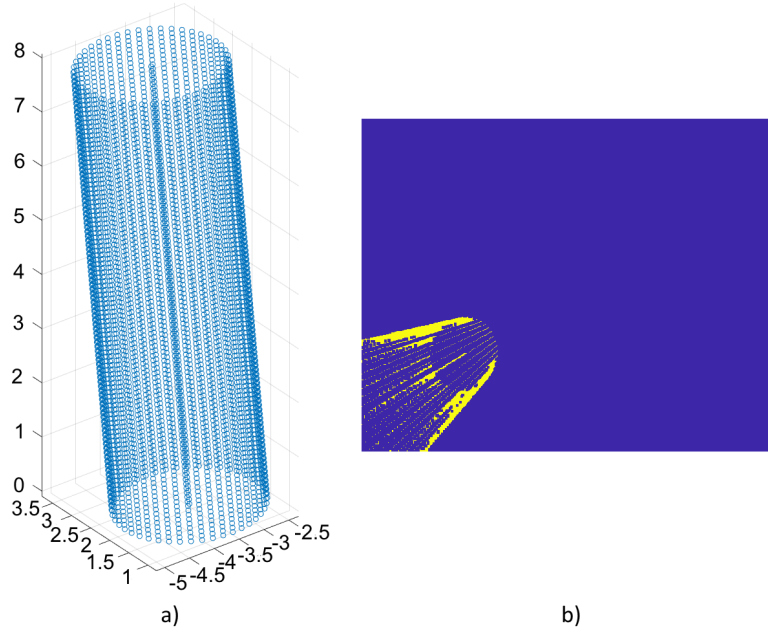


Figure 3.5: a) Point cloud of the modeled cylinder, b) Point cloud projection in 2D

modeling to the complete system.

3.4 Conclusion

We developed a virtual reality framework based on a complete modeling of our prototype in the form of a pair of cameras and a kinematic model to realistically simulate the movements of the fiberscope in the operating channel of the endoscope. We took advantage of the particular conditions found in our set-up to propose a simplified model which describes the relative pose estimation between the two cameras with only three degrees-of-freedom instead of six. Despite its simplicity, this model is totally adapted to realistically simulate the projection of the fiberscope's tip in the endoscopic view.

This virtual framework allows the generation of binary masks which mimic the contour of the fiberscope's tip observed in the endoscopic view. This model is required for the 2D/3D pose estimation algorithm from a single view presented in the following chapter. It also produces synthetic data sets with fully controlled kinematics to evaluate the accuracy and robustness of the proposed algorithm, which could not be done under real endoscopic examination conditions.

The process to extract the pose information from images and the accuracy assessment are detailed in the next chapter.

3.4. CONCLUSION

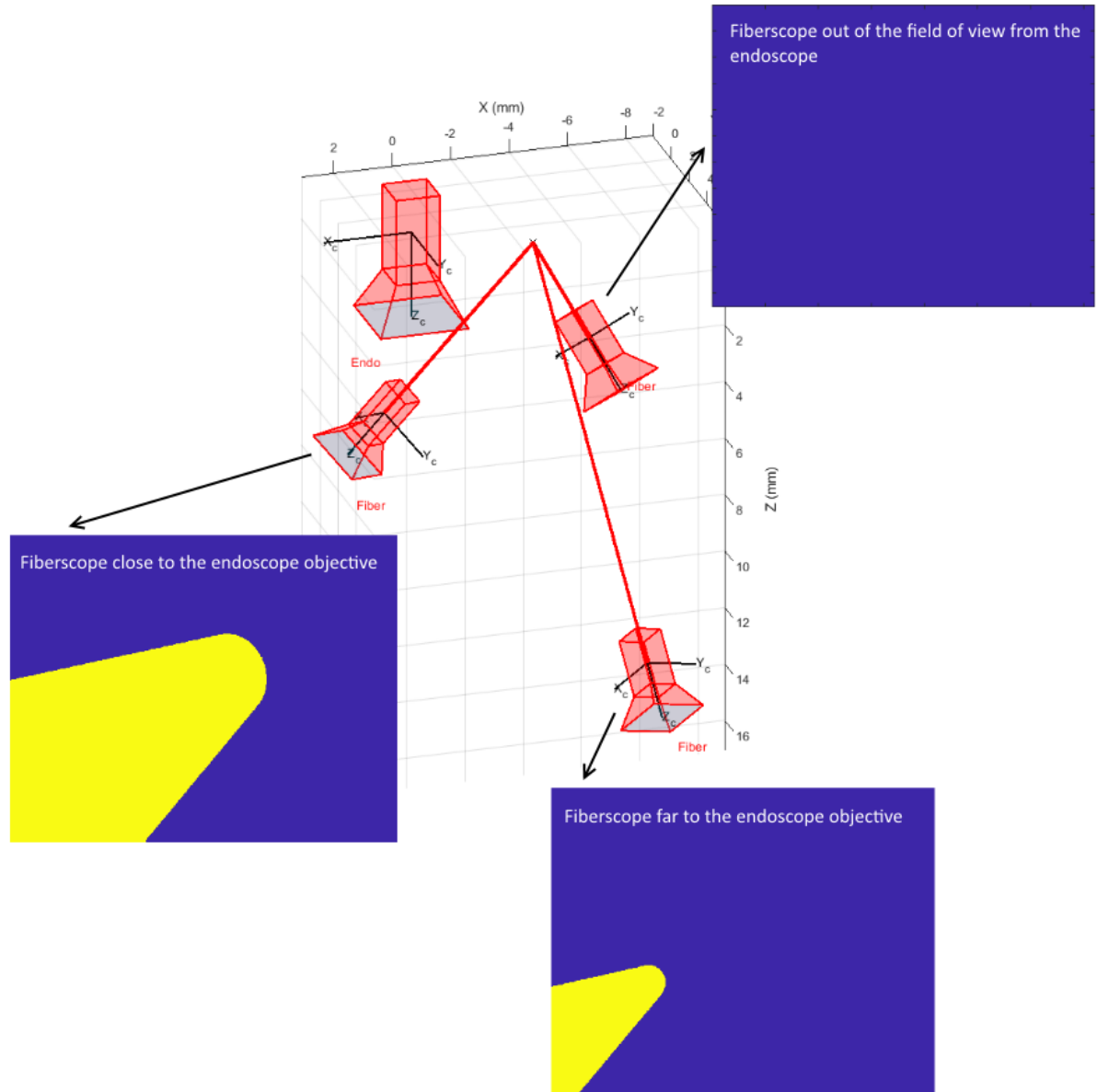


Figure 3.6: Kinematic deployment of the system a) Precession of the fiberscope through its original position's axis, different positions of the fiberscope tip are presented with their corresponding 2D projection

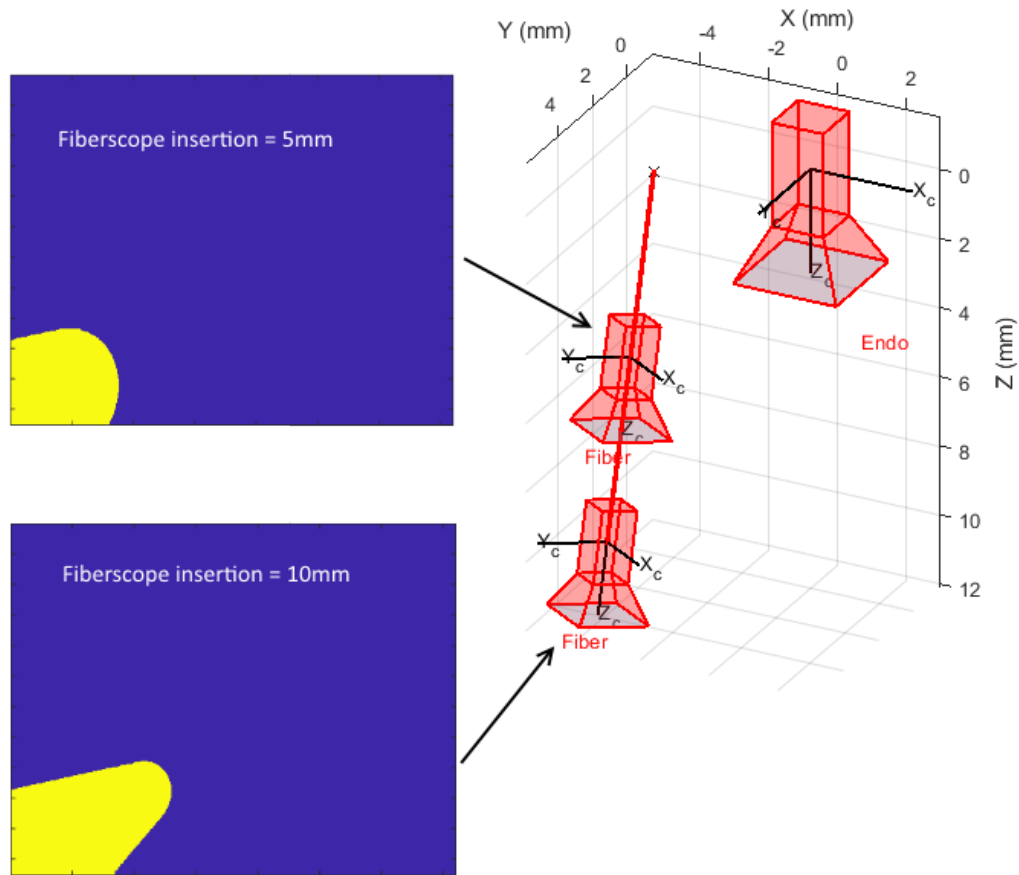


Figure 3.7: Kinematic deployment of the system a) Insertion of the fiberscope through its original position axis, different positions of the fiberscope tip are presented with their corresponding 2D projection

3.4. CONCLUSION

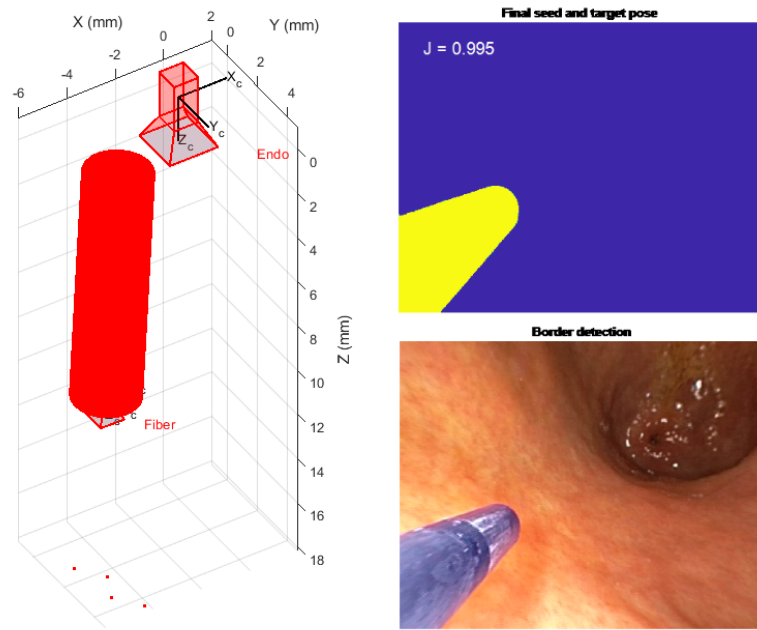


Figure 3.8: a) Point cloud of the modeled cylinder, b) point cloud projection in 2D, c) original endoscopic image

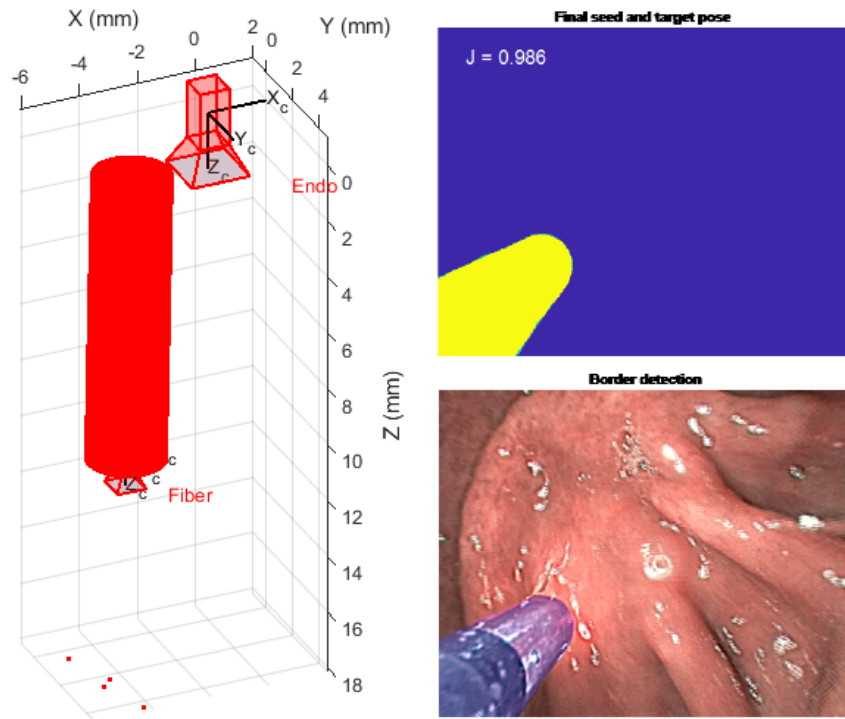


Figure 3.9: a) Point cloud of the modeled cylinder, b) point cloud projection in 2D, c) original endoscopic image

Chapter 4

Markerless fiberscope tracking based on binary mask similarity

In the previous chapter, we described the virtual modelization used to emulate the movements of the two endoscopic instruments. This allowed us to generate any pose and its respective projection on demand. In this chapter we explain how to use the aforementioned virtual framework to estimate the relative pose between both instrument's cameras based only on a single endoscopic image. The process is performed by the maximization of similarity between the segmented silhouette of the fiberscope in the image and that of a projected cylindrical model. This model-based pose estimation is applied image-by-image, thus a tracking-by-detection can be performed in real-time during examination.

4.1 Related work

Early methods of instrument tracking involved attaching electromagnetic or optical markers to the instruments and then estimating pose with a specialized tracking system [64, 65]. In [66], they designed an instrument equipped with a laser beam which projects a pattern onto the surface of the organs, allowing the system to safely bring the instrument to a desired location. This approach allows the use of simple, robust and quick image analysis routines, but raises cost issues and adds to the complexity of the laparoscopic set-up. Similarly [67] uses a RFID antenna connecting to a trocar cannula and an RFID tag attached to the surgical instrument to obtain its instantaneous pose. In [68], the authors worked with a DaVinci system (Intuitive Surgical) as an instrument holder and used kinematic information and template images of the instruments to detect them in stereo images. Some of these methods remain popular today [69]. However, the process of attaching markers to instruments as well as introducing tracking systems to the operating room encounter many regulation issues as they raise safety concerns and complicates

the surgical workflow.

Markerless solutions based on computer vision provide an alternative that can be entirely implemented in software, they do not interfere with the surgical workflow and they do not require modifications to the tracked instrument. This is hugely advantageous as methods can be easily translated to clinic use without an extensive process of distributing markers to hospitals and training medical staff how to attach them correctly. Comparisons between a marker-based and a markerless method used to update the state of a kinematic model of a bending instrument have been presented in [70, 71]. The results revealed a coherent performance between both cases which makes markerless methods a more attractive alternative.

4.1.1 Vision-based markerless tracking

Vision-based tracking of surgical tools in in-vivo scenarios remains challenging, as summarized by Bouget et al. [72], mainly due to endoscopic scenes being heterogeneous environments that feature a high variety of optical interactions that complicate the segmentation (i.e.; tissue specularity, partial occlusions, blurriness (due to fast motion and/or parts of the image out of focus) and presence of body fluids).

From the literature, two principal categories of image-based techniques arise: techniques performing region analysis for each new frame (tracking-by-detection) and techniques re-using the knowledge of tools from previous frames (recursive tracking) [73].

Recursive tracking

Recursive approaches have included tracking via particle filters [74] or Kalman filters [68, 75]; contour-tracking relying on the CONDENSATION algorithm [76]; region-based tracking using mutual information as similarity measure [77]; and feature matching from FAST corners [78]. Kumar et al [79] proposed an optimal fusion between outputs from various trackers, taking advantage of feature-based tracking robustness to small motion and region-based tracking robustness to significant motion.

Other works in the field rely on handcrafted features, such as Haar wavelets [80], gradient [81, 82] or color features [83, 84, 85], which come with their own advantages and disadvantages. While color features, for example, are computationally cheap, they are not robust towards strong illumination changes which are often present during the surgery. Gradients features, on the other hand, are not reliable to withstand the typical motion blur of the tools.

Several works have also made use of line detection algorithms to outline the apparent contours of laparoscopic instruments [86, 83]. The first used the temporal continuity

between consecutive frames and edge detection in conjunction with colour analysis of the laparoscopic tool. Initially, a point in the instrument tip is identified and then the axis of the instrument is estimated using the perpendicular plane which contains said point and the camera optical center. But again the effectiveness of the estimation depends on the proper tracking of the trocar insertion point. The latter estimated the location of the instrument's midline from contiguous frames and updates the system model which was initially set using color segmentation. This approach makes use of time information which increases the precision in the pose estimation, but is not robust enough for videos with low FPS.

Other works [87, 88] have performed a comparative analysis of recursive methods for tracking the stomach, liver and gall bladder. The main draw-back of recursive methods is the requirement of strong priors, something challenging in an endoscopic experimental frame. If something goes wrong between two consecutive frames, for example due to a complete occlusion of the target object or a very fast motion, the system can be lost and must be re-initialized in the same fashion. Nevertheless, all class specific tracking methods require a class-specific detector for (re-)initialisation of the tracking procedure.

Tracking-by-detection

Recent advances in computational power have enabled real-time tracking-by-detection, which has led to further research in this area and increases the viability of these techniques either implemented in a stand-alone fashion or for the purpose of tracking initialization,.

Voros et al. [89] performed a succession of mathematical morphology operations using the 3D tool insertion position in the abdominal cavity and the shape information to strongly constrain the detection search space. In Haase et al. [90], a three-layer approach encompassing clustering and Hough fitting operations has been proposed, assuming rigid tools with cylindrical shaft entering the scene from image boundaries. In Pezzementi et al. [91], a Gaussian mixture model using colour and texture features is used to perform the first stage, while a known by-part 3D model of the tool is iteratively projected (rotation and translation) on the resulting label mask to find the optimal object pose using maximum likelihood estimation.

When estimating the 3D-pose of the instruments, Allan et al. [92] performed pixel-wise classification using Random Forests based on a combination of colour, HOG, and SIFT features. Then the positions are retrieved from the semantic labels map using a flooding algorithm to identify the largest connected components. Underlying shapes are analysed using the moment of inertia tensor to retrieve principal orientation axes. The pose is refined within each region using an energy function and prior information of the tool geometry for the 2D to 3D lifting to obtain final 3D pose estimates. In Sznitman et

al. [93], an instrument-part detector has been proposed, with an early stopping scheme for speed efficiency. The multi-class classifier is combining the gradient boosting framework with edges features to assign an instrument-part or background label to each pixel of an image. Then, the different parts of the instrument are estimated by weighted averaging on the response scores. The overall instrument orientation is retrieved using RANSAC, by fitting the estimated shape of the instrument (i.e. a line) over the resulting labelled image.

As it's classic on visual servoing for laparoscopy, the instrument tracking problem is approached analogously to the pose estimation regarding tools as simple tubular shapes [65, 93] or solid cylinders with a tip alongside the center-line [92, 68, 90]. In this context, rough estimates of a tool shape have been expressed, either being represented by two edges symmetrically spaced from an axis-line containing the tip [89] or by two parallel side segments and a tip lying in-between [94]. On the other hand, highly detailed shapes with joint configurations can be leveraged from a rendering software [95]. A similar approach using the Hough transform to detect the straight lines of the instrument shaft is presented in [96]. This naturally leads to a dependency of the appearance and disappearance from the field-of-view, as tools must intersect with image boundaries to be visible.

4.1.2 Biopsy localization through optical probe tracking

The final objective of our instrument tracking implementation is to use the pose tracking to perform biopsy mapping. Optical probes are typically introduced through the instrument channel while holding the endoscope stationary. Since the probe needs to be placed in facing closely the tissue when the optical biopsy takes place, tracking the tip of the probe enables the localization of the biopsy site.

In Mountney et al. [97] the mapping of a sequence of spatio-temporally tracked biopsy sites are re-projected back onto the endoscopic video to provide a in-vivo augmented view, thus facilitating re-targeting and serial examination of potential lesions. The biopsy localization is based on Simultaneous Localisation And Mapping of the explored organ from the endoscopic camera. The 3D pose of the probe is estimated from 2 points detected in the endoscopic image (the gravity center and the tip), then expressed in the organ's world reference frame to locate the biopsy on the wall.

The current study differs from this work. First, we do not use SLAM and a 3D model of the organ. 3D reconstruction is very difficult in videos of actual patient examinations because singular point detection and matching fail due to lack of texture, blurring, or numerous specular reflections and other liquids. Secondly, we use Plücker's coordinates to estimate the 3D pose of the cylindrical probe from the contours detected in the endoscopic image. This method is much more robust than the estimation from 2 or 4 singular

points. Indeed, the detection of the points is much more sensitive to the segmentation result. We will see in the next chapter that this segmentation is far from obvious for the same reasons as mentioned above for the points matching. We introduce a deep learning approach for doing the segmentation. Finally, we propose a biopsy localization without 3D reconstruction taking into account the knowledge of an approximate of the distance of the endoscopic probe to the organ which is quite intuitive for the surgeon.

4.2 Instrument localization and tracking

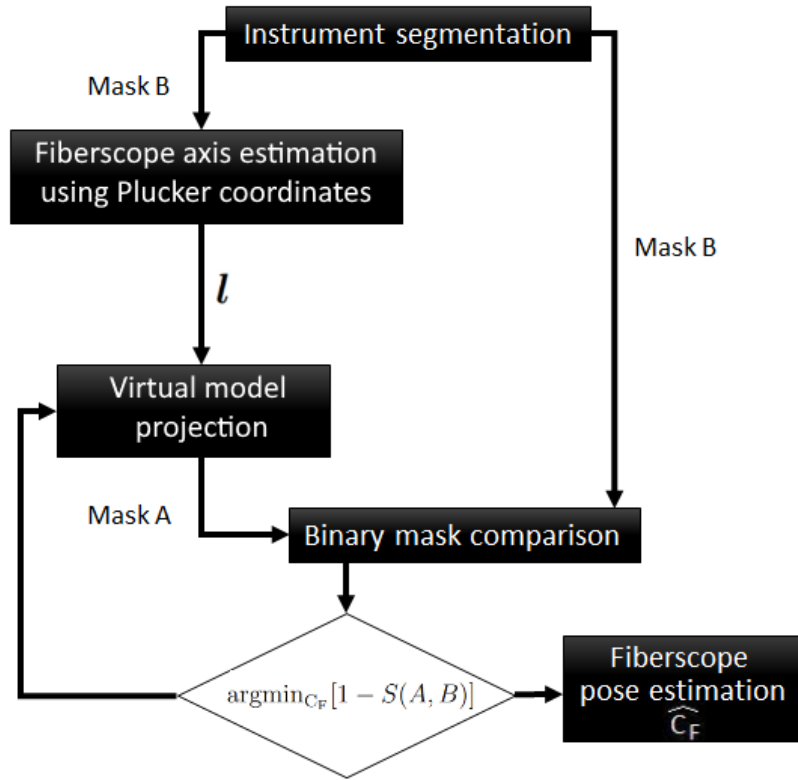


Figure 4.1: Block diagram of the minimization process

Figure 4.1 presents the localization pipeline in a block diagram. The process begins having as input the instrument segmentation (mask B) obtained from the endoscopic view. From mask B, we obtain the estimation of the 3D axis of the fiberscope based on Plücker coordinates (also called the orientation vector l in the endoscopic camera reference frame EC).

We initialize the iterative minimisation process for pose estimation using the orientation information and a fixed depth. At each iteration step, the silhouette observed by the endoscope (mask B) is compared to the projected virtual fiberscope model (mask A) based on the current pose C_F . The pose C_F is then iteratively refined by maximizing

the similarity between the two silhouettes. After convergence, the new optimal pose is obtained. As described in the previous chapter, the fiberscope camera pose relative to the endoscopic camera can be fully defined by only 3 degrees of freedom by assuming a fixed pivot point at the opening of the operating channel and without taking into account axial rotation (roll angle).

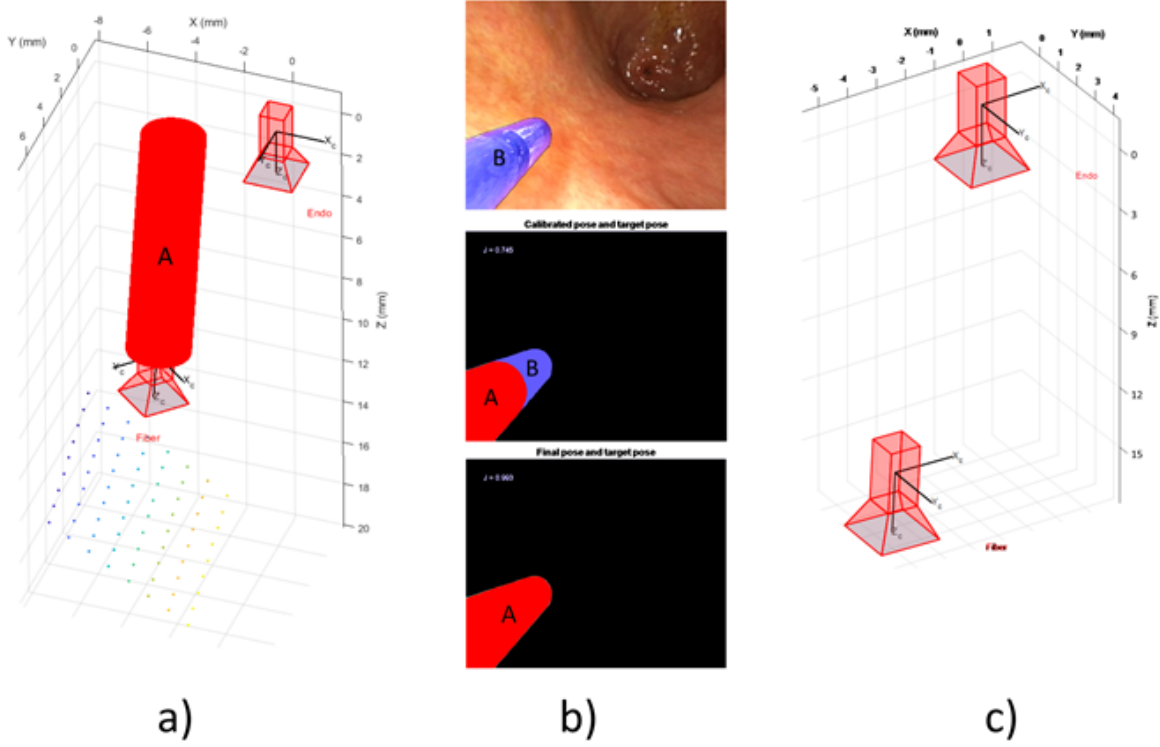


Figure 4.2: Graphical pipeline of the binary mask comparison process

The binary mask comparison or similarity maximisation process is depicted in Figure 4.2. The position of the fiberscope (outlined in red) is extracted from the extrinsic parameters of both cameras based on calibration chessboard pattern (4.2a). Then, with the pose information of both cameras, we design a realistic silhouette of the cylindrical probe as observed by the endoscope (mask A). On the other hand, the endoscopic image segmentation provides the real silhouette observed by the endoscope (mask B). The comparison process between the masks B and A is presented in 4.2b). At the end of convergence, the optimal position of the fiberscopic camera relative to the endoscopic camera by the 3 coordinates of the tip $[x_F; y_F; z_F]^T$ is located at the end of the fiberscope (4.2c).

In the remaining of this chapter we will go in deep through the 3 essentials parts of the aforementioned algorithm:

- Detecting the instrument contour in the endoscopic images (Sub-section 4.2.1)

- Identifying the 3D axis of the cylinder model corresponding to the fiberscopic instrument in each endoscopic image (Sub-section 4.2.2)
- Estimating the 3D pose of the fiberscope tip by the maximization of similarity measures (Sub-section 4.2.3)

4.2.1 Contour detection and pose estimation

Apparent contour of the fiberscopic instrument

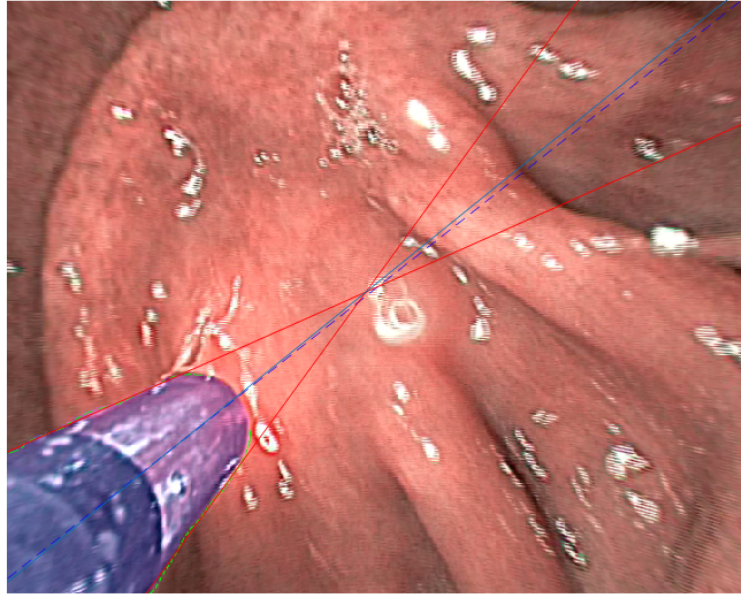


Figure 4.3: Fiberscope’s apparent contour lines estimation through RANSAC

The principal objective from this phase is to obtain the apparent lines that describe the contour of the instrument segmentation as shown in red on Figure 4.3. The input data will be the fiberscope segmentation and the output should be the apparent contour lines. We will assume segmentation can originate from any source (i.e., manual, automatic, etc) for now; a detailed section on the automatic segmentation of the fiberscope in endoscopic images is presented in Chapter 6.

One of the most commonly used approaches for vision-based line detection is a Canny-Hough combination. The outline of the instrument is extracted using Canny edge detector and the classical Hough transform is then used to extract a set of lines as candidates for the lane markers. While this approach shows good results in general, the detected lines are often skewed due to border irregularities (produced by bubbles or gastric fluid attached to the fiberscope) or light reflections in the fiberscopic tip. In this cases, Random Sample Consensus (RANSAC) proved to be robust against border irregularities. In addition, the instrument edge in our method is easily extracted from the binary mask.

The generic RANSAC algorithm robustly fits a model through the most probable data set or inliers while rejecting outliers [98]. Linear Least Squares Estimation (LSE) is then used to fit to a line on the inliers. In our implementation, we decided to apply RANSAC to the lower and upper halves of the instrument contour. Figure 4.3 illustrates the parametrization of the fitted line in terms of two apparent contour lines b^- and b^+ (depicted in red).

The process is divided in three major steps:

- Identifying the center of mass of the instrument segmentation and outline the main axis (blue dotted) of the segmentation to evaluate only one half of the instrument at the time (upper border and lower border).
- Applying RANSAC to the selected points (green) to estimate the line that fits best all the points (red) on each case. This way we obtain the two lines representing the upper and lower contour of the instrument
- Estimating the real axis of the instrument conformed by the vanishing point and equidistant points from both lines.

4.2.2 Cylinder axis estimation using Plücker coordinates

We made use of the linear algorithm for the pose estimation presented by [7]. In this work the authors considered that the fiberscope is a fixed radius cylinder with infinite high and that the endoscopic camera captures its 2D projection. Therefore, we can estimate its 3D orientation (cylinder's axis) using the fiberscope's apparent contour in the 2D image and Plücker coordinates.

Any line in the space has only 4 DoF. This is because it will remain the same line when undergoing a rotation about itself as the axis or a translation in its own direction. Therefore, a line can be represented by its direction vector \mathbf{l} and any point \mathbf{p} that it passes through.

The Plücker coordinates of the line are defined to be (\mathbf{l}, \mathbf{w}) , where $\mathbf{w} = \mathbf{p} \times \mathbf{l}$ is referred to as the moment vector (Fig 4.4). When \mathbf{l} is a unit vector and $\mathbf{l} \cdot \mathbf{w} = \mathbf{l} \cdot (\mathbf{p} \times \mathbf{l}) = 0$, $\mathbf{w} = \mathbf{p} \times \mathbf{l}$ is called the moment of the line. The norm $\|\mathbf{w}\|$ gives the distance from the origin to the line, achieved at the foot \mathbf{p}_\perp of the line's perpendicular through the origin, where $\mathbf{p}_\perp = \mathbf{l} \times \mathbf{w}$.

Any point M on the cylinder surface with coordinates (x, y, z) in the camera frame may be projected into the image plane at $\mathbf{m} = (u, v, 1)^T$. The apparent contour can be modeled as a set of two straight lines. Let \mathbf{b}^- and \mathbf{b}^+ be the two lines representing

the borders of the cylinder in the image and K^e the intrinsic camera matrix, we apply Equation 4.1.

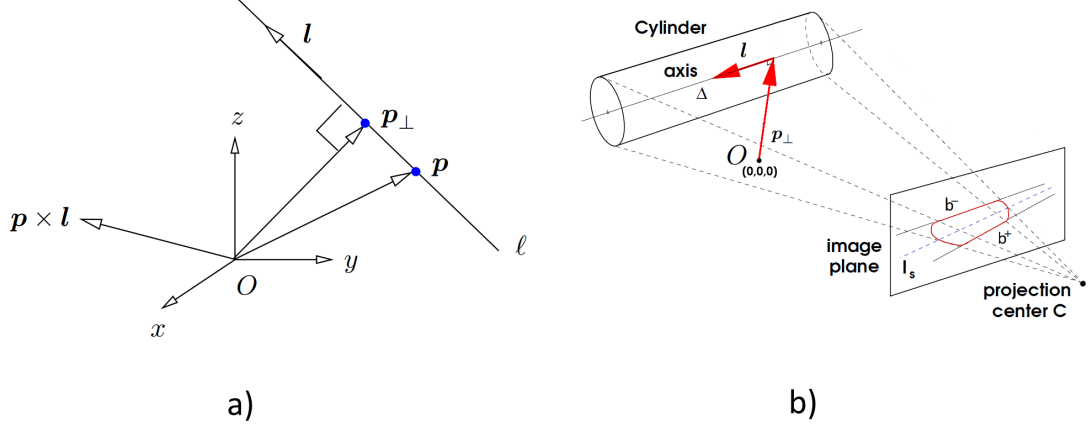


Figure 4.4: a) Plücker coordinates $(\mathbf{l}, \mathbf{p} \times \mathbf{l})$ of a line. Here \mathbf{p}_{\perp} is the closest point on the line to the origin \mathbf{O} b) Plücker representation of the cylinder's axis and its projection into the endoscopic image (Adapted from [7])

$$\begin{cases} (\mathbf{b}^-)^T \mathbf{m} \equiv \left\{ (K^e)^{-T} \times (I - \alpha [\mathbf{l}]_x) \mathbf{w} \right\}^T \mathbf{m} = 0 \\ (\mathbf{b}^+)^T \mathbf{m} \equiv \left\{ (K^e)^{-T} \times (I + \alpha [\mathbf{l}]_x) \mathbf{w} \right\}^T \mathbf{m} = 0 \end{cases} \quad (4.1)$$

With the matrix K^e as the endoscopic camera intrinsics, I the identity matrix and α being relative to the radius of the cylinder r_c :

$$\alpha = r_c / \sqrt{\|\mathbf{w}\|^2 - r_c^2} \quad (4.2)$$

Starting from equations 4.1, let \mathbf{b}^- and \mathbf{b}^+ be the two lines representing the borders of the cylinder in the image and K^e the intrinsics camera matrix, we apply Eq (2).

$$K^{eT} (\mathbf{b}^- \mathbf{b}^+ + \mathbf{b}^+ \mathbf{b}^-) K^e \equiv K^{eT} C K^e \quad (4.3)$$

Then, the decomposition of the matrix C can be related to the Plücker coordinates (\mathbf{l}, \mathbf{w}) [7]:

$$K^{eT} C K^e \equiv \begin{bmatrix} \mathbf{w} & z_u & \mathbf{l} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 - \sigma^2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{w}^T \\ z_u^T \\ \mathbf{l}^T \end{bmatrix} \quad (4.4)$$

Where $\mathbf{z} = \frac{\sqrt{1-\alpha}}{\|\mathbf{w}\|} [\mathbf{l}]_x \mathbf{w}$ while $\alpha = \frac{r_c}{\sqrt{\|\mathbf{w}\|^2 - r_c^2}}$ and the unity vector $\mathbf{z}_u = \frac{\mathbf{z}}{\sigma}$ while $\sigma = \|\mathbf{z}\|$.

Finally, the cylinder axis can be easily derivated from Plücker coordinates by finding the vector that connects \mathbf{p}_\perp to \mathbf{Z}_p . However this vector represents only the axis of the cylinder but not its actual insertion depth, for this reason we require extra processing to completely model the fiberoptic instrument.

4.2.3 Fiberscope pose estimation by binary mask similarity maximization

The relative pose estimation between the two heterogeneous cameras (endoscopic/fiberscopic) is therefore expressed as a minimization problem by fitting the projection of the virtual cylinder to the segmented silhouette of the fiberscope tip in the endoscopic image. The objective function consists in maximizing a similarity measure between the two binary masks (A and B) by iterative modification of the relative pose between cameras. If we achieve a perfect fit in the image plane we can obtain information from the 3D pose.

$$\hat{\mathbf{C}}_F = \operatorname{argmin}_{\mathbf{C}_F} [1 - S(A, B)] \quad (4.5)$$

To compare the two binary masks we tested three commonly used indexes [99]. All similarity indexes S comply the condition $0 \leq S \leq 1$. First, the Jaccard index J , also known as intersection over union is defined as the size of the intersection divided by the size of the union of the sample sets A and B:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}. \quad (4.6)$$

Secondly, Dice or Sørensen's measure is defined as twice the number of elements common to both sets divided by the sum of the number of elements in each set as presented on Eq.4.7:

$$DSC(A, B) = \frac{2|A \cap B|}{|A| + |B|} \quad (4.7)$$

where $|A|$ and $|B|$ are the cardinalities of the two sets (i.e. the number of elements in each set). The Dice coefficient is used in computer graphic lexicography to measure the lexical association score of two words. In contrast to Jaccard's index, it is not a distance since it does not check for triangular inequality. Whereas Jaccard is a true distance.

Finally, the F score represents how accurately the predicted boundary of an object matches the ground truth boundary. It is defined as the harmonic mean of the precision and recall values. Precision is the fraction of detections that are true positives rather than false positives. Recall is the fraction of true positives that are detected rather than missed.

$$F(A, B) = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (4.8)$$

$$\text{precision} = \frac{TP}{TP + FP} \quad \text{recall} = \frac{TP}{TP + FN} \quad (4.9)$$

The behavior of the complement of similarity indexes are observed varying each parameter deviation to an arbitrary fiberscope pose \mathbf{C}_F as shown in Figure 4.5, this way zero corresponds to the convergence objective. In addition, Figure 4.6 depicts graphically how the similarity measures vary in selected iteration frames.

Although the F score behaves as a strongly convex function while evaluating x_F and y_F parameters, it lacks of singular minimum values when analyzing z_F . Jaccard and Dice behave similarly in the three cases, describing a singular min value in the exact fit case. However, Jaccard's function presents itself as a better option due to the higher value of its local tangents which may lead to a faster convergence. Then, Equation 4.5 becomes:

$$\hat{\mathbf{C}}_F = \text{argmin}_{\mathbf{C}_F} [1 - J(A, B)] \quad (4.10)$$

where $\hat{\mathbf{C}}_F$ are the coordinates of the point representing the guessed pose of the cylinder's extreme point (fiberscope tip). Fig 4.7 depicts an example of convergence of $J(A, B)$.

We begin using a 2D projection of the virtual instrument with the orientation detected in the previous step but with an unknown insertion and we target the real image segmentation. For this purpose we use a simplex multidimensional unconstrained optimization introduced by Nelder and Mead [100].

One important process in any optimization method is the generation of a new set of search locations for the next iteration. The performance and complexity of any method is highly dependent on this process. The simplest approach is to use a predetermined uniform distribution of search locations. This approach is adopted by most convex optimization algorithms. There are some complex approaches like the use of crossover and mutation operators in genetic algorithms or the use of gradients in gradient-descent algorithms. Simplex is a simple optimization algorithm seeking the vector of parameters corresponding to the global extreme (maximum or minimum) of any n-dimensional function $F(x_1, x_2, \dots, x_n)$, searching through the parameter space.

The 3-dimensional simplex starts with 4 observations of the system response obtained with 4 different (x_1, x_2, x_3) parameter settings or guesses. These 4 observations correspond to the vertices of a tetrahedral body constituting the 1st simplex. From the evaluation of the response of each observation the position of next point to be evaluated (i.e. the new set of experimental conditions) is indicated by either reflection, expansion, or contraction operations.

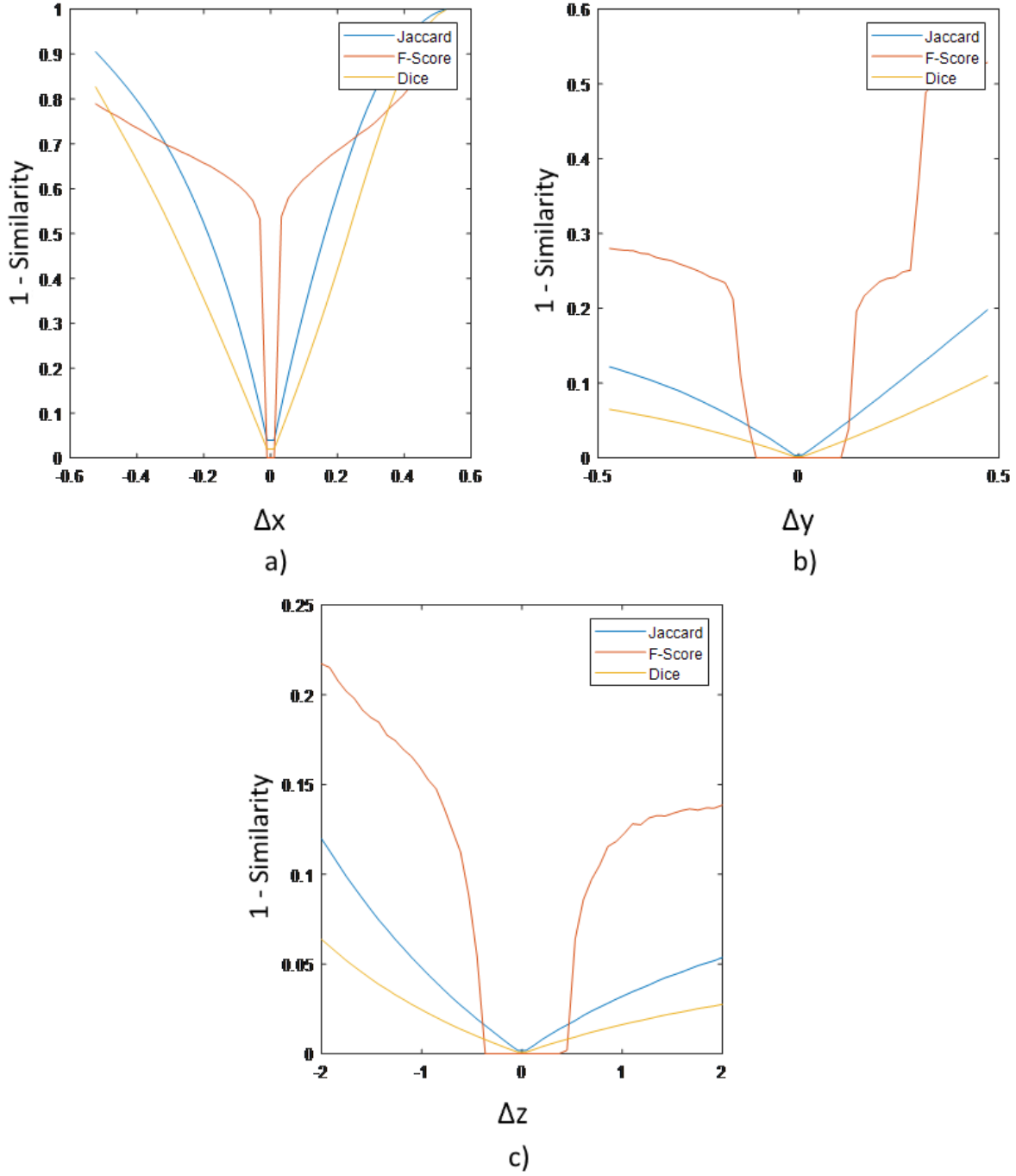


Figure 4.5: Behaviour of objective function using Jaccard, Dice and F-score similarity measures, according to each parameter deviation: a) x_F , b) y_F and c) z_F . Zero corresponds to the convergence objective (i.e. the perfect superposition of the two masks)

Given our problem is linear, the simplex method is an attractive choice for solving the function because it uses very simple equations while moving the point over the function. Although the simplex method has exponential complexity in the worst case, in practice it is still quite fast on most real world problems. A particular efficient implementation of

4.3. CONCLUSION

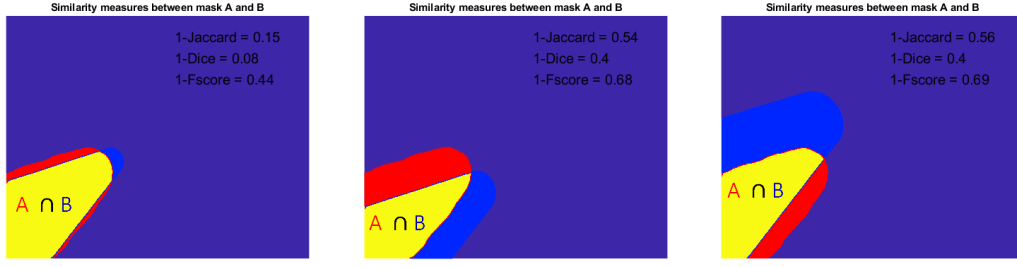


Figure 4.6: Binary mask similarity comparison. Mask A and B are colored and outlined by red and blue respectively.

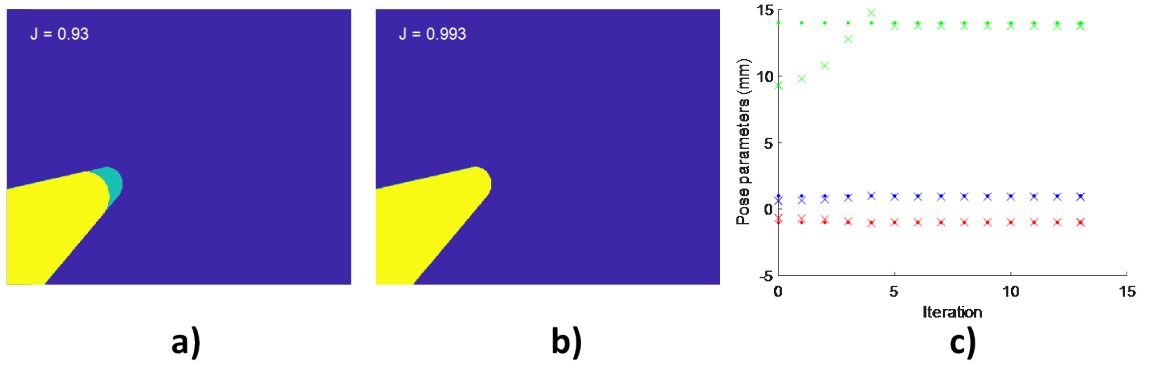


Figure 4.7: Binary mask similarity maximization: a) Initial state b) Final state c) Pose parameters evolution vs the number of iteration.

this method is presented by [101]

We can observe from Fig 4.7 that the closest the iterative method gets to convergence the higher the value of our similarity metric and thus the lower of its complement. Therefore when the maximisation of similarity between the two binary masks converges, we can return to the 3D space to estimate the insertion depth of the instrument in the virtual framework.

4.3 Conclusion

This chapter presented a landmark-free approach for tracking the motion of a fiberscope inserted in the accessory/biopsy channel of the endoscope. The ability to perform this type of pose estimation lifts the restriction other methods found in the literature have like depending on markers in the instrument or the temporal continuity of the frames. This is highly desirable in our application due the low ratio of image acquisition when comparing HS to RGB.

We took advantage of our prototype set-up to propose a simplified kinematic model

with only 3 parameters to describe the relative pose between the two cameras. The relative pose estimation is solved as a classical problem in computer vision of model-based object tracking. The technique relies on the silhouette of the fiberscope's tip in the endoscopic images considered as the projection of a 3D cylinder model to dynamically estimate the pose changes between the two cameras image-by-image. We used Plücker coordinates to backproject the cylinder axis from an initial segmentation performed in the RGB modality and obtain the final pose by maximizing the similarity between both silhouettes (i.e., probe segmentation and model projection). The cost function is necessary to estimate the insertion depth of the instrument since the backprojection only gives information of the fiberscope direction. Also, we explored different similarity measures and assessed their behavior in our proposed setup. Based on these results we decided to use Jaccard as the principal metric due to its convexity and singularity in the convergence points.

In the following chapter we will assess the performance of our proposed method in simulations and clinical data.

Chapter 5

Validation on synthetic and real images

In this chapter, we assess the accuracy and precision of our virtual framework which simulates the multispectral(MS)-augmented imaging prototype.

The virtual framework was set up using custom MATLAB from MathWorks (MA, USA) software and using the same dimensions and camera matrix as the optical instruments used in the prototype presented (i.e., Olympus EVIS EXERA III (Tokyo, Japan) endoscopic system and microflex m2.5-2500 fiberscope ITConcepts (Lahnau, Germany)).

Five set of tests are performed distributed over the chapter's section. Section 5.1 and 5.2 evaluate the propagation of error by contour estimation uncertainty. The first one is focused on evaluating the impact of the apparent contour on the backprojection of the cylinder axis, the second presents a similar analysis over the superior and inferior apparent borders of the instrument while varying the precision on the contour detection. Section 5.3 compares the pose estimation accuracy under a simulated environment and Section 5.4 assesses the pose tracking in a realistic trajectory scenario. Section 5.5 and 5.6 are focused in measuring the reliability of the estimates using real images of a chessboard pattern.

5.1 Propagation of error for axis estimation

To analyze the robustness of Plücker coordinates estimation, we performed a Monte Carlo simulation by introducing random errors in apparent contour line parameters to simulate errors in the border extraction. We used the parametric representation of a line:

$$\rho = x * \cos(\theta) + y * \sin(\theta) \quad (5.1)$$

Where, ρ is the distance from the origin to the line along a vector perpendicular to the line and θ is the angle of the perpendicular projection from the origin to the line

measured in degrees clockwise from the positive x-axis. The range of θ is $-90^\circ \leq \theta < 90^\circ$. The angle of the line itself is $+90^\circ$, also measured clockwise with respect to the positive x-axis as shown in Figure 5.1.

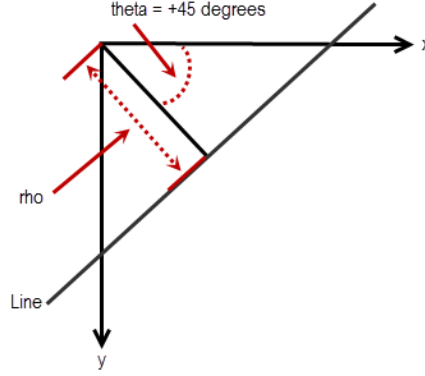


Figure 5.1: Parametric representation of a line on terms of θ and ρ

We performed the analysis only over the parameter θ because the error are usually produced by bubbles and light reflection which deform the contour of the fiberscope and altering the slope of the line. In addition, the fiberscope is almost always appearing on the image from the same origin and direction. One hundred trials of uniformly distributed random noise were added to the lines parameter θ of one border line without changing the second line, varying the standard deviation from 0 to 5 in steps of 0.1 degrees. A representation of the simulated lines is presented in Figure 5.2. The black lines represent the ground truth (GT) instrument contour and the colored lines the noise-affected simulated lines. Then, we evaluate the propagation of errors on the Plücker coordinates \mathbf{l} and \mathbf{w} .

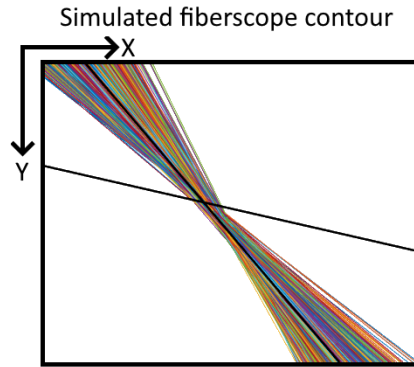


Figure 5.2: Simulated apparent contours on terms of θ from 0 to 5 in steps of 0.1 degrees

The cylinder axis is defined by the vector \mathbf{l} which is parallel to the fiberscope camera optical axis. We observed the angle deviation between the estimated vector \mathbf{l} and the GT (\mathbf{l}_{GT}) by the dot product:

$$AD = \text{acos}\left(\frac{\mathbf{l} \cdot \mathbf{l}_{GT}}{\|\mathbf{l}\| \times \|\mathbf{l}_{GT}\|}\right) \quad (5.2)$$

We also observed the Euclidian distance in millimeters between the estimated pose of the tip $\hat{\mathbf{C}}_F$ and the ground truth \mathbf{C}_{GT}

$$D_E = \|\hat{\mathbf{C}}_F - \mathbf{C}_{GT}\| \quad (5.3)$$

The results of the tests are presented in Figure 5.3 and 5.4

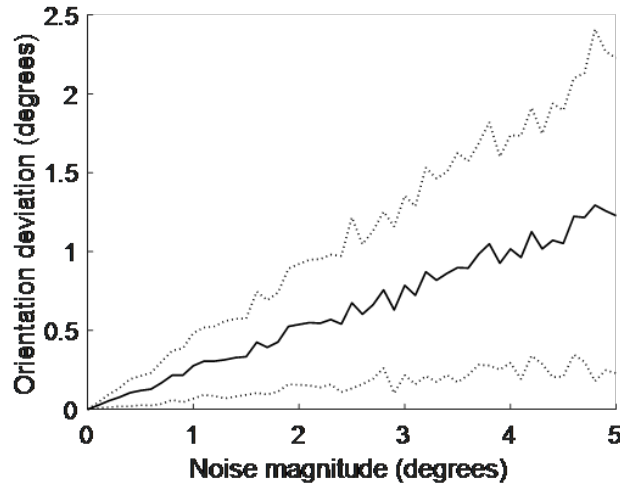


Figure 5.3: Monte Carlo test for 100 random trials : Mean and STD of the angular deviation between \mathbf{l} and \mathbf{l}_{GT} (in degrees) versus noise on θ .

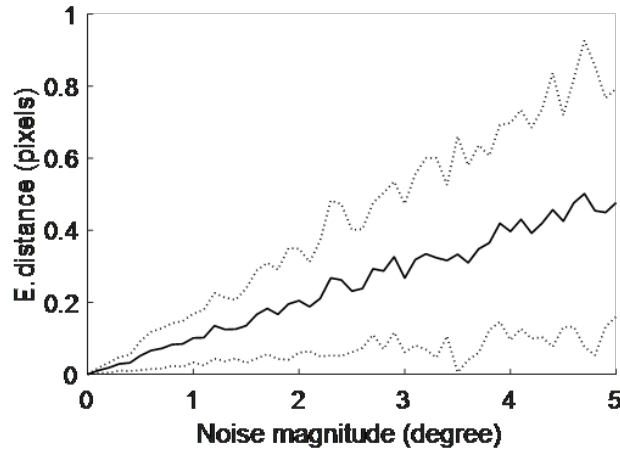


Figure 5.4: Monte Carlo test for 100 random trials : Mean and STD of E_D between $\hat{\mathbf{C}}_F$ and \mathbf{C}_{GT} (in mm) versus noise on θ .

We evaluated the consequence in deviation on the axis tolerance. In the first experiment, we observe that for a noise magnitude lesser than 5 degrees the mean error from point P over the axis of the cylinder is lower than 1 pixel. In addition, when measuring the deviation produced in the instrument's estimated axis in a range from 0 to 5 degrees of noise, the orientation deviation of the axis is lower than 1.5 in all cases.

5.2 Propagation of error for borders estimation

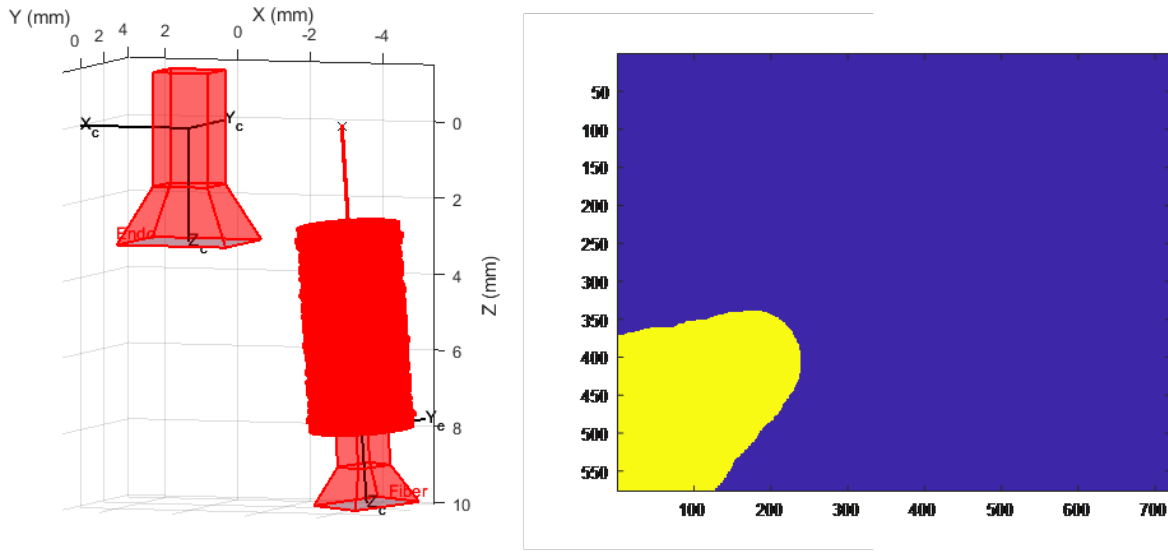


Figure 5.5: a) Noise corrupted virtual instrument b) Instrument's projection

To analyze the robustness of the apparent contour lines when the fiberscope mask does not perfectly reflects its cylindrical geometry, we performed a Monte Carlo simulation by introducing random errors in the cylinder's borders. We have already seen in Section 3.2.4 that the cylinder can be represented as a group of circles of the same radius extending themselves over the same axis $\overrightarrow{Z_p C_F}$. Therefore, we added random variations to the radius of these circles in a range of 0 to 0.2mm which represents 16% of the real fiberscope radius. This variation is reflected in the projected image as an irregular contour which resembles the variability of the contours obtained through the automatic segmentation. Figure 5.5 depicts a corrupted cylinder with 0.05 mm mean noise and its corresponding projection.

We evaluated the consequence in deviation on the upper and lower border tolerance. In both cases the slope of the apparent lines behaves robustly to the insertion of variability in the initial mask for estimation. In the upper border case, the maximum slope variation is lower than 0.1 while in the lower border case is lower than 0.4 (Figure 5.6).

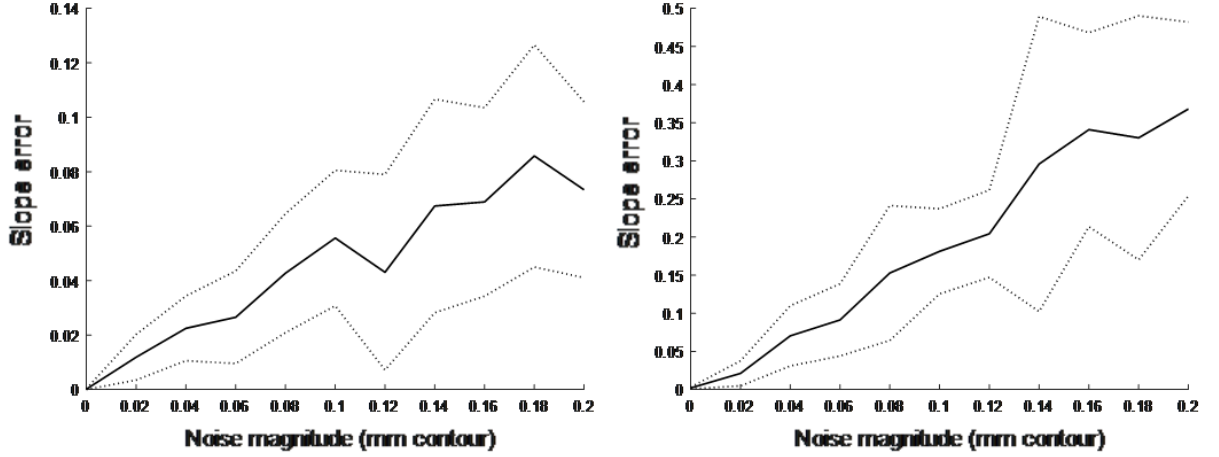


Figure 5.6: Monte carlo tests for 100 random trials : Mean and STD of the slope measurement error for the instrument borders a) Upper contour line slope b) Lower contour line slope

5.3 Test on synthetic data

To simulate the movement performed by the practitioner during endoscopic exploration, we created an helicoidal trajectory of the probe's tip, varying the insertion in a range from 8 mm to 18 mm, and the angular deviation inside a disk of 4 mm radius (see Figure 5.7).

This trajectory makes it possible to test a hundred different configurations between the optical probe and the endoscopic camera. These hundred samples can be considered individually for convergence tests or for simulating a frame-by-frame tracking, where the pose estimated in the previous frame is used as initialization for the current frame. We compared four variants for the pose estimation according to the initialization before minimization :

- "3DoF-min": for all samples we used $\mathbf{C}_{F_0} = [-3.80; 2.35; 5.27]$ as the calibrated position to produce the initial mask and did not made use of the Plücker based axis alignment ;
- "Plücker + 3DoF-min": for all samples, the started point for minimization is $\mathbf{C}_{F_P} = [x_p; y_p; 5.27]$, where x_p and y_p are calculated from the current frame by applying Plücker based axis alignment ;
- "Tracking without axis alignment": for each frame, the optimization algorithm is initialized by the pose estimated in the previous frame \mathbf{C}_{F-1} (Plücker based axis alignment is not activated);

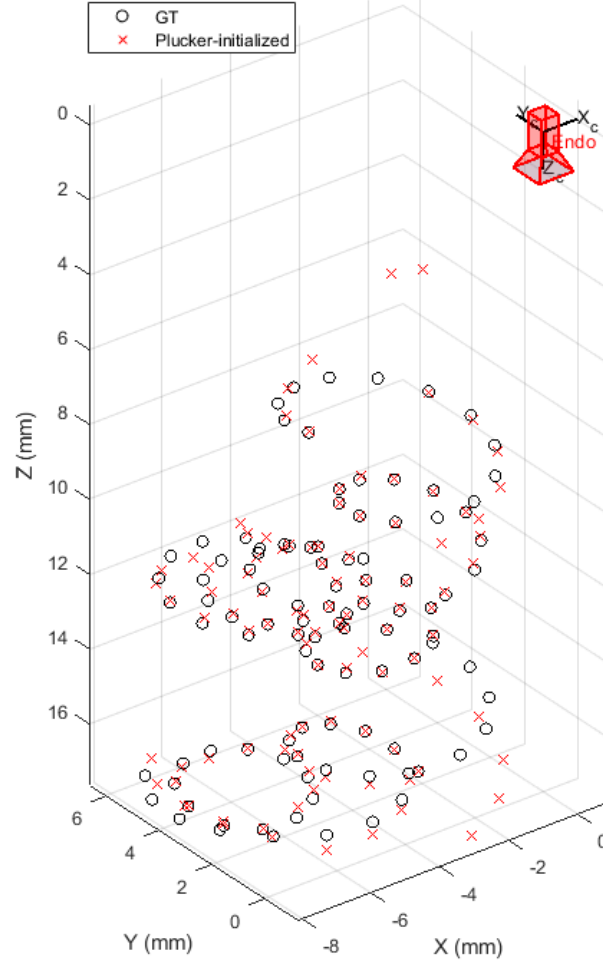


Figure 5.7: Simulated helicoidal trajectory mimicking an endoscopic exploration: comparison of estimated poses and ground truth.

- "Tracking with axis alignment": for each frame, we apply Plücker based axis alignment to calculate x_p and y_p , then the started point for minimization is $\mathbf{C}_{F_p} = [x_p; y_p; z_{F-1}]$, where z_{F-1} is from the pose estimated in the previous frame.

A summary of statistics is presented in Table 5.1. The accuracy of estimated pose is assessed from the ground truth by the mean, median and standard deviation of Euclidian distance and optical axis deviation (Equations 5.3 and 5.2) over the hundred samples of the helicoidal trajectory. The reprojection error in the endoscopic view due to pose deviation from ground truth is estimated on a virtual grid of size 1.5 mm x 2 mm placed at a distance of 6 mm in front of the optical probe. The convergence is also evaluated by the number of iterations and the final Jaccard index.

When comparing "3DoF-min" to "Plücker + 3DoF-min", we can observe that applying Plücker based axis alignment provides more accurate pose estimates. The difference

5.3. TEST ON SYNTHETIC DATA

Table 5.1: Error metrics over the helicoidal trajectory (100 positions)

Method		Pose distance (mm)	Angular Deviation (deg)	Reproj. Error (pixels)	Number of iterations	Final Jaccard index
3DoF-min	Mean	0.93	1.92	8.57	57	0.983
	STD	2.43	5.11	22.50	14	0.047
	Median	0.02	0.01	0.14	59	0.999
Plücker + 3DoF-min	Mean	0.52	1.02	4.58	55	0.991
	STD	1.67	3.24	14.34	14	0.030
	Median	0.02	0.01	0.10	55	0.999
Tracking without axis alignment	Mean	1.54	0.76	6.23	46	0.984
	STD	4.42	3.05	14.14	15	0.039
	Median	0.05	0.02	0.31	45	0.999
Tracking with axis alignment	Mean	0.33	0.32	2.57	24	0.996
	STD	0.68	1.73	9.43	13	0.007
	Median	0.11	0.02	0.43	20	0.998

between median and mean values is explained by the presence of outliers (see Fig. 5.8). The red points correspond to a convergence that does not lead to an optimal final value of the objective function (1-Jaccard). For information, the tolerance on the stopping criterion is 0.001. The error metrics indicate that initialization after Plücker based axis alignment gives clearly more precise and robust results.

When evaluating the tracking experiments, Plücker based axis alignment prior to minimization is still more efficient, accurate and faster (2 times). The number of outliers is greatly reduced and the mean and median values are close to each other. Finally, applying Plücker based axis alignment is clearly the best option.

Figure 5.9 presents some singular cases to illustrate the limitations in convergence during tracking with Plücker based axis alignment. The selected samples are marked on the trajectory in the centre of the figure (the red cross corresponds to the estimated pose of the blue circle to the ground truth). Four cases correspond to relatively good estimates (S7, S9, S67 and S73), and the other four to poorer estimates (S3, S5, S69 and S71). For each case, a comparison of the silhouettes is shown at the initialization and at the end of the optimization process. The red border corresponds to the silhouette projected with the initial position and the green one represents the current frame (target position). The differences between the silhouettes are shown in blue. In most cases, it can be observed that at the end of the optimization, the superposition of the silhouettes is very good with

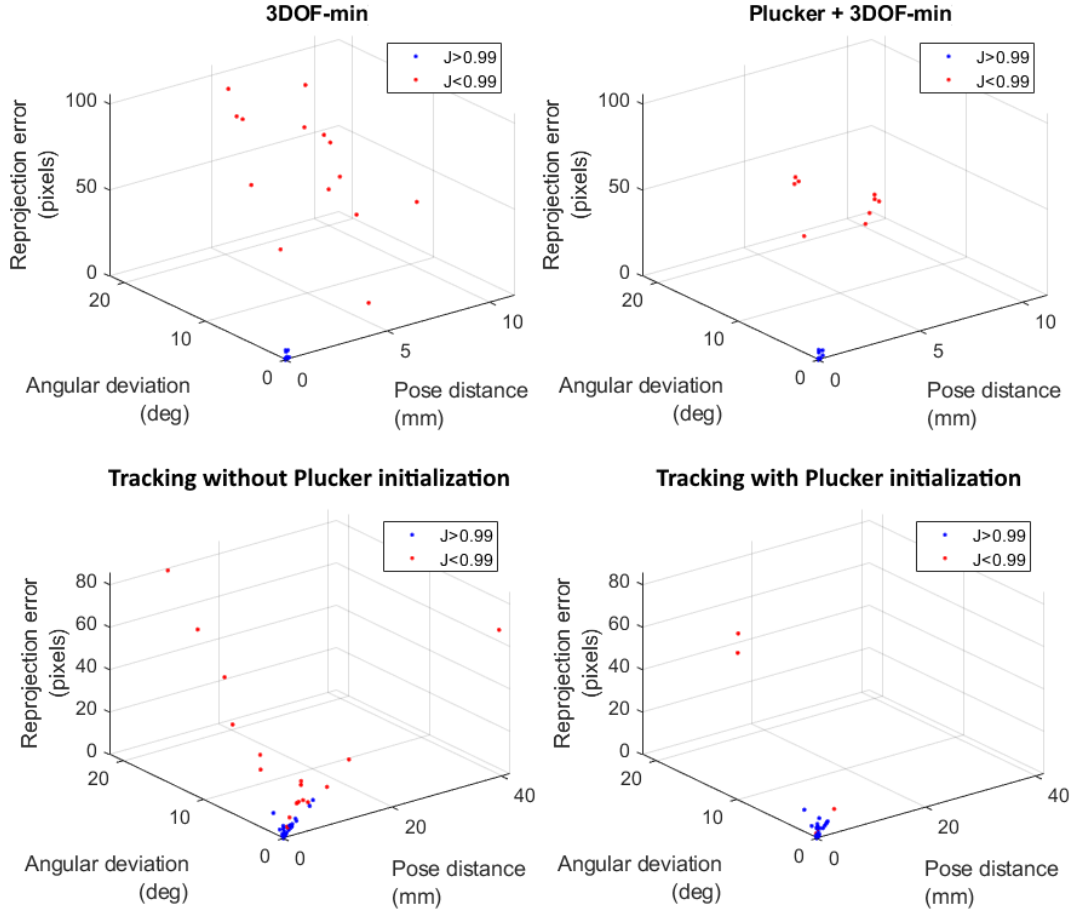


Figure 5.8: Distribution of error metrics according to Jaccard index greater or less than 0.99 with for the four variants: 3DOF-min (left-top), Plücker+3DOF-min (right-top), Tracking without Plücker (left-bottom) and Tracking with Plücker (right-bottom)

Jaccard index values higher than 0.99 except for S5 (0.962), which corresponds also to the worse estimate.

We observed there are two critical situations for our pose estimation method. First, when the silhouette of the fiberscope is truncated (cases S3 and S5): at least one of the straight edges of the cylinder is not visible enough to be correctly estimated and Plücker based axis alignment may then be defective. Sometimes, even when the initialization is close to the target, the minimization may not converge perfectly. For example, the final error for S3 remains small (less than 0.5 mm) but for S5 it is one of the largest errors in the trajectory (3.68 mm). This case happens if the tip of the fiberscope is not sufficiently inserted or with an tilting that brings it too close to the distal end of the endoscope, a situation that the surgeon spontaneously avoids during exploration. Also, the simulated trajectory pushes beyond the limits the possible tilting of the fiberscope (limited by the

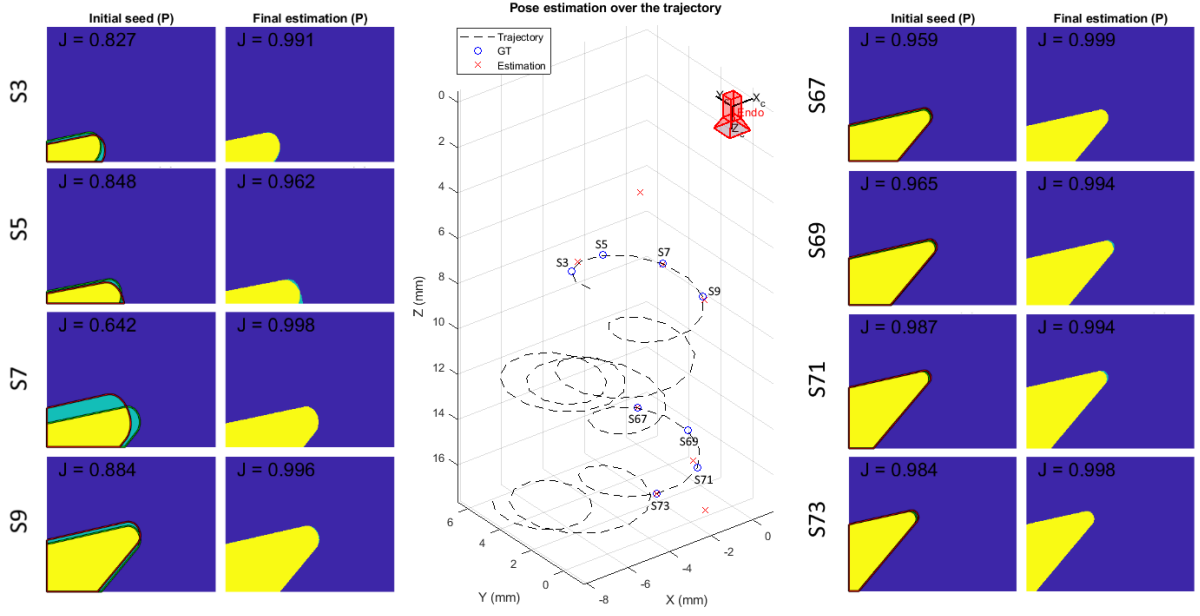


Figure 5.9: Some singular cases to illustrate the limitations of the method.

small gap between the diameter of the operating channel and that of the fiberscope) to test the method. Secondly, in the case of deep insertion of the fiberscope (beyond 14 mm), the discrimination potential of the Jaccard index is reduced by the perspective effect. Thus, even if the Jaccard index value is greater than 0.99, the distance error can be relatively high. This is the case for S69 and S71 with a Jaccard index value of 0.994, with distance error of 1.44 and 1.84 mm, respectively, but for S73 convergence gives a very accurate estimate.

Apart from these unstable conditions that are uncommon in actual exploration in patient, convergence is generally quite effective and the estimated pose accuracy is really good even when the starting point is far from the target (as in S7). The results indicate that our method can estimate the pose of the tip with an accuracy of less than 0.5 millimetre and the orientation of the optical probe with about 0.5 degree in a tracking scenario without considering the unrealistic extreme samples of this simulation. This is much better than what is reported by Xiongbiao et al. [102] for electromagnetic sensor tracking on a dynamic phantom (between 2.48 to 4.2 mm and 4.6 to 7.8 degrees), and in the same order of magnitude as vision tracking with photo-reflective markers reported in Shahidi et al. [103]. We also obtained better results than those reported in Mountney et al. [104] with vSLAM. Also, recent work confirms that the SLAM produces errors of 1.24 mm for the camera trajectories and RMSD of more than 2.5 mm for surface reconstruction [105].

Table 5.2: Estimated pose deviation on real images (mm)

Sample	Fiberscope's calibrated pose			Fiberscope's estimated pose			Euclidian distance
	x_{GT}	y_{GT}	z_{GT}	x_F	y_F	z_F	
1	-3.08	2.65	8.71	-3.08	2.65	8.71	0.0028
2	-4.89	2.70	9.12	-4.93	2.72	9.20	0.0992
3	-3.33	3.12	8.77	-3.33	3.12	8.77	0.0041
4	-4.39	2.33	9.06	-4.40	2.33	9.08	0.0244
5	-4.49	1.97	20.48	-4.46	1.96	20.32	0.1577
6	-4.71	3.47	9.27	-4.73	3.48	9.31	0.0416

5.4 Estimated pose error on chessboard pattern images

The experiments performed in the two following sections were performed using a set of corresponding HSI and RGB simultaneous acquired images of a 10x9 chessboard pattern with a grid resolution of 1 mm. We selected 6 fiberscope and endoscope image pairs and used the information from the chessboard pattern to estimate both cameras extrinsic parameters, this would be used as ground truth and will be addressed as calibrated pose. In addition, we estimated the fiberscope's pose using the RGB images and our pose estimation method using the instruments border information. Then we estimated the deviation between both measures using the euclidean distance.

The results are summarized in Table 5.2 and reveal a good accuracy between the estimated and calibrated pose with errors lower than 0.2 mm in all cases. In addition, the reprojection error produced by the registration using the estimated pose was always lower than 3 pixels. These numbers prove the reliability of the registration on real images while the plane of reference used as control points during the registration is properly estimated.

In addition, we measured the reprojection error of the registered images using the estimated fiberscope's pose. The calibrated pose and estimated pose of each case and their corresponding euclidean distance are presented in Table 5.3. An example of the resulting superimposed images, the transformed fiberoscopic image and the original RGB image are presented in Figure 5.10

5.5. PROPAGATION OF REPROJECTION ERROR BY PATTERN'S DISTANCE UNCERTAINTY

Table 5.3: Re-projection error as a function of estimated distance from plane of interest (pixels)

Sample	Pattern to camera distance (mm)	Reprojection Error	
		Mean	STD
1	18.02	0.68	0.27
2	18.68	1.30	0.72
3	18.66	0.74	0.36
4	19.02	1.57	0.78
5	35.04	2.75	1.20
6	19.60	2.78	1.55

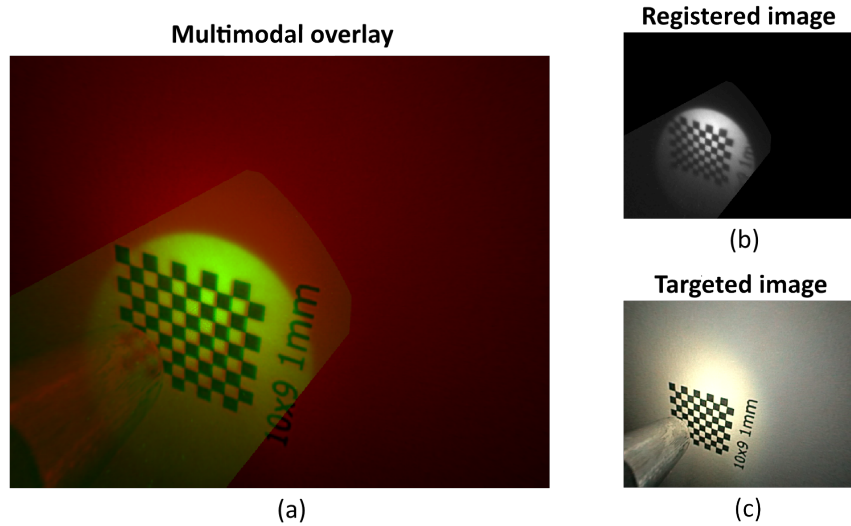


Figure 5.10: a) HS image from a chessboard pattern (green) registered to its corresponding RGB image (red), b) Transformed HS image, c) Original RGB image

5.5 Propagation of reprojection error by pattern's distance uncertainty

In this test, we measured how the misestimation of the pattern distance to the endoscopic reference frame affects the registration in terms of reprojection distance. During acquisition the doctor usually maintains a distance between 17 and 23 mm of the gastric wall with exceptions where the distance can go as high as 40 mm, but we cannot have a precise measurement of the actual distance of the gastric wall, which is why we require to

characterize the influence of this uncertainty.

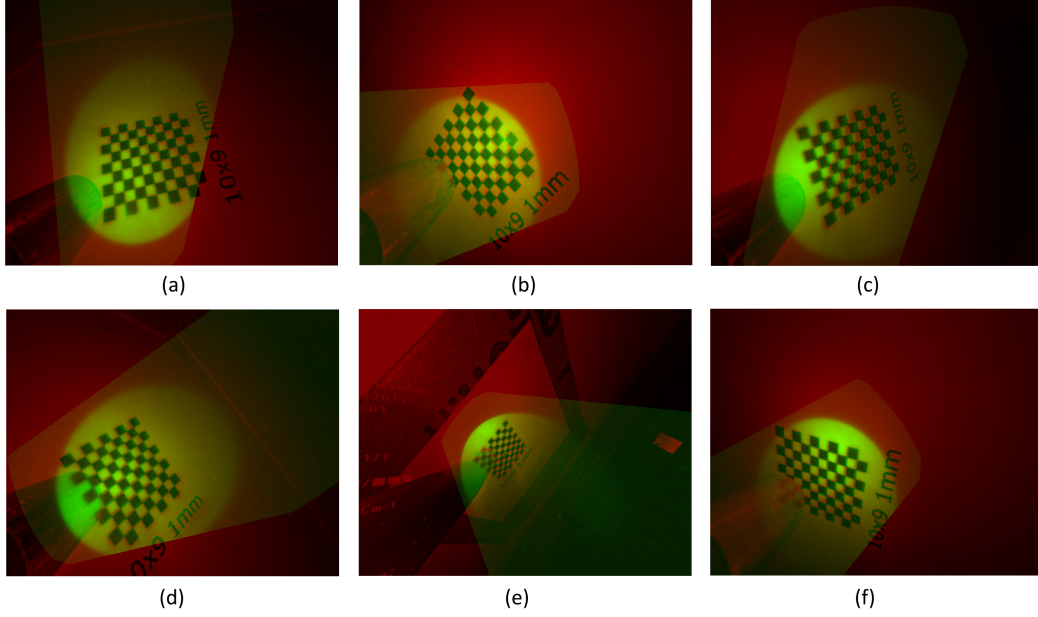


Figure 5.11: a,b,c,d,e,f) HS image from a chessboard pattern (green) registered to its corresponding RGB image (red). The images correspond to samples 1 to 6

We used the same sample as the previous section containing RGB and HS images of a chessboard pattern and used the registered images using the extrinsic values extracted from the patterns as reference (The complete set of registered images is depicted in Figure 5.11). Then, we projected the same pattern over the endoscopic camera varying its position over its z axis in a range of -5 and 5 mm with a step of 0.25mm, used the corners of the pattern as control points for registration and measured the Euclidean distance of the registered points vs the corners in the original image.

An example of the reprojected points and the original points at 3 different patterns positions (i.e., -2.5mm, 0mm and 2.5mm) is presented in Figure 5.12. In addition the propagation of error as a function of the distance from the center of the pattern is presented in Figure 5.13

Finally, we compare the coordinates of the pattern's corners corresponding to the registered HSI and original RGB through the euclidean distance. A summary of the pose values and final distance can be found in Figure 5.13. The results reveal a parabolic behavior as expected in which the mean error and standard deviation increase the farther the pattern is from its original position. It also increased faster when going closer to the camera than farther because of the distance perception of the camera. To maintain the reliability of the registration in acceptable limits and based on the results shown in Figure 5.13 we established the range of -4 to 3 mm as the trusted interval. This is, whenever the

5.6. CONCLUSION

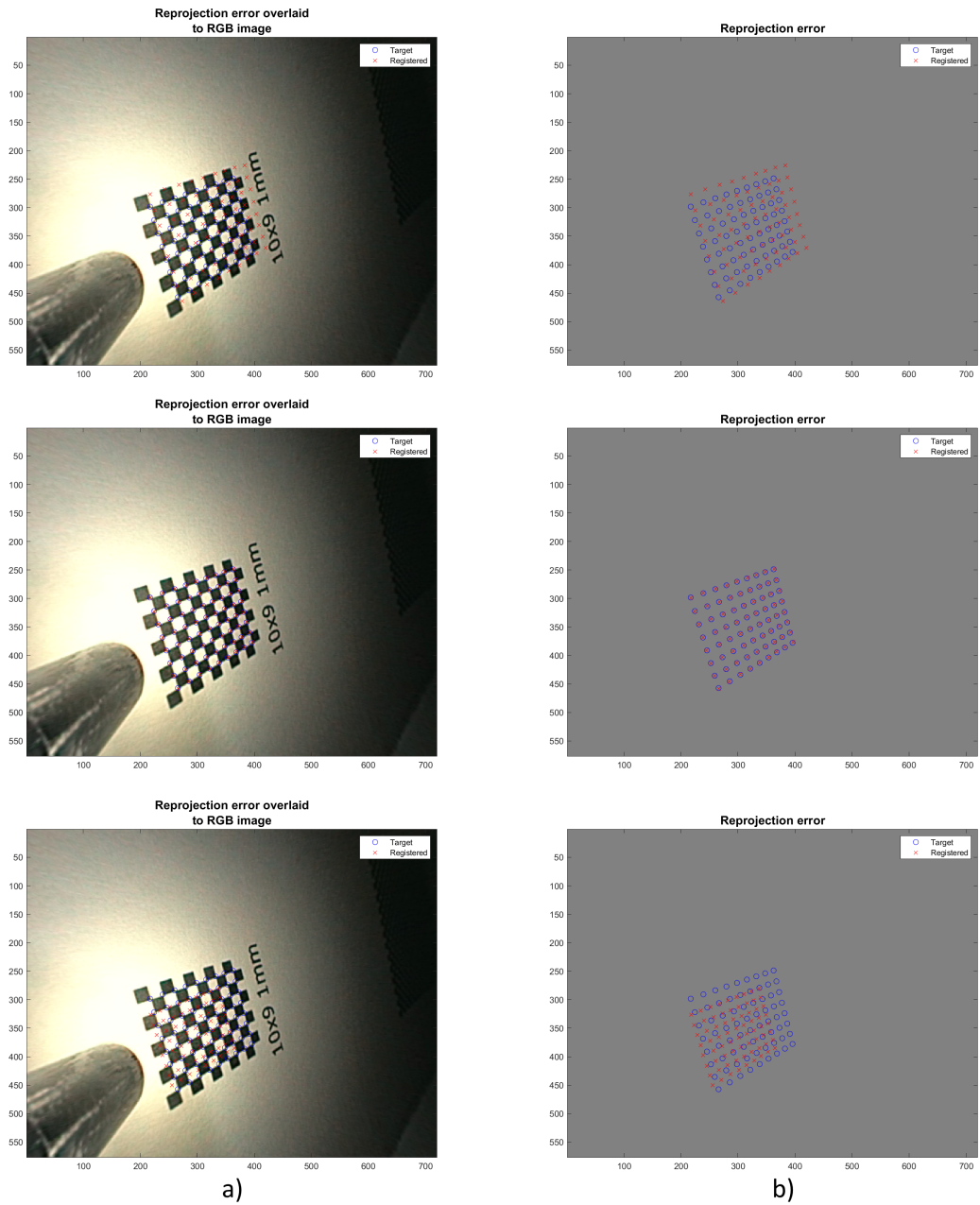


Figure 5.12: Reprojected points and the original points at 3 different patterns positions: a)-2.5mm b)0mm and c)2.5mm)

estimated distance from the gastric wall is on this range of uncertainty we can expect a mean reprojection error as high as 30 pixels)

5.6 Conclusion

We assessed the accuracy and precision of our virtual framework to emulate endoscopic images generated from a point cloud representing the fiber-optic instrument.

5.6. CONCLUSION

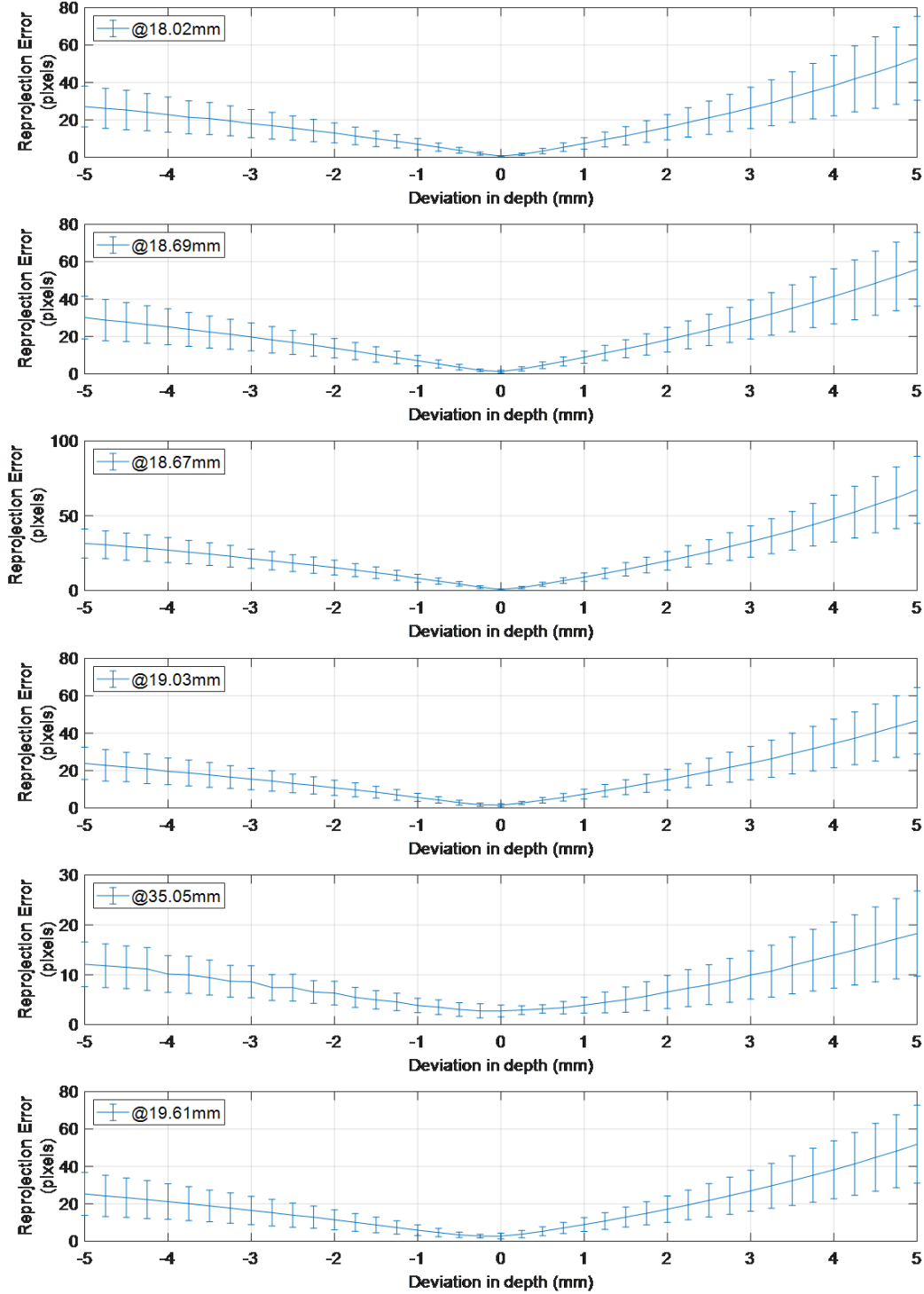


Figure 5.13: Re-projection error as a function of estimated distance from plane of interest in a range of ± 5 mm. The graphs correspond to samples 1 to 6 respectively

In the first two sections the tests revealed the robustness of the framework performance against variability in the measure of the instrument's apparent contour and geometry, accounting for mean values lower than 1.5 degrees and 0.6 pixels for angular deviation

and Euclidean distance respectively.

The third and forth sections assessed the pose tracking in different simulated scenarios. In both cases the Euclidean distance between the estimated and targeted pose and the angular deviation of the simulated cylinder were lower than three pixel,

The last two sections were focused in measuring the reliability of the estimates using real images of a chessboard pattern. The results revealed that the proposed method can effectively estimate the instrument pose when compared to the one obtained from the extrinsic parameters from a pattern image. Also we determined an interval of confidence for the misestimation of the distance from the endoscopic camera to the plane used for control point extraction for registration.

5.6. CONCLUSION

Chapter 6

Clinical application

In this chapter, we apply the algorithms previously introduced on real endoscopic images acquired during gastric screening procedures and the methodology followed for the multimodal registration with potential applications of HS-enhanced endoscopy in medicine.

The first section is focused in describing the methods used for the fiberoptic instrument segmentation on the endoscopic RGB images. We characterized the data base and present the results obtained while training two different convolutional neuronal networks. The second and third sections describe the characteristics of each modality and the procedure followed for their registration using the extrinsic values obtained from the modeled virtual framework. The last two sections are focused on the potential uses of the multimodal overlay during daily medical practice. This is the analysis of the localized hyperspectral data on independent frames and in a mosaic.

6.1 Segmentation of the fiber-optic instrument on RGB images

Contrary to the experiments presented in chapter 5 where we controlled the virtual framework or an image of a pattern to use as a reference, real life screening videos raise the limitations of a non referenced 3D space where is difficult to find corresponding features in the background (gastric wall). Therefore, to obtain the backprojection information of the instrument pose and orientation we require to outline its borders using the RGB image information.

There are a number of challenges that turn segmentation of surgical tools into a remarkably difficult task. Endoscopic scenes are heterogeneous environments that feature a high variety of optical interactions that complicate the segmentation (i.e.; tissue specularly, partial occlusions, blurriness (due to fast motion and/or parts of the image out

of focus), reflections (from instrument onto the patient’s tissue and vice versa) and body fluids. A possible option to ease the vision-based segmentation is to use fiducial markers like [106]. However, in the previous chapters we explained why we can’t use markers in our probe.

A prominent alternative in literature developed for instrument detection and tracking are vision-based methods [107]. Instrument background segmentation can be treated as a binary or instance segmentation problem for which classical machine learning algorithms have been applied using color and/or texture features [108, 85]. Later applications addressed this problem as semantic segmentation, aiming to distinguish between different instruments or their parts [91].

Moreover, deep learning-based applications to robotic instrument segmentation have demonstrated competitive performance in binary segmentation [109] and promising results in multiclass segmentation [110]. Also, modifications of deep neural networks that are suitable for use in mobile and embedded devices, including medical robots, are starting to emerge [111].

We trained two commonly used neural networks for segmentation and compared their results over our database to choose the optimal network for the initialization of the virtual instrument pose in our framework.

6.1.1 Acquisition and database

The acquisition protocol demanded to record the video endoscopy with white light and reflected light projected by a fibre in the operating channel. This reflected light was redirected to a hyperspectral camera for video acquisition. The examined sites were the gastric antrum and cardia if the latter presents an abnormality of position or contour (suspicion of Barrett’s oesophagus). Following this acquisition, gastric mucosa samples will be taken in the antrum and possibly in the cardia according to gastrointestinal endoscopy practice guidelines recommended by the Société Française d’Endoscopie Digestive (French Society of Gastrointestinal Endoscopy).

The database used to perform the test comprised 8 video acquisitions of different patients. The videos were captured at a rate of 25 images per seconds with a resolution of 720x576 pixels and they had an average of 55 seconds for a total of 10218 frames.

Our database was divided into 2 sets: Training-Validation and Test. The Training-Validation set correspond to two subsamples of the first four videos. The validation test comprises the last four videos. A detailed summary of the division of the data set and the corresponding number of frames for each one is presented in Table 6.1

Table 6.1: Distribution of the Training, Validation and Test datasets

Set	Number of frames				
	Seq1	Seq2	Seq3	Seq4	Total
Training	1107	2161	1647	245	5160
	Seq5	Seq6	Seq7	Seq8	Total
Test	1356	1203	2262	189	5010

6.1.2 Neural network training

A neural network is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates. The biggest advantage that Neural networks have is that it can adapt to changing input; so the network generates the best possible result without needing to redesign the output criteria.

Training a network consist in reality to learn the weights (between the neurons) by fitting the network to a training data. To learn the weights, it uses normally a backpropagation procedure. That means for each training data, we put it to the network which will calculate the output through some convolutional layers. At the output layer, the network computes the difference between the output value and the ground truth value as a loss. The network continues to put other data and update the weights until the loss function is optimized.

There are some optimization methods to minimize the loss like Stochastic gradient descent (SGD) between others. So, we usually need some parameters to configure the optimization method. In our application there are three main important set of parameters: Learning rate, momentum and weight decay

These parameters are related to the weight update equation. Normally, when we train a network from scratch (were all the weights are randomly initialized), we need a large dataset (like ImageNet) and set a high learning rate (for ex 0.01). After this process, we obtain a pre-trained model.

However, if we want to train again the network, for another application (for ex: contour detection, segmentation, semantic segmentation, etc) and on another dataset, we first load the pre-trained model (that means we load the weights from the pre-trained model instead of initializing randomly), then train and fit the network to the new dataset. This process is known as transfer learning. For example, in our case, when we trained RCF (which uses the VGG16 as backbone net), we loaded the VGG16 model pretrained on ImageNet for object detection to initialize the weight parameters and set: learning rate, momentum and weight decay after.

In our application we compared two of the commonly used networks: Deeplab and RCF

Deeplab

Deeplab [112] is a network which goal is to assign semantic labels (e.g., person, dog, cat and so on) to every pixel in the input image. Until now, this network has 4 versions: deeplab v1, v2, v3 and v3+. All the networks are always based on a backbone network (e.g. VGG, Resnet, Xception). We used deeplab v3+ because this version improves the accuracy of segmentation by determining an output stride. The effect of stride on the resolution is inversely proportional, this is if stride=2, the feature map decrease 2 times the width and height (resolution=1/2); if stride=4, resolution=1/4 and so on. If the resolution of feature map is too small: the prediction becomes fast, but less accurate. In contrast, if resolution is big, it needs more computations, but the segmentation accuracy will be improved. It's a trade-off between the speed and accuracy. We define firstly a fixed output stride (8 or 16) and after that, we adapt the atrous rates following the output stride.

The general parameter setup was: base learning rate 0.001 ; learning policy poly (that means the learning rate will be updated after each iteration, learning power 0.9, training number of steps 100k, train batch size 8, weight decay 0.00004 and crop size 513x513.

For the initial model, we used the Xception65 network which is pre-trained on PASCAL VOC dataset and for the ASPP parameters we used the deeplabv3+ structure with: output stride 16 (for encoder module); atrous rates [6, 12, 18] and decoder output stride 4.

Richer convolutional features

In RCF [113], features from all convolutional layers are well-encapsulated into a final representation which is amenable to training by back-propagation. As receptive field sizes of convolutional layers in VGG16 are different from each other, RCF endows a better mechanism than existing ones to learn multiscale information coming from all levels of convolutional features which are all pertinent for edge detection.

RCF is a versatile edge detection method which takes advantages of deep CNNs in combination with a more discriminative representation of complex data structures [113]. RCF encapsulates multiscale and multilevel information in a global way to perform the image-to-image prediction, amenable to training via backpropagation. The RCF network has been trained with binary masks from 54 manually segmented images. The training database has been extended to 864 images by applying horizontal, vertical, and combined

Table 6.2: Comparison between network performance in the Test set using Jaccard similarity measure as metric

		Jaccard			
		Seq5	Seq6	Seq7	Seq8
DeepLab	Mean	0.989	0.991	0.986	0.985
	STD	0.01	0.008	0.012	0.009
RCF	Mean	0.986	0.982	0.987	0.984
	STD	0.011	0.012	0.01	0.008

horizontal/vertical flipping and multiscale factors (1.0, 0.75, 0.5, 0.25) to the 54 original images. We keep settings proposed in the original paper (momentum 0.9, weight decay 0.0002, etc) except for the learning rate which has been fixed to $1e^{-8}$ instead of $1e^{-6}$. This value is important for the network to converge. Training is stopped after 40000 iterations.

To train RCF, we used SGD; other parameters for optimization procedure used were also: policy=step, gamma=0.1 and step size=10000. That means after 10k iterations, the learning rate decreases 10 times (for ex, at 0 iteration, learning rate = $1e^{-8}$; after 10k iterations, learning rate = $1e^{-9}$). To compute the loss, batch size=1 (how many images will be used for training at the same time), batch size was set to 1 to avoid the GPU memory problem), iteration size=10 (how many images will be used for calculating the loss), number of iterations to finish the training (normally, if we see the loss becomes stable, we can finish the training because the network is already converge).

6.1.3 Neural network deployment

Network comparison

We compared both networks over the Test data set using the Jaccard similarity measure. The ground truth consisted of a cylindrical approximation using our pose estimation method. This is, we roughly outlined the contour of the fiberscopic instrument in all images and then used this value as initialization for our maximization of similarity measure. The advantage of this method is that we can objectively compare the precision of the networks against a perfect cylindrical projection of the fiberscope.

The results for each set under evaluation are presented in Table 6.2

We observed that both implementations provided similar results in our Test set. However, DeepLab offers a better overall performance in 3 of 4 cases. In addition, the segmentation mask is directly obtained from the DeepLab network whereas in the RCF case we need to indirectly measure it from the network output which is a contour. Some examples

of the segmentation mask overlaid to their corresponding RGB images are presented in Figure 6.1

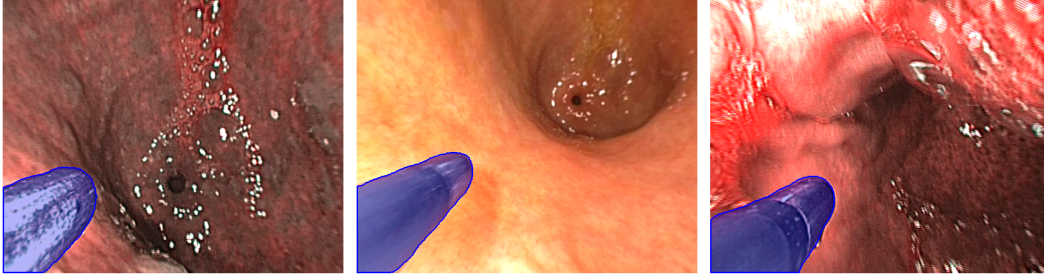


Figure 6.1: Example of a frames segmented using DeepLab.

Network limitations

The most challenging sequences were the ones that contained frames in which the fiberscope is in contact with the gastric wall. This produces reflection in the metallic part of the fiberscope and increased specularities due to the proximity of the fiberoptic light to the gastric wall. This problem occurred in several frames which did not converge correctly into a cylindrical projection and we decided to removed them from evaluation. Figure 6.2 illustrates an example of a removed frame.

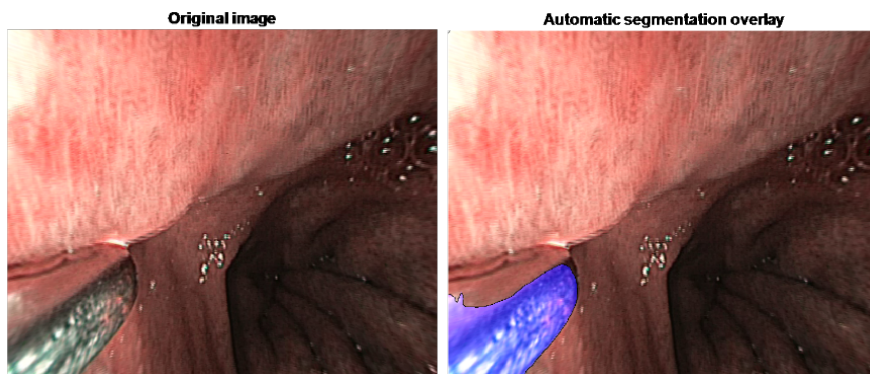


Figure 6.2: Example of a frame where segmentation fails due to the fiberscope's proximity to the gastric wall.

These images can be discarded for two reasons: first, the fiberscope is too close to the gastric wall, therefore the HSI will be oversaturated by the light emitted through the instrument itself, generating useless HS data. The second reason is that due to partial occlusion of the fiberscope tip or the highly reflective surface of the tip when proximal to the wall, the segmentation tends to fail

6.2 HS and RGB synergy

In the previous chapters we performed a registration between the two HS cameras allowing an exhaustive reflectance analysis of a single or group of points selected on the image. The HS spectral information and calibration of the camera allows us then to properly display all bands acquired through the fiberscope corresponding to a particular RGB frame (as seen in Figure 6.3).

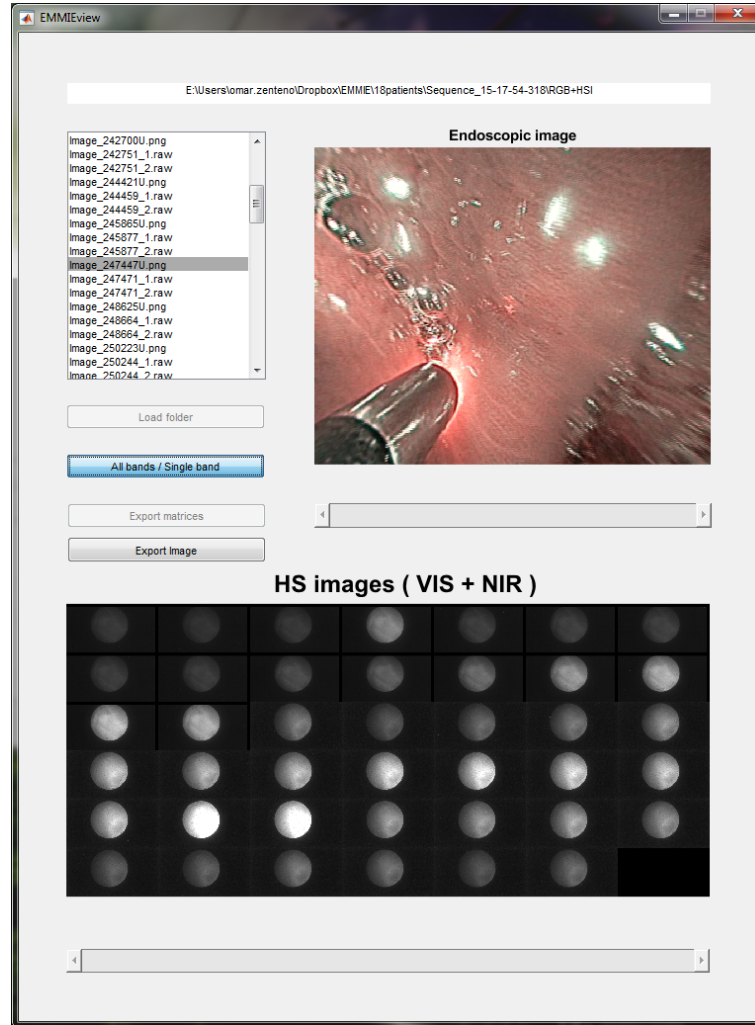


Figure 6.3: Interface to simultaneously view RGB and HS images

Also, we have estimated the relative pose between cameras and obtained their corresponding extrinsic parameters. If we try to portray graphically this process into the RGB image it could be shown as on Figure 6.4. We use the fiberscope instrument segmentation produced by the NN (a), then we perform the backprojection and minimization phase (b) and finally the pose is estimated (c). Figure 6.4c presents a projection of the virtual model depicting the accuracy of the modeling when compared to the RGB image on black

overlay.

We made use of the pose information and relative position between both cameras to perform an homography-based registration between modalities.

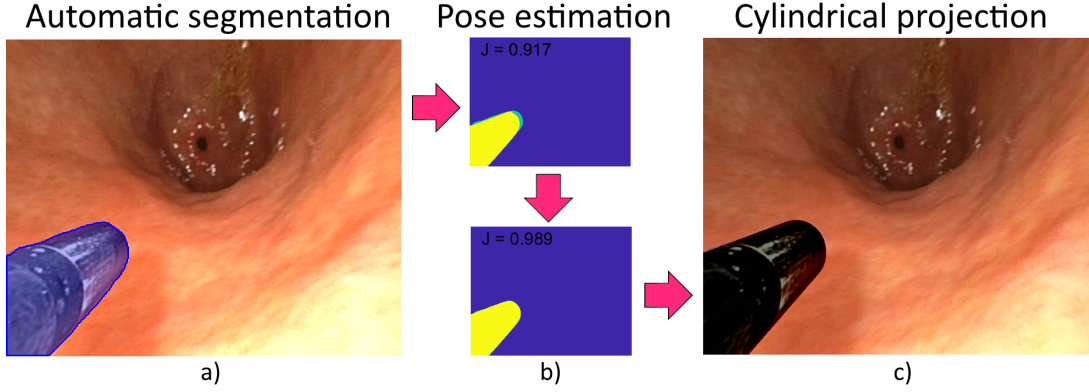


Figure 6.4: a) NN segmentation b) maximization of similarity c) virtual model projection

6.2.1 Enhanced endoscopy by image registration

Our image registration is performed in two steps. The first stage is in charge of registering the image data obtained from both XIMEA cameras while connected to the TwinCam system. The second stage is to register the complete Hypercube including VIS+NIR to the RGB image.

For the first part, the problem lies in the mirrored nature of the images captured by the TwinCam system. Since both modalities (VIS and NIR) are different in size and resolution and the fiberoptic image is not centered in the frame we cannot just mirror them. Nevertheless, both cameras are always fixed in the same position so the circle produced by the eye-cup connected to the TwinCam system can be used as control points for homographic registration. We segmented the circular capture zone using a level-sets segmentation and used random corresponding points on its circumference as control points.

Briefly, the planar homography which is a 3×3 matrix with 8 DoF relates the transformation between two planes (up to a scale factor):

$$s \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} \equiv H \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \equiv \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (6.1)$$

For the second stage, we already have the pose of both cameras and by extension their relative pose. To perform a scale-free registration we need to have a reference between

pixels and metric units. However it is really difficult due to septic practice to insert reference markers inside the patient.

In our virtual system, the endoscopic camera is our reference frame for two reasons. First, while scanning a suspicious area the instrument which moves more is the fiberscope, because the initial visual feedback is provided by the endoscope and after this the fiberscope is moved to target the point of interest. This produces a high variability in the insertion depth of the instrument while not being the case of the endoscope. Secondly, the FoV of the endoscope is higher than the fiberscope, with a bigger FoV the movements required to target a RoI are more subtle than with a smaller FoV like the fiberscope.

The average distance from the endoscope to the wall from the acquisitions in our data set was around $18\text{mm} \pm 4\text{mm}$. In the previous chapter we characterise the reprojection error when using a chessboard pattern depending of the estimated distance from the endoscope camera. The errors were acceptable under the same range of variability, therefore we decided to set a constraint of distance from the endoscope and the gastric wall to be 18mm. Its important to highlight that this constrain will affect the registration precision in a bigger manner the closer the gastric wall is to the camera than estimated.

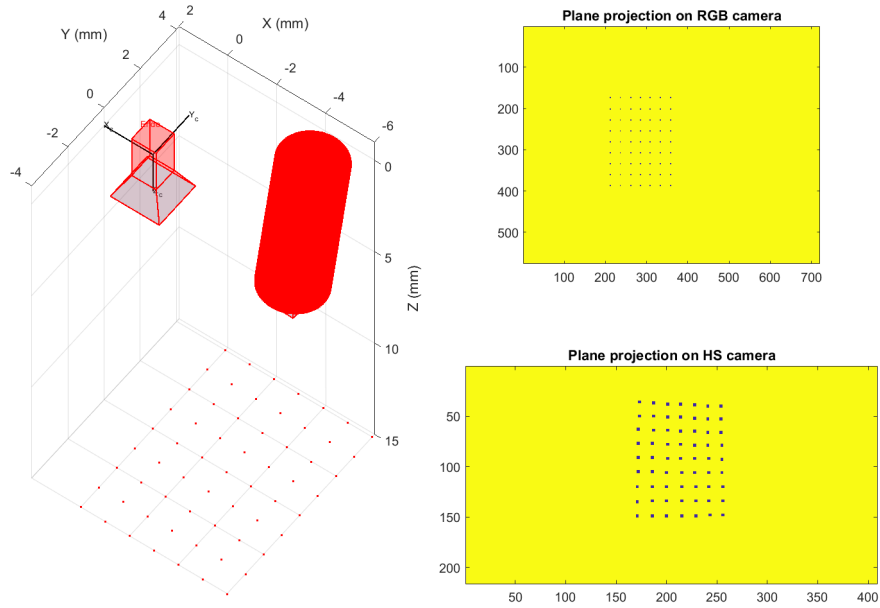


Figure 6.5: Example of a plane projection over a particular relative positions of cameras

We projected a 9×7 dot pattern with intervals of 1 mm in both dimensions. Then we projected virtual images of the patterns corresponding to each camera and used these points to generate a direct homographic information from the fiberscopic image to the endoscopic image. Process is depicted in Figure 6.5

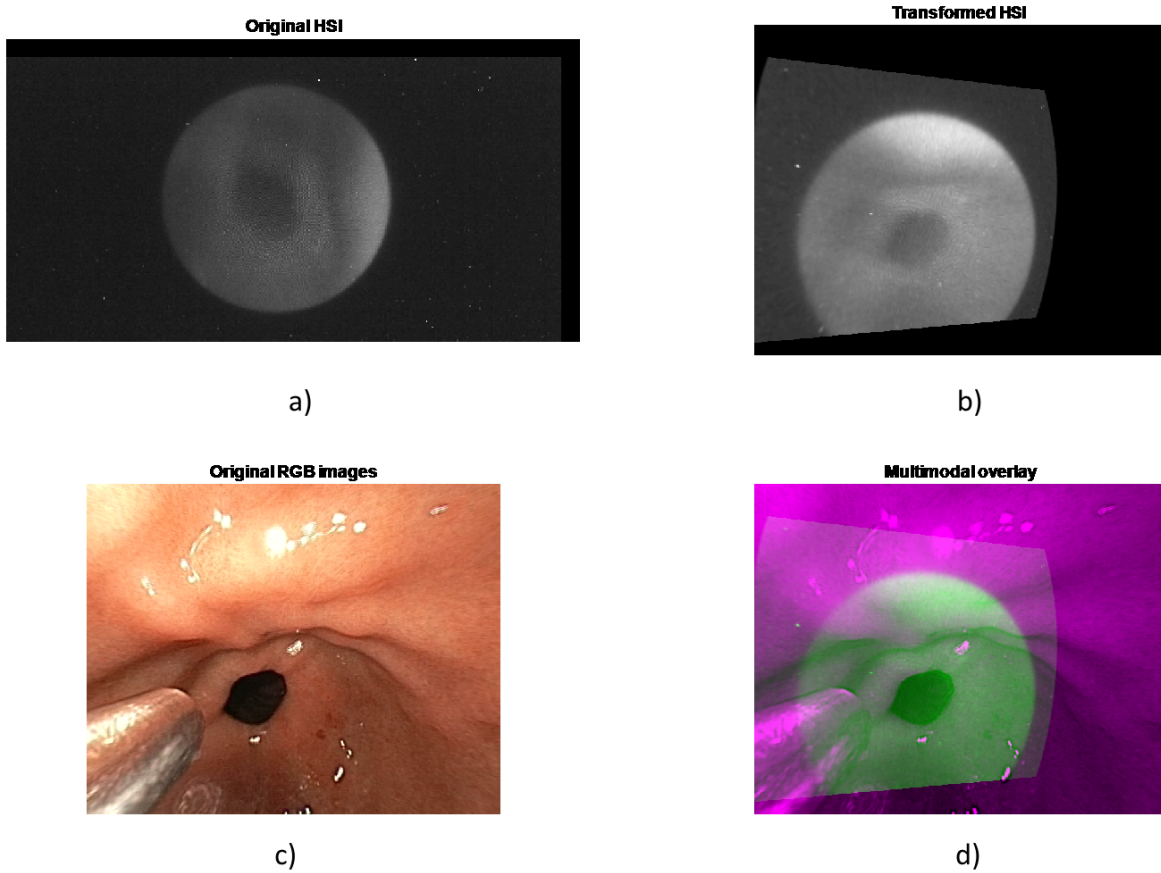


Figure 6.6: a) Original HS image, b) Transformed HS image, c) Original RGB image and d) Multimodal overlay .

The results were really encouraging, most of the images did fit into the estimated range without any scaling distortion effect. An example of a pair of images is presented in Figure 6.6. In this particular set we can observe how the features present in the HS images are overlaid to the same features presented in the RGB image. In the cases where the gastric wall was closer than estimated, there is an evident scaling artifact. Figure 6.7 depicts one of these cases where the projection of the HS image over the RGB is clearly bigger than the light projection of the fiberscope over the gastric wall.

6.3 Spectral bands content analysis from the HS images

The registration between modalities does not offer only visual information to the practitioner but also quantitative information of the tissues under screening. This data have the potential to be used for discrimination of tissue. However some processing is needed before its analysis.

The pixels in the registered images shown in the previous section represent the raw re-

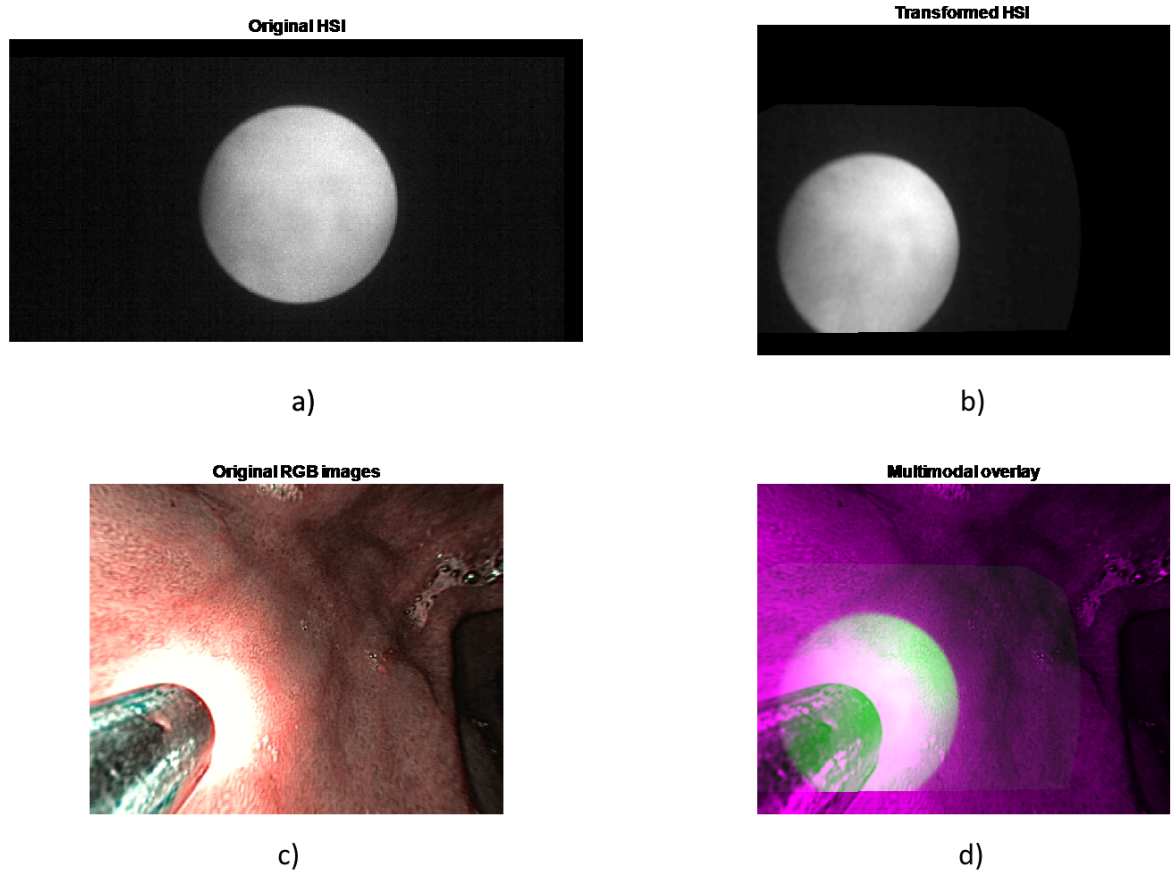


Figure 6.7: a) Original HS image, b) Transformed HS image, c) Original RGB image and d) Multimodal overlay .

flectance captured by each of the block from the acquiring matrix inside the sensor. These values still require to be corrected using the data obtained from the spectral calibration. An example of the data obtained from the raw images is presented in Figure 6.8.

To do this, we multiply the raw values by the correction function and generate a continuous spectra in the range of the bands the sensor is capable to capture data. However, not all pixels in the image are worth to be taken into account due to noise, saturation or overexposition. Therefore we selected the band with the highest brilliance on each image and performed a levelset segmentation to maintain only the pixels properly exposed. After this, the average spectra is presented as shown in Figure 6.9.

6.4 Biopsy location on the mosaic

Another potential application of the images could be to over-impose spectral data to a mosaic generated from the RGB images. This possibly offer several advantages over the analysis of single images. First the mosaic provides the clinician with a wider field of view

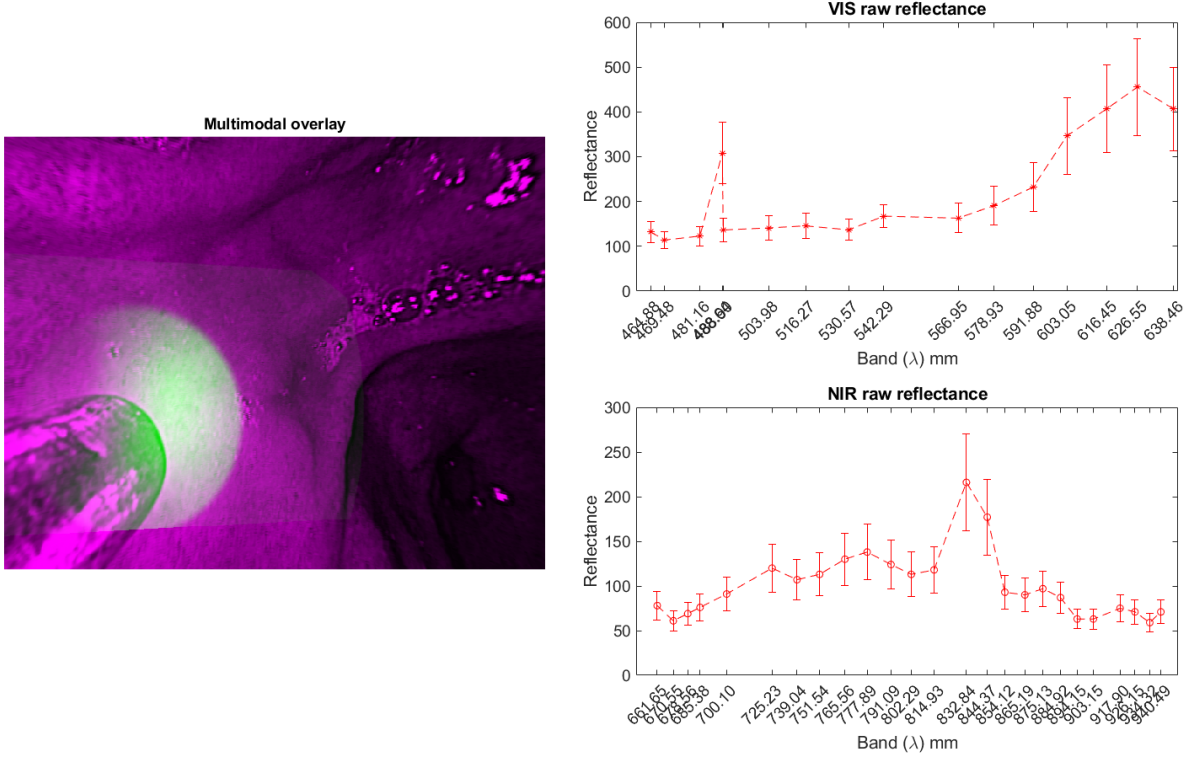


Figure 6.8: HS-RGB overlay and its corresponding raw reflectance data

of the analyzed area. Also, the doctor can choose to get HS data from localized areas while having visual feedback of all the scanned area. Finally, data redundancy can help to reduce variability in the estimated spectra. We worked in collaboration with the Centre de Recherche en Automatique de Nancy (CRAN) of the university of Nancy to generate the mosaics.

The method used is a variation of the method presented in [114]. And SfM-based pipeline for the construction of surfaces with few textures. The main contribution of this work lies in the integration of dense optical flow (DOF) fields in the SfM step for generating 2D point-groups under complex scene conditions.

The pipeline comprises five processing blocks:

- Pre-processing: This step selects the images which are minimally affected by specular reflections and motion blur [115]. The images are also undistorted using the optical centre coordinates and barrel distortion parameters estimated with a method dedicated to the short focal length lenses of endoscopes [116].
- SfM: This step provides both a sparse 3D point cloud (first representation of the surface to be recovered) and the camera pose parameters corresponding to the endoscope position and orientation for each image of set S . These results can only be

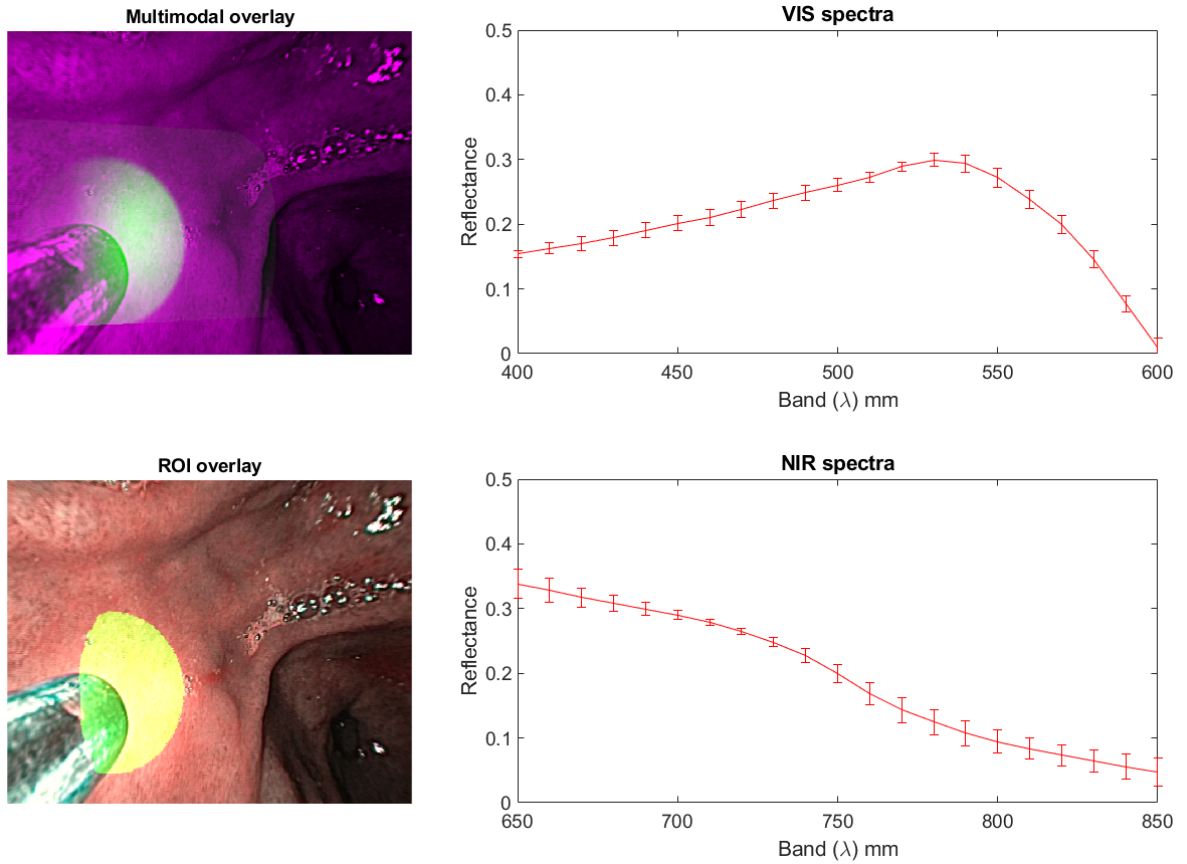


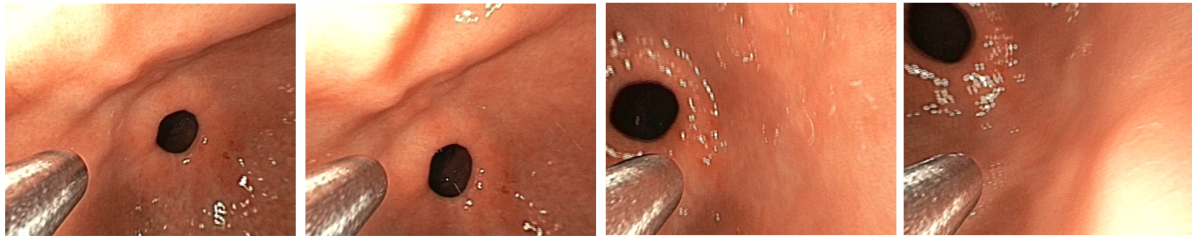
Figure 6.9: HS-RGB overlay and its corresponding reflectance data

obtained using groups of homologous 2D points (2D points of a group are all issuing from the projection of a same 3D point on the images of different viewpoints). These point-groups, together with a triangulation algorithm [43], are used for an initial estimation of the 3D points and camera poses which are all refined with a bundle adjustment technique [117].

- Dense reconstruction: This step uses the algorithm proposed in [118] to compute a dense point cloud by performing a completion of the surface represented by a sparse point cloud.
- Mesh generation: In this step, the mesh-generation algorithm given in [119] is used to pass from a dense point cloud to a surface approximated by small triangular facets.

Surface texturing: In this final step, the algorithm described in [120] is used to superimpose the textures seen in the images on the meshed surface.

The last three aforementioned steps are performed with the OpenMVS library [121].



a)



b)

Figure 6.10: Resulting mosaic image and the original images used

An example of the resulting mosaic and a subset of the images used for the procedure is depicted in Figure 6.10

Once the mosaic was produced we used the same homographic matrix applied to the original RGB images to the fiberoptic images and found the location of the center of the HS image in the mosaic. Figure 6.11 presents all the points where HS data is available. This provides the clinician with a higher degree of freedom to explore the gastric mucosa spectral property even off-line. Also any of the bands of the hyperspectral frames can be overlaid to the RGB image in the points shown. An example of this utility is presented in Figure 6.12 where two HS images from the points available were selected to display one of the VIS band values.

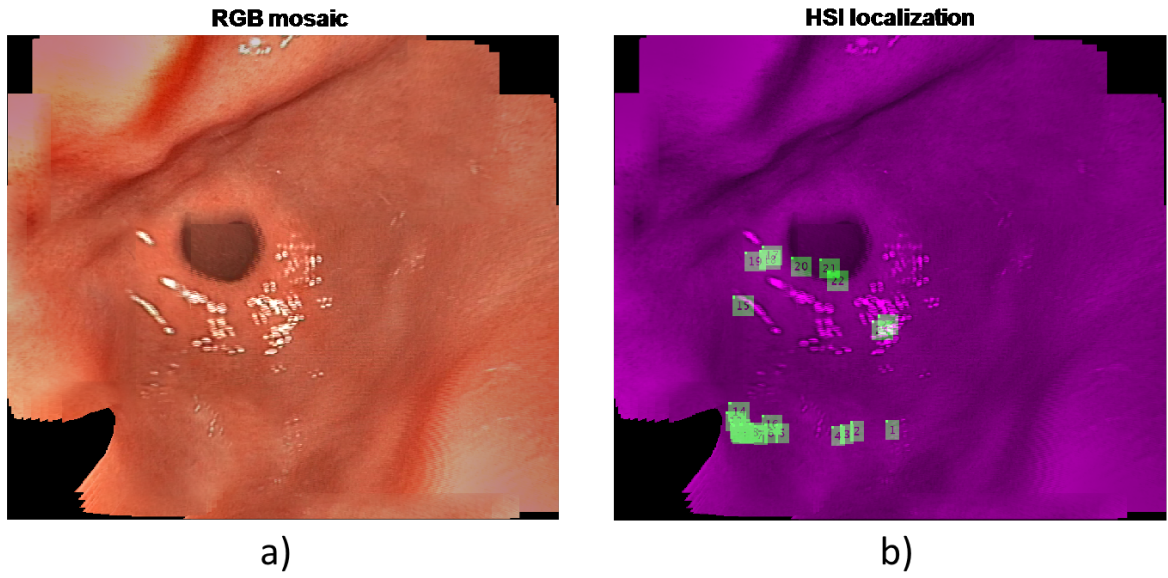


Figure 6.11: HSI localization into the RGB mosaic

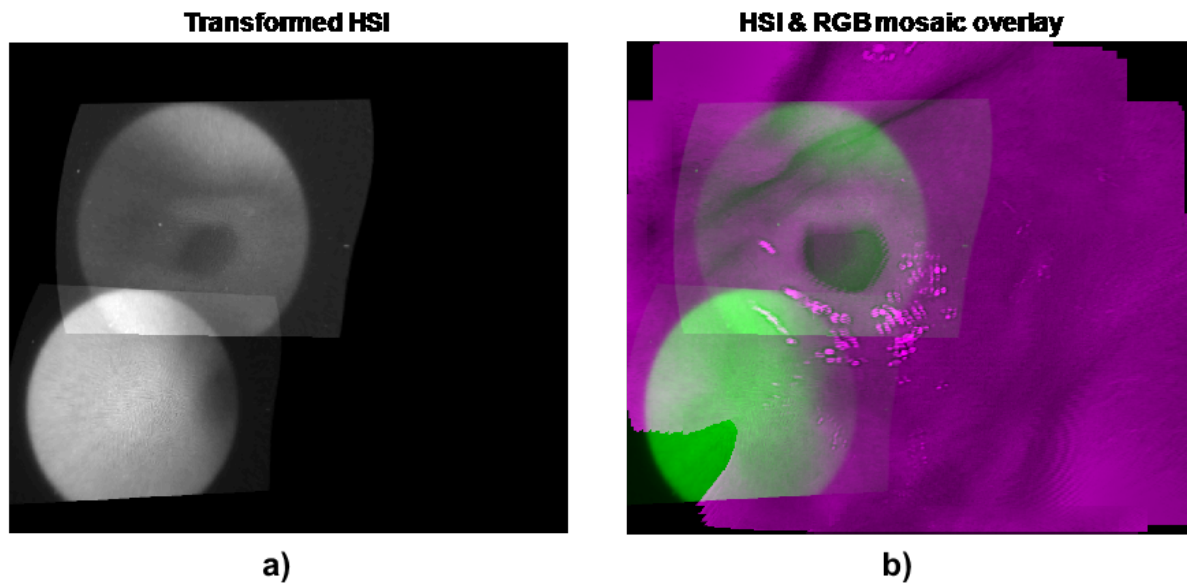


Figure 6.12: HSI overlaid to the RGB mosaic

6.5 Conclusion

We presented the implementation of the algorithms previously introduced on real endoscopic images acquired during gastric screening procedures and the methodology followed for the multimodal registration with potential applications of HS-enhanced endoscopy in medicine.

We compared the performance of two commonly used networks over our data set and

found that Deeplab had a better performance over RCF when outlining the instrument in the RGB images. Thanks to this we used the segmentation as initialization for the pose estimation procedure and were able to perform a registration between modalities under the constraints of a fixed average distance from the endoscope to the gastric wall.

We also, aboard the potential utilities of our results in clinical practice, showing that the spectra of areas under pathological suspicions could be analysed quantitatively to complement and enhance the current state of the art methods. Not only that, we did also shown the potential to build a HS mosaic over an RGB mosaic which will give also spatial information of the spectra in a easily visual distribution for the practitioners.

Conclusion and perspectives

The use of new multimodal imaging techniques to complement classical modalities have proved to be strongly promising for medical implementation. Gastroendoscopy exploration is no exception to this rule, image enhancement of RGB modalities are commonly used now in clinical practice with good results. Hyperspectral imaging has been used effectively in other medical domains, such as analysis of the skin. It is known that the light encodes chemical and architectural information from the tissue in a non-invasive way. In this sense, it could be used to provide quantitative information for objective diagnosis.

We have reviewed the benefits of the most used modalities in clinical practice. Chromoendoscopy and virtual chromoendoscopy, being the most widespread used, permit the detection and characterization of dysplasia in the gastrointestinal tract and identification of changes in the vascular system.

There is a strong interest in current clinical practice to improve diagnosis and visualization during gastroendoscopic screening procedures. Thanks to the development of new hyperspectral cameras, miniaturization of sensors and the capacity to handle high amount of data we have the opportunity now to characterize tissue in a quantitative way. We focused our efforts to provide an alternate measure of the gastric wall characteristics overlaid to endoscopic images, which is the common first-hand tool of clinicians.

Therefore the contribution of this study relays in the possibility to provide clinicians with a complementary measurement of the tissue which can be localized in the more familiar RGB images and potentially identify wavebands that could be used for differentiation between pathological and normal tissue. Up to our knowledge, we are the first to perform hyperspectral enhancement to gastroendoscopic images from the gastric mucosa using the complete spectrum. This is usually hardly achievable due to the constant movement of the stomach and we surpassed this limitation by using snapshot cameras, a newly introduced type of sensors to the HS imaging world.

We inspired ourselves in the NBI and Cellvizio systems to propose a prototype including a fiberscope connected to an HSI camera through the working channel of an endoscope. However, some limitations needed to be surpassed, including standardization of the various classification systems, incorporation of all these devices into a single easily

managed prototype, pose estimation of both cameras in the system, registration of both modalities and complete spectrum display.

Summary statement

MS-augmented endoscopic prototype

We introduced the MS-augmented endoscopic prototype and the elements it comprises. The system uses commercially available devices to assemble a complete solution. Also the experimental setup used for acquisition and the calibration of the optical system as a single multimodal image acquisition system. The spectral calibration of the system was performed to obtain reflectance information from the raw luminescence data captured from the sensors and the corresponding transformation function was applied before registration.

Prototype modeling and virtual simulation framework

We presented a virtual framework to generate synthetic data sets (i.e., point clouds and projections) of endoscopic images including a fiberscopic instrument tip into the scene. The movement of the cylinder is simplified to 3-DoF and is controlled by the direction of its axis using the interface between the endoscope and fiberscope as pivot points. This modelization allowed us to generate binary masks which mimic the outline of the instrument in the endoscopic images.

Pose tracking based on binary mask similarity

We presented a landmark-free approach for tracking the motion of a fiberscope inserted in the accessory/biopsy channel of the endoscope to reveal the location of optical biopsy onto the endoscopic image. The technique relies on the apparent contour of the fiberscope's tip in the endoscopic images considered as the projection of a 3D cylinder model to dynamically estimate the pose changes between the two cameras image-by-image. We controlled the virtual framework presented in chapter 3 to generate synthetic data sets to be used as input in the minimization function.

Pose validation assessment

We assessed the accuracy and precision of our virtual framework to emulate endoscopic images generated from a point cloud representing the fiber-optic instrument. The apparent contour on the backprojection of the cylinder axis proved to be robust to the variability in the contour input. We also assessed the pose tracking in different simulated scenarios. In both cases the euclidian distance between the estimated and targeted pose and the angular

deviation of the simulated cylinder were lower than one pixel. Finally we measured the reliability of the estimates using real images of a chessboard pattern. The results revealed that the proposed method can effectively estimate the instrument pose when compared to the one obtained from the extrinsic parameters from a pattern image. Also we determined an interval of confidence for the misestimation of the distance from the endoscopic camera to the plane used for control point extraction for registration.

Clinical application

We presented the implementation of the algorithms previously introduced on real endoscopic images acquired during gastric screening procedures and the methodology followed for the multimodal registration with potential applications of HS-enhanced endoscopy in medicine. We compared the performance of two commonly used neural networks over our data set and found that Deeplab had a better performance over RCF when outlining the instrument in the RGB images. We used the segmentation as initialization for the pose estimation procedure and were able to perform a registration between modalities under the constraints of a fixed average distance from the endoscope to the gastric wall. We also, aboard the potential utilities of our results in clinical practice, showing that the spectra of areas under pathological suspicions could be analysed quantitatively to complement and enhance the current state of the art methods. Not only that, we did also shown the potential to build a HS mosaic over an RGB mosaic which will give also spatial localization of the spectra in a easily visual distribution for the clinician. With our implementation the clinician can choose any of the points and analyse its spectra information.

Perspectives

The results from this study are promising and we consider that in the future, they could be used as a tool for exploration of the gastric wall characteristics. However, some actual limitations should be noted such as the small number of samples, the fact that this study included only patients from a single center, and the constraints of implementation using one specific manufacturer or the elements it comprises. With respect to the quantitative information, we consider making a better selection of the valid pixels to reduce the variability of the spectra (i.e., using the information of the spectra in all bands).

Some of the ideas to continue the developing cycle of our hypothesis are presented as follow.

Fiberscope axial rotation

The presented results are valid under some constraints. One of them is the assumption that the fiberscope axis does not rotate over itself while performing the exploration. This is not always true due the fact that we cannot control other movement than insertion in the fiberscope. Inevitably, the manipulation of the fiberscope during insertion and requirement of removal and sterilization after each patient due to aseptic practices makes replicating the same conditions of insertion each time hardly possible, specially if we are targeting a tracking without using markers.

We could use however tools found in the HS image to surpass these limitations. Due to the position of the light source in the fiberscope being not over the axis but displaced to the border of its circumference, the light projected over the gastric wall has always a deformation profile with an intensity gradient directed to the opposite side of the light source.

An exhaustive analysis over this phenomena and the use of image processing tools having into account the dynamic exposition time, which is the main challenge to to this, could allow us to have a consistent estimation of the fiberscope axis rotation

Depth estimation

The second constraint is the assumption of distance from the gastric wall to the endoscope reference frame. It is really difficult to obtain features from the specular and moving gastric mucosa. However, the same limitation was encountered before for the recostruction of 3D models using SFM. Indeed, the reconstruction pipeline proposed in [114] demonstrates the potential of this algorithm to find the relative position of a moving camera during endoscopic exploration. Even when the limitations of a scaling factor are still present the relative positioning of the cameras over a limited range of space (Stomachs are usually 15 cmm width in their wider parts) could help us estimate better the distance to the wall.

Spectral pathological discrimination

The spectral information has been normalized for display. This has the benefit to improve the easiness of interpretation of results due to the shape of the spectral function using relative magnitudes. However, for an exhaustive characterization of the tissue quantitative information is necessary. The main reason for the variability of the presented data is the dynamic exposition time required to acquire data due to the effect of illumination over the intensity values. Also, the selection of the pixels under analysis could be performed using information of the 3D reconstruction so we could avoid mixing textures in the selected ROI.

Real time processing

Finally, the experiments performed during this work has been performed using an experimental environment developed in MATLAB software. For its implementation in clinical practice an optimization of the procedures is required and its translation to a faster programming language which could allow us to perform the overlaid in real time allowing the doctor to have immediate feedback and leading to a better quality acquisition.

Final thoughts

We strongly believe that the tools presented in this work can generate the foundation for a minimally invasive exploration based on hyperspectral imaging to help in the diagnosis of pathologies during endoscopy.

The vision in the short term for this system is to be used as an additional tool that could help to highlight areas of tissues which are more likely to have precancerous lesions or inflammation, so the practitioner could avoid or reduce the number of biopsies extracted. In the long term, we believe in the potential of this technique to be extended to other modalities which could also be connected to a fiberoptic instrument.

Appendix A

Introduction (*traduction française*)

L'endoscopie est une technique médicale communément utilisée qui consiste à introduire un long tube fin et flexible équipé d'une lumière et d'une caméra pour examiner les organes internes d'une personne. La norme actuelle pour le diagnostic des maladies gastriques pendant l'examen endoscopique est une évaluation visuelle et des biopsies aléatoires ou ciblées. Cette approche présente des limitations pour détecter la maladie dans les premiers stades, lorsque la maladie est localisée sous la surface visible ou si elle ne présente pas de caractéristique visuelle pathognomonique. La majorité des endoscopes du commerce ne peuvent produire que des vidéos acquises en lumière blanche et les innovations récentes de ces dispositifs concernent essentiellement l'augmentation de la résolution de l'image couleur ou l'ajout de capacités de chromo-endoscopie sans colorant, comme l'imagerie à bande étroite NBI d'Olympus ou l'amélioration intelligente des couleurs chez Fujinon, par exemple. Ces techniques permettent d'analyser certaines zones suspectes en lumière blanche renforcée mais elles ne détectent pas les lésions telles que des gastrites inflammatoires ou des séquelles atrophiques. En outre, elles n'utilisent que les informations contenues dans les bandes visibles. Ces limitations justifient le développement de techniques d'imagerie innovantes pour répondre aux besoins de détection et de caractérisation précoces des maladies de l'œsophage et de l'estomac.

Dans cette thèse, nous proposons un système d'imagerie qui permet l'acquisition d'une vidéo composée de données hyperspectrales (HS) en plus de l'image couleur (RGB). Ce prototype a le potentiel de fournir de nouvelles informations pour la caractérisation des tissus tandis que le clinicien reste dans le cadre d'une exploration endoscopique conventionnelle. De nombreuses publications démontrent l'intérêt réel de combiner des informations spectrales pour la caractérisation des tissus gastriques à l'échelle macro ou mésoscopique [9, 10, 11].

Projet ANR EMMIE

C'est pourquoi, dans le cadre du projet de recherche collaboratif EMMIE (Endoscopie MultiModale pour les lésions Infflammatoires de l'Estomac) financé par l'ANR-15-CE17-0015, nous proposons un prototype d'endoscope amélioré par la technologie hyperspectrale. Ce système s'articule autour d'une sonde composée d'un fibroscope, d'une source d'éclairage et de caméras hyperspectrales pour équiper un endoscope du commerce.

La Figure A.1 illustre la répartition des tâches au sein du projet. Une analyse préliminaire sur un modèle animal a été effectuée pour caractériser le tissu gastrique par l'Institut Pasteur (Paris). L'acquisition d'images dans l'estomac humain a impliqué des patients de l'unité d'endoscopie de l'hôpital Ambroise Paré. Les analyses des échantillons prélevés serviront de diagnostic de référence.

La configuration matérielle du système, son étalonnage et son interface logicielle ont été conçus en collaboration avec le laboratoire Le2i (UMR CNRS 6306) de l'Université Bourgogne Franche-Comté. Les prétraitements, l'étalonnage spatial et le recalage des deux modalités ont été confiés au laboratoire PRISME de l'Université d'Orléans, tandis que l'étalonnage spectral et l'analyse hyperspectrale ont été confiés au Le2i. Enfin, le CRAN (UMR CNRS 7039) de l'Université de Lorraine avait la charge de construire une mosaïque 2D ou 3D d'images en utilisant les techniques de flot optique et de structure à partir du mouvement (SfM). Nous avons complété ces résultats en localisant l'information hyperspectrale dans la mosaïque.

Après acquisition et traitement, les données sont analysées en deux étapes pour obtenir différents niveaux de compréhension de la part du médecin :

- En ligne, une aide au diagnostic lors des procédures d'endoscopie en alertant le médecin lorsque l'analyse des images spectrales détecte des régions potentiellement inflammatoires.
- Hors ligne, une analyse plus détaillée de la réflectance spectrale, de la reconnaissance et de la reproductibilité grâce à la cartographie obtenue par mosaïquage d'images endoscopiques. L'intérêt est de donner un aperçu des lésions qui peuvent être réinterprétées lors des réunions de consultation médicale.

L'acquisition simultanée des deux canaux nécessite un alignement géométrique afin d'exploiter correctement les sources d'information complémentaires obtenues qui sont de dimensions et de résolutions différentes. Ce n'est pas une tâche triviale, car en raison des pratiques aseptiques lors des interventions chirurgicales, le fibroscope doit être stérilisé, ce qui oblige le médecin à l'enlever puis à réintroduire un nouveau dans le canal opérateur de l'endoscope à chaque examen. De plus, la texture de la paroi gastrique est pleine de

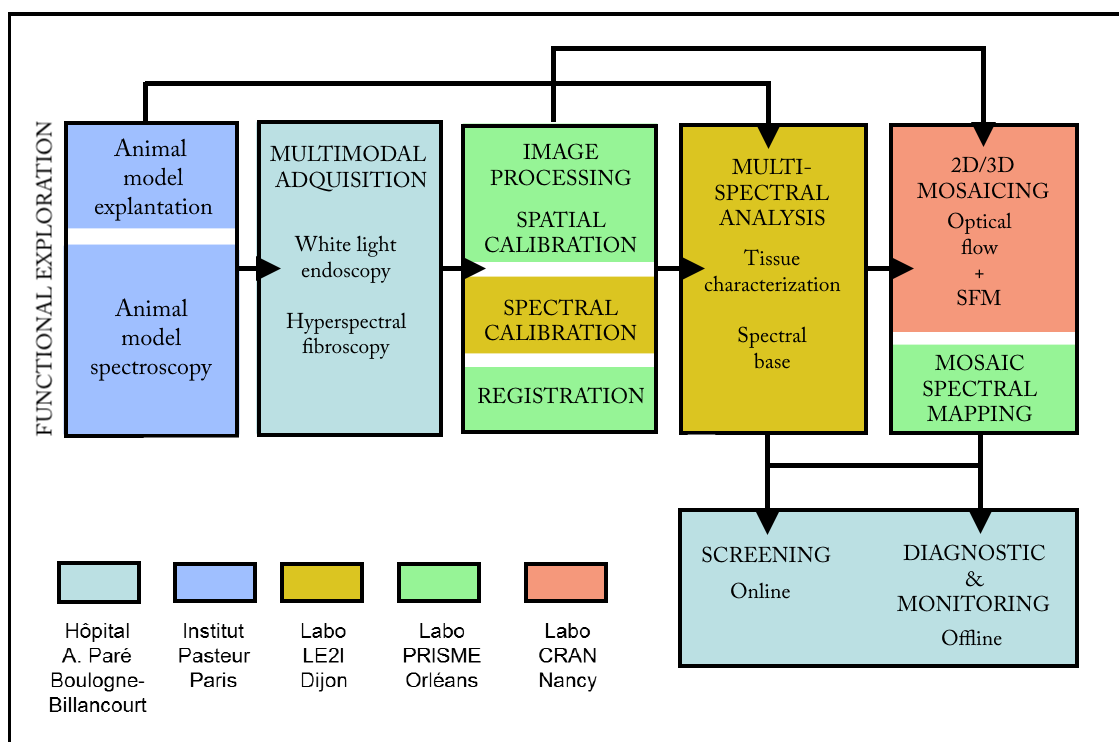


Figure A.1: EMMIE project block diagram

réflexions spéculaires et de liquide gastrique qui rendent difficile la tâche de trouver des caractéristiques communes pour un recalage d'images.

Contribution des travaux

Les trois principales contributions de ce travail sont : l'utilisation de caméras hyperspectrales « snapshot » pour effectuer le dépistage gastrique, une nouvelle approche de suivi sans marqueur de l'instrument dans l'image endoscopique et la localisation des données spectrales pendant l'analyse endoscopique.

Nous avons utilisé la présence de la sonde à fibres optiques dans les images couleur comme caractéristique de suivi. Nous avons estimé son mouvement en nous basant sur la mesure de similitude entre la projection d'un modèle virtuel et une segmentation automatique de l'instrument dans l'image endoscopique. Cette méthode est robuste face aux conditions qui entravent l'efficacité d'autres méthodes trouvées dans la littérature, comme par exemple, l'impossibilité d'utiliser des marqueurs sur la sonde ou l'acquisition avec une cadence d'images faible.

Le projet développe également une méthode de mosaïquage pour cartographier les zones ciblées, ce qui constitue une fonctionnalité supplémentaire pour le diagnostic. Grâce à notre méthode, nous pouvons localiser les données spectrales dans n'importe quelle

image RGB de la vidéo et améliorer l’exploration, ce qui permet aux cliniciens d’analyser les informations spectrales des zones localisées et de les traiter en temps différé. Cet aspect répond à une demande légitime du gastro-entérologue et de la communauté des anatomo-pathologistes.

Le prototype conçu au cours du projet EMMIE permet d’insérer dans le canal opérateur d’un endoscope un fibroscope connecté à des caméras hyperspectrales qui sont sensibles à la fois à la lumière visible et proche infra-rouge (NIR), qui fonctionne comme une sonde hyperspectrale localisée. Chose qui n’a jamais été mise en œuvre auparavant et qui ouvre donc une fenêtre d’opportunités pour l’exploration spectrale des tissus gastriques.

Organisation du manuscrit

Chaque chapitre de cette thèse contient une analyse bibliographique qui explore les travaux antérieurs et détaille davantage la logique qui sous-tend les expériences spécifiques présentées. La présentation des méthodes et résultats expérimentaux est suivie d’une discussion sur l’importance des résultats et l’identification des domaines de travail ultérieurs. Les principales conclusions de chaque section sont ensuite résumées. Voici un aperçu des chapitres de la thèse:

Le Chapitre 1 dresse un état de de l’art sur la biopsie optique et les systèmes actuels d’imagerie endoscopique multimodale ; le Chapitre 2 décrit les différents dispositifs qui composent le prototype de gastro-endoscopie hyperspectrale et l’étalonnage du système ; le Chapitre 3 détaille la modélisation du prototype et le logiciel de simulation cinématique développé ; le Chapitre 4 présente l’algorithme de suivi de l’instrument à fibres optiques proposé ; le Chapitre 5 évalue la validation du système sur des données synthétiques et réelles ; le Chapitre 6 présente l’application de notre méthode à des données cliniques. Le document se termine par une discussion finale et une analyse des travaux futurs dans la conclusion.

Appendix B

Résumé français

Suivi sans marqueur d'une sonde optique sous vision endoscopique : Vers une biopsie multispectrale

(Markerless tracking of a fiber-optical probe by endoscopic vision: Towards
multispectral enhanced biopsy)

Présentée par :

Omar ZENTENO

Sous la direction de :

Sylvie TREUILLET et Yves LUCAS

B.1 Introduction

L'endoscopie en lumière blanche (WLE) est la technique standard pour examiner à la fois les voies gastro-intestinales (GI) supérieures et inférieures, au prix d'un faible contrôle et d'un champ de vue limité. En général, l'endoscopie gastro-intestinale est associée à un faible traumatisme et des complications limitées. Cependant, elle utilise des biopsies incisionnelles avec une analyse histologique ex-vivo pour cibler les muqueuses pré-malignes, prendre l'évolution clinique ou suivre la progression pathologique.

Cette méthode d'évaluation retarde le diagnostic et risque d'introduire une erreur d'échantillonnage. De plus, l'absence de paramètres quantitatifs lors de l'inspection endoscopique entraîne souvent une variabilité importante inter-observateurs et intra-observateurs. Les récents progrès réalisés dans les fibres optiques, les sources de lumière, les détecteurs et la biologie moléculaire ont stimulé le développement de ces technologies d'imagerie

appelées biopsies optiques. Celles-ci effectuent une inspection non destructive de l'état histopathologique des muqueuses en utilisant la réponse des tissus à la lumière, ce qui permet une évaluation instantanée des tissus en alternative à la biopsie physique.

Nous proposons d'étendre les capacités de la WLE en insérant, dans le canal opérateur d'un endoscope, un fibroscope connecté à des caméras hyperspectrales (HS) dans les bandes du visible (WL) et du proche infra-rouge (NIR) en effectuant l'acquisition simultanée des deux flux vidéo (WLE et HS) et le recalage des deux modalités. Le prototype développé dans le cadre du projet ANR EMMIE va fournir aux médecins une aide au diagnostic par biopsie optique.

La précision de la méthode de localisation des biopsies a été évaluée sur un logiciel de simulation développé à cet effet et la chaîne complète des traitements validée sur des vidéos d'examen cliniques sur des patients.

Les principales contributions de ce travail sont:

1. Le développement et la caractérisation d'un prototype d'endoscope équipé d'une sonde hyperspectrale
2. Le suivi automatique sans marqueur de la sonde optique dans les images endoscopiques
3. Un recalage instantané des données spectrales sur l'image endoscopique RGB
4. La localisation des zones ciblées sur une cartographie partielle de l'estomac préalablement obtenue par mosaïquage des images extraites de la vidéo.

B.2 Caractérisation du système

Le prototype développé est un endoscope du commerce doté d'une sonde hyperspectrale (Figure B.1). Un fibroscope est introduit par le canal opérateur de l'endoscope et connecté via un séparateur optique à deux caméras (i.e., VIS et NIR). Le système est également équipé d'une source de lumière spécifique du fibroscope.

Le calibrage spectral du système est effectué pour obtenir des informations sur la réflectance à partir des données brutes de luminance mesurées par les capteurs. Le calibrage spatial démontre que le modèle optique de sténopé peut décrire efficacement le canal endoscopique et celui du fibroscope sans perte de précision par rapport à des modèles plus sophistiqués. Nous obtenons également les paramètres intrinsèques du système endoscopique sur la base du motif d'une mire en damier.

Notre prototype multimodal peut être considéré comme une paire de caméras non rigidement liées car le fibroscope est inséré dans le canal opérateur de l'endoscope pour

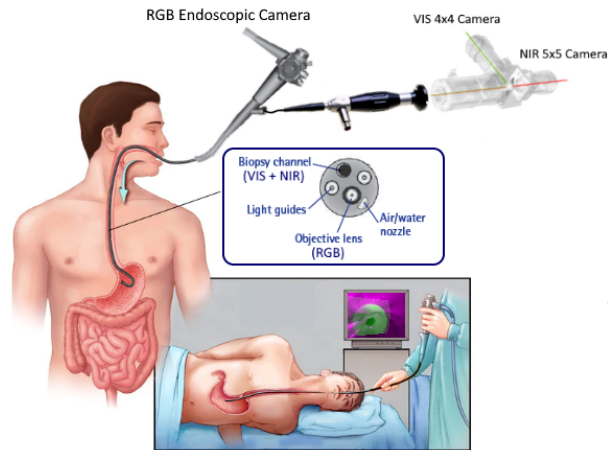


Figure B.1: Prototype d'endoscope augmenté d'une sonde hyperspectrale

chaque nouveau patient et peut glisser dans ce canal pendant l'examen. La pose relative entre les deux caméras s'apparente à un mouvement à six degrés de liberté. Cependant, en raison de notre configuration spécifique, le modèle cinématique entre les deux caméras peut être simplifié grâce aux deux hypothèses suivantes:

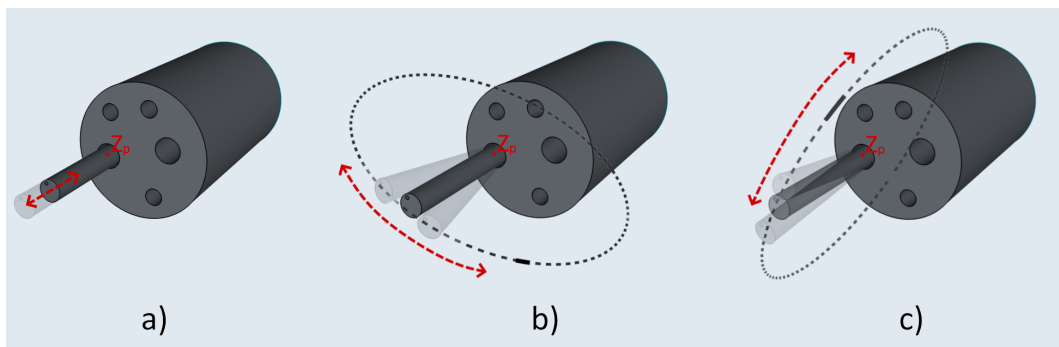


Figure B.2: Mouvements simulés du fibroscope : a) insertion (axe z), b et c) composantes de précession : lacet (axe y) et tangage (axe x)

- Le fibroscope est inséré dans le canal opérateur de l'instrument endoscopique, il peut facilement glisser longitudinalement mais son déplacement latéral est limité par le diamètre interne du canal ce qui permet d'introduire un point pivot fixe au niveau de l'ouverture du canal (Figure B.2).
- La rotation du fibroscope sur son propre axe est limitée car il est attaché au connecteur à l'autre extrémité. Mais il suffit d'exploiter la symétrie axiale parfaite et la rotation autour de son axe (roulis) ne pouvant pas être estimée sans l'utilisation de marqueurs sur la pointe.

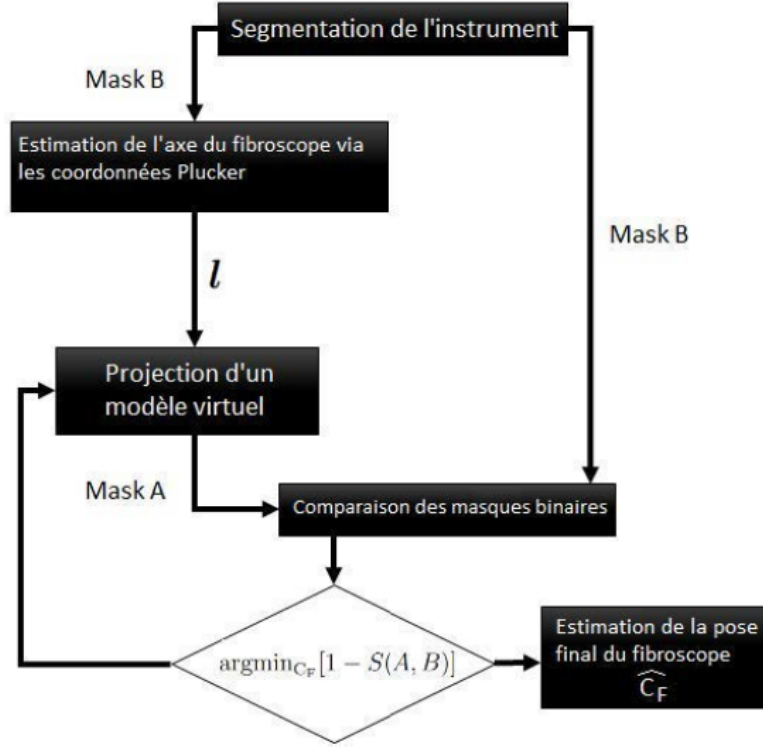


Figure B.3: Estimation de la pose des instruments

Sur la base de ces hypothèses, nous adoptons un modèle cinématique simplifié avec trois degrés de liberté pour décrire le mouvement relatif entre les deux caméras.

B.3 Estimation de la pose du fibroscope

Nous utilisons une méthode de vision basée modèle pour estimer la pose instantanée relative entre les 2 caméras [122]. Le fibroscope est un tube flexible de 2,5 mm de diamètre avec une extrémité distale rigide qui peut être modélisée comme un cylindre visible par l'endoscope. La méthode d'estimation de la pose s'effectue en plusieurs étapes présentées dans la Figure B.3 : Extraction de la silhouette de la sonde dans l'image endoscopique par apprentissage profond, estimation de l'orientation de l'axe 3D de la sonde fibroscopique par passage aux coordonnées de Plücker et la localisation de l'extrémité en maximisant l'adéquation de la silhouette extraite avec le modèle cylindrique.

La pose C_F est ensuite affinée de manière itérative en maximisant la similarité entre les silhouettes A et B. La pose optimale est obtenue après convergence.

$$\hat{C}_F = \text{argmin}_{C_F}[1 - S(A, B)] \quad (\text{B.1})$$

Table B.1: Comparaison des performances du réseau par mesure de similarité avec l'indice de jacquard pour métrique

		Jaccard index			
		Seq5	Seq6	Seq7	Seq8
DeepLab	Moyenne	0.989	0.991	0.986	0.985
	STD	0.01	0.008	0.012	0.009

B.3.1 Segmentation automatique du fibroscope

Nous avons utilisé le réseau Xception65 pré-entraîné sur la base de données PASCAL VOC. Pour l'entraînement du réseau, nous avons tracé le contour de l'instrument optique sur toutes les images afin de comparer objectivement la précision des réseaux par rapport à une projection cylindrique parfaite. Les résultats pour chaque sequence d'images évalué sont présentés dans le Tableau B.1 et quelques exemples de segmentation sont présentés dans la Figure B.4

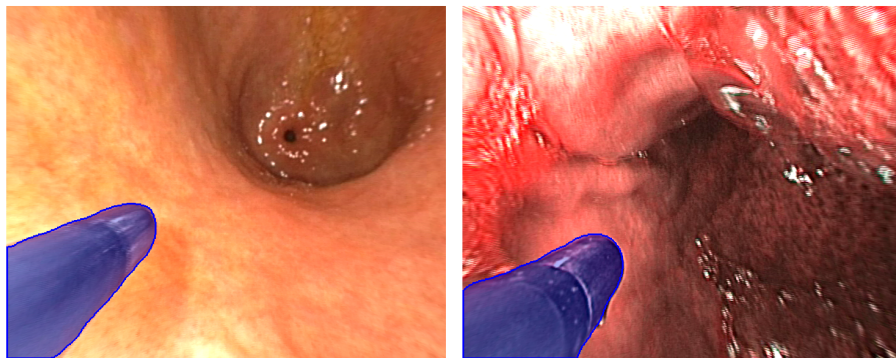


Figure B.4: Exemple d'un trames segmentées à l'aide de DeepLab

B.3.2 Estimation de l'axe du fibroscope via les coordonnées Plücker

A partir du contour de la silhouette du fibroscope obtenue par segmentation, nous extrayons les lignes génératrices du cylindre pour estimer l'orientation de son axe. Le processus se déroule en 3 étapes (voir Figure B.5) :

- Identifier le centre de masse de l'instrument et délimiter l'axe principal (pointillé bleu) de la segmentation pour évaluer seulement la moitié de l'instrument à la fois (bord supérieur et bord inférieur).
- Appliquer RANSAC aux points sélectionnés (vert) pour estimer la ligne qui correspond le mieux à tous les points (rouge) dans chaque cas. On obtient ainsi les deux lignes représentant les contours supérieur et inférieur de l'instrument

Apparent contour of the fiberoscopic instrument

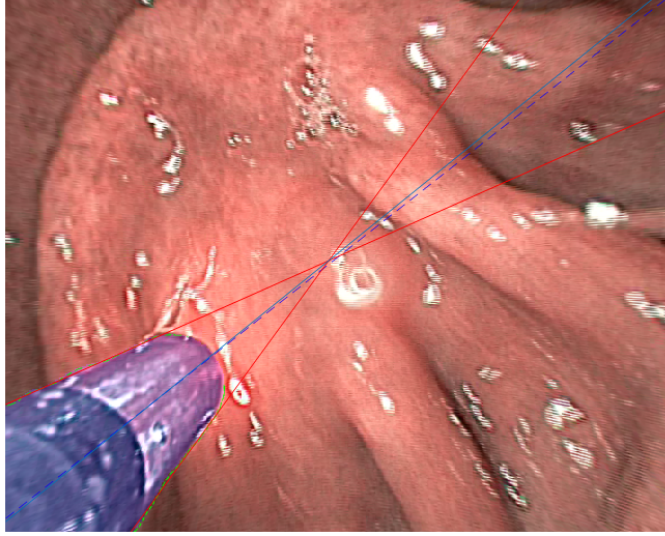


Figure B.5: Estimation des lignes de contour apparentes du fibroscope par RANSAC

- Estimer l'axe réel de l'instrument en exploitant les coordonnées de Plücker (Figure B.6).

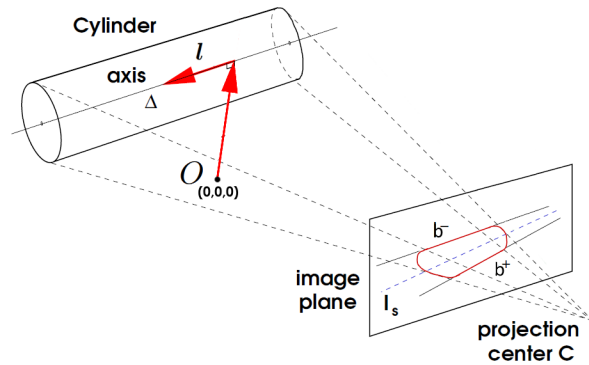


Figure B.6: Représentation par Plücker de l'axe du cylindre et sa projection dans l'image endoscopique (Adapté de [7])

On applique la méthode d'estimation proposée par [7]. Le rayon du cylindre et la matrice de calibrage de la camera K étant connus, la décomposition de la matrice $C = \mathbf{b}^-\mathbf{b}^+ + \mathbf{b}^+\mathbf{b}^-$ permet d'accéder aux coordonnées de Plücker (\mathbf{w}, \mathbf{l}) définissant l'axe du cylindre dans l'espace à partir des deux lignes génératrices projetées dans l'image.

$$K^{e^T} C K^e \equiv \begin{bmatrix} \mathbf{w} & \mathbf{z}_u & \mathbf{l} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 - \sigma^2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{w}^T \\ \mathbf{z}_u^T \\ \mathbf{l}^T \end{bmatrix} \quad (\text{B.2})$$

Où $\mathbf{z} = \frac{\sqrt{1-\alpha}}{\|\mathbf{w}\|} [\mathbf{l}]_x \mathbf{w}$ tandis que $\alpha = \frac{r_c}{\sqrt{\|\mathbf{w}\|^2 - r_c^2}}$ et le vecteur unité $\mathbf{z}_u = \frac{\mathbf{z}}{\sigma}$ tandis que $\sigma = \|\mathbf{z}\|$. Cependant, (\mathbf{w}, \mathbf{l}) ne représente que l'axe du cylindre et non pas sa profondeur d'insertion réelle, pour cela nous appliquons un processus d'optimisation de la pose par comparaison des masques binaires correspondant respectivement à la silhouette segmentée et à la projection d'un modèle géométrique de cylindre. La similarité entre les deux masques binaires est évaluée par l'indice de Jaccard qui présente un meilleur comportement de la fonction objectif comparé à Dice ou F-score

B.3.3 Recalage multimodal

Notre recalage d'images s'effectue en deux étapes:

- La première étape consiste à recaler les images des deux caméras XIMEA connectées au système de miroirs séparateurs Twin Cam. Le problème réside dans la nature différente des deux modalités (VIS et NIR) en taille et résolution et le fait que l'image issue de la fibre optique n'est pas centrée le recalage n'est donc pas un simple effet miroir. Toutefois, le cercle correspondant au champ de vue de la fibre optique peut être utilisé pour ce recalage géométrique. Nous avons segmenté la zone de capture circulaire par la méthode des level-set et nous avons utilisé des points correspondants aléatoires sur sa circonférence comme points de contrôle. On peut ainsi obtenir un hypercube complet des données (VIS+NIR) d'une résolution de 512x256 sur 41 bandes spectrales.
- Pour la deuxième étape, nous faisons l'hypothèse que la zone circulaire observée est quasi plane et à une distance approximative de un à deux centimètres des optiques pour estimer une homographie entre les deux images (WLE et HS) à partir de la pose relative des deux cameras.

B.4 Résultats expérimentaux

B.4.1 Test de simulation de suivi

Les séquences vidéos d'examens cliniques sur patients ne permettant pas de disposer d'une vérité terrain sur la position de la sonde, nous avons développé un logiciel de simulation pour valider la précision de la méthode de recalage.

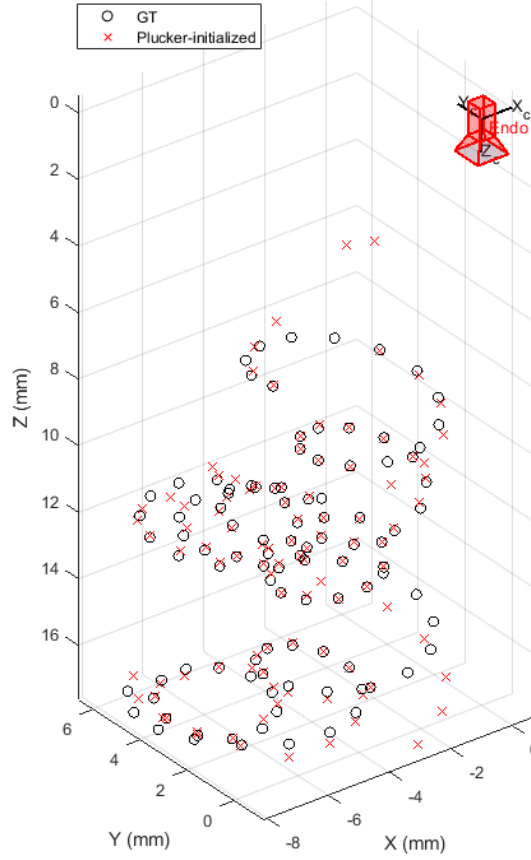


Figure B.7: Evolution de la pose du fibroscope (C_F) par rapport à la vérité terrain simulée sus une trajectoire de mouvement séquentiel de 100 iterations

La Figure B.7 présente la simulation d'une trajectoire quasi hélicoïdale avec des variations des trois degrés de liberté de mouvement (profondeur d'insertion et jeu selon les deux rotations) imitant le mouvement piloté par le praticien lors d'une exploration endoscopique : la plage d'insertion varie de 8 à 18 mm et les déviations angulaires sont circonscrites dans un disque de 4 mm de rayon.

Nous avons comparé quatre variantes pour l'estimation de la pose selon l'initialisation avant minimisation; un résumé des statistiques est présenté dans le Tableau B.2. La précision de la pose estimée est évaluée à partir de la vérité de terrain par la valeur moyenne, la médiane et l'écart-type de la distance euclidienne et de la déviation de l'axe optique sur la centaine d'échantillons de la trajectoire hélicoïdale. L'erreur de reprojection dans l'endoscopie due à la déviation de la pose par rapport à la vérité de terrain est estimée sur une grille virtuelle de taille 1,5 mm x 2 mm placée à une distance de 6 mm devant la sonde optique. La convergence est également évaluée par le nombre d'itérations et l'indice Jaccard final.

En comparant "3DoF-min" à "Plücker + 3DoF-min", nous pouvons observer que

B.4. RÉSULTATS EXPÉRIMENTAUX

Table B.2: Métriques d'erreur sur la trajectoire hélicoïdale (100 positions)

Méthode		Distance de pose (mm)	Déviati on angulaire (deg)	Erreur de Reproj. (pixels)	Nombre d'itérations	Index Jaccard final
3DoF-min	Moyenne	0.93	1.92	8.57	57	0.983
	Écart-type	2.43	5.11	22.50	14	0.047
	Médiane	0.02	0.01	0.14	59	0.999
Plücker + 3DoF-min	Moyenne	0.52	1.02	4.58	55	0.991
	Écart-type	1.67	3.24	14.34	14	0.030
	Médiane	0.02	0.01	0.10	55	0.999
Suivi sans alignement des axes	Moyenne	1.54	0.76	6.23	46	0.984
	Écart-type	4.42	3.05	14.14	15	0.039
	Médiane	0.05	0.02	0.31	45	0.999
Suivi avec alignement des axes	Moyenne	0.33	0.32	2.57	24	0.996
	Écart-type	0.68	1.73	9.43	13	0.007
	Médiane	0.11	0.02	0.43	20	0.998

l'application d'un alignement d'axe basé sur Plücker fournit des estimations de pose plus précises. La différence entre la médiane et la moyenne s'explique par la présence de valeurs aberrantes (voir Figure B.8), les points rouges correspondant à une convergence qui ne conduit pas à une valeur finale optimale de la fonction objectif (1-Jaccard). Les mesures d'erreur indiquent que l'initialisation après l'alignement des axes basé sur la méthode de Plücker donne des résultats nettement plus précis et plus robustes.

Lors de l'évaluation des expériences de suivi, l'alignement des axes basé sur la méthode de Plücker avant la tominimisation est encore plus efficace, plus précis et plus rapide (2 fois). Le nombre de valeurs aberrantes est considérablement réduit et les valeurs moyennes et médianes sont proches l'une de l'autre. Enfin, l'application de l'alignement des axes basé sur Plücker est clairement la meilleure option.

B.4.2 Tests de recalage sur une mire

Une mire observée simultanément par le fibroscope et endoscope permet d'estimer les paramètres extrinsèques des deux caméras utilisés pour obtenir la vérité terrain à comparer à la pose relative des caméras estimée par notre méthode. Les écarts sont résumés dans le Tableau B.3. Les résultats révèlent une bonne précision avec des erreurs inférieures à 0,2 mm dans tous les cas. En outre, l'erreur de réprojection du recalage à partir de la pose estimée est toujours inférieure à 3 pixels (Tableau B.4).

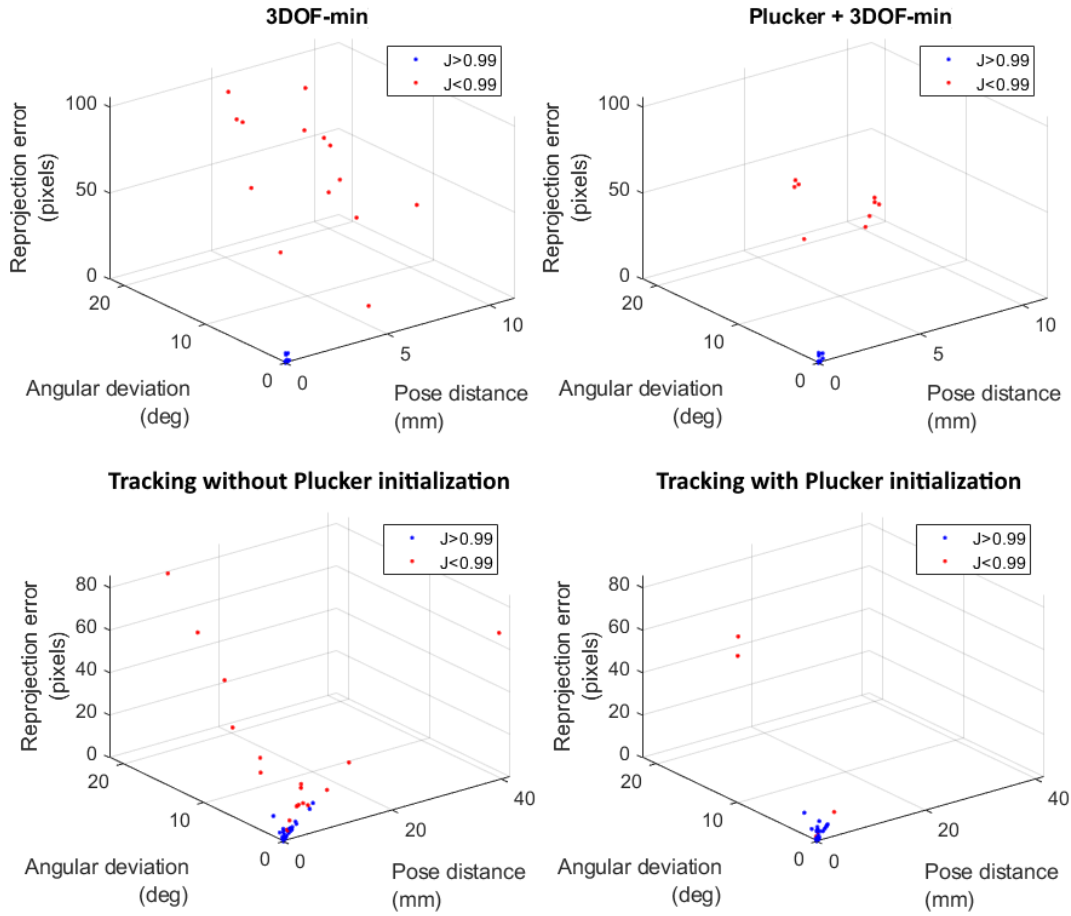


Figure B.8: Distribution des mesures d'erreur selon l'indice Jaccard supérieure ou inférieure à 0,99 avec pour les quatre variantes : 3DOF-min (en haut à gauche), Plücker+3DOF-min (en haut à droite), Tracking without Plücker (en bas à gauche) et Tracking with Plücker (en bas à droite)

Un exemple des images superposées résultantes, l'image transformée du fibroscope et l'image RGB originale sont présentés dans la Figure B.9.

B.4.3 Biopsie hyperspectrale

Le recalage entre les modalités offre au praticien une visualisation directe de la zone observée avec des informations quantitatives sur les tissus examinés. Ces données peuvent être utilisées pour la discrimination des tissus.

Un exemple de paire d'images est présenté en Figure B.10. Dans cette configuration particulière, nous pouvons observer comment les caractéristiques présentes dans les images HS sont superposées aux mêmes caractéristiques présentées dans l'image RGB.

Les spectres moyens sont présentés comme indiqué dans la Figure B.11. Les données

Table B.3: Estimation de l'écart de pose sur des images réelles (mm)

Sample	Pose calibrée du fibroscope			Pose estimée du fibroscope			Distance euclidienne
	x_{GT}	y_{GT}	z_{GT}	x_F	y_F	z_F	
1	-3.08	2.65	8.71	-3.08	2.65	8.71	0.0028
2	-4.89	2.70	9.12	-4.93	2.72	9.20	0.0992
3	-3.33	3.12	8.77	-3.33	3.12	8.77	0.0041
4	-4.39	2.33	9.06	-4.40	2.33	9.08	0.0244
5	-4.49	1.97	20.48	-4.46	1.96	20.32	0.1577
6	-4.71	3.47	9.27	-4.73	3.48	9.31	0.0416

Table B.4: Écart de rétroprojection en fonction de la distance estimée par rapport au plan d'intérêt (pixels)

Échantillon	Distance mire-caméra (mm)	Écart de rétroprojection	
		Moyenne	STD
1	18.02	0.68	0.27
2	18.68	1.30	0.72
3	18.66	0.74	0.36
4	19.02	1.57	0.78
5	35.04	2.75	1.20
6	19.60	2.78	1.55

brutes issues des caméras sont corrigées en utilisant la transformation établie lors du calibrage spectral. Seuls les pixels sans saturation sont sélectionnés pour calculer le spectre moyen.

B.4.4 Localisation sur la mosaïque

Nous avons travaillé en collaboration avec le Centre de Recherche en Automatique de Nancy (CRAN) pour générer les mosaïques panoramiques. La méthode utilisée est basée sur l'approche Optical-flow pour la construction de surfaces avec peu de texture. Une fois la mosaïque construite, nous avons utilisé la même matrice homographique que celle appliquée aux images RGB originales pour les images fibroscopiques et nous avons trouvé l'emplacement du centre de l'image HS dans la mosaïque.

La Fig B.12 présente tous les points où des données HS sont disponibles. Cela donne

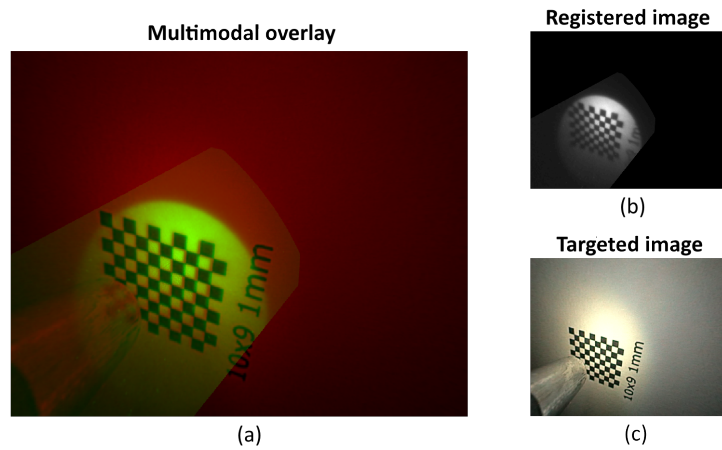


Figure B.9: a) Image HS d'une mire d'échiquier (en vert) recalée sur l'image RGB (en rouge), b) Image HS transformée, c) Image RGB originale

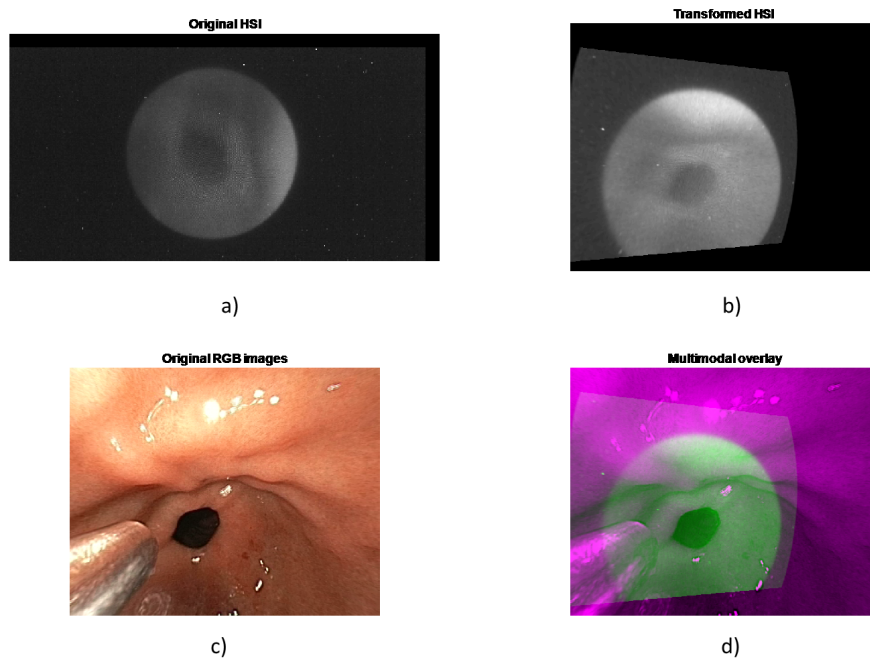


Figure B.10: a) Image HS originale, b) Image HS transformée, c) Image RGB originale et d) Superposition multimodale

au clinicien un degré de liberté plus élevé pour étudier les caractéristiques spectrales des muqueuses de l'estomac, même hors ligne.

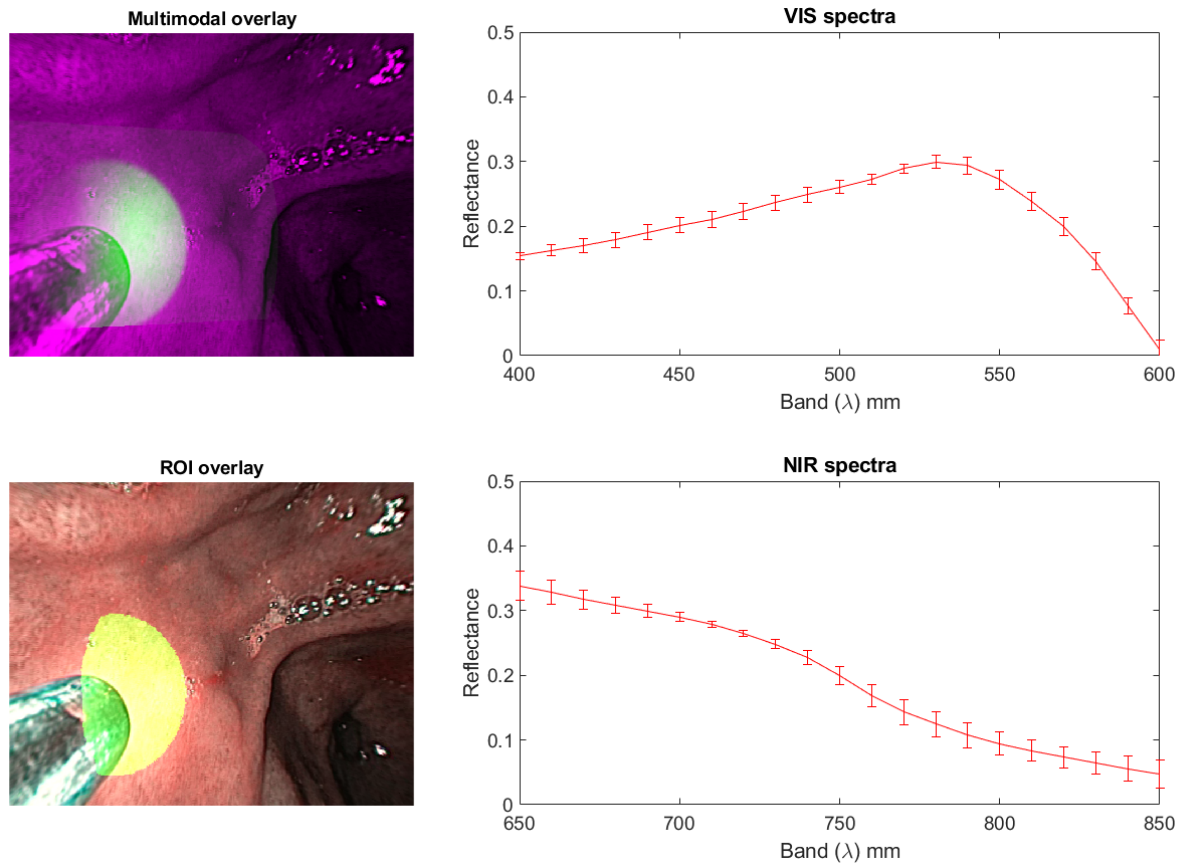


Figure B.11: Superposition HS-RGB et les données de réflectance correspondantes

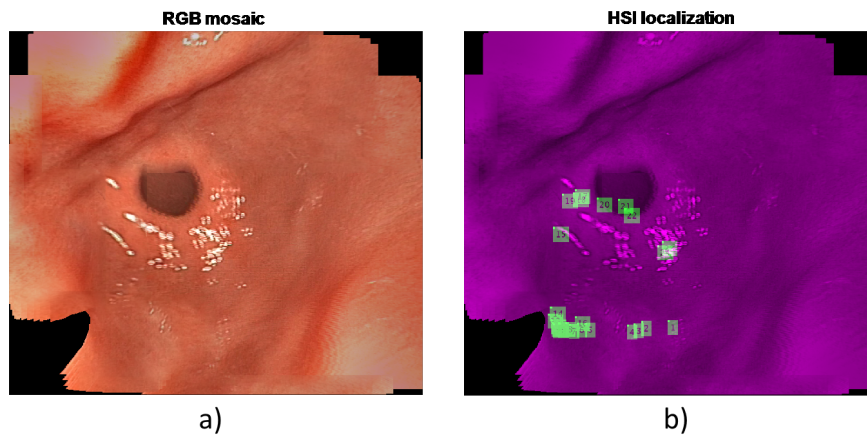


Figure B.12: Points de données HS dans la mosaïque 2D

B.5 Conclusion

L'amélioration du diagnostic et de la visualisation lors des procédures de dépistage gastro-endoscopique sont au cœur de l'évolution actuelle des pratiques cliniques. Grâce au développement de nouvelles caméras hyperspectrales, à la miniaturisation des capteurs

et à la capacité de traiter une grande quantité de données, nous avons maintenant la possibilité de caractériser les tissus de manière quantitative. Nous avons concentré nos efforts pour fournir une autre mesure des caractéristiques de la paroi gastrique superposée aux images endoscopiques, exploitée en routine par tous les cliniciens.

La contribution de cette étude permet donc de fournir aux cliniciens une mesure complémentaire des tissus qui peut être localisée dans les images RGB plus familières et d'identifier potentiellement des longueurs d'ondes qui pourraient être utilisées pour différencier les tissus pathologiques des tissus normaux. À notre connaissance, nous sommes les premiers à offrir une extension hyperspectrale des images gastro-endoscopiques de la muqueuse gastrique en utilisant le spectre complet. Cela est généralement difficilement réalisable en raison du mouvement constant de l'estomac et nous avons dépassé cette limite en utilisant des caméras instantanées, basées sur un nouveau type de capteur introduit dans le monde de l'imagerie HS.

Nous sommes convaincus que les outils présentés dans ce travail peuvent générer les bases d'une exploration peu invasive basée sur l'imagerie hyperspectrale par l'aide au diagnostic des pathologies lors d'un examen endoscopique.

L'exploitation clinique à court terme de ce système est celle d'un outil supplémentaire qui pourrait aider à mettre en évidence les zones de tissus qui sont les plus susceptibles de présenter des lésions précancéreuses ou une inflammation, de sorte que le clinicien pourrait éviter ou réduire le nombre de biopsies extraites. À long terme, nous pensons que cette technique pourrait être étendue à d'autres modalités qui pourraient également être reliées à un instrument à fibres optiques.

Bibliography

- [1] Merck and Co., Inc., Kenilworth, NJ, USA, “Viewing the digestive tract with an endoscope,” 2019. [Online; accessed October, 2019].
- [2] A. Breda, A. Territo, A. Guttilla, F. Sanguedolce, M. Manfredi, L. Quaresima, J. M. Gaya, F. Algaba, J. Palou, and H. Villavicencio, “Correlation between confocal laser endomicroscopy (cellvizio®) and histological grading of upper tract urothelial carcinoma: a step forward for a better selection of patients suitable for conservative management,” *European urology focus*, 2017.
- [3] O. R. Alharbi, N. S. Alballa, A. S. AlRajeh, L. S. Alturki, I. M. Alfuraih, M. R. Jamalaldeem, and M. A. Almadi, “Use of image-enhanced endoscopy in the characterization of colorectal polyps: Still some ways to go,” *Saudi journal of gastroenterology: official journal of the Saudi Gastroenterology Association*, vol. 25, no. 2, p. 89, 2019.
- [4] T. D. Wang and J. Van Dam, “Optical biopsy: a new frontier in endoscopic detection and diagnosis,” *Clinical gastroenterology and hepatology*, vol. 2, no. 9, pp. 744–753, 2004.
- [5] Olympus, *Gastrointestinal videoscope GIF-H185 spec sheet*, 2019.
- [6] Olympus, *Gastrointestinal videoscope GIF-1TH190 spec sheet*, 2019.
- [7] C. Doignon and M. de Mathelin, “A degenerate conic-based method for a direct fitting and 3-d pose of cylinders with a single perspective view,” in *Proceedings 2007 IEEE International Conference on Robotics and Automation*, pp. 4220–4225, IEEE, 2007.
- [8] H. Neumann, M. F. Neurath, and J. Mudter, “New endoscopic approaches in ibd,” *World journal of gastroenterology: WJG*, vol. 17, no. 1, p. 63, 2011.
- [9] G. Lu and B. Fei, “Medical hyperspectral imaging: a review,” *Journal of biomedical optics*, vol. 19, no. 1, p. 010901, 2014.

- [10] J. Yoon, J. Joseph, D. J. Waterhouse, A. S. Luthman, G. S. Gordon, M. Di Pietro, W. Januszewicz, R. C. Fitzgerald, and S. E. Bohndiek, “A clinically translatable hyperspectral endoscopy (hyse) system for imaging the gastrointestinal tract,” *Nature communications*, vol. 10, no. 1, pp. 1–13, 2019.
- [11] Z. Han, A. Zhang, X. Wang, Z. Sun, M. D. Wang, and T. Xie, “In vivo use of hyperspectral imaging to develop a noncontact endoscopic diagnosis support system for malignant colorectal tumors,” *Journal of biomedical optics*, vol. 21, no. 1, p. 016001, 2016.
- [12] M. Bruno, “Magnification endoscopy, high resolution endoscopy, and chromoscopy; towards a better optical diagnosis,” *Gut*, vol. 52, no. suppl 4, pp. iv7–iv11, 2003.
- [13] G. Zonios, R. Cothren, J. M. Crawford, M. Fitzmaurice, R. Manoharan, J. D. Van, and M. S. Feld, “Spectral pathology,” *Annals of the New York Academy of Sciences*, vol. 838, pp. 108–115, 1998.
- [14] S. L. Jacques, “Optical properties of biological tissues: a review,” *Physics in Medicine & Biology*, vol. 58, no. 11, p. R37, 2013.
- [15] T. Wolters, T. H. van der Kwast, C. J. Vissers, C. H. Bangma, M. Roobol, F. H. Schröder, and G. J. van Leenders, “False-negative prostate needle biopsies: frequency, histopathologic features, and follow-up,” *The American journal of surgical pathology*, vol. 34, no. 1, pp. 35–43, 2010.
- [16] C. H. Tae, J. H. Lee, B.-H. Min, K.-M. Kim, P.-L. Rhee, and J. J. Kim, “Negative biopsy after referral for biopsy-proven gastric cancer,” *Gut and liver*, vol. 10, no. 1, p. 63, 2016.
- [17] A. C. Croce and G. Bottiroli, “Autofluorescence spectroscopy and imaging: a tool for biomedical research and diagnosis,” *European journal of histochemistry: EJH*, vol. 58, no. 4, 2014.
- [18] A. P. Dhawan, B. D’Alessandro, and X. Fu, “Optical imaging modalities for biomedical applications,” *IEEE reviews in biomedical engineering*, vol. 3, pp. 69–92, 2010.
- [19] G. Galloro, “High technology imaging in digestive endoscopy,” *World journal of gastrointestinal endoscopy*, vol. 4, no. 2, p. 22, 2012.
- [20] J. E. East, J. L. Vleugels, P. Roelandt, P. Bhandari, R. Bisschops, E. Dekker, C. Hassan, G. Horgan, R. Kiesslich, G. Longcroft-Wheaton, *et al.*, “Advanced endoscopic imaging: European society of gastrointestinal endoscopy (esge) technology review,” *Endoscopy*, vol. 48, no. 11, pp. 1029–1045, 2016.

- [21] S. Beg and K. Ragunath, “Image-enhanced endoscopy technology in the gastrointestinal tract: what is available?,” *Best Practice & Research Clinical Gastroenterology*, vol. 29, no. 4, pp. 627–638, 2015.
- [22] L. M. W. K. Song, D. G. Adler, J. D. Conway, D. L. Diehl, F. A. Farraye, S. V. Kantsevov, R. Kwon, P. Mamula, B. Rodriguez, R. J. Shah, *et al.*, “Narrow band imaging and multiband imaging,” *Gastrointestinal endoscopy*, vol. 67, no. 4, pp. 581–589, 2008.
- [23] S. J. Chung, D. Kim, J. H. Song, M. J. Park, Y. S. Kim, J. S. Kim, H. C. Jung, and I. S. 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, pp. 136–142, 2010.
- [24] M. A. Manfredi, B. K. A. Dayyeh, Y. M. Bhat, S. S. Chauhan, K. T. Gottlieb, J. H. Hwang, S. Komanduri, V. Konda, S. K. Lo, J. T. Maple, *et al.*, “Electronic chromoendoscopy,” *Gastrointestinal endoscopy*, vol. 81, no. 2, pp. 249–261, 2015.
- [25] W. L. Curvers, R. Singh, L. W.-K. Song, H. C. Wolfsen, K. Ragunath, K. Wang, M. B. Wallace, P. Fockens, and J. Bergman, “Endoscopic tri-modal imaging for detection of early neoplasia in barrett’s oesophagus: a multi-centre feasibility study using high-resolution endoscopy, autofluorescence imaging and narrow band imaging incorporated in one endoscopy system,” *Gut*, vol. 57, no. 2, pp. 167–172, 2008.
- [26] T. S. Kirtane and M. S. Wagh, “Endoscopic optical coherence tomography (oct): advances in gastrointestinal imaging,” *Gastroenterology research and practice*, vol. 2014, 2014.
- [27] M. Goetz, A. Watson, and R. Kiesslich, “Confocal laser endomicroscopy in gastrointestinal diseases,” *Journal of biophotonics*, vol. 4, no. 7-8, pp. 498–508, 2011.
- [28] S. V. Panasyuk, S. Yang, D. V. Faller, D. Ngo, R. A. Lew, J. E. Freeman, and A. E. Rogers, “Medical hyperspectral imaging to facilitate residual tumor identification during surgery,” *Cancer biology & therapy*, vol. 6, no. 3, pp. 439–446, 2007.
- [29] L. C. Cancio, A. I. Batchinsky, J. R. Mansfield, S. Panasyuk, K. Hetz, D. Martini, B. S. Jordan, B. Tracey, and J. E. Freeman, “Hyperspectral imaging: a new approach to the diagnosis of hemorrhagic shock,” *Journal of Trauma and Acute Care Surgery*, vol. 60, no. 5, pp. 1087–1095, 2006.

- [30] R. Martin, B. Thies, and A. O. Gerstner, “Hyperspectral hybrid method classification for detecting altered mucosa of the human larynx,” *International journal of health geographics*, vol. 11, no. 1, p. 21, 2012.
- [31] P. Usenik, M. Bürmen, A. Fidler, F. Pernuš, and B. Likar, “Evaluation of cross-polarized near infrared hyperspectral imaging for early detection of dental caries,” in *Lasers in Dentistry XVIII*, vol. 8208, p. 82080G, International Society for Optics and Photonics, 2012.
- [32] F. Blanco, M. Lopez-Mesas, M. Valiente, S. Serranti, G. Bonifazi, and J. Havel, “Hyperspectral imaging based method for fast characterization of kidney stone types,” *Journal of biomedical optics*, vol. 17, no. 7, p. 076027, 2012.
- [33] J. A. Chin, E. C. Wang, and M. R. Kibbe, “Evaluation of hyperspectral technology for assessing the presence and severity of peripheral artery disease,” *Journal of vascular surgery*, vol. 54, no. 6, pp. 1679–1688, 2011.
- [34] N. T. Clancy, D. Stoyanov, D. R. James, A. Di Marco, V. Sauvage, J. Clark, G.-Z. Yang, and D. S. Elson, “Multispectral image alignment using a three channel endoscope in vivo during minimally invasive surgery,” *Biomedical optics express*, vol. 3, no. 10, pp. 2567–2578, 2012.
- [35] M. E. Martin, M. B. Wabuye, K. Chen, P. Kasili, M. Panjehpour, M. Phan, B. Overholt, G. Cunningham, D. Wilson, R. C. DeNovo, *et al.*, “Development of an advanced hyperspectral imaging (hsi) system with applications for cancer detection,” *Annals of biomedical engineering*, vol. 34, no. 6, pp. 1061–1068, 2006.
- [36] T. Dohi, K. Matsumoto, and I. Shimoyama, “The micro fabry-perot interferometer for the spectral endoscope,” in *Micro Electro Mechanical Systems, 2005. MEMS 2005. 18th IEEE International Conference on*, pp. 830–833, IEEE, 2005.
- [37] J. Galeano, R. Jolivot, Y. Benezeth, F. Marzani, J.-F. Emile, and D. Lamarque, “Analysis of multispectral images of excised colon tissue samples based on genetic algorithms,” in *Signal Image Technology and Internet Based Systems (SITIS), 2012 Eighth International Conference on*, pp. 833–838, IEEE, 2012.
- [38] S. Kiyotoki, J. Nishikawa, T. Okamoto, K. Hamabe, M. Saito, A. Goto, Y. Fujita, Y. Hamamoto, Y. Takeuchi, S. Satori, *et al.*, “New method for detection of gastric cancer by hyperspectral imaging: a pilot study,” *Journal of biomedical optics*, vol. 18, no. 2, pp. 026010–026010, 2013.

- [39] S. E. Martinez-Herrera, Y. Benezeth, M. Boffety, J.-F. Emile, F. Marzani, D. Lamarque, and F. Goudail, "Identification of precancerous lesions by multispectral gastroendoscopy," *Signal, Image and Video Processing*, vol. 10, no. 3, pp. 455–462, 2016.
- [40] F. Deligianni, *VISUAL AUGMENTATION FOR VIRTUAL ENVIROMNENTS IN SURGICAL TRAINING*. PhD thesis, Imperial College London, 2006.
- [41] M. Scholz, W. Konen, S. Tombrock, B. Fricke, L. Adams, M. Von Duering, A. Hentsch, L. Heuser, and A. Harders, "Development of an endoscopic navigation system based on digital image processing," *Computer Aided Surgery*, vol. 3, no. 3, pp. 134–143, 1998.
- [42] F. Sauer, A. Khamene, and S. Vogt, "An augmented reality navigation system with a single-camera tracker: System design and needle biopsy phantom trial," *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2002*, pp. 116–124, 2002.
- [43] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.
- [44] J. P. Helferty, C. Zhang, G. McLennan, and W. E. Higgins, "Videoendoscopic distortion correction and its application to virtual guidance of endoscopy," *IEEE transactions on medical imaging*, vol. 20, no. 7, pp. 605–617, 2001.
- [45] H. Aasen and A. Bolten, "Multi-temporal high-resolution imaging spectroscopy with hyperspectral 2d imagers—from theory to application," *Remote sensing of environment*, vol. 205, pp. 374–389, 2018.
- [46] K. J. Zuzak, M. D. Schaeberle, E. N. Lewis, and I. W. Levin, "Visible reflectance hyperspectral imaging: characterization of a noninvasive, in vivo system for determining tissue perfusion," *Analytical chemistry*, vol. 74, no. 9, pp. 2021–2028, 2002.
- [47] T. A. Clarke and J. G. Fryer, "The development of camera calibration methods and models," *The Photogrammetric Record*, vol. 16, no. 91, pp. 51–66, 1998.
- [48] F. Remondino and C. Fraser, "Digital camera calibration methods: considerations and comparisons," *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 36, no. 5, pp. 266–272, 2006.
- [49] J. Boland, T. Ager, E. Edwards, E. Frey, P. Jones, R. Jungquiet, A. Lareau, J. Lebaron, C. King, K. Komazaki, *et al.*, "Cameras and sensing systems," *Manual of Photogrammetry*, vol. 8, pp. 581–676, 2004.

- [50] P. Sturm and S. Ramalingam, *Camera Models and Fundamental Concepts Used in Geometric Computer Vision*. Now Publishers Inc, 2011.
- [51] J. Barreto, J. Roquette, P. Sturm, and F. Fonseca, “Automatic camera calibration applied to medical endoscopy,” 2009.
- [52] J. Tischendorf, C. Trautwein, T. Aach, D. Truhn, T. Stehle, *et al.*, “Camera calibration for fish-eye lenses in endoscopy with an application to 3d reconstruction,” in *2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 1176–1179, IEEE, 2007.
- [53] J. Heikkila, “Geometric camera calibration using circular control points,” vol. 22, pp. 1066–1077, IEEE, 2000.
- [54] M. Liu, D. Yan, X. Hu, Y. Zhou, and Z. Wu, “A convenient and fast method of endoscope calibration under surgical environment,” in *International Conference on Computer Analysis of Images and Patterns*, pp. 507–514, Springer, 2013.
- [55] D. Scaramuzza, A. Martinelli, and R. Siegwart, “A toolbox for easily calibrating omnidirectional cameras,” in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5695–5701, IEEE, 2006.
- [56] B. J. Urbina, E. J. Coto, O. C. Rodríguez, R. E. Miquilarena, and M. E. Cerrolaza, “A virtual system for laparoscopic surgery training,” in *II Conferencia Internacional de Bioingeniería Computacional*, 2005.
- [57] P.-A. Heng, C.-Y. Cheng, T.-T. Wong, Y. Xu, Y.-P. Chui, K.-M. Chan, and S.-K. Tso, “A virtual-reality training system for knee arthroscopic surgery,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 8, no. 2, pp. 217–227, 2004.
- [58] J. J. Pan, J. Chang, X. Yang, J. J. Zhang, T. Qureshi, R. Howell, and T. Hickish, “Graphic and haptic simulation system for virtual laparoscopic rectum surgery,” *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 7, no. 3, pp. 304–317, 2011.
- [59] P. A. Kenney, M. F. Wszolek, J. J. Gould, J. A. Libertino, and A. Moinzadeh, “Face, content, and construct validity of dv-trainer, a novel virtual reality simulator for robotic surgery,” *Urology*, vol. 73, no. 6, pp. 1288–1292, 2009.
- [60] T. S. Lendvay, P. Casale, R. Sweet, and C. Peters, “Initial validation of a virtual-reality robotic simulator,” *Journal of Robotic Surgery*, vol. 2, no. 3, pp. 145–149, 2008.

- [61] E. M. Bonrath, B. K. Weber, M. Fritz, S. T. Mees, H. H. Wolters, N. Senninger, and E. Rijcken, "Laparoscopic simulation training: testing for skill acquisition and retention," *Surgery*, vol. 152, no. 1, pp. 12–20, 2012.
- [62] G. Ahlberg, L. Enochsson, A. G. Gallagher, L. Hedman, C. Hogman, D. A. McClusky III, S. Ramel, C. D. Smith, and D. Arvidsson, "Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies," *The American journal of surgery*, vol. 193, no. 6, pp. 797–804, 2007.
- [63] C. Basdogan, C.-H. Ho, and M. A. Srinivasan, "Virtual environments for medical training: Graphical and haptic simulation of laparoscopic common bile duct exploration," *IEEE/Asme Transactions On Mechatronics*, vol. 6, no. 3, pp. 269–285, 2001.
- [64] M. K. Chmarra, C. Grimbergen, and J. Dankelman, "Systems for tracking minimally invasive surgical instruments," *Minimally Invasive Therapy & Allied Technologies*, vol. 16, no. 6, pp. 328–340, 2007.
- [65] S. Speidel, G. Sudra, J. Senemaud, M. Drentschew, B. P. Müller-Stich, C. Gutt, and R. Dillmann, "Recognition of risk situations based on endoscopic instrument tracking and knowledge based situation modeling," in *Medical Imaging 2008: Visualization, Image-Guided Procedures, and Modeling*, vol. 6918, p. 69180X, International Society for Optics and Photonics, 2008.
- [66] A. Krupa, J. Gangloff, C. Doignon, M. F. De Mathelin, G. Morel, J. Leroy, L. Soler, and J. Marescaux, "Autonomous 3-d positioning of surgical instruments in robotized laparoscopic surgery using visual servoing," *IEEE transactions on robotics and automation*, vol. 19, no. 5, pp. 842–853, 2003.
- [67] F. Miyawaki, T. Tsunoi, H. Namiki, T. Yaginuma, K. Yoshimitsu, D. Hashimoto, and Y. Fukui, "Development of automatic acquisition system of surgical-instrument information in endoscopic and laparoscopic surgery," in *2009 4th IEEE conference on industrial electronics and applications*, pp. 3058–3063, IEEE, 2009.
- [68] D. Burschka, J. J. Corso, M. Dewan, W. Lau, M. Li, H. Lin, P. Marayong, N. Ramey, G. D. Hager, B. Hoffman, *et al.*, "Navigating inner space: 3-d assistance for minimally invasive surgery," *Robotics and Autonomous Systems*, vol. 52, no. 1, pp. 5–26, 2005.

- [69] P. Cabras, F. Nageotte, P. Zanne, and C. Doignon, “An adaptive and fully automatic method for estimating the 3d position of bendable instruments using endoscopic images,” *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 13, no. 4, p. e1812, 2017.
- [70] R. Reilink, S. Stramigioli, and S. Misra, “3d position estimation of flexible instruments: marker-less and marker-based methods,” *International journal of computer assisted radiology and surgery*, vol. 8, no. 3, pp. 407–417, 2013.
- [71] C. Doignon, F. Nageotte, B. Maurin, and A. Krupa, “Pose estimation and feature tracking for robot assisted surgery with medical imaging,” in *Unifying perspectives in computational and robot vision*, pp. 79–101, Springer, 2008.
- [72] D. Bouget, M. Allan, D. Stoyanov, and P. Jannin, “Vision-based and marker-less surgical tool detection and tracking: a review of the literature,” *Medical image analysis*, vol. 35, pp. 633–654, 2017.
- [73] V. Lepetit, P. Fua, *et al.*, “Monocular model-based 3d tracking of rigid objects: A survey,” *Foundations and Trends® in Computer Graphics and Vision*, vol. 1, no. 1, pp. 1–89, 2005.
- [74] S. Speidel, E. Kuhn, S. Bodenstedt, S. Röhl, H. Kenngott, B. Müller-Stich, and R. Dillmann, “Visual tracking of da vinci instruments for laparoscopic surgery,” in *Medical Imaging 2014: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 9036, p. 903608, International Society for Optics and Photonics, 2014.
- [75] A. Reiter, P. K. Allen, and T. Zhao, “Feature classification for tracking articulated surgical tools,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 592–600, Springer, 2012.
- [76] R. Wolf, J. Duchateau, P. Cinquin, and S. Voros, “3d tracking of laparoscopic instruments using statistical and geometric modeling,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 203–210, Springer, 2011.
- [77] R. Richa, M. Balicki, E. Meisner, R. Sznitman, R. Taylor, and G. Hager, “Visual tracking of surgical tools for proximity detection in retinal surgery,” in *International Conference on Information Processing in Computer-Assisted Interventions*, pp. 55–66, Springer, 2011.

- [78] A. Reiter and P. K. Allen, “An online learning approach to in-vivo tracking using synergistic features,” in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3441–3446, IEEE, 2010.
- [79] S. Kumar, M. S. Narayanan, S. Misra, S. Garimella, P. Singhal, J. J. Corso, and V. Krovi, “Video-based framework for safer and smarter computer aided surgery,” in *The Hamlyn Symposium on Medical Robotics*, pp. 107–108, 2013.
- [80] R. Sznitman, R. Richa, R. H. Taylor, B. Jedynak, and G. D. Hager, “Unified detection and tracking of instruments during retinal microsurgery,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 5, pp. 1263–1273, 2012.
- [81] Y. Li, C. Chen, X. Huang, and J. Huang, “Instrument tracking via online learning in retinal microsurgery,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 464–471, Springer, 2014.
- [82] M. Ye, L. Zhang, S. Giannarou, and G.-Z. Yang, “Real-time 3d tracking of articulated tools for robotic surgery,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 386–394, Springer, 2016.
- [83] J. Zhou and S. Payandeh, “Visual tracking of laparoscopic instruments,” *Journal of Automation and Control Engineering Vol*, vol. 2, no. 3, 2014.
- [84] S. Speidel, J. Benzko, S. Krappe, G. Sudra, P. Azad, B. P. Müller-Stich, C. Gutt, and R. Dillmann, “Automatic classification of minimally invasive instruments based on endoscopic image sequences,” in *Medical Imaging 2009: Visualization, Image-Guided Procedures, and Modeling*, vol. 7261, p. 72610A, International Society for Optics and Photonics, 2009.
- [85] C. Doignon, F. Nageotte, and M. De Mathelin, “Segmentation and guidance of multiple rigid objects for intra-operative endoscopic vision,” in *Dynamical Vision*, pp. 314–327, Springer, 2006.
- [86] A. M. Cano, F. Gayá, P. Lamata, P. Sánchez-González, and E. J. Gómez, “Laparoscopic tool tracking method for augmented reality surgical applications,” in *International Symposium on Biomedical Simulation*, pp. 191–196, Springer, 2008.
- [87] N. Masson, F. Nageotte, P. Zanne, and M. de Mathelin, “In vivo comparison of real-time tracking algorithms for interventional flexible endoscopy,” in *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 1350–1353, IEEE, 2009.

- [88] N. Masson, F. Nageotte, P. Zanne, M. de Mathelin, and J. Marescaux, "Comparison of visual tracking algorithms on in vivo sequences for robot-assisted flexible endoscopic surgery," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5571–5576, IEEE, 2009.
- [89] S. Voros, J.-A. Long, and P. Cinquin, "Automatic detection of instruments in laparoscopic images: A first step towards high-level command of robotic endoscopic holders," *The International Journal of Robotics Research*, vol. 26, no. 11-12, pp. 1173–1190, 2007.
- [90] S. Haase, J. Wasza, T. Kilgus, and J. Hornegger, "Laparoscopic instrument localization using a 3-d time-of-flight/rgb endoscope," in *2013 IEEE Workshop on Applications of Computer Vision (WACV)*, pp. 449–454, IEEE, 2013.
- [91] Z. Pezzementi, S. Voros, and G. D. Hager, "Articulated object tracking by rendering consistent appearance parts," in *2009 IEEE International Conference on Robotics and Automation*, pp. 3940–3947, IEEE, 2009.
- [92] M. Allan, S. Ourselin, S. Thompson, D. J. Hawkes, J. Kelly, and D. Stoyanov, "Toward detection and localization of instruments in minimally invasive surgery," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 4, pp. 1050–1058, 2012.
- [93] R. Sznitman, C. Becker, and P. Fua, "Fast part-based classification for instrument detection in minimally invasive surgery," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 692–699, Springer, 2014.
- [94] S. McKenna, H. N. Charif, and T. Frank, "Towards video understanding of laparoscopic surgery: Instrument tracking," in *Proc. of Image and Vision Computing, New Zealand*, 2005.
- [95] A. Reiter, P. K. Allen, and T. Zhao, "Marker-less articulated surgical tool detection," in *Computer assisted radiology and surgery*, 2012.
- [96] J. Climent and R. A. Hexsel, "Particle filtering in the hough space for instrument tracking," *Computers in biology and medicine*, vol. 42, no. 5, pp. 614–623, 2012.
- [97] P. Mountney, S. Giannarou, D. Elson, and G.-Z. Yang, "Optical biopsy mapping for minimally invasive cancer screening," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 483–490, Springer, 2009.
- [98] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

- [99] G. Csurka, D. Larlus, F. Perronnin, and F. Meylan, “What is a good evaluation measure for semantic segmentation?,” in *BMVC*, vol. 27, p. 2013, Citeseer, 2013.
- [100] J. A. Nelder and R. Mead, “A simplex method for function minimization,” *The computer journal*, vol. 7, no. 4, pp. 308–313, 1965.
- [101] R. E. Bixby, “A brief history of linear and mixed-integer programming computation,” *Documenta Mathematica*, pp. 107–121, 2012.
- [102] L. Xiongbiao and M. Kensaku, “Robust endoscope motion estimation via an animated particle filter for electromagnetically navigated endoscopy,” *IEEE Transactions on Biomedical Engineering*, vol. 61, pp. 85–95, 2014.
- [103] R. Shahidi, M. Bax, M. Jr, J. Johnson, E. Wilkinson, B. Wang, J. West, M. Citardi, K. Manwaring, and R. Khadem, “Implementation, calibration and accuracy testing of an image-enhanced endoscopy system,” vol. 21, pp. 1524–35, 01 2003.
- [104] P. Mountney, S. Giannarou, D. Elson, and G.-Z. Yang, “Optical biopsy mapping for minimally invasive cancer screening,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 483–490, Springer, 2009.
- [105] C. Long, T. Wen, W. J. Nigel, R. W. Tao, and J. Z. Jian, “Slam-based dense surface reconstruction in monocular minimally invasive surgery and its application to augmented reality,” *IComputer Methods and Programs in Biomedicine*, vol. 158, pp. 135–146, 2018.
- [106] A. Reiter, R. E. Goldman, A. Bajo, K. Iliopoulos, N. Simaan, and P. K. Allen, “A learning algorithm for visual pose estimation of continuum robots,” in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2390–2396, IEEE, 2011.
- [107] B. Münzer, K. Schoeffmann, and L. Böszörményi, “Content-based processing and analysis of endoscopic images and videos: A survey,” *Multimedia Tools and Applications*, vol. 77, no. 1, pp. 1323–1362, 2018.
- [108] S. Speidel, M. Delles, C. Gutt, and R. Dillmann, “Tracking of instruments in minimally invasive surgery for surgical skill analysis,” in *International Workshop on Medical Imaging and Virtual Reality*, pp. 148–155, Springer, 2006.
- [109] L. C. García-Peraza-Herrera, W. Li, L. Fidon, C. Gruijthuisen, A. Devreker, G. Attilakos, J. Deprest, E. Vander Poorten, D. Stoyanov, T. Vercauteren, *et al.*, “Tool-net: holistically-nested real-time segmentation of robotic surgical tools,” in *2017*

- IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5717–5722, IEEE, 2017.
- [110] M. Attia, M. Hossny, S. Nahavandi, and H. Asadi, “Surgical tool segmentation using a hybrid deep cnn-rnn auto encoder-decoder,” in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 3373–3378, IEEE, 2017.
- [111] R. A. Solovyev, A. A. Kalinin, A. G. Kustov, D. V. Telpukhov, and V. S. Ruhlov, “Fpga implementation of convolutional neural networks with fixed-point calculations,” *arXiv preprint arXiv:1808.09945*, 2018.
- [112] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.
- [113] Y. Liu, M.-M. Cheng, X. Hu, K. Wang, and X. Bai, “Richer convolutional features for edge detection,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3000–3009, 2017.
- [114] T.-B. Phan, D.-H. Trinh, D. Lamarque, D. Wolf, and C. Daul, “Dense optical flow for the reconstruction of weakly textured and structured surfaces: Application to endoscopy,” in *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 310–314, IEEE, 2019.
- [115] D. H. Trinh, C. Daul, W. Blondel, and D. Lamarque, “Mosaicing of images with few textures and strong illumination changes: Application to gastroscopic scenes,” in *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 1263–1267, IEEE, 2018.
- [116] R. Miranda-Luna, W. Blondel, C. Daul, Y. Hernandez-Mier, R. Posada, and D. Wolf, “A simplified method of endoscopic image distortion correction based on grey level registration,” in *2004 International Conference on Image Processing, 2004. ICIP’04.*, vol. 5, pp. 3383–3386, IEEE, 2004.
- [117] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, “Bundle adjustment—a modern synthesis,” in *International workshop on vision algorithms*, pp. 298–372, Springer, 1999.
- [118] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, “Patchmatch: A randomized correspondence algorithm for structural image editing,” in *ACM Transactions on Graphics (ToG)*, vol. 28, p. 24, ACM, 2009.

- [119] M. Kazhdan, M. Bolitho, and H. Hoppe, “Poisson surface reconstruction,” in *Proceedings of the fourth Eurographics symposium on Geometry processing*, vol. 7, 2006.
- [120] M. Waechter, N. Moehrle, and M. Goesele, “Let there be color! large-scale texturing of 3d reconstructions,” in *European conference on computer vision*, pp. 836–850, Springer, 2014.
- [121] OpenMVS, “Openmvs: open multi-view stereo reconstruction library,” 2020.
- [122] O. Zenteno, P. VanTrung, S. Treuillet, and Y. Lucas, “Markerless tracking of micro-endoscope for optical biopsy in stomach,” in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4419–4422, IEEE, 2019.

Suivi sans marqueur d'une sonde optique sous vision endoscopique: vers une biopsie multispectrale

Un prototype d'endoscope augmenté d'une sonde multispectrale a été développé au moyen d'un fibroscope inséré dans le canal opérateur d'un système endoscopique commercial et connecté à une caméra multispectrale snapshot (41 bandes dans la gamme de 470 à 975 nm). La sonde multispectrale fonctionne comme une biopsie optique localisée pour une exploration médicale complémentaire avec un champ de vision beaucoup plus petit que l'endoscope. Cette thèse aborde le problème d'un recalage in vivo des images bimodales (RGB endoscope/MS fibroscope) afin de fournir au chirurgien une localisation précise de la biopsie dans l'image endoscopique pendant l'examen. Ce recalage est critique car l'absence de texture dans la paroi gastrique fait que les méthodes classiques exploitant une mise en correspondance des contenus des images échouent le plus souvent. De plus, par mesure d'asepsie, la sonde doit être retirée et nettoyée entre chaque patient. C'est pourquoi nous abordons le problème de recalage comme une estimation de pose relative entre deux caméras hétérogènes et non rigidement liées (endoscope et fibroscope). Les deux caméras peuvent être calibrées séparément à l'aide d'un motif en damier de sorte que leur pose relative instantanée puisse être estimée en suivant la caméra mobile (fibroscope). Pour ce faire, on profite de la présence du fibroscope dans les images endoscopiques et de sa géométrie. Ainsi, la pose du fibroscope est estimée en minimisant l'erreur de re-projection d'un cylindre virtuel en utilisant une pose estimée en quatre étapes : (i) la segmentation du fibroscope est réalisée à l'aide de modèles CNN formés sur nos séquences endoscopiques ; (ii) un axe de cylindre 3D est initialisé en utilisant la segmentation initiale et les coordonnées de Plücker ; (iii) l'estimation de la pose est optimisée en maximisant l'indice de similarité de Jaccard entre une segmentation automatique du fibroscope dans l'image et la projection du cylindre virtuel dans le cadre de référence endoscopique ; (iv) nous utilisons la pose des deux caméras pour estimer la transformation par homographie directe entre les points projetés virtuellement, ce qui permet d'obtenir un suivi sans marque. Cette approche a été validée en utilisant une simulation réaliste des mouvements du fibroscope à l'intérieur du canal opérateur (c'est-à-dire l'insertion et le béquillage) et les séquences in vivo acquises avec le prototype multimodal dans un centre médical. L'estimation en temps réel de la position relative entre les deux caméras permet au chirurgien de visualiser la zone ciblée par le fibroscope sur la paroi gastrique au cours de l'examen et de superposer les deux modalités d'images en donnant accès aux données spectrales pour la biopsie optique. Dans un processus hors ligne, il est également possible pour le suivi des patients de localiser les biopsies sur une carte panoramique construite par des techniques de mosaïquage d'images à partir de la vidéo endoscopique acquise lors de l'examen. D'une manière générale, la méthode proposée pour localiser la biopsie optique peut être étendue à n'importe quel type d'imagerie, en particulier pour une sonde sans marqueur et des images peu texturées.

Mots clés : Gastro-endoscopie, Imagerie hyperspectrale, Suivi de sonde

Markerless tracking of a fiber-optical probe by endoscopic vision: Towards multispectral enhanced biopsy

This work addresses the problem of bi-modal registration to provide the surgeon with an accurate optical biopsy location in endoscopic frames during image guided examination. A multispectral-augmented endoscopic prototype was developed based on a commercial endoscopic system: A fiberscope is inserted into the operating channel and connected to a snapshot multispectral camera (41 bands in the range of 470 to 975 nm), working as a localized optical biopsy for a complementary medical exploration with a much smaller field of view. The registration between modalities is challenging due to the lack of texture in gastric wall which causes direct image-content registration methods to fail. Therefore, we approach the registration problem as a relative pose estimation between two non-homogeneous cameras (endoscope and fiberscope) not rigidly linked. Both cameras can be separately calibrated using a chessboard pattern so their instantaneous relative pose can be estimated by tracking the moving camera (fiberscope). To do this, we take advantage of the fiberscope presence in the endoscopic images and its geometry. Therefore, the pose of the fiberscope is estimated by minimizing the re-projection error of a virtual cylinder using a predicted pose in four steps: (i) the segmentation of the fiberscope is achieved through CNNs models trained on our endoscopic sequences; (ii) a 3D cylinder axis is initialized using the initial segmentation and Plücker's coordinates; (iii) the pose estimate is optimized by maximizing Jaccard's similarity index between an automatic segmentation of the fiberscope in the image and the virtual cylinder projection in the endoscopic reference frame; (iv) we use the pose of both cameras to estimate the transformation through direct homography between virtually projected points achieving a markerless tracking. This approach was validated using a virtual framework to realistically simulate the movements of the fiberscope inside the operating channel (i.e, insertion and precession) and in-vivo sequences acquired with our multimodal prototype in a medical center. The final registered images offer the surgeon real time visualization of the area targeted by the fiberscope at the gastric wall and its corresponding spectral data. In an off-line process, it is also possible to locate the hyperspectral biopsy on a panoramic map constructed by image mosaicing techniques from the endoscopic video acquired during examination. In a general way, the proposed method for locating optical biopsy can be extended to any type of imaging, specially in the presence of poorly textured images or when markerless probe tracking is necessary, and could be valuable for patient monitoring.

Keywords : Gastro-endoscopy, Hyperspectral imaging, Instrument tracking