Target Tracking for Endoluminal Interventions

L. (Luuk) Voskuilen

MSc Report

Committee:
Prof. dr. I.A.M.J. Broeders
Dr. ir. F. van der Heijden
Drs. P.A. van Katwijk
Dr. E.J.F.M. ten Berge
N. van der Stap, MSc

September 2015

028RAM2015
Robotics and Mechatronics
TG-CES
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

Abstract

Flexible endoscopy has evolved from a strictly diagnostic modality to technique that also offers simple endoluminal interventions. In the future more complex interventions will be developed, however improvements to the flexible endoscopic platform are needed to facilitate this progression. One very promising improvement is the robotisation of flexible endoscopy. Robotisation also allows other techniques that may improve complex endoluminal interventions, such as target tracking. Target tracking is the use of computer vision software to follow a certain target in an image sequence. The displacement of the target can be returned to the robot to lock onto the target, creating a target lock. We aim to develop an accurate, reliable and real-time target lock.

The target tracking algorithm is based on a GPU-accelerated SURF (speeded-up robust features) feature detector and a brute force feature matcher using computer vision library OpenCV. The pre-processing, feature detection and matching, and post-processing were optimised using videos from regular clinical practice.

Subsequently, the tracking algorithm is validated using a series of videos of endoluminal interventions recorded at the Meander Medical Center. The location target in the videos has been manually annotated by expert endoscopists providing a ground truth. The target location calculated by the algorithm is compared to this ground truth. Overall the agreement between algorithm and endoscopist was good, but some weaknesses were identified, such as the inaccurate tracking during quick movements. However, the algorithm proved to be resistant to disturbances such as smoke or instruments in the field-of-view, and the throughput was in real-time.

Next, the tracking algorithm was validated in a controlled laboratory environment, using a robotic arm to consistently move the target. An optical tracking system was used to track the location and orientation of the target and the endoscope tip and provide a ground truth. The tracking was found to be in real time and very reliable, but improvements may still be made in the accuracy and reducing the latency. Using a controller developed by Frijnts, we were able to achieve a target lock under these controlled circumstances.

Preface

This report is the the result of a year-long clinical internship at the Meander Medical Center and the department of Robotics and Mechatronics at the University of Twente. This internship is the culmination of the master Technical Medicine offered by the University of Twente.

First, I would like to thank prof. Ivo Broeders for his support at the Meander Medical Center and the intriguing procedures I had the pleasure to attend at the operating theatre. Secondly, my thanks go to Ferdi van der Heijden for the help in optimising the algorithm and invaluable feedback. My gratitude also go to Stefan Frijnts, with whom I had many engaging discussions and his help was pivotal to the successful testing of the target lock. I would like to thank Paul van Katwijk and my intervision group, for being able to ventilate my doubts and worries, and providing very useful advice. I owe Guide de Jong thanks for being able to continue his research and his assistance in getting the algorithm to work and subsequently improve it. I would like to thank the entire department of Gastroenterology and Hepatology and most of all dr Thijs Schwartz for allowing me to record countless endoscopies and for enduring the manual annotation. I would also like to thank the department of general surgery at the Meander Medical Center for having me all these months and the ample opportunities I have had to participate in clinical activities. Thanks to the department of Robotics and Mechatronics for making me feel welcome and for the unforgettable Chrismas party. I am very grateful for dr Eric ten Berge for taking place in the graduation commission. Also, I would like to thank my family and friend for enduring me after I had an exhausting day in the clinic or the lab. But most of all I have to thank Nanda van der Stap for helping me plan the course of the project, for the challenging discussions, for the detailed and extensive feedback, and for looking after me during the entire internship.

Contents

List of Figures

List of Tables

List of Abbreviations

3D	Three-dimensional, page 35
APC	Argon Plasma Coagulation, page 30
CPU	Central Processing Unit, page 19
EMR	Endoscopic Mucosal Resection, page 11
ERCP	Endoscopic Retrograde Cholangiopancreaticography, page 11
ESD	Endoscopic Submucosal Dissection, page 11
EUS	Endoscopic Ultrasound, page 13
FNA	Fine-Needle Aspiration, page 13
FOV	Field-of-view, page 31
FPS	Frames per Second, page 15
GI	Gastrointestinal (Tract), page 10
GPU	Graphics Processing Unit, page 19
ICC	Intraclass Correlation, page 31
IR	Infra-Red, page 42
LED	Light-Emitting Diode, page 42
NOTES	Natural Orifice Transluminal Endoscopic Surgery, page 14
PCA	Principal Component Analysis, page 22
PDT	Photodynamic Therapy, page 12
RFA	Radio-Frequency Ablation, page 12
RMSE	Root Mean Squared Error, page 31
SIFT	Scale-Invariant Feature Tracking, page 19
SURF	Speeded-Up Robust Features, page 19

Chapter 1

Introduction

1.1 Endoscopy

1.1.1 History of endoscopy

Mankind has always been interested in endoscopy, as civilisations such as the Egyptians, Greeks, Romans, and Arabs have attempted to inspect human body cavities using simple instruments such as spatulas and specula [1;2]. However, the first attempt to an endoscope, the *Lichtleiter*, was invented by the German urologist Phillip Bozzini in 1806 [3], which conducted candlelight focused by concave mirrors through a rigid tube into a body cavity. Subsequently in 1877, improvements were made by adding lenses to improve the visualisation and by moving the light source to the tip of the scope, resulting in the first usable cystoscope. [1;4] Shortly thereafter, the first practical gastroscope was created in 1880, followed by the first practical sigmoidoscope in 1895. [1] These were all rigid endoscopes, which restricted the cavities that could be examined. For instance, only about 25 cm of the distal colon could be visualised. [1].

The next big advancement in endoscopy came with the establishment of fiberoptics. This allowed light to be conducted through flexible fibres, thus allowing flexible endoscopes to be developed. In 1957, the first successful flexible gastroscopy was performed by Hirschowitz after first testing the device on himself. [1] The first flexible colonoscope (created in 1963) however was not as successful initially, which was hindered by the anatomy of the colon, most notably the curvatures of sigmoid colon. However, several improvements such as the addition of air and fluid suction channels, the control of the endoscope tip, and the lengthening of the scope led to a successful integration into the clinic. At approximately the same time, a working channel was included, paving the way for endoscopic interventions, resulting in the flexible endoscope known to this day. In the remainder, we will focus on gastrointestinal flexible endoscopy.

1.1.2 The endoscope

An elaborate description of the modern flexible has been given by Schwab and Singh. [2] The flexible endoscopy system can be divided into three parts, the endoscope, the stack and the instruments. The endoscope is the actual 'tube' that enters the subjects body. The camera (imaging chip and lens) are situated on the tip, which can be controlled by the wheels on the body of the endoscope. The shaft of the endoscope contains fibreoptics to convey the light from the light source to the cavity. Furthermore, channels are present in the shaft for insufflation, irrigation and suction, and a working channel for the passage of instruments. The body contains apart from the abovementioned control wheels, locks for the wheels, buttons for the plumbing systems (insufflation, irrigation and suction), and various other buttons for functions such as the capture of a snap-shot and digital zooming.

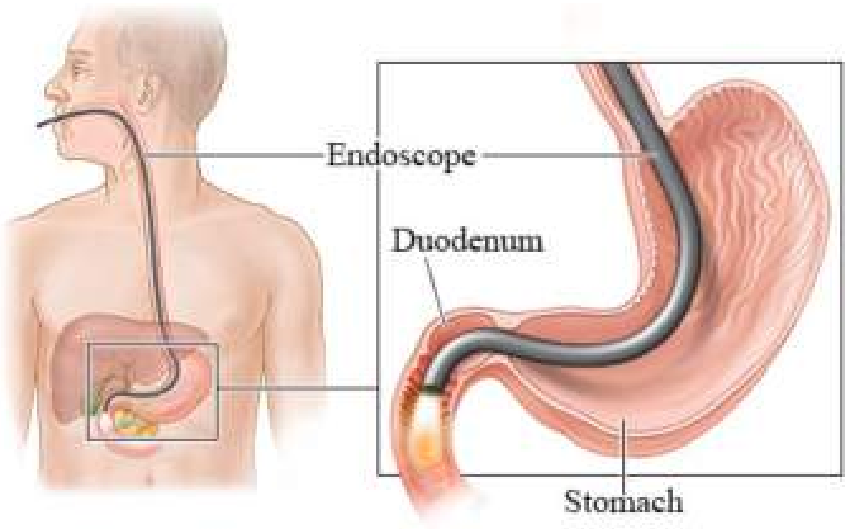


Figure 1.1: In this figure a schematic overview of a gastroscopy is depicted (from Queensland Centre for Digestive Medicine[6]). The endoscope is first advanced into the duodenum through the oesophagus and stomach. During retraction the mucosa of the duodenum, stomach and oesophagus are examined and possible biopsies are taken.

The flexible endoscope is connected to the stack. The stack contains the light source, an image processor, the actual plumbing systems and usually a screen on which to visualise the video stream captures by the camera.

Generally, a table is in close proximity containing instruments such as biopsy forceps, polypectomy snares and injection needles. Other units as electrocautery or argon plasma coagulation systems may be present on separate carts or on the endoscopy stack.

During certain endoscopic procedures, such as colonoscopies and ERCPs, or in specific cases, sedation and analgesia are indicated, which also have to be present. In The Netherlands, intravenous administration midazolam and fentanyl are preferred for sedation and analgesia respectively.[5] However, patient monitoring is mandatory during sedation, which is provided by pulse oximetry, blood pressure monitoring and a dedicated nurse.

1.1.3 Endoscopic interventions

As described in section 1.1.1, flexible endoscopy started as a purely diagnostic modality, but due to the addition of the working channel, therapeutic possibilities quickly emerged. In this section, we will provide a short review on endoluminal interventions favouring recent literature. We have focused our searched on endoluminal interventions of the upper and lower gastrointestinal (GI) tract.

Upper GI

Upper GI flexible endoscopy is comprised of oesophagogastroduodenoscopy (often shortened to gastroscopy, figure 1.1) and endoscopic retrograde cholongiopancreatography (ERCP, figure 1.2). During a gastroscopy generally random or targeted biopsies of the mucosa are performed. Random biopsies are a means of histologically proving gastritis[8] and/or the presence of helicobacter pylori. Targeted biopsies target mucosal defects such as Barrett's oesophagus for diagnosis and staging. Barrett's oesophagus is a condition affecting 1-2% of the population in which the distal end of the oesophagus undergoes intestinal metaplasia, which may progress into adenocarcinoma.[9] For submucosal defects such as gastrointestinal stromal tumours, keyhole biopsies can be used. During a keyhole

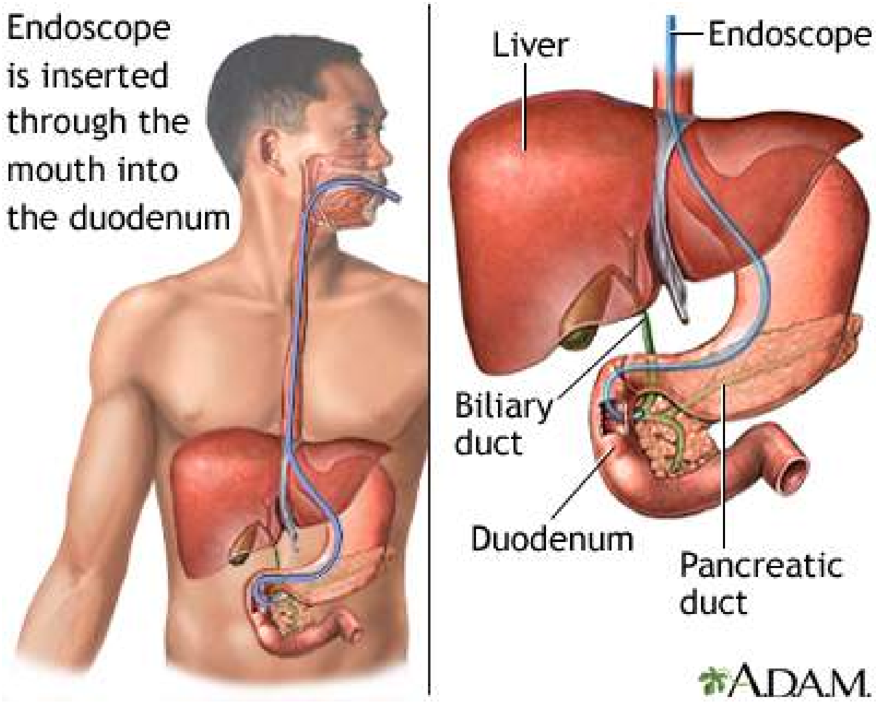


Figure 1.2: During an ERCP the endoscope is advanced into the duodenum and Vater's papilla is identified. After a papillotomy, a catheter can be inserted into the biliary and pancreatic ducts to visualise them using a contrast dye or to perform interventions such as stenting or lithotripsy. From MedlinePlus[7].

biopsy, a hole is created in the mucosa with a needle knife, through which a standard biopsy forceps can be introduced to obtain a tissue sample[10]. Gastroscopy can also be used in a more acute setting for the haemostasis of upper GI bleeding, for example due to peptic ulcer disease, a foreign body or gastroesophageal varices[11]. A solution of adrenaline can be used as an injectate, but there are also reports of applying a cyanoacrylate glue[12]. Non-bleeding gastroesophageal varices are generally first treated medicinally with beta-blockers, but a second line treatment can be rubber band ligation of the varices, which is an intervention also performed endoscopically[13]. Foreign bodies stuck in the GI track may also be removed endoscopically using for example forceps or snares.[14] There are even cases described in which migrated intragastric bands have been removed endoscopically using mechanical lithotriptors.[15] The final simple intervention noted here is the stenting of oesophagus and/or proximal stomach in a palliative setting[16–18], for oesophageal variceal bleeding[18] or after sleeve gastrectomy[19].

The first group of complex interventions includes resections in the upper GI tract. There is significantly more expertise present in the far East as opposed to the West, which is driven by the higher incidence of adenocarcinoma in the upper GI tract. However, diseases such as Barrett's oesophagus do have a significant prevalence in the West. David et al.[20] have written an extensive review on endoscopic interventions for Barrett's oesophagus while Hoppo et al.[21] focussed on gastric cancer. The first intervention is the endoscopic mucosal resection (EMR) which strictly speaking encompasses techniques ranging from polypectomies to piecemeal EMRs, but in practice EMR is used to refer to the piecemeal resection. During such a piecemeal EMR, saline enriched with a blue dye (and occasionally colloids such as glycerol) is injected into the submucosa to create a saline cushion. From this saline deposit, pieces of the mucosal defect can be removed by a snare. In such a way the entire lesion may be removed piece-by-piece. This has some disadvantages however, of which the most prominent are the difficult histological evaluation of the sample and the high recurrence rate[22]. Therefore, this technique can only be used for non-infiltrating carcinomas or precursors. For an en bloc resection, the endoscopic submucosal dissection (ESD) is used, which is even more complex than the EMR.

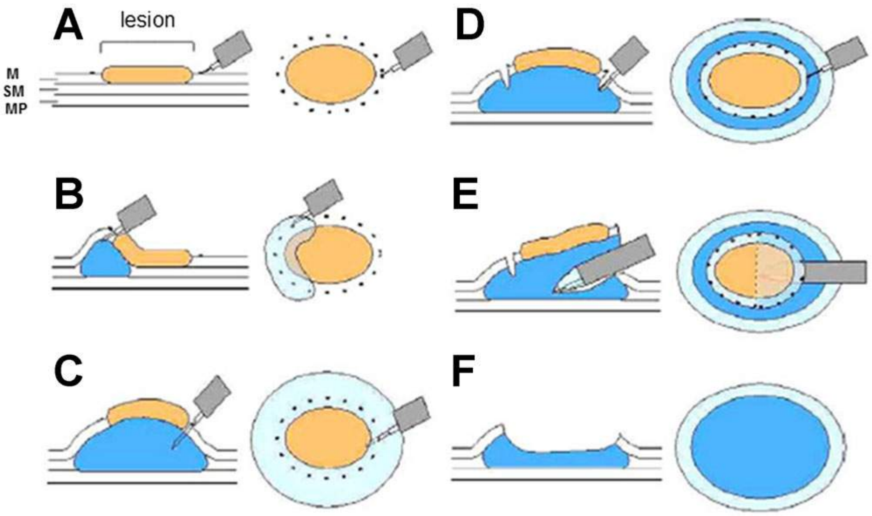


Figure 1.3: Schematic overview of an ESD procedure, copied from Hoppo et al.[21] M: Mucosa; SM: Submucosa; MP: Muscularis Propria. A: The mucosa around the lesion is marked with electrocautery for a complete resection even after tissue deformation; B and C: Injection of the saline solution in the submucosa creating a saline cushion and lifting the lesion; D: An incision is made in the mucosa along the previously created marks using a needle knife; E: Inserting the endoscope tip with a transparent cap into the incision enables further dissection through the submucosa; F: The lesion is removed en bloc.

During an ESD the same saline cushion is created to lift the lesion from the layers underneath (figure 1.3).[21;23] A needle knife is subsequently used to incise the mucosa around the lesion and a combination of a transparent cap and the needle knife are used to dissect the lesion from the submucosa/muscularis propria. In contract to EMR, this technique results in en bloc resections, on which a better histological analysis can be performed, and fewer recurrences.[23;24] However, more complications (delayed bleeding and perforations) are reported. An ESD is also much more complex resulting in a longer intervention time, and that a long training is required for endoscopists to perform ESDs. Therefore, ESDs are mainly performed in areas of high prevalence of early gastric cancer such as Japan.

Other interventions targeted specifically on Barrett's oesophagus are besides EMR, radio-frequency ablation (RFA)[9;20;25;26], and photodynamic therapy (PDT).[20;26] It is advisable that a Barrett's oesophagus with high-grade dysplasia (confirmed histologically and confined to the mucosa) is removed, for example by EMR, RFA, or PDT. RFA is a technique in which a high-frequency alternating current is administered to the dysplastic tissue, which is subsequently destroyed and replaced by normal mucosal tissue.[25] PDT uses an injected drug (a photosensitiser) that is converted by light of a certain wavelength into a cytotoxic substance.[20] Regarding short-term local recurrence, EMR, PDT and RFA appear to be equal. However EMR returns a tissue specimen for histological evaluation, but also results a higher complication rate.[20;25]

A completely different area of upper GI endoscopy is the ERCP (figure 1.2). During an ERCP, an endoscope with an angled tip is advanced into the duodenum. A catheter is inserted into Vater's papilla through which radiographic contrast dye is injected. By taking X-ray images, the system of biliary and pancreatic ducts as well as the gallbladder can be visualised. Subsequent interventions include sphincterotomy (also referred to as papillotomy), stenting and lithotripsy. However, the initial cannulation may already be difficult as cannulation of Vater's papilla fails in 5% of the cases and biliary cannulation fails in 15% to 35% of the cases.[27] Therefore it is recommended that a spincterotomy of the sphincter of Oddi is performed by a standard sphincterotome before cannulation.[27]. If the standard sphincterotomy fails, a precut or needle-knife sphincterotomy may be attempted.[27] Ultimately, the ERCP has to be repeated at a later time[28], or at a different location with a high-throughput of ERCPs, where more expertise is present.[29] indicating that even cannulation can be complex.

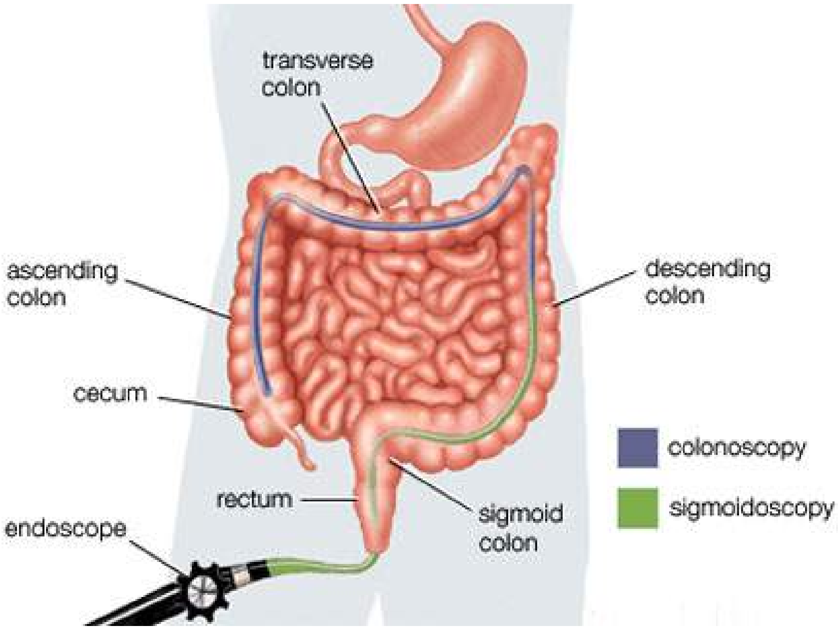


Figure 1.4: In this figure a schematic overview is given of a standard colonoscopy (from VanAmburg[37]). The difference in depth between colonoscopy and sigmoidoscopy is also visualised. During a colonoscopy, the endoscope is first advanced up to the caecum. Thereafter it is slowly retracted during which the mucosa of the colon is inspected.

A common intervention during ERCP is stenting. Mangiavillano et al. have written an excellent review on the indication and outcome of stenting during ERCP.[30] Indications for plastic stenting include benign biliary strictures, pancreatic duct obstructions, unsuccessful removal of biliary stones, biliary leaks, cholecystitis, the prevention of post-ERCP pancreatitis[31] and perforations.[30] Metallic stents are used in a palliative setting in order to prevent obstructions.[30;32]

The last ERCP-based intervention described here is lithotripsy or the destruction of stones summarising the extensive review of DiSario et al.[33] Mechanical lithotripsy is performed most frequently and involves a wire basket which may enclose the stone, and by retraction of the basket crush the stone.[33] In more specialised centres, electrohydraulic lithotripsy is used, in which a bipolar probe creates a shock wave to fragment the calculus, however direct choledochoscopic or pancreatoscopic vision and continuous plumbing are required.[33] Laser lithotripsy creates a similar shock wave as in electrohydraulic lithotripsy instead using a laser.[33] Extracorporeal shock wave lithotripsy may also be an option for which few complications are reported, but it does require a subsequent endoscopic removal of the stone fragments.[33;34]

Another form of upper GI endoscopy uses an entirely different mode of vision, namely endoscopic ultrasound (EUS). EUS can also be used to perform interventions in the GI tract, such as the drainage of pancreatic and biliary fluid collections by FNA (fine-needle aspiration).[35;36] Usage for malignancies include EUS guided brachytherapy and ethanol ablation, and for palliative pain management EUS guided celiac plexus block.[35;36]

Lower GI

Lower GI endoscopy can be divided into colonoscopy and sigmoidoscopy (figure 1.4, which is basically a colonoscopy which a shorter endoscope, during which only the right half of the colon may be observed. In contrast to upper GI interventions, lower GI interventions appear to be less diverse and overlap with upper GI interventions. Treatment for lower GI bleeding, from for example diverticula, is very similar to the treatment of peptic ulcer bleeding, namely adrenaline injection, electrocauterisation and or endoscopic clips.[38] Endoscopic resections in the lower GI also contain similar procedures (polypectomy, EMR, ESD) Two excellent reviews have been created by Kedia et al.[39] and Kaltenbach et al.[40]

Resections are again restricted to adenomas or early cancers and can be grouped into non-polypoid and polypoid lesions, the latter of which may be divided into sessile and pedunculated polyps.[40] Generally, polyps are removed using an endoscopic snare, which is a catheter containing a metallic loop. This snare can either be 'cold' or 'hot', referring to the application of electrocautery, which is dependent on polyp size.[39] This fairly simple procedure has been very successful in the prevention of colorectal cancer to the point that colonoscopies are indicated as a general screening program in many countries.[22;41;42]

Larger lesions require other approaches, such as EMR or ESD.[39;40] The procedures are similar to the procedures performed in the upper GI, as a saline cushion is first created to lift the lesion, after which it is resected using a snare (EMR) or a needle knife (ESD). However, care has to be taken as the intestinal wall is generally thinner in the lower GI compared to the upper GI.[43] Advantages and disadvantages are also comparable, with ESD having a higher en bloc resection rate and local recurrence rate, but having a higher intervention time and complexity.[39;40;44] Although EMR is performed relatively frequently, colorectal ESDs fall behind, even compared to gastric ESDs in Japan.[43;44] One may argue that this is the result of the complex nature of the intervention, thus requiring much training and expertise. In order to simplify complex endoscopic interventions, the robotisation of the flexible endoscope and the instruments has been attempted.

1.1.4 Robotisation

Most robotisations of flexible endoscopes do not focus on complex endoluminal interventions but on the even more complex transluminal interventions, commonly described as natural orifice transluminal endoscopic surgery (NOTES). NOTES is a form of minimally invasive surgery in which a single access point is created inside the body through which an endoscope and several instruments are inserted. A good review on robotic (and non-robotic) NOTES platforms has been written by Patel et al.[45] Most robotic platforms employ a master-and-slave system in which a group of servomotors controlling the endoscope and the instruments is the slave, which is controlled by the master, usually a telesurgical workstation, for example the MASTER[46]. However, they all use integrated optics, which may reduce the applicability of a robotic system versus a robotic system based on a readily available flexible endoscope, referring for example to the discussion on the cost-effectiveness of robotics in surgery.[47;48].

Therefore, the department of Robotics and Mechatronics of the Univeristy of Twente and DEMCON (Enschede, The Netherlands) have created in collaboration with the TeleFLEX system.[50] The TeleFLEX system (figure 1.5) uses a regular flexible endoscope which is fitted into a plastic interface, which in turn is driven by servomotors. The flexible endoscope can be easily inserted in and removed from the system. The system can be expanded by adding two robotically controlled instruments.[51] The user may control the endoscopy by using a joystick or touchpad which increases the efficiency of the advancement of the endoscope into the intestines.[52;53] If an endoluminal intervention is required, the user may switch to a telesurgical workspace for bimanual control of the instruments.

1.1.5 Advanced control

With the robotisation of the flexible endoscope also comes the opportunity to use advances algorithms to aid the user in the control of the endoscope. Examples include automated navigation[54–57], horizon stabilisation[58], pose estimation[59;60], and target locking,[61–64] of which target locking is the most promising for endoluminal interventions. Namely, target locking permits the user to focus on the instruments and the intervention instead of the continuous readjustment of the endoscope. However, current implementations all have their limitations, such as relying on periodical motions[61–63] or imaging beside the endoscope[64].

Figure 1.5: An overview of the TeleFLEX system (from Demcon[49]). To the left the endoscopy stack can be seem with the video processor, light generator, and the actuators for the endoscope and instruments. To the right, the telesurgical workstation can be seen.

1.2 This report

1.2.1 Objectives

In this master project, we aim to develop a target tracking algorithm based on previous work by De Jong.[65] Together with the controller developed by Frijnts[66], this will form the target locking module for the TeleFLEX system. The tracking algorithm will be validated using data from the regular clinical practice in the Meander Medical Center (Amersfoort, The Netherlands), and data from a tightly-controlled laboratory environment.

1.2.2 Requirements

Unlike current implementations, we do not want to rely on external imaging such as CT-scans, adding sensors to the endoscope or prediction of periodical motions, because these conditions constrain the clinical applicability of the target lock. Therefore, we have set up the following requirements for a target tracking algorithm.

- Accuracy
- Reliability
- Real-time
- Minimal input
- Ease-of-use

Accuracy refers to that the output of the tracking algorithm should be the same as the actual displacement of the target. Reliability is whether the target lock is able to perform correctly under various disrupting circumstances. Real-time means that the frame throughput of the algorithm is real-time (>15 FPS (frames per second)[67]). Minimal input can be realised by using solely the video feed from the endoscope and the initialisation of the target tracking by the user. Finally, ease-of-use refers to the intuitive initialisation of the target lock and that no additional hardware is required.

In chapter 2 the design of the algorithm will be described. Chapter 3 will focus on the validation of the algorithm using clinical data. The integration of the target tracking and the controller and subsequent testing is disclosed in chapter 4. Finally, a conclusion and points for future work will be given in chapter 5.

Chapter 2

Design

2.1 Requirements

In section 1.2.2, a series of requirements has been stated for a target tracking algorithm to improve complex endoluminal interventions is stated. Based on these requirements, a target tracking algorithm is designed, which will be described in this chapter.

There are several options available for tracking a target. However, many options can be dropped as they include either adding additional sensors to the endoscope. These options contradict with the latter two requirements, namely minimal input and ease-of-use. Adding extra hardware to the endoscope will make it harder to use, harder to clean and will take longer to set up. Therefore, we have chosen to develop a target locking technique based on the information provided by the endoscope alone.

The main source of information from the endoscope is the video stream. It contains not only information about the state of the intestines, but also about the movement of the endoscope tip relative to the bowel wall, as the camera is mounted on the tip of the endoscope. This movement can be extracted using the concept of optical flow.

2.2 Optical Flow

The concept of optical flow was coined by psychologist James J. Gibson[68]. He used optical flow to describe the sensed motion by an organism (egomotion) while moving through its environment. Later, this concept found much use in the computer vision community, as a means of detecting movement in an image sequence. The first group of algorithms calculated a dense optical flow field using the optical flow equation. This equation is derived in the following paragraph, based on the most well known dense algorithm described by Horn and Schunck[69].

2.2.1 Dense Optical Flow

Assume the image intensity $I(x, y, t)$, which is a function of the time (t) and location (x and y). After little time $(\mathrm{d}t)$, an object has travelled a short distance $(\mathrm{d}x, \mathrm{d}y)$, for which a Taylor series can be developed from the image intensity:

$$I(x+\mathrm{d}x, y+\mathrm{d}y, t+\mathrm{d}t) = I(x, y, t) + \frac{\partial I}{\partial x}\mathrm{d}x + \frac{\partial I}{\partial y}\mathrm{d}y + \frac{\partial I}{\partial t}\mathrm{d}t + \epsilon \tag{2.1}$$

The higher order terms of the Taylor series (ϵ) are omitted. Now, the brightness constancy constraint is introduced assuming that the object has had a constant brightness since the beginning of its movement:

$$I(x + \mathrm{d}x, y + \mathrm{d}y, t + \mathrm{d}t) = I(x, y, t) \tag{2.2}$$

Combining equations 2.1 and 2.2, we obtain:

$$\frac{\partial I}{\partial x}\mathrm{d}x + \frac{\partial I}{\partial y}\mathrm{d}y + \frac{\partial I}{\partial t}\mathrm{d}t = 0$$

and deviding by $\mathrm{d}t$ results in:

$$\frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \frac{\partial I}{\partial t} = 0 \tag{2.3}$$

in which $u = \frac{\mathrm{d}x}{\mathrm{d}t}$ and $v = \frac{\mathrm{d}y}{\mathrm{d}t}$ are the velocities of the object together forming a flow vector. Equation 2.3 is more generally known as the optical flow equation. However, this is one equation containing two unknown, which is unsolvable on its own. An example for this is a barber shop pole, for which is is unknown whether the lines are moving upwards or to the right. This is called the aperture problem and the various optical flow algorithms differ in the way they tackle this problem.

Horn-Schunck

Horn and Schunck[69] solved the aperture problem by introducing a constraint that the flow over entire image is smooth, meaning that neighbouring points have similar velocities. They have defined a parameter α, which is a weight factor in departure from the smoothness. Practically, this means that a larger α leads to a smoother optical flow field. Using this parameter they coudl define the following equations:

$$(\alpha^2 + I_x^2)u + I_x I_y v = \alpha^2 \overline{u} - I_x I_t$$
$$I_x I_y u + (\alpha^2 + I_y^2)v = \alpha^2 \overline{v} - I_y I_t$$

for which I_x, I_y, and I_t are the partial derivatives of I, while $\overline{u}$ and $\overline{v}$ are local averages of u and v. Thus two equations with two unknowns are left, whcih can be solved. The local averages are however dependent on the neighbouring values of u and v, which is solved by using an iterative algorithm.

Lucas-Kanade

Instead of the smoothness constraint from Horn-Schunck, Lucas and Kanade[70] use the assumption that displacements within a small section of the image are equal. This results in a series of equations for every pixel within the image section fow which equation 2.3 is true. Now, the are two unknowns for a number of equations equal to the number of pixels in the section of the image. Lucas and Kanade solve this by means of a least squares approximation.

Comparison

The Horn-Schunck algorithm tends to have more flow vectors than the Lucas-Kanade solution, but Horn-Schunck is more susceptible to noise[71]. However, both assume the optical flow equation (2.2, which is based on the brightness constancy assumption. We have noted however that this assumption is generally not true for endoscopic images as the brightness is heavily dependent on the distanc efrom the endoscope tip. Therefore, both solutions have limited applicabilty for our use and other algorithms are investigated.

2.2.2 Sparse Optical Flow

Kanade, Lucas and Tomasi were one of the first to make an attempt to feature tracking with the Kanade-Lucas-Tomasi feature tracker. However, we still noticed that this algorithm is very dependent on good illumination. Therefore, a more robust algorithm was required. Lowe created a feature tracking algorithm (SIFT)[72] that was more robust not only against illumination, but also against scaling and rotation. However, SIFT (scale-invariant feature tracking) is very computationally intensive, which contradicts the through-put requirement of the algorithm.

Therefore, a speeded-up variant of SIFT was used to compute our optical flow field, called SURF (speeded-up robust features)[73]. SURF uses some simplifications (such as integral images instead of image convolutions), which drastically speed up computing times, but slightly decrease robustness. It is also possible to accelerate the feature tracking by heterogeneous computing using readily available software libraries. In our case, hetergeneous computing means using the GPU (graphics processing unit) instead of the CPU (central processing unit) for the costly computations, such as feature detection. De Jong[65] has proven that a GPU-implementation accelerates the SURF algorithm to such an extent, that the computation time is within our requirements of real-time throughput. (>15 FPS[67]) The software library OpenCV concurrently with OpenCL are used in this instance.

In the following the SURF-based target tracking implementation created by De Jong[65] was improved on all fronts except feature detection and target region. Additionally the parameters of the algorithm are optimised for use in endoluminal interventions using a recorded video of an endoluminal intervention. This video is a different video from the others used in the remainder of the paper, such as the clinical validation (section 3). The over-all design is described in figure 2.1. In the following sections, the optimisation of the algorithm, divided into pre-processing (2.3), the actual feature detection and matching (2.4, and post-processing (2.5.2) are given.

2.3 Pre-processing

2.3.1 Histogram equalisation

When first looking at the low number of detected features in an example video of a colonoscopy, we decided the first step to be the enhancement of the input image. A low number of features will namely result in fewer vectors in the displacement field, which in turn will result in less reliable and (probably less accurate) target tracking. Therefore, we first looked for a method of increasing the number of detected features.

One way of increasing the number of features is by enhancing the image texture. The image texture is defined in this paper as the distribution of pixel intensities throughout the possible pixel intensities. Regularly, the information in an image is stored in a relatively narrow band of pixel intensities, resulting in a peak of the image histogram, see figure 2.2. To improved the image texture and, it is possible to redistribute the information over the entire possible range of intensities. This process is called histogram equalisation. Luckily, OpenCV has a readily available function to perform histogram equalisation. Note that the histogram equalisation is performed after the conversion from full-colour to grey-scale.

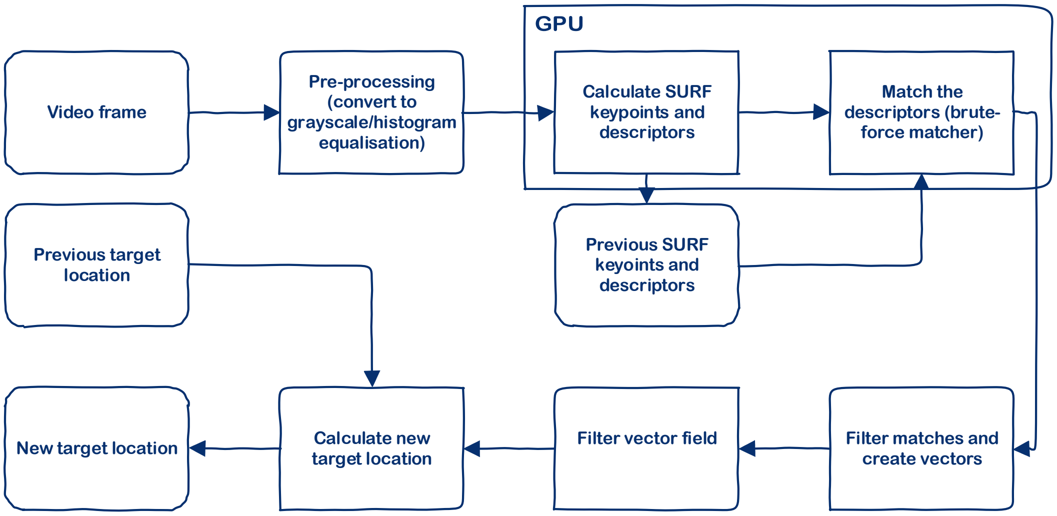


Figure 2.1: A flow chart of the current target tracking algorithm. Algorithm features are visualised by rectangular blocks, while variables are shown as rounded rectangles. (do rework)

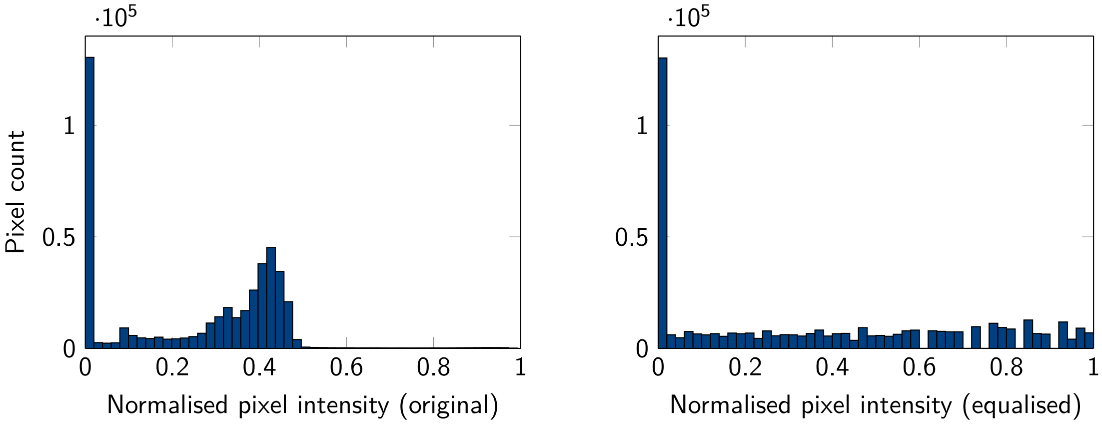
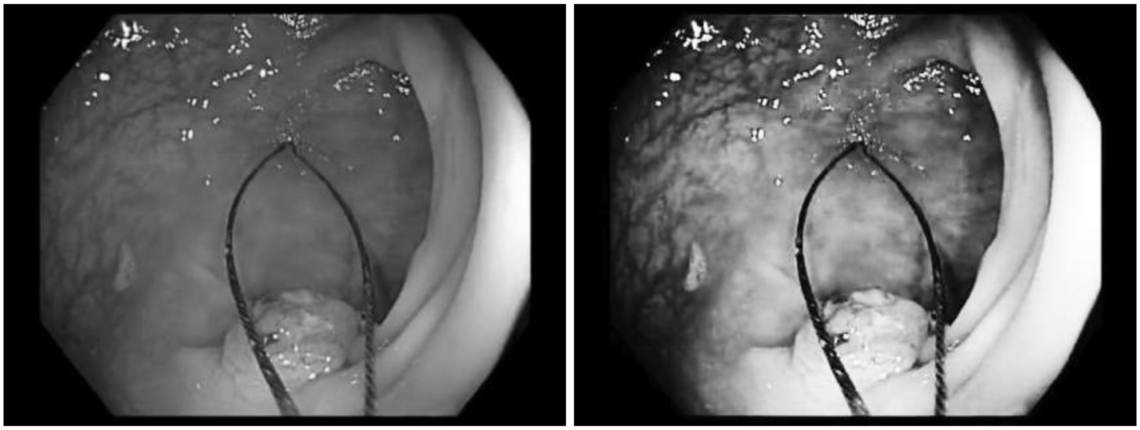
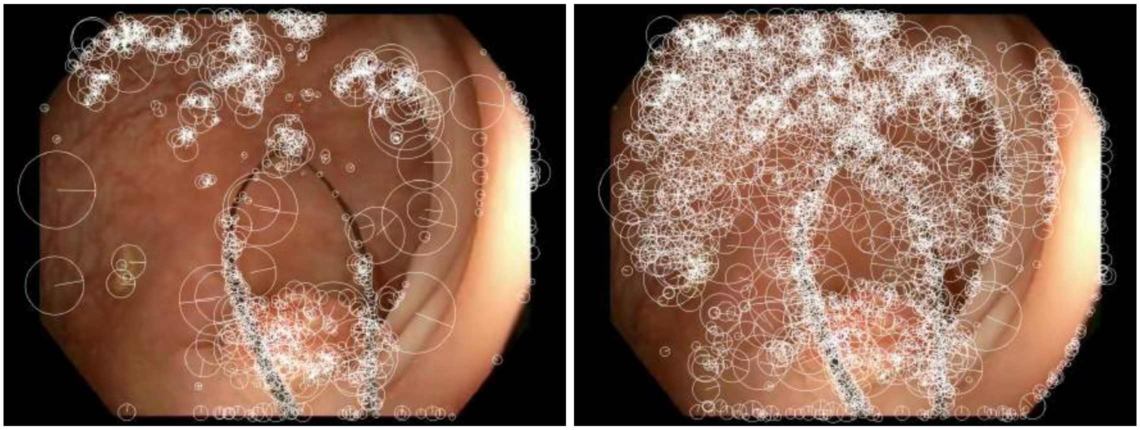


Figure 2.2: Histograms of a grey-scale sample picture before and after histogram equalisation. Note that overall, the pixel intensities have indeed equalised. However, a peak is still present at an intensity of zero (black) caused by the black area around the octagon containing the actual image. This peak will later be removed by the mask (see section 2.3.2), thus it is of no importance.



(a) Original **(b)** Equalised

Figure 2.3: The effect of the histogram equalisation on a frame from the sample video. Note that a wider range of image intensities is being used resulting in enhanced image texture.



(a) Original **(b)** Equalised

Figure 2.4: The features detected before and after histogram equalisation are depicted in white. The scale of the feature is visualised by the size of the circle and the orientation by the radius. The total number of features is evidently increased.

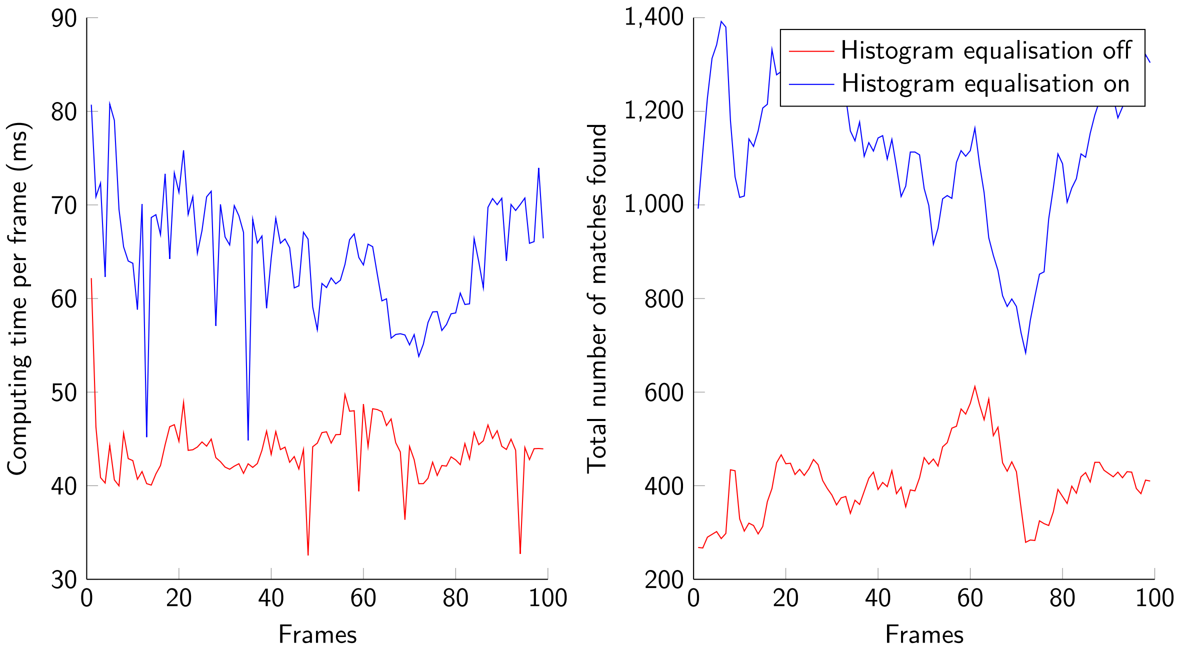


Figure 2.5: The difference in computing time and total number of feature matches per frame. There is a significant increase in computing time, due to the increase of the number of features, which in turn increases the time the SURF algorithm requires. (TMI, do remove?)

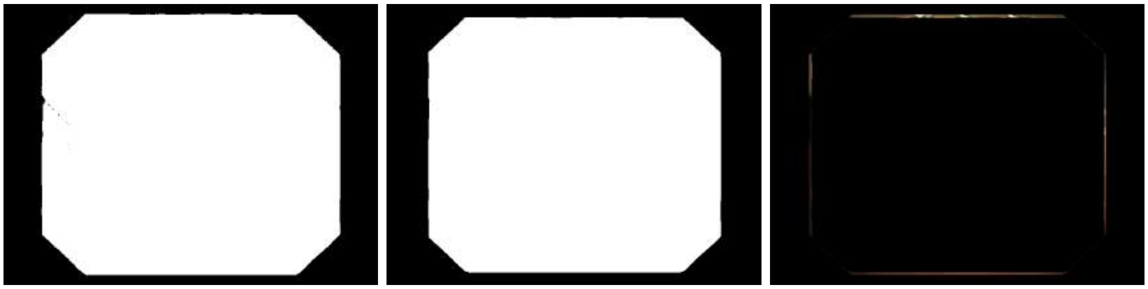
The effect of histogram equalisation on a frame from the image sequence is depicted in figure 2.3. Evidently, a wider range of image intensities is used, but more importantly, the total number of features has increased significantly after histogram equalisation (figure 2.4). This step comes at the cost of computation time, not only because of the rearrangement of the histogram itself, but also because of the resulting increased number of detected features. However the increase in computation time is not as large as the increase in detected features as can be seen in figure (do ref time vs number features), thus we deemed the histogram equalisation to be an effective tool for increasing the total number of features.

There are other methods of increasing the image texture besides histogram equalisation. Information is lost at the conversion of the input images from colour space to grey-scale. It is possible to execute a PCA (principal component analysis) on the three colour channels, and weigh the colour channels according to the PCA to increase the texture in the image. However, PCA requires even more computation time and it is not as easily implemented as histogram equalisation, which already returned excellent results. Therefore, we have chosen the histogram equalisation to enhance image texture.

2.3.2 Masking

The second thing we noticed when looking at the detected features, is that a large number of features is detected on the edge of the actual image of the colon shaped as an octagon and the surrounding black area. These features are useless for tracking, but they introduce outliers and absorb computing time. In order to remove these features a binary mask was developed.

The binary mask is created by a simple thresholding operation on the first received grey-scale image. This results in a raw binary image (see 2.6a) which is subsequently smoothened by two morphological operations. The first morphological operation is a dilation with a circular structuring element wit a diameter of 10 pixels. The second operation is an erosion using a diamond-shaped structuring element of 15 by 15 pixels. The resulting mask is



(a) After thresholding **(b)** Final mask **(c)** Information lost

Figure 2.6: Step in the development of a mask in order to exclude the edges of the octagon. First a binary threshold is applied, after which morphological operations are used to smoothen the binary mask. The last image shows the information lost due to the masking operation.

displayed in figure 2.6b and the part of the image which is effectively removed is visualised in figure 2.6c. In this last image can be clearly seen that the edges are removed with a minimal loss of information. The masking operation has been validated on an ERCP video (which has a slightly different octagonal) as well with similar results.

Note that the mask is not applied on the image before the SURF algorithm but directly after the displacement vector field is created. This means that keypoints and matches that will not be used anyway will be calculated by the SURF and matching algorithms. It is possible to enter a mask into a SURF algorithm as input, however this in not implemented in the OpenCL-accelerated version of SURF in OpenCV. In the future, this may be implemented resulting in an increase of the throughput, as less features are detected.

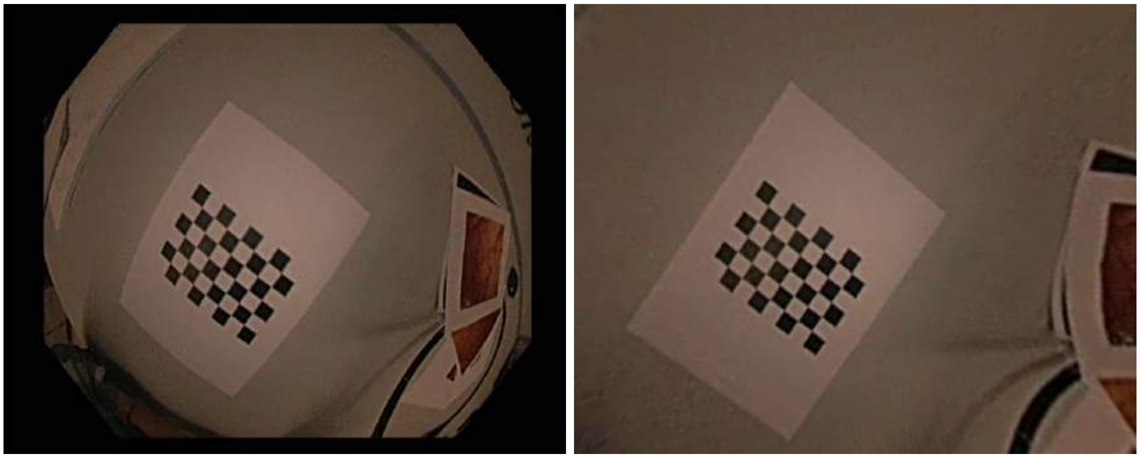
2.3.3 Camera Calibration

The lens system at the tip of the endoscope is hardly ideal. A significant amount of the fish-eye effect is present in endoscopic images. This may be intentional, as this fish-eye effect increases the surface area of the bowel that can be viewed in a single frame. However, this fish-eye distortion is detrimental for the optical flow, as displacement vectors cannot be compared reliably.

To counter the distortion, the camera can be calibrated using the toolbox provided by Bouguet[74]. This camera calibration is in our case performed by taking one-hundred images of a chessboard with the endoscope, with the chessboard at various positions and orientations in the field-of-view. If the dimensions and number of the squares of the chessboard are known, two matrices can be calculated, namely a matrix containing distortion coefficients and the camera matrix. For a more complete description we refer to section 4.1.2. These matrices are unique for the camera at the tip of the endoscope, have to be calculated only once and can be calculated long before any endoluminal intervention.

The two calculated matrices (camera matrix and distortion coefficients) are saved to a YAML-file by a separate calibration program, and can be loaded into the target tracking algorithm. We use the two calibration matrices in the initialisation of the target tracking to calculate a map, containing the deformation of an image needed to cancel the fish-eye effect. Subsequently, the map is applied to every frame (called undistorting) using a GPU-accelerated implementation in OpenCV.

The effect of the undistorting of the frames is visible in figures 2.7 and 2.8. Note the the bent lines in the image of the chessboard are straightened out effectively in the middle of the image, but that the image becomes blurry at the edges. This is due to the fact that individual pixels in the original image are smeared out over multiple pixels in the undistorted image. However this is the nature of the camera calibration an can hardly be changed.



(a) Distorted **(b)** Undistorted

Figure 2.7: An distorted and undistorted image of a chessboard recorded by the endoscope. Note that bent lines for example at the edges of the paper on which the chessboard is printed are straightened out. However, the undistortion is not perfect, especially at the edges of the image, because individual pixels in the distorted image are smeared out over the edge of the undistorted image. Accidentally, the black area around the octagonal has been removed.

(a) Distorted **(b)** Undistorted

Figure 2.8: An example of an original (distorted) and undistorted image of an inflamed bowel wall with a polyp. Note that the undistorted image was used for the locking validation in chapter4.

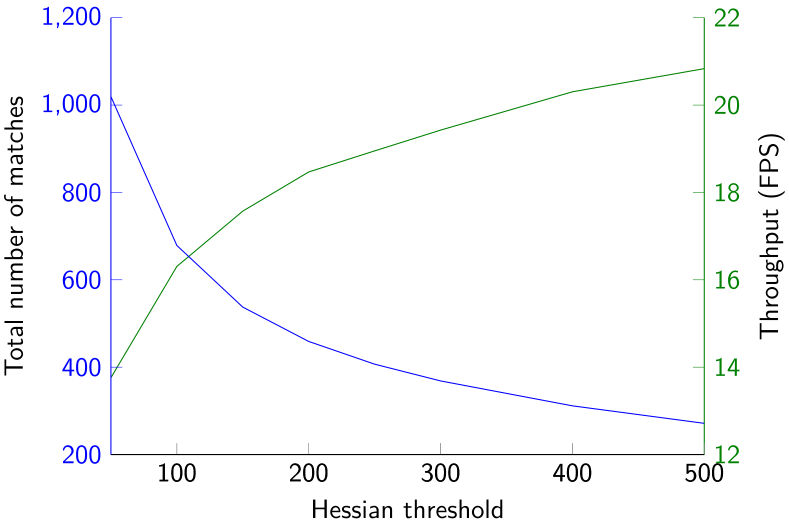


Figure 2.9: The effect of altering the Hessian threshold on the number of matches (blue) and throughput (green). A threshold of 250 was chosen as it provides a good throughput of approximately 20 FPS and enough features (approximately 400).

The effect of the camera calibration cannot be assessed with the training video's, thus this validation will be performed during the technical validation (see section 4 (do specify section further)).

It is possible to use another model of distortion instead of current model which may improve the undistorting. OpenCV has a specialised library for the compensation of fish-eye distortion, but it was not yet implemented in the OpenCV version we have used (2.4.9).

Advantages of the camera calibration are that pixel coordinates can be converted into real world coordinates (using millimetres) using the camera matrix, which may benefit the robotic control. Incidentally, the black edge, for which the masking operation (see section 2.3.2) was developed, is effectively removed rendering the masking operation useless if the camera is calibrated. However, we maintained the masking operation for when the camera calibration matrices are not present, for example during the clinical validation (section 3).

2.4 Feature Detection and Matching

2.4.1 Feature Detection

After the input images have been prepared by the pre-processing, they are passed on to the feature detector. The SURF detector (chosen in section 2.2) finds blob-like features based on the determinant of their Hessian matrix. An approximation of this determinant calculated and subsequently a threshold, we call the Hessian threshold, is applied to determine whether the keypoint is blob-like enough. Note that the actual calculation is more complex due to the calculation of the Hessians at multiple scale levels, for which we refer to the original paper[73].

The Hessian threshold can be adjusted to get the optimal number of features. A low Hessian threshold results in many features with many outliers and a high computing time and vice versa. The OpenCV documentation suggests a threshold value of 300-500, thus we adjusted the Hessian threshold from 50 to 500 in order to observe changes in number of matches and the computation time. Other parameters of the SURF algorithm were set to their defaults, namely the use 4 octaves with 2 octave layers per octave in the scale pyramid, and the use of 128-dimensional descriptors.

The results of the optimisation of the Hessian threshold are shown in figure 2.9. We can confirm that fewer keypoints are returned if the threshold is set to a higher value. We have chosen to continue with a threshold value of 250, because it returned a high number of matches (approximately 400 per frame) and a sufficient throughput of approximately 20 fps.

2.4.2 Feature Matching

There are several methods available to match features from subsequent frames. These can for example be based on complex decision trees to choose the best match quickly. However, the time needed for matching is nearly irrelevant regarding the time needed for the SURF algorithm. Therefore the simple brute-force matcher suffices for our implementation. This brute force matcher compares the descriptor of each feature in the first frame to the descriptor of each feature in the second frame, and returns the best matches. These matches can subsequently be converted into displacement field vectors by using the key points of the features.

Features can be matched using different norms. By using the norm, a response or goodness-of-match between descriptors is calculated on which the matches can be filtered. We have implemented a response ratio filter as proposed by Lowe[72] using the L2 norm (Euclidean distance). For this implementation, we require not only the best match, but also the second-best match from the brute-force matcher. The ratio between the response (L2 norm) of the best and second-best match may not exceed a certain number (for example, Lowe used 0.8[72]), because if this is the case, the matches are not very unique and thus the chance of the match being faulty is high. The brute-force matcher returning the best and second-best match is also GPU-accelerated.

The ratio threshold has been optimised by varying the ratio threshold from 0.1 to 0.9. The mean number of vectors inside the target region (do review, target region is not yet explained) is recorded, as well as the mean distance between the computed and manually annotated target location. We used the GPU-accelerated version of the brute-force matcher to minimise computation time. The target was manually annotated by an author on the used image sequence of 100 images.

The results are shown in figure 2.10. Note that as the ratio threshold increases, the mean number of vectors in the target region increases up until a ratio threshold of 0.8, after which it decreases. This decrease can be attributed to the target region erroneously moving to the edge of the image (due to a significant number of outliers) and staying there and where only few vectors are found inside the target region. The distance between computed and annotated target region starts to increase notably after a ratio of 6.5, due to an increase of outliers used in the calculation of the target location. We decided that the mean distance between computed and annotated target location is more important than the number of matches, thus it was decided that a ratio threshold of 0.65 is optimal.

Other methods of matching may decrease the computation time with respect to the brute-force matcher. The example of decision tree based matchers has already been stated, but it has not been evaluated as the brute-force matcher was sufficiently fast. We also experimented with a matcher that sorts matches based on their response and subsequently maintains only a certain top percentage. However we noted that this implementation is both less accurate (more outliers are retained) and costly (computation time is higher mainly because of the expensive sorting of the matches) compared to the ratio based matching filter.

2.5 Post-processing

2.5.1 Target Region

After the feature matching, a displacement field is generated of the entire frame (excluding the binary mask). However, as target tracking implies, we require only information of a certain target. Therefore, we use a region-of-interest, named the target region, to determine useful displacement vectors in the vicinity of the target. We have

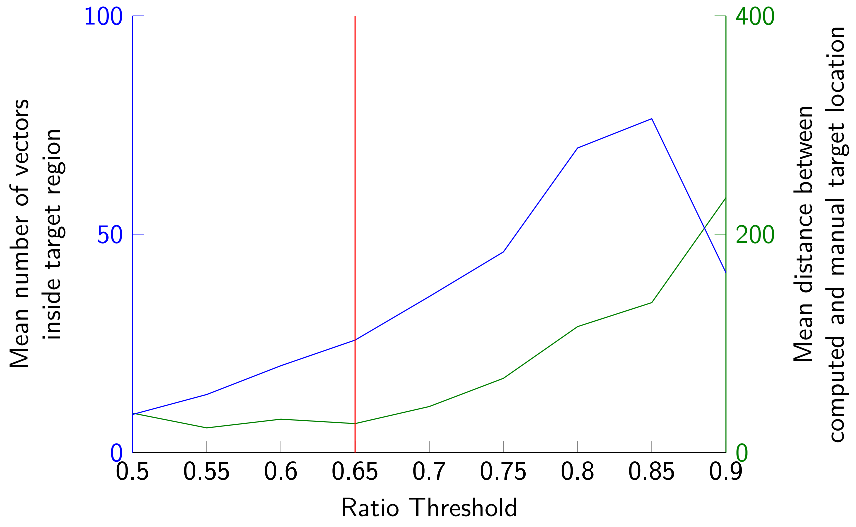


Figure 2.10: Increasing the ratio threshold as expected results in more matches, but also increased the computation time. We have found that a ratio threshold of 0.8 is a good middle ground between many matches, little outliers and adequate through-put.

chosen to use a circular target region, thus the target region can be described by a target location (the centre of the circular region of interest) and a target radius (the radius of the circle). Displacement vectors either beginning or ending in the target region are considered as inside the target region. Note that many times, the target radius as initialised is doubled to increase the number of displacement vectors inside the target region. This reduces the chance of loosing the target and increases the accuracy.

The vectors inside the target region are averaged, and the resulting mean displacement vector is considered the movement of the target between frames. Therefore, the target location is updated by adding the mean displacement vector to the previous target location.

The target region is initialised by the user at the initialisation of the target tracking algorithm, and if the target is lost (see section 2.5.4). There are two options for selecting a target region: a single click (determining the target location) while using a predefined target radius, or a click-and-drag to provide the user with control over both target location and radius. Feedback from expert endoscopists during the clinical validation step stated that the click-and-drag is preferred, thus this variation is implemented.

2.5.2 Median-based filtering

While visually reviewing the vectors inside the target region, see figure (do ref video?), a few very large outliers are still present interfering with the mean displacement vector. Therefore, an additional filtering step is added to remove these outliers.

We have implemented a median-based filter, which calculates the median length of the vectors located inside the target region and removes those vectors that are larger than the median times a certain factor. In order to optimise the filtering, the median filtering factor is varied from 1 to 9 and the mean number of vectors in the target region and the mean distance between computed and manually annotated target location are observed. Manual annotation was performed on the training image sequence of 100 images.

The results of this optimisation step are displayed in figure 2.11. Note that the target tracking reduces in accuracy (distance from annotated target) and reliability (number of vectors) if the median filtering factor is 1, as basically half of the vectors are removed. We see that already at a factor of 2 that the number of vectors in the target

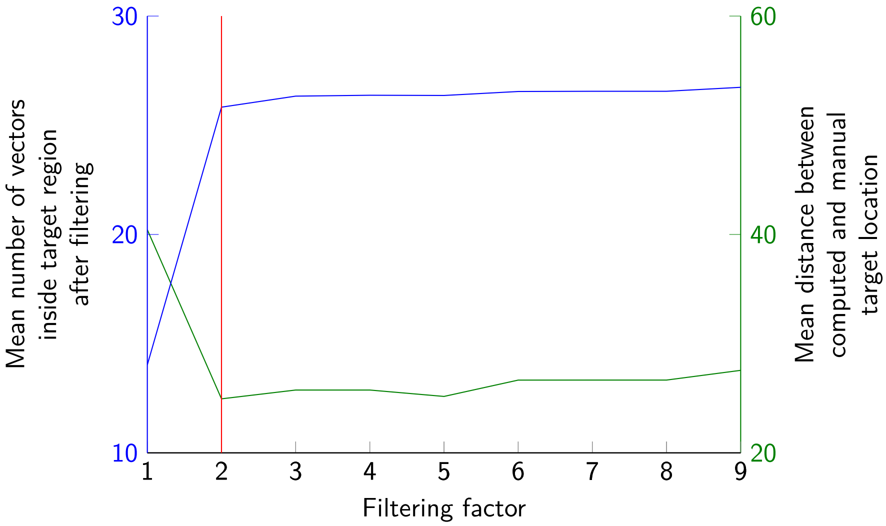


Figure 2.11: The effect of varying the median filter factor is shown here. It is evident that after a factor of 2 the filtering is stabilised with regards to the number of vectors, so hardly any reliability can be gained. Also, after the median filter factor of 2 the mean distance between annotated and computed target starts to increase. Therefore we have chosen to use a median filter factor of 2, visualised by the red line. (do more points between 1 and 2 for further fine-tuning, a factor 5+ is also pretty useless)

region has stabilised, and that the distance between computed and annotated is minimal. Increasing the factor will result in a minimal increase of the number of vectors, but increases the chance of outliers not being filtered out. Therefore, the median filtering factor of 2 has been deemed optimal.

We have implemented the median-based filtering as opposed to a mean-based filtering, because the median is affected less than the mean by a few large outliers. However, the calculation of the median is more costly, as the vector lengths have to be sorted. This increase in computing time is insignificant however regarding the time it takes to calculate SURF features. It is also possible that there are still outliers present with a length smaller than the median length of the displacement field, as only outliers that are too large are filtered out. However, these tiny vectors have very little effect on the resulting target displacement, thus filtering on outliers with a small length is not deemed necessary.

2.5.3 Zooming

During an endoluminal intervention the scope is advanced and retracted, effectively zooming in or out. This zooming has repercussions on the target radius as this radius may no longer be correct is too much zooming has occurred. Therefore, a zoom estimation is implemented. The zoom is defined here as the scale difference between two subsequent frames.

The zoom is estimated by calculating the sum of the distances from each feature in one frame, to each other feature in the same frame. The ratio between the sum of distances between features in the first and the sum of the second frame is the zoom. Note that only the features inside the target region after the two filtering steps (see sections 2.4.2 and 2.5.2) are used. Finally, the target radius is multiplied with the zoom, in the same way the target location is updated by the mean displacement vector.

In order to visualise how closely the zoom resembles the difference in scale of the target, the size of the target is manually annotated by drawing an ellipse over the laesion in all 100 images of the training sequence. The two radii of the ellipse are averaged and the ratio between two subsequent averaged radii is used to calculate the

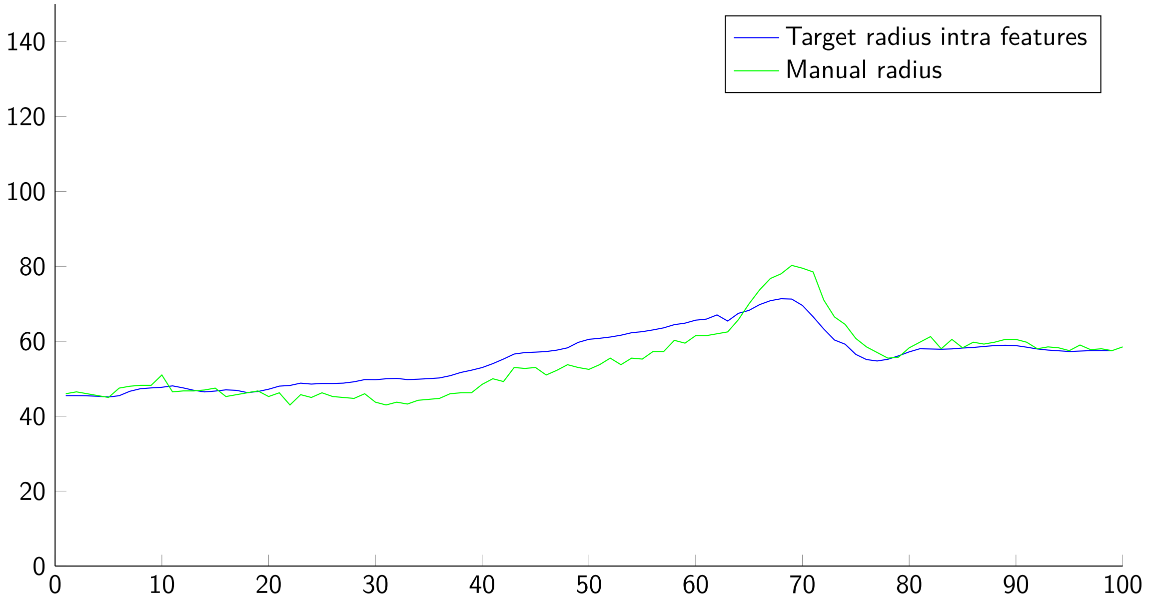


Figure 2.12: The target radius is displayed as calculated by the zoom estimation and by manual annotation of the target. A visible error between both radii is present.

zoom. The computed zoom and annotated zoom are compared visually and compared by calculating the ICC, for which more information can be found in section 3 (do specify reference).

The results of the zoom implementation are displayed in figure 2.12. The estimated zoom does resemble the manual zoom to some extent , but it is not very accurate. This can be attributed to a drift. This drift is present because the error in the zoom increases over each frame. Therefore, the zoom estimation may not very useful if target tracking is used for extended periods of time. Therefore, the accuracy of the zoom estimation has to be further evaluated, which is performed later in section 4.

2.5.4 Frame rejection

During quick movements, motion blur occurs resulting in little features found for a couple of frames. Therefore, a frame rejection is implemented. If less than five displacement vectors are present in the target region, the frame is rejected or skipped. Subsequently, the target tracking algorithm tries to detect matches between the last good (unrejected) frame, and up to ten subsequent frames. If all those ten frames are rejected, the target is considered lost. The number of ten subsequent frames is not an optimised number, as it is difficult to predict how many frames may be rejected until the reacquisition of the target is too inaccurate. Ten frames or 0.5 seconds (at 20 FPS) is in our opinion a reasonable number.

Chapter 3

Tracking Validation

Thus far, the current design of the tracking algorithm has only been tested on a single sample image sequence. Therefore, the logical next step is the validation of the target tracking on multiple varying image sequences. In contrast to the design phase, the algorithm initialisation and reference data will not be supplied by one of the authors, but by the end-users of the product. Note that a large proportion of this chapter will be published.[75]

3.1 Materials and Methods

The previously designed target tracking algorithm will be used. To summarise, the algorithm uses a GPU-accelerated SURF feature detector and subsequent matcher to calculate a sparse optical flow field, which is filtered and from which finally a target displacement is calculated. This displacement is used to track a target throughout an image sequence. The parameters of the algorithm have been set according to the values determined as optimal in chapter 2.

3.1.1 Image sequences

The tracking algorithm will be validated on image sequences or short video clips acquired during routine clinical practice. As the goal of the target tracking is to improve complex endoluminal interventions, image sequences of endoluminal interventions are used. This includes both simple, such as polypectomies, and complex, such as EMRs, interventions as in the Meander Medical Center, not enough complex interventions are performed to establish a large and diverse enough database. From all interventions named in section 1.1.3, the ones that are performed in the Meander Medical Center were identified. Interventions for which the target lock would have absolutely no added value, for example random gastric biopsies or FNA during EUS, were removed from the list. The procedures in which these intervention occur were identified, namely colonoscopies, EMRs (of the colon) and ERCPs. (Note that in this chapter the EMR has been promoted to a procedure, and divided into (sub)interventions such as dye injection and argon plasma coagulation (APC). Therefore, these procedures performed in the Meander Medical Center were recorded anonymously from October until December 2014. The videos have a resolution of 768×576, a frame rate of 25 FPS, and were deinterlaced. The endoscopes used are listed in table 3.2.

The recorded procedures were cut into smaller video clips compromised of endoluminal interventions. Disturbances in the clips of interventions such as bad illumination and blocked vision, and the targets of the interventions were noted. From the set of video clips of interventions a first group of six clips was chosen with maximal diversity in procedures, interventions and disturbances. The first 100 frames of these clips were used to create a set of six

Number	Procedure	Intervention	Target	Disturbances
1	Colonoscopy	Polypectomy	Polyp	Coarse movements
2	Colonoscopy	Polypectomy	Polyp	Bad illumination; Occlusions
3	ERCP	Cannulation	Vater's papilla	Target near edge of the FOV; Occlusions
4	ERCP	Sphincterotomy	Vater's papilla	Target near edge of FOV; Smoke development; Occlusions
5	EMR	Injection	Carcinoma	Large target; Colour change due to dye injection
6	EMR	APC	Residual lesion	Large area of removed mucosa; Small target; Sparks due to APC
7	Colonoscopy	Injection	Polyp	Colour change due to dye
8	Colonoscopy	Polypectomy	Polyp	Large polyp; Dirt on lens
9	ERCP	Cannulation	Vater's papilla	Endoscope motion
10	ERCP	Stent removal	Vater's papilla	Multiple instruments in view
11	EMR	Partial resection	Primary tumour	Large target; Coarse movements; Dye injection
12	Colonoscopy	APC	Polyp	Sparks; Bubbles; Small target; Target near edge of FOV

Table 3.1: Image sequences used in the validation. Note that in all sequences, instruments are present in the field of view (FOV).

image sequences[a]. Subsequently, a second set of six image sequences of 100 frames was created with the same maximal diversity in mind. The two sets of image sequences are described in table 3.1.

3.1.2 Manual annotation

Two expert endoscopists (>2000 procedures) subsequently manually annotated the location of the target in the images sequence in the first set of six interventions by selecting the centre point of the target. The annotators also defined an ellipse surrounding the target in the first frame which is used to initialise the target location and target radius in the tracking algorithm. The second set of interventions was annotated by only one expert endoscopist as we noticed a decrease in focus in a long session of manual annotation.

3.1.3 Analysis

The correlation between the annotations of the different endoscopist is evaluated using the intraclass correlation (ICC)[76]. The algorithm is initialised using the average ellipse and the target location supplied by the annotators. In order to assess the accuracy of the algorithm, the location according to the tracking algorithm is compared to the mean target location as determined by the expert endoscopists by calculating the root mean squared error (RMSE) of the distance between both targets and by calculating the ICC. The robustness of the algorithm is assessed by the mean number of optical flow vectors after filtering. Finally, the mean throughput (in FPS) will be calculated to determine whether the tracking is real-time (>15 FPS[67]).

---

[a] Video clips and image sequences are roughly equal to each-other, however in this report we use video clips to refer to a piece of video of undetermined length, while a image sequence has a fixed length of for example 100 frames.

Procedure	Endoscope	Description
Colonoscopy	Olympus CF-H180AL	Regular colonoscope
ERCP	Olympus TJF-160VR	ERCP-scope
EMR	Olympus PCF-PH190L	Paediatric colonoscope

Table 3.2: List of the endoscopes used from recording endoluminal interventions

Sequence	Vectors	Throughput (FPS)	ICC (%)	RMSE (pixels)
1	33.3	24.1	81.2	137.1
2	53.9	21.8	88.7	35.4
3	58.5	22.9	99.6	32.7
4	30.3	23.4	97.6	87.4
5	58.5	19.5	96.1	31.0
6	69.0	20.0	94.5	30.8
7	58.0	22.2	99.9	9.4
8	54.0	20.5	73.6	116.0
9	27.9	24.9	99.8	16.6
10	31.7	20.6	75.7	43.7
11	23.3	23.3	78.4	122.9
12	51.8	20.8	63.5	57.1

Table 3.3: Results for every sequence. Vectors: Mean number of optical flow vectors after filtering; ICC: Intraclass correlation; RMSE: Root mean squared error. The first six sequences were annotated by two annotators.

3.1.4 Hardware

The laptop on which the videos were recorded and target tracking is implemented, is a HP Elitebook 8570w mobile workstation running a 64-bit operating system The target tracking, based on the work done by De Jong[65], is written in C++ using Microsoft Visual Studio Express 2013 and using the libraries provided by OpenCV 2.4.9 which in turn incorporates OpenCL 1.1. Visualisation of the results was performed in MATLAB 2013a (Mathworks Inc., MA). Communication between Visual Studio and MATLAB was done by using the YAML-format and by loading functions from OpenCV into MATLAB by using MEX-files[77]. Videos of the procedures were recorded via a FireWire connection. The endoscopes used are listed in table 3.2.

3.2 Results

The ICC between annotators was very high (>97.5%) in all sequences. The ICC was relatively small if the target was large (97.6% versus 98.8% or higher in respectively sequences 4 versus the others).

The tracking results of the sequences are summarised in table 3.3. As expected the mean number of optical flow vector is larger in sequences with high image texture, such as sequence six in which the texture is increased due to the removal of a large section of the mucosa. Additionally, the mean number of vectors is large enough to ensure a robust target tracking. The mean throughput is larger than the minimal throughput of 15 FPS for every sequence, which confirms the real-time property of the target tracking.

The accuracy differs between image sequences. Examples of accurate (sequence 7) and inaccurate (sequence 1) tracking are elaborated in figures 3.1 and 3.2 respectively. In figure 3.1 the location of the target changes little which makes the tracking simpler for the algorithm. However, there was an instrument in view and the mucosa changed colour due to the injection of a dye. In figure 3.2, it is evident that a sudden increase in tracking

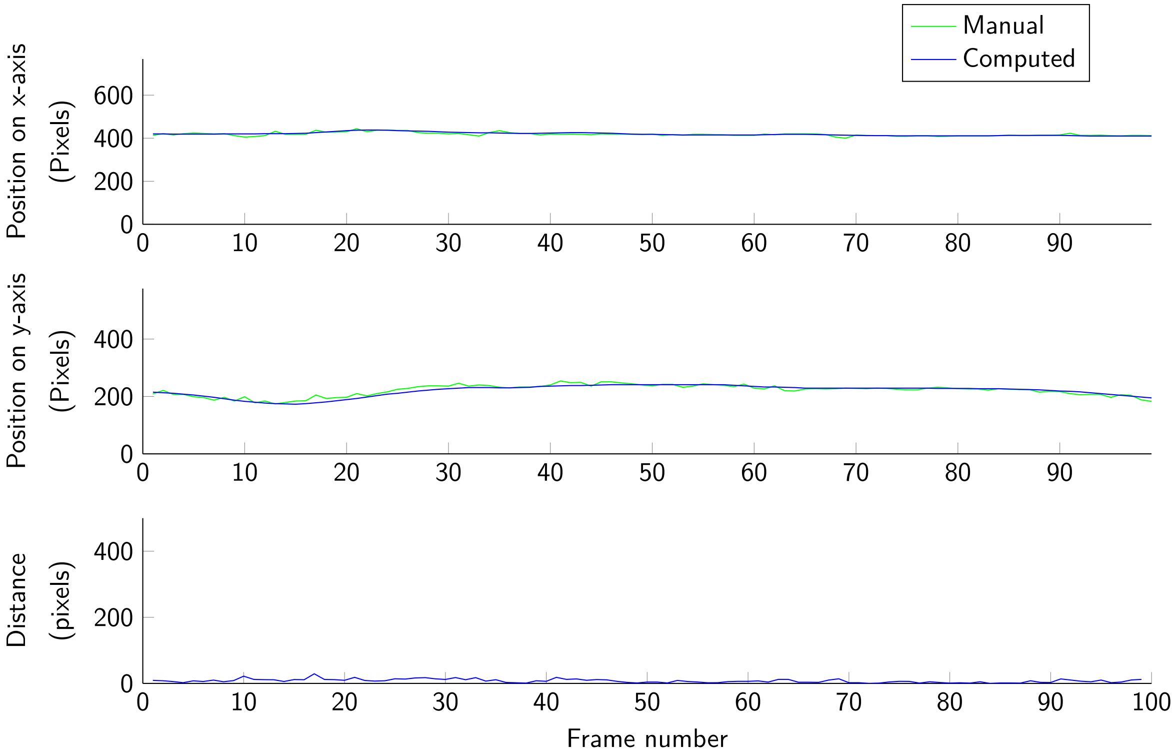


Figure 3.1: An example of accurate tracking (sequence 7).

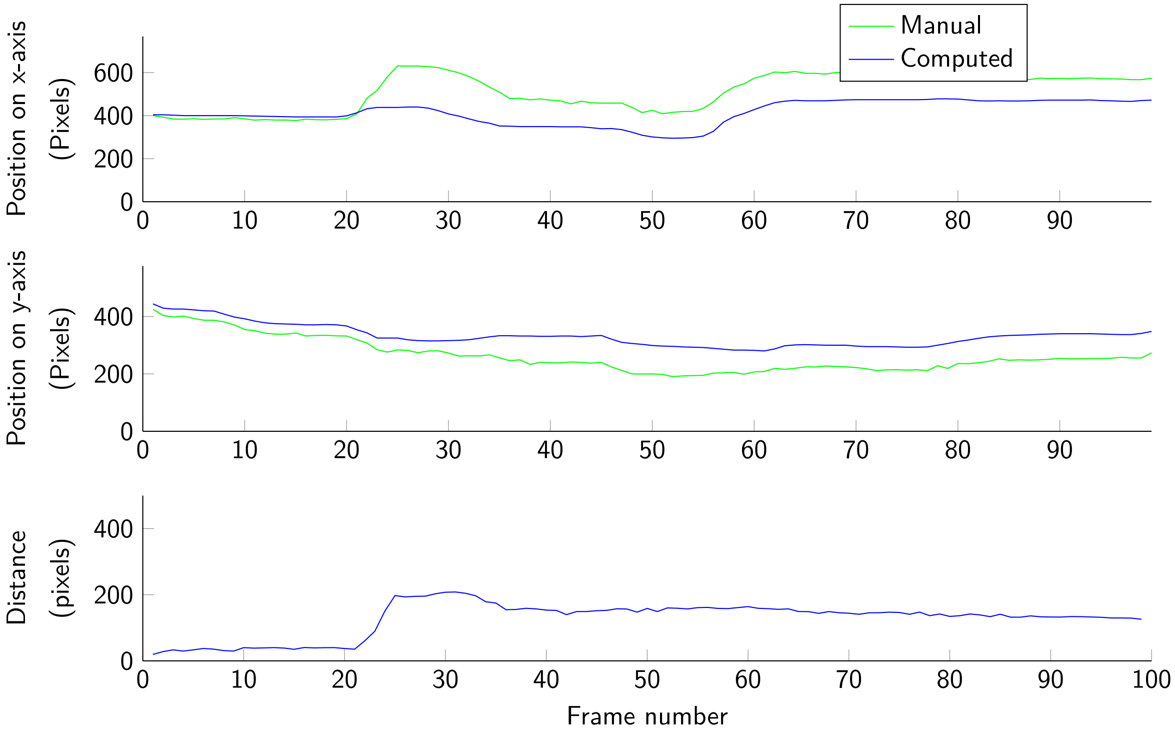


Figure 3.2: An example of inaccurate tracking (sequence 1). The disturbing factor has been motion blur which started around frame 20.

error occurs at frame 20. This is the result of quick movement of the endoscope tip at that frame. For the other inaccurately tracked sequences similar explanations may be found, namely endoscope rotation (sequence 8), target location near the edge of the field-of-view (FOV) (sequence 10), motion blur (sequence 11), and zooming (sequence 12).

3.3 Discussion

Certain disturbing factors have larger effect on the tracking accuracy than others. Instruments present in the FOV, smoke development due to electrocautlery, and dye injections hardly affect the accuracy. However, quick movements resulting in motion blur have a pronounced effect on the accuracy. This my be explained by the motion blur that occurs after quick motions, which decreases image texture and thus the number of features detected. However, the quick movement is most likely the result of the endoscopist adjusting the tip of the endoscope, which is impossible during the use of the target lock, as the endoscope is controlled robotically. Therefore, the endoscope has to be controlled in such a way that the tip does not exceed a certain speed limit.

Other disturbing factors such as zooming, rotation, and the target residing near the edge of the FOV are all related to the calculation of the displacement from the optical flow field. Currently, a simple mean of all vectors is calculated and used to update the target location. This calculation does not take into account rotations and depth changes. Therefore, it is recommended that a more advanced way of determining the target displacement from the optical flow field is implemented, based for example on quasi-spherical triangles[61].

Although the image sequences are diverse, they are not particularly long (approximately 3 seconds). The tracking validation may be improved by using longer image sequences of minutes instead of seconds. However, as we already noticed from using 12 sequences of 100 frames, the manual annotation is a relatively long and repetitive task, not eagerly performed by expert endoscopists. A solution however may be to have the endoscopist annotate only one in twenty frames (or less). This will increase the effective annotated image sequence to 2000 frames (approximately one minute). However, sudden changes in the target location may be missed by the annotators in this way.

In conclusion, the target tracking performs well under circumstances such as instruments in the FOV, dye injections, and smoke development. However, quick motions, rotations, and the target residing near the edge of the FOV have a clear effect of the accuracy of the tracking, which may be resolved by implementing a more advanced way of calculating the target displacement from the optical flow field. Finally, the robustness of the tracking is good as a large number of optical flow vectors is found and the tracking is performed in real-time.

Chapter 4

Locking Validation

After the successful validation of target tracking, the logical next step is the integration of the algorithm with the TeleFlex robotic endoscopy system. However, the control of the TeleFlex system has to be evaluated in order to know whether the information sent by the tracking algorithm is converted into the correct movement of the endoscope tip.

For this purpose, we have set up the following requirements on which the the experiments are based:

- Quick movement towards updated target location
- Slow enough movement to prevent motion blur
- Minimal overshoot from the controller

This chapter starts with a mathematical description of the problem, followed by the measurements preformed and the materials used. The results are divided into a section for the measurement calibration[a] and the actual validation measurements. The results are finally discussed.

4.1 Materials and methods

In order to compare the target location as calculated by the target tracking algorithm with the ground truth, we first need to reconstruct the target location on the image plane to a three-dimensional (3D) point in the real world. For this reconstruction, the homogeneous transformation matrices ${}^{c}\mathbf{H}_{e}$ and ${}^{b}\mathbf{H}_{t}$ are necessary, which are calculated during the system calibration. In the subsequent sections, first the homogeneous matrix will be explained, followed by the measurement calibration. Then, the reconstruction of the target location to a 3D point is described, and finally a description of the materials used is provided.

4.1.1 Homogeneous Transformation

If we assume rigid transformations between different coordinate systems, then the transformation for one coordinate system to another includes solely a rotation and a translation. Given the arbitrary coordinate systems a

---

[a] Note the three difference calibrations used in this chapter: the optical tracking calibration is the calibration of OptiTrack measurement system; the camera calibration is used to calculate the intrinsic and extrinsic parameters of the camera in the endoscope tip (section 4.1.2); and the measurement calibration is the overall calibration before the actual measurements in this chapter, containing both previous calibrations and results into calibration matrix ${}^{c}\mathbf{H}_{e}$.

Abbreviation	Coordinate system
b	Chess- or image Board
c	Camera
e	Endoscope tip trackable
o	Origin of the optical tracking system
t	Target trackable

Table 4.1: The abbreviations of the coordinate systems used in this chapter are summarised here. The calibration matrix ${}^{c}\mathbf{H}_{e}$ for example is a homogeneous transformation matrix from the coordinate system of the endoscope tip to the coordinate system of the camera. Appendix A contains a visual representation of the coordinate systems.

and b, and a three-dimensional point in coordinate system a described by column vector $\mathbf{u}_a$ and the same point in coordinate system b ($\mathbf{u}_b$) is given by

$$\mathbf{u}_b = {}^{b}\mathbf{R}_{a} \cdot \mathbf{u}_a + {}^{b}\mathbf{t}_{a}$$

in which ${}^{b}\mathbf{R}_{a}$ is defined as the 3-by-3 rotation matrix and ${}^{b}\mathbf{t}_{a}$ as the 3-by-1 translation vector from coordinate system a to b. Now, we simplify the equation to the following defining the point in homogeneous coordinates as $\hat{\mathbf{u}} = \begin{bmatrix} \mathbf{u} \\ 1 \end{bmatrix}$,

$$\hat{\mathbf{u}}_b = \begin{bmatrix} & {}^{b}\mathbf{R}_{a} & & {}^{b}\mathbf{t}_{a} \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \mathbf{u}_a \\ 1 \end{bmatrix} = {}^{b}\mathbf{H}_{a} \cdot \hat{\mathbf{u}}_a$$

where ${}^{b}\mathbf{H}_{a}$ is the homogeneous transformation matrix from coordinate system a to b. Thus, the rigid transformation is significantly simplified to a single matrix multiplication, from which two useful properties can be derived. Homogeneous transformation matrices can be cascaded

$${}^{c}\mathbf{H}_{a} = {}^{c}\mathbf{H}_{b} \cdot {}^{b}\mathbf{H}_{a}$$

(introducing another trivial coordinate system c) and inverted

$${}^{b}\mathbf{H}_{a}^{-1} = {}^{a}\mathbf{H}_{b} \quad \text{or} \quad {}^{a}\mathbf{H}_{b} \cdot {}^{a}\mathbf{H}_{b}^{-1} = \mathbf{I}$$

in which $\mathbf{I}$ is the 4-by-4 identity matrix. Note however that there are limitations to rigid transformations, namely in the assumptions that scaling and skewing between coordinate systems is absent.

The coordinate systems used in the remainder of this chapter are abbreviated and summarised in table 4.1. Next, the measurement calibration will be performed.

4.1.2 Measurement Calibration

Calibration matrix ${}^{c}\mathbf{H}_{e}$ cannot be measured directly, but it can be calculated by using the following equation:

$${}^{c}\mathbf{H}_{e} = {}^{c}\mathbf{H}_{b} \cdot {}^{b}\mathbf{H}_{t} \cdot {}^{t}\mathbf{H}_{o} \cdot {}^{o}\mathbf{H}_{e} \tag{4.1}$$

The first step in this equation is the calculation of the matrix ${}^{c}\mathbf{H}_{b}$ by means of a camera calibration.

Camera Calibration

A camera calibration is a method of quantifying the properties of a camera. We have used the model and method described by Bouguet[74]. This method consists of recording a set (i) of calibration images of a chessboard at different locations and orientations inside the field-of-view of the camera. The locations of inner chessboard corners are known in the chessboard coordinate system and these corners are matched to the corners extracted from the images. The intrinsic and extrinsic parameters of the camera can subsequently be calculated. First the intrinsic parameters are described.

The intrinsic parameters are parameters inherent of the camera, namely the camera matrix (containing the focal distance and principal point) and distortion coefficients. These intrinsic parameters can be used to transform the distorted image created by the endoscope camera to an image that fits the rigid transformation requirement. The camera calibration model used by Bouguet[74] takes into account radial and tangential distortions. Radial distortions are distortions in which the magnification increases or decreases with distance from the optical axis. This kind of distortion is intentional for the fish-eye lenses present in endoscopes, as it increases the field-of-view of the camera. Tangential distortion occurs when the lens and the imaging plate (or CCD-chip) are not perfectly aligned. Generally, endoscope cameras are well manufactured, thus we expect tangential distortion to play only a small role.

The distortions are modelled in the following way[74]. Let $\mathbf{x}_c = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix}$ be a point in the coordinate system c. The normalised image projection of $\mathbf{x}_c$ in coordinate system c is given by:

$$\mathbf{x}_n = \begin{bmatrix} x_c/z_c \\ y_c/z_c \end{bmatrix} = \begin{bmatrix} x_n \\ y_n \end{bmatrix} \tag{4.2}$$

The distorted normalised point in coordinate system c $\mathbf{x}_d$ is defined as:

$$\mathbf{x}_d = \begin{bmatrix} x_d \\ y_d \end{bmatrix} = (1 + k_1 r^2 + k_2 r^2 + k_3 r^6)\mathbf{x}_n + \mathrm{d}\mathbf{x}$$

where $r^2 = x_n^2 + y_n^2$, k_1, k_2, k_3 are radial distortion coefficients, and $\mathrm{d}\mathbf{x}$ is the tangential distortion vector defined as:

$$\mathrm{d}\mathbf{x} = \begin{bmatrix} 2p_1 x_x y_n + p_2(r^2 + 2x_n^2) \\ p_1(r^2 + 2y^2) + 2k_2 x_n y_n \end{bmatrix}$$

in which p_1 and p_2 are tangential distortion coefficients.

Now the distorted normalised point may be projected onto the image plane as point $\mathbf{x}_p$ (in a homogeneous vector in pixel coordinates) by using camera matrix $\mathbf{K}$:

$$\mathbf{x}_p = \begin{bmatrix} x_p \\ y_p \\ 1 \end{bmatrix} = \mathbf{K} \cdot \begin{bmatrix} \mathbf{x}_d \\ 1 \end{bmatrix} \quad \text{with} \quad \mathbf{K} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{4.3}$$

The camera matrix contains the focal distance $\mathbf{f}$ and the principal point $\mathbf{c}$ (the intersection of the image plane and the optical axis), both expressed in horizontal and vertical pixels. Therefore, the conversion from undistorted points in the real world (coordinate system c) to distorted points on the image plane is complete. This leaves the transformations from points in the chessboard coordinate system b to the undistorted points in coordinate system of the camera c.

These transformations are defined as the extrinsic parameters of the camera calibration, because they are not dependent on the camera itself, but on the position and orientation of the chessboard. Bouguet[74] assumes a rigid transformation between camera c and chessboard b, therefore the extrinsic parameters can be directly converted into homogeneous transformation matrices ${}^{c}\mathbf{H}_{b}(i)$.

Calibration Matrix

Because we have obtained a set of transformation matrices equation 4.1 is slightly modified to accommodate for the set of images.

$${}^{c}\mathbf{H}_{e}(i) = {}^{c}\mathbf{H}_{b}(i) \cdot {}^{b}\mathbf{H}_{t} \cdot {}^{t}\mathbf{H}_{o}(i) \cdot {}^{o}\mathbf{H}_{e}(i)$$

The first set of matrices ${}^{c}\mathbf{H}_{e}(i)$ is known from the camera calibration. The latter two are obtained by measuring the location and orientation of the endoscope tip trackable[b] (${}^{o}\mathbf{H}_{t}(i)$) and the target trackable using an optical tracking system, synchronously the acquisition of the calibration images. The final unknown matrix is the constant matrix ${}^{b}\mathbf{H}_{t}$. The translation part of ${}^{b}\mathbf{H}_{t}$ is measured: the distance from the origin of the optical tracking system to the origin of the chessboard b is measured by callipers and added to the distance from the origin of the optical tracking system to the target trackable. We assume that the rotation from t to b is described by right angles, which are derived visually. Finally, the set of matrices ${}^{c}\mathbf{H}_{e}(i)$ are calculated, which ought to be constant, assuming that the endoscope tip trackable does not move with respect to the camera. Therefore, the deviation in ${}^{c}\mathbf{H}_{e}(i)$ is a measure for how well the following measurements will be and thus is used to evaluate the measurement system.

The set of matrices ${}^{c}\mathbf{H}_{e}(i)$ are averaged to obtain the calibration matrix ${}^{c}\mathbf{H}_{e}$. The translation and rotation parts of the matrix are averaged individually. The translation is averaged for every direction (x, y, and z), for example the translation in the x-direction t_x:

$$\bar{t}_x = \frac{\sum\limits_{i=1}^{n} t_x(i)}{n}$$

The rotation part is converted to quaternions which can subsequently be averaged by using the method described by Markley et al[78]. An advantages of this method is that the orientation may be weighted and that quaternions cannot contain discontinuities as for example Euler angles can. This method is used instead of the singular value decomposition of the rotation matrices, as there is no need to check the orthogonality of the matrix. However, rotation matrices may be converted to quaternions incorrectly is certain cases, therefore it is advisable to the singular value decomposition in the future.

In the method described by Markley, we first calculate the 4-by-4 matrix $\mathbf{M}$:

$$\mathbf{M} = \sum_{i=1}^{n} w_{\mathbf{R}}(i)\mathbf{q}(i)\mathbf{q}(i)^{\mathrm{T}}$$

in which $\mathbf{q}$ are the quaternions and $w_{\mathbf{R}}$ are optional weights. We decided to weigh every measurement equally thus $w_{\mathbf{R}}$ were an array of ones. The eigenvector corresponding to the maximal eigenvalue of $\mathbf{M}$ is the average quaternion $\bar{\mathbf{q}}$ which is easily calculated in MATLAB. The average quaternion is subsequently converted to an average rotation matrix. The average rotation matrix and the average translation vector are combined into the required calibration matrix ${}^{c}\mathbf{H}_{e}$.

---

[b] A trackable is a virtual object created by the optical tracking system from the IR reflectors in order to obtain location and orientation of the endoscope tip and target board. (Section 4.1.4 and figure 4.2c)

4.1.3 Validation Measurements

For the validation measurements we need to know the position of the target location given by the tracking algorithm in the real world. Generally, to solve such a problem, multiple images are taken at different angles to reconstruct the 3D point. However, we only have one image for every 2D target location. We are still able to reconstruct the 3D point of the target location by using the transformation matrices calculated by the measurement calibration and the optical tracking system, and by assuming that the reconstructed point resides on the target board. The next section will explain the method of reconstruction.

3D Point Reconstruction

Given a single measurement of the experimental set-up, we first obtain from the tracking algorithm the target location in pixels $\mathbf{x}_p$. The target location is first converted into the normalised image projection $\mathbf{x}_n$ by

$$\hat{\mathbf{x}}_n = \mathbf{K}^{-1} \cdot \hat{\mathbf{x}}_p$$

in which $\mathbf{K}$ is the camera matrix from equation 4.3. (Again the circumflex is used to indicate a homogenised vector.) Note that the undistortion of the normalised image projection is not required, because the tracking algorithm already uses undistorted points to calculate the target location (section 2.3.3). During the measurements we will use the camera matrix and distortion coefficients calculated at the measurement calibration.

Secondly, the homogeneous transformation matrices of the endoscope tip and target board, ${}^{o}\mathbf{H}_e$ and ${}^{o}\mathbf{H}_t$ respectively, are returned by the optical tracking system. The next step is calculating matrix ${}^{c}\mathbf{H}_b$ given by the following equation:

$${}^{c}\mathbf{H}_b = {}^{c}\mathbf{H}_e \cdot {}^{e}\mathbf{H}_o \cdot {}^{o}\mathbf{H}_t \cdot {}^{t}\mathbf{H}_b \tag{4.4}$$

The two constant matrices (${}^{c}\mathbf{H}_e$ and ${}^{b}\mathbf{H}_t$) in this equation are calculated during the measurement calibration, while the other two matrices ${}^{o}\mathbf{H}_t$ and ${}^{o}\mathbf{H}_e$ are obtained by the optical tracking system, which are different for every measurement.

Now, we need to reconstruct the normalised image projection $\hat{\mathbf{x}}_n$ into a 3D point in the camera coordinate frame $\mathbf{x}_c$ using the inverse of equation 4.2:

$$\mathbf{x}_c = \begin{bmatrix} x_c/z_c \\ y_c/z_c \\ 1 \end{bmatrix} \cdot z_c = \hat{\mathbf{x}}_n \cdot z_c \tag{4.5}$$

Unfortunately, the depth z_c is not directly available, but we may calculate z_c by using a recursive algorithm (figure 4.1). We define the approximation of depth z_c as d and initialise the algorithm (at $n = 0$) by equalling $d(0)$ to the distance between coordinate system b and c in the z-direction, which is the twelfth element of matrix ${}^{c}\mathbf{H}_b$. We start the algorithm by calculating $\mathbf{x}_c(n)$ (equation 4.5) and subsequently transform $\mathbf{x}_c(n)$ to coordinate system b using

$$\hat{\mathbf{x}}_b(n) = {}^{b}\mathbf{H}_c \cdot \hat{\mathbf{x}}_c(n)$$

.

Assuming that point $\mathbf{x}_b(n)$ is located on the target board, $z_b(n)$ should equal zero. However, this is not the case as only an approximation of the depth d was used instead of the actual depth z_c. Therefore we correct $d(n)$ by adding $z_b(n)$:

Number	Type	Direction	Speed	Zooming	Tip control
1	Step	Horizontal	Constant	Off	Off
2	Periodical	Horizontal	Constant	Off	Off
3	Periodical	Horizontal	Constant	On	Off
4	Periodical	Horizontal	Increasing	Off	Off
5	Periodical	Vertical	Increasing	Off	Off
6	Continuous	Circular	Constant	Off	Off
7	Continuous	Circular	Constant	Off	On

Table 4.2: An overview of the seven measurements performed. The movement of the target board attached to the robotic arm is described. 'Type' describes the form of the movement and 'zooming' refer to whether the zoom estimation is activated. In the last measurement the endoscope tip is controlled by the TeleFlex system using input from the tracking algorithm.

$$d(n+1) = d(n) + z_b(n)$$

Using the improved guess $d(n+1)$, start again from equation 4.5 and calculate $\mathbf{x}_b(n+1)$. After multiple iterations, $z_b(n)$ will converge to zero

$$\lim_{n \to \infty} z_b(n) = 0 \quad \text{provided that} \quad z_b(n) < z_c - d(n)$$

resulting in a accurate estimation of point $\mathbf{x}_c$. We stop the recursion when the absolute error in the assumption that $\mathbf{x}_b$ resides on the target board is smaller than one millionth, $|z_b| < 10^{-6}$. The reconstruction algorithm is stable if $z_b(n)$ is smaller than $z_c(n) - d(n)$, which can be derived from figure 4.1.

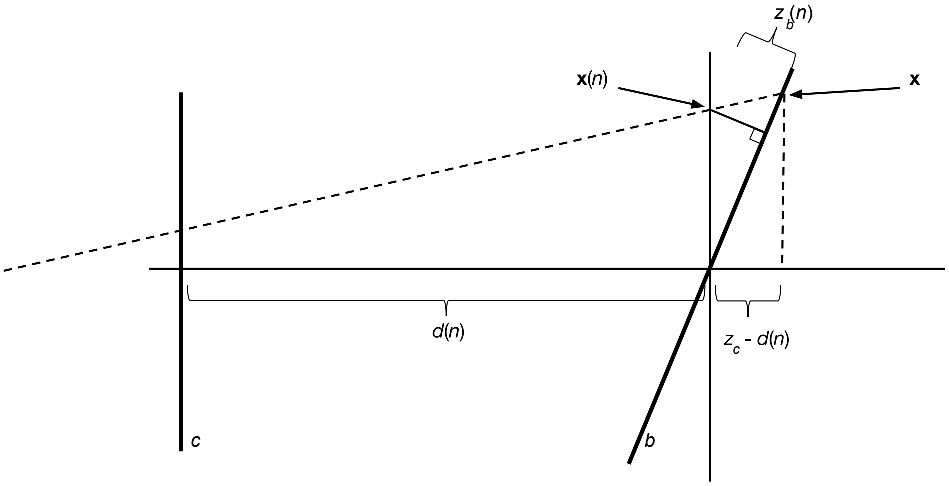
By using this reconstruction algorithm on the measurements, we are able to calculate the set of points $\mathbf{x}_b(i)$ which is the target location on target board b over time. If the target lock were perfect, the values of $\mathbf{x}_b(i)$ should be constant; the target should now move on the target board. By subtracting the first value $\mathbf{x}_b(1)$ of the initialisation of the tracking algorithm from the entire measurement $\mathbf{x}_b(i)$, we obtain the deviation from the initialised target. From this deviation, the Euclidean distance is calculated, resulting in the distance from the initialised target.

Six measurements are performed, summarised in table 4.2. Using these measurements we will evaluate latency (1), accuracy over time (2, 4, 5, and 6), accuracy while increasing target speed (4 and 5), zoom estimation from section 2.5.3 (2 and 3). The zoom estimation should be correlated to the the depth of the target location (z_c), which is determined by the intraclass correlation (ICC)[76]. Finally we will assess the effect of the control of the endoscope tip on the tracking.

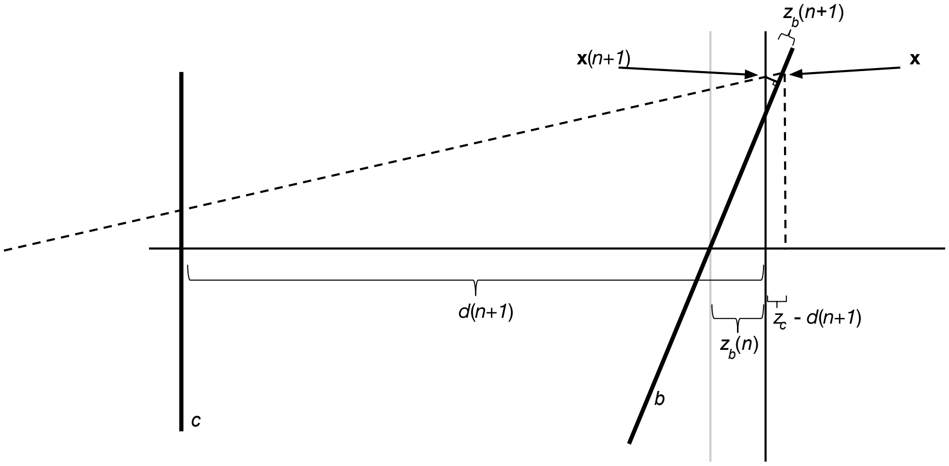
4.1.4 Materials

The TeleFlex control software has been written in Python 2.4, therefore the controller of the target lock is written in the same language. It is implemented as a separate module of the TeleFlex system, complying to the modular nature of the TeleFlex system. The tracking algorithm, written in C++, is wrapped into a Python extension, which makes for easy loading into the target lock controller.

A flexible colonoscope (Olympus CF-H180AL, Hamburg, Germany) together with a compatible video processor (Olympus, Hamburg, Germany) were used for video acquisition. The video stream was transported to a HP EliteBook 8570w via a FireWire connection. The endoscope was controlled by the TeleFlex system, which was connected to the laptop by means of an Ethernet connection. The endoscope was fixated on a table using tape (figure 4.2a).



(a) At $n = 0$



(b) At $n = 1$

Figure 4.1: The first and second step in the recursive reconstruction algorithm are depicted here. Only two dimensions are shown, but the algorithm remains valid in three dimensions. The fat lines represent the image plane in coordinate system c and the target board in coordinate system b which are not aligned. In the upper image, the estimated point $\mathbf{x}(n)$ and the error in the estimation $z_b(n)$ are given. Note that $z_b(n) < z_c - d(n)$ regardless of the orientation of b. Also z_b greatly decreases for every iteration. Pseudocode is available in appendix B.2

The target[c] image (figure 2.8b) to be tracked was a printed frame from a colonoscopy video recorded during the clinical validation (chapter 3). This frame was undistorted using the intrinsic camera parameters (section 4.1.2) from a similar scope, as is is very impractical to determine the intrinsic parameters of colonoscopes in the clinical routine. The frame chosen was a depiction of an inflamed colon to increase the image texture. An image with a high texture was chosen, because it will return more image features, which compensates for the loss of image features in the printed frame compared to the original. As the tip of the endoscope was not residing in a closed space such as the colon, obtaining sufficient and diffuse illumination was difficult with the light from the endoscope tip alone. Therefore, we have chosen to illuminate the target using another light source from behind the printed frame. We have built a wooden board filled with light emitting diodes (LEDS, with a total power of 30 W) on top of which a transparent plastic board is attached at a distance of 19.5 mm. The printed frame is easily attached to the transparent board by using binder clips. In order to have the target move in a predictable and reproducible manner, we have decided to fix the target to a robotic arm (KUKA LBR4+, Augsburg, Germany).

In order to validate the target lock module, the state of the target and the endoscope tip are required. These are recorded by using the OptiTrack (NaturalPoint, Inc., Corvallis, OR, Unites States) infra-red (IR) optical tracking system. IR reflectors were attached to the wooden board of the target using bolts, while at the endoscope tip, the reflectors were attached non-symmetrically via a 3D-printed template (figure 4.2c). Virtual objects called trackables were created in the OptiTrack software from the IR reflectors on the tip and target board. The optical tracking system was calibrated before the measurements. The optical tracking data was sent to the laptop by means of an Ethernet connection. Finally, in order to smoothen the tracking information, a Kalman filter was activated in the optical tracking software, set to 0.1 for each trackable.

4.2 Results

4.2.1 Measurement Calibration

The homogeneous transformation matrices acquired are visually assessed on validity. One example of this assessment is matrix ${}^{e}\mathbf{H}_t$, which is compromised of both matrices acquired by the optical tracking system, is depicted in figure 4.3.

The camera calibration had a reprojection error of no more than one pixel of any corner in either the x- or y-direction (figure 4.4). The following values for the camera matrix and distortion coefficients were recorded:

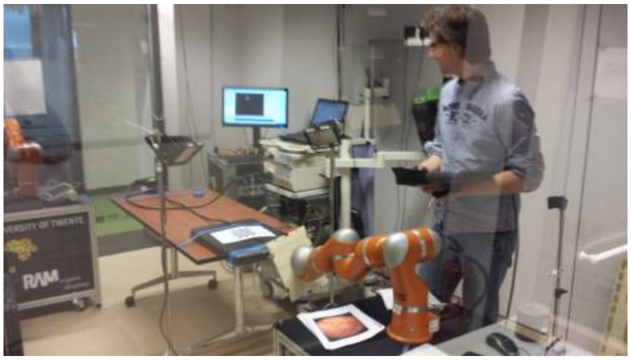
$$\mathbf{K} = \begin{bmatrix} 410.7 & 0 & 365.3 \\ 0 & 441.1 & 272.0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} k_1 \\ k_2 \\ p_1 \\ p_2 \\ k_3 \end{bmatrix} = \begin{bmatrix} -0.4739 \\ 0.1440 \\ 0.0049 \\ 0.0015 \\ 0 \end{bmatrix}$$

This means that the principal point (third and sixth element of $\mathbf{K}$) is close to the image centre ($\begin{bmatrix} 360 \\ 288 \end{bmatrix}$), and that the tangential distortion is nearly negligible.

With all homogeneous transformation matrices from equation 4.1 known, thus ${}^{c}\mathbf{H}_e(i)$ is calculated. The translation vectors and the Euler angles calculated from the rotation are displayed in figure 4.5. ${}^{c}\mathbf{H}_e(i)$ is fairly constant over the calibration measurements, with a variance of 1.93, 1.46, and 0.57 mm in the x-, y-, and z-directions of the translation vector respectively.

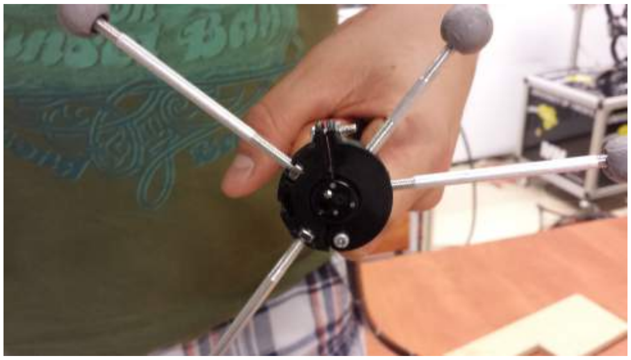
---

[c] Note the difference in the following three definitions using the term target: the target board is the board on which an image of an inflamed colon is placed; the target trackable is the trackable of the optical tracking system consisting of the three IR reflectors on the wooden board; the target location is the point in place calculated by the tracking algorithm which the set point for the controller.



(a) Overview

(b) Illumination



(c) Trackable tip

Figure 4.2: Three photographs of the experimental set up. Figure 4.2a displays the used set up as a whole, in which the TeleFlex system, endoscope, and KUKA-arm may be identified. Figure 4.2b shows the intricate pattern of LEDs used to obtain sufficient illumination of the target. Lastly, figure 4.2c shows the endoscope tip with the 3D-printed template holding the IR reflectors.

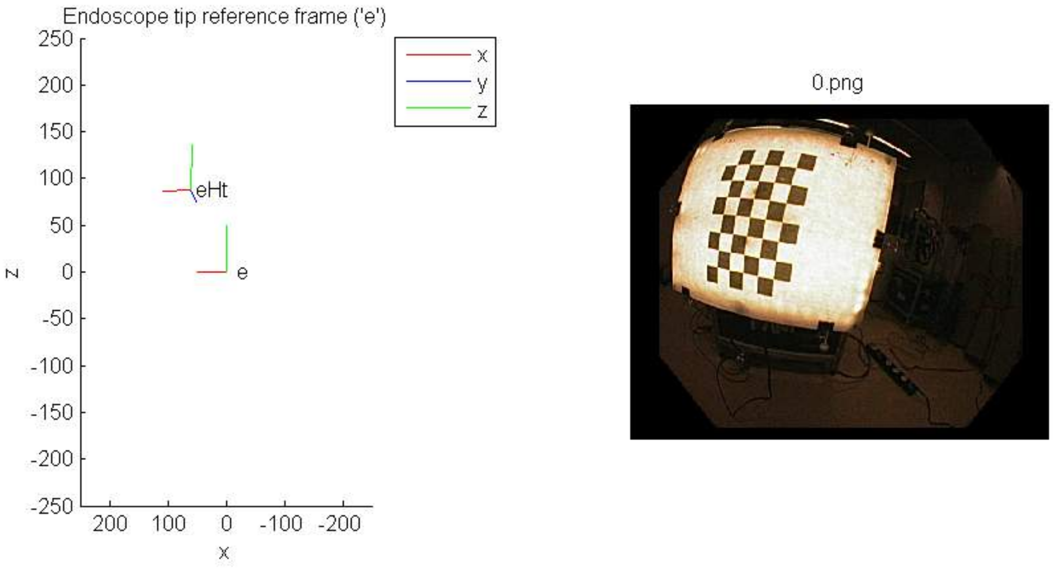


Figure 4.3: The matrix ${}^{e}\mathbf{H}_{t}$ is plotted to the left, next to the corresponding image by the endoscope. These figures are used for all images in the measurement calibration to visually check the validity of the acquired data.

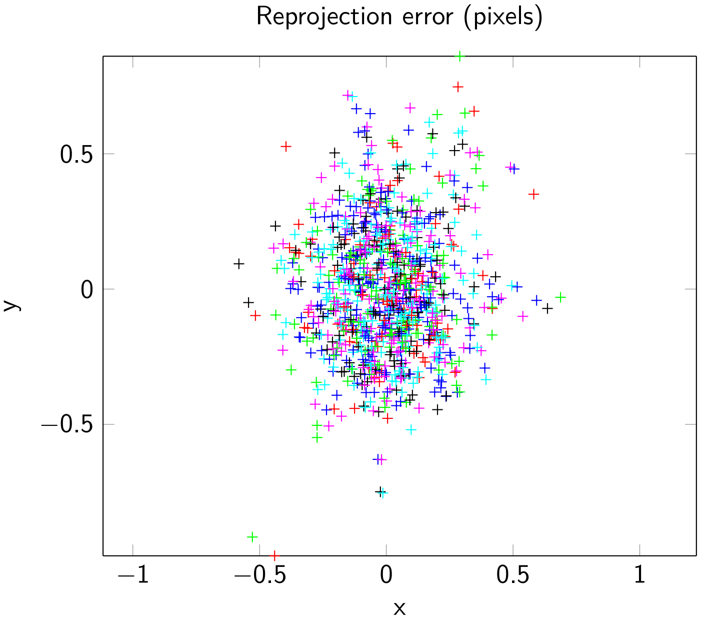


Figure 4.4: The corners of the chessboard as determined by the user input are subtracted from the corners calculated by reprojecting the corners onto the image plane. This reprojection error is given in pixels in the x- and y-directions. Different colours represent different images of the chessboard.

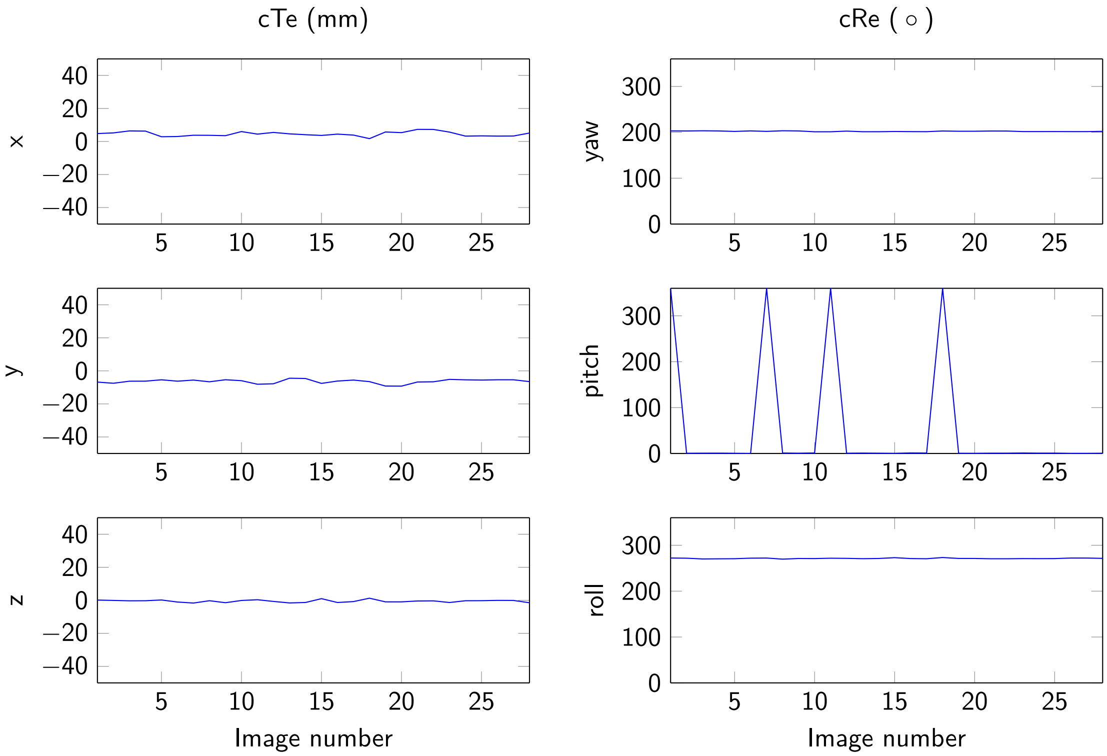


Figure 4.5: The left three plots display the translation of the matrix ${}^{c}\mathbf{H}_e(i)$, while the right three plots display the rotation of ${}^{c}\mathbf{H}_e(i)$ converted into Euler angles (convention ZYX). The constancy of the graphs is a measure for the accuracy of the system calibration. The peaks in the pitch can be explained by slightly deviating around the singularity of 0/360 degrees. This discontinuity is not present in the rotation matrices, however Euler angles are more intuitive and more easily visualised.

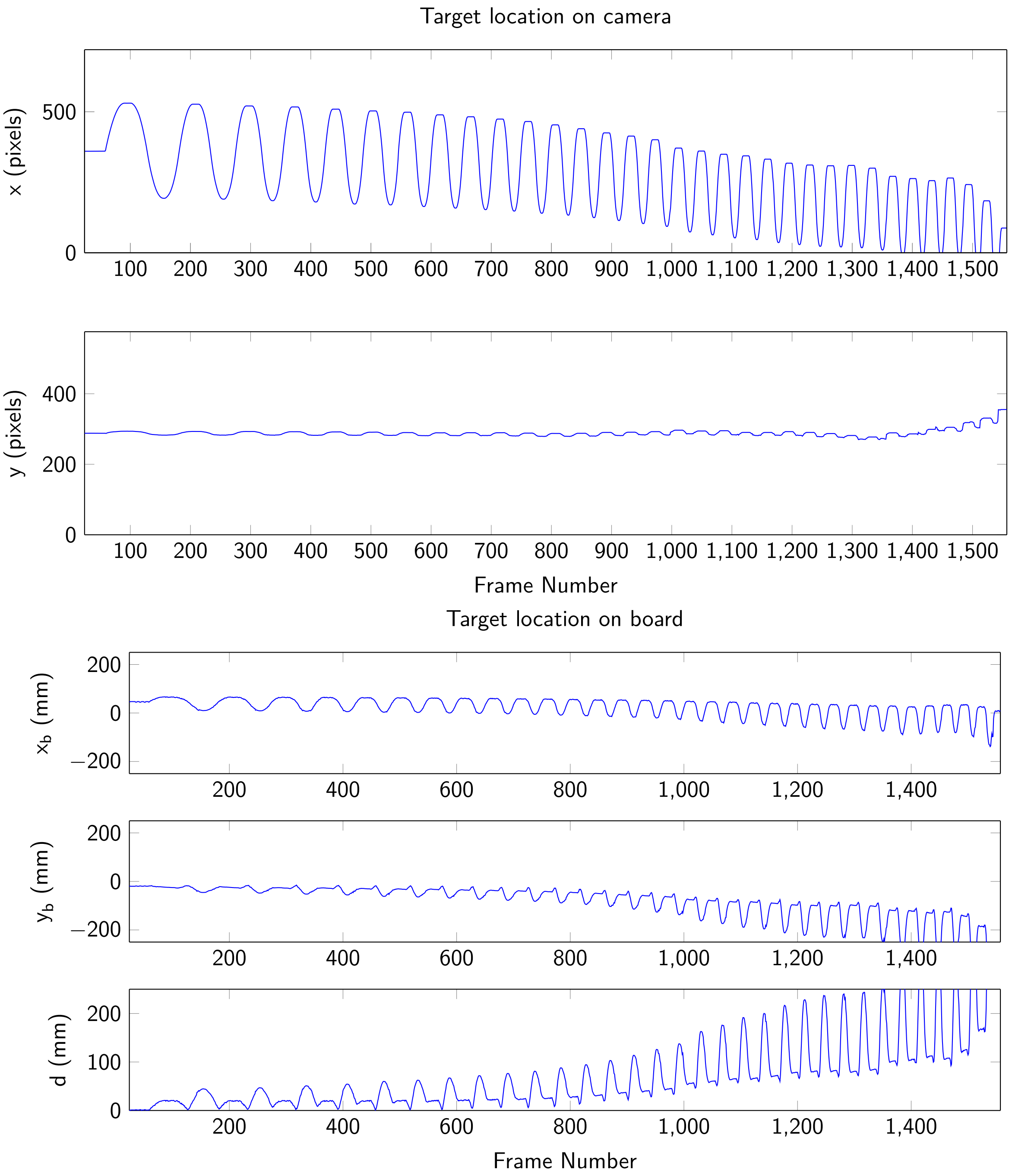
The set of matrices ${}^{c}\mathbf{H}_e(i)$ is averaged over the number of calibration images, returning the following matrix:

$$
{}^{c}\mathbf{H}_e = \begin{bmatrix} -0.926 & -0.377 & -0.009 & 4.558 \\ 0.016 & -0.017 & -1.000 & -6.342 \\ 0.377 & -0.926 & 0.022 & -0.502 \\ 0 & 0 & 0 & 1 \end{bmatrix}
$$

4.2.2 Validation Measurements

The reconstruction algorithm was tested on chessboard corners from the measurement calibration and showed errors smaller than 10^{-6} in x-, y-, and z-values) after less than 10 iterations. The latency between the movement of the target board and the movement of the target registered by the tracking algorithm ranges from 6 to 12 frames (mean 9.2 frames). The throughput of the tracking algorithm ranges from 17 to 22 FPS.

Subsequently, the tracking without control of the endoscope tip are displayed for a horizontal periodical motion (figure 4.6) and a vertical periodical motion (figure 4.7). In figure 4.6, the repeatability of the effect of the movement of the target board can be assessed. Regarding accuracy, the algorithm exponentially acquires drift as the speed of the target board is increased.



Target location on camera

Target location on board

Figure 4.6: Tracking without the control of the endoscope tip. The target board is moved in a periodical horizontal movement with increasing speed. The target location in *c* is shown in the first two graphs, while the location and the Euclidean distance from the initial value (*d*) in *b* is depicted below.

Similar results for a vertical periodical movement of the target board increasing in speed are displayed in figure 4.7. The accuracy of the tracking of vertical movement appears to be better than the tracking of horizontal movement, as there is little drift present, even at higher speeds.

The results of the actual target locking (in which the tip of the endoscope is controlled by the TeleFlex system) are visualised in figure 4.8. Again, a small drift occurs at higher target speeds, decreasing the accuracy. The mean distance between the actual target and the ground truth (d in figure 4.8) is 25.5 mm with a standard deviation of 28.7 mm.

The results of the evaluation of zooming are displayed in figure 4.9. A large drift is present in the zooming approximation by the tracking algorithm, which emerges before the actual movement of the target (at frame number of approximately 1130).

4.3 Discussion

Similarly to the results found in chapter 3, the found that the target tracking accuracy decreases with quick motions due to drift. Again, this may be attributed to the motion blur that occurs with quick motions, which reduces the number of detectable features. However, in measurements with little drift (figure 4.8) the mean distance between actual and calculated target is still significant at 25.5 mm.

This distance may be attributed to the camera calibration model. Although the camera calibration is quite accurate at the centre of the image, a significant error in the model is present at the edges. Other camera models specifically created for fish-eye lenses are developed (and implemented in OpenCV) that may increase the accuracy of the tracking, mainly at the edges of the frames.

Another way of improving the accuracy may be the use of a more sophisticated way of calculating the target displacement instead of the mean of the optical flow vectors. In the current implementation of the tracking algorithm, features that are close to the camera will return a larger flow vectors compared to further features when moved at the same speed. This results in an inaccurate target displacement if a simple averaging of the optical flow vectors is used. The previously mentioned method of Wong et al.[61] using a quasi-spherical triangle-based approach seems promising, however they use a stereoscopic endoscope to gather depth information.

We expected that the tracking accuracy of the vertical movement to be inferior the the accuracy of the horizontal movement, because the actual resolution in the vertical direction is halved due to interlacing. However, our results show the opposite effect, namely the vertical movement is more accurately tracked. This may be the result of image we used as a target, which was directed in the horizontal direction. Therefore, the chance of the tracking algorithm using features from beyond the target image was greater. These features such as the edge between image and board, are easier to track accurately, which may result in an overall increase of the accuracy in the vertical direction. Fortunately, this effect will not show up in the intestines as such edges are not present.

A mean distance between actual target and calculated target of 25.5 mm is very large regarding the size of for example the colon. In the colon, this value would probably be lower, as the distance between the target and the endoscope will be lower. This distance may also be reduced by reducing the latency of the target tracking. Changes in the target location are then more quickly compensated for resulting is a smaller overall distance. However, it is difficult to reduce this latency as we discovered that the latency is mainly caused in the endoscope/video processor/FireWire chain, which cannot be altered.

The zooming functionality, which might have been beneficial for an accurate control of the endoscope tip proved to be to unreliable. Therefore, we decided to turn the zooming off and advise to remove this implementation altogether. A more advanced manner of estimating the depth has to be found. An example is epipolar geometry used by Mori et al.[79], which is very similar to estimations of the focus of expansion[55;64].

Motions in the intestines are more complex than periodical motions in one direction. This has been simulated by a circular movement, which shows similar results to the results in a singular direction. Therefore, we expect

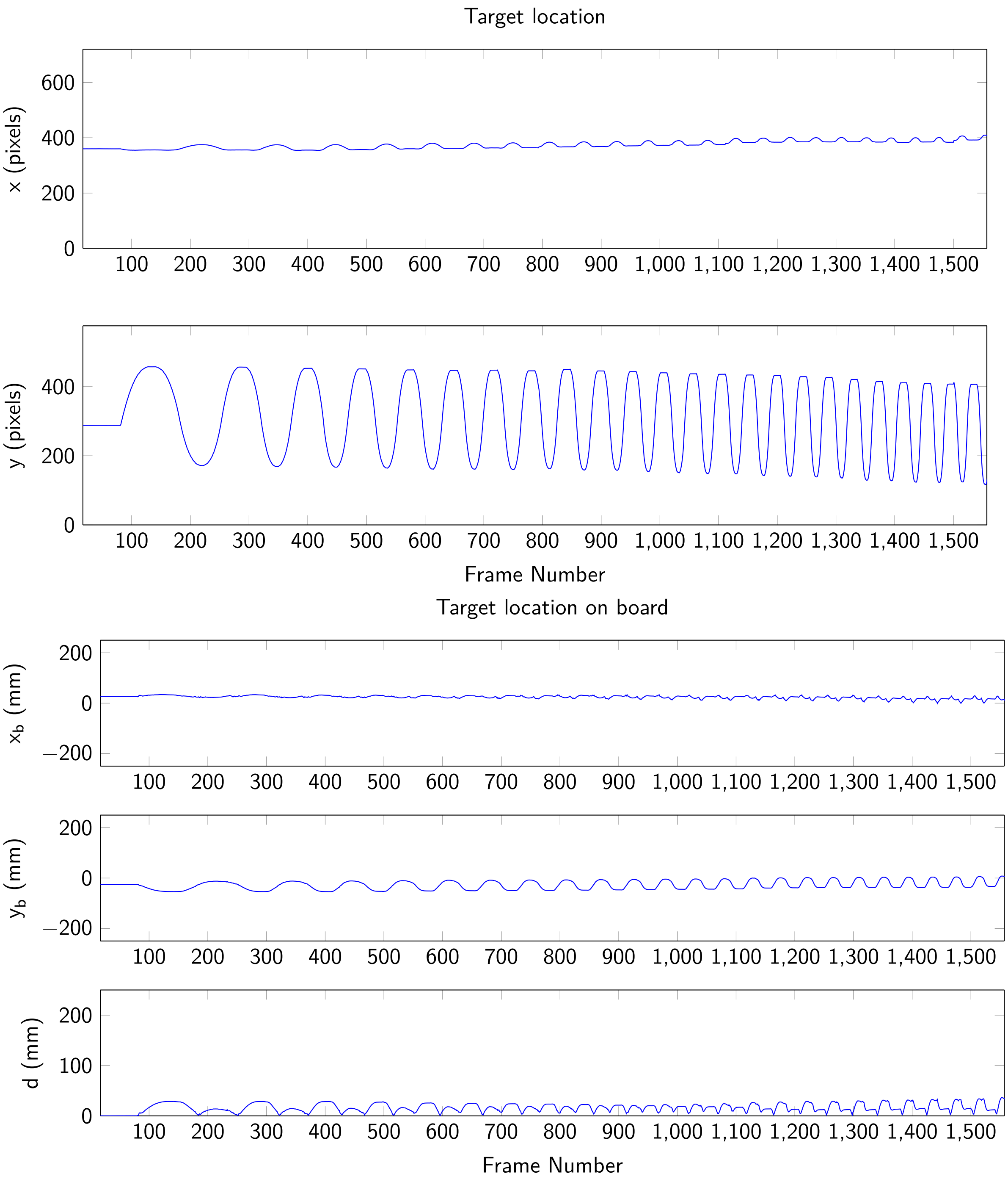


Figure 4.7: Tracking without the control of the endoscope tip. The target board is moved in a periodical vertical movement with increasing speed. The target location in *c* is shown in the first two graphs, while the location and the Euclidean distance from the initial value (d) in *b* is depicted below.

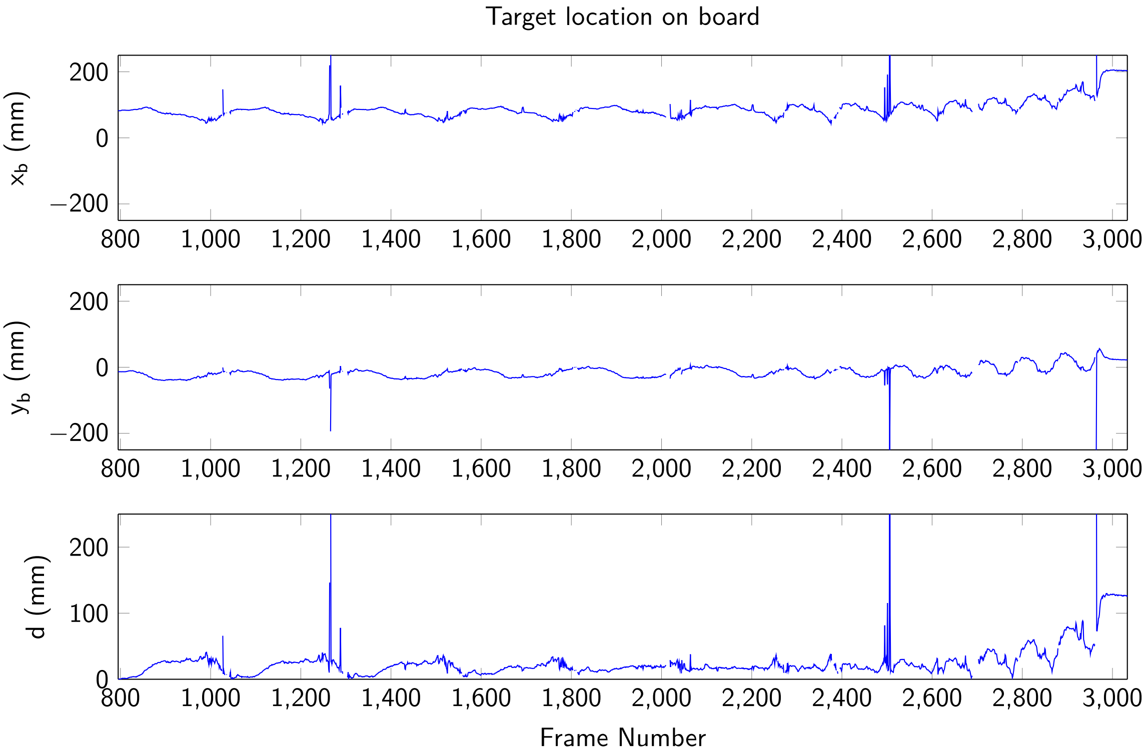


Figure 4.8: The results of the tracking of a circular motion with a control of the endoscope tip using data from the target tracking algorithm. The outliers present are caused by incorrect tracking by the optical tracking algorithm, as the circular motion occasionally resulted into the occlusion of IR reflectors.

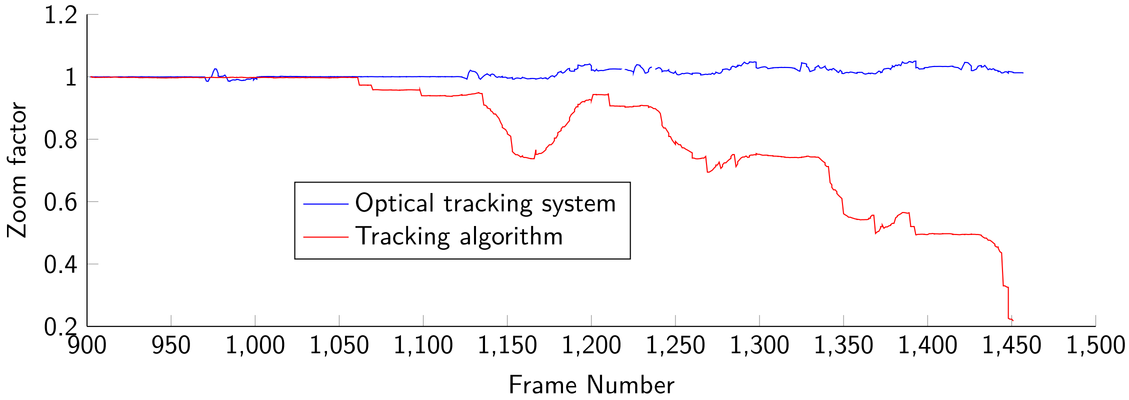


Figure 4.9: Zoom factor as calculated by the tracking algorithm (red) versus the zoom factor calculated by the optical tracking system (blue). If the zooming functionality were perfect, the curve of the tracking algorithm would align with the curve of the optical tracking algorithm, but a significant drift is present in the zoom factor.

the tracking algorithm to be equally accurate and reliable in an in vivo setting. However, this remains to be tested.

For the control of the endoscope tip in the target locking measurement, a controller based on inverse kinematics and hysteresis compensation was used, thanks to the work of Frijnts[66]. However, this controller was still in development, thus improvements may be made resulting in a more accurate control of the tip. This may result in the target being more in the centre of the image, probably resulting in a higher accuracy of the target tracking.

In conclusion, the target tracking was tested in a controlled environment. The reliability of the tracking is very good as shown by the consistent results after equal movements. The tracking remains real-time as the frame rate never falls below 15 FPS. The accuracy may still be improved by using more advanced ways of estimating the depth or calculating the target displacement from the optical flow vectors. Another concern is the latency present between movement of the target and the detection of the movement by the tracking algorithm.

Chapter 5

Conclusion

In this report we have developed a target tracking algorithm for endoluminal interventions using the TeleFlex system. First a review of endoluminal interventions for gastrointestinal endoscopy has been given, culminating into a set of requirements for a target tracking algorithm. Subsequently, the design of the tracking algorithm based on the work by De Jong[65] is given along with the optimisation of the parameters of this algorithm. Next, the tracking algorithm is validated using series of videos containing various disturbances, identifying weak points of the algorithm. Finally, the algorithm is validated in a controlled environment using an optical tracking system.

Overall, the developed tracking algorithm boasts a high reliability, real-time frame through-put, and resistance against disturbances such as instruments and smoke in view. However, the accuracy may be improved by increasing the resistance to quick movements, and by implementing a more advanced way for calculating the target displacement from the optical flow field. Additionally, a controller that uses data from the target tracking algorithm is being developed resulting in a target lock module for the TeleFlex system.

Our recommendations for future work would be the start of a clinical validation of the target lock module. Preferably, the validation will be carried out by expert endoscopists on ex vivo hollow visceral organs. Points of interest would not only be the accuracy and reliability of the target lock, but also ease-of-use and added clinical value. The ease-of-use may be increased by using different initialisations of the target lock, for example: Is a button on the endoscope or TeleFlex workstation, a tap on a screen of a foot switch preferred? The evaluation of the clinical may be measured by the time required to perform a certain complex intervention, or the damage caused to surrounding tissues.

For the future, it is also important to note that the feature detector used in the tracking algorithm, SURF, is patented. Therefore, commercial distribution of the target lock may be difficult. Fortunately, there are other free feature detection algorithms available such as BRISK[80] and FREAK[81]. However, switching to a new feature detector implies a revision of the design, optimisation, and validation.

Furthermore, the target tracking algorithm does not have to be confined to complex endoluminal interventions. Any instance of a robotically-controlled camera is potentially suitable for the target lock software. Examples include the DaVinci system or even drones.

Chapter 6

Bibliography

[1] C. B. Morgenthal, W. O. Richards, B. J. Dunkin, K. a. Forde, G. Vitale, and E. Lin. The role of the surgeon in the evolution of flexible endoscopy. *Surgical Endoscopy and Other Interventional Techniques*, 21(6):838–853, 2006. ISSN 09302794. doi: 10.1007/s00464-006-9109-4.

[2] Katie Schwab and Sukhpal Singh. An introduction to flexible endoscopy. *Surgery (Oxford)*, 29(2):80–84, February 2011. ISSN 02639319. doi: 10.1016/j.mpsur.2010.11.014. URL `http://linkinghub.elsevier.com/retrieve/pii/S0263931910002498`.

[3] R B Bush, H Leonhardt, I V Bush, and R R Landes. Dr. Bozzini's Lichtleiter. A translation of his original article (1806). *Urology*, 3(1):119–123, 1974. ISSN 00904295. doi: 10.1016/S0090-4295(74)80080-4.

[4] Harry W. Herr. Max Nitze, the Cystoscope and Urology. *Journal of Urology*, 176(4):1313–1316, 2006. ISSN 00225347. doi: 10.1016/j.juro.2006.06.085.

[5] Ned. Vereniging voor Anesthesiologie and Ned. Vereniging voor Kindergeneeskunde. Bijlagen Conceptrichtlijn Sedatie en/of analgesie (PSA) op locaties buiten de operatiekamer, 2009. URL `http://www.mdl.nl/richtlijnen2`.

[6] Queensland Centre for Digestive Medicine. Gastroscopy (Upper Endoscopy), 2010. URL `http://www.digestivemedicine.com.au/patientsgastro.php`.

[7] MedlinePlus. ERCP, 2015.

[8] S I Khakoo, a J Lobo, N a Shepherd, and S P Wilkinson. Histological assessment of the Sydney classification of endoscopic gastritis. *Gut*, 35(9):1172–1175, 1994. ISSN 0017-5749. doi: 10.1136/gut.35.9.1172.

[9] Eric S Orman, Nan Li, and Nicholas J Shaheen. Efficacy and durability of radiofrequency ablation for barrett's esophagus: Systematic review and meta-analysis, 2013. ISSN 15423565.

[10] Sjb Van Weyenberg and J Pohl. Gastric Gastrointestinal Stromal Tumor. *Video Journal and Encyclopedia of GI Endoscopy*, 1(1):168–169, 2013. ISSN 22120971. doi: 10.1016/S2212-0971(13)70069-4. URL `http://www.sciencedirect.com/science/article/pii/S2212097113700694`.

[11] Marcie Feinman and Elliott R. Haut. Upper Gastrointestinal Bleeding. *Surgical Clinics of North America*, 94 (1):43–53, 2014. ISSN 00396109. doi: 10.1016/j.suc.2013.10.004. URL `http://dx.doi.org/10.1016/j.suc.2013.10.004`.

[12] Ríos Eddy Castellanos, Pamela Seron, P Gisbert Javier, and Xavier Bonfill Cosp. Endoscopic injection of cyanoacrylate glue versus other endoscopic procedures for acute bleeding gastric varices in patients with portal hypertension. *Cochrane Database of Systematic Reviews*, 5, 2015. ISSN 1469-493X (Electronic).

doi: 10.1002/14651858.CD010180. URL `http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010180/abstract`.

[13] Lise Lotte Gluud and Aleksander Krag. Banding ligation versus beta-blockers for primary prevention in oesophageal varices in adults. *Cochrane database of systematic reviews (Online)*, 8(8):CD004544, 2012. ISSN 1469-493X. doi: 10.1002/14651858.CD004544.pub2. URL `http://www.ncbi.nlm.nih.gov/pubmed/22895942`.

[14] Kyong Hee Hong, Yoon Jae Kim, Jae Hak Kim, Song Wook Chun, Hee Man Kim, and Jae Hee Cho. Risk factors for complications associated with upper gastrointestinal foreign bodies. *World Journal of Gastroenterology*, 21(26):8125–8131, 2015. doi: 10.3748/wjg.v21.i26.8125.

[15] Iyad Khamaysi, Alain Suissa, Kamal Yassin, Ahmad Mahajna, Ahmad Assalia, Bishara Bishara, and Ian M. Gralnek. Endoscopic Removal of Partially Migrated Intragastric Bands Following Surgical Gastroplasty: a Prospective Case Series. *Obesity Surgery*, 25(5):942–945, 2015. ISSN 0960-8923. doi: 10.1007/s11695-015-1629-3. URL `http://link.springer.com/10.1007/s11695-015-1629-3`.

[16] Mehdi Tahiri, Pasquale Ferraro, André Duranceau, Melanie Berthiaume, Vicky Thiff, and Moishe Liberman. Self-expanding metallic stent placement with an exaggerated 5-cm proximal tumor covering for palliation of esophageal cancer. *Annals of Gastroenterology*, 28(3):347–352, 2015.

[17] Waleed Alazmi, Salman Al-Sabah, Daliya AlMohammad Ali, and Sulaiman Almazeedi. Treating sleeve gastrectomy leak with endoscopic stenting: the kuwaiti experience and review of recent literature. *Surgical Endoscopy*, 28:3425–3428, 2014. ISSN 09302794. doi: 10.1007/s00464-014-3616-5.

[18] Martin Müller, Thomas Seufferlein, Lukas Perkhofer, Martin Wagner, and Alexander Kleger. Self-Expandable Metal Stents for Persisting Esophageal Variceal Bleeding after Band Ligation or Injection-Therapy: A Retrospective Study. *Plos One*, 10(6):e0126525, 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0126525. URL `http://dx.plos.org/10.1371/journal.pone.0126525`.

[19] Maciej Matłok, Piotr Major, Michał Pędziwiatr, Marek Winiarski, Piotr Budzyński, Piotr Małczak, Leif Hynnekleiv, and Andrzej Budzyński. Endoscopic insertion of a self-expandable stent combined with laparoscopic rinsing of peritoneal cavity as a method for staple line leaks treatment in patients post laparoscopic sleeve gastrectomy. *Polski Przegląd Chirurgiczny2*, 87(2):238–244, 2015. doi: 20.1515/pjs-2015-0048.

[20] Waseem J. David, Bashar J. Qumseya, Yazen Qumsiyeh, Michael G. Heckman, Nancy N. Diehl, Michael B. Wallace, Massimo Raimondo, Timothy a. Woodward, and Herbert C. Wolfsen. Comparison of endoscopic treatment modalities for Barrett's neoplasia. *Gastrointestinal Endoscopy*, 2015. ISSN 00165107. doi: 10.1016/j.gie.2015.03.1979. URL `http://linkinghub.elsevier.com/retrieve/pii/S0016510715022907`.

[21] Toshitaka Hoppo and Blair a. Jobe. Endoscopy and role of endoscopic resection in gastric cancer. *Journal of Surgical Oncology*, 107(3):243–249, 2013. ISSN 00224790. doi: 10.1002/jso.23126.

[22] Andrea Anderloni, Manol Jovani, Cesare Hassan, and Alessandro Repici. Advances , problems , and complications of polypectomy. *Clinical and Experimental Gastroenterology*, 7:285–296, 2014.

[23] Noriya Uedo, Yoji Takeuchi, and Ryu Ishihara. Endoscopic management of early gastric cancer: endoscopic mucosal resection or endoscopic submucosal dissection: data from a Japanese high-volume center and literature review. *Annals of gastroenterology : quarterly publication of the Hellenic Society of Gastroenterology*, 25(4):281–290, 2012. ISSN 1108-7471. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3959406&tool=pmcentrez&rendertype=abstract`.

[24] Young-Mi Park, Eun Cho, Hye-Young Kang, and Jong-Mann Kim. The effectiveness and safety of endoscopic submucosal dissection compared with endoscopic mucosal resection for early gastric cancer: a systematic review and metaanalysis. *Surgical endoscopy*, 25(8):2666–2677, 2011. ISSN 0930-2794. doi: 10.1007/s00464-011-1627-z.

[25] Georgina Chadwick, Oliver Groene, Sheraz R. Markar, Jonathan Hoare, David Cromwell, and George B. Hanna. Systematic review comparing radiofrequency ablation and complete endoscopic resection in treating dysplastic Barrett's esophagus: A critical assessment of histologic outcomes and adverse events. *Gastrointestinal Endoscopy*, 79(5):718–731.e3, 2014. ISSN 10976779. doi: 10.1016/j.gie.2013.11.030. URL `http://dx.doi.org/10.1016/j.gie.2013.11.030`.

[26] A Mocanu, R Bârla, P Hoara, and S Constantinoiu. Current endoscopic methods of radical therapy in early esophageal cancer. *Journal of Medicine and Life*, 8(2):150–156, 2015.

[27] Pier Alberto Testoni, Sabrina Testoni, and Antonella Giussani. Difficult biliary cannulation during ercp: How to facilitate biliary access and minimize the risk of post-ercp pancreatitis. *Digestive and Liver Disease*, 43 (8):596–603, 2011. ISSN 15908658. doi: 10.1016/j.dld.2011.01.019. URL `http://dx.doi.org/10.1016/j.dld.2011.01.019`.

[28] Michael P. Swan, Michael J. Bourke, Stephen J. Williams, Sina Alexander, Alan Moss, Rick Hope, and David Ruppin. Failed biliary cannulation: Clinical and technical outcomes after tertiary referral endoscopic retrograde cholangiopancreatography. *World Journal of Gastroenterology*, 17(45):4993–4998, 2011. ISSN 10079327. doi: 10.3748/wjg.v17.i45.4993.

[29] Michael Pavlides, Ashley Barnabas, Nilesh Fernandopulle, Adam A Bailey, Jane Collier, Jane Phillips-Hughes, Anthony Ellis, Roger Chapman, and Barbara Braden. Repeat endoscopic retrograde cholangiopancreaticography after failed initial precut sphincterotomy for biliary cannulation. *World Journal of Gastroenterology*, 20(36):13153, 2014. ISSN 1007-9327. doi: 10.3748/wjg.v20.i36.13153. URL `http://www.wjgnet.com/1007-9327/full/v20/i36/13153.htm`.

[30] Benedetto Mangiavillano. Outcome of stenting in biliary and pancreatic benign and malignant diseases: A comprehensive review. *World Journal of Gastroenterology*, 21(30):9038, 2015. ISSN 1007-9327. doi: 10.3748/wjg.v21.i30.9038. URL `http://www.wjgnet.com/1007-9327/full/v21/i30/9038.htm`.

[31] Qing Qing Shi, Xiao Yi Ning, Ling Ling Zhan, Guo Du Tang, and Xiao Ping Lv. Placement of prophylactic pancreatic stents to prevent post-endoscopic retrograde cholangiopancreatography pancreatitis in high-risk patients: A meta-analysis. *World Journal of Gastroenterology*, 20(22):7040–7048, 2014. ISSN 22192840. doi: 10.3748/wjg.v20.i22.7040.

[32] Ryuichi Yamamoto. Comparison of endoscopic stenting for malignant biliary obstruction: A single-center study. *World Journal of Gastrointestinal Endoscopy*, 7(9):889, 2015. ISSN 1948-5190. doi: 10.4253/wjge.v7.i9.889. URL `http://www.wjgnet.com/1948-5190/full/v7/i9/889.htm`.

[33] James DiSario, Ram Chuttani, Joseph Croffie, Julia Liu, Daniel Mishkin, Raj Shah, Lehel Somogyi, William Tierney, Louis M. Wong Kee Song, and Bret T. Petersen. Biliary and pancreatic lithotripsy devices. *Gastrointestinal Endoscopy*, 65(6):750–756, 2007. ISSN 00165107. doi: 10.1016/j.gie.2006.10.002.

[34] Manu Tandan. Extracorporeal shock wave lithotripsy for pancreatic and large common bile duct stones. *World Journal of Gastroenterology*, 17(39):4365, 2011. ISSN 1007-9327. doi: 10.3748/wjg.v17.i39.4365.

[35] Vishal Sharma, Surinder S Rana, Deepak K Bhasin, Vishal Sharma, Surinder S Rana, and Deepak K Bhasin. Endoscopic ultrasound guided interventional procedures. *World Journal of Gastrointestinal Endoscopy*, 7(6): 628–642, 2015. doi: 10.4253/wjge.v7.i6.628.

[36] Carlo Fabbri, Carmelo Luigiano, Andrea Lisotti, Vincenzo Cennamo, Clara Virgilio, Giancarlo Caletti, and Pietro Fusaroli. Endoscopic ultrasound-guided treatments: Are we getting evidence based - A systematic review. *World Journal of Gastroenterology*, 20(26):8424–8448, 2014. ISSN 22192840. doi: 10.3748/wjg.v20.i26.8424.

[37] VanAmburg Surgery Care. Colonoscopy/Endoscopy, 2014. URL `http://www.vanamburgsurgery.com/colonoscopy/`.

[38] Kavin A Ghassemi and Dennis M Jensen. Lower GI bleeding: Epidemiology and Management, 2013.

[39] Prashant Kedia and Jerome D Waye. Colon polypectomy: a review of routine and advanced techniques. *Journal of clinical gastroenterology*, 47(8):657–65, 2013. ISSN 1539-2031. doi: 10.1097/MCG.0b013e31829ebda7. URL `http://www.ncbi.nlm.nih.gov/pubmed/23948754`.

[40] Tonya Kaltenbach and Roy Soetikno. Endoscopic resection of large colon polyps. *Gastrointestinal Endoscopy Clinics of North America*, 23(1):137–152, 2013. ISSN 10525157. doi: 10.1016/j.giec.2012.10.005. URL `http://dx.doi.org/10.1016/j.giec.2012.10.005`.

[41] Gezondheidsraad. Bevolkingsonderzoek naar darmkanker. Technical report, Gezondheidsraad, Den Haag, 2009. URL `http://www.gezondheidsraad.nl/nl/adviezen/preventie/bevolkingsonderzoek-naar-darmkanker`.

[42] Klaus Mönkemüller, Helmut Neumann, Peter Malfertheiner, and Lucia C. Fry. Advanced Colon Polypectomy. *Clinical Gastroenterology and Hepatology*, 7(6):641–652, 2009. ISSN 15423565. doi: 10.1016/j.cgh.2009.02.032. URL `http://dx.doi.org/10.1016/j.cgh.2009.02.032`.

[43] Toshio Uraoka, Adolfo Parra-Blanco, and Naohisa Yahagi. Colorectal endoscopic submucosal dissection: Is it suitable in western countries? *Journal of Gastroenterology and Hepatology (Australia)*, 28(3):406–414, 2013. ISSN 08159319. doi: 10.1111/jgh.12099.

[44] Jing Wang, Xiao-Hua Zhang, Jian Ge, Chong-Mei Yang, Ji-Yong Liu, and Shu-Lei Zhao. Endoscopic submucosal dissection vs endoscopic mucosal resection for colorectal tumors: A meta-analysis. *World journal of gastroenterology : WJG*, 20(25):8282–7, 2014. ISSN 2219-2840. doi: 10.3748/wjg.v20.i25.8282. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4081704&tool=pmcentrez&rendertype=abstract$\delimiter"026E30F$nDRUCK`.

[45] Nisha Patel, Carlo Seneci, Guang-zhong Yang, Ara Darzi, Julian Teare, and United Kingdom. Flexible platforms for natural orifice transluminal and endoluminal surgery. *Endoscopy International Open*, 02:117–123, 2014. ISSN 2364-3722. doi: 10.1055/s-0034-1377171.

[46] S. J. Phee, S. C. Low, V. a. Huynh, a. P. Kencana, Z. L. Sun, and K. Yang. Master and slave transluminal endoscopic robot (MASTER) for natural orifice transluminal endoscopic surgery (NOTES). *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Engineering the Future of Biomedicine, EMBC 2009*, pages 1192–1195, 2009. ISSN 1557-170X. doi: 10.1109/IEMBS.2009.5333413.

[47] J Heemskerk. *Robot-Assisted Laparoscopic Surgery*. PhD thesis, Universiteit Maastricht, 2014.

[48] Yair Lotan. Is robotic surgery cost-effective: no. *Current opinion in urology*, 22(1):66–9, January 2012. ISSN 1473-6586. doi: 10.1097/MOU.0b013e32834d4d76. URL `http://www.ncbi.nlm.nih.gov/pubmed/22080872`.

[49] Demcon. Advanced Endoscopy, 2015. URL `http://www.demcon.nl/showcase/advanced-endoscopy-project/`.

[50] Jeroen Ruiter and Esther Rozeboom. Design and evaluation of robotic steering of a flexible endoscope. *Biomedical Robotics ...*, 2012. URL `http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6290272`.

[51] Rob Reilink and AML Kappers. Evaluation of robotically controlled advanced endoscopic instruments. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 9(April):240–246, 2013. doi: 10.1002/rcs. URL `http://onlinelibrary.wiley.com/doi/10.1002/rcs.1496/full`.

[52] J. G. Ruiter, G. M. Bonnema, M. C. Voort, and I. a. M. J. Broeders. Robotic control of a traditional flexible endoscope for therapy. *Journal of Robotic Surgery*, 7(3):227–234, April 2013. ISSN 1863-2483. doi: 10.1007/s11701-013-0405-4. URL `http://link.springer.com/10.1007/s11701-013-0405-4`.

[53] Esther Rozeboom, Jeroen Ruiter, Michel Franken, and Ivo Broeders. Intuitive user interfaces increase efficiency in endoscope tip control. *Surgical endoscopy*, pages 1–6, March 2014. ISSN 1432-2218. doi: 10.1007/s00464-014-3510-1. URL `http://www.ncbi.nlm.nih.gov/pubmed/24671354`.

[54] N Van Der Stap, R Reilink, S Misra, IAMJ Broeders, and F van der Heijden. The use of the focus of expansion for automated steering of flexible endoscopes. In *The Fourth IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics*, pages 13–18, 2012. ISBN 9781457711985. URL `http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6290804`.

[55] N van der Stap, R Reilink, S Misra, IAMJ Broeders, and F van der Heijden. A feasibility study of optical flow-based navigation during colonoscopy. *International Journal of Computer Assisted Radiology and Surgery*, 7 (1):229–235, 2012. URL `http://eprints.eemcs.utwente.nl/22038/`.

[56] N Van Der Stap, C H Slump, I a M J Broeders, and F Van Der Heijden. Image-based Navigation for a Robotized Flexible Endoscope. *Computer-Assisted and Robotic Endoscopy*, 8999:77–87, 2014. ISSN 16113349. doi: 10.1007/978-3-319-13410-9_8.

[57] Nanda van der Stap, Ferdinand van der Heijden, and Ivo A M J Broeders. Towards automated visual flexible endoscope navigation. *Surgical endoscopy*, 27(10):3539–47, October 2013. ISSN 1432-2218. doi: 10.1007/s00464-013-3003-7. URL `http://www.ncbi.nlm.nih.gov/pubmed/23670745`.

[58] Mikael H Sodergren, Alexander Warren, Jean Nehme, James Clark, Sonja Gillen, Hubertus Feussner, Julian Teare, Ara Darzi, and Guang-Zhong Yang. Endoscopic horizon stabilization in natural orifice translumenal endoscopic surgery: a randomized controlled trial. *Surgical innovation*, 21(1):74–9, February 2014. ISSN 1553-3514. doi: 10.1177/1553350613489187. URL `http://www.ncbi.nlm.nih.gov/pubmed/23686394`.

[59] Charreau S Bell, Keith L Obstein, and Pietro Valdastri. Image partitioning and illumination in image-based pose detection for teleoperated flexible endoscopes. *Artificial intelligence in medicine*, 59(3):185–96, November 2013. ISSN 1873-2860. doi: 10.1016/j.artmed.2013.09.002. URL `http://www.ncbi.nlm.nih.gov/pubmed/24188575`.

[60] Kensaku Mori, Daisuke Deguchi, Jun-ichi Hasegawa, and Yasuhito Suenaga. A Method for Tracking the Camera Motion of Real Endoscope by Epipolar Geometry Analysis and Virtual Endoscopy System. *Miccai*, pages 1–8, 2001. URL `http://www.springerlink.com/index/vfrgc18axd5hh184.pdf`.

[61] Wai Keung Wong, Bo Yang, Chao Liu, and Philippe Poignet. A quasi-spherical triangle-based approach for efficient 3-D soft-tissue motion tracking. *IEEE/ASME Transactions on Mechatronics*, 18(October 2012): 1472–1484, 2013. ISSN 10834435. doi: 10.1109/TMECH.2012.2203919.

[62] Laurent Ott, Florent Nageotte, Philippe Zanne, Michel De Mathelin, and Senior Member. Robotic Assistance to Flexible Endoscopy by Physiological-Motion Tracking. *Robotics, IEEE Transactions on*, 27(2):346–359, April 2011. doi: 10.1109/TRO.2010.2098623.

[63] L. Ott, F. Nageotte, P. Zanne, and M. de Mathelin. Physiological motion rejection in flexible endoscopy using visual servoing and repetitive control : Improvements on non-periodic reference tracking and non-periodic disturbance rejection. *2009 IEEE International Conference on Robotics and Automation*, pages 4233–4238, May 2009. doi: 10.1109/ROBOT.2009.5152655. URL `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5152655`.

[64] Jianfei Liu, Kalpathi R Subramanian, and Terry S Yoo. An optical flow approach to tracking colonoscopy video. *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society*, 37(3):207–23, April 2013. ISSN 1879-0771. doi: 10.1016/j.compmedimag.2013.01.010. URL `http://www.ncbi.nlm.nih.gov/pubmed/23490235`.

[65] G A de Jong. A Real-Time Target Lock for Endoscopic Interventions, 2014.

[66] Stefan D Frijnts. *Target Lock Controller (work in progress)*. Master thesis, University of Twente, 2015.

[67] Anthony G Gallagher and Gerald C O'Sullivan. Human Factors in Acquiring Medical Skills. In *Fundamentals of Surgical Simulation*, pages 89–121. Springer-Verlag London, London, 1 edition, 2012. doi: 10.1007/978-0-85729-763-1.

[68] K Nakayama. James J. Gibson: An appreciation. *Psychological Review*, 101(2):329–335, 1994. URL `http://psycnet.apa.org/journals/rev/101/2/329/`.

[69] BK Horn and BG Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981. URL `http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1231385`.

[70] BD Lucas and T Kanade. An iterative image registration technique with an application to stereo vision. *IJCAI*, 130(1981):121–130, 1981. URL `http://flohauptic.googlecode.com/svn-history/r3/trunk/optic_flow/docs/articles/LK/Barker_unifying/lucas_bruce_d_1981_2.pdf`.

[71] JL Barron, DJ Fleet, and SS Beauchemin. Performance of optical flow techniques. *International journal of computer vision*, 12(1):43–77, 1994. URL `http://link.springer.com/article/10.1007/BF01420984`.

[72] David G Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 2004.

[73] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359, June 2008. ISSN 10773142. doi: 10.1016/j.cviu.2007.09.014. URL `http://linkinghub.elsevier.com/retrieve/pii/S1077314207001555`.

[74] Jean-Yves Bouguet. Camera Calibration Toolbox for Matlab, 2014. URL `http://www.vision.caltech.edu/bouguetj/calib_doc/`.

[75] Nanda van der Stap, Luuk Voskuilen, Guido de Jong, Hendrikus J M Pullens, Matthijs P Schwartz, Ivo Broeders, and Ferdi van der Heijden. A Real-Time Target Tracking Algorithm for a Robotic Flexible Endoscopy Platform. In *2nd International Workshop on Computer Assisted and Robotic Endoscopy*, 2015.

[76] E. M. Jellinek. On the use of the intraclass correlation coefficient in the testing of the difference of certain variance ratios. *Journal of Educational Psychology*, 31:60–63, 1940. ISSN 0022-0663. doi: 10.1037/h0062703.

[77] Kota Yamaguchi. mexopencv, 2011. URL `http://kyamagu.github.io/mexopencv/`.

[78] F. Landis Markley, Yang Cheng, John L. Crassidis, and Yaakov Oshman. Averaging quaternions. *Advances in the Astronautical Sciences*, 127:1675–1682, 2007. ISSN 00653438. doi: 10.2514/1.28949.

[79] K. Mori, D. Deguchi, J. Sugiyama, Y. Suenaga, J. Toriwaki, C. R. Maurer, H. Takabatake, and H. Natori. Tracking of a bronchoscope using epipolar geometry analysis and intensity-based image registration of real and virtual endoscopic images. *Medical Image Analysis*, 6:321–336, 2002. ISSN 13618415. doi: 10.1016/S1361-8415(02)00089-0.

[80] Stefan Leutenegger, Margarita Chli, and Roland Y. Siegwart. BRISK: Binary Robust invariant scalable keypoints. *2011 International Conference on Computer Vision*, pages 2548–2555, November 2011. doi: 10.1109/ICCV.2011.6126542. URL `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6126542`.

[81] a. Alahi, R. Ortiz, and P. Vandergheynst. FREAK: Fast Retina Keypoint. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 510–517, June 2012. doi: 10.1109/CVPR.2012.6247715. URL `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6247715`.

Appendix A

Coordinate Systems

In the following figures, the coordinate systems (also known as reference frames) are visualised. We refer to table 4.1 for the abbreviations. *Plz redo these despicable images*

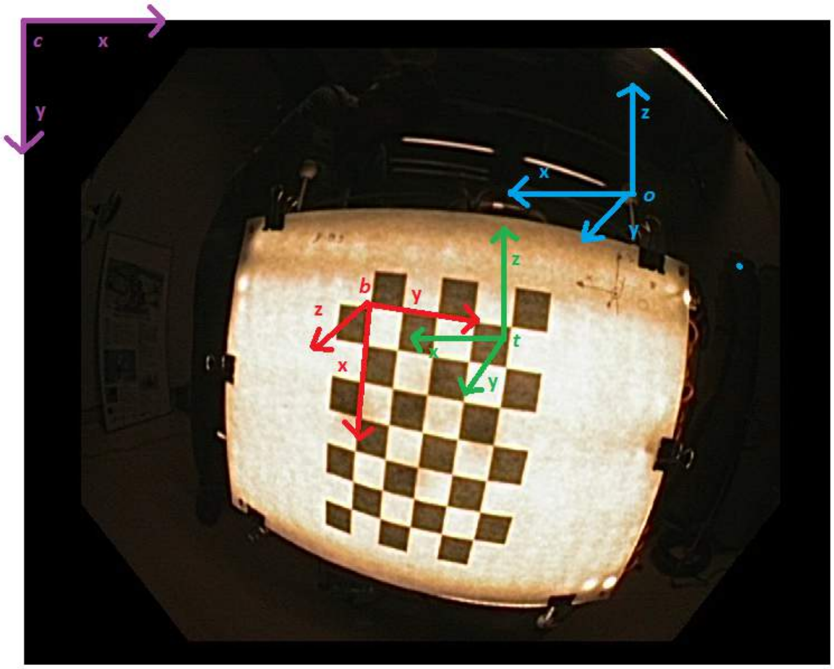


Figure A.1: An image of the chessboard taken by the endoscope. The following coordinate systems are (badly) drawn: chessboard *b* (red); camera *c* (purple); target *t* (green); and origin of the optical tracking system *o* (blue). Note that the origin of *o* coincides with an IR reflector. This is intentional, as it makes for easy measurement of the translation vector trom the trackable to the IR reflector of the origin of *o*. (This can no longer be used when the robot arm is moving during measurements for example.)

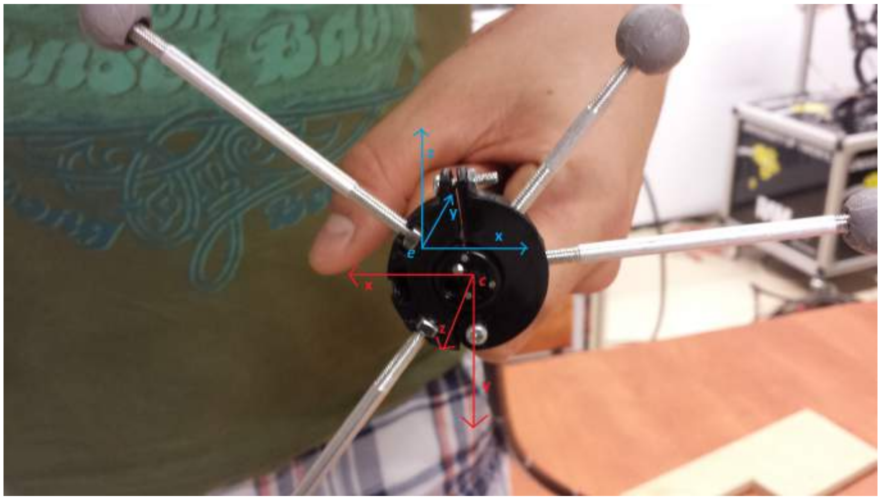


Figure A.2: A photograph of the tip of the endoscope with the frame containing the IR reflectors attached. The coordinate frame of the camera *c* (red) and the coordinate ssystem of the trackable on the endoscope tip *e* (blue) are visualised. The origin of *c* is assumed to be at the lense on the endoscope.

Appendix B

MATLAB Pseudocode

B.1 testTechnicalValidation1

Note that there are a lot of steps (mainly visualisation steps) in the pseudocode that are not necessary for the ultimate result ${}^{c}\mathbf{H}_{e}$. These are merely meant to validate the individual transformation matrices.

B.1.1 Definitions

Abbreviation	Coordinate system
c	Camera
e	Endoscope tip
t	Target board
b	Chessboard
o	Origin of OptiTrack
s	Shaft of endoscope

Homogeneous transformation matrix ${}^{b}\mathbf{H}_{a}$ transforms a column vector from coordinate system a to system b. Furthermore, a distinction is made between every image (images for which there is OptiTrack data available) and calibrated images (later the collection i), the images for which the camera calibration was successful, generally less than every image.

1. Run the Camera Calibration Toolbox for Matlab by Bouguet[74] on the images from the endoscope
2. Load the results of the camera calibration and calculate ${}^{c}\mathbf{H}_{b}$ and the weights $w_{\mathbf{R}}$ and $\mathbf{w_t}$ for calibrated images
3. Load data from the OptiTrack trackables into ${}^{o}\mathbf{H}_{e}$ and ${}^{o}\mathbf{H}_{t}$ for every image
4. Determine matrix ${}^{b}\mathbf{H}_{t}$ in the following way:
    (a) Determine the rotation of ${}^{o}\mathbf{R}_{b}$ assuming right angles between b and o and not rotation between o and t
    (b) Measure ${}^{o}\mathbf{t}_{b}$
    (c) Write down ${}^{o}\mathbf{t}_{t}$ from the OptiTrack system at initialisation
    (d) Create ${}^{o}\mathbf{H}_{b}$ from ${}^{o}\mathbf{R}_{b}$ and ${}^{o}\mathbf{t}_{b}$

```
    (e) Calculate ${}^{t}\mathbf{H}_b = {}^{o}\mathbf{H}_b - {}^{o}\mathbf{t}_t$
    (f) Calculate ${}^{b}\mathbf{H}_t = {}^{t}\mathbf{H}_b^{-1}$
5. Plot ${}^{o}\mathbf{H}_b$ in coordinate system $o$
6. Calculate ${}^{c}\mathbf{H}_t = {}^{c}\mathbf{H}_b \cdot {}^{b}\mathbf{H}_t$ for calibrated images
7. Plot ${}^{c}\mathbf{H}_b$ and ${}^{c}\mathbf{H}_t$ in coordinate system $c$
8. Calculate ${}^{e}\mathbf{H}_t = {}^{o}\mathbf{H}_e^{-1} \cdot {}^{o}\mathbf{H}_t$ for every image
9. Plot ${}^{e}\mathbf{H}_t$ in coordinate system $c$
10. Calculate ${}^{c}\mathbf{H}_e = {}^{c}\mathbf{H}_b \cdot {}^{b}\mathbf{H}_t \cdot {}^{o}\mathbf{H}_t^{-1} \cdot {}^{o}\mathbf{H}_e$ for calibrated images
11. Calculate the weighted average translation ${}^{c}\bar{\mathbf{t}}_e$ using $\bar{\mathbf{t}} = \frac{\sum_{i=1}^{n} \mathbf{w}_{\mathbf{t},i}\mathbf{t}_i}{\sum_{i=1}^{n} \mathbf{w}_{\mathbf{t},i}}$
12. Calculate the weighted average standard deviation of ${}^{c}\mathbf{t}_e$ using $\sigma = \sqrt{\frac{\sum_{i=1}^{n} \mathbf{w}_{\mathbf{t},i}(\mathbf{t}_i - \bar{\mathbf{t}})^2}{\frac{M-1}{M}\sum_{i=1}^{n} \mathbf{w}_{\mathbf{t},i}}}$ in which scalar $M$ is the number of non-zero weights.
13. Transform ${}^{c}\mathbf{R}_e$ onto quaternions $\mathbf{q}_i$ for calibrated images
14. Calculate matrix $\mathbf{M}$ by $\mathbf{M} = \sum_{i=1}^{n} w_{\mathbf{R},i}\mathbf{q}_i\mathbf{q}_i^{\mathrm{T}}$ according to Markley et al.[78]
15. Calculate the average quaternion $\bar{\mathbf{q}}$ being the eigenvector corresponding to the largest eigenvalue
16. Calculate the average ${}^{c}\mathbf{H}_e$ from $\bar{\mathbf{q}}$ and $\bar{\mathbf{t}}$
17. Convert ${}^{c}\mathbf{R}_e$ into Euler angles using the $zyx$-convention for every image
18. Plot the Euler angles and translation for calibrated images
```

B.2 testValidation2

The following piece of pseudocode from script testValidation2 is the recursive reconstruction of $\mathbf{x}_b$, the subsequent visualisation of the target location and finally the evaluation of the zooming functionality. For missing definitions return to section 4.1.3.

```
1. Load ${}^{c}\mathbf{H}_e$ and ${}^{b}\mathbf{H}_t$ from the calibration data
2. Read ${}^{o}\mathbf{H}_t(i)$, ${}^{o}\mathbf{H}_e(i)$, target location and target radius from the OptiTrack data file.
3. For $i = 1 : n$ do:
    (a) Calculate ${}^{c}\mathbf{H}_b(i) = {}^{c}\mathbf{H}_e \cdot {}^{o}\mathbf{H}_e^{-1}(i) \cdot {}^{o}\mathbf{H}_t(i) \cdot {}^{b}\mathbf{H}_t^{-1}$
    (b) Calculate $\hat{\mathbf{x}}_n = \mathbf{K}^{-1} \cdot \hat{\mathbf{x}}_p(i)$
    (c) Initialise $d(1) = {}^{c}\mathbf{H}_b(3, 4, i)$ and $j = 1$
    (d) While $|z_c| > 10^{-6}$ do:
        i. Calculate $\mathbf{x}_c = \hat{\mathbf{x}}_n \cdot d(j)$
        ii. Calculate $\hat{\mathbf{x}}_b = {}^{c}\mathbf{H}_b^{-1} \cdot \hat{\mathbf{x}}_c$
```

        iii. Calculate $d(j+1) = d(j) + z_b$

    (e) Return $\mathbf{x}_b(i)$

    (f) Calculate $\hat{\mathbf{x}}_o(i) = {}^{o}\mathbf{H}_t \cdot {}^{b}\mathbf{H}_t^{-1}(i) \cdot \hat{\mathbf{x}}_b(i)$

4. Plot $\mathbf{x}_b(i)$ of which only the x- and y-values are relevant
5. Calculate the displacement from the initialised value $\mathbf{x}_d(i) = \mathbf{x}_b(i) - \mathbf{x}_b(1)$
6. Calculate the Euclidean norm of the displacement $d_{\mathbf{x}}(i) = \sqrt{x_d(i)^2 + y_d(i)^2}$
7. Plot $\mathbf{x}_d(i)$ and $d_{\mathbf{x}}(i)$
8. Calculate the depth ratio $r_d(i) = \frac{y_o(i)}{y_o(1)}$
9. Recalculate the scale factor $s(i) = \frac{\text{target radius}(1)}{\text{target radius}(i)}$
10. Plot $r_d(i)$ and $s(i)$ to evaluate the zooming functionality