Hybrid RGB/Time-of-Flight Sensors in Minimally Invasive Surgery

Hybride RGB/Time-of-Flight Sensoren für die Minimal-invasive Chirurgie

Der Technischen Fakultät
der Friedrich-Alexander-Universität
Erlangen-Nürnberg

zur

Erlangung des Doktorgrades Dr.-Ing.

vorgelegt von

Sven Haase

aus

Fürth, Deutschland

Als Dissertation genehmigt
von der Technischen Fakultät
der Friedrich-Alexander-Universität Erlangen-Nürnberg

Tag der mündlichen Prüfung:	09.06.2016
Vorsitzende des Promotionsorgans:	Prof. Dr. P. Greil
Gutachter:	Prof. Dr.-Ing. J. Hornegger
	Prof. Dr.-Ing. R. Koch

Abstract

Nowadays, minimally invasive surgery is an essential part of medical interventions. In a typical clinical workflow, procedures are planned preoperatively with 3-dimensional (3-D) computed tomography (CT) data and guided intraoperatively by 2-dimensional (2-D) video data. However, accurate preoperative data acquired for diagnose and operation planning is often not feasible to deliver valid information for orientation and decisions within the intervention due to issues like organ movements and deformations. Therefore, innovative interventional tools are required to aid the surgeon and improve safety and speed for minimally invasive procedures. Augmenting 2-D color information with 3-D range data allows to use an additional dimension for developing novel surgical assistance systems. Here, Time-of-Flight (ToF) is a promising low-cost and real-time capable technique that exploits reflected near-infrared light to estimate the radial distances of points in a dense manner. This thesis covers the entire implementation pipeline of this new technology into a clinical setup, starting from calibration to data preprocessing up to medical applications.

The first part of this work covers a novel automatic calibration scheme for hybrid data acquisition based on barcodes as recognizable feature points. The common checkerboard pattern is overlaid by a marker that includes unique 2-D barcodes. The prior knowledge about the barcode locations allows to detect only valid feature points for the calibration process. Based on detected feature points seen from different points of view a sensor data fusion for the complementary modalities is estimated. The proposed framework achieved subpixel reprojection errors and barcode identification rates above 90% for both the ToF and the RGB sensor.

As range data of low-cost ToF sensors is typically error-prone due to different issues, e.g. specular reflections and low signal-to-noise ratio (SNR), preprocessing is a mandatory step after acquiring photometric and geometric information in a common setup. This work proposes the novel concept of hybrid preprocessing to exploit the benefits of one sensor to compensate for weaknesses of the other sensor. Here, we extended established preprocessing concepts to handle hybrid image data. In particular, we propose a nonlocal means filter that takes an entire sequence of hybrid image data into account to improve the mean absolute error of range data by 20%. A different concept estimates a high-resolution range image by means of super-resolution techniques that takes advantage of geometric displacements by the optical system. This technique improved the mean absolute error only by 12% but improved the spatial resolution simultaneously. In oder to tackle the issue of specular highlights that cause invalid range data, we propose a multi-view scheme for highlight correction. We replace invalid range data at a specific viewpoint with valid data of another viewpoint. This reduced the mean absolute error by 33% compared to a basic interpolation.

Finally, this thesis introduces three novel medical applications that benefit from hybrid 3-D data. First, a collision avoidance module is introduced that exploits range data to ensure a safety margin for endoscopes within a narrow operation side. Second, an endoscopic tool localization framework is described that exploits hybrid range data to improve tool localization and segmentation. Third, a data

fusion framework is proposed to extend the narrow field of view and reconstruct the entire situs.

This work shows that hybrid image data of ToF and RGB sensors allows to improve image based assistance systems with more reliable and intuitive data for better guidance within a minimally invasive intervention.

Kurzübersicht

Minimal-invasive Chirurgie ist heutzutage ein essentielles Teilgebiet medizinischer Eingriffe. In einem typischen klinischen Arbeitsablauf werden Operationen präoperativ anhand von 3-dimensionalen (3D) Daten der Computertomographie (CT) geplant und intraoperativ mittels 2-dimensionaler (2D) Videodaten unterstützt. Präzise präoperativ aufgenommene Daten für die Diagnose und Operationsplanung sind aufgrund von Problemen wie der Organbewegung und Organverformung meist nicht in der Lage valide Informationen für die Orientierung und Entscheidungen während der Operation zu vermitteln. Daher werden innovative interventionelle Werkzeuge benötigt um den Chirurgen zu unterstützen und die Sicherheit und Dauer von minimal-invasiven Eingriffen zu verbessern. 2D Farbdaten mit 3D Entfernungsdaten zu überlagern erlaubt es eine zusätzliche Dimension zu nutzen um neuartige chirurgische Assistenzsysteme zu entwickeln. Dabei stellt Time-of-Flight (ToF) ein kostengünstiges und echtzeitfähiges Verfahren dar, das reflektiertes nahinfrarotes Licht ausnutzt um radiale Distanzen in einem dichten Gitter zu erfassen. Diese Arbeit behandelt die komplette Implementierung dieser neuen Technologie in ein medizinisches Umfeld, angefangen von der Kalibrierung über die Datenvorverarbeitung bis hin zu medizinischen Anwendungen.

Der erste Teil der Arbeit behandelt ein neuartiges automatisches Kalibrierungsverfahren für hybride Datenaufnahmen. Es basiert auf Barcodes als identifizierbare Merkmalspunkte. Das gewöhnliche Schachbrett Muster ist überlagert mit einem Marker, der unverwechselbare 2D Barcodes beinhaltet. Das Vorwissen über die Positionierung der Barcodes erlaubt es nur valide Merkmalspunkte für die Kalibrierung zu erfassen. Basierend auf den aus verschiedenen Ansichten erkannten Merkmalspunkten wird eine Datenfusion der sich ergänzenden Modalitäten ermittelt. Das vorgestellte System erreicht Rückprojektionsfehler im Subpixelbereich und Barcode Identifikationsraten von über 90% für ToF und RGB Sensoren.

Da Entfernungsdaten von kostengünstigen ToF Sensoren aufgrund verschiedener Probleme, z.B. spiegelnde Reflektionen und ein niedriges Signal-zu-Rausch Verhältnis (SNR), üblicherweise fehleranfällig sind, ist Datenvorverarbeitung ein essentieller erster Schritt nach der Aufnahme photometrischer und geometrischer Informationen in einem gemeinsamen Aufbau. Diese Arbeit stellt das Konzept der hybriden Datenvorverarbeitung vor, um mit den Vorteilen des einen Sensors die Nachteile des anderen zu kompensieren. Hierbei erweiterten wir etablierte Vorverarbeitungverfahren um hybride Bilddaten zu verarbeiten. Im speziellen stellen wir einen nichtlokalen Mittelwertfilter vor, der eine ganze Sequenz von hybriden Bilddaten verwendet, um das den mittleren absoluten Fehler der Entfernungsdaten um 20% zu verbessern. Ein anderes Verfahren schätzt ein hochauflösendes Bild von Entfernungsdaten mittels Super-Resolution und nutzt dabei geometrische Verschiebungen des optischen Systems aus. Diese Technik verbessert den mittleren absoluten Fehler zwar nur um 12%, verbessert aber gleichzeitig auch die räumliche Auflösung. Um die Problematik der spiegelnden Glanzlichter, an deren Stelle ungültige Entfernungsdaten ermittelt werden, anzugehen stellen wir ein Konzept mit Bildern verschiedener Ansichten vor um die Glanzlichtdaten zu korrigieren. Wir ersetzen ungültige Daten von einem bestimmten Beobach-

tungspunkt durch gültige Daten einer anderen Ansicht. Dies verminderte den mittleren absoluten Fehler um 33% verglichen mit einer konventionellen Interpolation.

Abschließend führt diese Thesis drei neuartige medizinische Anwendungen ein, die von hybriden 3D Daten profitieren. Zunächst wird ein Modul zur Kollisionsvermeidung vorgestellt, das die Entfernungsdaten nutzt, um einen Sicherheitsabstand des Endoskops in dem engen Arbeitsfeld sicher zu stellen. Als zweites wird ein System zur Lokalisierung von endoskopischen Werkzeugen beschrieben, das die hybriden Entfernungsdaten ausnutzt, um die Ortsbestimmung und die Segmentierung der Werkzeuge zu verbessern. Als drittes wird ein System der Datenfusion vorgestellt, das das eingeschränkte Sichtfeld erweitert und dabei den gesamten Situs rekonstruiert.

Diese Arbeit zeigt, dass hybride Bilddaten von ToF und RGB Sensoren es ermöglichen bildbasierte Assistenzsysteme zu verbessern, indem zuverlässigere und intuitivere Daten zur Unterstützung während minimal-invasiven Eingriffen genutzt werden können.

Acknowledgment

First of all I want to thank my family for supporting me from the beginning of my interest in computer science in school up to finishing this thesis. They always backed me up financially and mentally the entire time. Even in times of high stress and little success they kept me motivated and had faith in my work.

Special thanks go to Prof. Dr. Joachim Hornegger, who offered me the opportunity to work at his lab and share innovative concepts and ideas with colleagues in Erlangen and at conferences all over the world. Achim always added nice ideas to my concepts and algorithms to make them convincing and interesting for the entire community of computer assisted surgery. Besides the nice atmosphere in the entire lab I want to mention the great help of Sebastian Bauer, Jakob Wasza, Thomas Köhler and Andreas Maier. Without their support this whole research project would not have been possible. Discussing papers and ideas, improving my papers and spending time with them at conference was always a great pleasure for me.

A very important part of my time was spent at the "Klinikum rechts der Isar". Here, I thank Prof. Dr. Hubertus Feußner and his team for all the inspiring discussions and their tremendous help in developing applications closely related to realistic medical requirements. The teams was always available when I needed advices from the medical point of view and invited me to spend time in Munich. I always felt like I was part of their team.

I also acknowledge the great support from my project partners in Heidelberg at the DKFZ. Especially PD Dr. Lena Maier-Hein and Thomas Kilgus helped me a lot to bring this work to the great piece that it has become. They always had time to discuss important topics and support me with helpful advices.

Furthermore, I acknowledge the nice collaboration with WOLF GmbH and Metrilus GmbH that supported us with the necessary hardware to evaluate our algorithms with up-to-date prototype hardware. Furthermore, both companies shared their knowledge with us and thereby together we were able to improve the endoscope prototype by collaborating our common ideas.

Finally, special thanks also goes to the DFG for funding the project and the conference stays within the last 4 years. For the interdisciplinary discussions and financial support I also thank the Graduate School of Information Science in Health. Within their workshops I always met friendly researchers with interesting projects besides computer science. I am also grateful for the financial support of the FAZIT foundation.

Contents

Part I

Introduction

CHAPTER 1

Structure of the Thesis

This thesis focuses on fundamental research and first applications of novel miniature hybrid range imaging devices in minimally invasive abdominal surgery. The applications are based on data of a 3-dimensional (3-D) Time-of-Flight (ToF)/RGB endoscope and on a concept of a miniature ToF based satellite camera. The work covers the whole pipeline of investigating a new imaging device: First, the setup including a range imaging sensor and a color sensor is calibrated. Second, the measured data is preprocessed considering different sources that disturb the ToF signal. Third, medical applications, e.g. endoscopic tool localization and collision avoidance, are investigated and evaluated. An important point of this work is to take advantage of the combination of sensors in a hybrid imaging system to build more robust and intuitive applications for image guidance in minimally invasive surgery.

This work will not investigate different optimization techniques of the hardware, e.g. in terms of the optical systems or the sensors. Though several error sources of ToF devices are known, in this thesis all problems are addressed in a matter that the ToF sensor could generically be replaced by a different range image acquiring sensor.

1.1 Contributions

The scientific focus of this work is to investigate novel miniature hybrid range imaging devices in minimally invasive procedures and their potential as guidance systems for endoscopic interventions. The applications are to be evaluated with medical experts and on realistic data gained either from porcine studies or for quantitative evaluation gained from range image simulation. As published in different scientific papers, the main contributions of this work are as follows:

- Development of a fast and robust calibration framework for hybrid range imaging setups. The focus lies on intrinsic and extrinsic calibration of color and range sensors. Sensor data fusion of color and range data is investigated for different setups. Due to the low image resolution of data measured by

miniature range sensors the investigated calibration techniques are based on enhanced checkerboards and automatic corner detection [Haas 12, Haas 13b].

- Investigation of different preprocessing techniques required for robust applications and intuitive visualization. These algorithms consider the low signal-to-noise ratio (SNR) [Lind 14], invalid range measurements induced by specular reflections [Haas 14] and the low spatial resolutions [Kohl 13, Koeh 14a, Koeh 14b] of miniature range sensors.

- Development of three concepts for computer guidance systems driven by range images. First, a novel enhancement module for endoscope holders to avoid collision within the human body is engineered [Haas 13c]. Second, endoscopic tool localization is investigated for 3-D endoscopy [Haas 13e]. Third, initial situs reconstruction using ToF satellite cameras for improved orientation is developed [Haas 13a].

Besides these major aspects, this work also contributed to other publications, e.g. to investigate different range imaging techniques in minimally invasive surgery [Groc 12, Maie 14], to develop a range imaging software framework [Wasz 11a] and to investigate real-time feasibility of super-resolution [Wetz 13].

1.2 Outline

This thesis is divided into three major parts: Background information on range imaging in minimally invasive surgery, hybrid preprocessing for data quality improvement and medical applications for image guidance within endoscopic interventions. In detail, the work is structured by:

Range Imaging in Abdominal Surgery Part I outlines the medical and technical background of this work. Initially, the workflow and the medical scenario of minimally invasive abdominal surgery is introduced. Then, existing image guidance systems are described briefly and the benefits of additional range imaging sensors are elaborated. This chapter also describes different range image acquisition techniques including structured light, stereo vision and ToF technology.

Hybrid Range Data Preprocessing Part II begins with describing the sensor data fusion of range images and color images in detail. Here, the focus is set on the calibration framework using a self-encoded marker. For the fusion itself, two different setups are analyzed: A stereo setup with two different viewing points and a setup including a beam splitter for a single viewing point. After calibration, a variety of techniques to overcome technological limitations of range imaging devices in general with an evaluation on ToF data is described. All concepts exploit the fact that hybrid data, i.e. complementary data of two different sensors, is acquired and aligned to each other. First, a denoising approach based on nonlocal means is evaluated. Second, this part investigates a multi-sensor super-resolution approach for denoising, deblurring and upsampling in a joint manner. Third, a

pipeline for restoration of defect range image regions caused by specular highlights is described.

Applications in Abdominal Surgery Part III covers three concepts for range image driven applications to aid surgeons within intervention. A first application performs distance supervision and correction between the endoscope and the observed tissue. The second application shows the feasibility to automatically localize endoscopic tools with 3-D ToF/RGB endoscopy. In a third application the initial reconstruction of the operation situs for better orientation and navigation is investigated.

Summary and Outlook Part IV summarizes the entire work and its general conclusion is drawn. The key concepts and major issues are pointed out. Furthermore, a short outlook for future work and potential of next-generation range imaging devices is outlined.

The remaining part of this chapter introduces the notation necessary to describe the mathematically formulations of the algorithms developed within this thesis.

1.3 Notation

As this thesis deals with cameras, we have to introduce points in different coordinate systems and dimensions first. A 3-D point is denoted as:

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \qquad \text{or} \qquad \widetilde{\boldsymbol{x}} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{pmatrix}, \tag{1.1}$$

if described in homogeneous coordinates. The coordinate system is denoted by the subscript, where $\boldsymbol{x}_\mathrm{W}$ describes a point in the common world coordinate system and $\boldsymbol{x}_\mathrm{C}$ describes a point in a local camera coordinate system. To allow transformations from world coordinates to camera coordinates and vice versa, rotation matrices and translation vectors need to be introduced. A rotation matrix is described as $\boldsymbol{R} \in \mathbb{SO}_3$ and a translation vector is denoted as $\boldsymbol{t} \in \mathbb{R}^3$:

$$\boldsymbol{R} = \begin{pmatrix} r_{1,1} & r_{1,2} & r_{1,3} \\ r_{2,1} & r_{2,2} & r_{2,3} \\ r_{3,1} & r_{3,2} & r_{3,3} \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{t} = \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix}. \tag{1.2}$$

A transformation of a 3-D point $\boldsymbol{x}$ from the world coordinate system to a camera coordinate system is thereby described by $\boldsymbol{x}_\mathrm{C} = \boldsymbol{R}(\boldsymbol{x}_\mathrm{W} - \boldsymbol{t})$.

As this thesis deals with common CMOS/CCD cameras, we reduce there model to a usual pinhole camera. Therefore, we describe a projection from 3-D coordi-

nates in the camera coordinate system onto the 2-dimensional (2-D) image plane to compute sensor pixel coordinates by a perspective transformation [Hart 04]:

$$s\widetilde{\boldsymbol{u}} = s\begin{pmatrix} u_1 \\ u_2 \\ 1 \end{pmatrix} = \begin{pmatrix} f_{x_1}x_1 + c_{x_1}x_3 \\ f_{x_2}x_2 + c_{x_2}x_3 \\ x_3 \end{pmatrix} = \begin{pmatrix} f_{x_1} & 0 & c_{x_1} & 0 \\ 0 & f_{x_2} & c_{x_2} & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{pmatrix} = \boldsymbol{K}\widetilde{\boldsymbol{x}}_{\mathrm{C}}, \qquad (1.3)$$

with f and c describing the focal length and the central point of the pinhole camera, respectively. Here, both parameters are given in pixel units and therefore might vary in x_1 and x_2 dimension depending on the aspect ratio of an individual pixel on the sensor. To compute 2-D pixel coordinates u_1 and u_2 of the projection, the result needs to be scaled by s. By choosing $s = \frac{1}{x_3}$ it is already shown that uncertainty in the x_3 coordinate has a direct influence on the pixel coordinates.

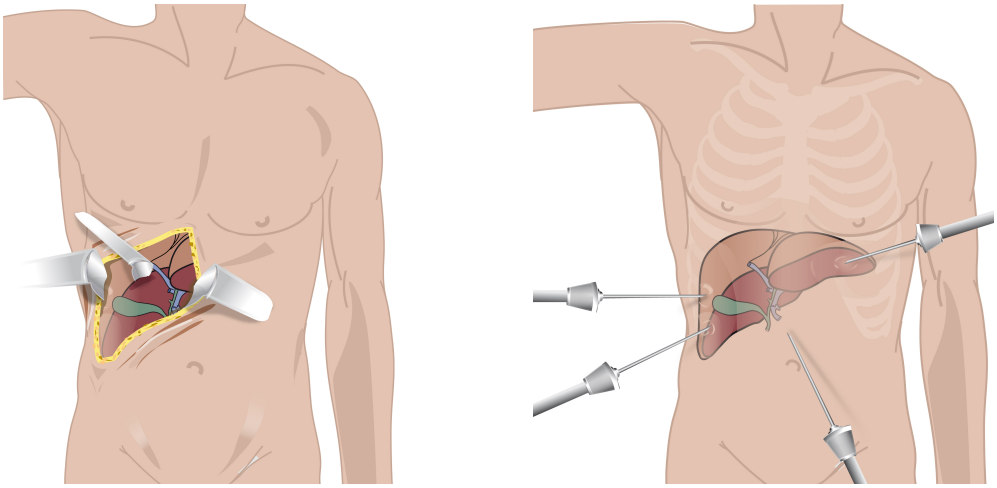
The computed pixel coordinate $\boldsymbol{u} \in \mathbb{R}^2$ describes a point on the image plane that delivers a measured intensity i. The subscript denotes the image source and the superscript describes the pixel position, e.g. $i^{\boldsymbol{u}}_{\mathrm{Gray}}$ denotes the intensity of a pixel in a grayscale image. A 2-D image with M_1 pixels in u_1 direction and M_2 pixels in u_2 direction is written as a vector $\boldsymbol{i} \in \mathbb{R}^M$ with $M = M_1 \cdot M_2$ by concatenating all pixels. As this thesis deals with sensors that acquire a sequence of frames, successive frames are described by their time step t. The set of all frames up to time T is denoted as $Q = \{\boldsymbol{i}_1, ..., \boldsymbol{i}_{\mathrm{T}}\}$.

Further variables and formulations that are only necessary in a particular part of this thesis are introduced where they are needed.

CHAPTER 2

Background on Abdominal Surgery

The term *Surgery* is derived from the Greek word *Χειρουργικη* and describes a medical procedure to improve a patient's quality of life. These procedures are either required to investigate or treat a pathology or in modern surgery also to change the patient's appearance. Considering only inpatient surgery, 51.5 million surgical procedures were performed in 2010 in the U.S. [Unit 10]. More than 6 million of those were operations on the digestive system constituting this to be one of the most important fields of surgical interventions. This highlights the need for further improvement on abdominal surgery involving the stomach, kidneys, liver etc. Abdominal surgery can be sub-divided by different criteria, e.g. by the invasiveness, by the affected body regions or by the particular pathology. In this thesis, abdominal surgery is split into two categories considering their invasiveness: Conventional surgery, also known as open surgery, and minimally invasive surgery, also called endoscopy. According to [Unit 10], in 2010 1.6 million endoscopic interventions were performed in the U.S. treating the digestive system, i.e. other endoscopic interventions as bronchoscopy or cystoscopy are not included. Therefore, to assist surgeons in difficult medical scenarios, a variety of guidance systems evolved over years, in particular for minimally invasive interventions. Those system usually acquire images with different modalities that visualize certain aspects of the human body to help making a diagnosis or to ease the actual treatment of a pathology [Grim 99]. This chapter points out the key aspects of minimally invasive surgery with particular focus on abdominal surgery. The comparison between minimally invasive and conventional open surgery is illustrated and several procedures are detailed. Moreover, this chapter introduces available assistance systems for endoscopy. Finally, this leads to the motivation why range imaging is of special interest for the next step of modern surgery.



(a) (b)

Figure 2.1: Illustrations of a cholecystectomy. 2.1a shows an open surgery with direct access and 2.1b shows a minimally invasive approach with endoscopic tools.

2.1 Minimally Invasive Surgery and Open Surgery

One criterion to categorize medical interventions is the degree of invasiveness. Therefore, this work distinguishes between minimally invasive and open surgery which is illustrated in Fig. 2.1. As the notation suggests, minimally invasive procedures describe medical procedure with little operative trauma. In comparison to conventional open surgery this leads to a shorter recovery time for the patient and thereby to a reduced hospital stay. In recent years, a variety of minimally invasive alternatives to conventional open surgery have evolved with special focus on pathologies of the heart [Gold 04] and the abdomen [Krem 01]. In minimally invasive surgery the physician has no direct access to the organs or structures of the human body. On the one hand this means fewer and smaller scars and less pain for the patient. On the other hand, without direct access the physician has a limited sense of orientation and usually has to rely on additional imaging techniques. For some medical procedures minimally invasive alternatives are not available as the incision is just too small, e.g. the removal of larger organs or transplantations. For smaller organs such as the kidney or gallbladder, laparoscopic interventions are already performed as a common routine. In terms of operative time, minimally invasive surgery usually takes longer due to the smaller incision and worse orientation. Both open and in most cases minimally invasive surgery require anesthesia within the intervention. Statistical comparison of open and minimally invasive surgery in terms of quality of life was published by Velanovich [Vela 00] and shows an overall improved result for laparoscopic interventions evaluated on four different procedures. Surgeons have applied endoscopic interventions as treatments for a multitude of different areas of the animal and human body, e.g. bronchoscopy covers examinations of the trachea and the lung, laparoscopy covers examinations of the abdominal organs and otoscopy covers examinations of the ears.

2.2 Minimally Invasive Abdominal Surgery

As one of the most important fields of minimally invasive procedures, the diagnosis and treatment of abdominal pathologies is the main medical application of this thesis. Here, a variety of important special instruments are required, see Fig. 2.2:

- *Endoscopes*: Depending on the procedure these devices are non-rigid (flexible endoscopes) or rigid (laparoscopes) and serve as a camera inside the human body. Rigid endoscopes have the benefit that the navigation is way more intuitive although the degrees of freedom during the navigation is reduced compared to non-rigid endoscopes.
- *Trocars*: To allow a fast exchange of different instruments a trocar is placed in the human body as a port to the abdominal cavity. For different procedures different sizes ranging from several mm up to a few cm are available.
- *Surgical instruments*: For the actual procedure different endoscopic tools are required, e.g. clamps or scissors. As illustrated in Fig. 2.2, those instruments have a scissor-like grip to control the action at the top of the tool.

The workflow of an endoscopic procedure in general is described by three major steps, i.e. preparation, procedure and recovery [Krem 01]. Starting up to several days before the actual intervention, the patient has to stop taking medication that is not prescribed due to the procedure. Furthermore, he is not allowed to drink or eat 6-12 hours in advance. After removing all clothes and all jewelry the patient is getting dressed in operation clothes and brought into the operation room. The actual procedure starts with attaching several sensors to the patient to continuously monitor his health status. Usually, then, the patient is anesthetized until the whole procedure is finished. After removing the hair and cleaning the surgical site, the actual procedure starts with a small incision, where the first trocar is inserted. Through this trocar, the abdomen is insufflated with carbon dioxide gas. This allows the physician to have more room for the procedure. Then, additional cuts for additional trocars might be performed dependent on the complexity of the procedure. These trocars are then used as a port to the abdominal cavity for a laparoscope or endoscopic instruments within the actual treatment. In a final step of the procedure, the carbon dioxide gas has to be removed and all the incisions are closed with stitches. The recovery phase starts after the patient woke up. He is then instructed how to keep the wounds clean and follow-up appointments are set up, e.g. to have the stitches removed. On average the recovery takes up to 12 weeks.

2.3 Assistance Systems for Endoscopy

In modern surgery, companies develop a variety of assistance systems to ease the navigation or to reduce the required manpower for minimally invasive procedures. Usually, besides the physician several assistants are required within the intervention. As the surgeon performs the actual procedure with endoscopic tools,

Figure 2.2: Different instruments for minimally invasive abdominal surgery. From left to right: A rigid endoscope, a non-rigid endoscope, a trocar, a set of sterile surgical instruments.

one assistant has to hold the endoscope, one has to hand him the required instruments over, one has to supervise the patient and often a few more are involved for general organization. The remaining part of this section introduces the most relevant of the available assistance systems in detail.

A very basic and intuitive assistance system is an endoscope holder, see Fig. 2.3. These medical devices are available in different complexities, ranging from simple non electronic static holders to automatic flexible holder that are navigated by a joystick. In general, these endoscope holders allow a more stable image acquisition as they exclude any jitter induced by a human. However, those systems are often very basic and still have to be navigated by a surgeon. Fig. 2.3 shows the SOLOASSIST©[1] [Hart 09], an electronic assistance arm that is navigated by an joystick and simulates a human arm.

Besides endoscope holders fully automatic robotic assistance systems are already commercially available [Sung 01]. These systems allow the surgeon to be at a separate workstation as illustrated in Fig. 2.3. All commands are directly transmitted to the robot allowing the surgeon to be at a distant place while performing the procedure. To describe one system in more detail, the da Vinci©[2] system in particular is navigated by grips and pedals to enable various degrees of freedom. For intuitive visualization this robot acquires stereoscopic images and thereby gives the surgeon a 3-D impression of the scene.

2.4 Motivation for Range Images in Endoscopy

Although assistance systems for minimally invasive surgery, e.g. endoscope holders, are commercially available, one major issue of endoscopic interventions is the orientation in the human body. Therefore, the navigation is dependent on the experience of the surgeon and usually based on endoscopic 2-D video images. Due to the lack of intuitive visual comparison to the environment, the narrow field of view induces a loss of depth and size estimation in the abdominal cavity. However, this information is required for diagnosis, e.g. the size of a polyp, and for decision, e.g. choosing the most reasonable endoscopic instrument. Schoen et

[1] AKTORmed GmbH, Barbing, Germany
[2] Intuitive Surgical Incorporated, Sunnyvale, USA

Figure 2.3: Photos of three surgical assistance systems. First, a static endoscope holder from Novid Surgical. Second, the SOLOASSIST, a joystick navigated endoscope holder. The third photo shows the da Vinci assistance system with the workstation on the left and the distant robotic system on the right.

al. [Scho 97] have shown that about 20% of the polyps in their experiments were estimated inaccurately. Therefore, a variety of different approaches to compensate for this loss of information were investigated [Vaki 94], e.g. adding grids to the endoscope lens or estimate sizes by comparison with known instruments. The rising importance of minimally invasive procedures in abdominal surgery and the novel assistance systems allow the integration of additional sensors for faster and safer interventions. Both previously described assistance systems lack the support for real 3-D metric data acquisition to either react on changes in the environment or to allow measurements within the human cavity. An additional 3-D sensor offers new applications reaching from collision avoidance to robust tool localization and 3-D situs reconstruction. The long-term goal of those assistant systems is to ensure safety while reducing manpower, costs and complexity of abdominal interventions.

This thesis focuses on range image acquiring devices as additional 3-D sensors in minimally invasive abdominal surgery due to their non-harmful acquisition technique. Range image sensors measure topographic surface data of the observed scene. Usually, this data is delivered as a 2-D image with metric radial or orthographic information as intensity values for each pixel. With metric range data and known camera parameters it is possible to reconstruct the surface $\mathcal{S}$ of the operation site in 3-D by inverting Eq. (1.3):

$$\widetilde{\boldsymbol{x}}_C = \begin{pmatrix} (u_1 - c_{x_1}) \frac{x_3}{f_{x_1}} \\ (u_2 - c_{x_2}) \frac{x_3}{f_{x_2}} \\ x_3 \\ 1 \end{pmatrix}, \tag{2.1}$$

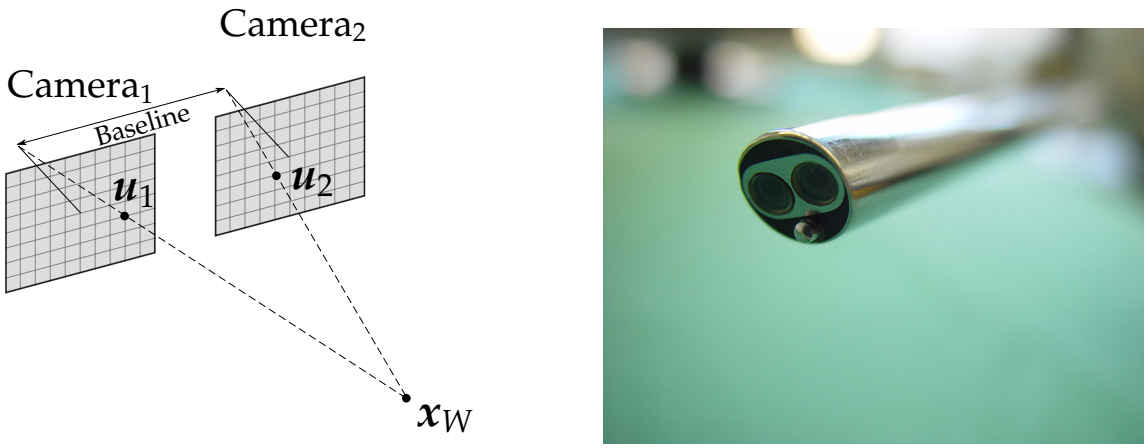
where x_3 is given by the range data. Different concepts for range image acquisition have been proposed. For abdominal surgery three major techniques have evolved as described in Chapter 3. In this work, in particular, ToF sensors are utilized to deliver topographic data in a fast manner. Detailed information on the ToF measurement technique is given in Section 3.3.

CHAPTER 3

Range Sensors for Endoscopy

Different hardware setups for range image acquisition are already implemented in a multitude of different applications, e.g. entertainment systems like the Kinect©[1] or radiotherapy systems like Cyberknife©[2]. Besides several *Shape-from-X* approaches, e.g. Shape-from-Shading [Wu 09, Mour 01], the three most popular acquisition techniques are using *Stereo Vision*, *Structured Light* or *Time-of-Flight*. With increased precision and reduced manufacturing costs, these devices gained interest for new application as well, e.g. driver assistance system or minimally invasive surgery. Recently, the three major concepts were adopted and implemented into endoscopic hardware and thereby face the conflict of delivering data with high accuracy and miniaturizing the setup for minimally invasive surgery. Today, only stereo endoscopes are commercially available, but setups using structured light and ToF technology are highly investigated by different researchers. Maier-Hein et al. [Maie 13, Maie 14] have published a first comparison between all three range acquisition techniques and evaluated their performance on real ex-vivo data. A long-term goal for range imaging in abdominal surgery was proposed by Su et al. [Su 09], where data registration of intra operative range data and preoperative computed tomography (CT) data is described for augmented reality. This combination would allow to have a detailed visualization of the high resolution volumetric CT dataset while it is correctly aligned to the current position of the endoscope. This improves orientation and thereby allows better navigation within the abdominal cavity. This chapter will focus on the working principle and the state-of-the-art of available range imaging systems and conclude with a first comparison of these approaches as proposed in [Maie 14]. We only consider established range image setups that acquire data in a single shot procedure, i.e. all Shape-from-X techniques are not included.

[1] Microsoft Corporation, Redmond, USA
[2] Accuray Incorporated, Sunnyvale, USA



(a) (b)

Figure 3.1: (a) describes a stereo vision setup with two cameras Camera_1 and Camera_2 that observe a 3-D point $\boldsymbol{x}_W$. The projection of this point onto each image plane results in an pixel index $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ which can be computed by Eq. (1.3). (b) shows the front view of a stereo endoscope with two apertures to acquire images from two different points of view [Spei 09].

3.1 Stereo Vision

Stereo endoscopy is the most commonly used 3-D range image acquiring technology in minimally invasive surgery. Besides its application in rigid endoscopes, stereo vision is also implemented in the da Vinci assistance system. Stereo vision describes an intuitive acquisition technique that is similar to the human vision and depth estimation.

Working Principle The core concept behind stereo endoscopy is to estimate range information by observing a scene from two different perspectives. Given a known baseline, the framework has to detect the 2-D projections of a 3-D point in both image planes. In theory, using basic trigonometry the range information of these points can then be computed by triangulation, see Fig. 3.1. In practice, both lines will probably not intersect and minimizing the distance of both lines will estimate the position of the 3-D point.

The requirements for stereo endoscopy are on the one hand a precisely calibrated device and on the other hand a diversified texture information of the observed scene. Accuracy is increased with a wider baseline between both sensors. As this baseline is limited by the diameter of the endoscope, the improved accuracy has to be gained by computing the corresponding points in both images with higher precision. Corresponding points are computed by detecting features in both images, e.g. by applying the scale-invariant feature transform (SIFT) [Lowe 04] or by computing speeded-Up robust features (SURF) [Bay 06]. Matching those feature points results in point pairs that correspond to the same 3-D point in the observed scene. Therefore, the output of a stereo endoscope highly depends on the quality of the two images and on the speed and robustness of the

feature detection and matching. As the estimated range images of a stereo setup are computed from disparity maps and highly depend on the texture variety in the observed scene, the range data resolution is spatially varying and can not be considered constant in the entire image. In areas with homogeneous color information no range data can be estimated.

As the bottleneck of this range image acquisition technique is the computation of corresponding feature points, hardware manufacturers tackle this problem by increased image resolution in the sensor domain. This leads to more details even in almost homogeneous regions in the acquired images but also induces more computational effort to compute features on both images. Therefore, estimating accurate 3-D range data in real-time is a major issue in stereo endoscopy.

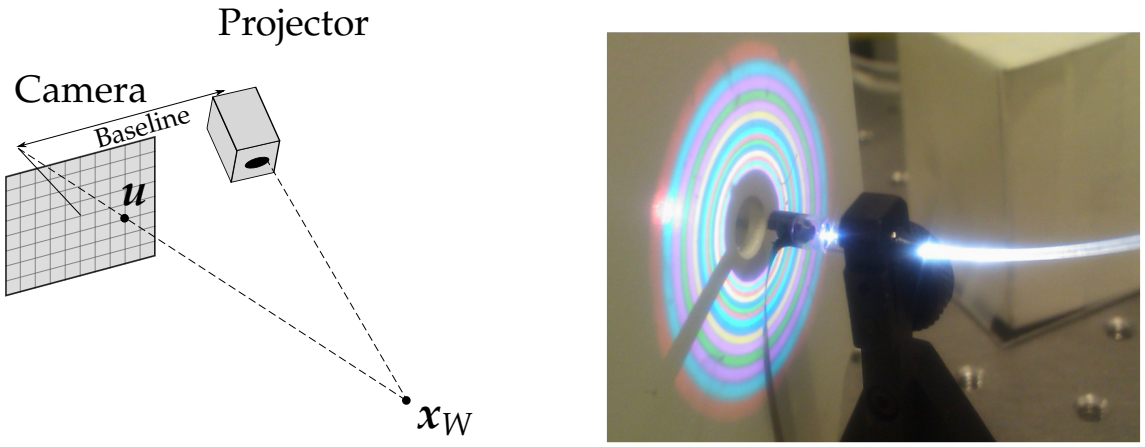
State of the Art Stereo endoscopes are currently the only commercially available and CE certified 3-D endoscopes. However, concrete benefits of stereoscopic reconstructions in endoscopy have not been proved yet. New applications and algorithms for disparity maps and stereo endoscopic reconstructions are still investigated. Mueller et al. [Muel 04] summarized the possibilities and limitation of those setups for minimally invasive surgery. Two major trends in stereo endoscopy have evolved, recently. On the one hand computation of feature points and thereby computation of the disparity map is transferred onto general purpose computation on graphics processing units (GPGPU) as proposed by Röhl et al. [Rhl 12] and Stoyanov et al. [Stoy 10]. On the other hand, extending the field of view by registering successive stereo range image frames in an extended simultaneous localization and mapping (SLAM) approach is an important topic to improve the surgeon's orientation as published by Totz et al. [Totz 11].

3.2 Structured Light

Structured light endoscopy is a novel technique based on stereo vision concepts but with artificially created feature points instead of those given by textural information. In minimally invasive surgery, structured light systems are not yet commercially available and only few publications have addressed this technique so far, see [Clan 11, Schm 12].

Working Principle The working principle of structured light sensors is very similar to stereo vision systems. In comparison to those, structured light systems do not require two cameras observing the scene. The second camera is replaced by a projector that generates a known pattern onto the observed scenario, see Fig. 3.2. Still, the known baseline and corresponding points in the acquired image and the known projection pattern are used to reconstruct 3-D points by triangulation. As the projector generates the pattern that is used to compute the feature points, structured light is also called an active triangulation technique.

Similar to stereo vision, the baseline between the sensor and the projector and an accurate calibration is required for high quality measurements. As long as the pattern is clearly visible, this technique is independent of texture information of



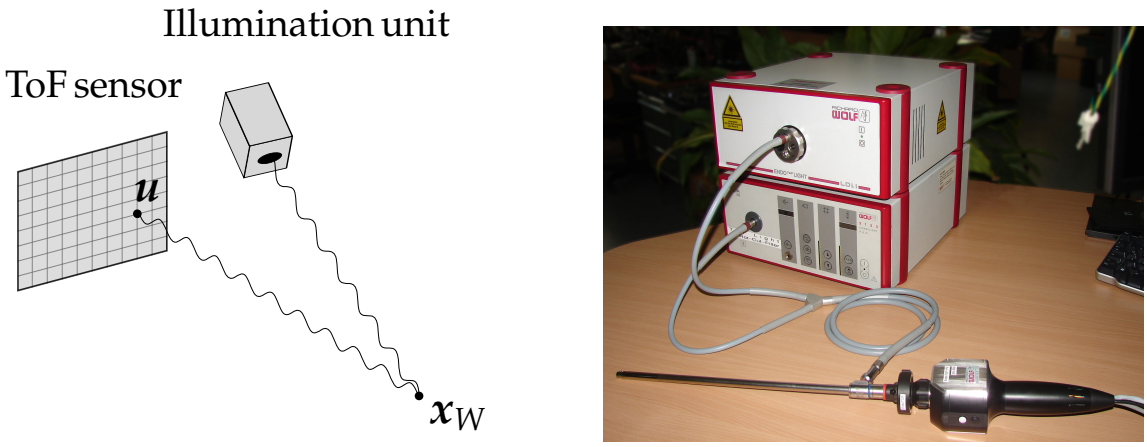
(a) (b)

Figure 3.2: (a) describes a structured light setup with one camera and one projector that generates a known pattern onto the scene. The projection of this point onto the image plane results in an pixel index $\boldsymbol{u}$. Together with the projection pattern, the 3-D world coordinates $\boldsymbol{x}_W$ can be reconstructed using triangulation. (b) shows the prototype proposed by Schmalz et al. [Schm 12].

the observed scene. Opposed to stereo vision systems, the feature detection framework can be highly adapted to the projection pattern, as the structure of the feature points is known. The projection pattern should be easy to detect and hard to disturb by the texture of the scene. In conventional structured light setups, stripes or sinusoidal patterns are popular. In minimally invasive setups, color coded dots [Clan 11] and circles [Schm 12] are employed.

As the core concept for structured light is similar to stereo vision, it shares the same bottleneck of detecting and identifying the feature points of the projected pattern. Furthermore, the smaller the structures of the projected pattern are, the more 3-D points can be reconstructed, but the harder it is to identify those structures in the acquired images.

State of the Art Only few prototypes of structured light devices for minimally invasive procedures exist. Clancy et al. [Clan 11] proposed a device that is capable to reconstruct only approx. 50 3-D points, but with high accuracy. They manufactured a fiber-based structured light probe that projects color coded dots onto the scene. A different setup was proposed by Schmalz et al. [Schm 12], where the projected pattern consists of color coded circles. This allows a reconstruction of approx. 5000 high accurate 3-D points. Beside the lack of a CE certification, the disadvantage of this prototype is missing photometric information. Current structured light technology as it is implemented in the Kinect© has tackled the issue of the visible projected pattern by using near-infrared wavelength.



(a) (b)

Figure 3.3: (a) describes a ToF setup with a single camera and a illumination unit that sends out the modulated light. The reflected signal is then received by the ToF sensor at pixel index $\boldsymbol{u}$. After calculating the phase shift, the radial distance is computed. (b) shows the prototype described by Haase et al. [Haas 13b]. The setup includes two separated light sources for the color and the ToF acquisition.

3.3 Time-of-Flight

ToF technology tackles the topic of 3-D reconstruction from a completely different point of view. Instead of using acquired color images to find any distinctive structures, reflection characteristics are exploited to physically measure the distances of the observed scene [Lang 01].

Working Principle The concept behind ToF technology is to measure a frequency modulated light ray that is sent out by an illumination unit and received by a ToF sensor, see Fig. 3.3. The received sinusoidal signal is sampled at four timestamps to estimate the phase shift Φ between the emitted and received signal. The radial distance d is then computed by:

$$d = \frac{c}{2f_{\text{mod}}} \cdot \frac{\Phi}{2\pi}, \tag{3.1}$$

where c denotes the speed of light and f_{mod} the modulation frequency [Lang 01].

As the illumination unit can be realized by LED diodes and the sensor is a simple CMOS or CCD chip, production costs of ToF sensors are rather low. However, due to their novelty compared to stereo vision, current ToF devices exhibit low data quality and low image resolution. Besides the range image, most ToF devices provide additional data, e.g. photometric data, often denoted as the amplitude image, and a binary validity mask. Due to its measurement technique ToF setups do not require a baseline between the illumination unit and the measuring sensor, which is beneficial for the use in minimally invasive surgery.

Besides the low resolution, systematic errors reduce the data quality extremely. Kolb et al. [Kolb 10] proposed a report on ToF sensors and described their error

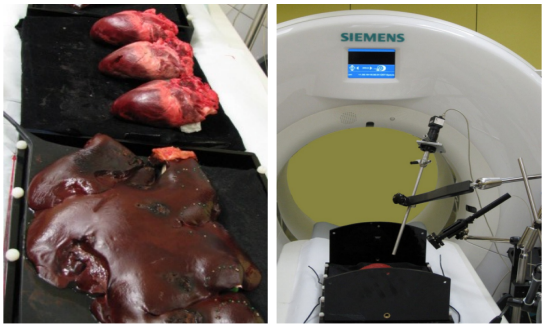


Figure 3.4: Two photos of the experimental setup in [Maie 14]. Different organs were examined in realistic medical scenarios, including surgical cuts, smoke and blood.

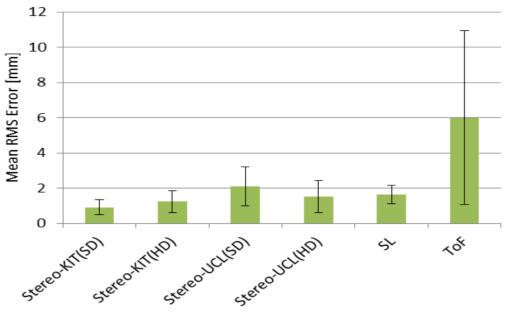


Figure 3.5: A boxplot comparing the mean root-mean-square error of four stereo endoscopic systems, a structured light setup and a Time-of-Flight endoscope [Maie 14].

sources, e.g. color dependent range measurements, temperature issues of the devices and flying pixels at object boundaries. In minimally invasive procedures two major issues occur. First, multiple reflection within the abdominal cavity corrupt ToF measurements. Second, inhomogeneous illumination caused by the endoscopic optic hinders accurate range measurements.

State of the Art For minimally invasive surgery, only very few prototypes of ToF endoscopes are described in literature. Penne et al. [Penn 09] proposed their first prototype of a rigid 3-D ToF endoscope in 2009 as a university research project. In [Penn 10] they extended that hardware by an attached RGB endoscope for a first feasibility study of ToF/RGB data fusion for 3-D endoscopy. In 2012, Haase et al. [Haas 12] published a first approach to fuse ToF range data and RGB color data acquired with a novel hybrid ToF/RGB endoscope manufactured by Wolf GmbH, Knittlingen, Germany. This hardware acquires both complementary data through a single optical channel and is further described in Section 3.6. In [Haas 13a] a first approach of an ToF based satellite camera was introduced, see Section 3.6.

3.4 Comparison on Real Data

As mentioned in the previous sections, all three single shot range image acquisition techniques have their benefits and drawbacks. In [Maie 14] a first quantitative comparison study has been presented. A single structured light sensor, one ToF setup and four stereo vision setups were investigated. As illustrated in Fig. 3.4, the study considered several different organs and included challenging scenarios, e.g. a surgical cut, smoke and blood. For quantitative comparison all scenarios were also measured in a CT scanner before and after the experiments. This data served as ground truth data for the study. To register the acquired range images and the ground truth data, small markers were attached to the regions of interest. The evaluation only considered the region inside those markers.

As illustrated in Fig. 3.5, the stereo endoscopes and the structured light setup achieved a similar mean root-mean-square error, whereas the ToF device exhibits a rather high error. However, this device is still in an early prototype status and thereby is hardly comparable with the other setups. Due to its compact housing and on-chip acquisition technique the ToF technology was chosen for further research described in this thesis. Furthermore, ToF sensors have less issues in terms of occlusions due to non existing baseline in the setup, in comparison to the stereo vision and structure light.

3.5 Time-of-Flight Sensor Issues

All range image acquiring systems exhibit similar error sources. As this thesis utilizes ToF technology for data acquisition, we focus this section on ToF issues. Due to its novelty ToF imaging technology still suffers from a multitude of different drawbacks and error sources [Kolb 10]. A variety of correction methods from calibration techniques to filtering were investigated and summarized by Reynolds et al. [Reyn 11]. This chapter describes three major issues regarding ToF sensors in general and their application in minimally invasive surgery, in particular. First, the low SNR makes image preprocessing a mandatory first step. Second, the large pixel size of those sensors limits the measured range image resolution. Third, specular highlights, known to be an issue in medical applications for conventional RGB endoscopes [Zimm 06, Arno 10], result in invalid range measurements as well.

Low Signal-to-Noise Ratio As mentioned in Section 3.3, ToF data acquiring sensors face several different error sources that lead to temporal noise, systematic offsets and invalid or incorrect range data. Temporal noise is most notably at dark areas, where the reflected signal is of poor quality. In static scenarios temporal noise can be tackled with averaging successive frames. However, in dynamic scenes this leads to motion artifacts and thereby to incorrect data. Different reflectivity also leads to an amplitude related offset. Other systematic offsets are caused by the sensor's temperature and the integration time as described by Mersmann et al. [Mers 13]. Another issue of ToF devices is the correct measurement of object boundaries where different distance levels adjoin each other. As described by Sabov et al. [Sabo 10], flying pixels are randomly assigned to the foreground object or background. In the abdominal cavity a major problem occurs from multiple reflections. Several emitted light rays are reflected at multiple objects before being received by the sensor. Here, the chip is not able to separate these signals and thereby computes incorrect range measurements.

Low Spatial Resolution Pixel spacing, pixel sizes and the chip dimensions of an imaging system define the spatial resolution of the sensor and thereby the quality of details that are visible in the acquired data. ToF sensors, in particular, are still manufactured with rather low spatial resolution which leads to an image resolution of 512×424 px at the maximum for the Kinect One© down to 64×48 px for the ToF/RGB endoscope prototype described in Section 3.6. In hybrid imaging setups,

the complementary photometric sensor usually exhibit a higher image resolution for a similar field of view. This leads to the concept to increase the low-resolution (LR) ToF data not only by naive upsampling, e.g. using bilinear or bicubic interpolation, but by including high-resolution (HR) photometric information into the upsampling process of the acquired range data. Incorporating information with higher SNR and higher image resolution allows to reconstruct these features in the range domain. This leads to higher accuracy in identification and localization of important image structures.

Specular Highlights In conventional endoscopy specular reflections are a known issue, as in those areas pixels are oversaturated and only deliver white color information. Here, light rays hit the object perpendicular to its surface plane. Different approaches have been published to detect specular reflections and to replace invalid photometric information, e.g. by data provided by a technique based on a normalized convolution [Arno 10], anisotropic diffusion [Grge 01] or temporal registration [Stoy 05]. Range image data is equally affected by specular reflections. Some sensors are able to detect affected regions and deliver this information in its validity mask. However, highlight boundaries with invalid data often remain and small specular regions are often ignored. Here, it is hard to identify small specular highlights as incorrect data in the range domain as they might just result in small topographic offsets. Therefore, additional photometric data, either acquired by another sensor or delivered by the ToF device itself can be exploited to detected specular highlights and mark them invalid. Removing invalid range data in those areas can either be performed by interpolation techniques as proposed by Wasza et al. [Wasz 11a] or by replacing it with correct data of different images as proposed in Chapter 7.

3.6 Employed Hardware

For the evaluation of the algorithms on real data as described in the following chapters, we examined two different sensors, dependent on the application. On the one hand we had access to a 3-D ToF/RGB endoscope with low SNR but high-quality RGB information in addition to range data. On the other hand, we used a miniature ToF sensor acquiring range data with improved image resolution and data quality as a reference for a range data acquiring satellite camera.

Hybrid 3-D Endoscope The ToF/RGB endoscope is manufactured by Richard Wolf GmbH, Knittlingen, Germany. The prototype acquires ToF (64×48 px) and RGB (640×480 px) data simultaneously through one optical system at a frame rate of 30 frames per second (fps). In addition, the photometric amplitude image and a flag image is delivered to indicate for each pixel if its measured range is reliable or erroneous. Nevertheless, most valid pixels show a low SNR due to several error sources, e.g. temperature related or amplitude related offsets [Kolb 10]. Since this endoscopic system is in an early prototype stage, all experiments were only performed in ex-vivo studies with porcine organs or realistic human organ phantoms.

3-D Satellite Camera Our reference hardware for a range data acquiring satellite camera is a PMD CamBoard nano manufactured by pmdtechnologies GmbH, Siegen, Germany. This device acquires data with up to 90 fps and delivers an amplitude, flag and range image with an image resolution of 160×120 px. The sensor is integrated into a miniature housing with a size of $37 \times 30 \times 25$ mm. This still excludes the use through a trocar in minimally invasive procedures, but it identifies the trend towards smaller devices that will fulfill the requirements for real-time applications in minimally invasive surgery.

Range Image Simulator For quantitative evaluations, a range image simulator was implemented that allows to imitate different range data acquiring setups. This data source behaves like a common range image device and delivers data according to its real counterpart. It allows simulation of the endoscope and the satellite camera, i.e. range images, amplitude images, validity maps and color images are generated. Furthermore, realistic lighting behavior is implemented including specular reflections. For simplicity, additive Gaussian noise and blur is used to obtain noisy data and simulate the optical system. To simulate the amplitude related error described in Section 3.5 the variance of Gaussian noise depends on the simulated amplitude data. The operation scenes are composed by medical experts. Based on CT scans, textured 3-D meshes of human organs are observed by a virtual range image device. The depth information of this 3-D rendered scene is then used as a ground truth range image. As different devices show different individual error sources, the range image simulator was designed to provide the basic behavior that most devices have in common.

The next chapter describes the required calibration of hybrid range imaging setups. With the knowledge of error-prone range data in most scenarios it also elaborates techniques to tackle the issues mentioned in Section 3.5.

Part II

Hybrid Range Data Preprocessing

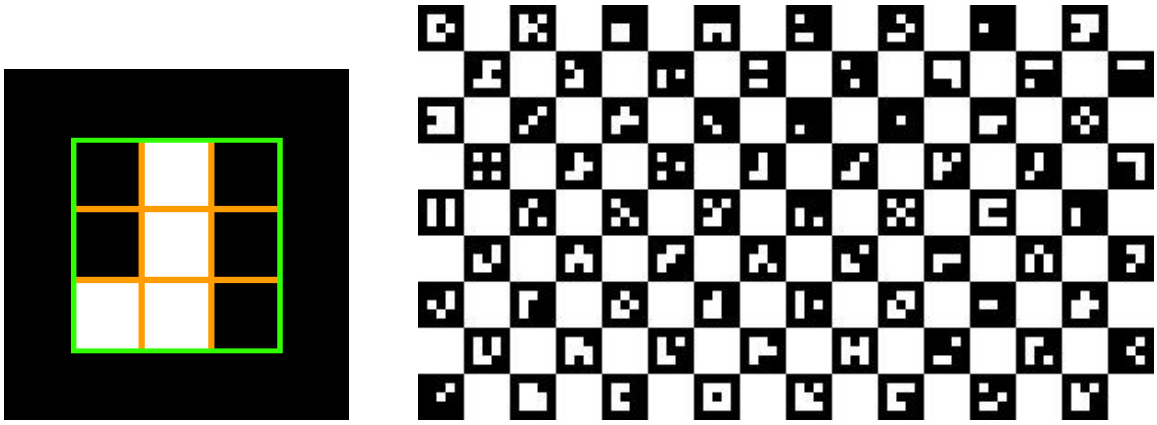
CHAPTER 4

Calibration and Data Fusion

Data fusion is one of the major topics in medical imaging. The key idea is that different sensors acquire complementary data and thereby are able to provide an augmented view for guidance during an intervention. As surgeons are used to endoscopic 2-D color data, without photometric information an intuitive representation of range data alone is hardly feasible. Therefore, an automatic calibration scheme for data fusion of photometric RGB information and ToF range data is investigated in [Haas 12, Haas 13b].

For ToF/RGB data, sensor fusion has been demonstrated for two individual devices in a stereo setup [Lind 07, Gudm 11] and in particular for ToF endoscopy by Penne et al. [Penn 10]. Compared to these setups, the device described by Haase et al. [Haas 12, Haas 13b] acquires data through a common optical system. This improvement allows to use a robust homographic mapping instead of exploiting the error-prone range information to project 3-D points onto the RGB sensor. Both mapping techniques require a calibration of the RGB and ToF sensor beforehand.

Conventionally, the corners of neighboring patches of a checkerboard pattern are used as features for camera calibration. These feature points can be detected automatically with established calibration frameworks. For ToF endoscopy, calibration is a highly recurrent task as a recalibration of the system must be performed each time the endoscope optics are changed. Due to inhomogeneous illumination in the ToF images, conventional checkerboard detection algorithms result in a high error rate. Since the RGB and the ToF chip do not share the same geometries both sensors need to be calibrated separately. For estimating a relative transformation from the coordinate system of one sensor into the coordinate system of the other sensor, corresponding feature points in both views have to be detected. Therefore, 2-D barcodes are used for feature point identification as proposed by Fiala et al. [Fial 08]. Conventional checkerboard detection systems have to identify the complete checkerboard in each image. For extrinsic calibration in particular, it is not sufficient to detect feature points in each view, but it is also required to identify their position in the known ground truth checkerboard.



(a) (b)

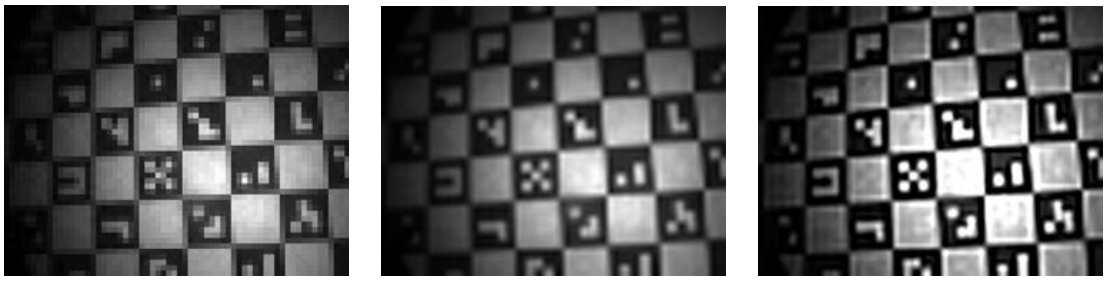
Figure 4.1: Detailed illustration of the self-encoded marker. (a) shows a single patch with green lines separating the barcode and the border and orange lines separating the barcode boxes. (b) is the set of all barcodes used for calibration.

This chapter describes an approach that uses the entire field of view at all distances even if the checkerboard is only partially visible. This allows to compute parameters with equal accuracy for the entire volume of interest, as the feature points throughout all images are evenly distributed across the whole field of view. For higher robustness in low resolution ToF images (64×48 px), we adopted the marker of Forman et al. [Form 11] with a reduced barcode size of 3×3. The detected corner points are used for camera calibration to estimate the intrinsic and extrinsic parameters.

4.1 Camera Calibration with Self-Encoded Marker

Camera calibration in general exploits known ground truth 3-D world coordinates and corresponding 2-D image coordinates to estimate parameters of a projection matrix by minimizing an error metric between the detected 2-D image coordinates and the results of projecting the 3-D points onto the image plane. Therefore, a plane calibration pattern with easy detectable feature points, e.g. black circles on white background or a checkerboard, is utilized to be observed from multiple different views to estimate the projection parameters robust for the whole volume of interest. In addition to a basic camera calibration we need to compute corresponding points in both sensor domains, this section starts with a detection scheme of a self encoded marker that allows to compute same feature points for both sensors.

Self-Encoded Marker As proposed by Fiala et al. [Fial 08], 2-D barcodes can be used for feature point identification, which are required for camera calibration. In our approach, we use a checkerboard marker with unique barcodes embedded in the checkerboard patches for a recognition of the feature points independently of the rotation of the entire marker [Form 11]. The 2-D barcode is described by 3×3 blocks as depicted in Fig. 4.1. Even in low-resolution images, a robust detection



(a) (b) (c)

Figure 4.2: Illustration of the image enhancement pipeline. (a) shows the original amplitude image. (b) is upsampled data generated with bicubic interpolation. (c) shows the image after applying an unsharp mask.

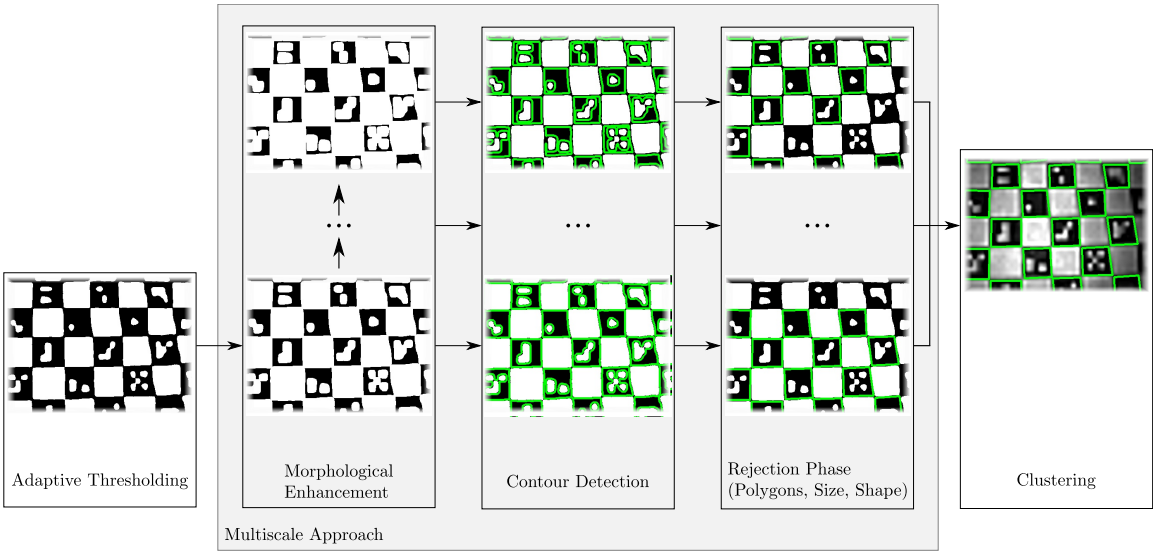


Figure 4.3: Workflow of the marker detection process. The different scales of the patch recognition phase vary in the number of erosions and dilations applied on the binary image. The examples show an image section of a Time-of-Flight amplitude image. Green pixels denote the detected contours.

of the barcodes is feasible. All barcodes are embedded into a checkerboard patch and thus surrounded by a black border. The feature points for calibration are the checkerboard corners identified by the barcodes.

Time-of-Flight Image Enhancement As this approach utilizes an implementation of a pixel-accurate framework, upsampling the ToF amplitude images is the initial step. Next, preprocessing these images as shown in Fig. 4.2 is required to compensate for inhomogeneous illumination. After bicubic resampling, unsharp masking is performed for local contrast enhancement [Mali 77]: $\hat{i}^{\boldsymbol{u}}_{\text{Amp}} = i^{\boldsymbol{u}}_{\text{Amp}} - i^{\boldsymbol{u}}_{\text{Amp,blur}}$, where $\hat{i}^{\boldsymbol{u}}_{\text{Amp}}$ is the contrast enhanced photometric amplitude value at pixel position $\boldsymbol{u}$ and $i^{\boldsymbol{u}}_{\text{Amp,blur}}$ is a blurred version of the input image. As illustrated in Fig. 4.2c, the contrast was improved after applying our preprocessing pipeline.

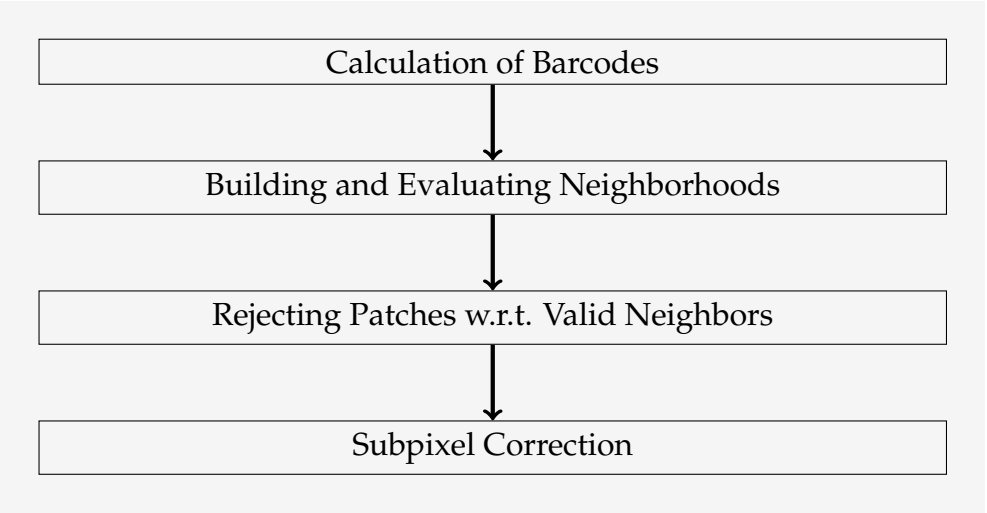


Figure 4.4: Workflow of the marker identification process describing the phases leading from barcode calculation to checkerboard corner detection.

Marker Detection The marker detection process is illustrated in Fig. 4.3 and demonstrated for a ToF amplitude image. For marker detection, we use a binarized version of the ToF amplitude and RGB input image. Therefore, an adaptive thresholding technique is performed calculating the threshold individually for each pixel depending on its neighborhood [Well 93]. In order to retrieve the contours of each checkerboard patch, morphological erosion and dilation is applied on the binary images to separate the blurred patches. Enhancing the detection algorithm proposed by Haase et al. [Haas 12], the inhomogeneous illumination in the acquired images is addressed by using a multiscale approach. Within each scale a morphological opening is performed, which is a dilation of the eroded image. Across multiple scales the number of erosions applied on the image before applying the same number of dilations defines the individual scale. The output of the morphological enhancement is used for contour detection as proposed by Freeman [Free 70] to find the checkerboard patches. Subsequently, a shape analysis is performed on all contours. First, the contours are approximated by polygons [Doug 73] and rejected if an approximation by four points is not achieved within ten iterations. Then, the contours are analyzed by their shape and length. Contours with a non-square convex hull or with an unexpected size are rejected. Finally, a clustering, similar to [Corm 01], is performed across all scales to combine contours of different scales describing the same checkerboard patch.

Marker Identification The marker identification process is depicted in Fig. 4.4. First, all detected patches are identified by their barcode. The barcode is calculated by dividing each patch in 5×5 blocks and analyzing the inner 3×3 blocks. The barcode is represented by a unique hash value calculated by the number of black blocks and their position. Therefore, all barcodes have to be unique in terms of rotation in order to prevent wrong identifications if the marker is rotated by more than 45°. Subsequently, these identified barcodes are associated in a common structure describing their neighboring patches. The same structure is constructed

for the ground truth image, respectively. In order to verify the identified barcodes, each associated neighbor of an identified patch is compared to the ground truth neighbor. A score calculated by the validity of all four neighbors of each patch is finally used to reject an incorrect identified barcode. For the upcoming calibration process all identified checkerboard corners are then corrected with subpixel accuracy by gradient analysis.

Camera Calibration Finally, the previously identified checkerboard corners are utilized for camera calibration. For estimating the intrinsic parameters, the corners in all views are associated to their real world coordinates using prior knowledge about the checkerboard geometry. Following the approach of Zhang et al. [Zhan 04], the focal length (f_{x_1}, f_{x_2}) and the principal point (c_{x_1}, c_{x_2}) assembling the camera matrix $\boldsymbol{K} \in \mathbb{R}^{3\times3}$ are estimated dependent on the pixel spacing. The matrix is calculated by minimizing the reprojection error e^{rep} using a Levenberg-Marquardt optimization:

$$[\hat{\boldsymbol{R}}, \hat{\boldsymbol{t}}, \hat{\boldsymbol{K}}] = \arg\min_{\boldsymbol{R},\boldsymbol{t},\boldsymbol{K}} \quad e^{\text{rep}}(\boldsymbol{X}_{\text{det}}, \boldsymbol{X}_{\text{est}}), \tag{4.1}$$

where $\boldsymbol{X}_{\text{det}}$ is a matrix composed by all detected feature points in the acquired image. $\boldsymbol{X}_{\text{est}}$ is composed by the 2-D points calculated by projecting the 3-D ground truth coordinates onto the camera plane with the estimated camera parameters $\boldsymbol{R}$, $\boldsymbol{t}$ and $\boldsymbol{K}$ by $\widetilde{\boldsymbol{x}}_{\text{est}} = \boldsymbol{K}(\boldsymbol{R}\boldsymbol{x}_W + \boldsymbol{t})$. From now on the equation sign is used for homogeneous coordinates, which means that both sides are equal up to a scale factor. The error e^{rep} is defined by the mean of the euclidean distance between each point pair.

4.2 Color and Range Data Fusion

Stereo Setup Previously proposed approaches for ToF/RGB image fusion estimate the relative transformation using the extrinsic parameters of both sensors, see Fig. 4.5. First, the 3-D world coordinates are calculated based on ToF range data. Second, all 3-D points are transformed into the RGB sensor coordinate system. Third, the derived 3-D points are projected onto the RGB image plane. The relative transformation between both sensors is described by:

$$\boldsymbol{R}_{\text{rel}} = \boldsymbol{R}_{\text{RGB}}\boldsymbol{R}_{\text{ToF}}^{\top}, \quad \boldsymbol{t}_{\text{rel}} = \boldsymbol{t}_{\text{RGB}} - \boldsymbol{R}_{\text{rel}}\,\boldsymbol{t}_{\text{ToF}}, \tag{4.2}$$

where $\boldsymbol{R} \in \mathbb{R}^{3\times3}$ denotes a rotation matrix and $\boldsymbol{t} \in \mathbb{R}^3$ a translation vector and the index denotes the modality. Considering a pinhole camera model, a 2-D coordinate on the RGB sensor plane is computed by first applying Eq. (2.1) to reconstruct $\widetilde{\boldsymbol{x}}_{\text{ToF,C}}$, transform it into the RGB camera coordinate system and then apply Eq. (1.3):

$$\widetilde{\boldsymbol{x}}_{\text{RGB}} = \boldsymbol{K}_{\text{RGB}}(\widetilde{\boldsymbol{R}}_{\text{rel}}^{-1}(\widetilde{\boldsymbol{x}}_{\text{ToF,C}} - \widetilde{\boldsymbol{t}}_{\text{rel}})). \tag{4.3}$$

This deduces that the range data accuracy has direct influence on the mapping quality of the color information.

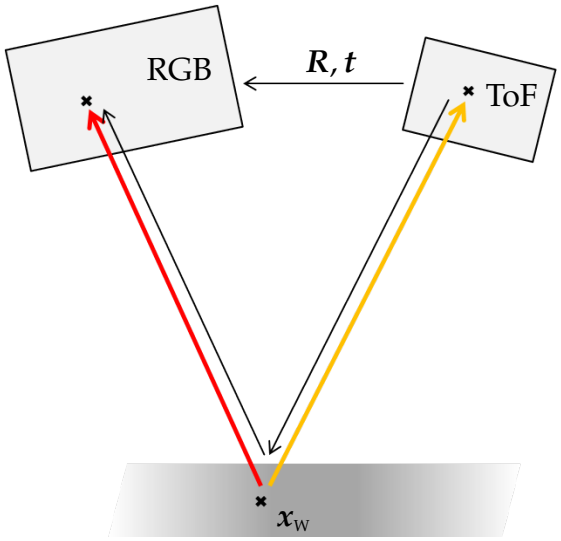


Figure 4.5: Mapping RGB color data and range data in a stereo setup. Data fusion is performed by reconstructing the 3-D points, transforming them according to the relative position of the sensors and projecting them down onto the RGB sensor plane.

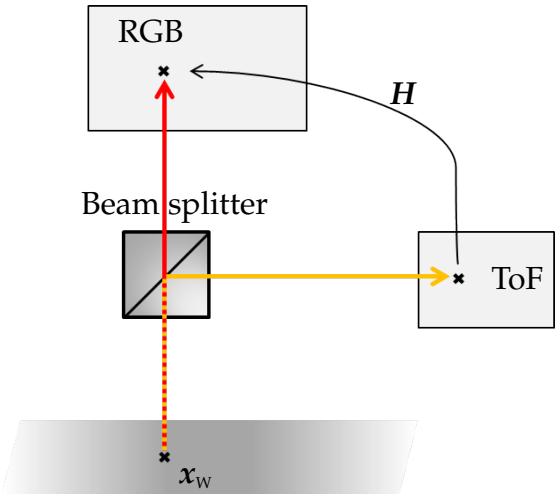


Figure 4.6: Mapping RGB color data and range data in a beam splitter setup. Data fusion is performed by applying a homography on each 2-D pixel position in the color domain to compute the corresponding pixel position in the Time-of-Flight domain.

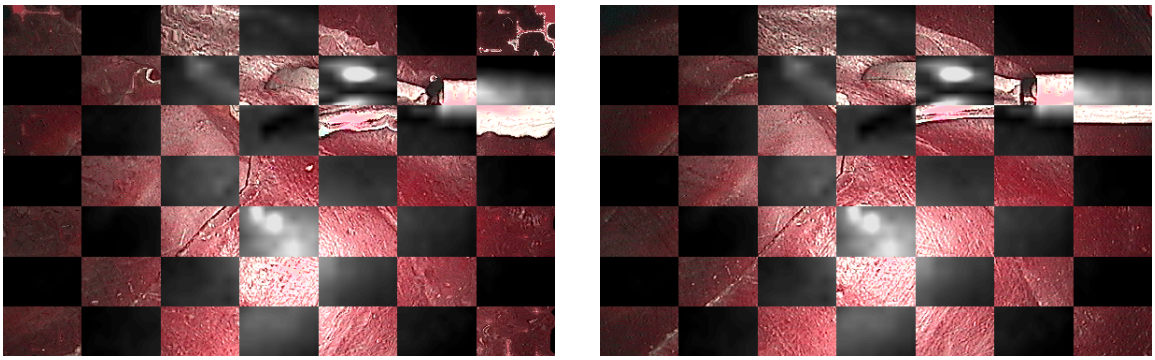
Homographic Setup In the described hybrid ToF/RGB endoscope both sensors acquire the scene through the same optical system, see Fig. 4.6. A beam splitter separates the incoming signal into near-infrared light for the ToF chip and the residual for the RGB chip. Since ToF and RGB images share the same center of projection [Ben 00], this allows to use a homographic mapping for transforming a 2-D RGB pixel $\widetilde{x}_{RGB}$ onto the ToF chip following the equation:

$$\widetilde{\boldsymbol{x}}_{\text{ToF}} = \boldsymbol{H}\widetilde{\boldsymbol{x}}_{\text{RGB}}, \tag{4.4}$$

where $\boldsymbol{H} \in \mathbb{R}^{3\times 3}$ denotes the homography between both sensor planes. This homography is estimated by employing the point correspondences in an image pair of both sensors.

4.3 Experiments

Hardware All experiments were performed using a 3-D endoscope prototype as described in Section 3.6 that acquires ToF and RGB data through a single endoscopic system. As the calibration pattern is printed on a flat plane, detecting and identifying the checkerboard patches in the range domain is not feasible. However, as ToF sensors also compute an amplitude image that provides photometric information, the framework is applied on this image to estimate parameters for the ToF sensor.



(a) (b)

Figure 4.7: Two checkerboard views of the Time-of-Flight/RGB fusion result. (a) shows a stereo vision mapping by projecting the reconstructed 3-D world coordinates on the RGB chip. (b) illustrates the improved homographic mapping with more robust results noticeable at the border of the endoscopic tool.

Calibration Data To enable a reliable evaluation, the calibration pattern was observed from 100 different views. The views were shifted in all directions and acquired from different angles and thereby contain a varying amount and size of checkerboard patches. For evaluating the robustness of our calibration algorithm, the reprojection error for a setup using 70 different views was computed for both sensors in 30 repetitions. Furthermore, the focal length and the principal point were calculated for different numbers of views for 30 repetitions each. The views were chosen randomly from all 100 views without using any view twice within one repetition. Evaluating the barcode identification process was based on a gold standard in all images labeled by an expert.

Sensor Fusion For sensor fusion evaluation, we constructed a realistic medical scenario and used the stereo vision mapping according to Eq. (4.3) as well as the homographic mapping according to Eq. (4.4). In order to obtain a quantitative comparison for both techniques, the normalized mutual information (NMI) [Stud 99] was computed as a similarity measurement using the RGB image and the amplitude image, which represents the ToF domain. NMI is derived from information theory and computes the inherent dependence of two random variables and thereby allows to compute a similarity measurement between images acquired from different modalities. A checkerboard representation of both input images as depicted in Fig. 4.7 shows qualitative improvements.

In a second part we evaluate the data fusion by measuring distances on our 3-D reconstruction in the color domain and transferring those points into the ToF domain for distance computation. Compared to Field et al. [Fiel 09], the ToF surface mesh does not rely on corresponding feature points in a stereo setup and therefore provides measurable points in a dense manner all over the observed scene. Given a ground truth distance, e.g. the length of an endoscopic tool, we compare the known length and the computed distance. This comparison was repeated for 30 successive frames acquired in a realistic medical scenario with a liver phantom and endoscopic tools, see Fig. 4.8.

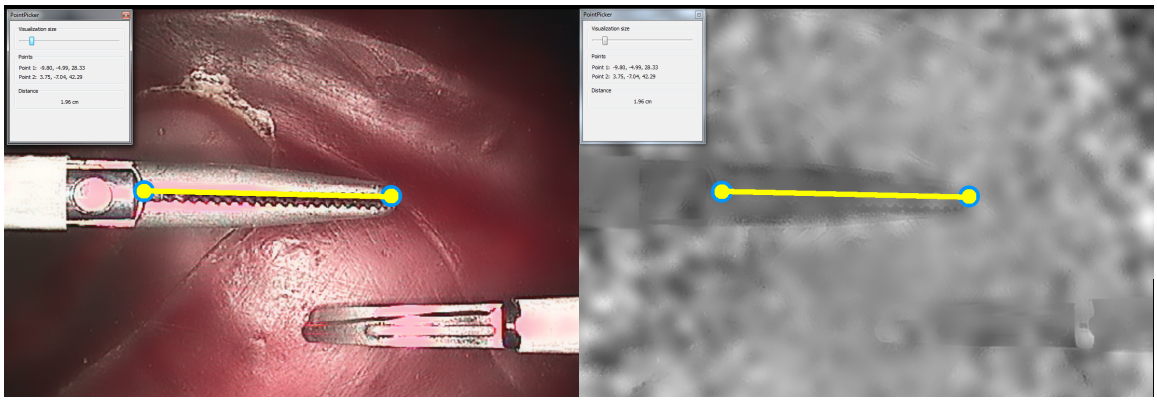


Figure 4.8: Measuring the length of the endoscopic tool in a realistic scenario. The yellow line denotes the measured length. Left: Mapped RGB image. Right: Time-of-Flight range image.

4.4 Evaluation and Discussion

Calibration Data The advantage of our multiscale approach is noticeable in the rejection phase of Fig. 4.3. On the last scale, our algorithm was able to detect different patches compared to the first scale. Only choosing one of those scales always leads to missing patches. In terms of marker identification, we achieved an identification rate of 92.7% for the RGB images. The small improvement from an identification rate of 92.3% in the single scale approach in [Haas 12] is due to the fact that almost all completely visible barcodes were identified using the conventional approach. The residual were expert labeled barcodes that were only partially visible and thereby not identified by our algorithm. In terms of the ToF data, we improved the identification rate of the barcodes from 92.0% [Haas 12] to 96.4% using our multiscale approach. Partially visible barcodes are less an issue in the ToF images due to the fact that the expert was not able to identify those barcodes either. Note that all identified barcodes are verified beforehand. Therefore, no erroneous identified barcodes are retained for calibration. As shown in Fig. 4.9, increasing the number of different views and thereby increasing the amount of checkerboard corners improves the robustness of intrinsic parameter calibration. Please note that 70 different views seems sufficient to result in a reliable calibration output as all relative standard deviations were below 5%. For 30 repetitions using 70 images for calibration the mean reprojection error resulted in 0.63 px for the ToF sensor and 0.49 px for the RGB sensor. The reprojection error allows a first interpretation of the quality of the estimated parameters, where a high reprojection error indicates that the parameters fit poorly to the input data. On the other hand, a low reprojection error combined with a huge collection of input data indicates high quality parameters. We managed to keep the reprojection error in subpixel accuracy for all input data. Comparing the ToF results and the RGB results in Fig. 4.9 leads to the conclusion that due to the higher SNR and the higher image resolution the RGB sensor requires less calibration images for robust results.

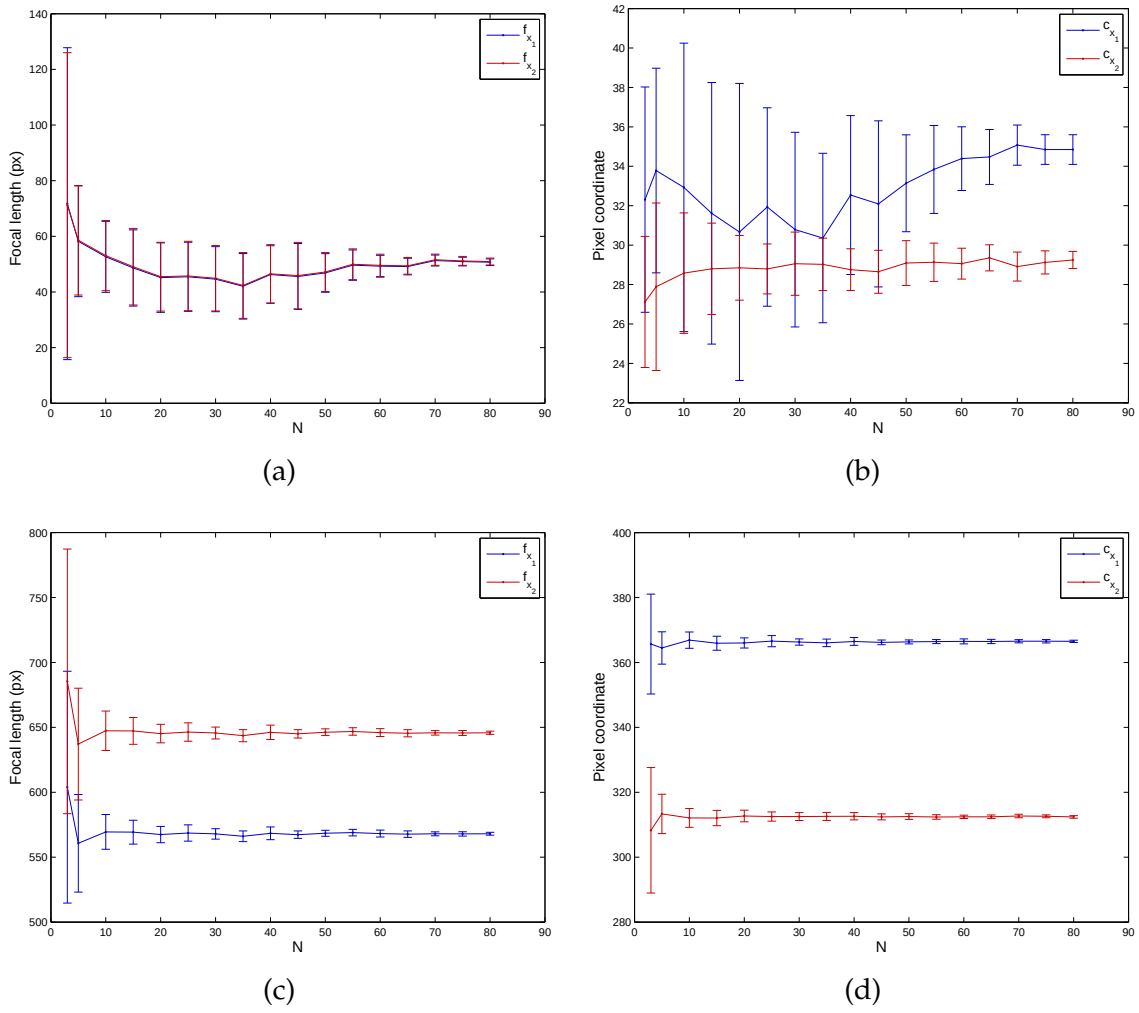


Figure 4.9: Plots of the mean and the standard deviation of the focal length (f_{x_1}, f_{x_2}) and the principal point (c_{x_1}, c_{x_2}) for different number of checkerboard views N. (a) and (b) illustrate the calibration output for Time-of-Flight data. (c) and (d) illustrate the calibration output for RGB data.

Sensor Fusion As shown in Fig. 4.7, the homographic mapping is independent of the error-prone ToF range values and, therefore, is more reliable. The improved results are noticeable along the border of the observed tool and at the corners of the image where range data is usually less reliable due to inhomogeneous illumination. The NMI between the amplitude image and the raw RGB image resulted in 0.92. After stereo vision mapping it was improved to 0.93. With our homographic transformation we achieved an NMI of 0.95. This leads to the conclusion that the initial misalignment of both input images is not huge, but can be further improved by our approach.

The tool tip measured as depicted in Fig. 4.8 has a length of 20.0 mm. Averaging the values using the range data for calculating this length resulted in a mean length of 18.0 mm with a standard deviation of 3.3 mm. This accuracy enables rough estimations within the human body. However, due to the early development stage

of this ToF/RGB endoscope, the accuracy is not yet sufficient enough for assisting surgeons during minimal invasive procedures, but it demonstrates the feasibility of measurements with improved hardware.

4.5 Conclusion and Future Work

An easy-to-use calibration technique for hybrid ToF/RGB miniature devices was presented and demonstrated for a 3-D endoscope. The evaluation has shown that the estimated intrinsic and extrinsic parameters are reliable and that metric measurements are possible in both the ToF and the RGB domain. The search for corresponding feature points in data of different imaging sensors to estimate a relative transformation was eased by utilizing a self-encoded marker with unique barcodes and an automatic detection and identification scheme.

Further research needs to investigate if different calibration patterns, e.g. circles, are able to provide more reliable results. Furthermore, it should be analyzed how the size of the barcodes influences the calibration accuracy.

CHAPTER 5

Hybrid Nonlocal Means Filtering

This chapter addresses the low SNR of ToF range data mentioned in Section 3.5. In image processing a variety of different approaches to reduce the noise level of acquired data have been proposed. Local filtering, such as the bilateral filter [Toma 98] or the guided filter [He 13], smooth the image while preserving edges. Here, a pixel is denoised by analyzing the local neighborhood. Then, the filtered pixel value is usually computed by a weighted average of the surrounded pixel values with the weights controlling the smoothness along and across edges. As each pixel can be calculated separately these local filters exhibit a fast computation time if parallelized for GPGPU. Both the bilateral filter and the guided filter have only few parameters, which eases their adaption to specific scenarios. Both techniques have been applied to ToF range data [Lenz 13, Wasz 11b] and have shown to reduce noise notably. In contrast to these preprocessing techniques based on data of a single modality, our setups allow hybrid approaches, like it is shown in [Kopf 07] for the joint bilateral filter. Here, high quality color information is utilized to denoise low quality range data. Nevertheless, those methods only exploit local structures to preserve image details while smoothing. However, repetitive structures distributed over the entire image can be exploited to denoise measured image data with a similar neighborhood.

Buades et al. [Buad 05, Buad 08] and Danielyan et al. [Dani 12] proposed nonlocal filtering techniques, where similar neighborhoods all over the image are used to denoise a specific region. Both approaches do not only try to find similar pixels in the local neighborhood of a pixel, but identify same structures across the entire image. A denoised pixel value can be computed by a weighted average of all other image points, with the weights describing the similarity of the neighborhoods. Hu et al. [Hu 13] have compared nonlocal filtering techniques for range data. However, the low quality of ToF range data hinders a robust similarity measure.

We introduce a hybrid approach for nonlocal filtering based on the example of the nonlocal means (NLM) filter [Buad 08]. Due to the low SNR of ToF range data, mapped high quality RGB data is used to compute neighborhood similarities for range data patches. Furthermore, the NLM filter was extended into the temporal domain as published in [Lind 14] to utilize an entire sequence of range images for further denoising.

5.1 Nonlocal Means Filtering

Buades et al. [Buad 05] proposed the NLM filter for 2-D images as a denoising technique that exploits the fact that in most scenarios similar structures appear all over the image. Later, Schall et al. [Scha 07] applied the same technique on range data and showed that nonlocal similarity measurements result in higher quality range data than conventional local measurements, e.g. used in the bilateral filter.

Originally, the NLM filter was proposed for monochrome images, but can easily be calculated for multichannel images by applying the technique on each channel. Each intensity i at pixel position $\boldsymbol{u}$ in a monochrome image $\boldsymbol{i}$ is denoised by:

$$\hat{i}^{\boldsymbol{u}} = \sum_{\boldsymbol{v} \in \omega^{\boldsymbol{u}}} \frac{1}{k^{\boldsymbol{u}}} w(\boldsymbol{u}, \boldsymbol{v}) i^{\boldsymbol{v}}, \tag{5.1}$$

where $\omega^{\boldsymbol{u}}$ denotes the search window around $\boldsymbol{u}$ and $w(\boldsymbol{u}, \boldsymbol{v})$ defines the similarity weight between the neighborhood around $\boldsymbol{u}$ and the neighborhood around $\boldsymbol{v}$. $k^{\boldsymbol{u}}$ is a normalizing constant for each weight. The weights $w(\boldsymbol{u}, \boldsymbol{v})$ are computed by:

$$w(\boldsymbol{u}, \boldsymbol{v}) = \exp\left(-\frac{1}{h} \sum_{\boldsymbol{v}' \in \omega'} \exp\left(-\frac{||\boldsymbol{v}'||_2^2}{\sigma^2} \right) |i^{\boldsymbol{u}+\boldsymbol{v}'} - i^{\boldsymbol{v}+\boldsymbol{v}'}|^2 \right), \tag{5.2}$$

where h is the filtering parameter that limits the influence of both the pixel distance and the intensity difference, σ is the standard deviation of the Gaussian kernel and ω' is the similarity window. The points $\boldsymbol{v}'$ are relative offsets describing the neighborhood of $\boldsymbol{u}$ and $\boldsymbol{v}$, e.g. $\boldsymbol{v}' \in [-1, 0, 1] \times [-1, 0, 1]$ for a 3×3 similarity window. The normalizing constant is computed for the entire search window by:

$$k^{\boldsymbol{u}} = \sum_{\boldsymbol{v} \in \omega^{\boldsymbol{u}}} w(\boldsymbol{u}, \boldsymbol{v}). \tag{5.3}$$

In Fig. 5.1 the NLM filter is illustrated for different pixels at a single time step t_0.

5.2 Hybrid Nonlocal Means Filtering

Schall et al. [Scha 07] have shown that directly applying a nonlocal filter on range data notably increases its SNR and achieves better results compared to local bilateral filtering [Toma 98]. However, their experiments were based on data captured with a laser and structured light scanner. In contrast, our setups have a very fast acquisition time but due to the novelty of ToF devices show a rather low quality

output. This leads to unreliable results by simply using Eq. (5.2) for calculating the weights necessary for averaging all pixels within the search window appropriately. Noisy data might influence the filter output by pretending structures that do not exist in ground truth data. Therefore, the NLM filter was extended in a hybrid manner by using a more reliable data source for calculating the similarity weights. In our endoscopic setup, in addition to the ToF data source we have additional RGB information of higher resolution and higher quality available. Similar to Huhle et al. [Huhl 10], we calculate the similarity weight of corresponding neighborhoods not only in the range domain but also in the grayscale converted photometric domain. This is based on our texture mapping as described in Chapter 4. In minimally invasive procedures usually this is valid, as important structures that differ in the range domain, e.g. organ boundaries, are also clearly visible in the color domain. As the pixel density of both color and range images differ by a factor of ten we simply adopt the color data dimensions to the image resolution of the range data. Based on Eq. (5.1), this leads to the following equation for denoising a range value i_{ToF} at position $\boldsymbol{u}$ considering weights in both data sources:

$$\hat{i}^{\boldsymbol{u}}_{\text{ToF}} = \sum_{\boldsymbol{v}\in\omega^{\boldsymbol{u}}} w_{\text{Hyb}}(\boldsymbol{u},\boldsymbol{v}) i^{\boldsymbol{v}}_{\text{ToF}}, \tag{5.4}$$

where w_{Hyb} denotes the hybrid similarity weight computed by:

$$w_{\text{Hyb}}(\boldsymbol{u},\boldsymbol{v}) = \alpha \frac{1}{k^{\boldsymbol{u}}_{\text{ToF}}} w_{\text{ToF}}(\boldsymbol{u},\boldsymbol{v}) + (1-\alpha)\frac{1}{k^{\boldsymbol{u}}_{\text{RGB}}} w_{\text{RGB}}(\boldsymbol{u},\boldsymbol{v}), \tag{5.5}$$

where $k^{\boldsymbol{u}}_{\text{ToF}}$ is calculated by Eq. (5.3) with the intensity being replaced by the corresponding range value i_{ToF}. w_{ToF} and w_{RGB} denote the similarity weights computed in the ToF domain and the color domain, respectively. The range reliability parameter α is chosen empirically and denotes the influence of the range similarity weight w_{ToF}. For simplicity, the color weights in the grayscale converted image are calculated equally to Eq. (5.2) and in the range domain by:

$$w_{\text{ToF}}(\boldsymbol{u},\boldsymbol{v}) = \exp\left(-\frac{1}{h}\sum_{\boldsymbol{v}'\in\omega'} \exp\left(-\frac{||\boldsymbol{v}'||_2^2}{\sigma^2}\right) |i^{\boldsymbol{u}+\boldsymbol{v}'}_{\text{ToF}} - i^{\boldsymbol{v}+\boldsymbol{v}'}_{\text{ToF}}|^2\right). \tag{5.6}$$

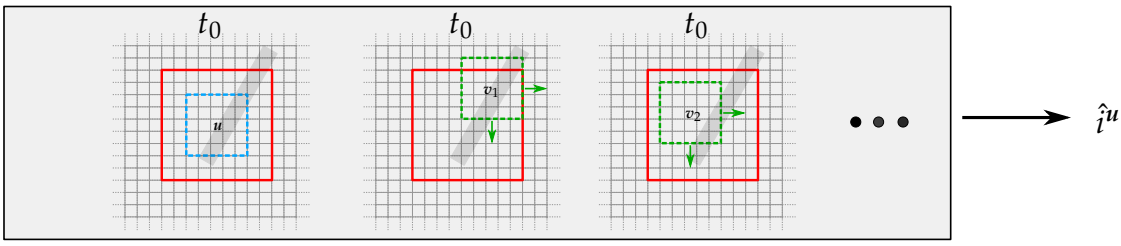


Figure 5.1: Workflow of the NLM filter. Red denotes the search window, blue the neighborhood of $\boldsymbol{u}$ and green the neighborhood of other pixels within the search window. The neighborhood in the middle shows a very high similarity with the blue one which will result in a high weighting $w(\boldsymbol{u},\boldsymbol{v}_1)$. The similarity of $\boldsymbol{u}$ and $\boldsymbol{v}_2$ is rather low which will result in a low weighting $w(\boldsymbol{u},\boldsymbol{v}_2)$.

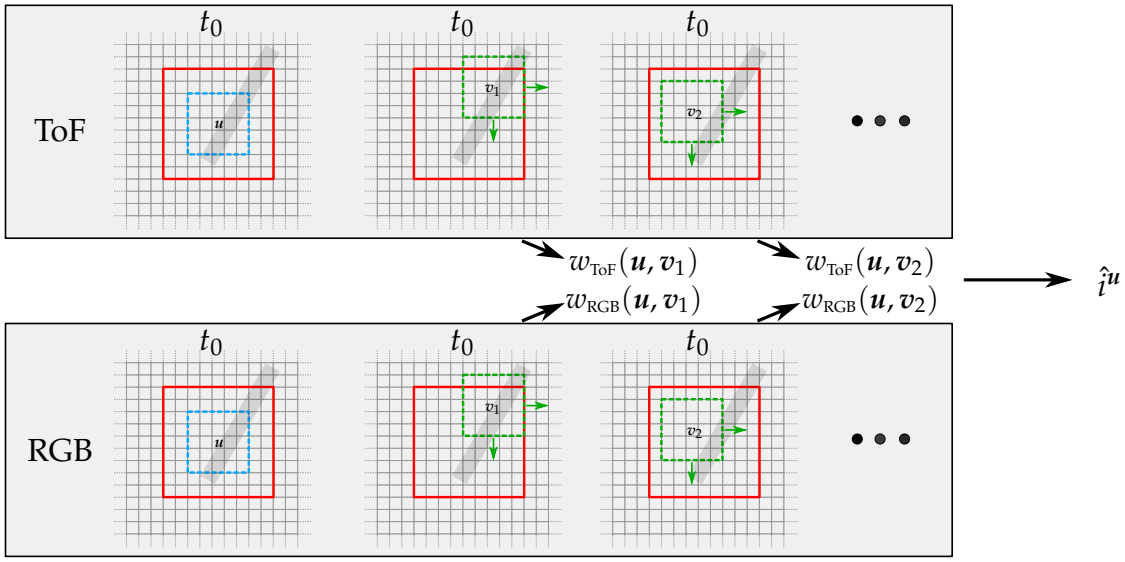


Figure 5.2: Workflow of the hybrid NLM filter. Red denotes the search window, blue the neighborhood of $\boldsymbol{u}$ and green the neighborhood of other pixels within the search window. The neighborhood in the middle shows a very high similarity with the blue one which will result in a high weighting $w(\boldsymbol{u}, \boldsymbol{v}_1)$. The similarity of $\boldsymbol{u}$ and $\boldsymbol{v}_2$ is rather low which will result in a low weighting $w(\boldsymbol{u}, \boldsymbol{v}_2)$. All similarities are computed in the ToF and RGB domain, simultaneously.

In a more general scenario, the similarity measurements can be replaced by any other function and for color images in particular, could be a function considering all three channels. Jingjing et al. [Dai 13] published an article about multichannel consideration in NLM.

In Fig. 5.2 the hybrid NLM filter is illustrated for different pixels for a single time step t_0. As calculations are performed in both domains the computational effort is increased. However, details not visible in the noisy ToF data can be induced into the output range image by the higher quality color information. Nevertheless, both the original NLM filter and the hybrid formulation need to have repetitive structures within the search window to achieve satisfying results.

5.3 Multi-Frame Hybrid Nonlocal Means Filtering

In images obtained by cameras covering a wide field of view, repetitive structures are a common feature. However, in the small field of view of endoscopic devices, this is not always guaranteed. Therefore, we enhanced the hybrid nonlocal means filter into the temporal domain. By extending the search window of similar structures from a single frame into a sequence of frames at different time steps, it is ensured that at least the same scene point is seen several times. Due to the temporal noise mentioned in Section 3.5 averaging corresponding pixels in a range image sequence can already increase ToF data quality. However, as organs and endoscopic tools move during medical procedures, keeping the search window fixed at a certain position for only 30 frames already means that motion is ignored for an

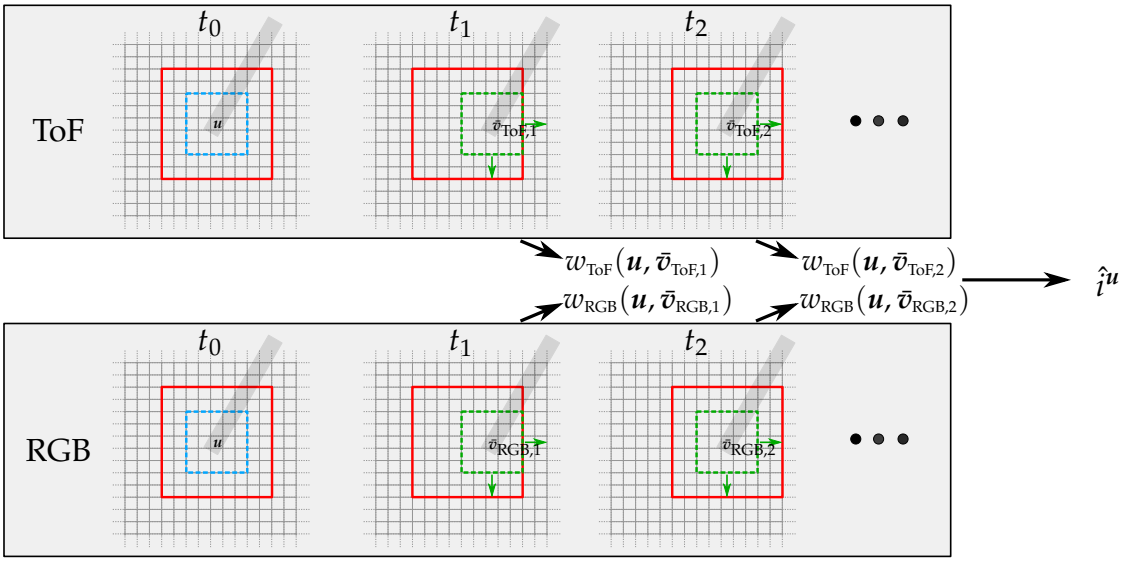


Figure 5.3: Workflow of the temporal hybrid NLM filter. Red denotes the search window, blue the neighborhood of the $\boldsymbol{u}$ and green the neighborhood of other pixels within the search window. In time step t_1 the tool has moved to a new position within the current search window of t_0. The search window for the upcoming time step t_2 is now shifted to be centered around the most similar pixel $\bar{v}_1$. If the tool moves outside the search window of t_0 this shifts keep the relevant pixels available for denoising $i^{\boldsymbol{u}}$.

entire second. Extending the search window by another dimension also leads to a tremendous computational effort. To address the motion issue, we adopt the position of the search window for a particular pixel for each frame by the relative offset of the most similar pixel in the previous frame. This implicitly tracks the motion for each scene point in the temporal domain. In a usual scenario most scene points will be seen in most of the frames within the sequence. This leads to the feasibility to rather denoise a particular pixel by its representation in other frames than by similar but different pixels within one single image. As the most similar pixel in the color domain and the ToF domain might be at different pixel positions at a single time step in both domains, we have to introduce a consolidation step to merge both results. Therefore, w_{Hyb} in Eq. (5.4) is substituted by Eq. (5.5) and expanded. Furthermore, the weighted average is not computed across the search window ω, but across an image sequence of T frames, where t_0 is the current frame and t_n are the previous frames. The denoised range value is computed by:

$$\hat{i}^{\boldsymbol{u}}_{\text{ToF}} = \sum_{t=1}^{T} \alpha \frac{1}{k^{\boldsymbol{u}}_{\text{ToF}}} w_{\text{ToF}}(\boldsymbol{u}, \bar{\boldsymbol{v}}_{\text{ToF,t}}) i^{\bar{\boldsymbol{v}}_{\text{ToF,t}}}_{ToF} + (1-\alpha) \frac{1}{k^{\boldsymbol{u}}_{\text{RGB}}} w_{\text{RGB}}(\boldsymbol{u}, \bar{\boldsymbol{v}}_{\text{RGB,t}}) i^{\bar{\boldsymbol{v}}_{\text{RGB,t}}}_{ToF}, \tag{5.7}$$

where $\bar{v}$ denotes the point v with the most similar neighborhood ω' within the search window ω at time step t compared to ω' around $\boldsymbol{u}$ at time step t_0. This point should describe the same point at different time steps and thereby allows temporal averaging without the issue of motion artifacts. Even if a point appears for the first time, it will at least be denoised by averaging points in a similar structure. The

h	0.001	0.33	0.66	1.0
err	7.84	7.03	7.38	7.53
α	0.00	0.33	0.66	1.0
err	7.84	7.27	6.98	6.89
σ	0.001	1.00	2.00	3.00
err	7.24	7.07	6.90	6.90
ω	11×11	13×13	15×15	17×17
err	6.89	6.90	6.92	6.96
ω'	5×5	7×7	9×9	11×11
err	7.04	6.94	6.90	6.89

Table 5.1: Influence of the different parameters. For each parameter test all other parameters were kept fixed at the optimal output of the grid search. These numbers are only a subset of the grid search results to show that the parameters only have little influence on the output if not set to extreme values. The error is given in mm describing the mean absolute differences of the intensities in the filter output and the ground truth data.

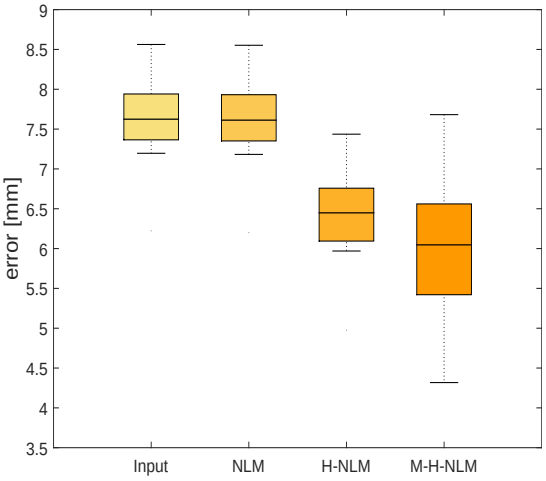
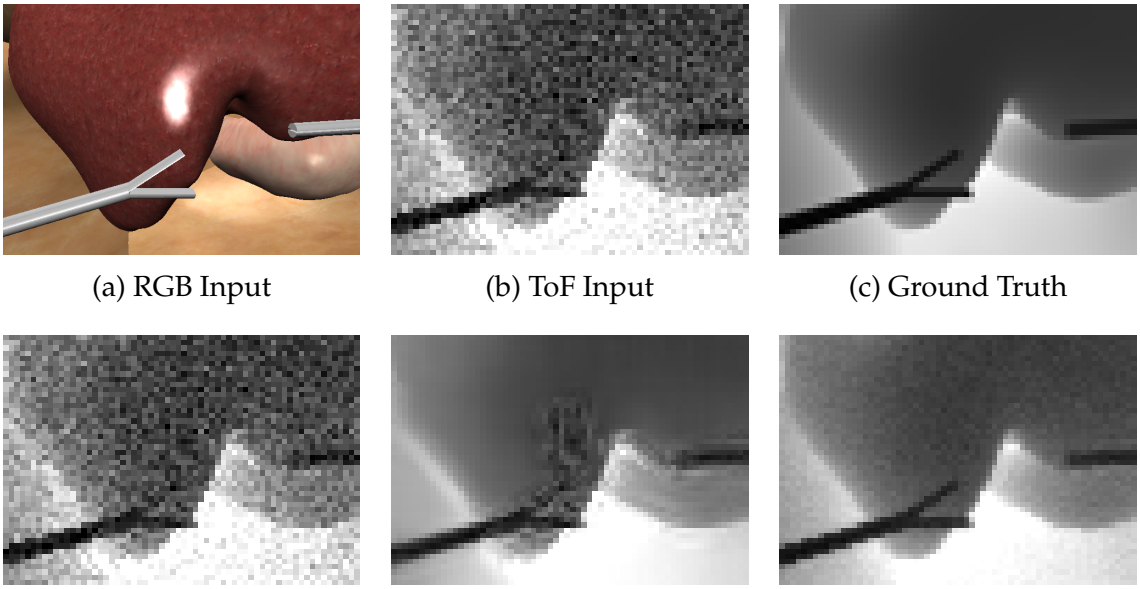


Figure 5.4: Boxplot of 10 evaluation sequences. It shows the mean absolute difference of the intensities between the ground truth data and the input data (Input), the output of the conventional NLM filter, the output of the hybrid nonlocal means (H-NLM) filter and the output of the multi-frame hybrid nonlocal means (M-H-NLM).

weightings w_{RGB} and w_{ToF} are calculated according to Eq. (5.2) and Eq. (5.6). Applying this technique in a sliding window allows denoising in a continuous way. In contrary to the single-frame approaches described in Section 5.1 and Section 5.2 in the temporal extension the size of the search window ω can be set to a rather low value as it only needs to cover the shift of pixels within two time steps.

5.4 Evaluation and Discussion

For evaluating the different variations of the NLM filter on ToF range data, we divided our experiments into two groups. For qualitative evaluation on real data we acquired ToF and RGB data of a porcine liver in a box that imitated the abdominal cavity with the hybrid 3-D endoscope introduced in Section 3.6. For qualitative and quantitative evaluations in a controlled environment we created 11 scenarios with the range image simulator described in Section 3.6 and applied all three techniques to denoise the simulated low quality ToF range images. The synthetic range image results were then compared to ground truth input data to calculate absolute errors. Random motion of the camera was used to simulate movements of the endoscope held by a surgeon for the sequence required in the multi-frame hybrid NLM filter. To find optimal parameters, a grid search was applied considering all parameters on the first simulated dataset. The first dataset was then excluded from further evaluations. The grid search was performed on the original NLM filter and on the hybrid NLM as the best parameters might change consider-



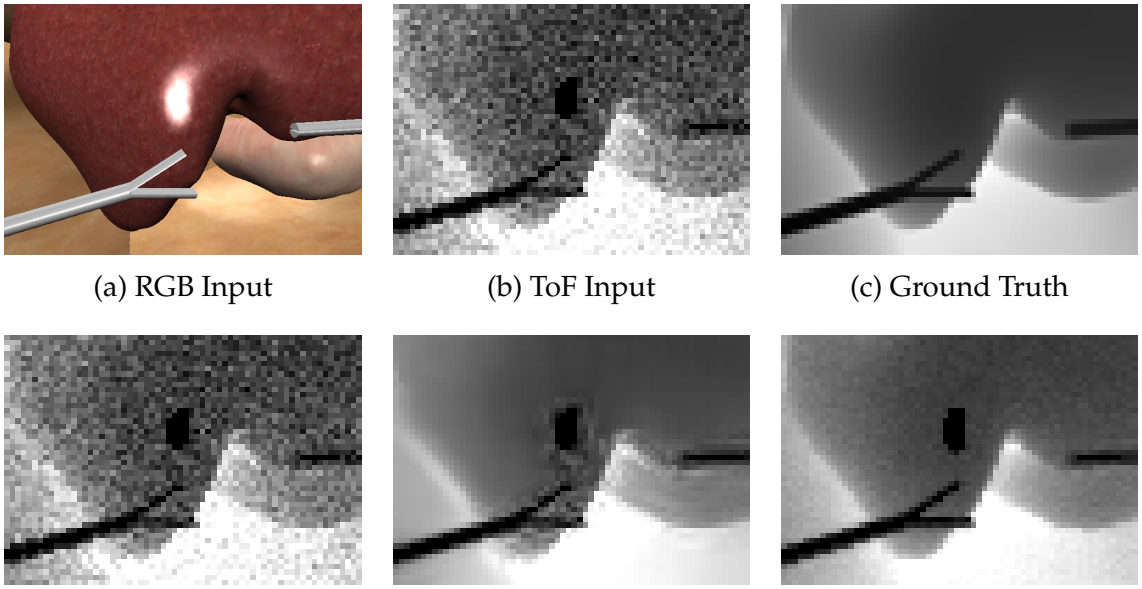
(a) RGB Input (b) ToF Input (c) Ground Truth

(d) NLM (e) H-NLM (f) M-H-NLM

Figure 5.5: A synthetic scene (S2) without defect range data induced by specular reflections. The first row shows the color input data, the range input data and the ground truth data. The second row illustrates the output of the original NLM filter (d), the output of the single-frame hybrid NLM filter (e) and the output of the multi-frame hybrid NLM filter (f).

ing similarity calculations in the color domain. The obtained optimal parameters were then kept fixed for all other datasets. To show the influence of each parameter in the hybrid and the multi-frame hybrid NLM filter Table 5.1 shows a subset of the grid search results for different parameter combinations. The relevant parameter varies within a realistic interval and the other parameters were kept fixed at their optimal value. The optimal values for the hybrid approaches are $h = 0.2$, $\alpha = 0$, $\sigma = 2.1$. The optimal size of the search window ω is 11×11 and of the neighborhood ω' is 11×11. Note that setting the reliability of the range sensor to $\alpha > 0$ increases the mean error. This shows that the current quality of the input range images is not sufficient for detecting similar neighborhoods. But with increasing progress in sensor development α might increase. Considering Eq. (5.1) to Eq. (5.7) show that in comparison to most local filters, e.g. the bilateral filter, the quantity of parameters in the NLM filter hardens the optimization for different applications. However, for our medical application the parameters have shown to be easy to adjust due to their rather small influence on the mean absolute error.

Considering 10 different simulated scenarios the mean of all absolute errors when compared to ground truth data was 7.58 mm for the raw input data. Applying the conventional NLM filter reduced the error only to 7.56 mm due to the low SNR of the ToF range data. The hybrid approach reduced further it to 6.39 mm and with the extension into the temporal domain the mean of all absolute errors was 6.06 mm. In Fig. 5.4 we also evaluated the output by comparing the original NLM filter, the hybrid and the multi-frame hybrid approach to the raw input data. Here,

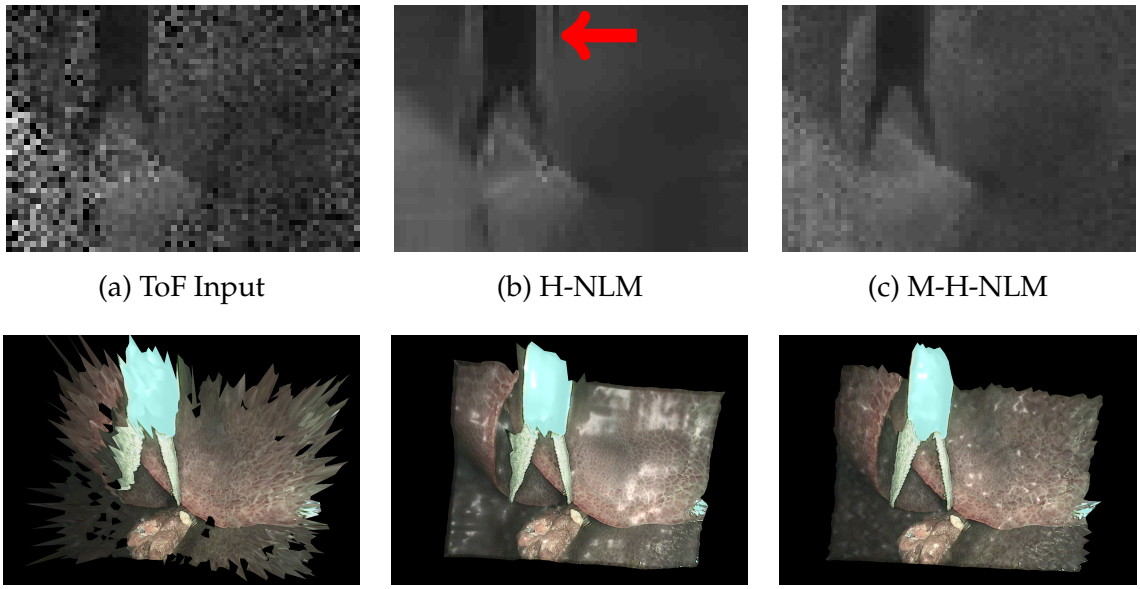


(a) RGB Input (b) ToF Input (c) Ground Truth

(d) NLM (e) H-NLM (f) M-H-NLM

Figure 5.6: A synthetic scene (S2) with defect range data induced by specular reflections. The first row shows the color input data, the range input data and the ground truth data. The second row illustrates the output of the original NLM filter (d), the output of the single-frame hybrid NLM filter (e) and the output of the multi-frame hybrid NLM filter (f).

the mean absolute errors of all 10 evaluation datasets were taken into account to compute boxplots for all approaches. Note that the multi-frame hybrid NLM filter achieved the lowest error across all datasets. Fig. 5.5 and Fig. 5.6 show one dataset for qualitative comparison. Although, the hybrid NLM results seem smoother, this technique also provokes more artifacts and copies the specular highlight visible in the color domain into the range domain. This issue is most notably in the dataset, where the specular highlight is only visible in the color domain. Note that pixels with defect range data induced by specular reflections are tracked and thereby will stay defect in all NLM filter variations. For qualitative evaluation on real data we reconstructed two datasets of porcine organs measured by our ToF/RGB endoscope prototype. One scene was composed with endoscopic instruments and one showed the porcine organs only, see Fig. 5.7 and Fig. 5.8. Note that smoothness was better when applying the hybrid NLM filter, but the temporal approach exhibits less artifacts induced by copying texture information into the final range image, e.g. shadow of the tool in the first sequence and smoothness between the background and an organ in the second scenario. We used a sequence length of 30 frames to have a realistic acquisition time of one second. The conventional and hybrid NLM filter were able to have an increased search window which results in a huge amount of pixels considered for denoising. Increasing the amount of pixels in the temporal approach would mean to extend the sequence. Therefore, please consider that a direct comparison of the hybrid NLM and the temporal hy-



(a) ToF Input (b) H-NLM (c) M-H-NLM

(d) ToF Input 3-D (e) H-NLM 3-D (f) M-H-NLM 3-D

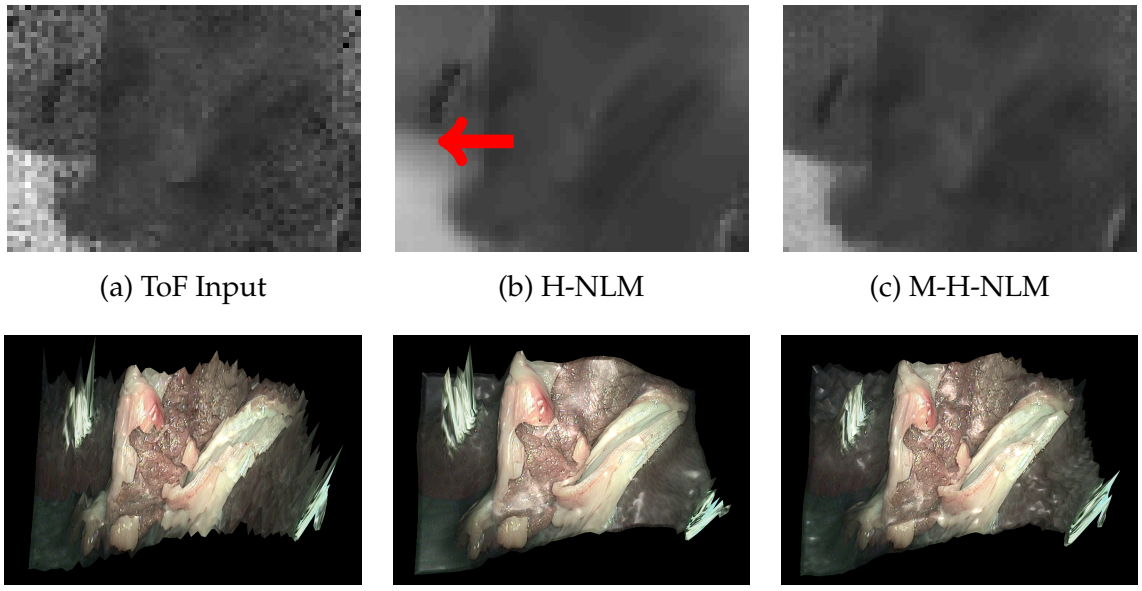
Figure 5.7: Real data results of the hybrid and multi-frame hybrid NLM filter. (a) and (d) show the raw input range images, (b) and (e) show the output of the single-frame hybrid NLM filter and (c) and (f) show the output of the multi-frame NLM approach. The red arrow marks a texture issue of the single-frame hybrid approach, i.e. shadows are copied into the range domain.

brid NLM is only marginally feasibly as the latter one is a multi-frame approach in contrary to the single-frame hybrid NLM.

5.5 Conclusion and Future Work

This chapter has described a hybrid nonlocal filtering technique for denoising low quality ToF range images using the example of the NLM filter. Instead of analyzing neighborhood similarities in the range domain, we have measured the similarity in the high quality color images and used this information to increase the SNR of the range data. Furthermore, a multi-frame extension was introduced using pixel tracking that allowed averaging of image sequences without having to deal with motion artifacts. Both NLM variations showed improved data quality in comparison to noisy input data and to the naive NLM filter.

Regarding the NLM filter for range images, future work should evaluate if combining the multi-frame approach with the conventional single-frame NLM filter could build an improved denoising technique. Furthermore, different sensors should be evaluated, where the range sensor reliability can be taken into account. Nonlocal filters in general should be compared for ToF and hybrid ToF setups. Here, the BM3D [Dani 12] is of great interest as it combines the correlations of different neighborhoods known from NLM filters with the correlations within a neighborhood known from wavelet shrinkage [Bals 05]. Nevertheless, for all hy-



(a) ToF Input (b) H-NLM (c) M-H-NLM

(d) ToF Input 3-D (e) H-NLM 3-D (f) M-H-NLM 3-D

Figure 5.8: Real data results of the hybrid and multi-frame hybrid NLM filter. (a) and (d) show the raw input range images, (b) and (e) show the output of the single-frame hybrid NLM filter and (c) and (f) show the output of the multi-frame NLM approach. The red arrow marks a texture issues of the single-frame hybrid approach, i.e. smoothness in the color domain is transferred into the range domain, where an organ and the background should be strictly differentiated.

brid preprocessing it is highly recommended to analyze the correlation of both sensor data beforehand. Often color and range data exhibit similar structures, similar edges and smooth regions. However, specular highlights, strong vessels or blood flow might create distinctive photometric features that should not be transferred into the range image. Therefore, it is advisable to inspect corresponding patches of color and range data by a data independent similarity measurement, e.g. the normalized mutual information [Stud 99]. Then it is possible to adapt the sensor reliability based on the correlation of both datasets. This allows to take high quality color images into account, where similar structures are visible, and to prevent copying incorrect photometric structures.

CHAPTER 6

Hybrid Super-Resolution

This chapter addresses the low SNR of ToF range data mentioned in Section 3.5 and improves the low image resolution as described in Section 3.5, simultaneously. A joint framework to increase ToF range image resolution and increase the data quality has been introduced by Park et al. [Park 11]. Although, their results were quite convincing, this approach was composed by several separate steps including color data segmentation and a NLM term. Therefore, we decided to solve both issues in a joint super-resolution framework. Multi-frame super-resolution methods recover a HR image from a sequence of LR frames with known subpixel displacements [Park 03]. Compared to single image upsampling, such techniques also increase the SNR and preserve edges essential for noisy range data. Approaches for color images were also adopted to ToF imaging [Schu 09]. For super-resolution in general and ToF range data in particular, accurate estimation of subpixel displacements is a challenging task to be solved before applying any super-resolution technique [Fars 04]. In literature, several robust methods were proposed [Tipp 03, Fran 07]. Here, super-resolution and motion estimation are formulated as joint optimization which is computationally demanding [Fran 07] or restricted to simplified motion models such as rigid motion [Tipp 03] being an invalid assumption for the medical applications considered in this work.

In this chapter a novel super-resolution framework for range data acquired in the multi-sensor setup described in Section 3.6 is introduced. Movements of the 3-D endoscope held by the surgeon are used as a cue for super-resolution. Our approach is based on sensor fusion of complementary RGB and range data, which is described in Chapter 4. Motion is estimated by computing optical flow on RGB data to obtain accurate displacements for range images as published in [Kohl 13]. This novelty of our method enables robust motion estimation without computationally demanding joint optimization whereas optical flow avoids restrictions of simplified models essential for realistic laparoscopic scenes.

6.1 Maximum A-Posteriori Framework

For this chapter a single LR image $\boldsymbol{i}$ with intensities i at time step t from a sequence of T frames is denoted as vector $\boldsymbol{i}_t \in \mathbb{R}^M$ with $M = M_1 \cdot M_2$ by concatenating all pixels. Each $\boldsymbol{i}_t$ is related to a reference frame $\boldsymbol{i}_{\text{ref}}$ by a geometric transformation modeling 3-D displacements. The goal of super-resolution is to determine an HR range image $\boldsymbol{i}^{\text{HR}} \in \mathbb{R}^N$, $N = s^2 \cdot M$ from T LR frames in Sequence Q for the magnification factor $s \in \mathbb{R}^{N^+}$.

As published in [Park 03], the MAP framework is based on a generative image model, which describes the image acquisition in mathematical terms. To recover an HR image, super-resolution is implemented by energy minimization based on this generative model.

6.1.1 Generative Image Model

To obtain the HR image $\boldsymbol{i}^{\text{HR}}$, the relation to each LR frame $\boldsymbol{i}_t$ at time step t is described by the generative image model according to:

$$\boldsymbol{i}_t = \boldsymbol{W}_t \boldsymbol{i}^{\text{HR}} + \epsilon_t. \tag{6.1}$$

The system matrix $\boldsymbol{W}_t$ covers the following three deformation. First, the geometric displacements between $\boldsymbol{i}^{\text{HR}}$ and $\boldsymbol{i}_t$ is modeled. Second, blur induced by the camera point spread function (PSF) is described. Third, downsampling of the HR image to its LR representation is covered. Additionally, spatially invariant noise is modeled by $\epsilon_t \in \mathbb{R}^M$. For a space invariant Gaussian PSF of width σ, the matrix elements are obtained by:

$$\boldsymbol{W}_{mn} = \exp\left(-\frac{||\boldsymbol{v}_n - \boldsymbol{u}'_m||_2^2}{2\sigma^2}\right), \tag{6.2}$$

where $\boldsymbol{v}_n \in \mathbb{R}^2$ are the coordinates of the n^{th} pixel in $\boldsymbol{i}^{\text{HR}}$ and $\boldsymbol{u}'_m \in \mathbb{R}^2$ are the coordinates of the m^{th} pixel in $\boldsymbol{i}_t$ mapped to the HR grid [Tipp 03]. To be efficient in terms of memory management, we truncate W_{mn} for $||\boldsymbol{v}_n - \boldsymbol{u}'_m||_2^2 > 3\sigma$. The algorithm shown in Section 6.2 gives further details on the parameterization of this model for range image super-resolution.

6.1.2 MAP Estimator

For the MAP estimation $\hat{\boldsymbol{i}}^{\text{HR}}$ of the HR image $\boldsymbol{i}^{\text{HR}}$, a data term and a regularizer weighted by $\lambda > 0$ is required:

$$\hat{\boldsymbol{i}}^{\text{HR}} = \arg\min_{\boldsymbol{i}^{\text{HR}}} \left(\sum_{t=1}^{T} ||\boldsymbol{i}_t - \boldsymbol{W}_t \boldsymbol{i}^{\text{HR}}||_2^2 + \lambda \sum_{n=1}^{N} h_\tau \left((\boldsymbol{D}\boldsymbol{i}^{\text{HR}})_n \right) \right), \tag{6.3}$$

where $\boldsymbol{D}$ is a high-pass filter and h_τ is the pseudo Huber loss function [Hube 64] used for regularization and described by:

$$h_\tau(z) = \tau^2 (\sqrt{1 + (z/\tau)^2} - 1). \tag{6.4}$$

Algorithm 6.1 hybrid super-resolution (HSR)

Input: Range data $\boldsymbol{i}_{ToF,t} \in Q$, RGB data $\boldsymbol{i}_{RGB,t}$, reference frame with ref $= \lceil T/2 \rceil$
Output: Super-resolved range image $\hat{\boldsymbol{i}}^{\mathrm{HR}}_{\mathrm{ToF}}$
for $t = 1 \ldots T$ **do**
$\boldsymbol{i}_{\mathrm{RGB,t}} := \mathrm{Fuse}(\boldsymbol{i}_{\mathrm{ToF,t}}, \boldsymbol{i}_{\mathrm{RGB,t}})$ ▷ see Chapter 4
$\boldsymbol{w}_{\mathrm{RGB}}(\boldsymbol{u}_{\mathrm{RGB}}) := \mathrm{OpticalFlow}(\boldsymbol{i}_{\mathrm{RGB,t}}, \boldsymbol{i}_{\mathrm{RGB,ref}})$
$\boldsymbol{w}_{\mathrm{ToF}}(\boldsymbol{u}_{\mathrm{ToF}}) := \Delta\big(\begin{pmatrix} l_1 \cdot w_{\mathrm{RGB},1}(\boldsymbol{u}_{\mathrm{RGB}}) & l_2 \cdot w_{\mathrm{RGB},2}(\boldsymbol{u}_{\mathrm{RGB}}) \end{pmatrix}^\top\big)$ ▷ see Eq. (6.6)
$\boldsymbol{W}_t := \mathrm{ComposeSystemMatrix}(\boldsymbol{w}_{\mathrm{ToF}}(\boldsymbol{u}_{\mathrm{ToF}}))$ ▷ see Eq. (6.2)
$\gamma_t^m, \gamma_t^a := \mathrm{MSAC}(\boldsymbol{i}_{\mathrm{ToF,ref}}, \mathrm{Warp}(\boldsymbol{i}_{\mathrm{ToF,t}}, \boldsymbol{w}_{\mathrm{ToF}}(\boldsymbol{u}_{\mathrm{ToF}})))$ ▷ see Section 6.2.2
$\boldsymbol{i}^{\mathrm{HR}}_{\mathrm{ToF},0} := \mathrm{BicubicUpsampling}(\boldsymbol{i}_{\mathrm{ToF,ref}})$ ▷ initial guess
$\hat{\boldsymbol{i}}^{\mathrm{HR}}_{\mathrm{ToF}} := \mathrm{SCG}(\boldsymbol{i}^{\mathrm{HR}}_{\mathrm{ToF},0}, \{\boldsymbol{i}_{\mathrm{ToF,t}}\}, \{\boldsymbol{W}_t\}, \{\gamma_t^m, \gamma_t^a\})$ ▷ see Eq. (6.3)

To enforce smoothness for the HR image $\boldsymbol{i}^{\mathrm{HR}}$, $\boldsymbol{D}$ is chosen to ba a Laplacian. Since the regularizer is based on the Huber function, it penalizes outliers less strictly and thereby preserves edges more reliable compared to a Tikhonov regularization [Tikh 77] using the L_2 norm.

6.2 Hybrid Super-Resolution

As introduced in Section 6.1 for intensity images, for the hybrid ToF/RGB setup we denote $\boldsymbol{i}_{\mathrm{RGB}}$ and $\boldsymbol{i}_{\mathrm{ToF}}$ as vectors of concatenated color and range pixels, respectively. In our framework, each $\boldsymbol{i}_{\mathrm{RGB,t}}$ is aligned to $\boldsymbol{i}_{\mathrm{ToF,t}}$ after sensor fusion as described in Chapter 4. Motion estimation is performed on the sequence of color images by employing optical flow and the displacement fields are projected to the range image domain to compose all system matrices $\boldsymbol{W}_t$. We obtain $\hat{\boldsymbol{i}}^{\mathrm{HR}}_{\mathrm{ToF}}$ by minimizing the refined version of Eq. (6.3) as described in Eq. (6.8) using scaled conjugate gradients (SCG) optimization [Nabn 02] with a bicubic upsampled version of reference frame $\boldsymbol{i}_{\mathrm{ToF,ref}}$ coincident with $\hat{\boldsymbol{i}}^{\mathrm{HR}}_{\mathrm{ToF}}$ as initial guess. See Algorithm 6.1 for further details on our approach.

6.2.1 Range Image Registration

Based on optical flow, we determine displacement vector fields $\boldsymbol{w}_{\mathrm{RGB}} : \Omega_{\mathrm{RGB}} \rightarrow \mathbb{R}^2$ between a reference frame $\boldsymbol{i}_{\mathrm{RGB,ref}}$ and every other frame of the image sequence Q. For each pixel $\boldsymbol{u}_{\mathrm{RGB}}$ the displacement vector $\boldsymbol{w}_{\mathrm{RGB}}(\boldsymbol{u}_{\mathrm{RGB}})$ is given by:

$$\boldsymbol{w}_{\mathrm{RGB}}(\boldsymbol{u}_{\mathrm{RGB}}) = (w_{\mathrm{RGB},1}(\boldsymbol{u}_{\mathrm{RGB}}), w_{\mathrm{RGB},2}(\boldsymbol{u}_{\mathrm{RGB}}))^\top. \tag{6.5}$$

This transforms each point $\boldsymbol{u}_{\mathrm{RGB}}$ from a color image $\boldsymbol{i}_{\mathrm{RGB,t}}$ of time step t to its position $\boldsymbol{u}'_{\mathrm{RGB}}$ in the reference image $\boldsymbol{i}_{\mathrm{RGB,ref}}$ according to $\boldsymbol{u}'_{\mathrm{RGB}} = \boldsymbol{u}_{\mathrm{RGB}} + \boldsymbol{w}_{\mathrm{RGB}}(\boldsymbol{u}_{\mathrm{RGB}})$. The central frame $\boldsymbol{i}_{\mathrm{RGB,ref}}$ with ref $= \lceil T/2 \rceil$ is chosen as reference to minimize the expected displacements between $\boldsymbol{i}_{\mathrm{RGB,ref}}$ and $\boldsymbol{i}_{\mathrm{RGB,t}}$ for robust flow estimation. Optical flow is computed in a course-to-fine manner using the method proposed by Liu [Liu 09]. After estimating a displacement vector field $\boldsymbol{w}_{\mathrm{RGB}}$ in the color domain,

it is transformed into the range domain to obtain the range displacement vector field $\boldsymbol{w}_{\text{ToF}} : \Omega_{\text{ToF}} \rightarrow \mathbb{R}^2$ by:

$$\boldsymbol{w}_{\text{ToF}}(\boldsymbol{u}_{\text{ToF}}) = \Delta \begin{pmatrix} l_1 \cdot w_{\text{RGB},1}(\boldsymbol{u}_{\text{RGB}}) & l_2 \cdot w_{\text{RGB},2}(\boldsymbol{u}_{\text{RGB}}) \end{pmatrix}^\top . \tag{6.6}$$

Here, $\Delta : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ describes the resampling operator. It is implemented as the median of corresponding displacement vectors $\boldsymbol{w}_{\text{RGB}}$ in both coordinate directions. To obtain $\boldsymbol{w}_{\text{ToF}}$ in the dimension of range data, rescaling by l_x, with $0 < l_x \leq 1$, is required, where l_x denotes the ratio of resolutions between $\boldsymbol{i}_{\text{ToF,t}}$ and $\boldsymbol{i}_{\text{RGB,t}}$. After this registration step, we follow Eq. (6.2) to compose the system matrix for each frame.

6.2.2 Range Correction

As the 3-D movements of the endoscope also include out-of-plane translations, the successive acquired range images show an offset due to the distance measurement. Former super-resolution approaches, such as [Schu 09], have neglected this effect and thereby reconstruct less reliable images in our scenarios. However, since this is comparable to the fusion of intensity images with different illumination, we adopt a photometric registration scheme to correct the range values. With the assumption that the frames are geometrically aligned by warping them according to the precomputed optical flow displacement vector field, we denote $i_{\text{ToF,ref}}$ as a range value in reference frame $\boldsymbol{i}_{\text{ToF,ref}}$ and $i_{\text{ToF,t}}$ as the corresponding range value at another time step t. This corresponding range value is then computed according to the affine model $i_{\text{ToF,t}} = \gamma^m \cdot i_{\text{ToF,Ref}} + \gamma^a$.

An M-estimator sample consensus (MSAC) [Torr 00] is utilized for robust estimation of γ^m and γ^a as suggested by Capel [Cape 04] for photometric registration. These parameters are plugged into the generative image model (6.1) for γ_t^m and γ_t^a. For the reference $\boldsymbol{i}_{\text{ToF,ref}}$ we set $\gamma_{ref}^m = 1$ and $\gamma_{ref}^a = 0$ to obtain a super-resolved image having the same measurement range as the reference frame. Including both correction parameters γ^a and γ^m in Eq. (6.1) results in:

$$\boldsymbol{i}_{\text{ToF,t}} = \gamma_t^m \boldsymbol{W}_t \boldsymbol{i}^{\text{HR}} + \gamma_t^a \mathbf{1} + \epsilon_t. \tag{6.7}$$

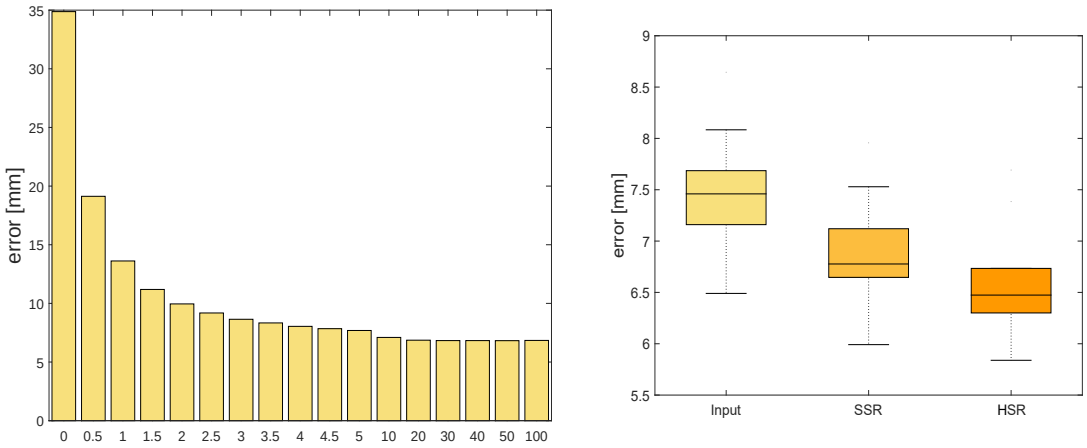
The objective function Eq. (6.3) applied on range data is then extended into:

$$\hat{\boldsymbol{i}}_{\text{ToF}}^{\text{HR}} = \arg\min_{\boldsymbol{i}_{\text{ToF}}^{\text{HR}}} \left(\sum_{t=1}^{T} ||\boldsymbol{i}_{\text{ToF,t}} - \gamma_t^m \boldsymbol{W}_t \boldsymbol{i}_{\text{ToF}}^{\text{HR}} - \gamma_t^a \mathbf{1}||_2^2 + \lambda \sum_{n=1}^{N} h_\tau \left((\boldsymbol{D} \boldsymbol{i}_{\text{ToF}}^{\text{HR}})_n \right) \right). \tag{6.8}$$

The minimization problem is then solved using an SCG optimization framework.

6.3 Evaluation and Discussion

Our HSR technique is compared to the conventional single-sensor super-resolution (SSR) approach where optical flow is estimated on range data directly. The PSF width was set to $\sigma = 0.5$ and for regularization using the Huber function we set $\lambda = 50$ and $\tau = 5 \cdot 10^{-3}$ determined empirically. SCG was used with termination



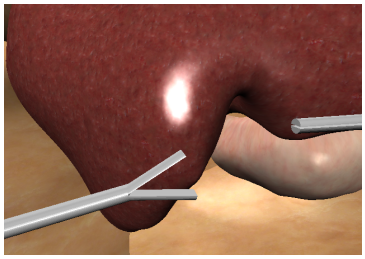
(a) (b)

Figure 6.1: (a) shows the absolute error of the entire image in dependency of the regularizer weight λ. (b) shows a boxplot of the absolute errors for all 10 evaluations sequences.

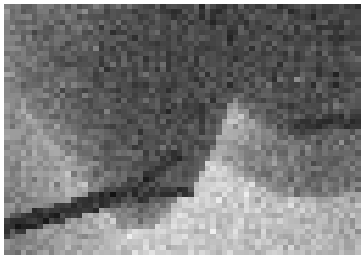
tolerance 10^{-3} for pixels of $\boldsymbol{i}_{\mathrm{ToF}}^{\mathrm{HR}}$ and the objective function value. The maximum iteration number was set to 50. Super-resolution was applied with magnification $s = 4$ using $T = 30$ frames (29 template and one reference frame).

For computing reliable and repeatable quantitative results we used the range image simulator described in Section 3.6. Each LR frame is a downsampled version of the HR ground truth data and disturbed by a Gaussian PSF and by additive, zero-mean, Gaussian noise. Random motion of the camera was used to simulate movements of the endoscope held by a surgeon. Small displacements of endoscopic tools and organs simulated minimally invasive surgery, see Section 5.4. As quality metric we employed the mean absolute error. See Fig. 6.1b for comparison of absolute error measures averaged over ten sequences. Applying conventional super-resolution the mean absolute error over all sequences was improved from 7.48 mm to 6.88 mm. With our proposed multi-sensor we reduced the error further to 6.58 mm. Fig. 6.1a evaluates the mean absolute error for the training set with different values for the regularizer weight λ. Note that any value between 20 and 100 seems to be reasonable. For qualitative evaluation on the synthetic datasets, Fig. 6.2 and Fig. 6.3 show the output of the single-sensor and the multi-sensor super-resolution as well as the input data. The organ boundaries in general and the endoscopic tools in particular are better reconstruct in the multi-sensor approach due to the improved registration.

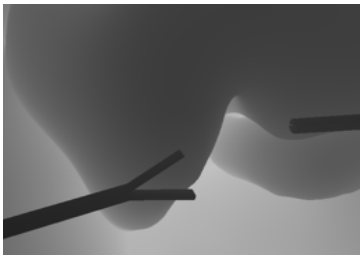
For evaluation on real datasets, we used the same datasets as in Section 5.4. The results are illustrated in Fig. 6.4 and Fig. 6.5. Note that the trade-off between smoothness and preserving edges is improved in our hybrid approach. The endoscopic tool as well as organ boundaries are reconstructed more accurately.



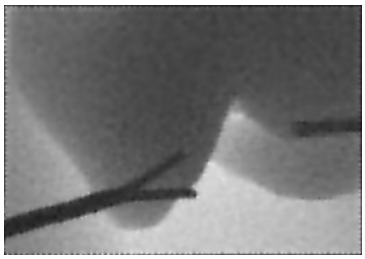
(a) RGB Input



(b) ToF Input



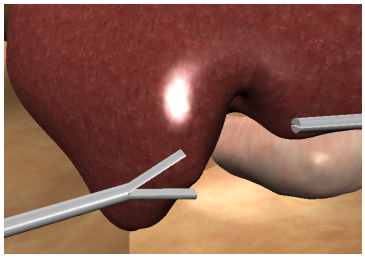
(c) Ground Truth



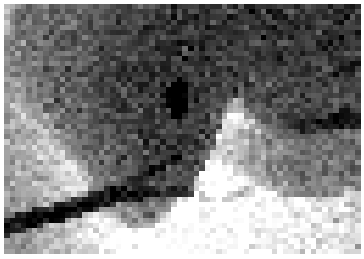
(d) SSR

(e) HSR

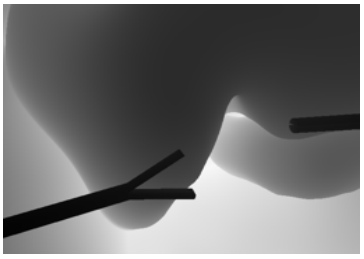
Figure 6.2: A synthetic scene with realistic lighting and texture information. (a), (b) and (c) show the color input data, the range input data and the ground truth data. (d) shows the output of the SSR and (e) shows the output of the HSR.



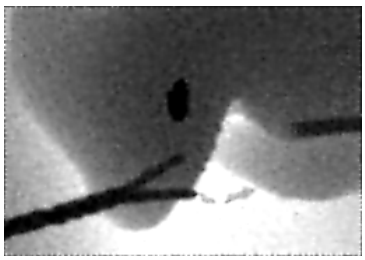
(a) RGB Input



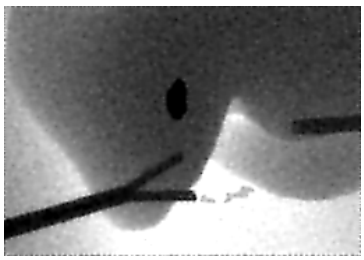
(b) ToF Input



(c) Ground Truth

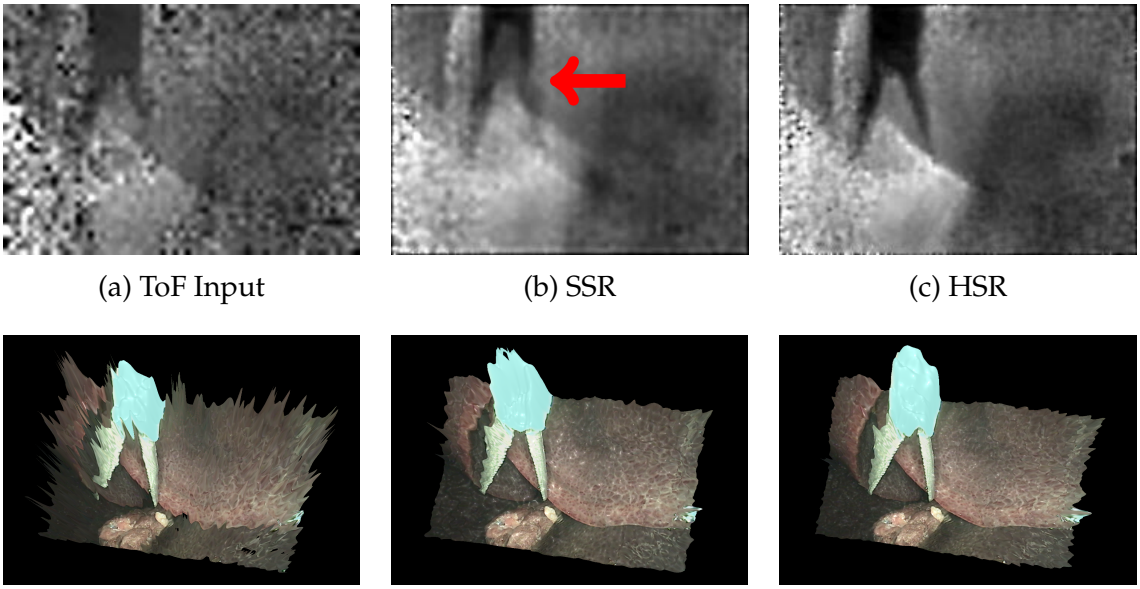


(d) SSR



(e) HSR

Figure 6.3: These images show the same data as described in Fig. 6.2, but with defect range values induced by specular highlights. Note that both approaches suffer from this issue as the specular highlights are visible in both RGB and ToF data.



(a) ToF Input (b) SSR (c) HSR

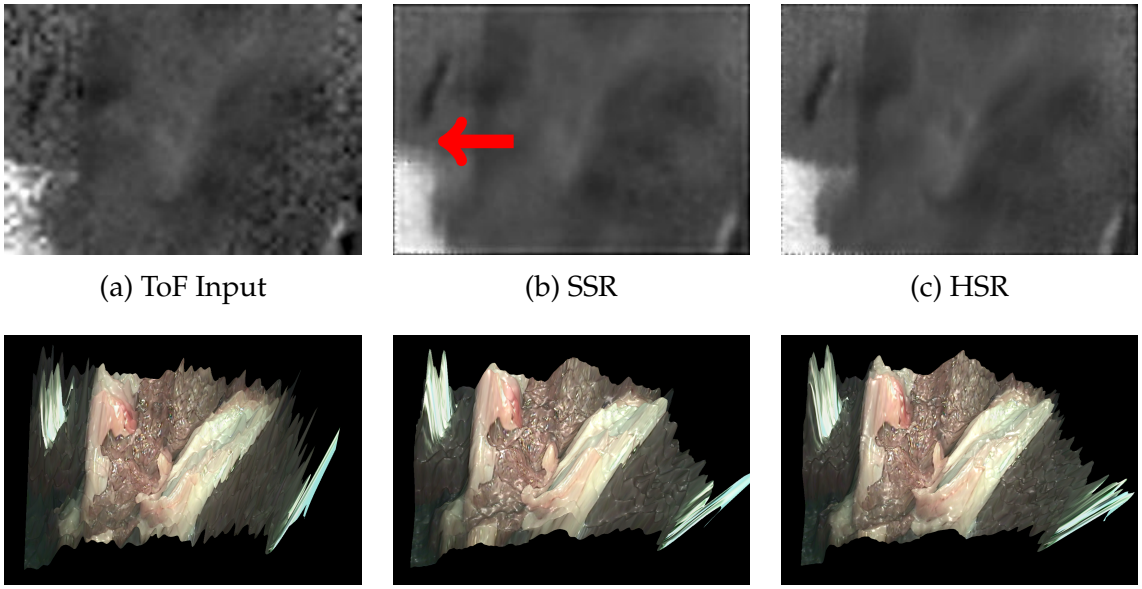
(d) ToF Input 3-D (e) SSR 3-D (f) HSR 3-D

Figure 6.4: Real data results of the single-sensor super-resolution and the multi-sensor super-resolution. (a) and (d) show the raw input range images, (b) and (e) show the output of the single-sensor super-resolution and (c) and (f) show the output of the multi-sensor super-resolution. The red arrow marks a registration issue of the single-sensor approach, which causes the endoscopic tool to be reconstructed poorly.

6.4 Conclusion and Future Work

This chapter introduced a novel super-resolution framework designed for multi-sensor setups. In our experiments we increased the SNR and the image resolution of ToF range data. The approach was evaluated on range data of realistic synthetic scenarios and real data of porcine organs acquired with a ToF/RGB endoscope. The crucial task of image registration is solved in the color domain, where high resolution and more reliable data is available. This allows a more accurate motion estimation compared to the error-prone range image registration. Due to a direct mapping of both sensor data, we used the registration output of the HR color images for the super-resolution framework applied on range data. The results showed improved image resolution and higher quality range data. Details hardly visible in the raw input range data were reconstructed in the output while reducing the noise, simultaneously.

Future work should analyze possibilities to use the HR color images for regularization in the super-resolution framework. Additional structures only visible in the color images could help to prevent smoothing across edges and keep details that are not visible in the raw low-resolution range data. Furthermore, the optical flow output needs to be verified and could be improved by a reliability term. In some cases, e.g. if blood flows over an organ, optical flow applied on the color data might result in wrong transformation estimations. Further work on



(a) ToF Input (b) SSR (c) HSR

(d) ToF Input 3-D (e) SSR 3-D (f) HSR 3-D

Figure 6.5: Real data results of the single-sensor super-resolution and the multi-sensor super-resolution. (a) and (d) show the raw input range images, (b) and (e) show the output of the single-sensor super-resolution and (c) and (f) show the output of the multi-sensor super-resolution. The red arrow marks a registration issues of the single-sensor approach, which blurs an edge induced by a organ and the background that should be strictly differentiated.

multi-sensor super-resolution was published by Köhler et al. [Koeh 14a, Koeh 14b]. Furthermore, real-time capability needs to be investigated. First results for super-resolution on the GPU were already published by Wetzl et al. [Wetz 13].

CHAPTER 7

Specular Highlight Removal

Specular highlights are common issues in laparoscopic images [Arno 10]. These areas provide no information as the majority of the light emitted by the light source is directly reflected at the surface. This saturates the sensor pixels at those areas and thereby only white intensity information is acquired. As described in Section 3.5, range sensors in general and ToF sensors in particular are equally error-prone to those effects. Specular highlights usually appear at round objects and mirror-like surfaces, as it is seen on wet organs in the abdominal cavity.

In color images specular highlights are characterized in the HSV color space by low saturation and high value intensities whereas in ToF based range images those areas show incorrect or missing measurements, see Fig. 7.4. Different approaches for specular reflection removal have been proposed [Arno 10, Grge 01, Stoy 05, Xu 10, Wasz 11a]. A method for conventional endoscopy proposed by Arnold et al. [Arno 10] uses an inpainting technique based on a normalized convolution [Knut 93]. Gröger et al. [Grge 01] reconstructed image structures by employing anisotropic diffusion. Stoyanov et al. [Stoy 05] published an approach based on temporal registration. However, as there is no evidence that replacing specular reflections in color images improves the impression for surgeons, our approach only tackles the problem of replacing invalid range information. Satisfying interpolation results for defect pixel restoration in range image were published by Wasza et al. [Wasz 11a]. However, interpolation techniques suffer from missing information in the defect regions. Therefore, Xu et al. [Xu 10] described an approach where two images from slightly different points of view are registered using SURF matching. Then, specular reflections of the first view are replaced by non specular areas in the second view based on global intensity adjustment of both images. However, this global approach suffers from misaligned edges after image registration. Especially in our medical application, the camera movement is not controllable which might result in structures reconstructed at wrong positions.

This chapter is based on [Haas 14] and describes a combination [Xu 10] and [Arno 10]. The registration in [Xu 10] is extended by a robust patch based approach and the global intensity correction is replaced by a local adjustment. Then specular reflections in a range image are replaced by non specular pixels of another

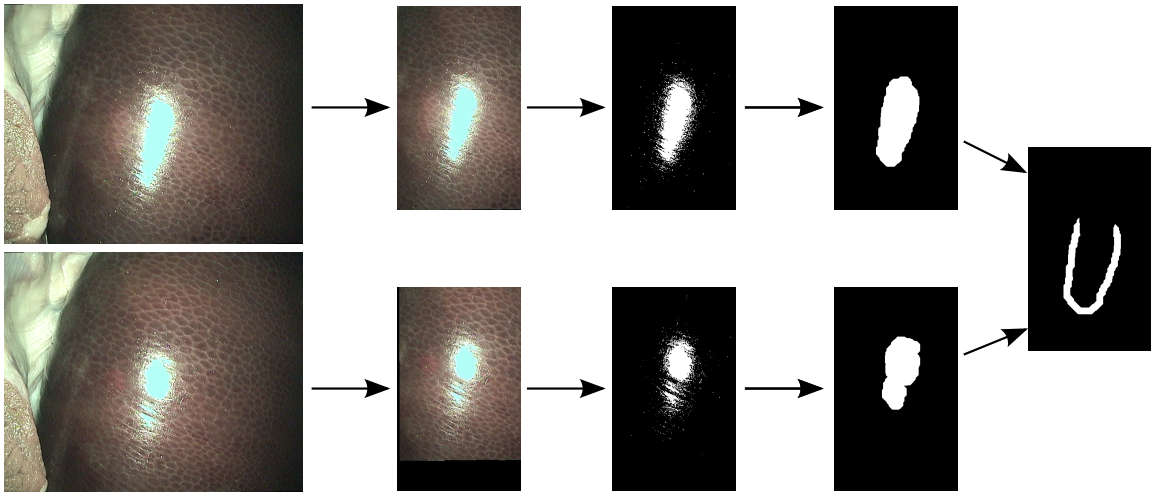


Figure 7.1: Workflow of our mask images. (a) shows the two different color input frames $\boldsymbol{i}_{\text{RGB},1}$ and $\boldsymbol{i}_{\text{RGB},2}$, (b) denotes the two highlight patches after registration, (c) shows the direct result of Eq. (7.2), (d) denotes the result after applying morphological opening and closing, (e) shows the margin for Section 7.2.2.

view. Remaining defect data will be interpolated using normalized convolution [Knut 93]. In comparison to conventional interpolation techniques, our approach is able to recover structures that are completely marked as invalid in a single range image.

7.1 Specular Highlight Detection

Specular reflections are detected in the color images by analyzing the color components in the HSV color space. Hue represents the color, which is not of further importance for our application. The saturation intensity is denoted by $i^{\boldsymbol{u}}_{\text{Sat}}$. The value intensity is denoted by $i^{\boldsymbol{u}}_{\text{Val}}$ and describes the brightness. Value and saturation intensities are given by:

$$i^{\boldsymbol{u}}_{\text{Sat}} = 1 - \frac{3}{i^{\boldsymbol{u}}_{\text{R}} + i^{\boldsymbol{u}}_{\text{G}} + i^{\boldsymbol{u}}_{\text{B}}} \min(i^{\boldsymbol{u}}_{\text{R}}, i^{\boldsymbol{u}}_{\text{G}}, i^{\boldsymbol{u}}_{\text{B}}) \qquad \text{and} \qquad i^{\boldsymbol{u}}_{\text{Val}} = \max(i^{\boldsymbol{u}}_{\text{R}}, i^{\boldsymbol{u}}_{\text{G}}, i^{\boldsymbol{u}}_{\text{B}}), \tag{7.1}$$

with $\min(i^{\boldsymbol{u}}_{\text{R}}, i^{\boldsymbol{u}}_{\text{G}}, i^{\boldsymbol{u}}_{\text{B}})$ and $\max(i^{\boldsymbol{u}}_{\text{R}}, i^{\boldsymbol{u}}_{\text{G}}, i^{\boldsymbol{u}}_{\text{B}})$ denoting the minimal and maximal intensity of the three color channels red, green and blue. Specular reflections typically show a low saturation and a high value intensity. We apply the reflection detection proposed by Zimmermann-Moreno et al. [Zimm 06] by marking all pixel positions $\boldsymbol{u}$ as specular reflections according to:

$$i^{\boldsymbol{u}}_{\text{Mask}} = \begin{cases} 1, & \text{if} \quad i^{\boldsymbol{u}}_{\text{Sat}} \leq \alpha \cdot i^{\max}_{\text{Sat}} \quad \wedge \quad i^{\boldsymbol{u}}_{\text{Val}} \geq \beta \cdot i^{\max}_{\text{Val}} \\ 0, & \text{else} \end{cases}, \tag{7.2}$$

where $i^{\max}_{\text{Sat}}$ and $i^{\max}_{\text{Val}}$ denote the maximal intensity in the saturation and the value channel, respectively. $i^{\boldsymbol{u}}_{\text{Sat}}$ and $i^{\boldsymbol{u}}_{\text{Val}}$ are given by Eq. (7.1). As this techniques leads

to fuzzy segmentation boundaries, morphological operations such as opening and closing are subsequently applied to the reflection masks to ensure that all reflections are covered completely and noise is removed, see Fig. 7.1.

7.2 Specular Highlight Removal

Similar to the NLM filter described in Chapter 5, this section describes a single-frame and a novel multi-frame approach to replace invalid information in range images corrupted by specular highlights. First, a single-frame approach is described as proposed by Arnold et al. [Arno 10] for endoscopic interventions and proposed by Wasza et al. [Wasz 11a] for medical range data. Second, a novel multi-frame approach is introduced as proposed in [Haas 14].

7.2.1 Single-Frame Defect Pixel Interpolation

Knutsson et al. [Knut 93] proposed the normalized convolution (NC) as an interpolation technique, to smear valid neighborhood information into invalid regions. An output pixel of the NC is computed by:

$$i^{\boldsymbol{u}}_{\mathrm{NC}} = \frac{\sum_{\boldsymbol{v}' \in \omega'} \exp(-\frac{||\boldsymbol{v}'||_2^2}{\sigma^2})(1 - i^{\boldsymbol{u}}_{\mathrm{Mask}}) i^{\boldsymbol{u}+\boldsymbol{v}'}}{\sum_{\boldsymbol{v}' \in \omega'} \exp(-\frac{||\boldsymbol{v}'||_2^2}{\sigma^2})(1 - i^{\boldsymbol{u}}_{\mathrm{Mask}})}, \tag{7.3}$$

where $i^{\boldsymbol{u}}_{\mathrm{Mask}}$ is calculated by Eq. (7.2) and denotes whether a pixel $\boldsymbol{u}$ is part of a specular highlight. Similar to the NLM $\boldsymbol{v}'$ describes the relative pixel offsets in the neighborhood. The final output of the normalized convolution replaces invalid pixels with the output of the NC by:

$$\hat{i}^{\boldsymbol{u}}_{\mathrm{NC}} = (1 - i^{\boldsymbol{u}}_{\mathrm{Mask}}) i^{\boldsymbol{u}} + i^{\boldsymbol{u}}_{\mathrm{Mask}} i^{\boldsymbol{u}}_{\mathrm{NC}}. \tag{7.4}$$

7.2.2 Multi-Frame Defect Area Restoration

Hybrid range imaging in minimally invasive procedures, e.g. by a 3-D endoscope, is usually embedded in a workflow, where the device is moved within the abdominal cavity. This induces that a sequence of images is acquired from different perspectives. Similar to Chapter 6, this section describes a correction technique for specular highlights, where multiple frames are fused to restore a range image.

Matching Two Range Image Patches

Endoscopic ToF images suffer from a high noise level. Therefore, a direct registration of two range images $\boldsymbol{i}_{\mathrm{ToF},1}$ and $\boldsymbol{i}_{\mathrm{ToF},2}$ is not feasible. However, as we have corresponding high quality photometric color information $\boldsymbol{i}_{\mathrm{RGB},1}$ and $\boldsymbol{i}_{\mathrm{RGB},2}$, we estimate the transformation in the RGB domain. Due to the homographic alignment between the range images and corresponding color images, a local transformation between $\boldsymbol{i}_{\mathrm{RGB},1}$ and $\boldsymbol{i}_{\mathrm{RGB},2}$ can be mapped directly to the range image domain similar to the registration process in Chapter 6. As proposed in [Xu 10] we use a feature

based registration approach as conventional optical flow techniques would try to match the specular reflections. Our technique calculates SURF in both images and detects corresponding points in both views [Bay 06]. The SURF detector is inspired by SIFT but is more robust and faster to compute by the use of integral images. The feature description is based on 2-D Haar wavelet responses. To reject erroneous matching feature points we apply the hierarchical multi-affine algorithm proposed by Puerto-Souza et al. [Puer 13]. This approach estimates different transformations by including different clusters of feature points randomly sampled by means of random sampling and consensus (RANSAC) [Fisc 81]. All clustered correspondences that result in a similar transformation are then included into the final point correspondence dataset. Based on this feature matching an affine transformation is estimated. However, an affine transformation is often only an unsatisfying estimation of the actual motion, the proposed approach improves the concept of Xu et al. [Xu 10] to feature a more robust registration. This is achieved by only estimating local transformations and by analyzing and restoring each specular highlight by a patch with an edge length three times the size of the highlight's bounding box in each dimension. These individual patches can be computed by different techniques after computing the binary mask according to Eq. (7.2), e.g. by a connected component analysis [Same 88] or clustering techniques such as k-means clustering [Hart 79].

Range Correction

To cope with different range values in $\boldsymbol{i}_{\text{ToF},1}$ and $\boldsymbol{i}_{\text{ToF},2}$ after registration due to viewpoint changes, e.g. out-of-plane movements, a correction of range values is mandatory, see Section 6.2.2. We extend the global intensity correction proposed in [Xu 10] by correcting each non specular region in $\boldsymbol{i}_{\text{ToF},2}$ locally. Here, we consider a margin of the specular reflection that is computed by dilating the reflection mask and then subtracting the original reflection mask, see Fig. 7.1. This margin contains the range data closest to the defect area that is not part of a specular highlight in both views. To achieve a smooth transition between the content of $\boldsymbol{i}_{\text{ToF},1}$ and the replaced regions of $\boldsymbol{i}_{\text{ToF},2}$ we correct each pixel in $\boldsymbol{i}_{\text{ToF},2}$ by:

$$i^{\boldsymbol{u}}_{\text{ToF},2'} = i^{\boldsymbol{u}}_{\text{ToF},2} + \frac{1}{\sum_{\boldsymbol{v}\in M} w(\boldsymbol{u},\boldsymbol{v})} \sum_{\boldsymbol{v}\in M} w(\boldsymbol{u},\boldsymbol{v})(i^{\boldsymbol{v}}_{\text{ToF},1} - i^{\boldsymbol{v}}_{\text{ToF},2}), \tag{7.5}$$

where M is the set of all margin pixels and $w(\boldsymbol{u},\boldsymbol{v})$ is a weight that takes the distance of a margin pixel at position $\boldsymbol{v}$ to the pixel position $\boldsymbol{u}$ into account. The weights are thereby derived by:

$$w(\boldsymbol{u},\boldsymbol{v}) = \exp\left(-\frac{||\boldsymbol{u}-\boldsymbol{v}||_2^2}{\sigma^2}\right), \tag{7.6}$$

where σ controls the influence of the distance as a weight.

Reflection Restoration

In a final step we replace all specular reflection in $\boldsymbol{i}_{\text{ToF},1}$ by non specular corresponding areas in $\boldsymbol{i}_{\text{ToF},2}$ based on the previously estimated transformation and on range

correction as proposed in the previous paragraphs. The restored range value $i^{\boldsymbol{u}}_{\mathrm{ToF},1'}$ at pixel position $\boldsymbol{u}$ is given by:

$$i^{\boldsymbol{u}}_{\mathrm{ToF},1'} = (1 - i^{\boldsymbol{u}}_{\mathrm{Mask},1}) i^{\boldsymbol{u}}_{\mathrm{ToF},1} + i^{\boldsymbol{u}}_{\mathrm{Mask},1} i^{\boldsymbol{u}}_{\mathrm{ToF},2'}. \tag{7.7}$$

Depending on the viewpoint this may result in remaining pixels that are marked as reflections in both images. Thereupon, a last mask image with $i^{\boldsymbol{u}}_{\mathrm{Mask},1'} = i^{\boldsymbol{u}}_{\mathrm{Mask},1} i^{\boldsymbol{u}}_{\mathrm{Mask},2}$ is calculated. This mask image serves as input to replace the remaining defect pixels with a version of $\boldsymbol{i}^{\boldsymbol{u}}_{\mathrm{ToF},1'}$, where the defect areas are interpolated by the normalized convolution, see Eq. (7.3). Therefore, the final output is given by:

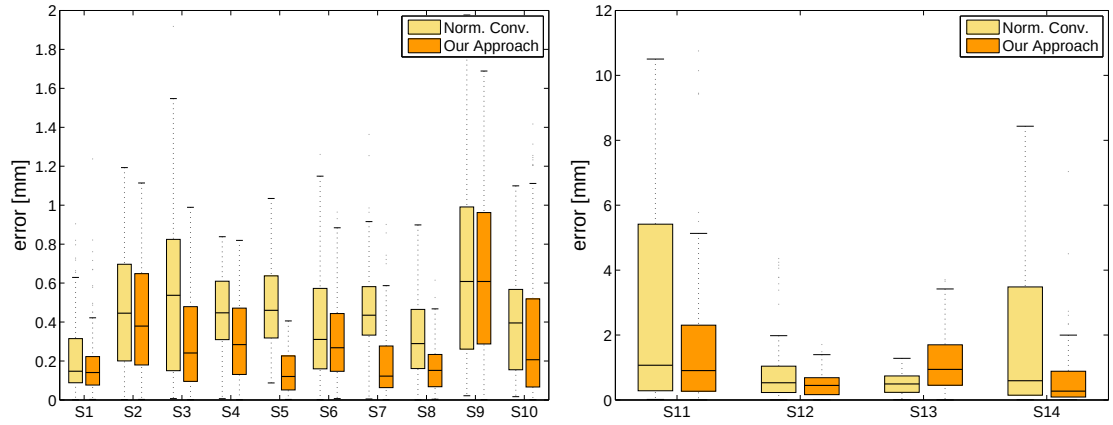
$$\hat{i}^{\boldsymbol{u}}_{\mathrm{ToF},1} = (1 - i^{\boldsymbol{u}}_{\mathrm{Mask},1'}) i^{\boldsymbol{u}}_{\mathrm{ToF},1'} + i^{\boldsymbol{u}}_{\mathrm{Mask},1'} i^{\boldsymbol{u}}_{\mathrm{ToF},1',\mathrm{NC}}, \tag{7.8}$$

where $i^{\boldsymbol{u}}_{\mathrm{ToF},1',\mathrm{NC}}$ denotes the range value obtained from the normalized convolution applied on the corrected range image $\boldsymbol{i}_{\mathrm{ToF},1'}$ obtained according to Eq. (7.7).

7.3 Evaluation and Discussion

The proposed method was evaluated on synthetic datasets acquired with the range device simulator introduced in Section 3.6. Based on ground truth data, absolute distance errors were calculated and qualitative results are given. To evaluate our approach only in terms of invalid range data, the simulator created perfect low resolution range images solely corrupted by specular highlights. For real data scenarios, we used the 3-D endoscope described in Section 3.6. However, due to the early prototype status of this device quantitative evaluations are not yet feasible on real data. In terms of patch computation, we dealt with typical simplified scenarios, where only a single specular highlight was detectable.

In Fig. 7.2 we compare the absolute errors between a conventional interpolation technique and our approach. Here, 10 common datasets and four challenging datasets, i.e. range images where important structures are covered by specular highlights, are evaluated. Note that in three of the four challenging cases our approach was able to restore the important structure, see Fig. 7.3. The higher error in dataset S13, was induced by a poor registration of both views. Overall, the error is reduced in 13 of the 14 datasets. In the common scenarios, the mean absolute error across all datasets was reduced from 0.45 mm using a normalized convolution to 0.30 mm using our approach. Regarding the mean absolute error in dataset S14 the error was reduced from 2.2 mm to 0.7 mm. Qualitative results for real data are depicted in Fig. 7.4 that shows two 3-D meshes overlaid with RGB information before and after applying an NC interpolation and our restoration technique. In both cases the spikes of the raw input data were reduced by applying the defect pixel restoration. Due to the low SNR of the acquired data of the current prototype no notable difference between both restoration techniques is observable. However, the synthetic results show that our approach will be beneficial when data quality will improve.



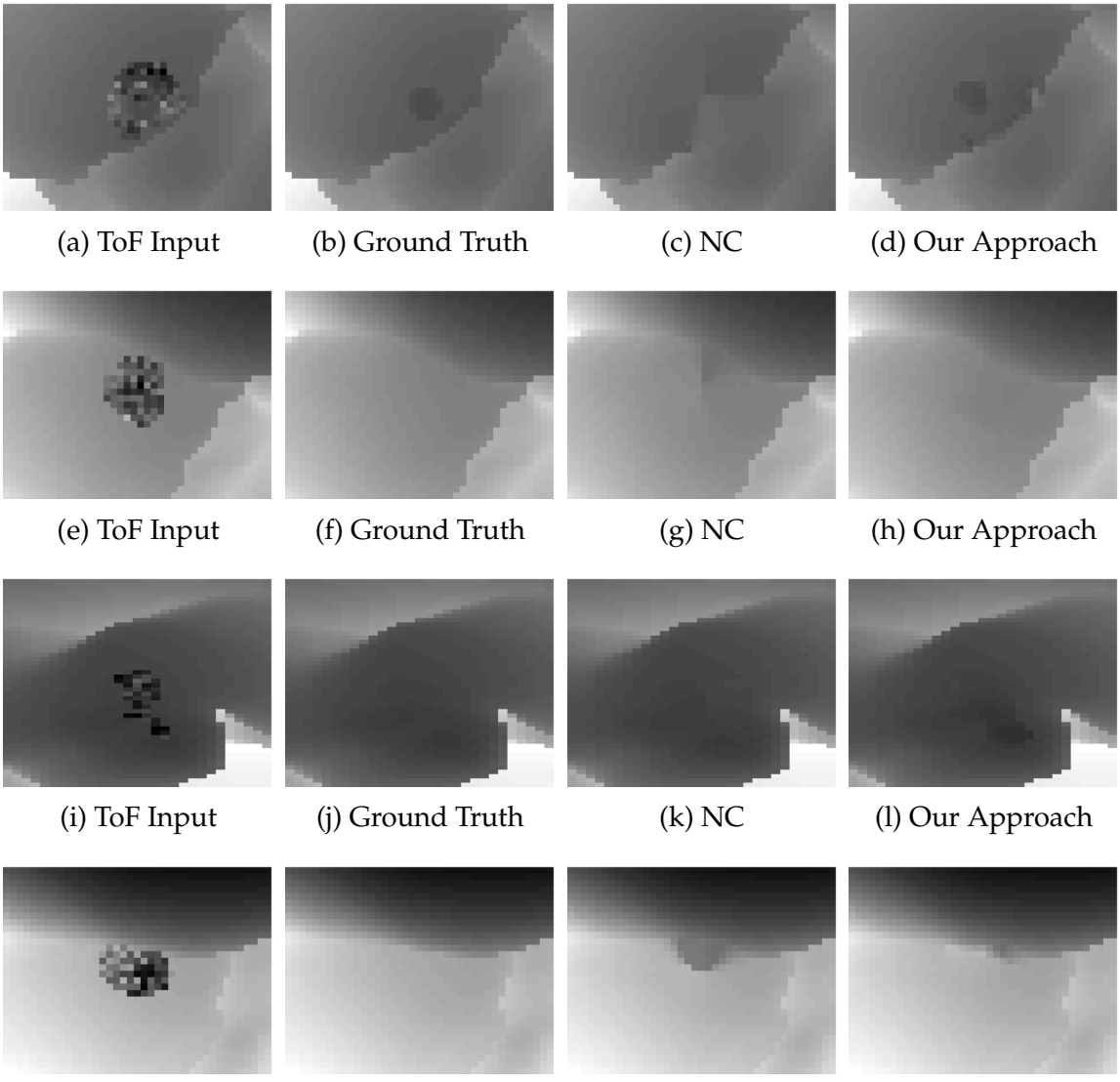
(a) Common Scenarios (b) Challenging Scenarios

Figure 7.2: Boxplots for the pixelwise absolute error in the specular region considering synthetic data: 10 conventional datasets (left) and four challenging datasets (right). Note that only in dataset S13 our approach failed due to a poor registration of the affected image patches.

7.4 Conclusion and Future Work

This chapter has presented a novel technique to restore valid range data in invalid regions caused by specular highlights. As proposed in [Haas 14] we apply a patch based registration for each highlight and fuse valid information of different viewpoints into the current view. We have shown that our approach outperforms conventional interpolation in 13 of our 14 datasets. In one dataset a poor registration led to a higher mean absolute error in comparison to an interpolation approach. In other challenging scenarios we restored important structures where conventional interpolation lost the information due to its single frame concept.

Future work has to cover an evaluation on more datasets to show benefits in other scenarios, e.g. a surgical cut in the organ that is covered by a reflection. Additionally we have to investigate the behavior in the presence of laparoscopic tools. For real-time requirements in a medical environment the framework has to be parallelized using GPGPU. In particular, the computational expensive registration can be parallelized. The individual steps of the workflow could also be replaced, e.g. the feature detection or the range correction, and should be evaluated for the best outcome.



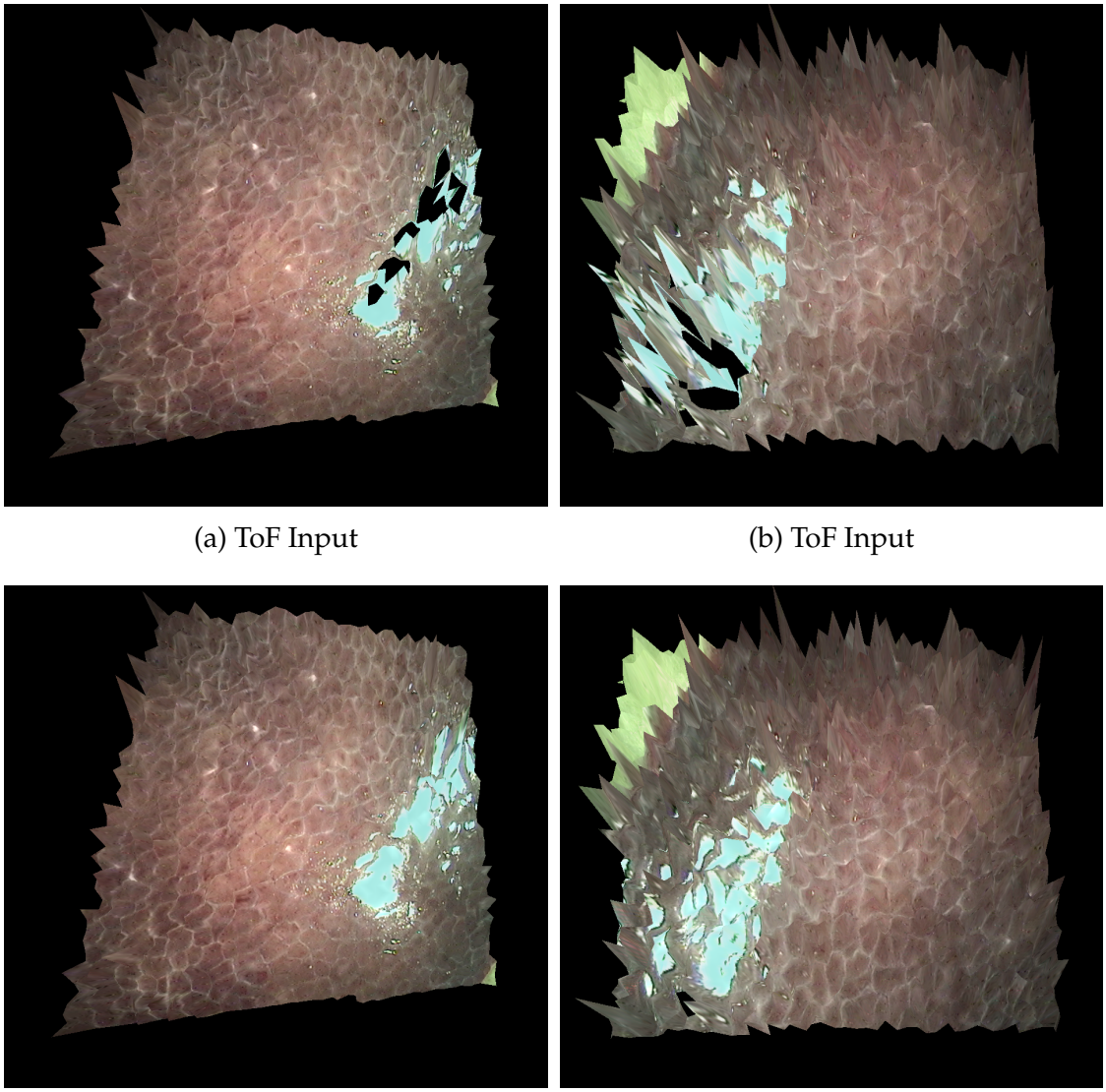
(a) ToF Input (b) Ground Truth (c) NC (d) Our Approach

(e) ToF Input (f) Ground Truth (g) NC (h) Our Approach

(i) ToF Input (j) Ground Truth (k) NC (l) Our Approach

(m) ToF Input (n) Ground Truth (o) NC (p) Our Approach

Figure 7.3: The four rows show the challenging datasets of Fig. 7.2. In the first dataset, a specular highlight covered a simulated polyp. The second and fourth dataset show highlights near organ boundaries. In the third scenario the highlight covers a rather smooth and homogeneous areas, which hardens the registration process.



(a) ToF Input (b) ToF Input

(c) Final Output (d) Final Output

Figure 7.4: Two real ex-vivo datasets. (a) and (b) show both 3-D reconstructions of the raw input data. (c) and (d) show the output after specular highlight removal. Black areas inside the surface shows regions where no valid information is given by the sensor.

Part III

Applications in Abdominal Surgery

CHAPTER 8

Collision Avoidance

After addressing the data issues of range imaging in minimally invasive procedures, this chapter describes the first application to assist surgeons in endoscopic interventions. In conventional minimally invasive procedures instruments and endoscopes are navigated by a surgeon and his team. Especially during long interventions this includes the risk of a jitter as an additional source of error and leads to unstable blurry images during the intervention. By erroneous navigation of instruments or the endoscope the surgeon may harm surrounding healthy tissue. To compensate for the issue of jitters robotic assistance systems have been proposed to allow indirect navigation by the use of joysticks [Hart 09, Aion 02, Pole 08]. Although, jitters are avoided by these systems, navigation with the joystick is even less intuitive. Hence, for direct and indirect navigation the problem of erroneous movements caused by misinterpreted images and an insufficient field of view remains. The avoidance of risk situations in minimally invasive procedures has been addressed by several research groups [Spei 08, Haas 13e]. Speidel et al. propose an approach using a stereo endoscope and a knowledge representation system [Spei 08]. The endoscopic tools are tracked in 2-D and located in 3-D. Based on the defined logic the surgeon is warned in case of any risk situations. In [Haas 13e] Haase et al. describe a 3-D tool localization algorithm based on a ToF/RGB endoscope that holds potential for avoidance of risk situations using 3-D metric information. Nevertheless, both approaches require a specific 3-D endoscope and a learning phase for interpreting the additional information.

Our approach is published in [Haas 13c] and integrates seamlessly into the current workflow without the need of expensive hardware and any further learning phase concerning new software. We propose a supervision module for any robotic endoscope holder that keeps a safety margin between the endoscope and the operation site by extending or retracting a telescope that is directly attached to the endoscope. The adjustment is based on range images acquired at high frame rates. For clinical scenarios real-time constraints are fulfilled by using state-of-the-art hardware and software optimization. This module holds the potential to improve the safety for patients and simultaneously ease the navigation for surgeons.

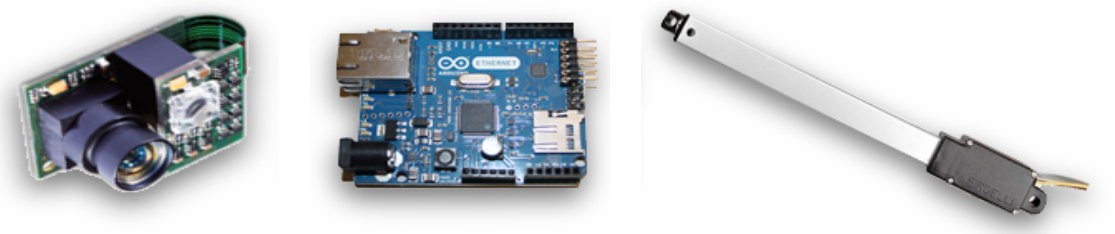


Figure 8.1: The three components of our enhancement module. From left to right: PMD CamBoard nano, Adruino Uno micro controller, L12 linear servo motor.

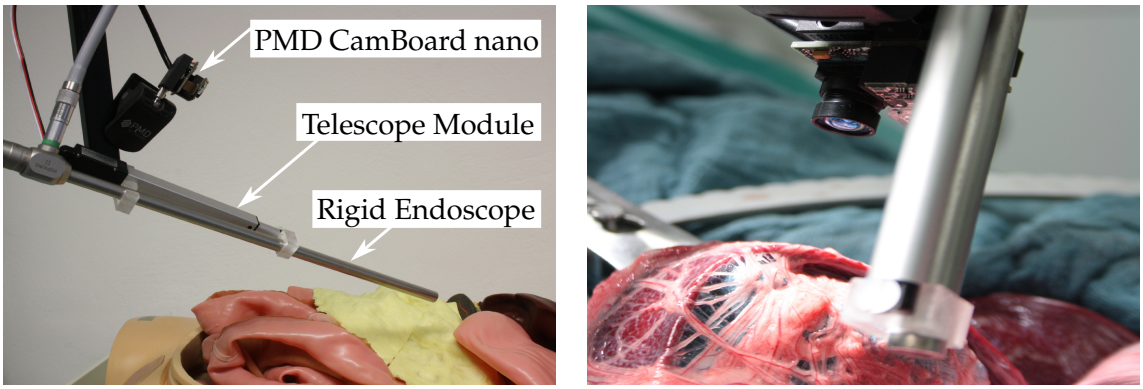


Figure 8.2: The prototype Time-of-Flight based module in a phantom study on the left and in an in-vivo study on a pig on the right. Due to the prototype status the in-vivo experiments were performed in an open surgery manner.

8.1 Time-of-Flight Guidance Module

The enhancement module is composed of three parts: The distance measuring sensor, the telescope module and a micro controller for communication. Fig. 8.1 illustrates those three components and Fig. 8.2. depicts our assembled prototype attached to a robotic endoscope holder. Though our setup is generic for different endoscope holders, for all our experiments we used the SOLOASSIST [Hart 09] robotic endoscope holder, which imitates a human arm and is navigated by small joystick that allows free movements in all three dimensions. For distance measurement a ToF camera acquires range information in real-time [Lang 01], see Section 3.3. For our experiments the CamBoard nano was chosen as a reference device as it combines an adequate resolution (160×120 px) in a small housing ($37 \times 30 \times 25$ mm). However, due to a low SNR, preprocessing range images is an essential step. To satisfy real-time constraints we use the RITK[1] [Wasz 11a] software framework to build a preprocessing pipeline on the graphics card using CUDA. As ToF sensors suffer from temporal and spatial noise, preprocessing in both domains is required. Otherwise, the noise forces the telescope to adjust the distance constantly which results in a rather instable video image. As the prepro-

[1] http://www5.cs.fau.de/research/software/ritk/

cessing techniques described in Part II are not yet real-time capable, we perform a temporal averaging on a few consecutive frames first and then apply the bilateral filter [Toma 98] for edge-preserving spatial denoising:

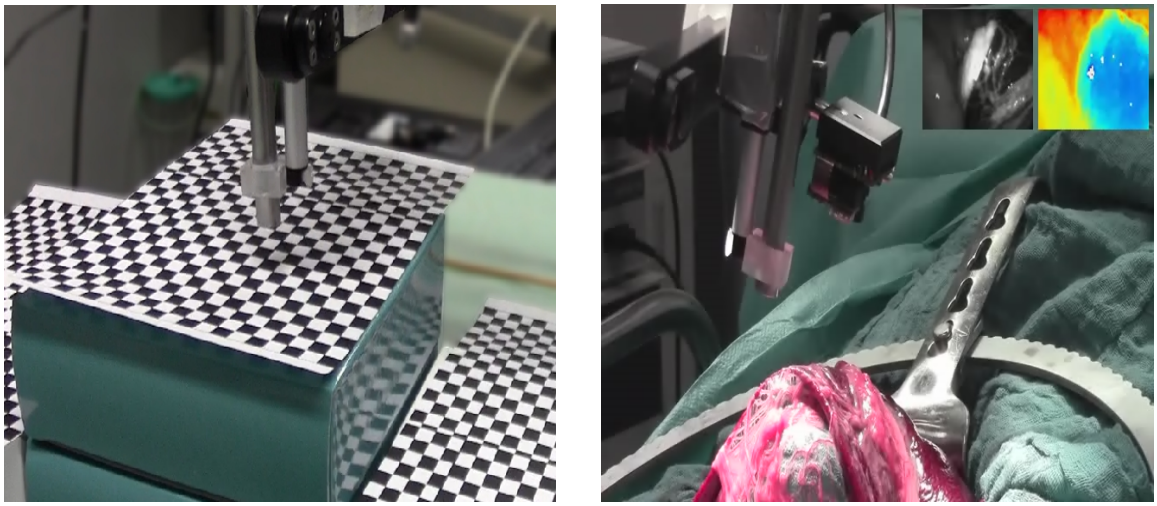
$$\hat{i}^{\boldsymbol{u}}_{\text{tmp}} = \frac{1}{T}\sum_{t=1}^{T} i^{\boldsymbol{u}}_{\text{t}} \text{ and } \hat{i}^{\boldsymbol{u}}_{\text{bil}} = \frac{1}{k}\sum_{\boldsymbol{v}'\in\omega'} \exp(-\frac{||\boldsymbol{v}'||_2^2}{\sigma_1^2}) \exp(-\frac{|i^{\boldsymbol{u}} - i^{\boldsymbol{u}+\boldsymbol{v}'}|^2}{\sigma_2^2}) i^{\boldsymbol{u}+\boldsymbol{v}'}, \quad (8.1)$$

where $\hat{i}^{\boldsymbol{u}}_{\text{tmp}}$ denotes the result of the temporal averaging, $\hat{i}^{\boldsymbol{u}}_{\text{bil}}$ denotes the result of the bilateral filter and k is a normalizing constant. The sigmas describe the influence of the spatial and intensity distance to the final output. For robust results the median distance value of a region of interest is calculated as input for the distance correction. On the operation site, the telescope module executes the actual distance adjustment. Depending on the range information of the ToF device we adjust the length of the telescope to fit the safety margin. A fast length adaption and a small housing is an essential requirement in a clinical setup. For our prototype we have attached an L12 linear servo motor[2] to the robot assistance system. It allows adjustments at a speed of 23 mm/s and a maximum extension of 100 mm. Communication of the telescope and the computer that acquires range data using the ToF sensor is handled by a micro controller offering a Labview[3] interface. In our prototype module the open source hardware micro controller Adruino Uno [Banz 11] is used for simple data processing and instructing the telescope. The data exchange between the computer and the guidance module was performed over a serial communication interface.

8.2 Workflow Integration

An initial software setup is required before using our module the first time. Due to the generic framework all configurations in terms of preprocessing can be set up once and kept for further interventions. For the preprocessing steps we have to adjust the number of frames T used for temporal averaging, the kernel radius of the neighborhood ω' and both sigmas in bilateral filtering. Before the intervention, the supervision module needs to be attached to the endoscope holder. Instead of the actual endoscope we attach the telescope to the assistance arm and attach the endoscope to the telescope. This allows navigation of the robotic arm and correction of the distance between the endoscope and the operation site without the need of manipulating the actual endoscope holder. The ToF device is then attached to the fixed part of the telescope. In a final version of our module all components will be kept in one housing for easier usage. During the intervention the surgeon navigates the endoscope using the joystick of the robotic assistance system. Depending on the range image of the ToF sensor the telescope then automatically adjusts its length to protect healthy tissue by avoiding collisions. Furthermore, this guarantees a sufficient field of view by keeping a maximal distance to the observed surface.

[2] Firgelli Technologies Incorporate, Victoria BC, Canada
[3] http://www.ni.com/labview/



(a) (b)

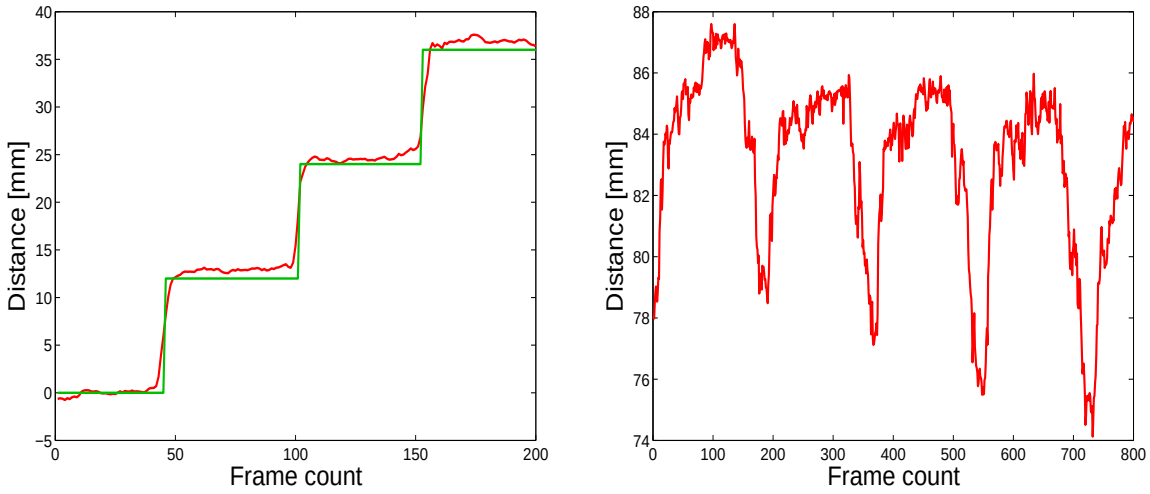
Figure 8.3: (a) shows an experimental setup to evaluate the module's behavior with different photometric intensity values. (b) shows the setup in an in-vivo pig study. The upper left image is a grayscale image acquired by the Time-of-Flight sensor. The upper right image is the corresponding color coded range image.

8.3 Evaluation and Discussion

The experiments are split into two parts. First, we measure the accuracy of the ToF sensor in a quantitative manner. Second, we demonstrate the ability of our module in an in-vivo pig study. However, due to size limitations of our prototype the experiments are performed in an open surgery manner. For all experiments the entire data processing pipeline using RITK operates at a frame rate above 20 fps on an off-the-shelf mobile graphics card (Nvidia Quadro FX 1800M). For our experiments we set $T = 4$, ω' to describe the 5×5 neighborhood and $\sigma_1 = 6$ and $\sigma_2 = 1.6$.

Fig. 8.4a demonstrates the accuracy of the PMD CamBoard nano. We measured a wooden step phantom with a given step heights of 12 mm. The 1-D signal of the ToF sensor was computed by measuring median distances in a region of interest acquired by the ToF sensor after applying the described preprocessing pipeline. Note that our measurements follow the ground truth data with a mean distance offset of less than a millimeter.

For qualitative evaluation we utilize our prototype in a pig study in an open surgery manner as illustrated in Fig. 8.3b. During the intervention the endoscope is navigated across the situs resulting in several different range plateaus. The dark blue area in Fig. 8.3b denotes an organ close to the sensor. After navigating to the dark red area, the telescope extends to keep the desired distance and shortens after returning to the blue spot. Besides the change of distance due to navigation to different operation spots we also address respiratory motion in the pig study. Fig. 8.4b illustrates the median value of a fixed region of interest for several seconds. Within this experiment, the respiration amplitude was increased artificially



(a) (b)

Figure 8.4: (a) shows a plot of a measured step phantom. In red the preprocessed Time-of-Flight data and in green the ground truth distances with 0 mm denoting the initial distance. (b) shows the median distance values during artificial respiration. Note that for interpretation of the breathing amplitude the vertical axis needs to be flipped.

using a ventilator. The plot shows the increasing amplitude by a decreasing distance to the sensor. Note that the maximal exhale state remains almost constant.

Our experiments have shown that the proposed module is feasible to supervise minimally invasive interventions and ensure a safety margin by adjusting the telescope length and thereby adjust the distance between the endoscope and the operation site. The change of distance can either be induced by navigation of the endoscope or by organ movements due to respiratory motion. The robustness of our range acquisitions depends on the size and the position of the region of interest. Due to occlusion artifacts the module is not yet capable to guarantee safety in all directions. In terms of module size upcoming ToF device are expected to satisfy the required dimensions to allow further experiments in realistic scenarios for minimally invasive surgery. The speed of the telescope motor is sufficient for smooth navigation but is expected to be increased with upcoming hardware.

8.4 Conclusion and Future Work

In this chapter we proposed a new guidance module for robot assistance systems for minimally invasive interventions. We enhanced an endoscope holder by a ToF sensor to measure the distance of the observed tissue and used a telescope to adjust the distance of the endoscope. This ensures a safety margin from healthy tissue and thereby eases the navigation for surgeons. An in-vivo pig study in an open surgery manner has shown that our module adjusts the distance to the surface and additionally allows compensating respiratory motion.

Future work will address further miniaturization and a single housing for the module to allow first in-vivo experiments in a minimally invasive manner. Furthermore, different range cameras have to be evaluated. However, even with improved range data one issue remains, i.e. at least two cameras are required to handle occlusions caused by the endoscope itself. In the current state, only one side of the endoscope can be observed to adjust the distance. Approximating healthy tissue from the opposite direction will still cause collisions. Further investigations concerning this issue should try to integrate the ToF endoscope into this module, as it covers the entire field of view directly.

CHAPTER 9

Endoscopic Tool Localization

This chapter addresses image guided surgery for robot assisted interventions. In particular, the task to localize instrument tips and thereby to ease autonomous field of view correction has always been a highly investigated field of research. Approaches to localizing endoscopic tools can be split into two groups: Tool segmentation and tracking based on color features [Doig 05] or by using prior knowledge about the tool geometry [Clim 04, Spei 08, Wolf 11]. Climent and Mares [Clim 04] proposed a technique that relies on the Hough transformation to find straight lines indicating the presence of endoscopic instruments. This approach combined with an heuristic filter achieves robust results being capable to detect the tool tip in 99% of all evaluated cases. However, to reach interactive frame rates and high robustness, several restrictions have been made, e.g. only radial lines are considered as tool candidates. Furthermore, this technique processes color images only and is thereby error-prone to inhomogeneous illumination. Doignon et al. [Doig 05] proposed another technique for recognizing endoscopic tools based on a joint hue saturation color feature. Their approach takes prior knowledge about the color of the endoscopic tool into account to perform an adaptive region growing with the seed point automatically detected by the algorithm. For this purpose, the fact that the endoscopic tool enters the scene from the boundary is utilized to search the border for minima in the joint color feature space as seed candidates. However, due to possible occlusions this assumption does not always hold true for color images. Recently, a technique relying on statistical and geometric modeling was proposed by Wolf et al. [Wolf 11]. They utilize the insertion point in 3-D and the conventional 2-D color information for more robust tool tracking and enabled tracking in 3-D compared to Climent [Clim 04]. Another method for 3-D localization using a stereo endoscope was proposed in [Spei 08] where the tool tip was first located in 2-D and in a second step its 3-D location was estimated using stereo vision.

Our approach extends those concepts by combining range and color information and utilizing a scoring system for higher robustness. The algorithm is published in [Haas 13e] and is based on prior knowledge about the color as well as the geometry of instruments used in minimally invasive procedures. The novelty of our approach is to exploit both color and range information of the new ToF/RGB endoscope to increase robustness and to enable 3-D localization of en-

doscopic tools. To increase reliability, a scoring system of intermediate results is introduced. This allows to assess the results between both modalities and enables the automatic adoption for subsequent steps to be performed on the best results of the previous calculations. In contrast to the real-time capable tool localization proposed in [Haas 13e], we published a tool segmentation technique in [Haas 13d] that requires a sophisticated preprocessing as described Chapter 6. Tool segmentation is desired if data needs to be used for registration as in Chapter 10. Due to the complicated preprocessing, we demonstrate a rather basic hybrid segmentation technique but show that the hybrid super-resolution is able to improve the segmentation output notably.

9.1 Tool Tip Localization Framework

In this section we describe our approach to locate endoscopic tools in a multimodal manner using color and range data. In the following sections the scoring value is defined within $[0, 1]$ and denoted by S with the subscript denoting the current step. Finally, $\boldsymbol{u}_{\text{Can}}$ marks a candidate point of the current step and $\hat{\boldsymbol{u}}$ denotes a point indicating a reliable result for the next step. In this thesis we denote the part of the instrument attached to the shaft as the *tool tip*. Regions of interest are abbreviated as *ROI*s. As this chapter describes a hybrid approach that works entirely equal on both modalities all equations are not associated to a specific sensor.

9.1.1 Preprocessing Pipeline

As described in Section 3.5, ToF data is not suitable to be used as the raw sensor output. Therefore, preprocessing the range images is essential for robust tool localization. Due to the high frame rate of our endoscope and time constraints in a medical environment, we use a real-time preprocessing pipeline similar to [Wasz 11a] for the range images instead of the novel preprocessing approaches described in Part II. The color information was denoised using edge-preserving guided filtering [He 13]. These color images were thereupon used for guided upsampling the ToF images in a joint manner. In detail, we apply a bilinear upsampling of the range data and then use the color data similar to Chapter 5 to denoise the data. Upsampling the range data allows to work in the same domain for both modalities within the entire framework and precludes scaling operations in each step. In terms of color data, the saturation space is a suitable representation of the color image to distinguish between instruments and body tissue as endoscopic tools are in general grayish. The saturation image is calculated as described in Eq. (7.1)

9.1.2 Generic Localization Algorithm

To increase robustness we developed a generic algorithm that can be applied to both color and range information likewise. The algorithm is divided into three steps, each followed by a rejection phase, a scoring phase and a consolidation of all intermediate results as depicted in Fig. 9.1. First, ROIs are extracted that indicate potential locations of laparoscopic tools. Second, the two lines defining the shaft of

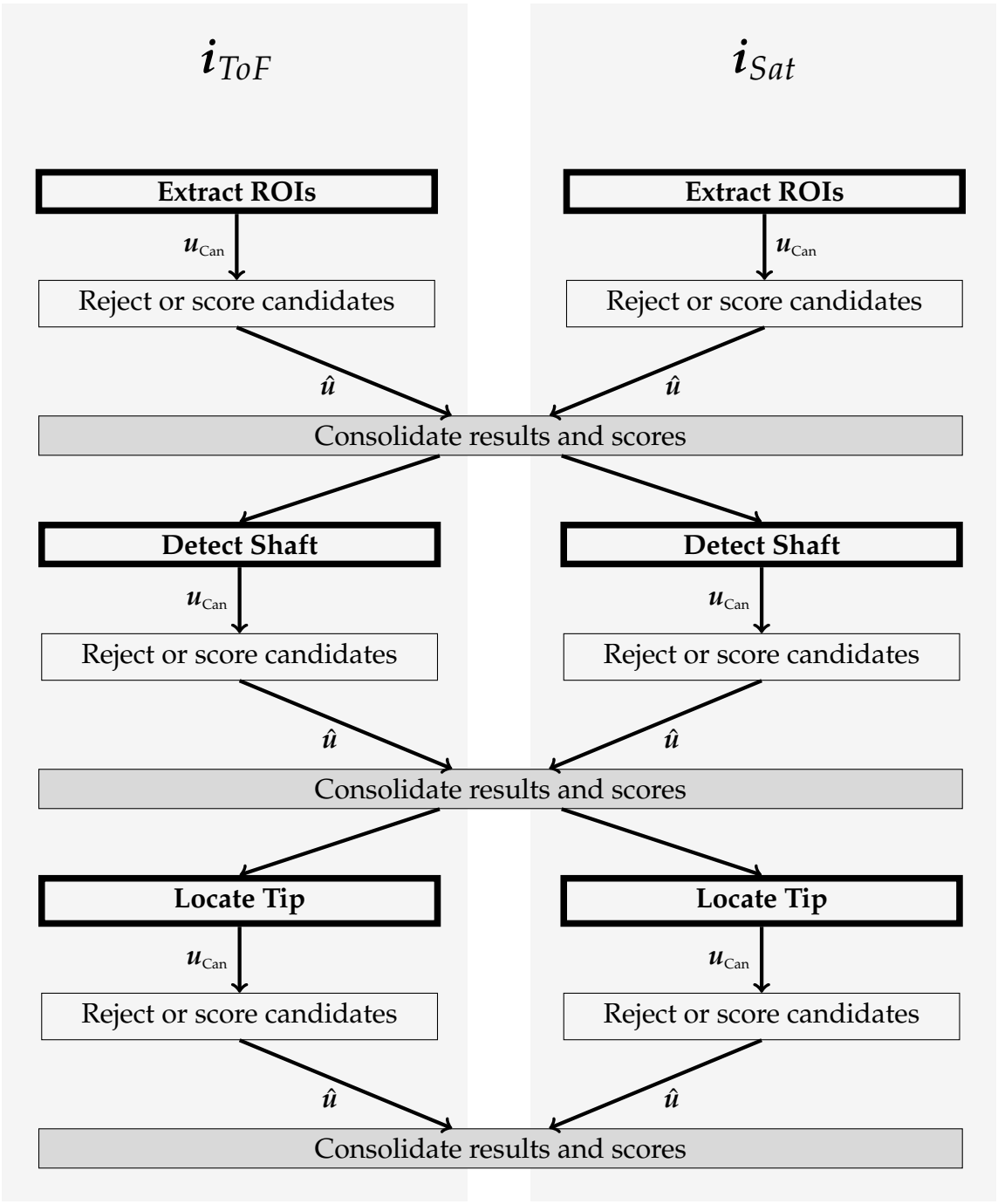


Figure 9.1: Tool localization is computed for both modalities denoted by $\boldsymbol{i}_{\text{ToF}}$ for the range image and by $\boldsymbol{i}_{\text{Sat}}$ for the saturation image. After rejecting false candidates and scoring, the results are fused for the next step.

the endoscopic tools are detected. Third, the tool tip is located along the centerline of the shaft.

ROI Extraction To reduce computation time and increase robustness all ROIs are to be found where endoscopic tools are expected. We exploit the fact that tools enter the scene from the boundary of the image. This allows us to reduce the search space for the ROIs by analyzing pixels along the border of the images only. As values in range images represent distances to the sensor, small values indicate close points. In saturation images low intensity values indicate uncolored pixels being a typical property of laparoscopic instruments. Thus, detecting local minima along the border for both modalities results in a hypothesis generation $\boldsymbol{u}_{\text{Can}}$ of an endoscopic tool. After finding all $\boldsymbol{u}_{\text{Can}}$ a twofold rejection phase is performed. First, the neighborhood of all $\boldsymbol{u}_{\text{Can}}$ is expected to have a similar value and therefore is analyzed by its variance. Second, candidates that have an intensity value $i^{\boldsymbol{u}_{\text{Can}}}$ above the mean μ of the corresponding input images are determined as unreliable and therefore rejected. Then a simple clustering is performed to fuse candidates of close mutual proximity that refer to the same tool. The representative of each cluster is denoted by $\hat{\boldsymbol{u}}$. The scoring value for each $\hat{\boldsymbol{u}}$ in the input image $\boldsymbol{i}_{\text{Inp}}$ is calculated by:

$$S_{\text{ROI}}(\hat{\boldsymbol{u}}) = 1 - \frac{i_{\text{Inp}}^{\hat{\boldsymbol{u}}}}{\mu_{\text{Inp}}} \tag{9.1}$$

for the range image candidates and equally for the saturation image candidates. This scoring value converges to 1 the lower the located minimum is compared to the mean of the whole image. The size of the ROIs is defined by δ denoting a fraction of the input image size, e.g. $\frac{1}{4}$ of the input image. The ROIs are determined with the previously found initial points $\hat{\boldsymbol{u}}$ being the center.

Shaft Detection On the saturation as well as the range images we apply the Sobel operator to find edges as depicted in Fig. 9.2. For each of the previously detected ROIs the gradient image is then transformed into Hough space to detect noticeable lines in polar coordinates. This step requires the instrument to be rigid with a cylindrical shaft, which is a valid assumption for laparoscopic tools. We assume that calculating the Hough transformation of an ROI results in two high peaks in Hough space that point to the location of the two lines describing the boundary of the shaft. The crucial step of this part is finding these maxima. To find two separated peaks, first the global maximum in Hough space is found and then the second maximum outside the neighborhood of the first maximum is located.

To reject false candidates $\boldsymbol{u}_{\text{Can}}$ in Hough space the angle between the corresponding lines is analyzed. Within a tolerable range due to perspective distortion shaft lines are expected to be parallel. Thus, if the angle exceeds a threshold $\Delta\varphi$ the line referred by the second maximum in Hough space is rejected. Otherwise both candidates $\boldsymbol{u}_{\text{Can}}$ are accepted as reliable $\hat{\boldsymbol{u}}$.

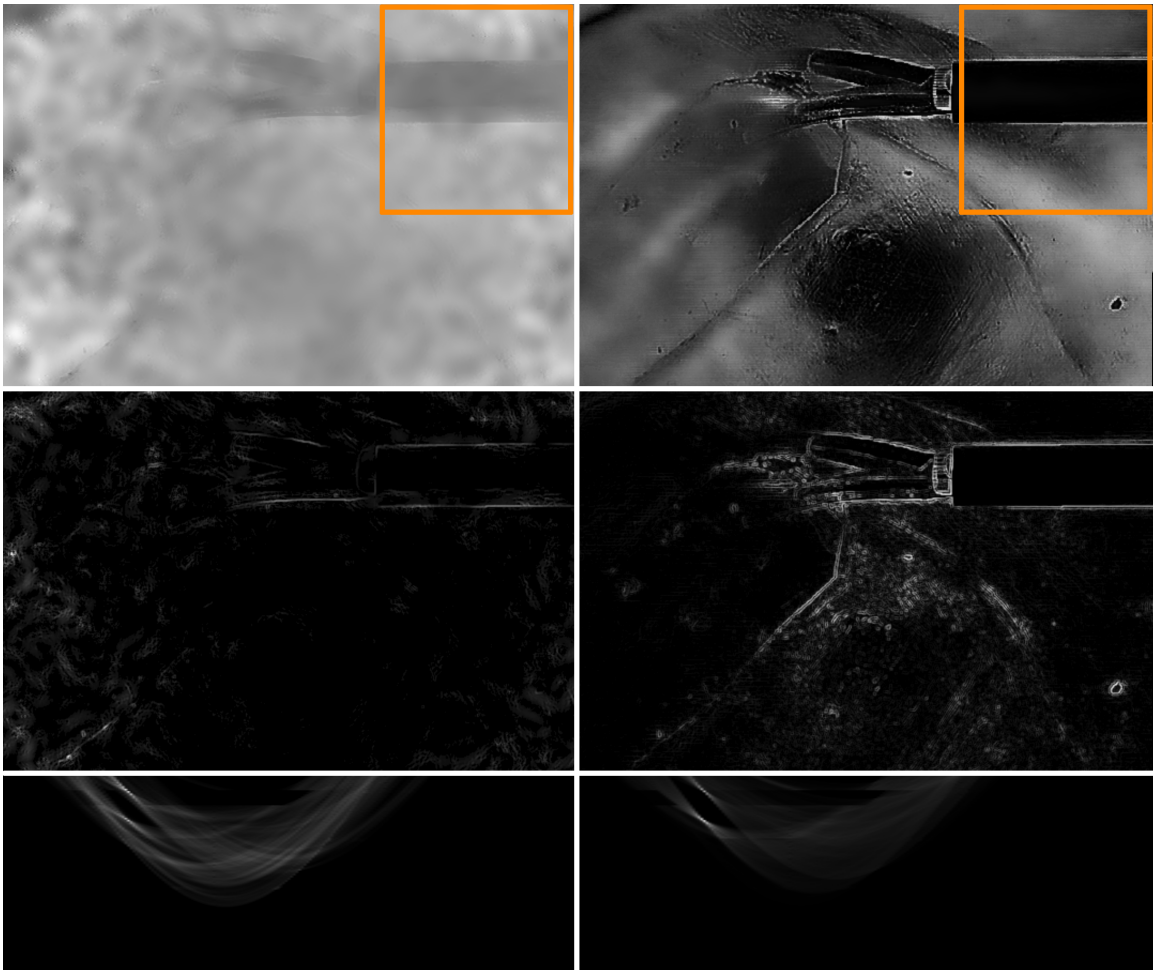


Figure 9.2: Range image outputs (first column), RGB image outputs (second column). Range and modified saturation input (first row), corresponding edge images (second row), Hough image for the calculated regions of interest denoted by the orange boxes (third row).

For scoring shaft lines the intensity of the peaks $i^{\hat{\boldsymbol{u}}}_{\text{Hough}}$ are evaluated, as a higher intensity value in Hough space indicates more points being assigned to that line. Therefore, the score is computed as:

$$S_{\text{Shaft}}(\hat{\boldsymbol{u}}) = 1 - \frac{\mu_{\text{Hough}}}{i^{\hat{\boldsymbol{u}}}_{\text{Hough}}}, \tag{9.2}$$

with μ_{Hough} denoting the mean value of votes in Hough space. A higher peak compared to the mean results in an increased score.

Tool Tip Localization Finally, we assume that the tip of the endoscopic tool needs to be located along the centerline between both detected shaft lines and is detectable by a steep gradient on this line. The point with the highest gradient indicates a step from the tool to the background.

As we assume smooth movements between successive frames, the criterion for rejection of the tip location is its distance to the location of the previous frame.

Therefore, a located point $\boldsymbol{u}_{\text{Can}}$ is considered as a reliable point $\hat{\boldsymbol{u}}$ if the 3-D distance to its location in the previous frame is below a threshold ϵ. The reliability of $\hat{\boldsymbol{u}}$ at the tip of the endoscopic tool is then computed by:

$$S_{\text{Tip}}(\hat{\boldsymbol{u}}) = 1 - \frac{\mu_{\text{Sobel}}}{i^{\hat{\boldsymbol{u}}}_{\text{Sobel}}}, \tag{9.3}$$

with μ_{Sobel} denoting the mean of all edge pixels of the input image. A strong gradient compared to all other edges results in a score converging to 1.

9.1.3 Combining Range and Color Localization

To increase robustness the result of each step is combined for both imaging modalities. For consolidation the scoring is analyzed and a weighting α denoting the reliability of the range sensor is used for each step. α depends on the hardware and scenario. The combined results serve as an input for the next step for both modalities. If the euclidean distance of their final pixel positions $\hat{\boldsymbol{u}}_{\text{Fin}}$ is low, the final score $\hat{S}_{\text{Fin}}$ for this step is calculated as:

$$\hat{S}_{\text{Fin}}(\hat{\boldsymbol{u}}_{\text{Fin}}) = \alpha S_{\text{ToF,Tip}} + (1-\alpha) S_{\text{Sat,Tip}}, \tag{9.4}$$

and the consolidated point $\hat{\boldsymbol{u}}_{\text{Fin}}$ is given as:

$$\hat{\boldsymbol{u}}_{\text{Fin}} = \alpha \hat{\boldsymbol{u}}_{\text{ToF}} + (1-\alpha) \hat{\boldsymbol{u}}_{\text{Sat}}. \tag{9.5}$$

Otherwise, if S of a single modality weighted by α still exceeds a threshold γ, $\hat{S}_{\text{Fin}}$ is calculated as:

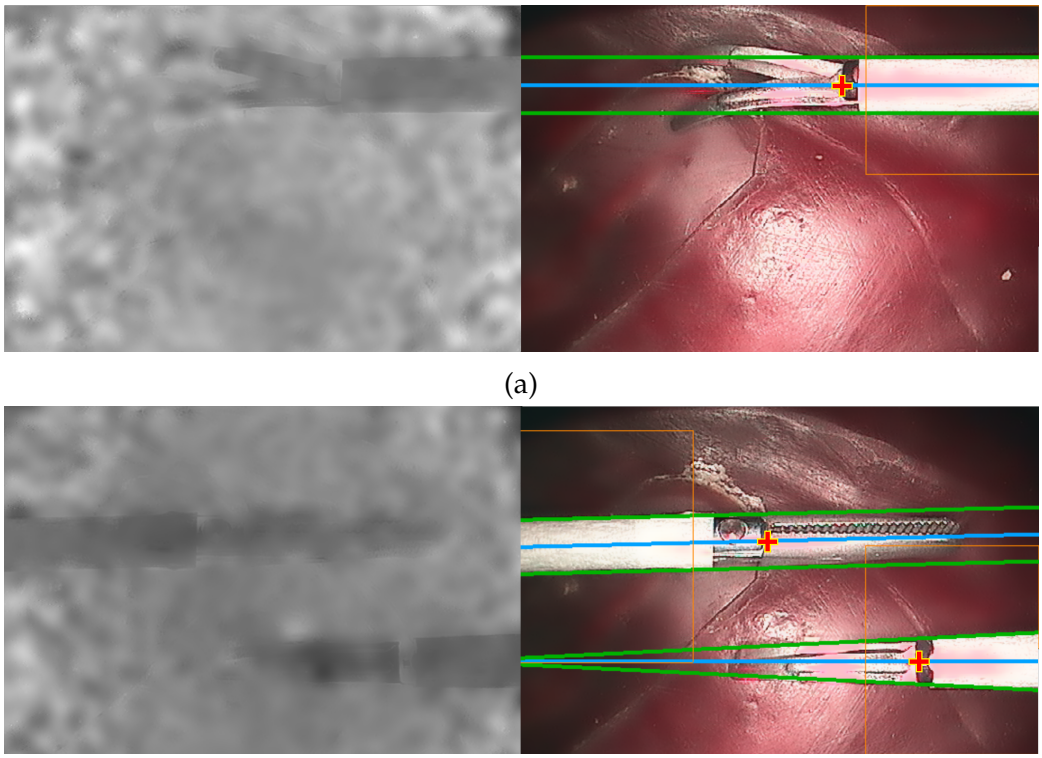
$$\hat{S}_{\text{Fin}}(\hat{\boldsymbol{u}}_{\text{Fin}}) = \begin{cases} \alpha S_{\text{ToF,Tip}} & \text{, if } \alpha S_{\text{ToF,Tip}} \geq (1-\alpha) S_{\text{Sat,Tip}} \\ (1-\alpha) S_{\text{Sat,Tip}} & \text{, if } (1-\alpha) S_{\text{Sat,Tip}} > \alpha S_{\text{ToF,Tip}} \end{cases} \tag{9.6}$$

The final result $\hat{\boldsymbol{u}}_{\text{Fin}}$ is then equal to $\hat{\boldsymbol{u}}_{\text{Fin}}$ of this modality. If in any step neither similar values nor dominant scores for a single modality are found the possible candidate for an endoscopic tool is rejected and the localization procedure for this point is aborted.

9.1.4 Evaluation and Discussion

All experiments were performed using the 3-D endoscope prototype described in Section 3.6. As this endoscope is still in an early prototype stage, all experiments are performed ex-vivo using a liver phantom and real endoscopic tools.

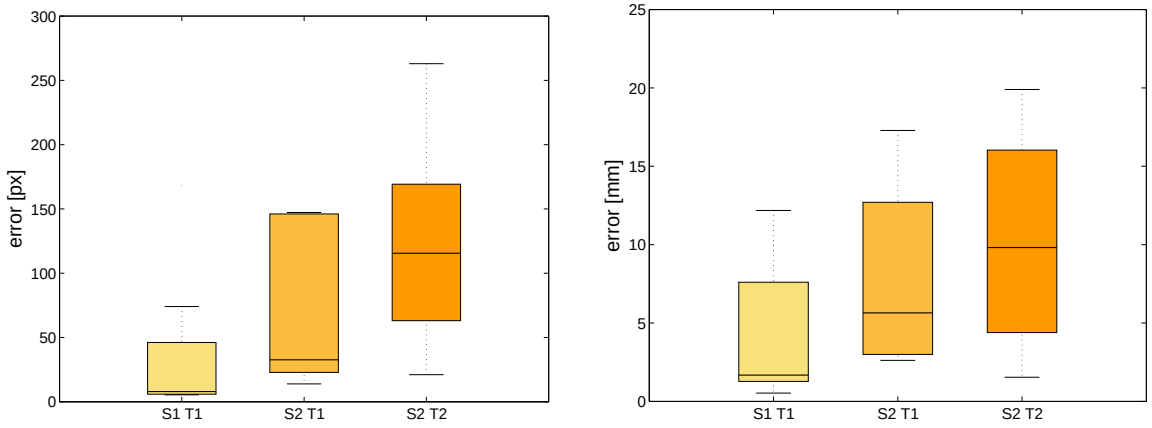
For all scenarios we set the angular threshold $\Delta\varphi$ to 25°, the scoring threshold γ to 0.15 and the size of each ROI denoted by δ to $\frac{1}{4}$ of the input image size. The threshold ϵ defining smooth movements was set to 5 mm. As in our experiments usually the RGB data showed more reliable results for the initial point the weighting α_{ROI} for the first step was set to 0.40. α_{Shaft} was set to 0.50. Due to material properties of the instruments, tool tips in the saturation image are usually located at the beginning of the tool tip, whereas the step in range data is expected at the very end of the tool. Therefore, both final tool tip locations differ even though they may describe the same tool tip. As the gradient information of the color image showed more reliable edges α_{Tip} was set to 0.



(a)

(b)

Figure 9.3: Range and color images of sequences S1 (a) and S2 (b) acquired for evaluation of the distance error. The red cross marks the detected tool tip, green the shaft boundaries, blue the centerline, orange the regions of interest.



(a) (b)

Figure 9.4: Distance errors in 2-D and 3-D between manually labeled and automatic located tool tips for both sequences (S1, S2) in boxplots.

Intersection	Sat	Range	Comb
S_{ROI} (T1)	0.55	0.20	0.41
S_{Shaft} (T1)	0.93	~~0.61~~	0.46
S_{Tip} (T1)	0.94	0.61	0.94
S_{ROI} (T2)	~~0.42~~	0.79	0.32
S_{Shaft} (T2)	~~0.69~~	0.95	0.48
S_{Tip} (T2)	0.94	0.61	0.94

Blood	Sat	Range	Comb
S_{ROI} (T1)	0	0.41	0.17
S_{Shaft} (T1)	0.83	~~0.66~~	0.42
S_{Tip} (T1)	0.92	0.64	0.92
S_{ROI} (T2)	~~0.15~~	0	0
S_{Shaft} (T2)	0	0	0
S_{Tip} (T2)	0	0	0

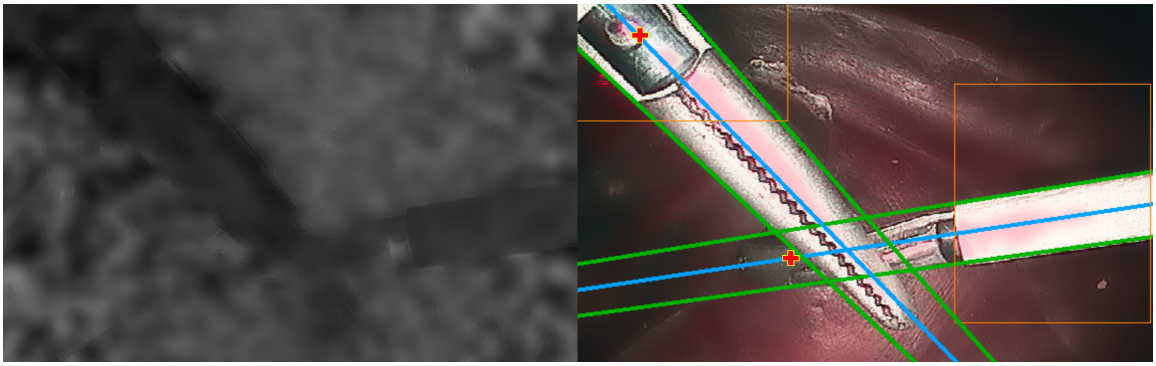
Occlusion	Sat	Range	Comb
S_{ROI} (T1)	0	0.72	0.29
S_{Shaft} (T1)	0.83	~~0.71~~	0.42
S_{Tip} (T1)	0.96	0.78	0.96
S_{ROI} (T2)	0.40	0	0.24
S_{Shaft} (T2)	0	~~0.95~~	0
S_{Tip} (T2)	0	0	0

Absence	Sat	Range	Comb
S_{ROI} (T1)	0.25	0.11	0.20
S_{Shaft} (T1)	0	~~0.78~~	0
S_{Tip} (T1)	0	0	0
S_{ROI} (T2)	~~0.11~~	0	0
S_{Shaft} (T2)	0	0	0
S_{Tip} (T2)	0	0	0

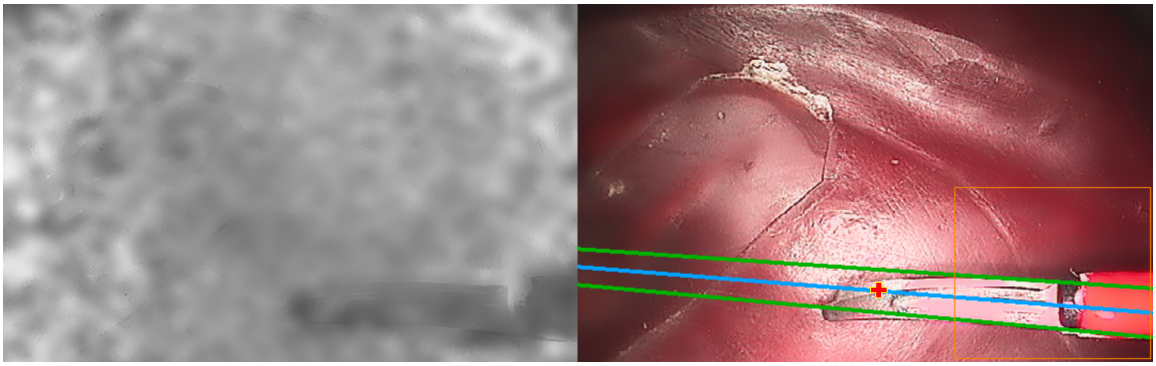
Table 9.1: The scoring results S of each intermediate step in the four challenging scenarios. The scores are calculated for both saturation and range image separately and combined. Crossed out values, denote candidates that where rejected. T1 and T2 denote two different endoscopic tools. Note that even in the absence of any tool initially candidates are detected. However, these are rejected due to our consolidation.

Sequences We evaluated the accuracy of our approach on 10 frames for two different scenarios. These scenarios include a scene with a single endoscopic tool in Fig. 9.3a and a scene with two endoscopic instruments inserted from different directions in Fig. 9.3b. For quantitative results, the 2-D and 3-D euclidean distance was calculated between the located tool tip and the ground truth data manually labeled on the fused sensor data. In Fig. 9.4a and Fig. 9.4b the euclidean distances between the manually labeled ground truth and the automatically detected tool tips are shown. Note that for all tools the median 3-D error was below 10 mm, which seems sufficient for most applications, e.g. field of view adjustment.

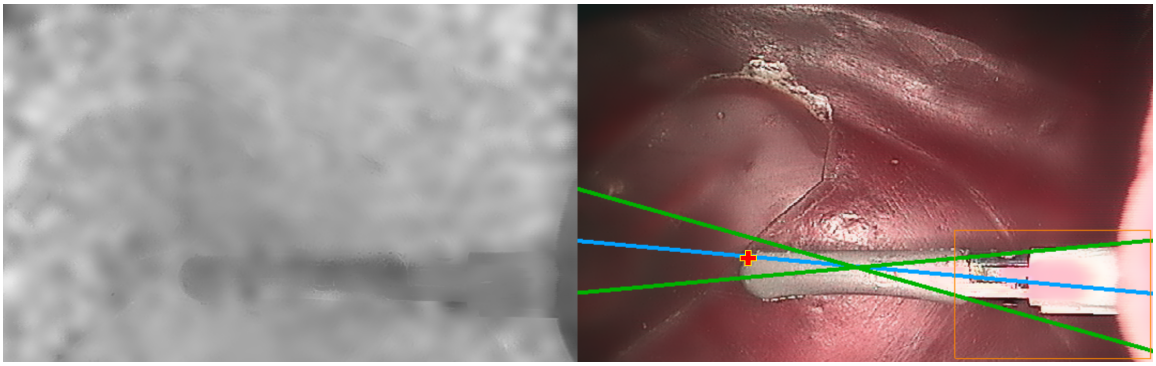
Challenging Scenarios The robustness of our algorithm was evaluated on single frames showing challenging scenarios. These scenarios include intersection of two tools, blood splatter on a tool, occlusion by the surrounding tissue and the absence of any endoscopic tool, see Fig. 9.5. Note that in all scenarios the existing endoscopic instruments were found. These experiments are evaluated by their scoring to show the influence and reliability of each modality in Table 9.1. As described in Section 9.1.3 the final scoring value is either a weighted average of both modalities if both results are close to each other or a weighted scoring of a single modality if this scoring exceeds γ. If a field contains 0s, no potential candidates are detected. Note that in the second scene the blood on the tool caused our algorithm to find only a wrong initial point in the color image due to the increased saturation on the tool. Weighting the initial point detected in the range image by α_{ROI} we determine



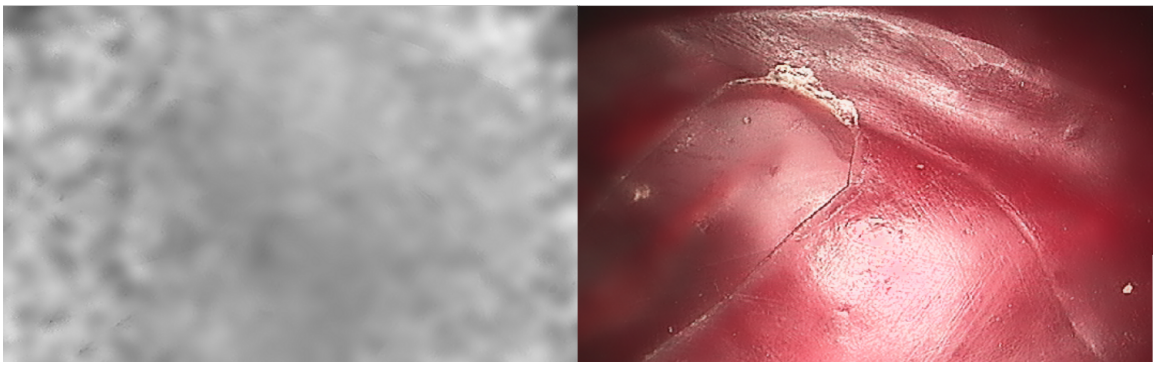
(a)



(b)



(c)



(d)

Figure 9.5: Four challenging scenarios: (a) intersection of tools, (b) blood on the shaft, (c) occlusion at the border, (d) the absence of tools. The red cross marks the detected tool tip, green the shaft boundaries, blue the centerline, orange the regions of interest.

the correct initial point while the false candidate of the color image is rejected. In the absence of any tools, a common initial point is wrongly detected in the first step but correctly rejected within our scoring phase.

9.2 Case Study: Tool Segmentation

In various other applications, e.g. range image registration as included in Chapter 10, the accurate localization of the tool tip is not of particular importance. Hence, these applications require a segmentation of the entire tool as a mask to exclude these areas from further processing. This section describes in a first preliminary case study a hybrid thresholding technique as proposed in [Haas 13d] and simultaneously shows a comparison of the current real-time capable preprocessing technique to the previously discussed hybrid super-resolution, see Chapter 6.

9.2.1 Hybrid Segmentation Framework

Based on the output of the preprocessing we apply instrument segmentation on data of both modalities. We distinguish between instruments and background by different thresholding techniques [Doig 05]. For our segmentation we exploit the fact that instruments are usually closer to the sensor and that instruments are usually grayish. Due to the data fusion in our hybrid 3-D endoscope, we can not only exploit the range data but also incorporate the color information into the segmentation process similar to [Spei 08]. Range values $i^{\boldsymbol{u}}_{\text{ToF}}$ are considered as instruments pixels if $i^{\boldsymbol{u}}_{\text{ToF}} \leq i^{min}_{\text{ToF}} + \alpha_{\text{Seg}}$. This assumption requires the minimum range value to be located on a potential tool. The parameter α_{Seg} describes a tolerance margin for range values on an instrument. We have chosen an additive term here, as independent of the global minimum value i^{min}_{ToF} α_{Seg} should describe the radius of the endoscopic tool, if entered perpendicular to the endoscope as in our case study. In the color domain we exploit the value and the saturation channel of the HSV color space to segment the instrument similar to Section 9.1.2. Here, pixels $\boldsymbol{u}$ are considered as instrument pixels if $i^{\boldsymbol{u}}_{\text{Sat}} \leq \beta_{\text{Seg}} \cdot i^{\text{max}}_{\text{Sat}}$ and $i^{\boldsymbol{u}}_{\text{Val}} \geq \gamma_{\text{Seg}} \cdot i^{\text{max}}_{\text{Val}}$, where $i^{\boldsymbol{u}}_{\text{Sat}}$ and $i^{\boldsymbol{u}}_{\text{Val}}$ denote the saturation channel and the value channel of the color image, respectively. Both binary results are then consolidated into a common segmentation mask by multiplication. For outlier removal caused by noisy data we apply morphological operators to close small holes.

9.2.2 Evaluation and Discussion

For quantitative evaluation, a dataset of Chapter 6 was manually labeled by an expert for ground truth data. The parameters were set empirically on a separate dataset to $\alpha_{\text{Seg}} = 0.03$, $\beta_{\text{Seg}} = 0.4$ and $\gamma_{\text{Seg}} = 0.6$ for data scaled to $[0, 1]$. For robustness, the minimum i^{min}_{ToF} was computed on a median filtered version with a kernel size of 15×15. Table 9.2 shows statistical properties of the hybrid segmentation compared to segmentation results for each modality individually. Furthermore, the real-time capable preprocessing of Section 9.1 is compared to the multi-sensor

super-resolution technique of Chapter 6. Fig. 9.6 shows the qualitative results of the evaluated segmentation techniques. Note that due to wrongly copied texture information in the range image, e.g. the smooth transition on the tool tip, caused by the guided filter [He 13], the F-score of the guided output is rather low, even though the output looks smoother. Nevertheless, in a hybrid approach the output still benefits from the preprocessed range data. However, with the novel super-resolution technique the F-score has the best result. Hence, a hybrid approach allows to compensate oversegmented areas in a single modality by a correct segmentation mask in the other modality and thereby increases the robustness of the entire segmentation framework.

9.3 Conclusion and Future Work

This chapter introduced a robust localization approach for endoscopic instruments and showed in a first case study a simple segmentation approach for comparison of different preprocessing techniques. Both techniques are based on the concept of hybrid 3-D imaging, i.e. acquisition of photometric information and topographic range data simultaneously. Both applications have shown a higher accuracy when applied on hybrid range data in comparison to an application driven by the data of a single modality only.

Future work on this topic is divided into two different tasks. First, further research has to cope with the issue of different data preprocessing to gain an increased SNR for the input data of tool localization and segmentation. Second, both approaches can be improved in their algorithm. For the tool localization, finding the initial point is a very important task that requires further improvement. Within the segmentation framework, the basic thresholding could be replaced by advanced techniques, e.g. k-means clustering [Ryu 12] or the segmentation technique proposed by Doignon et al. [Doig 05]. Furthermore, both techniques could benefit from combining the algorithms and using the result of one technique as input for the other algorithm.

	HSV	Guided	HSR	HSV Guided	HSV HSR
Sensitivity	0.94	0.06	0.47	0.63	0.45
Specificity	0.83	0.13	0.97	0.94	1.0
F-Score	0.46	0.09	0.51	0.53	0.60

Table 9.2: Comparison of the single segmentation results for color (HSV), guided filter preprocessing and super-resolution (HSR) preprocessing and the hybrid approaches for preprocessed data combined with color information.

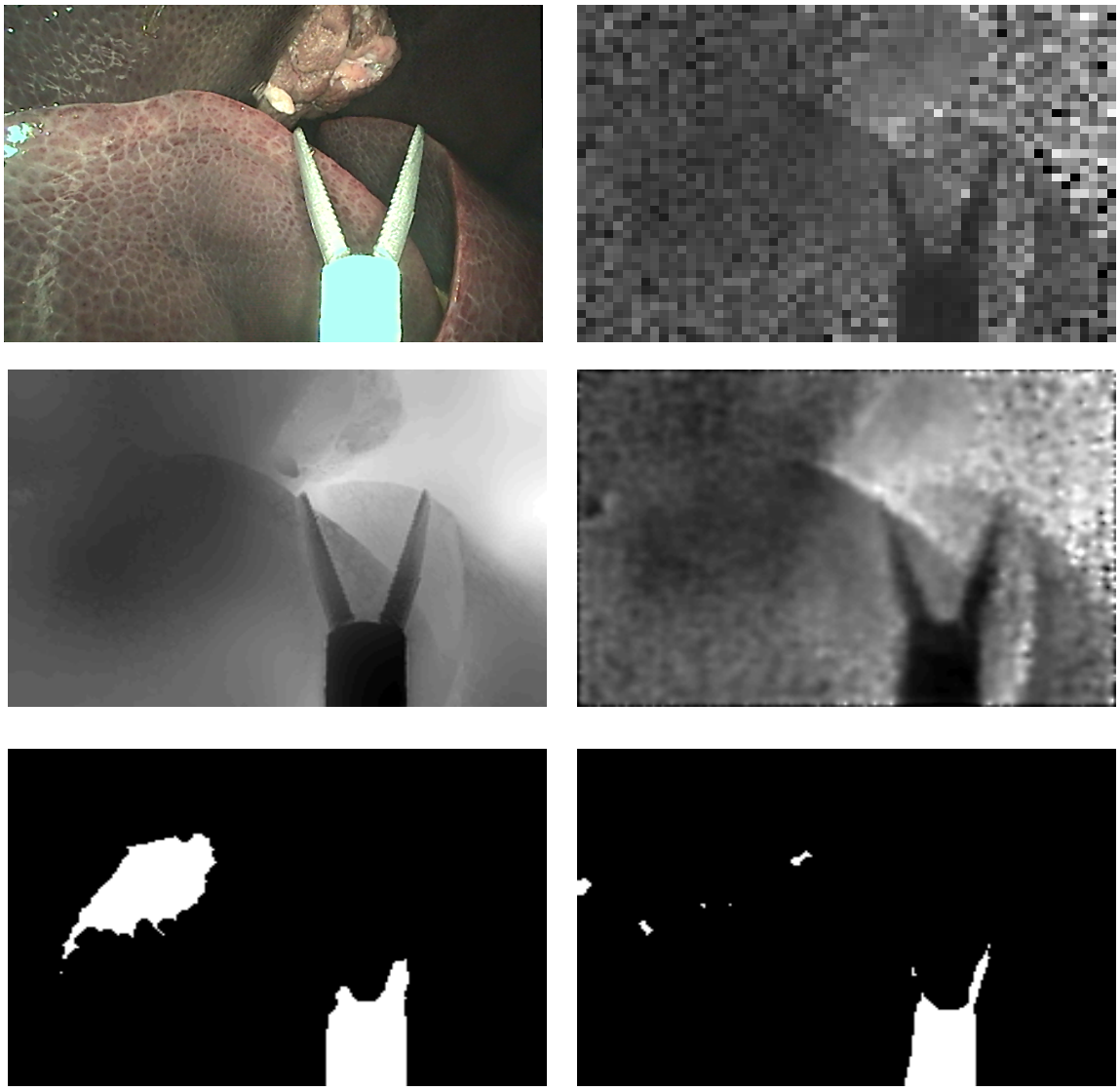


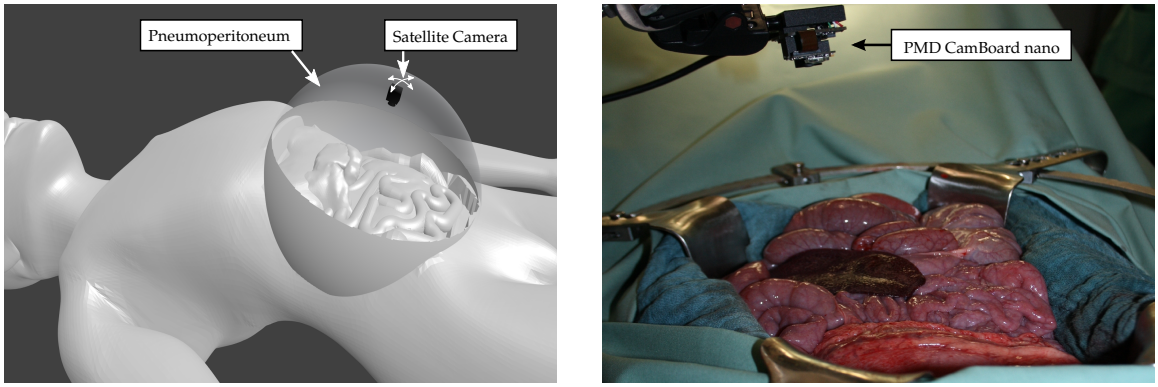
Figure 9.6: The first row denotes the input color and range image. The second row shows the output of the guided upsampled range data and the super-resolved range data. The last row shows the hybrid segmentation results for both preprocessing techniques.

CHAPTER 10

Situs Reconstruction

Navigation and orientation are of particular relevance for the surgeon in minimally invasive surgery due to the limited field of view with conventional endoscopes. To improve both, different concepts to insert additional cameras have been proposed [Oley 05, Cade 09]. For instance, Cadeddu et al. describe a video camera that is positioned on the posterior abdominal wall and guided by an anterior magnetic device [Cade 09]. Instead, we propose the concept of *3-D satellite cameras* as illustrated in Fig. 10.1a. These cameras are inserted into the abdomen via a trocar and positioned at the top of the pneumoperitoneum. Here, the imaging device can survey the operation field. Nevertheless, due to size limitations in endoscopic procedures, satellite cameras have shortcomings related to the hardware and optical systems. One of these is a narrow field of view. To expand the limited field of view the camera will reconstruct the entire situs initially by rotating and acquiring images from different areas for data fusion and then focus on the operation field. With no further repositioning of the patient the assumption of rigidity is acceptable for navigation assistance. Opposed to related work, our satellite camera delivers ToF 3-D surface and photometric information instead of pure 2-D video data. This enables a broad field of medical applications, e.g. collision detection, automatic navigation or registration with preoperative data.

Different approaches for data fusion with real-time capability have been proposed recently [Warr 12, Moun 09, Rhl 12, Newc 11]. Warren et al. proposed a simultaneous localization and mapping based approach for natural orifice transluminal endoscopic surgery [Warr 12]. For stereo endoscopy, Röhl et al. presented a novel hybrid recursive matching algorithm that performs matching on the disparity map and the two input images [Rhl 12]. Areas with little textural diversity are challenging scenarios for those color-based approaches regarding 3-D reconstruction. Instead of using conventional endoscopes we propose to navigate a 3-D satellite camera for reconstruction of the whole situs to enable a better orientation within the pneumoperitoneum. A ToF sensor acquires photogeometric data, i.e. both range data and intensity images encoding the amplitudes of the measured signal. By exploiting both complementary information we are able to reconstruct surfaces in areas with low textural diversity as well as areas with low topological diversity. The framework is based on the implicit surface representation as



(a) (b)

Figure 10.1: (a) Illustration of the 3-D ToF satellite camera hovering above the situs at the zenith of the pneumoperitoneum. (b) Experimental setup for acquiring in-vivo data in a pig study. Note the physical dimension of the miniature ToF camera.

implemented in KinectFusion [Newc 11]. In-vivo experiments on real data from a miniature ToF camera indicate the feasibility of using 3-D satellite cameras for situs reconstruction during minimally invasive surgery.

10.1 Photogeometric Data Fusion Framework

We use a truncated signed distance function (TSDF) [Curl 96] to reconstruct the interior abdominal space. The advantage of this approach is threefold. First, by incorporating successive frames, details are refined. Second, the TSDF allows incorporating additional information for regions that were seen from different perspectives comparable to super-resolution techniques. This allows implicit denoising of data with lower quality. Third, the TSDF representation is computational efficient with both constant run time and memory. Inspired by the work of Whelan et al. [Whel 13], we enhanced the traditional TSDF from 3-D to 4-D to incorporate the amplitude domain. In this context, a major contribution is the incorporation of confidence weights derived from ToF characteristics into the TSDF reconstruction. To cope with real-time requirements in medical environments we apply a GPU-based photogeometric registration approach [Baue 13]. Below we detail the initial preprocessing for ToF data. As the TSDF is an implicit surface representation raycasting techniques can be applied to obtain a range image [Park 98]. Raycasting typically describes a rendering technique to visualize 3-D volumes. However, by tracing each ray of a camera pixel, we can also fill a range image with corresponding distance values.

10.1.1 Preprocessing Pipeline

To compensate for the low SNR of ToF devices, we apply a real-time capable data enhancement pipeline that is split into three processes. First, we interpolate invalid pixels by a normalized convolution [Knut 93]. Second, we decrease the tem-

poral noise by averaging successive frames, which is possible owing to the high acquisition speed of the ToF sensor. Third, we perform bilateral filtering for edge-preserving denoising. The amplitude data depend not only on the material but also on the distance to the light source. Therefore, correcting this data is necessary for incorporating the photometric data into the registration process. We correct amplitude data according to a simplified physical model $i^{\boldsymbol{u}}_{\bar{\text{Amp}}} = i^{\boldsymbol{u}}_{\text{Amp}}(i^{\boldsymbol{u}}_{\text{ToF}})^2$. Here, $i^{\boldsymbol{u}}_{\text{Amp}}$ denotes the amplitude value at pixel position $\boldsymbol{u}$ and $i^{\boldsymbol{u}}_{\text{ToF}}$ denotes the measured radial distance [Opri 07]. Furthermore, we also apply edge-preserving denoising in the amplitude domain. Nevertheless, photometric registration is still affected by glare lights, which we detect by basic thresholding and label as invalid pixels to exclude them for further processing, see Section 7.1.

10.1.2 Range Image Registration

The preprocessed data deliver range corrected photometric and denoised geometric information of the situs from different points of view. For estimating the rotation $\boldsymbol{R}_k$ and the translation $\boldsymbol{t}_k$ between the camera coordinate system of frame k and the global world coordinate system we align two successive frames by applying an approximate iterative closest point (ICP) implementation [Baue 13]. The approach extends the traditional 3-D nearest neighbor search within ICP to higher dimensions, thus enabling the incorporation of additional complementary information, e.g. photometric data. It is based on the random ball cover acceleration structure for efficient nearest neighbor search on the GPU [Cayt 11]. The reference point set is denoted as fixed point set $\mathcal{F}$ and the point set of successive frames acquired with a moving camera are denoted by $\mathcal{M}$. We compute the closest point $\hat{\boldsymbol{x}}^f_{\text{W}}$ in the fixed dataset by:

$$\hat{\boldsymbol{x}}^f_{\text{W}} = \min_{\boldsymbol{x}^f_{\text{W}}} d(\boldsymbol{x}^f_{\text{W}}, \boldsymbol{x}^m_{\text{W}}). \tag{10.1}$$

For 4-D data considered in this chapter, the photogeometric distance metric d is defined as:

$$d(\boldsymbol{x}^f_{\text{W}}, \boldsymbol{x}^m_{\text{W}}) = \left((1-\chi)\|\boldsymbol{x}^f_{\text{W}} - \boldsymbol{x}^m_{\text{W}}\|^2_2 + \chi |i^{\boldsymbol{u}^f}_{\bar{\text{Amp}}} - i^{\boldsymbol{u}^m}_{\bar{\text{Amp}}}|^2 \right), \tag{10.2}$$

where $\chi \in [0,1]$ is a non-negative constant weighting the influence of the photometric information. $\boldsymbol{x}^f_{\text{W}}$ and $\boldsymbol{x}^m_{\text{W}}$ denote an individual 3-D point in the fixed and the moving point set, respectively. $\boldsymbol{u}^f$ and $\boldsymbol{u}^m$ denote the corresponding pixel coordinate on the sensor plane and $i_{\bar{\text{Amp}}}$ denote the range corrected amplitude value given as a scalar value, see Section 10.1.1.

For improved reconstruction, e.g. in terms of loop closures, we fuse our data in a frame-to-model manner [Newc 11], i.e. the current frame with the moving point set is not registered to the previous frame directly but to a raycasted image of the reconstructed model seen from the camera of the previous frame. Due to our high acquisition frame rate the rigid assumption for frame-to-model transformation estimation is tolerable.

10.1.3 Photogeometric Data Fusion

Our reconstruction is based on a volumetric model defined by a TSDF along the lines of [Newc 11]. The TSDF is based on an implicit surface representation given by the zero level set of an approximated signed distance function of the acquired surface. For each position $\boldsymbol{x}_{\mathrm{W}} \in \mathbb{R}^3$, the TSDF $\mathcal{T}_{\mathcal{S}}$ holds the distance to the closest point on the current range image surface $\mathcal{S}$ w.r.t. the associated inherent projective camera geometry:

$$\mathcal{T}_{\mathcal{S}}\left(\boldsymbol{x}_{\mathrm{W}}\right) = \eta\left(\|\boldsymbol{x}_{\mathrm{ToF,C}}\|_2^2 - \left\|\boldsymbol{x}_{\mathrm{C}}\|_2^2\right.\right), \tag{10.3}$$

where $\boldsymbol{x}_{\mathrm{C}} = \boldsymbol{R}_k \boldsymbol{x}_{\mathrm{W}} + \boldsymbol{t}_k$ denotes the transformation of $\boldsymbol{x}_{\mathrm{W}}$ from world space into the fixed local camera space. As described in Eq. (1.3), the 2-D pixel index $\boldsymbol{u}$ is computed by utilizing the intrinsic camera parameters $\boldsymbol{K}$ to perform the projection of each 3-D point $\boldsymbol{x}_{\mathrm{C}}$ into the image plane. $\boldsymbol{x}_{\mathrm{ToF,C}}$ is the 3-D reconstructed 2-D point $\boldsymbol{u}$ by utilizing the measured range data, see Eq. (2.1). Therefore, $\boldsymbol{x}_{\mathrm{ToF,C}}$ represents the closest point to $\boldsymbol{x}_{\mathrm{C}}$ on the surface $\mathcal{S}$. The truncation operator η controls the support region, i.e. outside this region the distance function is cut off.

To enable photogeometric reconstruction in a frame-to-model manner, our approach stores and fuses amplitude information. The amplitude value $\mathcal{T}_{Amp}$ is described by:

$$\mathcal{T}_{\mathrm{Amp}}\left(\boldsymbol{x}_{\mathrm{W}}\right) = i^{\boldsymbol{u}}_{\bar{\mathrm{Amp}}}, \tag{10.4}$$

where $\boldsymbol{u}$ is the associated 2-D pixel coordinate of $\boldsymbol{x}_{\mathrm{W}}$ computed by Eq. (1.3). For robust data fusion we assign a confidence weight to each TSDF value to describe the reliability of the new measurement. In particular, we introduce the confidence weight w of a new measurement as:

$$w(\boldsymbol{u}) = \exp(-\frac{\alpha}{i^{\boldsymbol{u}}_{\bar{\mathrm{Amp}}}}) \exp(-\frac{\|\boldsymbol{u}-\boldsymbol{c}\|_2^2}{\beta}) i^{\boldsymbol{u}}_{\mathrm{Flag}}, \tag{10.5}$$

with α and β controlling the influence of the first terms and $\boldsymbol{c}$ denoting the pixel position of the center in the range image as included in $\boldsymbol{K}$. Here, we exploit three characteristics of ToF cameras. With lower amplitude values or higher distances to the center the confidence decreases. The binary validity information $i^{\boldsymbol{u}}_{\mathrm{Flag}}$ is provided by the ToF sensor and combined with the result of our glare light detection.

To provide temporal denoising we benefit from different frames that acquired the same spots by:

$$\hat{\mathcal{T}}_t = w_t(\boldsymbol{u})\gamma\mathcal{T}_t + w_{t-1}(\boldsymbol{u})(1-\gamma)\mathcal{T}_{t-1}, \tag{10.6}$$

where $\hat{\mathcal{T}}_t$ denotes the temporal denoised result and $\mathcal{T}_t$ denotes the current result of Eq. (10.3) and Eq. (10.4). The weight γ describes the influence of the previously reconstructed result $\mathcal{T}_{t-1}$. The confidence weights $w_t(\boldsymbol{u})$ and $w_{t-1}(\boldsymbol{u})$ ensure that unreliable new data has no influence on the current surface representation.

10.2 Evaluation and Discussion

The experiments are split into two parts. First, for quantitative evaluation, we utilized the range image simulator described in Section 3.6 and reconstructed a

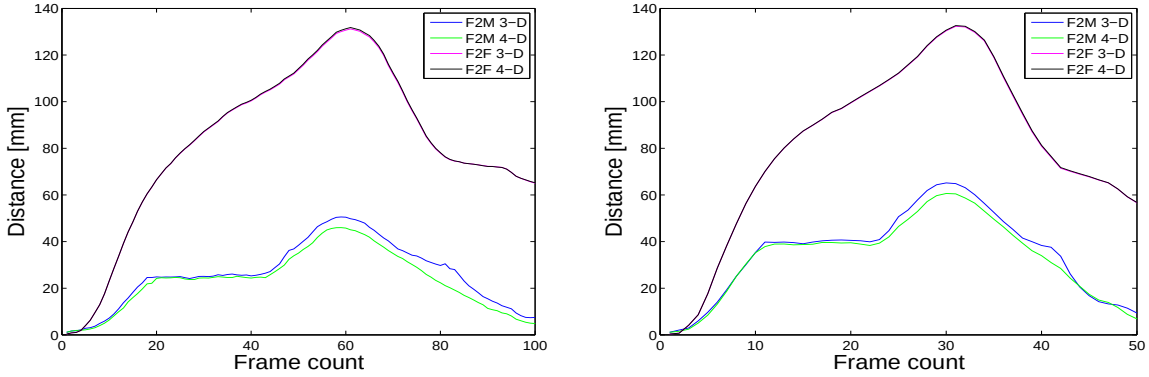
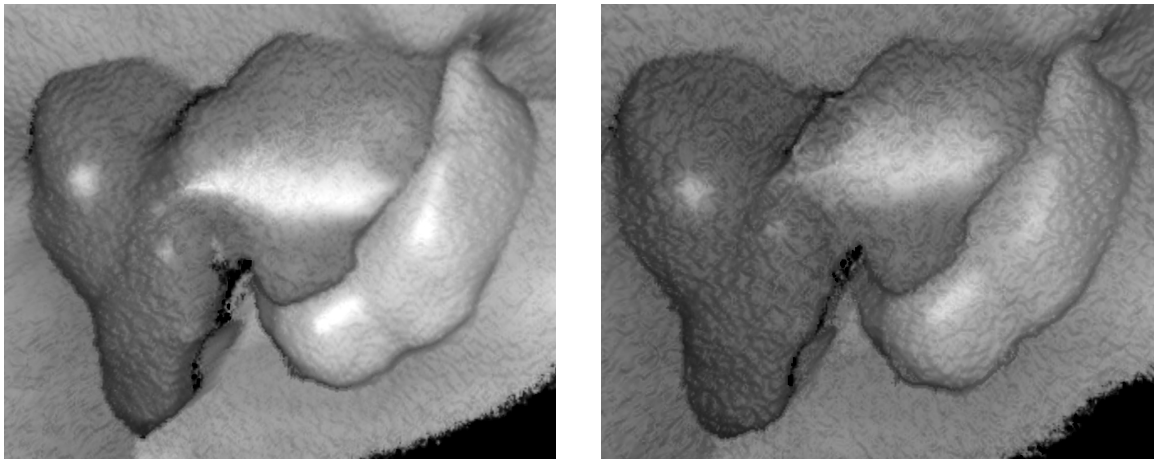


Figure 10.2: Comparison of data fusion based on frame-to-frame 3-D registration (F2F 3-D), based on frame-to-frame 4-D registration (F2F 4-D), based on frame-to-model 3-D registration (F2M 3-D) and based on our approach (F2M 4-D). The left plot shows the mean absolute error of a point-to-point distance metric for a sequence of 100 frames. The right image shows the same distances for a sequence length of 50 frames.

human abdomen with a rotating virtual satellite camera with realistic noise characteristics from different points of view. Based on the known camera path a direct comparison of ground truth data and reconstructed data is possible. Second, we acquired real in-vivo data in a pig study. In both experiments the satellite camera was moved across the situs at a typical measuring distance of 20 cm, while reconstructing the 3-D geometry of the operation field. The temporal denoising parameter was set to $\gamma = 0.95$. The weightings of the confidence terms were set to $\alpha = 2 \cdot 10^3$ and $\beta = 5$. The photometric weighting was set to $\chi = 2 \cdot 10^{-7}$. Regarding the scale of the parameter, range of the amplitude value has to be taken into account that exceeds $1 \cdot 10^4$. In the considered scenario, the texture is rather homogeneous. Hence, we set χ comparably low. Nonetheless, it guides the registration in flat regions.

Considering the quantitative evaluation, we compared the point-to-point distance of data fusion based on frame-to-frame registration to data fusion based on frame-to-model registration. In addition, we applied data fusion for pure 3-D data and with additional photometric data to prove its benefit. As illustrated in Fig. 10.2 and Fig. 10.3, we additionally showed that the camera speed within a sequence influences the reconstruction accuracy. However, in both scenarios the data fusion based on a 4-D frame-to-model registration achieves the best results. Furthermore, the plots in Fig. 10.2 highlight the great loop-closure behavior of the frame-to-model approach as the end of the sequence coincide with its beginning due to the rotational movement of the satellite camera. The median absolute error along all datasets was more than 50.00 mm for both frame-to-frame approaches, 26.01 mm for the frame-to-model 3-D approach and further reduced to 24.64 mm for our proposed frame-to-model 4-D approach. For the real data experiments we used 25 frames for data fusion. In particular, we acquired a scene of 250 frames and fused every 10th frame to obtain a sufficient frame-to-frame movement. We



(a) (b)

Figure 10.3: (a) shows our final result of the dataset evaluated in Fig. 10.2a. (b) shows the our final result of the dataset evaluated in Fig. 10.2b. Note the improved edge-preservation between both organs in (a).

averaged data over three successive frames to reduce temporal noise for the registration process. Note that even considering two additional frames for temporal denoising of each frame used for situs reconstruction, the entirety of required frames is 75, which is still an acquisition time of below 1 *s*. The parameters for the bilateral filter and the normalized convolution were set empirically. Fig. 10.4 shows in-vivo data reconstructed with our proposed technique. In Fig. 10.5 we show that the introduction of confidence weights allows to reconstruct the situs properly from the single frames. The upper right frame shown in Fig. 10.5 is clearly reconstructed wrong in the data fusion without confidence weights.

10.3 Conclusion and Future Work

This chapter introduced a miniature ToF device as a 3-D satellite camera for minimally invasive surgery to reconstruct the operation situs. To extend the camera's field of view, we introduced a fusion framework that allows to reconstruct the operation situs for better orientation and navigation using both geometric and photometric information. Our proof-of-concept GPU implementation runs at 2 Hz on an off-the-shelf laptop. Experiments on simulated and real data showed that we benefit from our proposed confidence weights and resulted in a median absolute mesh-to-mesh distance of less than 25 mm compared to ground truth data. However, current ToF cameras do not yet fit the required size for minimally invasive surgery and exhibit an insufficient SNR for real medical environments.

Future work will investigate the upcoming generation of miniaturized ToF cameras that are expected to feature a geometry that fits through a trocar. Furthermore, additional terms for the confidence maps, such as surface normals, need to be evaluated. In terms of robustness an additional color sensor with higher quality photometric data could have a beneficial impact on the reconstruction.

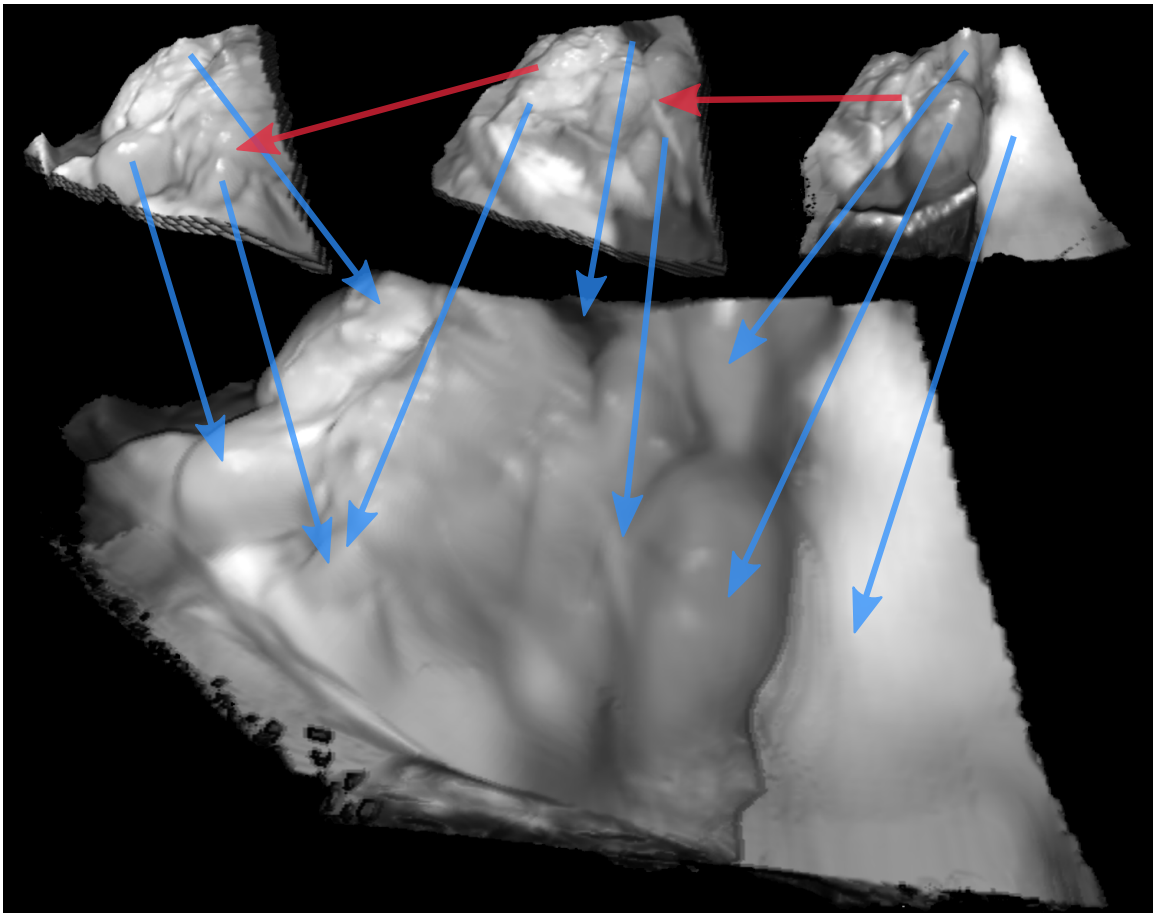
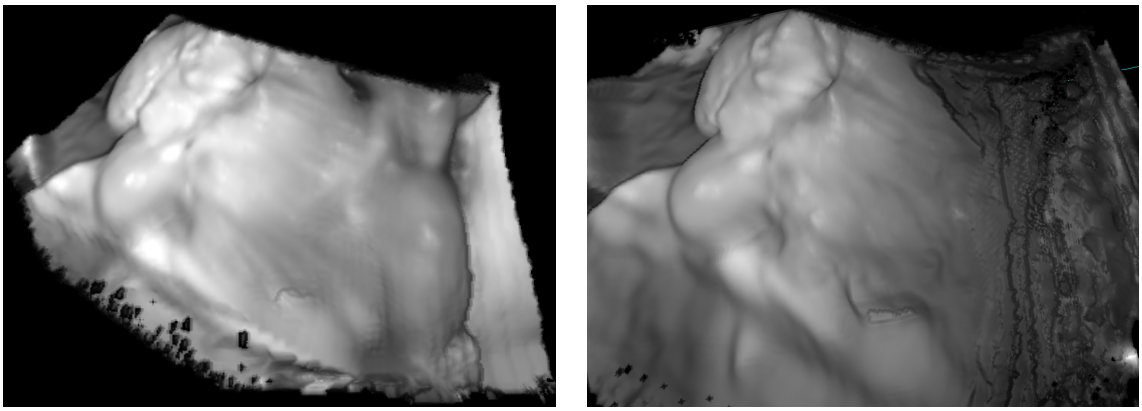


Figure 10.4: This image shows three individual frames of the pig study and the final situs reconstruction below. Blue arrows point to landmarks of individual frames and the reconstruction. Red arrows denote common landmarks in the individual frames to point out that the data fusion was performed in the correct order.



(a) (b)

Figure 10.5: (a) shows the reconstruction with and (b) without the use of the confidence maps to compensate for sensor issues.

Part IV

Summary and Outlook

CHAPTER 11

Summary

This thesis investigated the entire data processing pipeline for novel hybrid range image acquiring devices with the focus to assist in a medical environment for modern minimally invasive surgery. The term *hybrid* refers to two different imaging technologies that are utilized to acquire complementary data. In our scenario this data was composed by range images delivered by a ToF sensor and additional photometric data, i.e. either a grayscale representation of the entire scene also acquired by the ToF sensor or color data acquired by an additional RGB sensor, see Section 3.6. Not only improves the photometric information the visual impression and eases the diagnosis, the proposed preprocessing algorithms and medical applications have also shown that complementary range and photometric data has the potential to improve the actual output of image guided interventions in abdominal surgery. The aim is to reduce the duration of an intervention while keeping or improving the level of safety. The medical applications presented in this work are all together targeting the tremendous challenge of minimally invasive surgery, i.e. the navigation and orientation within the narrow field of view in the abdominal cavity.

After explaining the medical background in Chapter 2 in terms of workflow and currently available devices for minimally invasive surgery, we motivated the use of range image guided assistance systems for endoscopic interventions. Then, we introduced the three major single-shot range image acquiring technologies: stereo vision, structured light and Time-of-Flight. Consecutively, we motivate our decision to use ToF technology. Although all the work presented in this thesis was evaluated on ToF data, please note that the concepts are applicable to any other hybrid range image acquiring sensor. Subsequently, we described the current issues of range images in general and ToF technology in particular, i.e. low SNR, low spatial resolution and invalid or missing data due to specular highlights. In Chapter 3, we introduced our three data acquisition systems for experimental evaluations. In this thesis we acquired qualitative evaluation data with a ToF/RGB endoscope prototype and the compact PMD CamBoard nano as a reference hardware for a satellite camera. The latter device performs better in terms of range data quality, but does not acquire color data and is not yet compact enough for endoscopic interventions. Quantitative evaluation was performed with a range image simulator that delivers high quality ground truth range images and RGB color data as well as realistic ToF data. The synthetic scenes were composed by textured 3-D meshes of human organs based on CT scans. Our framework is able to simulate both the endoscope prototype and the satellite camera.

The first part of this work presented a fast calibration technique to estimate camera parameters, e.g. the focal length and the central point, and simultaneously align the complementary data to enable joint data processing in hybrid range image acquiring setups, see Chapter 4. In contrast to conventional checkerboard detection techniques, the proposed calibration scheme is based on a self-encoded marker with embedded 2-D barcodes that allows to use the entire field of view for automatic feature point detection. Furthermore, the simplistic embedded barcodes enable feature point detection even on low quality image data as acquired by most currently available compact ToF sensors. Using a multiscale barcode identification framework, we achieved identification rates $> 90\%$ for the RGB and ToF sensor. Based on feature points located at the corners of those barcodes in 70 different images the reprojection error using our estimated camera parameters was kept below 1 px. Due to the fact that the unique barcodes are observed and recognized from both imaging devices, i.e. the RGB and the ToF sensor, the coordinates of common feature points are thereupon used to estimate a relative transformation between both sensors, either based on a conventional stereo setup or a homographic transformation in a beam splitter setup.

The second part of the this work tackled the major issues of range imaging devices in general and for their applicability in minimally invasive surgery in particular: low signal-to-noise ratio and spatial resolution, due to size limitations in the abdominal cavity, and invalid measurements for areas affected by specular reflections, due to wet surfaces and direct light irradiation of the imaging device. In Chapter 5 we addressed the low SNR by a novel denoising approach based on the nonlocal means filter. For higher robustness in hybrid imaging, we extended the NLM concept by color weights. Here, we exploited the benefits of hybrid range image acquiring devices to use the advantages of the one sensor, i.e. high quality color data, to enhance the data of the other sensor by calculating similarity measurements in the HR color domain. Furthermore, the multi-frame hybrid NLM filter applies a tracking of image points within a sequence instead of conventionally averaging similar points within a single image. Our novel technique reduced the mean absolute distance errors compared to ground truth data by 20%. In Chapter 6 a hybrid super-resolution concept was introduced, where we estimated sensor movements based on a sequence of high-quality color images and thereupon used this information to improve the spatial resolution and data quality of the ToF sensor. The framework estimates an HR range image reconstructed by a sequence of LR range images. Here, the improvement of the mean absolute distance errors by 12% was slightly higher compared to the M-H-NLM filter. However, super-resolution increases spatial resolution simultaneously and thereby reconstructs important structures not visible in LR raw data. The evaluation has proven that both of our algorithms for data quality improvement allow to reconstruct important structures, such as tool tips of endoscopic instruments and organ boundaries, in the range domain that where not visible in the raw data. The last preprocessing concept, introduced in Chapter 7, addressed the issue of missing or invalid data due to specular reflections. Here, we also reconstructed missing structures by exploiting a sequence of hybrid ToF/RGB data. Considering movements of the device within a sequence and thereby movements of the

specular highlights, we replaced invalid regions of one frame with valid data of another frame, after aligning both frames based on robust feature points. For a reliable mask image, the highlights are detected in the HSV representation of the color images by analyzing the saturation and value channel. Considering only the effect of specular highlights, our approach reduced the mean absolute distance error by 33% compared to a basic interpolation technique.

In the final part, this thesis introduced first medical applications that hold the potential to improve and ease image guided minimally invasive surgery. The first approach avoids the need for the surgeon to be able to interpret any acquired data of the range sensor, see Chapter 8. The range data is utilized to automatically ensure a safety margin between the endoscope and the observed tissue. Based on the range data, an additional hardware module adjusts the distance with a telescope motor attached to an endoscope holder. Due to our generic composition of the module, it is applicable to any endoscope holder. In Chapter 9 we described a hybrid tool tip localization framework based on the prior knowledge of endoscopic instruments. Exploiting both the range and the color domain allowed to improve the robustness of the framework especially for challenging scenarios. Conventional instrument detection systems are based on color data only and thereby are unreliable when color of the instrument is altered, e.g. by blood or occlusion. Our hybrid approach compensates for the weaknesses of one sensor by the strengths of the other sensor. In addition to this localization approach, we showed in another feasibility study a basic segmentation framework for the entire tool based on hybrid range/RGB data. Here, we also demonstrated the advantage of our hybrid super-resolution for improved image details compared to conventional preprocessing. The last medical application is described in Chapter 10 and proposed the use of a 3-D satellite camera that acquires range and grayscale images of the scene simultaneously. The chapter introduced a framework to use a sequence of those images from different field of views to reconstruct the entire situs in 3-D for better navigation and orientation. The fusion of different range images was based on the ICP algorithm applied in a frame-to-model manner to reduce the accumulation of errors for successive frames. In a pig study, we evaluated that a sequence of range images was fused correctly to show the entire situs and furthermore showed that additional reliability weights based on photometric information improved the situs reconstruction.

In its entirety, this work has given fundamental concepts to implement range image acquiring systems for a medical environment. The three major steps for novel assistance system in modern surgery were addressed: system calibration, data preprocessing, medical applications. Although state-of-the-art ToF devices are not yet capable of acquiring data with high accuracy, we have shown that in a hybrid setup with complementary photometric information, these range image acquiring sensors hold potential for future medical applications in minimally invasive surgery. For preprocessing as well as for the final applications it is always beneficial to make use of all available data of both sensors. Algorithms solely based on 2-D color data or 3-D range data are often error-prone, as both modalities exhibit strengths and weaknesses. Conveniently, in several scenarios those

characteristics can be exploited in a joint framework to improve the final outcome driven by the strengths of both modalities.

The overall message of this thesis is the idea of using complementary data of different modalities in a joint manner to compose reliable medical assistance systems. The same idea is already in the focus of research regarding single photon emission computed tomography (SPECT) and CT data or positron emission tomography (PET) and CT data, and was in this thesis shown for hybrid range and photometric data acquiring systems.

CHAPTER 12

Outlook

As this thesis addressed the general question whether conventional color video data driven minimally invasive procedures would benefit from complementary metric range data, future work should build upon the conclusions drawn from the previous chapters. The important topics for future research cover the general assumption of similarity of structures in hybrid imaging, further investigations in novel hardware devices and development of real-time capable medical software solutions.

In all proposed algorithms augmented range data showed a beneficial effect on the output data. However, the assumption of a direct congruence between structures in color images and complementary range images can not always be taken for granted. A typical problem occurs if any texture information in the color image appears that is not present in the range data, e.g. blood flow due to a surgical cut, or vice versa, e.g. two organs at different distance with similar photometric appearance. A first approach would be to utilize the mutual information, which is an established distance metric for the similarity measurement of two datasets acquired from different modalities. If applied patch wise the mutual information could be considered as a confidence term that shows the agreement of the observed structures in both images.

In terms of hardware, with the introduction of the Kinect One© new generations of range sensors are expected. Especially, the low image resolution and data quality of ToF sensors will certainly be addressed by manufactures for future devices. As a feasibility study, the proposed algorithms were only evaluated on ToF data and simulated range images. However, future research should investigate different acquisition techniques, e.g. stereo vision setups and structured light setups, see Chapter 3. As shown in [Maie 14] all acquisitions techniques exhibit certain benefits and should thereby be evaluated. With upcoming miniature range acquiring devices, future work should also evaluate the proposed techniques on more in-vivo datasets and analyze additional issues that may occur due to the closed abdominal cavity, e.g. bad lighting conditions.

As a general outlook for all proposed algorithms, the requirement of real-time capable frameworks should hereby be reinforced. With new GPUs and new SDKs available, even the super-resolution approach is feasible to be performed in real-time. Individual parts have already been shown to be implementable under real-time restrictions [Wetz 13]. Especially in medical environments, the frame rate of a software determines its value. Here, with the rising importance of the OpenCL community, both OpenCL and CUDA should be investigated for real-time capable medical solutions.

List of Figures

List of Tables

Bibliography

[Aion 02] S. Aiono, J. Gilbert, B. Soin, P. Finlay, and A. Gordan. "Controlled trial of the introduction of a robotic camera assistant (Endo Assist) for laparoscopic cholecystectomy". *Surgical Endoscopy And Other Interventional Techniques*, Vol. 16, No. 9, pp. 1267–1270, 2002.

[Arno 10] M. Arnold, A. Ghosh, S. Ameling, and G. Lacey. "Automatic segmentation and inpainting of specular highlights for endoscopic imaging". *Journal on Image and Video Processing*, Vol. 2010, pp. 9:1–9:12, 2010.

[Bals 05] E. J. Balster, Y. F. Zheng, and R. L. Ewing. "Feature-based wavelet shrinkage algorithm for image denoising.". *IEEE Transactions on Image Processing*, Vol. 14, No. 12, pp. 2024–2039, 2005.

[Banz 11] M. Banzi. *Getting Started with Arduino*. *Make: projects*, O'Reilly Media, 2011.

[Baue 13] S. Bauer, J. Wasza, F. Lugauer, D. Neumann, and J. Hornegger. *Real-Time RGB-D Mapping and 3-D Modeling on the GPU Using the Random Ball Cover*, pp. 27–48. *Advances in Computer Vision and Pattern Recognition*, London, UK, 2013.

[Bay 06] H. Bay, T. Tuytelaars, and L. Gool. "SURF: Speeded Up Robust Features". In: *Computer Vision - ECCV 2006*, pp. 404–417, 2006.

[Ben 00] M. Ben-Ezra. "Segmentation with Invisible Keying Signal". In: *Computer Vision and Pattern Recognition (CVPR)*, pp. 1032–1037, 2000.

[Buad 05] A. Buades, B. Coll, and J. M. Morel. "A non-local algorithm for image denoising". In: *Computer Vision and Pattern Recognition (CVPR)*, pp. 60–65, 2005.

[Buad 08] A. Buades, B. Coll, and J.-M. Morel. "Nonlocal Image and Movie Denoising". *International Journal of Computer Vision*, Vol. 76, No. 2, pp. 123–139, 2008.

[Cade 09] J. Cadeddu, R. Fernandez, M. Desai, R. Bergs, C. Tracy, S.-J. Tang, P. Rao, M. Desai, and D. Scott. "Novel magnetically guided intra-abdominal camera to facilitate laparoendoscopic single-site surgery: initial human experience". *Surgical Endoscopy*, Vol. 23, pp. 1894–1899, 2009.

[Cape 04] D. Capel. *Image mosaicing and super-resolution*. PhD thesis, Oxford, 2004.

[Cayt 11] L. Cayton. "Accelerating Nearest Neighbor Search on Manycore Systems". *Computing Research Repository - CoRR*, Vol. abs/1103.2635, 2011.

[Clan 11] N. T. Clancy, D. Stoyanov, L. Maier-Hein, A. Groch, G.-Z. Yang, and D. S. Elson. "Spectrally encoded fiber-based structured lighting probe for intraoperative 3D imaging". *Biomedical Optics Express*, Vol. 2, No. 11, pp. 3119–3128, 2011.

[Clim 04] J. Climent and P. Mares. "Automatic instrument localization in laparoscopic surgery". *Electronic Letters on Computer Vision and Image Analysis*, Vol. 4, No. 1, pp. 21–31, 2004.

[Corm 01] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Data structures for disjoint sets*, Chap. 21, pp. 498Ű–524. The MIT Press, 2 Ed., 2001.

[Curl 96] B. Curless and M. Levoy. "A Volumetric Method for Building Complex Models from Range Images". In: *Conference on Computer Graphics and Interactive Techniques*, pp. 303–312, ACM, New York, NY, USA, 1996.

[Dai 13] J. Dai, O. Au, L. Fang, C. Pang, F. Zou, and J. Li. "Multichannel Nonlocal Means Fusion for Color Image Denoising". *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 23, No. 11, pp. 1873–1886, 2013.

[Dani 12] A. Danielyan, V. Katkovnik, and K. Egiazarian. "BM3D Frames and Variational Image Deblurring". *IEEE Transactions on Image Processing*, Vol. 21, No. 4, pp. 1715–1728, 2012.

[Doig 05] C. Doignon, P. Graebling, and M. de Mathelin. "Real-time segmentation of surgical instruments inside the abdominal cavity using a joint hue saturation color feature". *Real-Time Imaging*, Vol. 11, No. 5-6, pp. 429–442, 2005.

[Doug 73] D. H. Douglas and T. K. Peucker. "Algorithms for the reduction of the number of points required to represent a digitized line or its caricature". *Cartographica: The International Journal for Geographic Information and Geovisualization*, Vol. 10, No. 2, pp. 112–122, Oct. 1973.

[Fars 04] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar. "Advances and Challenges in Super-Resolution". *International Journal of Imaging Systems and Technology*, Vol. 14, No. 2, pp. 47–57, 2004.

[Fial 08] M. Fiala and C. Shu. "Self-identifying patterns for plane-based camera calibration". *Machine Vision and Applications*, Vol. 19, No. 4, pp. 209–216, 2008.

[Fiel 09] M. Field, D. Clarke, S. Strup, and W. Seales. "Stereo endoscopy as a 3-D measurement tool.". In: *IEEE Conference on Engineering in Medicine and Biology Society*, pp. 5748–5751, 2009.

[Fisc 81] M. A. Fischler and R. C. Bolles. "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography". *Communications of the ACM*, Vol. 24, No. 6, pp. 381–395, June 1981.

[Form 11] C. Forman, M. Aksoy, J. Hornegger, and R. Bammer. "Self-encoded marker for optical prospective head motion correction in {MRI}". *Medical Image Analysis*, Vol. 15, No. 5, pp. 708–719, 2011.

[Fran 07] R. Fransens, C. Strecha, and L. Van Gool. "Optical flow based super-resolution: A probabilistic approach". *Computer Vision and Image Understanding*, Vol. 106, No. 1, pp. 106–115, 2007.

[Free 70] H. Freeman. "Boundary encoding and processing". *Picture Processing and Psychopictorics*, pp. 241–266, 1970.

[Gold 04] D. Goldstein and M. Oz. *Minimally invasive cardiac surgery*. *Contemporary cardiology*, Humana Press, 2004.

[Grge 01] M. Gröger, W. Sepp, T. Ortmaier, and G. Hirzinger. "Reconstruction of Image Structure in Presence of Specular Reflections". In: *Pattern Recognition*, pp. 53–60, 2001.

[Grim 99] W. Grimson, R. Kikinis, F. A. Jolesz, and P. Black. "Image-guided surgery". *Scientific American*, Vol. 280, No. 6, pp. 54–61, 1999.

[Groc 12] A. Groch, S. Haase, and M. Wagner. "Optimierte endoskopische Time-of-Flight Oberflächenrekonstruktion durch Integration eines Struktur-durch-Bewegung Ansatzes". In: T. Tolxdorff and T. M. Deserno, Eds., *Bildverarbeitung für die Medizin*, pp. 39–44, Berlin, 2012.

[Gudm 11] S. Á. Gudmundsson and J. R. Sveinsson. "TOF-CCD image fusion using complex wavelets". In: *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1557–1560, 2011.

[Haas 12] S. Haase, C. Forman, T. Kilgus, R. Bammer, L. Maier-Hein, and J. Hornegger. "ToF/RGB Sensor Fusion for Augmented 3-D Endoscopy using a Fully Automatic Calibration Scheme". In: T. Tolxdorff, T. M. Deserno, H. Handels, and H.-P. Meinzer, Eds., *Bildverarbeitung für die Medizin*, pp. 111–116, Berlin / Heidelberg, 2012.

[Haas 13a] S. Haase, S. Bauer, J. Wasza, T. Kilgus, L. Maier-Hein, A. Schneider, M. Kranzfelder, H. Feußner, and J. Hornegger. "3-D Operation Situs Reconstruction with Time-of-Flight Satellite Cameras Using Photogeometric Data Fusion". In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 356–363, 2013.

[Haas 13b] S. Haase, C. Forman, T. Kilgus, R. Bammer, L. Maier-Hein, and J. Hornegger. "ToF/RGB Sensor Fusion for 3-D Endoscopy". *Current Medical Imaging Reviews*, Vol. 9, No. 2, pp. 113–119, 2013.

[Haas 13c] S. Haase, J. Hornegger, A. Schneider, M. Kranzfelder, H. Feußner, T. Kilgus, and L. Maier-Hein. "Time-of-Flight Based Collision Avoidance for Robot Assisted Minimally Invasive Surgery". In: P. Fiorini and G. Ferrigno, Eds., *Evaluating effectiveness and acceptance of robots in surgery: user centered design and economic factors*, pp. 000–000, 2013.

[Haas 13d] S. Haase, T. Köhler, T. Kilgus, L. Maier-Hein, J. Hornegger, and H. Feußner. "Instrument Segmentation in Hybrid 3-D Endoscopy using Multi-Sensor Super-Resolution". In: W. Freysinger, Ed., *Computer- und Roboter Assistierte Chirurgie*, pp. 194–197, 2013.

[Haas 13e] S. Haase, J. Wasza, T. Kilgus, and J. Hornegger. "Laparoscopic Instrument Localization using a 3-D Time-of-Flight/RGB Endoscope". In: *IEEE Workshop on Applications of Computer Vision (WACV)*, pp. 449–454, 2013.

[Haas 14] S. Haase, J. Wasza, M. Safak, T. Kilgus, L. Maier-Hein, H. Feußner, and J. Hornegger. "Patch based specular reflection removal for range images in hybrid 3-d endoscopy". In: *ISBI*, p. TODO, 2014.

[Hart 04] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second Ed., 2004.

[Hart 09] F. Härtl, J. Maifeld, A. Schneider, and H. Feußner. "Prospective Evaluation of a fluid driven electromagnetic support system for solo-surgery". In: O. Dössel and W. Schlegel, Eds., *World Congress on Medical Physics and Biomedical Engineering, September 7 - 12, 2009, Munich, Germany*, pp. 278–281, Springer Berlin Heidelberg, 2009.

[Hart 79] J. A. Hartigan and M. A. Wong. "Algorithm AS 136: A k-means clustering algorithm". *Applied Statistics*, Vol. 28, No. 1, pp. 100–108, 1979.

[He 13] K. He, J. Sun, and X. Tang. "Guided Image Filtering". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 6, pp. 1397–1409, 2013.

[Hu 13] W. Hu, X. Li, G. Cheung, and O. C. Au. "Depth map denoising using graph-based transform and group sparsity". In: *MMSP*, pp. 1–6, 2013.

[Hube 64] P. J. Huber. "Robust Estimation of a Location Parameter". *The Annals of Mathematical Statistics*, Vol. 35, No. 1, pp. 73–101, 03 1964.

[Huhl 10] B. Huhle, T. Schairer, P. Jenke, and W. StraSSer. "Fusion of range and color images for denoising and resolution enhancement with a non-local filter". *Computer Vision and Image Understanding*, Vol. 114, No. 12, pp. 1336–1345, 2010.

[Knut 93] H. Knutsson and C.-F. Westin. "Normalized and Differential Convolution: Methods for Interpolation and Filtering of Incomplete and Uncertain Data". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 515–523, Jun 1993.

[Koeh 14a] T. Köehler, S. Haase, T. Kilgus, L. Maier-Hein, H. Feußner, and J. Hornegger. "Bibtex not yet available". In: T. Tolxdorff, T. M. Deserno, H. Handels, and H.-P. Meinzer, Eds., *Bildverarbeitung für die Medizin*, pp. 111–116, Berlin / Heidelberg, 2014.

[Koeh 14b] T. Köehler, S. Haase, T. Kilgus, L. Maier-Hein, H. Feußner, and J. Hornegger. "Bibtex not yet available". *Medical Image Analysis*, Vol. 16, No. 5, pp. 1063–1072, 2014.

[Kohl 13] T. Köhler, S. Haase, S. Bauer, J. Wasza, T. Kilgus, L. Maier-Hein, H. Feußner, and J. Hornegger. "ToF Meets RGB: Novel Multi-Sensor Super-Resolution for Hybrid 3-D Endoscopy". In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 139–146, 2013.

[Kolb 10] A. Kolb, E. Barth, R. Koch, and R. Larsen. "Time-of-Flight Cameras in Computer Graphics". *Computer Graphics Forum*, Vol. 29, No. 1, pp. 141–159, 2010.

[Kopf 07] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele. "Joint Bilateral Upsampling". *ACM Transactions on Graphics*, Vol. 26, No. 3, 2007.

[Krem 01] K. Kremer. *Minimally Invasive Abdominal Surgery*. *Minimally Invasive Abdominal Surgery*, Georg Thieme Verlag, 2001.

[Lang 01] R. Lange and P. Seitz. "Solid-state time-of-flight range camera". *IEEE Journal of Quantum Electronics*, Vol. 37, No. 3, pp. 390–397, Mar 2001.

[Lenz 13] F. Lenzen, K. Kim, H. Schäfer, R. Nair, S. Meister, F. Becker, C. Garbe, and C. Theobalt. "Denoising Strategies for Time-of-Flight Data". In: *Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications*, pp. 25–45, Springer Berlin Heidelberg, 2013.

[Lind 07] M. Lindner and A. Kolb. "Data-Fusion of PMD-Based Distance-Information and High-Resolution RGB-Images". In: *International Symposium on Signals, Circuits and Systems (ISSCS)*, pp. 121–124, 2007.

[Lind 14] T. Lindenberger, S. Haase, T. Kilgus, L. Maier-Hein, H. Feußner, and J. Hornegger. "Bibtex not yet available". In: T. Tolxdorff, T. M. Deserno, H. Handels, and H.-P. Meinzer, Eds., *Bildverarbeitung für die Medizin*, pp. 111–116, Berlin / Heidelberg, 2014.

[Liu 09] C. Liu. *Beyond Pixels: Exploring New Representations and Applications for Motion Analysis*. PhD thesis, Massachusetts Institute of Technology, 2009.

[Lowe 04] D. G. Lowe. "Distinctive Image Features from Scale-Invariant Keypoints". *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91–110, Nov. 2004.

[Maie 13] L. Maier-Hein, P. Mountney, A. Bartoli, H. Elhawary, D. Elson, A. Groch, A. Kolb, M. Rodrigues, J. Sorger, S. Speidel, and D. Stoyanov. "Optical techniques for 3D surface reconstruction in computer-assisted laparoscopic surgery". *Medical Image Analysis*, Vol. 17, No. 8, pp. 974–996, 2013.

[Maie 14] L. Maier-Hein, A. Groch, A. Bartoli, S. Bodenstedt, G. Boissonnat, P.-L. Chang, N. Clancy, D. Elson, S. Haase, E. Heim, J. Hornegger, P. Jannin, H. Kenngott, T. Kilgus, B. Müller-Stich, D. Oladokun, S. Röhl, T. dos Santos, H.-P. Schlemmer, A. Seitel, S. Speidel, M. Wagner, and D. Stoyanov. "Comparative Validation of Single-shot Optical Techniques for Laparoscopic 3D Surface Reconstruction". *IEEE Transactions on Medical Imaging*, Vol. 33, No. 10, pp. 1913–1930, Oct 2014.

[Mali 77] D. F. Malin. "Unsharp Masking". *AAS Photo-Bulletin*, No. 16, pp. 10–13, 1977.

[Mers 13] S. Mersmann, A. Seitel, M. Erz, B. Jähne, F. Nickel, M. Mieth, A. Mehrabi, and L. Maier-Hein. "Calibration of time-of-flight cameras for accurate intraoperative surface reconstruction". *Medical Physics*, Vol. 40, No. 8, pp. –, 2013.

[Moun 09] P. Mountney and G.-Z. Yang. "Dynamic view expansion for minimally invasive surgery using simultaneous localization and mapping.". In: *IEEE Conference on Engineering in Medicine and Biology Society*, pp. 1184–7, 2009.

[Mour 01] F. Mourgues, F. Devemay, and E. Coste-Maniere. "3D reconstruction of the operating field for image overlay in 3D-endoscopic surgery". In: *IEEE and ACM International Symposium on Augmented Reality*, pp. 191–192, 2001.

[Muel 04] U. Mueller-Richter, A. Limberger, P. Weber, K. Ruprecht, W. Spitzer, and M. Schilling. "Possibilities and limitations of current stereo-endoscopy". *Surgical Endoscopy And Other Interventional Techniques*, Vol. 18, No. 6, pp. 942–947, 2004.

[Nabn 02] I. T. Nabney. *NETLAB: Algorithms for Pattern Recognition*. *Advances in Pattern Recognition*, Springer, 1^{st} Ed., 2002.

[Newc 11] R. A. Newcombe, A. J. Davison, S. Izadi, P. Kohli, O. Hilliges, J. Shotton, D. Molyneaux, S. Hodges, D. Kim, and A. Fitzgibbon. "KinectFusion: Real-time dense surface mapping and tracking". In: *IEEE International Symposium on Mixed and Augmented Reality*, pp. 127–136, Oct. 2011.

[Oley 05] D. Oleynikov, M. Rentschler, A. Hadzialic, J. Dumpert, S. R. Platt, and S. Farritor. "Miniature robots can assist in Laparoscopic cholecystectomy". *Surgical Endoscopy*, Vol. 19, No. 4, pp. 473–476, April 2005.

[Opri 07] S. Oprisescu, D. Falie, M. Ciuc, and V. Buzuloiu. "Measurements with ToF Cameras and Their Necessary Corrections". In: *International Symposium on Signals, Circuits and Systems (ISSCS)*, pp. 1–4, July 2007.

[Park 03] S. C. Park, M. K. Park, and M. G. Kang. "Super-resolution image reconstruction: a technical overview". *IEEE Signal Processing Magazine*, Vol. 20, No. 3, pp. 21–36, 2003.

[Park 11] J. Park, H. Kim, Y.-W. Tai, M. Brown, and I. Kweon. "High quality depth map upsampling for 3D-TOF cameras". In: *IEEE International Conference on Computer Vision (ICCV)*, pp. 1623–1630, 2011.

[Park 98] S. Parker, P. Shirley, Y. Livnat, C. Hansen, and P.-P. Sloan. "Interactive ray tracing for isosurface rendering". In: *Conference on Visualization*, pp. 233–238, Oct 1998.

[Penn 09] J. Penne, K. Höller, M. Stürmer, T. Schrauder, A. Schneider, R. Engelbrecht, H. Feußner, B. Schmauss, and J. Hornegger. "Time-of-Flight 3-D Endoscopy". In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 467–474, 2009.

[Penn 10] J. Penne, C. Schaller, R. Engelbrecht, L. Maier-Hein, B. Schmauss, H.-P. Meinzer, and J. Hornegger. "Laparoscopic Quantitative 3D Endoscopy for Image Guided Surgery". In: *Bildverarbeitung für die Medizin*, pp. 16–20, 2010.

[Pole 08] R. Polet and J. Donnez. "Using a laparoscope manipulator (LAPMAN) in laparoscopic gynecological surgery.". *Surgical Technology International*, Vol. 17, No. , pp. 187–191, 2008.

[Puer 13] G. Puerto-Souza and G.-L. Mariottini. "A Fast and Accurate Feature-Matching Algorithm for Minimally-Invasive Endoscopic Images". *IEEE Transactions on Medical Imaging*, Vol. 32, No. 7, pp. 1201–1214, July 2013.

[Reyn 11] M. Reynolds, J. Dobos, L. Peel, T. Weyrich, and G. Brostow. "Capturing Time-of-Flight data with confidence". In: *Computer Vision and Pattern Recognition (CVPR)*, pp. 945–952, 2011.

[Rhl 12] S. Röhl, S. Bodenstedt, S. Suwelack, H. Kenngott, B. P. Müller-Stich, R. Dillmann, and S. Speidel. "Dense GPU-enhanced surface reconstruction from stereo endoscopic images for intraoperative registration". *Medical Physics*, Vol. 39, No. 3, pp. 1632–1645, 2012.

[Ryu 12] J. Ryu, J. Choi, and H. C. Kim. "Endoscopic vision based tracking of multiple surgical instruments in robot-assisted surgery". In: *International Conference on Control, Automation and Systems (ICCAS)*, pp. 2195–2198, Oct 2012.

[Sabo 10] A. Sabov and J. Krüger. "Identification and Correction of Flying Pixels in Range Camera Data". In: *Spring Conference on Computer Graphics*, pp. 135–142, 2010.

[Same 88] H. Samet and M. Tamminen. "Efficient Component Labeling of Images of Arbitrary Dimension Represented by Linear Bintrees". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 10, No. 4, pp. 579–586, July 1988.

[Scha 07] O. Schall, A. Belyaev, and H.-P. Seidel. "Feature-preserving Non-local Denoising of Static and Time-varying Range Data". In: *ACM Symposium on Solid and Physical Modeling*, pp. 217–222, ACM, 2007.

[Schm 12] C. Schmalz, F. Forster, A. Schick, and E. Angelopoulou. "An endoscopic 3D scanner based on structured light". *Medical Image Analysis*, Vol. 16, No. 5, pp. 1063–1072, 2012.

[Scho 97] R. E. Schoen, L. D. Gerber, and C. Margulies. "The pathologic measurement of polyp size is preferable to the endoscopic estimate". *Gastrointestinal Endoscopy*, Vol. 46, No. 6, pp. 492–496, 1997.

[Schu 09] S. Schuon, C. Theobalt, J. Davis, and S. Thrun. "LidarBoost: Depth superresolution for ToF 3D shape scanning". In: *Computer Vision and Pattern Recognition (CVPR)*, pp. 343–350, 2009.

[Spei 08] S. Speidel, G. Sudra, J. Senemaud, M. Drentschew, B. P. Müller-Stich, C. Gutt, and R. Dillmann. "Recognition of risk situations based on endoscopic instrument tracking and knowledge based situation modeling". In: *Society of Photographic Instrumentation Engineers (SPIE)*, pp. 69180X–69180X–8, 2008.

[Spei 09] S. Speidel, J. Benzko, S. Krappe, G. Sudra, P. Azad, B. P. Müller-Stich, C. Gutt, and R. Dillmann. "Automatic classification of minimally invasive instruments based on endoscopic image sequences". In: *Society of Photographic Instrumentation Engineers (SPIE)*, pp. 72610A–72610A–8, 2009.

[Stoy 05] D. Stoyanov and G.-Z. Yang. "Removing specular reflection components for robotic assisted laparoscopic surgery". In: *IEEE International Conference on Image Processing (ICIP)*, pp. 632–635, 2005.

[Stoy 10] D. Stoyanov, M. Scarzanella, P. Pratt, and G.-Z. Yang. "Real-Time Stereo Reconstruction in Robotically Assisted Minimally Invasive Surgery". In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 275–282, 2010.

[Stud 99] C. Studholme, D. Hill, and D. Hawkes. "An overlap invariant entropy measure of 3D medical image alignment". *Pattern Recognition*, Vol. 32, No. 1, pp. 71–86, 1999.

[Su 09] L.-M. Su, B. P. Vagvolgyi, R. Agarwal, C. E. Reiley, R. H. Taylor, and G. D. Hager. "Augmented Reality During Robot-assisted Laparoscopic Partial Nephrectomy: Toward Real-Time 3D-CT to Stereoscopic Video Registration". *Urology*, Vol. 73, No. 4, pp. 896–900, 2009.

[Sung 01] G. T. Sung and I. S. Gill. "Robotic laparoscopic surgery: a comparison of the da Vinci and Zeus systems". *Urology*, Vol. 58, No. 6, pp. 893–898, 2001.

[Tikh 77] A. Tikhonov and V. Arsenin. *Solutions of ill-posed problems*. *Scripta series in mathematics*, Winston, 1977.

[Tipp 03] M. E. Tipping and C. M. Bishop. "Bayesian Image Super-resolution". In: *Advances in Neural Information Processing Systems*, pp. 1303–1310, MIT Press, 2003.

[Toma 98] C. Tomasi and R. Manduchi. "Bilateral Filtering for Gray and Color Images". In: *International Conference on Computer Vision (ICCV)*, pp. 839–846, 1998.

[Torr 00] P. H. S. Torr and A. Zisserman. "MLESAC: A New Robust Estimator with Application to Estimating Image Geometry". *Computer Vision and Image Understanding*, Vol. 78, p. 2000, 2000.

[Totz 11] J. Totz, P. Mountney, D. Stoyanov, and G.-Z. Yang. "Dense Surface Reconstruction for Enhanced Navigation in MIS". In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 89–96, 2011.

[Unit 10] United States Department of Health and Human Services. Centers for Disease Control and Prevention. National Center for Health Statistics. "National Hospital Discharge Survey, 2010". 2010.

[Vaki 94] N. Vakil, W. Smith, K. Bourgeois, E. Everbach, and K. Knyrim. "Endoscopic measurement of lesion size: Improved accuracy with image processing". *Gastrointestinal Endoscopy*, Vol. 40, No. 2, pp. 178–183, 1994.

[Vela 00] V. Velanovich. "Laparoscopic vs open surgery". *Surgical Endoscopy*, Vol. 14, No. 1, pp. 16–21, 2000.

[Warr 12] A. Warren, P. Mountney, D. Noonan, and G.-Z. Yang. "Horizon Stabilized-Dynamic View Expansion for Robotic Assisted Surgery (HS-DVE)". *International Journal of Computer Assisted Radiology and Surgery*, Vol. 7, No. 2, pp. 281–288, 2012.

[Wasz 11a] J. Wasza, S. Bauer, S. Haase, M. Schmid, S. Reichert, and J. Hornegger. "RITK: The Range Imaging Toolkit - A Framework for 3-D Range Image Stream Processing". In: P. Eisert, J. Hornegger, and K. Polthier, Eds., *VMV 2011: Vision, Modeling & Visualization*, pp. 57–64, 2011.

[Wasz 11b] J. Wasza, S. Bauer, and J. Hornegger. "High Performance GPU-based Preprocessing for Time-of-Flight Imaging in Medical Applications". In: H. E. Handels, Ed., *Bildverarbeitung für die Medizin*, pp. 324–328, Berlin Heidelberg, 2011.

[Well 93] P. D. Wellner. "Adaptive Thresholding for the Digital Desk". Tech. Rep., 1993.

[Wetz 13] J. Wetzl, O. Taubmann, S. Haase, T. Köhler, M. Kraus, and J. Hornegger. "GPU Accelerated Time-of-Flight Super-Resolution for Image-Guided Surgery". In: T. Tolxdorff and T. M. Deserno, Eds., *Bildverarbeitung für die Medizin*, pp. 21–26, 2013.

[Whel 13] T. Whelan, H. Johannsson, M. Kaess, J. Leonard, and J. McDonald. "Robust real-time visual odometry for dense RGB-D mapping". In: *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5724–5731, May 2013.

[Wolf 11] R. Wolf, J. Duchateau, P. Cinquin, and S. Voros. "3D Tracking of Laparoscopic Instruments Using Statistical and Geometric Modeling". In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 203–210, Springer Berlin Heidelberg, 2011.

[Wu 09] C. Wu, S. G. Narasimhan, and B. Jaramaz. "A Multi-Image Shape-from-Shading Framework for Near-Lighting Perspective Endoscopes". *International Journal of Computer Vision*, February 2009.

[Xu 10] Y. Xu, F. Wang, and Y. Zhao. "Matching based Highlight Removal". In: *International Conference on Multimedia Technology (ICMT)*, pp. 1–4, 2010.

[Zhan 04] Q. Zhang and R. Pless. "Extrinsic calibration of a camera and laser range finder (improves camera calibration)". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2301–2306, 2004.

[Zimm 06] G. Zimmerman-Moreno and H. Greenspan. "Automatic detection of specular reflections in uterine cervix images". In: *Society of Photographic Instrumentation Engineers (SPIE)*, pp. 61446E–61446E–9, 2006.