Master Thesis Report

Image-matching based navigation system for robotic ureteroscopy in kidney exploration

by

Zhili Huang

Student Name	Zhili Huang
Student number	5205557

Supervisor:	Prof. Dr. Jenny Dankelman
Assistant supervisor:	Chun-Feng Lai
Project Duration:	Feb, 2022 - Sep, 2022
Faculty:	Faculty of Mechanical, Maritime and Materials Engineering, Delft
Track:	BioMechanical Design

TU Delft

Acknowledgements

This thesis marks the end of my MSc in Mechanical Engineering at Delft University of Technology. It has been a great and meaningful experience, and I would like to thank the people who helped me during the project.

First, I would like to thank my supervisor Jenny Dankelman for providing this opportunity to work on this interesting project and for guiding me on the right track. I would also like to give a big thank to Chun-Feng Lai for his support. Whenever I encountered a problem and stagnated, he would always help me to sort it out patiently and provide comprehensive advice. We respect each other, and it is a great pleasure to work with him.

Secondly, I would like to thank my parents for their support not only financially, but also mentally. A special thank goes to my wife, Ye Qiu, who bravely accompanied me to an unfamiliar city and bravely went back to China alone and faced a long separation for the sake of our future. She always supports me wholeheartedly, guides me when I'm down, and rejoices in my achievements.

Finally, thanks to all the friends I met at the TU Delft. They make my life colorful and let me respect each individual. Also, thanks to all the people who provide support work for TU Delft. Without them, there will be no comfortable learning environment, complete hardware facilities, and sufficient food and supplies.

Abstract

Kidney stone disease has become the most common urinary tract disease. Around 12% of the world population was influenced by kidney stone diseases. Among the main treatments, flexible ureteroscopy (fURS) is always the first choice when facing intermediate-size stones, because it can remove the stones of various compositions with a flexible endoscope inserted through a natural orifice, and is safer with less bleeding.

One of the main problems when surgeons perform the fURS operation on the urinary system, especially for kidney exploration, is to **navigate the ureteroscope**. Although fURS can provide surgeons views inside the kidney, because of the narrow space in the kidney and the limited field of view (FOV) of a ureteroscope camera, surgeons find it hard to tell which calyx is the image representing by only using a ureteroscope. So the problem can be defined as follows: Which calyx in the kidney is being observed, and where is the location and orientation of the ureteroscope tip inside the kidney?

To overcome these problems, this thesis aims to propose an **image-matching-based navigation system** to help surgeons know where the ureteroscope tip is. The whole system framework consists of two steps: pre- and post-operation.

Pre-operation: Firstly, a 3-Dimension (3D) kidney model of a patient can be generated using computed tomography (CT) scan data. Secondly, the 3D kidney model can be used to render a virtual ureteroscopy (VURS) environment to simulate the ureteroscopy procedure. Finally, VURS images of different calyxes in the kidney model taken by the virtual camera (VC) are collected with the corresponding position and orientation (pose) data of the VC as the Virtual image database (VID).

Post-operation: Firstly, real calyx images (RCIs) inside the kidney are acquired through a robotic ureteroscope. Secondly, edge features are extracted from these images as the input for the matching algorithm. Thirdly, the best matching image in the VID can be found after performing the matching algorithm. Finally, according to the RCI-corresponding pose data of the VC, the location of the real ureteroscope can be inferred, which can be shown in the VURS environment.

This thesis uses an open-source 3D kidney STL model to replace the data from CT scans due to resource limitations. Unity software is used to generate the VURS, where the VID is collected. A robotic ureteroscope prototype and a 3D-printing kidney phantom based on the STL model are used to simulate the fURS procedure to get RCIs. After applying image processing techniques to those images, the shape context (SC) matching method is used to find the best matching image from the VID. From the pose data of the VC corresponding to the resulting image, surgeons can know the real ureteroscope tip location information.

The simulation and experiment are conducted to test the feasibility and performance of the system. Some simulation images are collected from the same VURS environment with a point light interference and additional noise. Those simulation images are treated as the input images to test the SC matching algorithm. In the experiment, coordinates registration between the EMT and the VURS coordinate needs to be performed to match results validation. By using the robotic ureteroscope prototype with an EMT sensor controlled by a gamepad, RCIs and their corresponding EMT position data can be collected in the kidney phantom. Those RCIs are matched with VID images to find the best matching image. The correctness of the resulting image is judged by both the ground truth provided by the author and the EMT system. In addition, the VC location of the result in the VURS environment are visualized to provide an intuitive understanding.

The experiment results show the system performances on both simulation images and RCIs. For simulation images, the system feasibility is verified with around 90% accuracy. The performance is reduced for the RCIs, but the system is still feasible. Because of the limitation of the EMT sensor in the experiment, the verification is different from the ground truth. However, this problem can be solved by choosing a more suitable EMT sensor.

The whole system still has many aspects that can be improved, for example, automatic VID collection, optimizations on the design of the robot, the camera, and the light source, further researching the image matching algorithm to improve accuracy or reduce computing time, system performance in non-rigid environments, and applying them in vivo.

Contents

Nomenclature

Abbreviations

Abbreviation	Definition
3D	3-Dimension
BRIEF	Binary Robust Independent Elementary Features
CT	Computed Tomography
CV	Computer Vision
DOF	Degrees of Freedom
EBUS	Endobronchial Ultrasound
EKF	Extended Kalman Filter
EM	Expectation Maximization
EMT	Electromagnetic Tracking
ESWL	Extracorporeal Shock Wave Lithotripsy
fURS	Flexible Ureteroscopy
FOV	Field of View
GM	Graph Matching
ICP	Iterative Closest Point
MDCT	Multi detector Computed Tomography
MIS	Minimally Invasive surgery
NOTES	Natural Orifice Transluminal Endoscopic Surgery
ORB	Oriented FAST and Rotated BRIEF
PCNL	Percutaneous Nephrolithotomy
PSR	Point Set Registration
RCI	Real Calyx image
RCM	Remote Center of Motion
RMSE	Root Mean Square Error
SC	Shape Context
SfM	Structure from Motion
SLAM	Simultaneous Localization and Mapping
STD	Standard Deviation
URS	Ureteroscopy
VB	Virtual Bronchoscopy
VC	Virtual Camera
VID	Virtual Image Database
VURS	Virtual Ureteroscopy

1 Introduction

1.1. Urolithiasis

Urolithiasis, also known as kidney stone disease, has plagued humankind for thousands of years from 4000 B.C. [1], and has become the most common urinary tract disease [2]. According to [3], around 12% of the world population was influenced by kidney stone diseases, regardless of their ages, sexes, and races [4, 5]. The paper published in 2013 [6] shows that kidney stone disease affects more men than women in the 20–49 age group. The prevalence of kidney stone disease is also different among countries. Developed countries have a higher incidence than developing countries [7]. In the US, the overall prevalence reached 10.1% in 2016, while the number in 1980 was 3.2% [8]. The same trend was also found in the U.K. [9]. Kidney stone disease is highly related to genetic and metabolic facts, dietary habits, and even climatic factors [7, 10, 5]. In addition, approximately 50% of the people who once suffered from the disease face the recurrence during their lifetime [11].

Various methods to treat kidney stone disease are developed and concluded according to the clinical situations. When the stones are smaller than 5 mm, active surveillance is the best way because they are likely to pass spontaneously [7]. However, surgical interventions are needed to deal with stones larger than 5 mm, which may cause severe pain and block the ureter. In the very beginning, open surgery was one of the few options. It was not until the 1980s that the minimally invasive approach became the gold standard for reducing the risk of infection, recovery time, pain, and discomfort [12].

Thanks to the development in medical devices, nowadays, there are three main techniques to treat kidney stones: Extracorporeal Shock Wave Lithotripsy (ESWL), Percutaneous Nephrolithotomy (PCNL), and Flexible Ureteroscopy (fURS). ESWL creates ultrasound shock waves to break up the stones into small segments that can be expelled through the urinary system. However, when dealing with the large stones or the stones that obstruct the urinary system and those composed of calcium oxalate monohydrate, ESWL is not a good choice [13]. These stones are more resistant and more challenging to break [14, 15]. PCNL removes the stones by inserting a scope through a small incision from the back of a patient. This method can remove large stones with a high success rate but leaves incisions that cannot be ignored. Unlike the previous two methods, fURS removes the stones with a flexible endoscope inserted through a natural orifice, urethra, and the bladder up to the ureter and the kidney. This approach can effectively remove the large stones and those stones that ESWL cannot break up. At the same time, fURS is safer with less bleeding and can be performed in patients with renal anomalies, morbidly obese, or pregnant [7, 16].

1.2. Ureteroscopy

Ureteroscopy (URS) is a procedure that a surgeon inserts a ureteroscope through the patient's urethra to find the location of the stones in the urinary system. The first ureteroscopic procedures were reported in 1977 [17] and 1978 [18]. Both used a rigid endoscope (A in Fig. 1.1) to do the procedure. However, the rigid ureteroscope cannot access some positions inside the urinary system, especially in male patients [19]. To solve this problem, a flexible ureteroscope was developed. It was the first time in 1964 that Marshall used a passively steerable flexible ureteroscope without a working channel [20]. Then, fiber-optic fURS were developed to avoid damaging tissues when accessing the ureter [20].

Since, in most patients, the diameter of the ureter on the asymptomatic side is 3 mm or smaller [21], the diameter of the ureteroscope should be smaller than 3 mm, which is 9 Fr (French gauge unit system). Nowadays, many manufacturers have proposed their ureteroscope products, like Karl Storz (B in Fig. 1.1) and Olympus (C in Fig. 1.1). Most ureteroscopes are equipped with a digital camera and can deflect around 270 degrees with a tip diameter less than 9 Fr [22]. Surgeons can access more places inside the urine system and get higher image or video quality because of the improvements in those ureteroscope products.

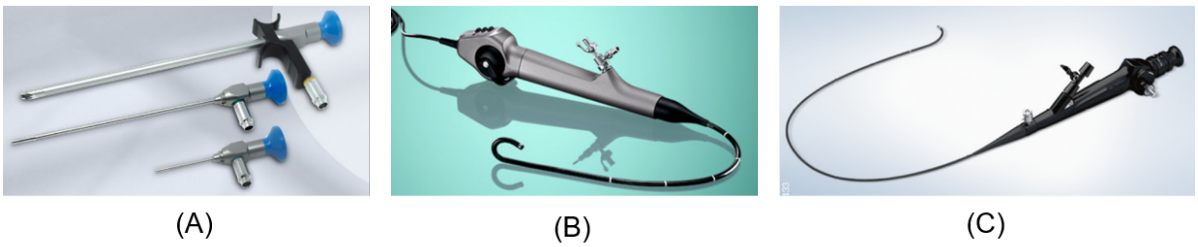

Figure 1.1: Some examples of the ureteroscope product. (A) ATMOS rigid endoscope product [23]; (B) The Flexible Video Cystoscope HD-VIEW™ from Karl Storz [24]; (C) URF-P7/P7R Olympus [25].

1.3. Robotic techniques

When performing the URS intervention, the doctors' dexterity and sensory feedback are limited due to the limited space in the urinary system. In addition, holding a ureteroscope for too long can cause fatigue. Robotic techniques that cover from end-effector structure design to system control algorithms, from visualization enhancement to user-friendly interface improvement, have the potential to address these problems [26]. The most well-known robotic system for the URS is the Avicenna Roboflex System (ELMED™ medical system), which is the first robot system for retrograde intrarenal surgery and fURS and lasertripsy [27]. The system (Fig. 1.2) is based on a master-slave control system, which reduces the working load of surgeons by letting them use joysticks to control the ureteroscope. It is a fact that robotic techniques have greatly improved diagnostic and surgical procedures.

Figure 1.2: The Avicenna Roboflex System overview [27].

Although the robotic fURS has been proven its feasibility by the Avicenna Roboflex System, it still has room for improvement. Because the insertion method of a ureteroscope is through only one natural orifice, it limits the surgeons from doing the precise operation in a narrow and small space. In addition,

because of the size limitation, most ureteroscopes are designed to be monocular. This reduces the quality of the visualization inside the urinary system.

1.4. Clinical problems

One of the main problems when surgeons perform the URS operation on the urinary system, especially for kidney exploration, is to navigate the ureteroscope. More specifically, surgeons find it difficult to know which calyx is being observed and where the ureteroscope tip reaches. The problem is caused by the complex structure inside the kidney with multiple calyxes and the limitation of the ureteroscope. Fig. 1.3 shows the structure inside a kidney, and Fig. 1.4 shows an actual image captured by a ureteroscope inside a kidney. Because of the limited space in the kidney and the limited field of view of a ureteroscope camera, the images captured by the ureteroscope camera, when facing one of the calyxes, only show a dark area that may be similar to other calyxes. This fact makes surgeon hard to tell which calyx is the image represented by only using a ureteroscope. Thus, when doing kidney exploration to find kidney stones, there may be unchecked missing calyxes, leading to incomplete inspection or more effort and time to re-exploration. So the problems can be defined as:

- Which calyx inside the kidney is being observed?
- Where is the location and orientaion of the ureteroscope tip inside the kidney?

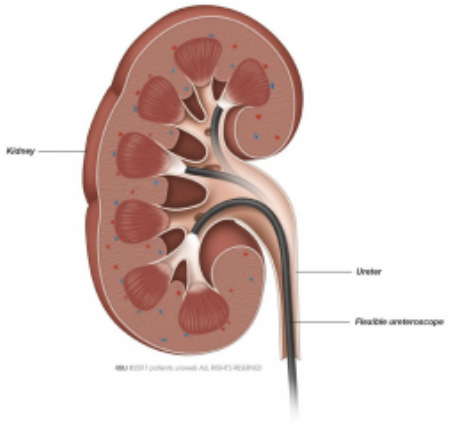

Figure 1.3: An simple visualization to show the inside structure of a kidney [28].

Figure 1.4: An example view form a URS procedure, provided from [29].

These problems can be seen as a navigation problem. According to [30], navigation can be treated as three fundamental competencies: self-localization, path planning, and map building and interpretation. Self-localization means knowing the location of the system itself. Path planning selects a proper route to reach the target place. Map building and interpretation are to know the information from the environment to interact with, which also helps with localization. Using these competencies as a reference, navigation techniques solve two fundamental questions: “where am I” and “how to get there”. In this scenario, the problem in kidney exploration can be transferred to a “where am I” question. Thus, a navigation system for the ureteroscope can solve this problem.

1.5. Goal of this thesis

Trying to solve the kidney exploration problem, this thesis aims to propose an image-matching-based navigation system to help surgeons know where the ureteroscope tip is. The main idea of the system is to match the images from a real ureteroscope camera with the images from Virtual Ureteroscopy (VURS) to get the position information inside the kidney. The system can be separated into two steps:

- *Pre-operation.* This step is doing some preparations for the navigation system. In the pre-operation step, the target kidney of the patient can be simulated as a 3-Dimension (3D) model using CT scan data. The 3D model is rendered to generate VURS, where virtual images inside

the kidney can be achieved. Those virtual calyxes images are the virtual image database (VID) for the matching step. At the same time, the virtual camera (VC) position and orientation are recorded.

- *Post-operation.* After using the ureteroscope to explore the kidney, real calyx images (RCIs) are collected. The most matching virtual calyx image can be found by matching them with the VID collected in the previous step. The system will notify the operator where the ureteroscope tip locates from the VC position and orientation.

The rest of the report is organized as follows. In Chapter 2, related state of the art is discussed, most attention is placed on navigation systems for endoscopes and image matching techniques. Chapter 3 describes the proposed image-matching-based navigation system in detail. Chapter 4 shows the whole experimental activities to verify the system. The experiment results are shown in Chapter 5. Chapter 6 and Chapter 7 present the discussions and conclusion of the whole thesis, respectively.

2

State of the Art

2.1. Navigation system for endoscopes

Various methods are presented to help operators do the endoscopic navigation in different medical applications. These methods can be mainly divided into two categories based on data acquisition methods: Electromagnetic-Tracking (EMT) -based and vision-based methods. EMT-based methods refer to doing the localization based on EMT sensors, usually combined with the 3D model built from preoperative Computed Tomography (CT) images. Vision-based methods mainly used the endoscopic videos or images to localize the endoscope.

2.1.1. EMT-based methods

EMT-based methods are used because of their stability [31] compared with image-based tracking methods. In the medical field, EMT sensors are usually used to estimate endoscope movements. Most of the cases use the navigation technique framework that does the data fusion or registration between models generated from pre-operative CT images and the data acquired from an EMT sensor [32, 33, 34, 35, 36].

Most navigation systems are designed for the bronchoscopy which is a typical application scenario that needs accurate localization because of the complexity of the bronchial tree. After the first human trial of the EMT-based bronchoscopy system presented by Schwarz et al. [37], many efforts were made to register between the EMT tracking system and CT images to improve the trajectory and the localization accuracy of the bronchoscope. Klein et al. [32] produced a fiducial-free registration algorithm by maximizing the fitness between the trajectory of the bronchoscope measured by an EMT sensor and the segmented airways of the CT volume. Although the algorithm achieved higher registration accuracy than common approaches, like point-based registration and iterative-closest-point (ICP)-based registration, it is heavily influenced by its initialization during optimization. To reduce the impact of initial registration between EMT coordinate and CT coordinate, Luo and Mori [33] proposed an adaptive fiducial-free registration method that uses a multiple point selection strategy based on sensor orientation and endoscope radius information. However, both methods cannot deal with dynamic conditions caused by breathing, patient movement, or tissue deformation where the registration accuracy decreases. To alleviate this negative impact of the variant working environment, Hofstad et al. [34] presented an intra-operative local registration method to update the initial registration by optimization in a limited region instead of the global area. The method is still rigid and based on the ICP algorithm with the positions and orientations of the bronchoscope and the direction of the bronchial centerline information. Another non-rigid method proposed by Navaei Lavasani et al. [35] also solves the dynamic problems. They applied a centerline-guided Gaussian mixture model to register the moving bronchoscope position with the static pre-operative CT space automatically in real-time, which preserves both the local and global structures and is formulated and solved under the Expectation-Maximization (EM) framework.

EMT-based methods are also used in colonoscopy navigation systems. Similar to the navigation framework in bronchoscopy, Oda et al. [36] proposed a colonoscope navigation system based on a colonoscope tracking method to guide inexperienced physicians. The method employs the ICP al-

gorithm to register the CT coordinate and EMT sensor coordinate. The method used a line-based registration between the colon centerline generated from CT images and the curved line obtained by the EMT sensor. The algorithm performs the registration by finding the correspondence between landmarks on two lines. A threshold of the distance is applied to eliminate the correspondence between two landmarks that are not closed to each other. The relation between the distance threshold and tracking accuracy was further evaluated by Oda et al. [38]. They changed the distance thresholds for each position in the colon phantom and measured tracking errors at 52 points, finding that decreasing the distance thresholds can improve the colonoscope tracking accuracy. However, these methods lead to large tracking errors when colon deformations occur. Estimation methods of the colon shape can reduce this problem [39].

However, one of the biggest problems of EMT-based methods is that an extra sensor is needed, which adds to the cost of the solutions. In addition, EMT-based methods cannot provide surgeons with intuitive visual results on endoscope locations.

2.1.2. Vision-based methods

Endoscopes are regarded as the eye that enable people to look inside human bodies. Therefore, endoscopy vision techniques are crucial when performing diagnosis and treatments.

One widely discussed topic is how to combine the endoscopic images or videos with the data from other modalities to realize accurate localization. The basic workflow for this combination in bronchoscopy is shown in Fig 2.1. Based on this workflow, Khare et al. [40] presented an image-guided intervention system to help physicians to know the location of the bronchoscopy. By taking the endoluminal virtual bronchoscopy (VB) views generated from the pre-operative multi-detector computed tomography (MDCT) chest scan as references, physicians can localize the bronchoscopy through a global registration algorithm between real bronchoscopic images and virtual images. Byrnes and Higgins [41] improved the workflow by automatically analyzing bronchoscopic video stream information instead of manually using images. The video stream is parsed into segments based on the initial manually marked video frame. Then by using these video segments and a 3D airway tree model from MDCT images, a linkage between key video frames and airway path locations can be generated, and the bronchoscope history path can be created. To improve the localization accuracy, Zang et al. [42] proposed a system that integrates endobronchial ultrasound (EBUS) into the conventional image-guided intervention system to do the data fusion between EBUS, MDCT, and bronchoscopy video. Based on the previous system, Higgins et al. [43] presented a multimodel system for bronchoscopy navigation which combines all modalities for completing lung-cancer staging workflow. The system combines various automatic and semi-automatic image-processing tools for data fusion to improve navigation efficiency.

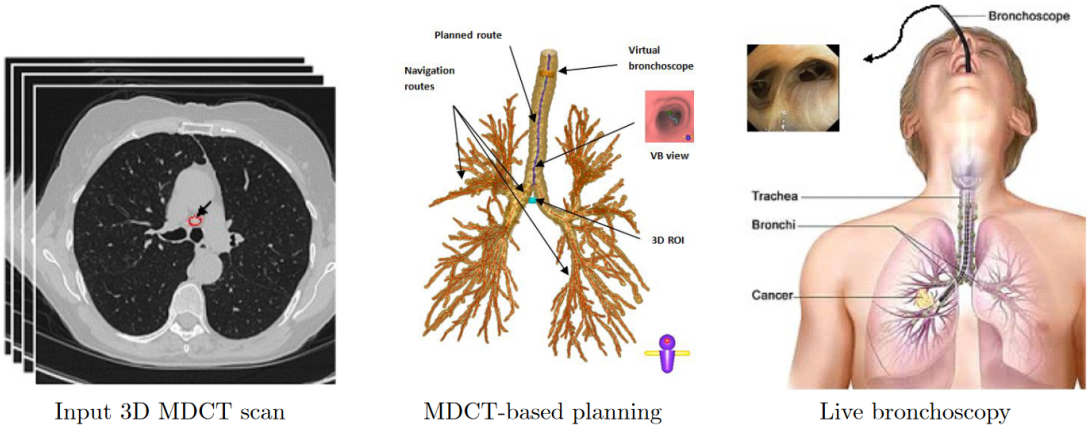

Figure 2.1: A typical work flow of image-guided bronchoscopy navigation from [40]: Firstly, a region of interest is localized from 3D MDCT slices obtained from patient. Secondly, the 3D airway tree, the 3D region of interest, VB endoluminal views, navigation paths, are computed. Thirdly, a live bronchoscope is used with the guidance and registration with VB view.

Recently, machine-learning-based methods have been used to improve the endoscope localization accuracy. Visentini-Scarzanella et al. [44] presented a decoupled deep learning architecture to get the depth information of bronchoscopic images for improving the bronchoscope navigation system. The

method contains two networks. The first one, the RGB transcoder network, takes bronchoscopic images as input and outputs textureless representations resembling CT renderings. The second network maps renderings to corresponding depth information. Shen et al. [45] applied a conditional generative adversarial learning framework to estimate the context-aware depth information through monocular bronchoscopic videos for camera localization. The final poses of the bronchoscope are generated by maximizing the similarity with the depth information from CT scan, based on the method proposed previously [46], which uses normalized cross-correlation and mutual information for the similarity measure. Zhao et al. [47] took the bronchoscopy localization as a learning-based global localization problem and solved it by linking the domain information between the real bronchoscopic videos and VB views through deep neural networks, which also achieved good localization accuracy. The method that used dual convolutional neural networks to generate depth images was validated by Liu et al. [48]. In addition, to deal with the uncertainty of each global pose which indicates the confidence level of the bronchoscope position, the variational inference was used through sampling within the underlying probability distribution.

Endoscopic vision with advanced CV technologies also has been gradually studied. Visual simultaneous localization and mapping (SLAM) is one of the most popular topics, which uses vision sensors as input to do the structure recovery and localization. The feasibility of both stereo and monocular extended Kalman filter (EKF) SLAM was validated in medical scenes [49] [50]. Speidel et al. [51] provided a robust feature tracking method through the Lukas-Kanade method [52] during SLAM for a stereo endoscope pose estimation. A 3D feature-based map is updated with each subsequent stereo pair images captured by the stereo endoscope. During this procedure, unstable features will be excluded from the database which represents the 3D map. Finally, the endoscope pose is estimated by an iterated EKF. Grasa et al. [53] presented a monocular SLAM algorithm to get a 3D map for the cavity and the endoscope trajectory. The algorithm combines EKF monocular SLAM with 1-point random sample consensus to cope with the high outlier rate when exploiting the correlation between pairs of matches in real-time. The method was validated with human in-vivo endoscopic image sequences. Mahound et al. [54] applied Oriented FAST and Rotated BRIEF (ORB) SLAM to estimate the monocular endoscope location and reconstruct a semi-dense map of soft organs. To deal with the practical situation in MIS, for example, illumination changes, poor textures, and small deformations of tissues, they proposed a monocular quasi-dense reconstruction algorithm with a depth propagation algorithm by using keyframe images [55]. The method, however, sacrifices the surface reconstruction accuracy for densifying map points. Chen et al. [56] proposed a robust global 3D surface reconstruction framework for a dense surface, at the same time improving the reconstruction accuracy. The reconstruction framework uses only unorganized sparse point clouds acquired from the SLAM and the moving least squares smoothing algorithm to process the point clouds data set in real-time.

The point cloud generated by SLAM techniques can be used to perform the registration between VB and real airway. Cheng et al. [57] proposed a visual SLAM based method to build a bronchoscopic navigation system (Fig 2.2) by registering two point-cloud models which were generated from bronchus segmentation after the pre-operative CT scan and real bronchoscope images by using ORB-SLAM algorithm, achieving low tracking errors. An improved version of the previous method was proposed to further reduce the tracking errors by constraining the range of candidate frames in a local region to lower the possibility of mismatching [58]. Piccinelli et al. [59] also used scaled ORB-SLAM with the ICP algorithm to obtain a sparse point cloud that can be registered with the pre-operative model of the patient. The final 3D anatomy after reconstruction was used to guide a robotic arm to perform some representative surgical tasks. However, the method only suits rigid scenarios. To deal with the non-rigid reconstruction, Lamarca et al. [60] presented a monocular SLAM approach, DefSLAM. The method performs real-time tracking and mapping for non-rigid scenes by combining both shape from template and non-rigid SfM techniques. However, challenges, including uneven illumination, poor visual texture, and non-isometric deformations still remain.

Most methods mentioned above provide a navigation guide system for an endoscopy procedure where operators still need to compare real images with virtual ones. Although the systems have brought convenience and reliable reference to operators during the endoscopy procedures, they can still be improved by adding the automatic matching process to further reduce the workload of operators.

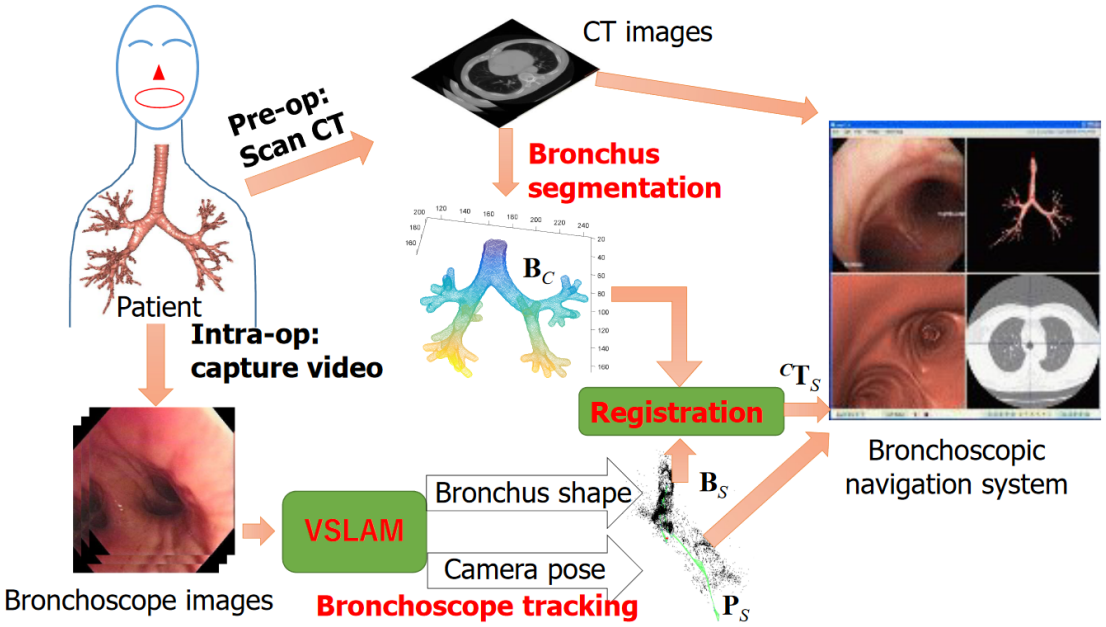

Figure 2.2: A work flow of Visual SLAM based bronchoscopic navigation system from [57].

2.2. Image matching

Image matching is defined as the process that aims to identify and then correspond the same or similar structure/content from two or more images[61]. This technique is widely used in various fields, including pattern recognition, image mosaic, image fusion, image retrieval, and target recognition and tracking. There are two popular frameworks for imaging matching: area-based frameworks and feature-based frameworks [62]. Area-based frameworks use the information of entire images to do the image matching under a similarity metric that measures the accuracy of image alignment. In contrast, feature-based frameworks perform image matching using the features that are extracted from the images. By comparison, feature-based frameworks are more efficient in computation, better at handling geometrical deformation, and more flexible in a wide range of applications, especially when doing multi-modal image matching where images are acquired from different modalities or situations, for example, image-paint and day-night [61, 62].

Feature-based frameworks consist of two steps: feature extraction and feature matching. The former selects and extracts the features from the images according to the application requirement. The features are then transferred into different descriptors, mainly referring to the matching algorithms' data types. The feature matching step uses a specific metric or formula to calculate the similarity between the feature descriptors extracted from both sides of the matching images.

2.2.1. Feature extraction

Various types of information in an image can be treated as features, like color, texture, key points, edges in the images, etc. Since the goal of this thesis is to navigate the ureteroscope in the kidney by focusing on the calyx in the images, color and texture features are ineffective because they have little difference from the calyx to the calyx. These two features are better choices when detecting the lesions on the surface of the tissues. Therefore, how the calyx looks like, in other words, the geometric features of calyxes, are more suitable for this scenario. The geometric features can be generally divided into points feature and edge features.

Points feature

Points features are the most popular because of their easy extraction, simple description, and fast computation. The common idea of the points features detection is to treat the interest points or critical points as helpful information to represent the images, for instance, intersections or a high curvature point of an edge.

The earliest points feature detector can be traced back to the paper published in 1977 [63] by Moravec, where they treated corners as the interest points and recognized them by using the auto-correlation of the local intensity. This method measures the similarity between the patch centered on each pixel in the image and its nearby patches in eight directions. The similarity is calculated by taking the sum of squared differences between the corresponding pixels of two patches. The corner is detected because of its large intensity variation if the similarity is larger than a given threshold. However, this method is not invariant to the image rotation. Harris et al. [64] introduced a corner detector invariant to orientation and illumination to solve this problem. It uses a two-order moment matrix or an auto-correlation matrix to find the grey value changes. The method has reliable repeatability and distinctiveness. Another famous corner points detection method is the FAST detector [65]. It assumed that the image intensity change of corner points should be large in all directions. By calculating the corner response function, a minimum change of the intensity for a pixel among all directions can be known, which is used to determine whether it is a corner point. This method is efficient in finding corner features. Rublee et al. [66] proposed an integrated feature detector and descriptor, ORB, which is built on the FAST key-point detector and the Binary Robust Independent Elementary Features (BRIEF) descriptor to combine both reliability and efficiency. The method is also rotation invariant and can achieve good matching performance under noise with low computational cost.

Besides the corner feature detectors, blob feature detectors are also widely used because of their accuracy, and robustness [61]. The famous SIFT detector [67, 68] is one of the blob feature detectors. It uses the difference of Gaussians, filtered by the Hessian matrix of the image intensity values, to find the local extremum point with and detect the final stable scale-invariant blob features. An improved blob feature detector, SURF, was proposed by Bay et al. [69] which makes the SIFT more efficient by approximating the Hessian matrix using Haar wavelet calculation, combining an integral image strategy. By using this method, the computational time for constructing a second-order differential can be reduced.

In endoscopy applications, these points feature detection methods are usually used in visual SLAM and 3D reconstructions, where key points in every consecutive video frame are extracted and used to calculate the endoscope trajectories or to do the image stitching. In these scenarios, processed images are all from the same modality, the real endoscope camera. At the same time, the color and texture information of the tissue is also taken into consideration. These conditions allow point feature detection methods to succeed. However, this study uses images from different modalities, an image sensor, and rendered VURS. In addition, since localizing the endoscope tip and recognizing the calyxes are the main goals, the color and texture information of the tissue are not valuable information. According to Jiang et al. [62], feature-based matching may be totally failed in multi-modal image matching in some scenarios because of the nonlinear intensity variance of the images. A more robust type of feature with more information should be used in this study.

Edge feature

Since every shape of the calyx hole is different, when doing the calyx image matching from endoscopy, the edge feature is a better choice because it has more geometry information for every calyx, improving the matching accuracy. The depth difference between the tissue walls inside and outside calyxes can be reflected in the endoscopy image from the strong intensity contrasts between adjacent pixels. Therefore, using edge detection techniques, those calyx shapes can be extracted and then described by point sets. The edge detectors can be divided into two categories: Gradient and Laplacian [70, 71].

The earliest Gradient edge detector is Roberts edge detector [72], which uses discrete differentiation to calculate the addition of squares of the differences between adjacent pixels along the diagonal in an image. It uses 2-by-2 kernels (Fig. 2.3) for both vertical direction and horizontal direction. However, this detector is inaccurate and very sensitive to noise. In 1970, Prewitt edge detector [73] was introduced. The detector uses 3-by-3 kernels (Fig. 2.4), unlike the ones that the Roberts edge detector uses. Therefore, the Prewitt edge detector performs better in detecting the edges in horizontal and vertical directions. Another famous edge detector, Sobel edge detector [74], was first introduced in 1968. The detector also uses two 3-by-3 kernels (Fig. 2.5) to calculate the gradient approximation of the image intensity function to detect edges. It combines Gaussian smoothing and differential derivation and introduces the concept of weight to consider the different influences between adjacent and distant pixels. Therefore, the Sobel detector is more accurate than the previous two detectors and can deal with the noise.

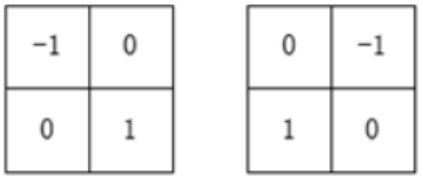

Figure 2.3: 2-by-2 kernels for the Roberts edge detector.

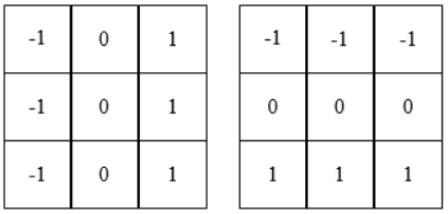

Figure 2.4: 3-by-3 kernels for the Prewitt edge detector.

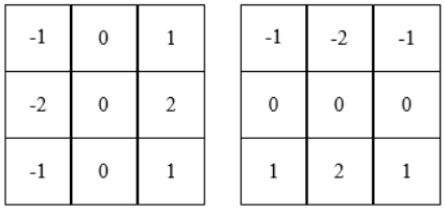

Figure 2.5: 3-by-3 kernels for the sobel edge detector.

Another edge feature detector is the Laplacian detector which calculates second-order derivations in an image to detect edges. Laplacian of Gaussian [75] is one of these edge detectors. It works well when there is a sudden change in the transition of the grey level. The maximum is found when the second-order derivative crosses zero, which means the location can be seen as the edge. It has the same performance in all directions when detecting edges. However, it is sensitive to noise and usually is applied after using a Gaussian Smoothing filter to reduce the image noise. Canny edge detector [76] is another famous Laplacian edge detector. It considers two thresholds to detect strong edges and weak edges. Strong edges refer to the points with dramatic changes in grayscale. Weak edges are the points with a slight change of grayscale. The second derivative zero crossing point calculates both strong and weak edges. This detector is widely used because of its robustness and accuracy. At the same time, it can extract the edges without affecting or changing those features.

The canny edge detector is also popular in endoscopy image processing. Kang and Doraiswami [77] proposed a edge-extraction based real-time image processing system. In this system, the Canny edge detector is used to do the segmentation to make the polyp identification. Hafner et al. [78] also used a canny edge detector to do the image segmentation and automatically classified these segments to help colon cancer detection. Angiectasia of the small bowel and bleeding small bowel can also be detected using the Canny edge detector, which is reported by Alizadeh et al. [79]. Therefore, the canny edge detector is used in this thesis to extract the features of calyx images.

2.2.2. Feature matching

When doing feature matching, the feature points are extracted from the images on both sides and used to do the matching. The geometrical spatial relations, calculated by optimization methods, of these feature point sets are usually criteria for judging their correspondences or similarities. The methods that generate the geometrical relations between two feature point sets can be classified into graph matching and point set registration [61].

Graph matching

Graph matching (GM) methods define nodes and edges in each point set to generate a graph correspondingly. For example, there are two graphs $G_1 = \{V_1, E_1\}, G_2 = \{V_2, E_2\}$, where V_1, V_2 are the vertex or node sets representing feature points, and E_1, E_2 are the edge sets consisting of all edges over a pair of nodes. The number of the nodes in G_1, G_2 are represented by n_1, n_2 respectively. Each node is noted as v_i for V_1, or v_j for V_2. The overall affinity relationship between two graphs, which maximizes the graph structure similarity, can be generated using various methods. Then the node-to-node correspondences between two graphs representing two point sets can be established [61]. GM usually treats point set matching as a quadratic assignment problem [80]. Most of the studies are based on Lawler's quadratic assignment problem, whose formulation of GM can be written as:

$$\max J(\mathsf{X}) = vec(\mathsf{X})^T \, \mathsf{K} \, vec(\mathsf{X}) \quad s.t. \mathsf{X}_{ij} \in \{0, 1\}, \tag{2.1}$$

where X is a permutation matrix, i.e. $X_{ij} = 1$ indicates that the node $v_i \in V_1$ corresponds to the node $v_j \in V_2$, $vec(X)$ denotes the column-wise vectorization of X. K is the affinity matrix encoding the nodes and edge affinities between two graphs. Then the matching is formulated as an optimization-similar problem that finds the optimal node-to-node correspondences X to maximize the overall affinity score $J(X)$. Based on that assumption, many solvers are presented. Most of them try to relax stringent constraints and provide approximate solutions [61]. Some of them regard GM as the eigenvector problem[69, 81], others focus on convex relaxations [82, 83] and convex-to-Concave relaxations[84, 85]. Recent years, learning based methods [86, 87] are also used to solved the GM problems.

Although GM has the potential to find the proper point correspondences by optimizing the corresponding overall matrix, it is limited because of its high computational cost and outlier sensitivity.

Point set registration

In most cases, point set registration (PSR) aims to find a global transformation relation between two point sets to achieve point-to-point matching by using iterative estimations.

One of the most famous methods is ICP [88]. The main idea of ICP is to perform a point-by-point matching between two point sets. Since there are many possibilities for the correspondence of points, ICP uses an iterative method to continuously estimate the rotation and translation between two point sets by minimizing the root mean square point-to-point distance metric based on the previous matching result until convergence. It is widely used because of its simplicity. However, the performance of the method heavily depends on the initialization, which may lead ICP to local optima. There are many variants of the ICP algorithm to improve its performance. For example, Granger and Pennec [89] integrate the expectation maximization (EM) principles with ICP (EM-ICP) to solve the matching problems, which are formulated into a maximum-likelihood estimation of the transformation and the matches. Compared with the standard ICP, EM-ICP is more robust and time-saving, maintaining similar performances in precision. Another famous variant is LM-ICP proposed by Fitzgibbon [90], which uses a nonlinear optimization method, the Levenberg– Marquardt algorithm, to minimize the registration error directly. The method also improves the robustness and computational efficiency. Robust point matching algorithm [91] is another method to perform the point set matching. It uses deterministic annealing, and soft assignment of point sets correspondences to perform a soft registration. Unlike the ICP, where the matching result of two points is either true or false, the robust point matching algorithm defines a soft correspondence, ranging from 0 to 1, to present the matching degree of two points, and finally will converge to 0 or 1. This method makes the matching more adaptive, flexible, and robust to those noises, spurious and missing points, and outliers.

Although the methods that use only points to do the matching are simple and efficient, their performances concerning accuracy are not satisfying in this thesis, because of the large nonlinear intensity variance between real images and VURS images, and non-negligible edge deformations of the calyxes under the different angle of view. In this case, a different method that is robust to the shape deformation is needed.

Shape context (SC) matching [92] is one of the candidates. This method proposes a SC descriptor to characterize the relative positions and orientations between a reference point and other edge points. Fig. 2.6 shows the details of the SC matching. (a) and (b) are two edge point sets, representing two different shapes of the letter “A”, where an obvious shape deformation exists. (c) is a diagram of the log-polar histogram bin used to compute the SCs. Since full sets of vectors contain too much information and will be greatly influenced by the sampled points, point distribution is used to present a compact descriptor. The bin in (c) is then used to calculate the points distribution relative to a reference point. The bin is uniformly divided in terms of direction and for distances, is uniform in log-polar space in order to make the descriptor more sensitive to the points which are closer to the reference point. (d), (e) and (f) show the three SC histogram examples of the marked reference points in (a) and (b), circle, triangle, and diamond, respectively. Those histograms are based on a log-polar coordinate, which shows relative distance and direction relations between the rest of the point set and the reference point. From the visualization, the histograms of (d) and (e) are similar, which is consistent with the similarity of the location of both reference points in the respective shape. By contrast, the histogram in (f) is quite different from (d). (g) shows the matching between two shapes. The similarity of two shapes can be treated as the distances of SC descriptors.

In this thesis, the edges that need to be matched are from different modalities. Therefore, the edge of the same calyx differs from the two modalities in both the shape and number of the points. These

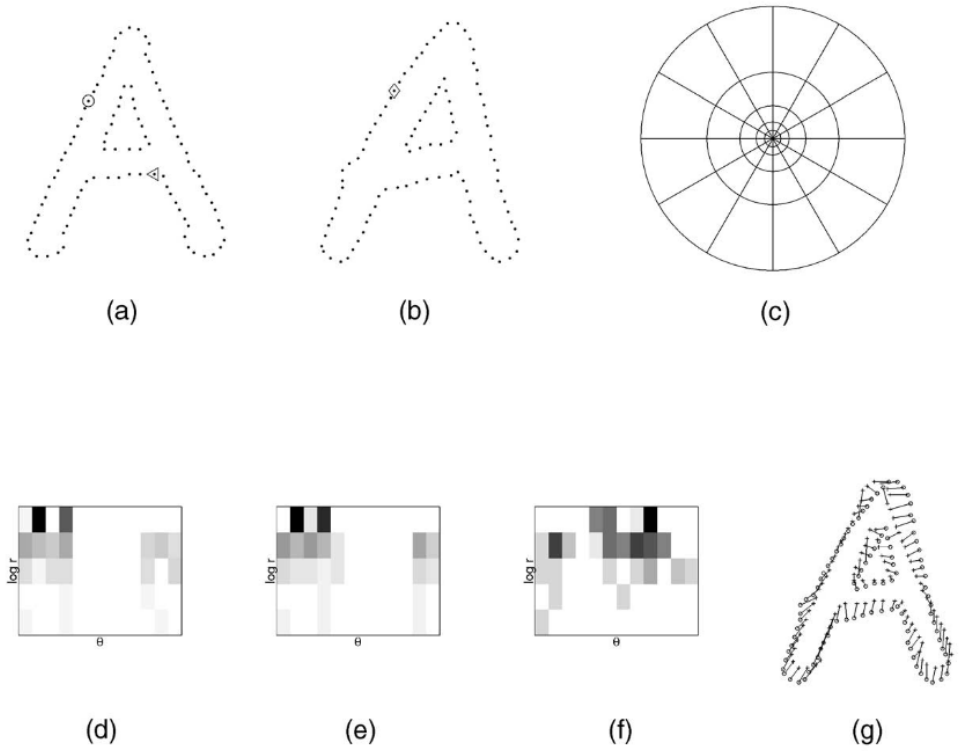

Figure 2.6: SC computation and matching from [92]. (a) and (b) are two point sets representing the contours of letter “A” with different shapes; (c) is the log-polar bin used to calculate the SC; (d), (e), (f) show the histograms of the marked reference points in (a) and (b), circle, triangle, and diamond; (g) matching result between two shapes.

facts result from the difference between camera poses, illumination conditions, and image qualities. Since the SC matching is invariant to translation, scaling, rotation [92], and robust to the deformations, the method is used in this thesis to do the image matching.

3

Materials and Methods

3.1. Introduction to the solutions

fURS is one of the choices for diagnosing and treating kidney stones because it is safe for people of all ages. During the intervention, one problem is exploring the kidney thoroughly. More specifically, knowing which calyx in the kidney is being observed and which calyxes have been observed. Currently, surgeons only use a ureteroscope to explore the kidney, which is difficult due to the limited view.

For this reason, this thesis proposes a novel image-matching-based navigation system that helps surgeons know the approximate location of the ureteroscope and the calyx they are examining. The navigation system performs an image matching algorithm between the real URS and VURS images to navigate the real ureteroscope tip from the VURS environment. As shown in Fig. 3.1, the system framework consists of two stages: pre- and post-operation.

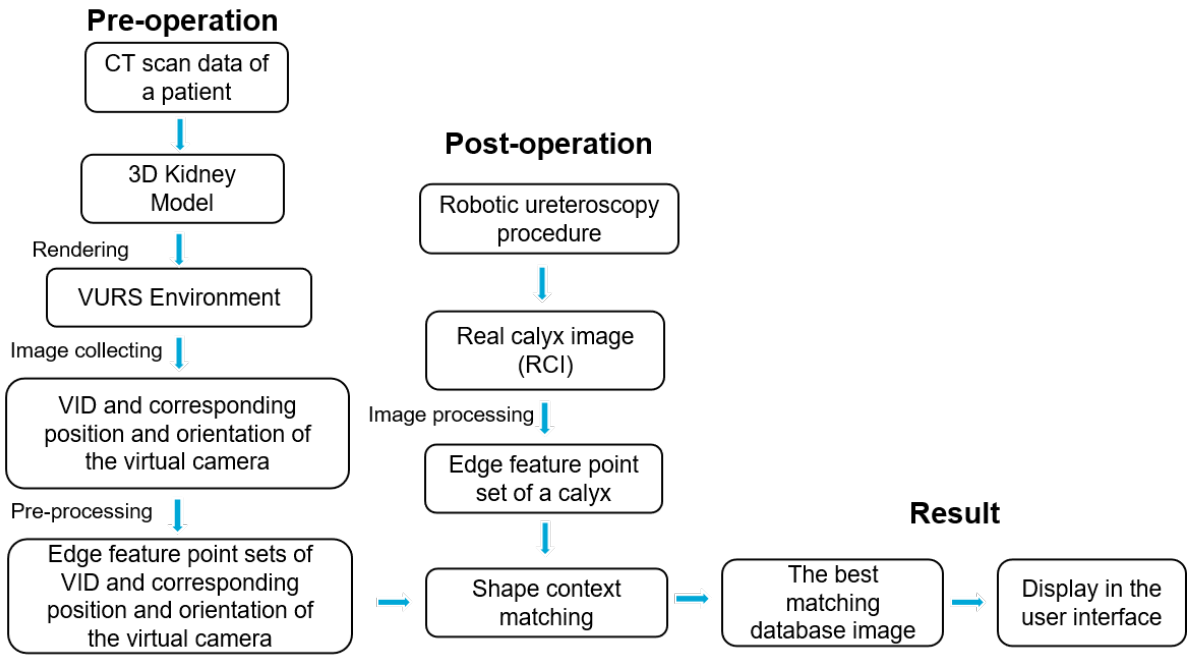

Figure 3.1: The proposed navigation system framework overview.

Pre-operation: Firstly, a 3D kidney model of a patient can be generated by using CT scan data. Secondly, the 3D kidney model is used to render a VURS environment to simulate the URS procedure. Thirdly, VURS images of different calyxes in the kidney model taken by the VC are collected with the corresponding position and orientation data of the VC. Finally, those VURS images are pre-processed to extract their edge features and be described in point sets as the VID for the matching algorithm.

post-operation: Firstly, RCIs inside the kidney are acquired through a robotic ureteroscope. Secondly, edge features are extracted from these images as the input for the matching algorithm. Thirdly, the best matching calyx image from the VID can be found after performing the SC matching algorithm. Finally, according to the image's corresponding position and orientation data of the VC, the location of the real ureteroscope can be inferred, which can be shown in the VURS environment through a user interface.

However, because of limited resources and time, acquiring actual CT data from a patient is not possible throughout the thesis period. An open-source 3D kidney STL model [93] (Fig. 3.2) is used to replace the data from CT scans to verify the feasibility of the system framework. At the same time, the robotic URS procedure has to be performed in a real kidney phantom based on the 3D kidney model with a cable-driven robot prototype.

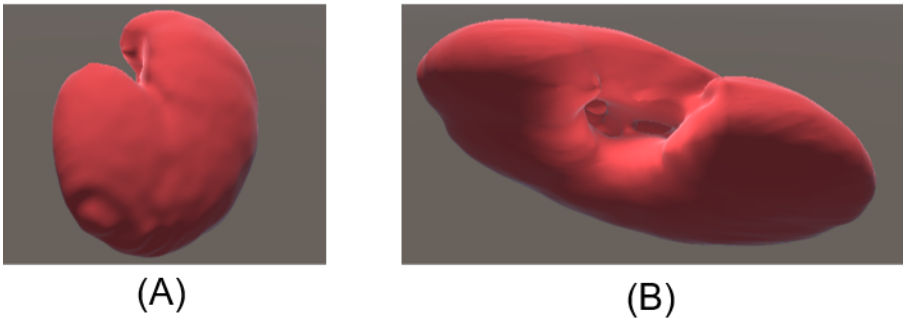

Figure 3.2: (A) Front view of the kidney model; (B) Left view of the kidney model.

The rest of this chapter explains Kidney phantom generation, VURS rendering, VID collecting, Image processing, and Imaging matching algorithm.

3.2. Kidney phantom

To realize the navigation system, a kidney phantom is needed. Since this thesis only considers rigid conditions without tissue deformations, 3D printing technology is used to fabricate the kidney phantom because it is easy to implement and is suitable for fabricating complex structures.

3.2.1. Devices and software

Form 3B+ (Fig. 3.3) from Formlabs is used to do the 3D printing. The product is ideal for anatomical models and medical device prototypes with high printing precision [94]. Model V2 Resin (Fig. 3.4) is chosen as the printing material for the kidney phantom.

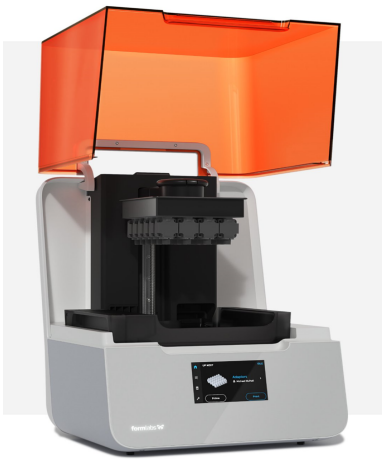

Figure 3.3: Form 3B+ 3D-printing device from Formlabs [94]. **Figure 3.4:** Printing material, Model V2 Resin form Formlabs [95].

Considering the size of the robot prototype, whose body diameter is larger than the entrance of the kidney calyxes in its original size, the scale of the 3D-printing kidney phantom is set to 1.8 times the original. This makes the printer not possible to print the whole 3D phantom at once. Therefore, the phantom is divided into two parts and printed separately. There are two ways to divide the kidney model into two parts: horizontal (follow the red line in Fig. 3.5) and vertical (follow the blue line in Fig. 3.5). The vertical way is selected to reduce the influence on calyxes. Since there are inevitable errors on the surfaces of two printed parts, there is always a seam (for example, it is shown by B in Fig. 3.7) between the two parts when connecting them, which will significantly influence the later image processing and matching. The seam will cross almost all the calyxes if two parts are divided horizontally. While only one calyx is influenced when using the vertical way.

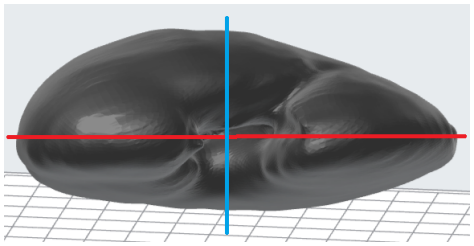

Figure 3.5: Two possible ways to divide the kidney model. The red line divides the phantom horizontally, while the blue line divides the phantom vertically. Since calyxes mainly distribute along the red line, dividing the phantom in the vertical way influences less calyxes.

The separated STL files of the 3D kidney model are imported into the 3D printing software PreForm, which is also produced by Formlabs [96]. The printing setup for both parts are shown in Fig. 3.6.

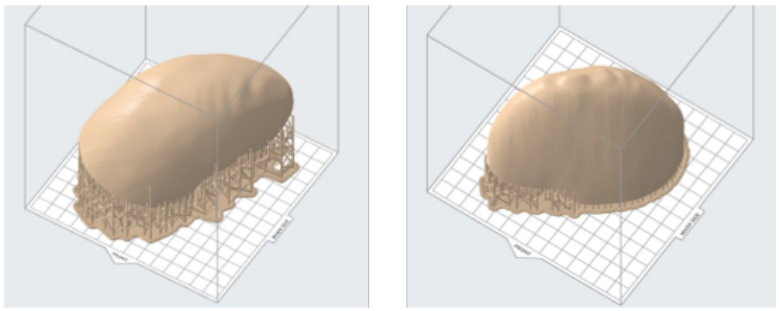

Figure 3.6: 3D printing setups generated from a 3D printing software, PreForm, for two parts of the phantom separated from the vertical way. The left image shows the part on the left side of the blue line in Fig. 3.5, and the right one corresponds to the other side.

3.2.2. Finished phantom

After removing the supports of the printed parts and polishing the connecting surface, those two parts can be connected into one phantom. Since model accuracy is not important in this thesis, the tape is used to fix and connect those two parts. The final phantom that is prepared for the experiment is generated (Fig. 3.7).

(A) (B)

Figure 3.7: (A) Top view of the kidney phantom; (B) Front view of the kidney phantom.

3.3. VURS rendering

VURS is a method to simulate real ureteroscopic views to help surgeons get familiar with the inside structure of the urinary system. The same principle was widely used in bronchoscopy and colonoscopy. In this thesis, a VURS for the kidney is developed from an open source 3D kidney STL file by using Unity software. There are three main goals to generate a VURS:

- Collecting VURS images as the VID for image matching with the RCIs.
- Providing the position and the orientation of the VC, which can be used to locate the ureteroscope tip.
- Showing the intuitive ureteroscope tip location result inside the kidney.

3.3.1. Unity software

Unity is a software that is developed for a game engine where people can create 3 Dimensional scenes and interactive simulations and other experiences [97]. Due to this property, Unity has been extended into various fields, such as film, engineering, architecture and geography [98]. In 2020, Unity took 61% of the game engine market share among the surveyed developers [99]. The number of the Unity creators increased 31% in 2021 [100]. In this thesis, Unity is selected to generate the VURS because of its ease of use, rich resources and active community to support new developers. The editor version of Unity in this thesis is “2020.3.7f1”.

3.3.2. VURS generation

The STL file of the 3D kidney model was imported into the unity software. Because color and texture information is unimportant in this thesis, a standard material with a similar color to tissues was used to render the model (A in Fig. 3.8). The model size in the VURS environment is 10.8 times larger than the original size, which is six times the actual phantom size. This is because, in the Unity environment, the camera cannot see geometry if the distance between the camera and the object is less than the distance of the near-clipping plane. Since the original kidney size is relatively small for the Unity environment in meters, the camera is prone to get too close to the object and cannot see the geometry in front of it. This will cause much trouble to the work of collecting pictures later. Thus, the model size is enlarged.

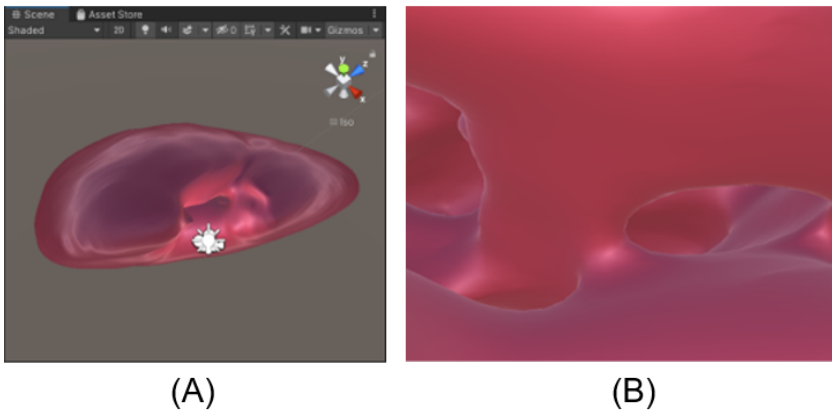

Figure 3.8: (A) The overview of the VURS environment, where the white icon represents the VC; (B) An example view from the VC in VURS environment.

A VC was created and used to collect the VURS images as the VID for calyxes. By using scripts, the position and orientation of the VC can be controlled by a gamepad. With the position and orientation changes of the VC, the view captured by the camera can cover all calyxes from different angles of view (an example is shown by B in Fig. 3.8).

3.4. VID collection

This step generates the virtual images of all calyxes inside the kidney in the VURS environment, which is then processed to be the VID for the matching algorithm. Although there are many calyxes inside this

kidney model, eight of them are taken into account in this thesis workflow. This is because the design of the robotic ureteroscope in the experiment limits its range of motion, allowing the robot to observe only eight calyxes.

Since the VID is where all calyx images are matched with the RCI to get the most similar calyx image result, the image quality should be guaranteed. In other words, the images in the VID should show the edge information of calyxes as clearly as possible and are easily processed for edge feature extraction. Therefore, no light is added when collecting the VID to eliminate the effect of light sources. Fig. 3.9 shows those eight VID images of different calyxes without light sources.

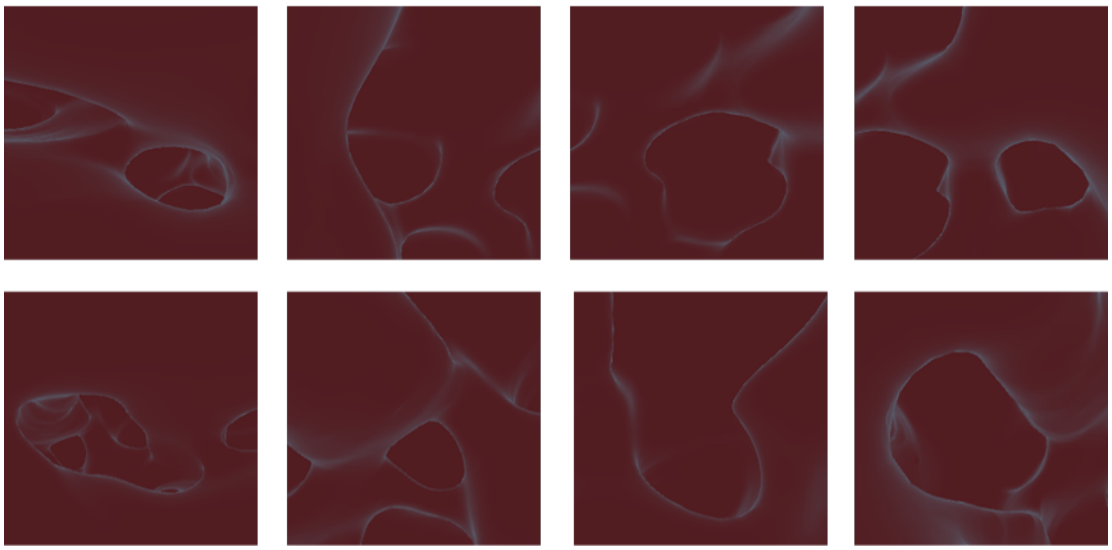

Figure 3.9: The sample images in the VID, captured by the VC in VURS environment. Each image represents different type of calyx. Eight types of calyx in total in the VID.

Considering that the shape of the calyx edges may change because of the ureteroscope camera angles, five images are taken in different angles for every calyx. Therefore, the number of images in the VID is 40. Taking the bottom right calyx in Fig. 3.9 as an example, five images are presented in Fig. 3.10. The difference between those five images is insignificant because the limitations of the robotic ureteroscope used in this thesis are considered. In the experiment, significant orientation changes of the robot tip can hardly be achieved due to the motion limitation of the cable-driven robot caused by large bending radii. Thus, the camera angles for capturing one calyx are about the same.

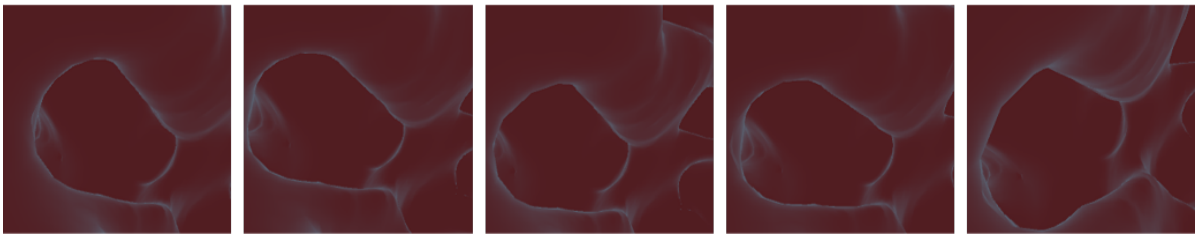

Figure 3.10: Five different images for the same type of calyx (bottom right calyx in Fig. 3.9). All images are in the VID, captured by the VC in VURS environment with different angle of view.

At the same time, the position and orientation data of the VC in the VURS environment corresponding to every image are also recorded in a “CSV” file. The position data consist of three values: x-axis, y-axis, and z-axis in the VURS coordinate. The orientation data are stored as quaternions. These data are used to judge whether the matching result is correct or not by being compared with the data from the EMT sensor after registration.

3.5. Image matching algorithm

The image matching algorithm calculates the similarities between all images in the VID and the RCI. By comparing the similarity values, the smallest value means the most matching image with respect to a particular RCI. The algorithm consists of two steps: feature extraction and feature matching.

3.5.1. Feature extraction

Before performing the feature matching to find the best matching image from the VID, edge features should be extracted not only for VID images but also for RCIs that are collected from the real ureteroscope camera inside the kidney phantom. Unlike the images from the VID, the RCIs have lower quality and more noise. Therefore, pre-processing is needed before extracting the edge features for RCIs.

Pre-processing

The RCIs are collected from the actual camera inside the kidney phantom. Therefore, the quality of those images are limited by the camera performance, illumination from the light source and the surface condition inside the phantom. Pre-processing is needed to make the edge feature extraction more accurate, including noise removal, contrast enhancement, and highlight removal. The result for each step of the pre-processing for an example of RCIs is shown in Fig 3.11.

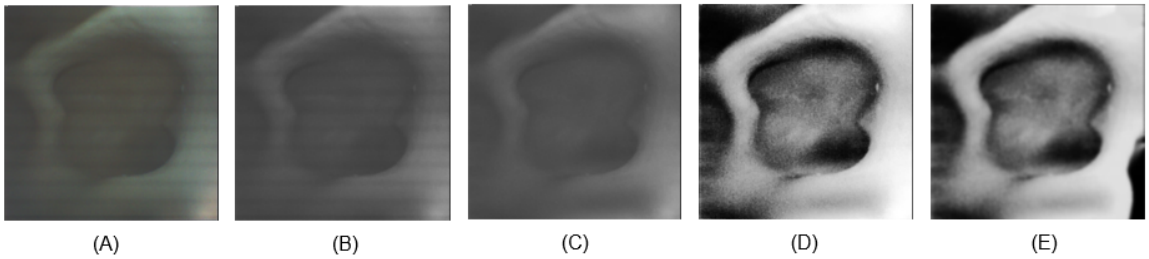

Figure 3.11: The resulting images in each step of pre-processing. (A) A raw image of a calyx taken from the real kidney phantom; (B) The image after converting to the grayscale; (C) The image after horizontal stripes removal using Fourier transformation; (D) The image after contrast enhancement using histogram equalization; (E) The image after highlight and noise removal using Gaussian blurring.

Picture A in Fig. 3.11 is an original RCI of a calyx. From this example image, several problems in RCIs can be found. Firstly, there are apparent horizontal stripes all over the image, which is caused by the diffraction of the light source, a LED light. This is inevitable in this experiment because of the limited relative position between the light and the camera. The detail will be explained in Chapter 4. Secondly, the contrast of the image is low, which adds difficulty to edge feature detection later. Thirdly, the camera noise influences the image to be grainy. This may lead to edge detection errors where the algorithm treats those intensity differences caused by noise as edges. All processing steps are explained as follows:

- Converting the original image into grayscale (picture B). This step reduces the dimension of image data from 3 dimensions (images in RGB) to 2 dimensions, making later processing easier and faster.
- Eliminate the horizontal stripes by Fourier transformation (picture C). Since horizontal stripes are regular noise, they can be easily recognized in the frequency domain. According to [101], those horizontal stripes can be removed by setting their values to zero in the frequency domain.
- Performing the equalization of the histogram [102] to strengthen the contrast (picture D). The high contrast of the image can make those edge features easier to be extracted by increasing the intensity difference.
- Using Gaussian blur to reduce the noise (picture E). Through this step, the image is smoothed, avoiding detecting fragmented edges caused by noise.

The aforementioned pre-processing flow is implemented on the RCIs. It is not necessary for the images from the VID to go through these steps because they are of high quality and noiseless. Therefore, converting them into grayscale images is the only step in pre-processing VID images.

Edge extraction

After pre-processing, images are ready for edge feature extraction. The Canny edge detector [76] is used to extract features because of its robustness and accuracy. Two thresholds need to be considered when performing the Canny operator, “highThreshold” and “lowThreshold”. The “highThreshold” determines strong edge pixels, which refers to the pixels whose gradient value is higher than the “highThreshold”. If the gradient value of a pixel is smaller than then “highThreshold” and larger than the “lowThreshold”, that pixel is seen as a weak edge pixel. Strong edge pixels are used to extract those noticeable edges and are treated as the base of the edge features. Weak edge pixels are mainly used to smooth the strong edge features or make the those features continuous.

Although in the pre-processing stage, grayscale value histograms of all RCIs are equalized, the distribution of the values of each image is still different. Thus, different thresholds of the Canny operator for different RCIs are needed. There are many studies discussing the adaptive thresholds for the Canny edge detector [103, 104, 105, 106]. One of them is by using the Otsu method to decide the high threshold. The low threshold is set around one-third of the high threshold according to the suggestion in OpenCV documentation [107]. Otsu method determines the threshold by minimizing intra-class intensity variance or by maximizing inter-class variance [108]. All edge extraction steps are as follows:

- Using the Otsu algorithm to get a threshold value ac coding to the RCIs.
- Using Canny edge detector to extract the edge information. The high threshold is the value calculated by the Otsu algorithm, and the low threshold is set to one-third of the high threshold.
- Eliminate the small-area distinct edges by setting a threshold value. If the area of the pixels which belong to some distinct edge is smaller than the threshold value, this kind of edge will be ignored.
- Doing the dilation to make key edges continuous and apparent (picture A in Fig. 3.12).
- Extracting the skeletons of those edges to choose main features (picture B in Fig. 3.12).
- Transferring the edge image into points set (picture C in Fig.3.12). The point set can be generated by getting the image coordinates of the pixels whose values are not zero.

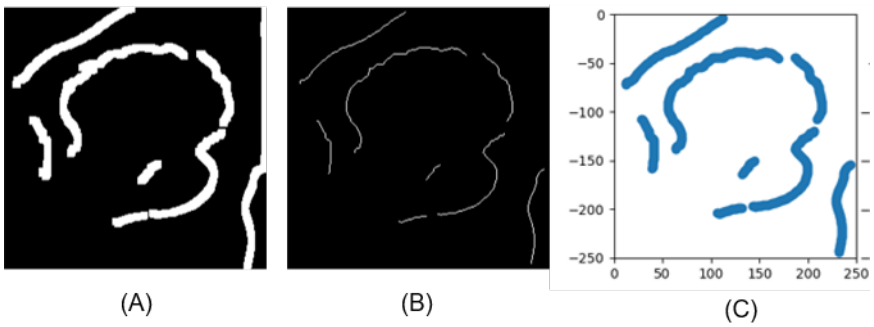

Figure 3.12: Images that represent the edge feature detection process. Based on the Fig. 3.11 (E), edge features can be extracted using Canny edge detector. (A) The dilated edge feature image after using Canny edge detector; (B) The edge feature image after skeletonization; (C) The plot of the final edge points set that is generated by getting the nonzero pixels and is used to do the SC matching.

Those steps are performed on VID images in advance to save computational time. The edge features data of all the VID images are saved as point sets in a “txt” file. Before performing the feature matching algorithm, those data will be read from the file and transferred to arrays type.

3.5.2. Feature Matching

This step is to find the most matching image from the VID to the RCI. SC matching algorithm is selected to perform the matching process because it contains the relative positional and rotational relation information among points, improving the robustness when facing shape deformations. The whole workflow of the image matching algorithm is shown in Algorithm 1.

Firstly, loading the point sets of all the VID images from the “txt” file (pts_db_all) is processed in advance. Secondly, processing the RCI (img_input) and extracting the edge features (explained in the ’Feature extraction’ section), which are then described by a point set (pts_input). Thirdly, each

Algorithm 1 Feature matching algorithm

Require: pts_db_all, the edge point sets corresponding to VID images
Require: img_input
  $dist\_list \leftarrow []$ ▷ An empty list to save the distances
  $pts\_input \leftarrow$ feature_extraction(img_input)
  for each $pts\_db \in pts\_db\_all$ **do**
    $dist \leftarrow$ compute_shape_context_distance(pts_input, pts_input)
    $dist\_list$.append($dist$)
  end for
  $idx\_result \leftarrow dist\_list$.index(min($dist\_list$)) ▷ Find the index of the point set with the minimum

point set in the VID is used to do SC matching with the point set of the RCI. The distances between SC descriptors from the RCI and VID images are saved as a list ($dist\_list$). Finally, the VID image corresponding to the smallest distance in the list is regarded as the result that best matches the RCI.

3.5.3. Result validation and performance metrics

The resulting image in the VID from the image matching algorithm will be displayed. Although through the ground truth, provided by the author, the correctness of the result can be validated, another method need to be applied to deal with the situation without knowing the ground truth. Therefore, the EMT sensor will be used to verify the correctness of the result further. This validation method is based on the assumption that when the VC in the VURS environment and the real URS camera acquire the same calyx image in the same coordinate system, their positions and orientations (pose) are similar or have relatively small errors.

The EMT system can provide the pose information of the EMT sensor in the electromagnetic field. By performing registration between the EMT coordinate and the VURS coordinate, the pose data obtained by the EMT in the EMT coordinate can be transformed into the VURS coordinate. At the same time, each VID image has the pose data of the corresponding VC in the VURS coordinate. The result can be validated by comparing the pose information of the EMT sensor and the VC corresponding to the resulting image in the VURS coordinate. However, in this thesis, only position data are used to do the comparison because the selected 5 degree-of-freedom (DOF) EMT sensor cannot provide the rotation around z-axis.

For better understanding, there are two points: ${}^{E}p_{in}$ representing the real input EMT sensor position in EMT coordinate. ${}^{V}p_{db}$ represents the VC position corresponding to the resulting image from the VID in the VURS coordinate. Let ${}^{V}\mathrm{R}_{E}$ and T represent the rotation matrix and the translation vector that can transform the EMT coordinate to the VURS coordinate after registration. Thus, the actual input EMT sensor position in the VURS coordinate can be written as:

$$ {}^{V}p_{in} = {}^{V}\,\mathrm{R}_{E}\,{}^{E}p_{in} + \mathrm{T}. \tag{3.1} $$

Theoretically, whether in the VURS environment or the real experiment, the camera positions corresponding to the same calyx pictures should be similar in the same coordinate system. Of course, it is unrealistic to match the coordinates of the two camera positions completely, considering that there may be differences in the camera field of view, phantom processing and splicing errors, position measurement errors, and registration errors. However, the result can be judged by calculating whether the distance between two points in the same coordinate system is less than a threshold. This can be formulated as:

$$ \sqrt{|{}^{V}p_{in} - {}^{V}\,p_{db}|^{2}} \leq \epsilon, \tag{3.2} $$

where ϵ represents the threshold, which will be decided according to the error accumulation in the experiment.

4

Experiments

4.1. Simulation Images for matching algorithm testing

In order to test the feasibility of the algorithm before applying it in real situations, some simulation images are used. The specific operation is as follows: change the color of the rendering material of the kidney model in the VURS environment, and collect a virtual image similar to the RCI through the VC with the point light source as the input image for matching with the VID. At the same time, these simulation images will be added with gaussian noise to mimic the images from the real camera. Fig. 4.1 shows the simulation image examples of each calyx.

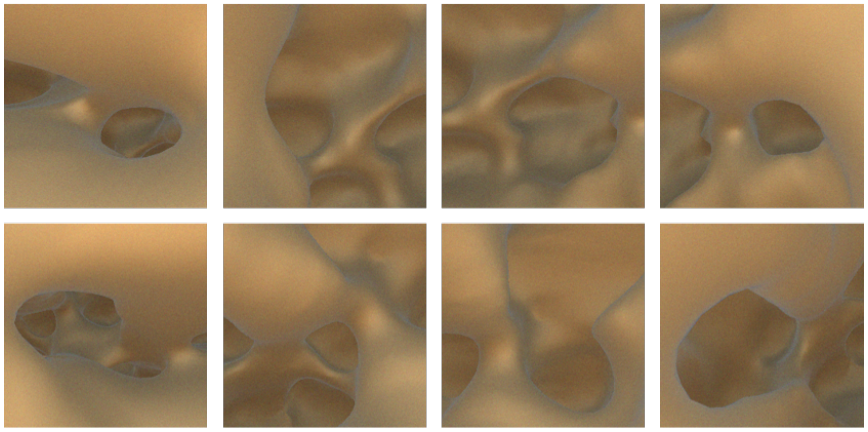

Figure 4.1: The sample images of the simulation images, captured by the virtual camera in VURS environment with different material color, a point light and Gaussian noise. Each image represents different type of calyx. Eight types of calyx in total.

Since both the simulation and the VID images are taken from the same environment, the location and edge features of each calyx are known during the acquisition process. Therefore, the result obtained after using the simulation image to perform image matching does not need to be verified by the position information of the VC, and its performance can be judged by ground truth provided by the author.

4.2. Experimental protocol

To validate the whole navigation system, an experiment has been conducted. The goal of the experiment is to test the feasibility and performance of the navigation system in a real environment. More specifically, the system can find the correct calyx, calyx branch, and approximate location of the real ureteroscope from the result of the image matching algorithm. The whole experiment workflow is shown in Fig. 4.2.

In the RCI image collection part, coordinates registration between the EMT sensor and VURS needs to be performed firstly, which is used for matching results validation. Then the robotic ureteroscope

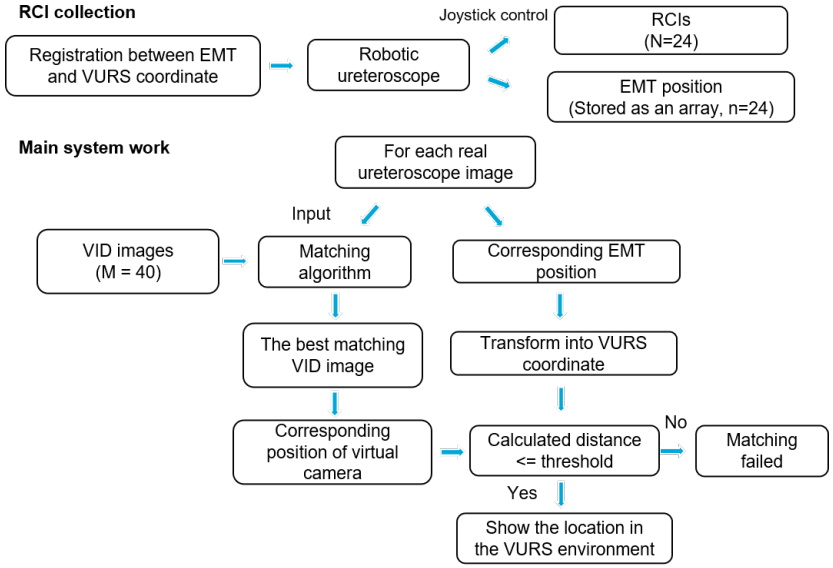

Figure 4.2: The experiment workflow consists of two parts: RCI collection and main system work.

prototype with an EMT sensor are controlled by a gamepad to insert into the kidney phantom to get RCIs and their corresponding EMT position data.

In the main system work part, those RCIs are matched with VID images to find the best matching result image from the VID. At the same time, corresponding position and orientation data of the VC in the VURS environment of the resulting image are also found. Finally, according to the transformation between the EMT coordinate and the VURS coordinate, the RCI-corresponding position data of the EMT sensor can be transformed into the VURS coordinate. Those data are compared with the position data of the VC in the VURS coordinate of the resulting image to validate the performance. In addition, the result data are used to visualize the VC location in the VURS environment to make the system user-friendly.

In the rest of this chapter, the experiment setting, including robot setup, EMT sensor and software, is explained. Then, the registration method is described. In the end, a detailed validation method is shown.

4.3. Experimental setting

This section will introduce the robot system, EMT system setup, host computer and software used to implement the algorithm of the experiments, and the final setup to perform the experiment.

4.3.1. Robot system

The robot system is shown in Fig. and consists of three parts: the robot body, drive motors (four stepper motors and one linear slider) and the control board.

Robot prototype

The robot prototype used in this experiment draws on the design of the HelicoFlex robot[109] developed by the BITE group [110] at TU Delft, which has high torsional and axial stiffness and low bending stiffness.

The prototype is made up of a tip, two steerable helicoid segments and a rigid shaft (Fig. 4.4). Each segment can be independently controlled by two cables and is 20 mm long. However, in this experiment, only the distal segment is controlled to steer to simulate the standard ureteroscope product with only two DOFs. The diameter of the prototype is 5 mm, and the maximum bending angle is 120°. It is a cable-driven soft robot.

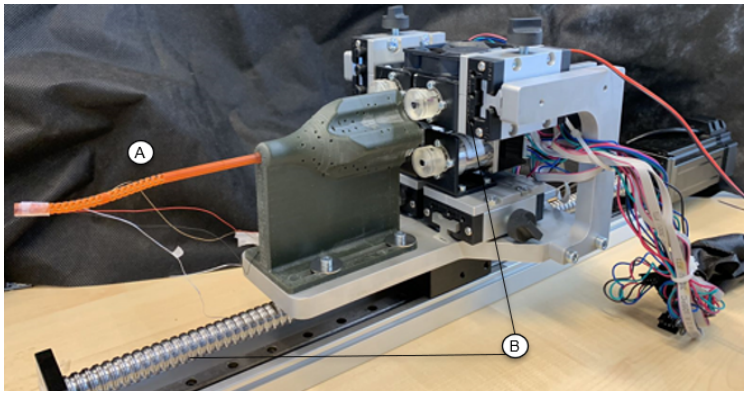

Figure 4.3: The robot prototype used in the experiment without the control board. (A) the robot body. (B) four stepper motors and a linear slider.

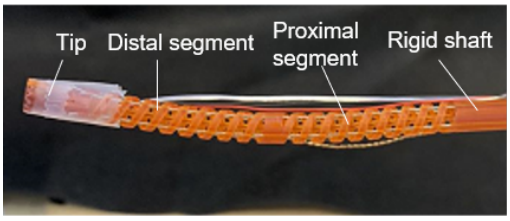

Figure 4.4: The robot body overview. It consists of tip, two helicoid segments (distal and proximal) and a rigid shaft.

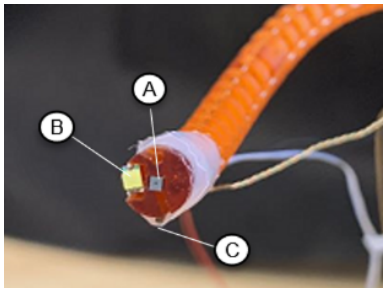

Figure 4.5: The structure of the robot tip. (A) 1X1 mm digital camera. (B) LED light. (C) EMT sensor.

At the tip of the prototype, a digital camera, a LED light, and an EMT sensor is fixed (Fig. 4.5). The digital camera is "NE2D_RGB_V160F2.4_2M" manufactured by ams-OSRAM AG. [111]. The size of this camera is 1 mm by 1 mm. It provides 250 $\times$ 250 image resolution with 3 um by 3 um pixel size, and its field of view (FOV) is 160°. The digital camera is selected because of its suitable size for an endoscope, wide FOV, and low cost. The camera is fixed in the center of the robot tip to provide images without offsets.

The LED light is placed close to the camera for illumination. Typically, optical fibers are used as light sources for illumination in endoscopic devices, such as LithoVue [112], a disposable endoscope produced by Boston Scientific Corporation. However, due to the material and texture of the phantom, and the camera limitations, the brightness of the image captured by the camera, when using an optical fiber for illumination, is very low, so the image processing cannot be performed. Therefore, a LED light together with an Arduino board to control its brightness is directly used as the light source. Because the size of the LED light is much larger than that of the optical fiber, due to the limitation of the tip size, the position of the LED light has little choice, which leads to the appearance of diffraction fringes in the subsequent image acquisition. However, this interference is acceptable because it can be removed by image processing.

The EMT sensor is taped directly under the camera. As initially designed, there is a slot reserved for holding the sensor. Nevertheless, due to the length limitation of the tip, when installing the EMT sensor in the slot, most of the sensor will extend beyond the tip. As a result, cameras with wide FOV will also capture the EMT sensor into the image, making subsequent image processing impossible. In this experiment, the data of the EMT sensor is only used to verify the results and does not require very high position accuracy. Therefore, it is feasible to fix the EMT sensor with tape directly.

Drive motors

Stepper motors are used to control the robot steering. For example, the cable's length on one side is changed by the rotation of the motor, thereby forming a pull on that side of the robot. According to the principle of the cable-driven robot, the robot will bend with the change of cable length and drive the robot to steer this way.

In this robot system, the four stepper motors are Nema 11 Closed-loop Geared Stepper (Fig. 4.6). Every two motors control a segment to achieve 2 DOFs steering. In this experiment, only one segment is used to complete the steering. To solve the forward insertion problem, a linear slider (RATTMMOTOR 400 mm Travel Length Linear Rail shown in Fig. 4.7) is used. The slider provides a translation DOF. So the whole robot system has three DOFs: two rotations and one translation.

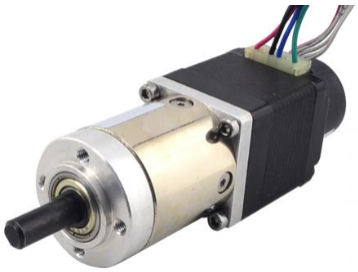

Figure 4.6: The stepper motor used in the robot prototpye [113].

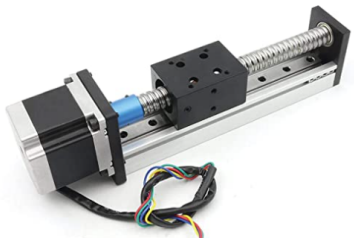

Figure 4.7: The linear slider used in the robot prototpye [114].

Control board

The control board overview is shown in Fig. 4.6.

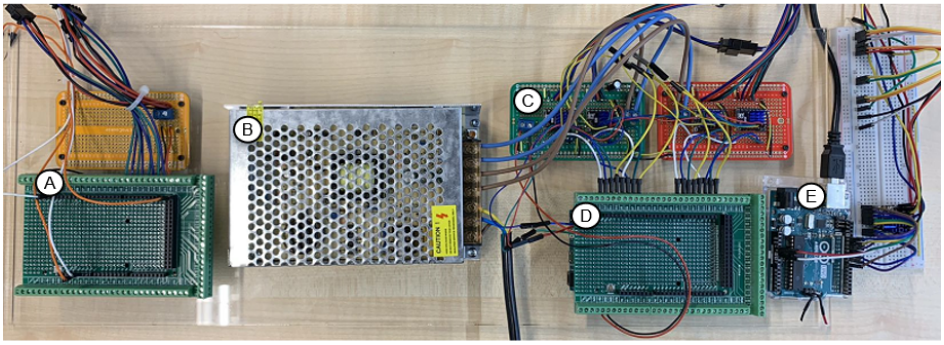

Figure 4.8: The overview of the control board. (A) An encoder for motors; (B) A power supply; (C) Two motor divers (each driver board drives two motors); (D) Arduino Due board [115] for the drivers controller; (E) Arduino Uno board for the slider controller.

4.3.2. EMT system

The EMT system obtains the position data of the EMT sensor in its own coordinate by generating the electromagnetic field. This system is often used in computer-assisted interventions [116], where a free line-of-sight does not exist and the positioning of surgical instruments cannot be performed visually. This technique was proposed firstly by Raab et al. [117] and Kuipers [118] in 1970s, and began to be applied in medical field in 2000s [119]. Nowadays, there are two main EMT systems: Aurora® [120], produced by Northern Digital Inc. and the Polhemus EMT system [121].

In this experiment, Medical Aurora® V3 from Northen Digital Inc (NDI) was used [120]. The EMT system consists of four parts: a field generator, EMT sensors and a system control unit and a sensor interface unit (shown in Fig. 4.9).

Among these four components, the selection of sensors is the most. EMT sensors are classified according to the DOF of sensed data, 5 DOFs and 6 DOFs. A 5 DOFs sensor ignores the DOF of rotation around the z-axis. For many medical instruments, like needles, this DOF is less important than the rest of them. In addition, if the sensor needs to measure 6 DOFs, then it needs two inductors. This is why the diameter of the 6 DOFs sensor is much larger than that of the 5 DOFs sensor. In this experiment, 5 DOFs sensors were used because of the size constraints of the robot.

Data from the sensor are gained by using the software, NDI Track. The system was used in the cube volume configuration where the root means square of the position data acquired from a 5 DOFs sensor is 0.7 mm with 95% confidential interval of 1.4 mm [120]. If the EMT system is operating with

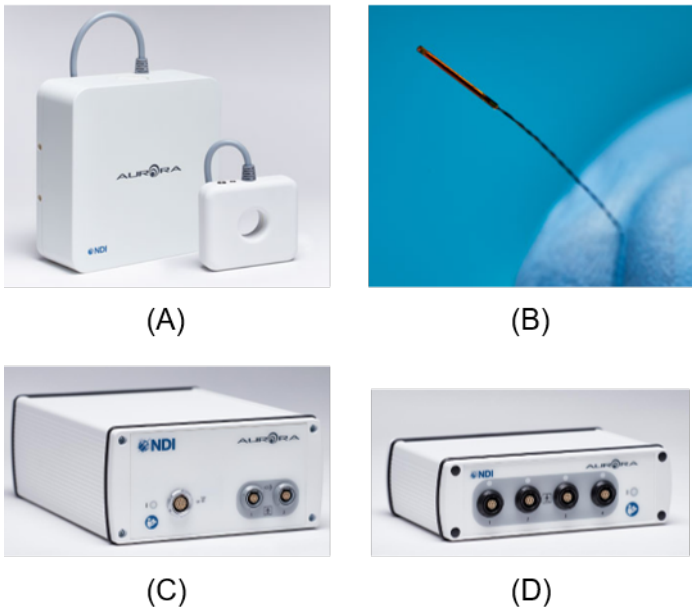

Figure 4.9: The components of the Aurora EMT system [120]. (A) The Aurora electromagnetic field generator; (B) The Aurora 5DOF EMT sensor; (C) The Aurora System Control Unit; (D) The Aurora Sensor Interface Unit.

ferromagnetic devices nearby or metals that disturb the magnetic field, the EMT system will be greatly influenced in accuracy and performance or even unusable. Therefore, in order to avoid this kind of situation, in the experiment, the racks and auxiliary parts used to fix the phantom near the field generator are non-ferromagnetic and non-metallic.

4.3.3. Host computer and software

The host computer has three functions in this experiment: providing the control system of the robot to control the motions of the robot by using a gamepad, receiving the data of the EMT sensor, acquiring RCIs inside the kidney phantom, and performing subsequent image processing and the image matching algorithm. The host computer specifications are as follows:

- Computer product: ThinkPad P1 2^{nd} Gen (laptop).
- CPU processor: Intel® Core™ i7-9750H, 6 cores, 2.6GHz base frequency.
- Operating system: Windows® 10 Home 64.
- Graphics: NVIDIA® Quadro® T1000 with 4GB GDDR5 memory.
- Memory: DDR4-2666 ECC 16GB.

The EMT part has been explained in the “EMT system”. Therefore, the software used to control the system for the robot and image acquiring and processing are explained below.

Robot system control

The robot control system based on the Robot Operating System (ROS) [122], an open-source robotics middleware suite [123], is built to establish the communication between the gamepad and the Arduino board that controls the motors. In this way, the instructions of the robot motion are conveyed through the buttons or joysticks on the gamepad. These instructions are published in the form of a rostopic, and a script based on Arduino-ROS-API subscribes to the rostopic and transfers them into low-level instructions for controlling motors.

Since ROS is built on the Linux operating system and the host computer is a windows system, a virtual machine is needed to run two systems on the same host computer. The VMware Workstation 16 Player platform [124] is used to run a Linux-based virtual machine. Ubuntu 18.04 is the running version of the Linux system. The key parameters of the virtual machine are 4GB of memory and 2 processor cores. ROS-kinetic is used as the ROS version in this experiment.

Image acquiring and processing
Based on the camera used in the experiment, its developer provides the supporting image acquisition software, NanEye Viewer [125]. The software can read the images in real-time through the Nano board of the camera and the USB transmission of the host computer. When the digital camera displays the desired image by controlling the robot, it can be manually triggered to acquire the image and save it for subsequent image processing. The software runs in the windows system. The version of NanEye Viewer is 6.4.3.3.

After obtaining the RCI, it is necessary to obtain edge information through image processing, convert it into a point set, and match it with the pre-processed VID. All of the above functions are implemented in python. Among them, “OpenCV” [126] is the most frequently used package, which provides a very comprehensive basic image processing algorithm. “scikit-image” [127] and “plantcv” [128] are also used in skeletonization and edge pruning to remove small branches, respectively. When doing image matching, “ShapeContextDistanceExtractor” class in OpenCV is used as the implementation method of SC matching [92]. The version of “OpenCV’, “scikit-image” and “plantcv” are 4.5.5, 0.19.3 and 3.14, correspondingly.

4.3.4. Final setup
Figure 4.10 shows the overall experimental setup.

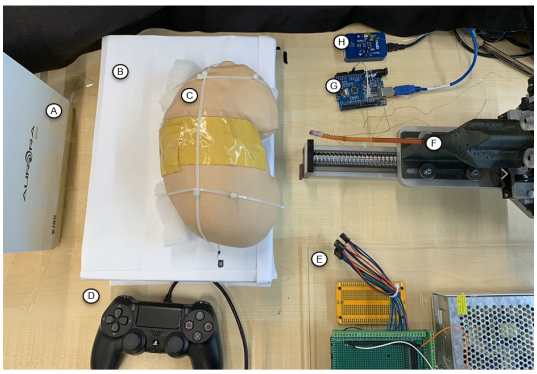

Figure 4.10: The overview of the experiment setting. (A) Electromagnetic field generator. (B) The bracket for fixing the kidney phantom. (C) The kidney phantom. (D) The joystick that is used to control the robot. (E) The control board for the robot. (F) The robot prototype. (G) The arduino board for LED light. (H) The arduino board for digital camera.

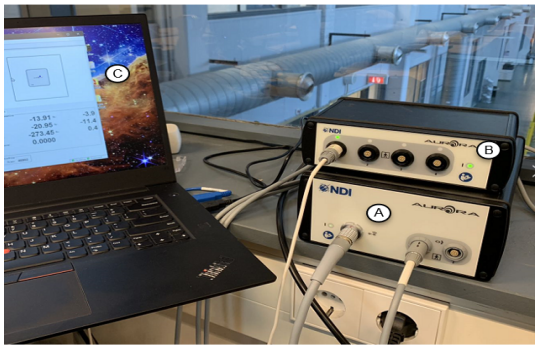

Figure 4.11: (A) Aurora® System Control Unit; (B) Sensor Interface Unit; (C) the host computer.

The Aurora® System Control Unit and Sensor Interface Unit are placed close to the host computer 4.11, as far as possible from the field generator to avoid interference with the magnetic field. Because of the need to ensure the repeatability of the experimental data collected by EMT, the relative positions of the field generator and kidney phantom cannot be changed. However, due to the low requirements for precision and the limitation of project time, the field generator and the kidney phantom are directly fixed on a workbench with tape. The space for the movement of the EMT sensor was guaranteed to be covered by the magnetic field space.

The robot is controlled by a gamepad. Fig. 4.12 shows working conditions and the direction of the EMT coordinate. According to the previous subsection “Robot system”, the robot has only three DOFs, two rotations and one translation. The two rotating motions can be controlled by a joystick. The left or right operation of the joystick can make the robot rotate around the x-axis (turning left or right), while the up or down operation of the joystick corresponds to the rotations around the y-axis (turning up or down). The translation of the linear guide is controlled by the shoulder buttons “L” and “R” on the gamepad. The L button corresponds to the forward translation (along the z-axis) of the linear slider, and the “R” key controls the backward translation.

4.4. Coordinates registration
In order to verify whether the image matching result is correct, the position of the EMT sensor corresponding to the RCI and the position of the VC corresponding to the resulting image needs to be

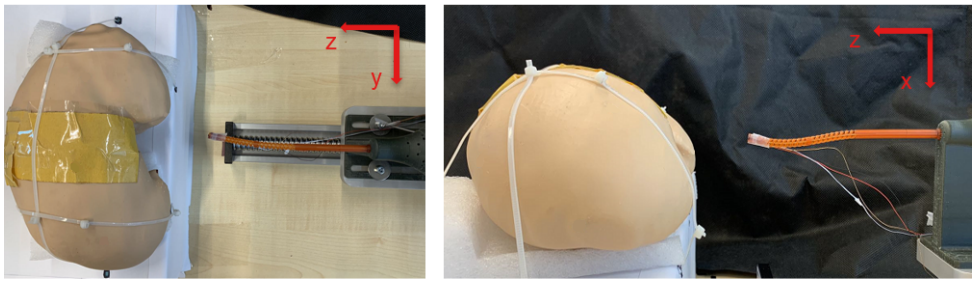

Figure 4.12: The EMT coordinate directions. The left image shows the top view of the experiment setup, where the positive z direction is along the robot body, pointing the kidney phantom, and the positive y direction is perpendicular to the z direction, pointing down in the image. The right image shows the front view of the experiment setup, where positive x direction is also perpendicular to the z direction, pointing down in the image.

compared in the same coordinate system. This requires the registration of two different coordinate systems to determine the transformation relationship.

The registration is essential to find the transformation relationship between the two coordinate systems, including rotation and translation. For the rigid scenario, the most direct way is to find two sets of points that correspond in different coordinate systems and optimize the most suitable transformation relationship by Least-Squares fitting [129]. This method is based on the singular value decomposition of a 3-by-3 matrix to find the rotation matrix. Compared with the other two methods, quaternion [130] and iteration [131], it can be drawn that the calculation speed of this method is the fastest when the number of corresponding points is large.

Since our experiments are all rigid and do not consider the deformation of the kidney, the above coordinates registration method is suitable for this experiment. By finding the corresponding feature points on the actual phantom and the kidney model in the VURS environment and recording their position information in their respective coordinate systems, two sets of corresponding point sets can be collected. This coordinates registration algorithm can calculate the rotation and translation matrices of the two coordinate systems.

Feature points selection

Firstly, eight points are taken out on the surface of the phantom. The reason for not using an additional object to do the calibration is to avoid errors due to the relative position of that object to the kidney model in both the VURS environment and in the real world. Since the kidney phantom is almost curve-like in shape, it is difficult to find enough distinct feature points. Also, the lower half of the kidney phantom is mainly inaccessible because it is fixed to the stand. So there are not many optional feature points on the phantom. The first six points are relatively obvious feature points, and the last two are artificially selected points to make the points as scattered as possible. Secondly, the EMT sensor is

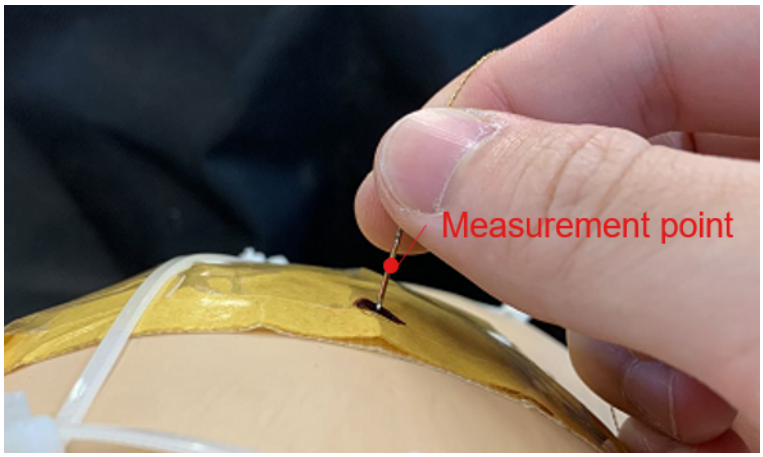

Figure 4.13: Measuring a feature point using the EMT sensor. The sensor direction aligns with the normal direction of the surface. The measurement point is at the center of the sensor.

used to make contact with the selected eight points to obtain the position information of the sensor in the EMT coordinate system. Since what the sensor measures is not the position data of its tip, but the position in the center of the sensor, in the process of contacting the phantom, it is necessary to adjust the direction of the sensor as much as possible to the normal direction of the surface where the feature point is located (shown in Fig. 4.13). This provides a standard for the later selection of feature points in the VURS environment.

Finally, all corresponding feature points are identified on the kidney model in the VURS environment, and the position information is recorded. Spheres can be used to simulate the EMT sensors thanks to the ability to place objects on surfaces in Unity. When the radius of the sphere is set equal to half the length of the EMT sensor, the position data of the sphere center are corresponding to those of the EMT sensor. This is why, as mentioned earlier, the EMT sensor needs to be placed in the normal direction of the surface. Fig. 4.14 show all eight selected feature points in both environment and their correspondence.

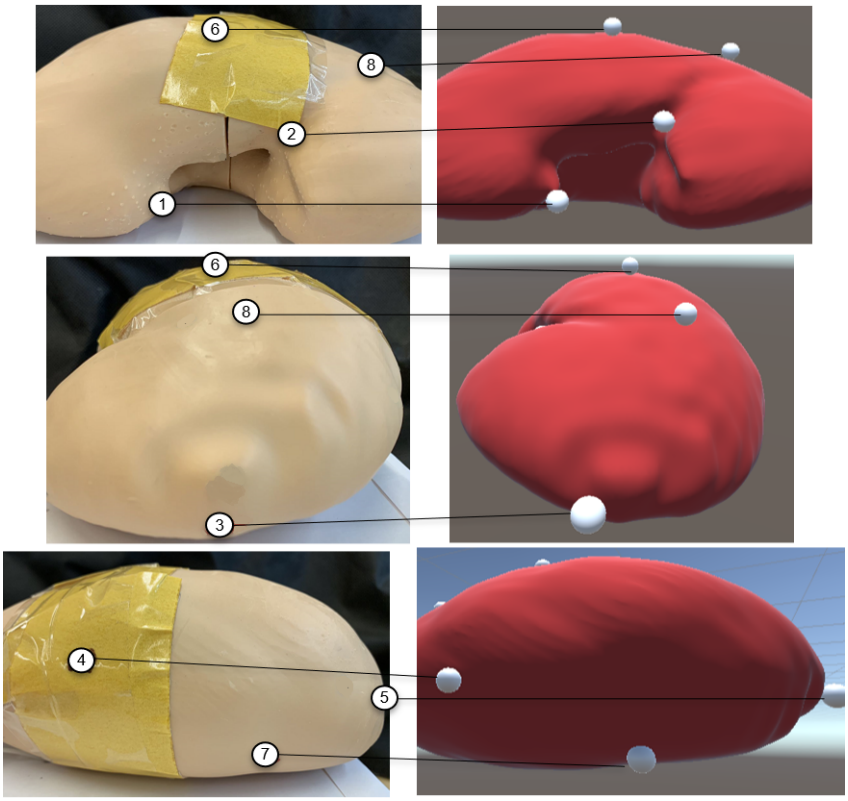

Figure 4.14: The views of feature points fitting between real phantom and VURS. Those eight point groups are used to do the coordinates registration.

Registration implementation

Since the kidney model in the VURS environment is scaled 6 times to the size of the real phantom, the sizes of spheres also need to be scaled 6 times. In addition, the position data collected from VURS need to be restored to the actual size, which means that all the data from VURS need to be divided by 6.

After collecting all the point data, the coordinates registration algorithm calculates the transformation relationship between the two coordinate systems. A simple way is to use least-squares fitting to calculate the transformation relation [129]. [132] has implemented the algorithm into a python script and is verified to be correct. In this experiment, the rotation and translation matrices between the EMT coordinate system and the VURS coordinate system can be calculated by using this python script.

4.5. Result validation

The transformation relationship between the two coordinates can be obtained through the registration of the EMT coordinate and the VUE coordinate. Using this transformation relationship, any point in the EMT coordinate can be represented in the VURS coordinate. According to this basis, the position data of the EMT sensor when the RCI is collected can be transferred into the VURS coordinate and compared with the position data of the VC corresponding to the matching result in the VURS coordinate to verify if the resulting image is correct or not.

As mentioned in Chapter 3, a threshold should be determined to compare with the distance between the positions of the EMT sensor and the VC corresponding to the resulting image in the VURS coordinate. The initial threshold ϵ can be defined as:

$$\epsilon = e_{cam} + e_{reg} + e_{EMT}, \tag{4.1}$$

where e_{cam} is the error caused by the difference between the FOV of cameras; e_{reg} is the error of coordinates registration; e_{EMT} is the error caused by the EMT sensor itself. As for the error caused by phantom fabrication and connection, it has already been involved in the error of coordinates registration.

The error caused by the difference between the FOV of cameras (e_{cam}) can be calculated from a manual calibration. The basic assumption behind this error is that the distances between the camera and the object are different from the VC and the real URS camera because of their different FOVs. Therefore, taking one of the RCIs as the reference, a virtual image is taken from the VURS environment to be as consistent as possible with the reference. The error can be seen as the distance between the EMT sensor position and the VC position of that image in the VURS coordinate system by using the results of the coordinates registration. It can be shown as:

$$\begin{aligned} {}^{V}p_e &= {}^{V}\,\mathsf{R}_E\,{}^{E}p_e + \mathsf{T}, \\ e_{cam\_n} &= \sqrt{|{}^{V}p_e - {}^{V}\,p_v|^2}, \end{aligned} \tag{4.2}$$

where ${}^{V}p_e$ and ${}^{E}p_e$ represent the EMT position data corresponding to the reference RCI in the VURS and the EMT coordinate, respectively, Vp_v represents the VC position data of the virtual image in the VURS coordinate, and e_{cam_n} represents the resulting error calculated by one reference. To prevent bias from a single result, multiple images are used. Therefore, the final error can be defined as:

$$e_{cam}^{'} = \frac{\sum_{i=1}^{n} e_{cam\_n}}{n} + \sigma, \tag{4.3}$$

where n represents the number of reference images, σ represents the standard deviation (STD) of errors. Since SC matching is scale-invariant, there is a difference between the calyx size in the resulting image and the RCI, which is caused by the difference in camera object distance. Considering the above situation, an STD is added to the mean value to improve tolerance for distance.

When calculating $e_{cam}^{'}$, the error of coordinates registration (e_{reg}) is already taken into account by using the rotation and translation matrices. Thus, e_{reg} should be ignored in this method. The root mean square (e_{EMT}) of the 5DOFs EMT sensor for position measurement can be found in [120], which is the known error of the standard product. Therefore, the final threshold, ϵ, can be defined as:

$$\epsilon = e_{cam}^{'} + e_{EMT}. \tag{4.4}$$

5 Results

5.1. Registration Result

As explained in Chapter 3 and 4, coordinates registration between the EMT and the VURS coordinate needs to be performed. By collecting 8 corresponding points in both coordinates and using the least-squares fitting, both rotation and translation matrices can be calculated. Table 5.1 shows the original position data of all 8 registration points in both VURS and EMT coordinates.

Table 5.1: The 8 corresponding points data from two coordinates for coordinates registration.

Points No.	**VURS** (m)			**EMT** (mm)		
	x	y	z	x	y	z
1	-0.418	-0.813	-1.189	-22.80	8.48	-274
2	-0.492	-0.672	-0.952	-48.66	-34.39	-255
3	-0.507	-0.795	-0.434	-15.28	-112.78	-234
4	-1.121	-0.861	-0.993	-25.25	-16.25	-149
5	-0.723	-0.887	-1.644	-30.20	86.87	-228
6	-0.756	-0.515	-1.040	-80.90	-11.75	-216
7	-0.776	-0.594	-0.713	-57.46	-76.03	-205
8	-0.988	-1.002	-1.311	-6.80	43.03	-175

Before using the algorithm to solve the registration problem, the data from the VURS coordinate need to be pre-processed. Since the kidney model in the VURS environment is 6 times the real phantom, those point data in the VURS coordinate should divide by 6 to keep the same scale as the EMT coordinate. In addition, a left-handed coordinate system is used in unity, while the EMT coordinate is right-handed. Thus, the VURS coordinate system needs to transform into a right-handed coordinate system first. This can be achieved by flipping the z-axis of the VURS coordinate system. After that, the data should be multiplied by 1000, because the Unity environment is measured in meters, and they need to be represented in millimeters. These steps can be formulated as:

$$ {}^{V}\mathrm{P}_{V} = \frac{1}{6}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} {}^{V}\mathrm{P}_{V\_ori}, \tag{5.1} $$

where ${}^{V}\mathrm{P}_{V}$ and ${}^{V}\mathrm{P}_{V\_ori}$ are the matrices that present the pre-processed position data and the origin data of the VURS environment. They are 3 by 8 matrices, where 3 rows are for x, y, z components and 8 columns are for 8 points.

By using the code provided by [132], the rotation matrix and the translation matrix that transform

the EMT coordinate into the VURS coordinate are calculated:

$$ {}^{V}\mathrm{R}_{E} = \begin{bmatrix} 0.109 & -0.118 & -0.987 \\ -0.987 & -0.134 & -0.093 \\ -0.121 & 0.984 & -0.130 \end{bmatrix}, \quad \mathrm{T} = \begin{bmatrix} -332 \\ -185 \\ 153 \end{bmatrix}. \tag{5.2} $$

Fig. 5.1 visualizes the registration result by using a 3D plot. The root-mean-square-error (RMSE) can be calculated by using the pre-processed VURS points and corresponding points generated from applying rotation and translation matrices to the data in the EMT coordinate to check its accuracy. The RMSE is 5.20 mm.

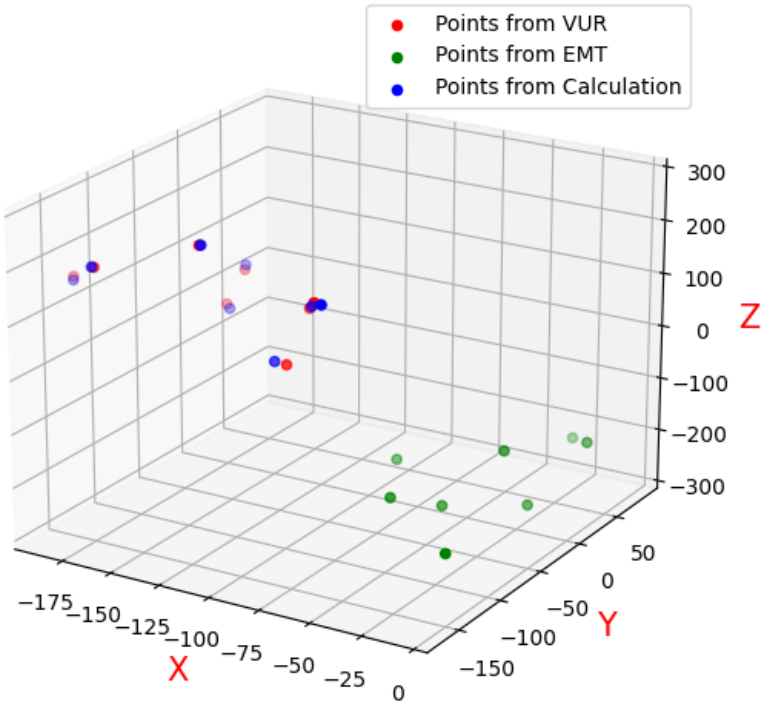

Figure 5.1: The visualization of the registration result on those eight feature points. The green points represent the eight EMT points in the EMT coordinate, while the red points represent the eight corresponding VC points in the VURS coordinate. The blue ones are transformed from the green points by using the rotation and translation matrices (Equation 5.2) generated by coordinates registration.

5.2. Algorithm performance

5.2.1. Threshold determination

According to equation 3.2, it is possible to judge whether the matching result is correct by comparing the distance between the VC and the EMT sensor in the VURS coordinate system with a specific threshold. The result can be considered correct when the distance is less than the threshold. Otherwise, it is wrong. From equation 4.4, two factors affect the threshold.

To calculate $e^{'}_{cam}$, five images are selected ($n = 5$ in equation 4.3). Fig. 5.2 shows the chosen images to calculate the error. The VC position in the VURS coordinate (${}^{V}p_{v}$) and the EMT sensor position in the EMT coordinate (${}^{E}p_{e}$) for all groups of images are shown in Table 5.2. Their corresponding errors are also calculated through Equation 4.2 after doing the pre-processing on ${}^{V}p_{v}$, according to Equation 5.1. Finally, their mean value and STD equal to 13.70 mm and 0.70 mm. Thus, based on Equation 4.3, $e^{'}_{cam}$ is 14.40 mm.

According to [120], the RMS of the 5DOFs EMT sensor for position measurement (e_{EMT}) equals to 0.7 mm.

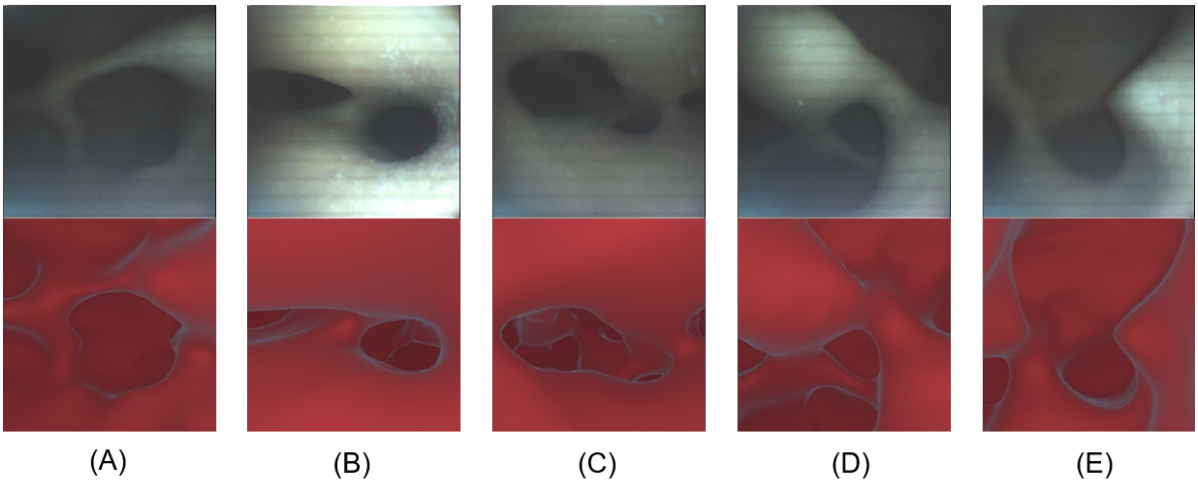

Figure 5.2: Five images are selected to calculate the error caused by the difference between the FOV of cameras. The images in the first row are captured by real camera. The images in the second row are manually collected in VURS environment, as similar as possible to their corresponding real images. (A) to (E) represent the five different groups of images.

Table 5.2: The position data for all images in Fig. 5.2 with their corresponding e_{cam} and their mean value and STD.

No. of image group	${}^{V}p_v$ (m)			${}^{E}p_e$ (mm)			Error (mm)
	x	y	z	x	y	z	
A	-0.649	-0.784	-0.893	-35.05	-45.62	-217.87	12.62
B	-0.547	-0.794	-1.067	-22.49	-24.2	-246.98	14.09
C	-0.483	-0.846	-1.078	-26.56	1.59	-254.93	13.64
D	-0.665	-0.780	-1.180	-35.05	19.42	-220.09	14.50
E	-0.627	-0.777	-1.179	-28.68	14.81	-226.68	13.67
Mean	/	/	/	/	/	/	13.70
Std	/	/	/	/	/	/	0.70

To sum up, the threshold can be determined as:

$$\begin{aligned} \epsilon &= e^{'}_{cam} + e_{EMT} \\ &= 14.40 + 0.70 \\ &= 15.10\,\text{mm}. \end{aligned} \tag{5.3}$$

5.2.2. Performance on simulation images

The simulation images are regarded as the input to test the image matching algorithm. Those images are taken from the VURS environment with another rendering material color, a point light and additional Gaussian noise to simulate the real situation as much as possible. There are 24 images in total with 8 calyxes, three images per calyx. The results after image matching for those simulation images are shown from Fig. 5.3 to 5.10.

5.2.3. Performance on RCIs

The edge features

Before doing image matching to find the most similar calyx image in the VID, the edge features of the RCI are extracted first. The edge features corresponding to each image can be obtained and converted into point sets through the image processing method mentioned in Chapter 3. Fig. 5.11 to 5.14 show the result of edge extraction for each image. In each figure, two groups of images are displayed, which correspond to two calyxes. For each calyx, there are three images, where the left part is the original RCI and the right part is the final edge feature for this RCI.

As seen from the above figures, for different calyx, the edge information that can be extracted is

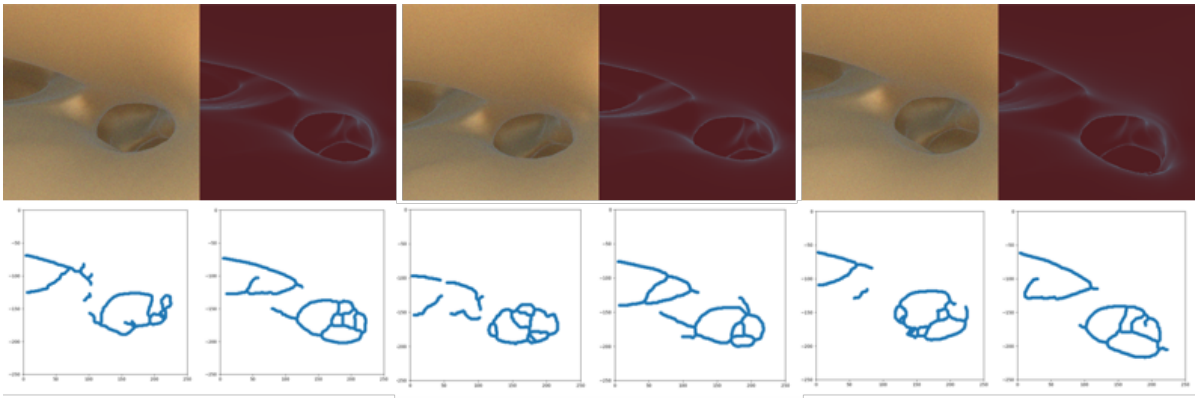

Figure 5.3: The results of calyx A in simulation after the image matching. There are three groups for each calyx. In each group, the left-top image is the simulation image; the left-bottom image is its edge feature image; the right-top image is the resulting image in VID, and the right-bottom image is its edge feature image.

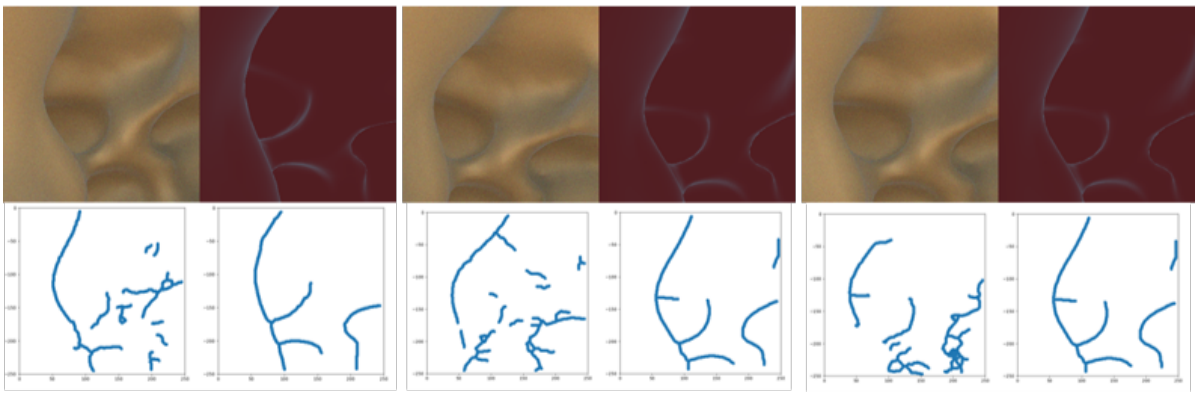

Figure 5.4: The results of calyx B in simulation after the image matching. The same structure to Fig. 5.3.

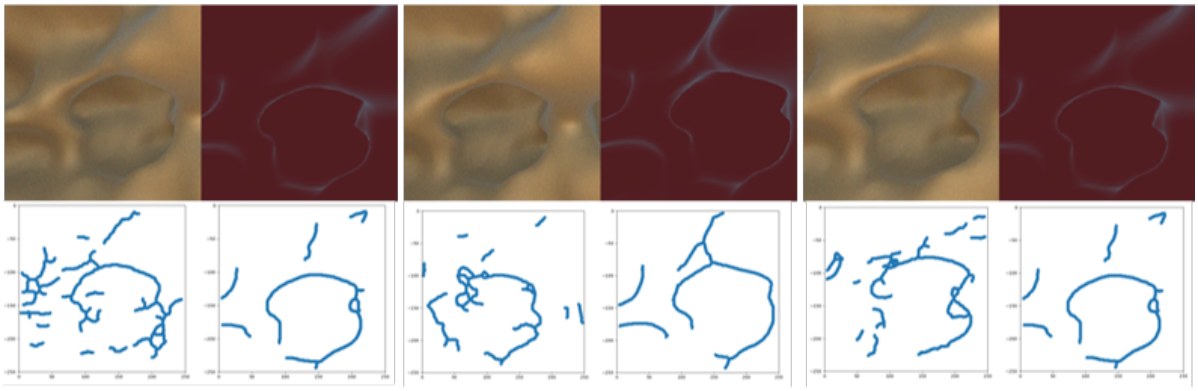

Figure 5.5: The results of calyx C in simulation after the image matching. The same structure to Fig. 5.3.

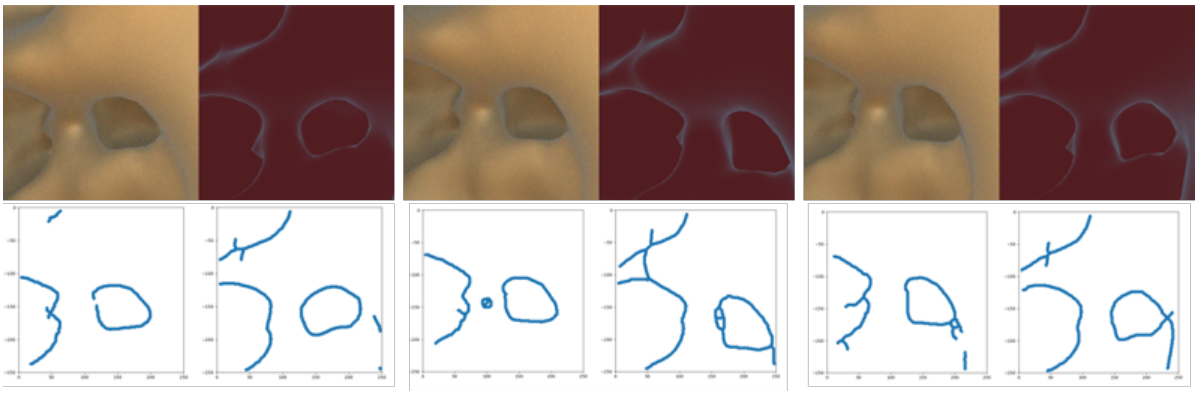

Figure 5.6: The results of calyx D in simulation after the image matching. The same structure to Fig. 5.3.

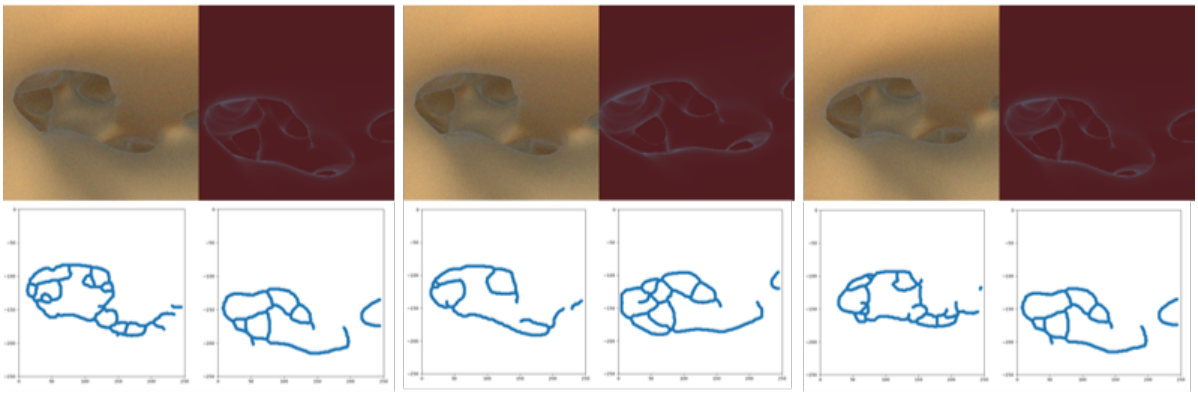

Figure 5.7: The results of calyx E in simulation after the image matching. The same structure to Fig. 5.3.

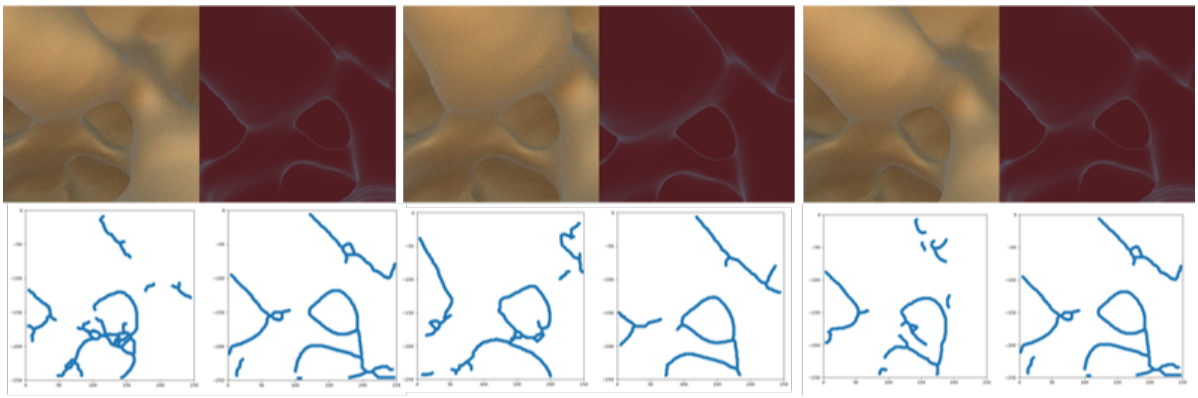

Figure 5.8: The results of calyx F in simulation after the image matching. The same structure to Fig. 5.3.

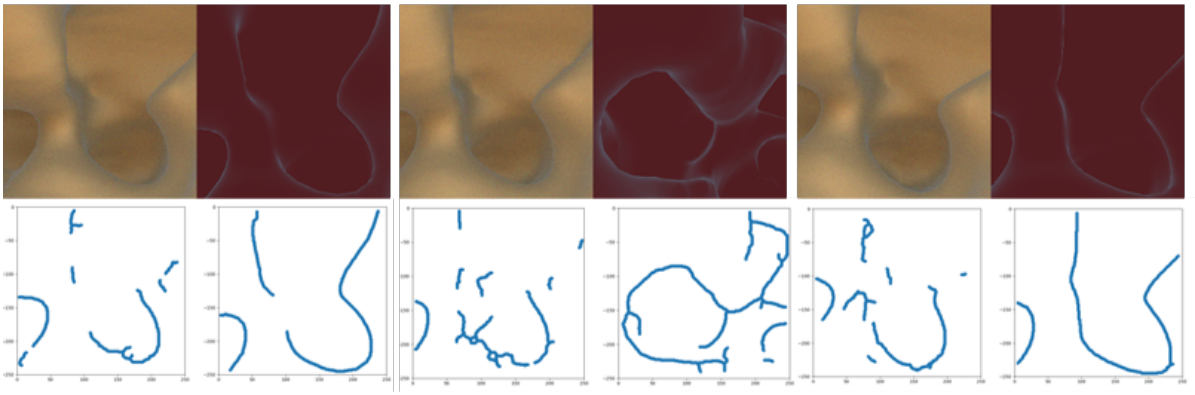

Figure 5.9: The results of calyx G in simulation after the image matching. The same structure to Fig. 5.3.

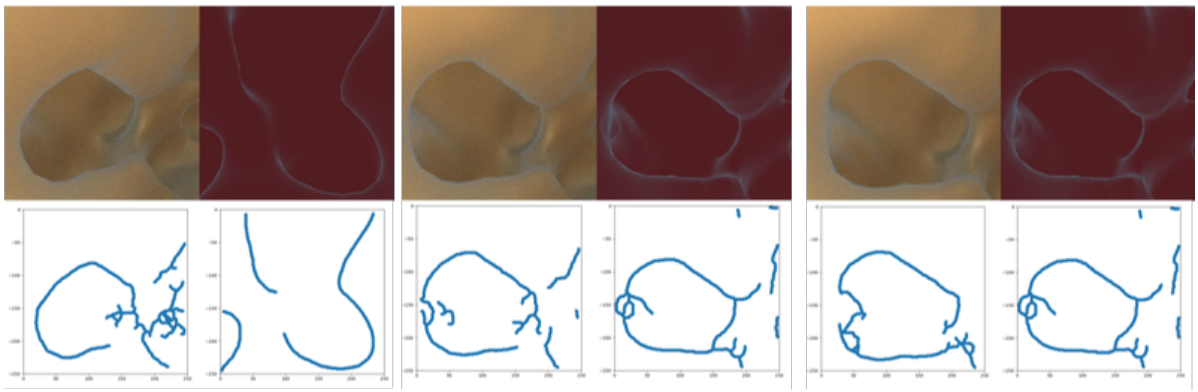

Figure 5.10: The results of calyx H in simulation after the image matching. The same structure to Fig. 5.3.

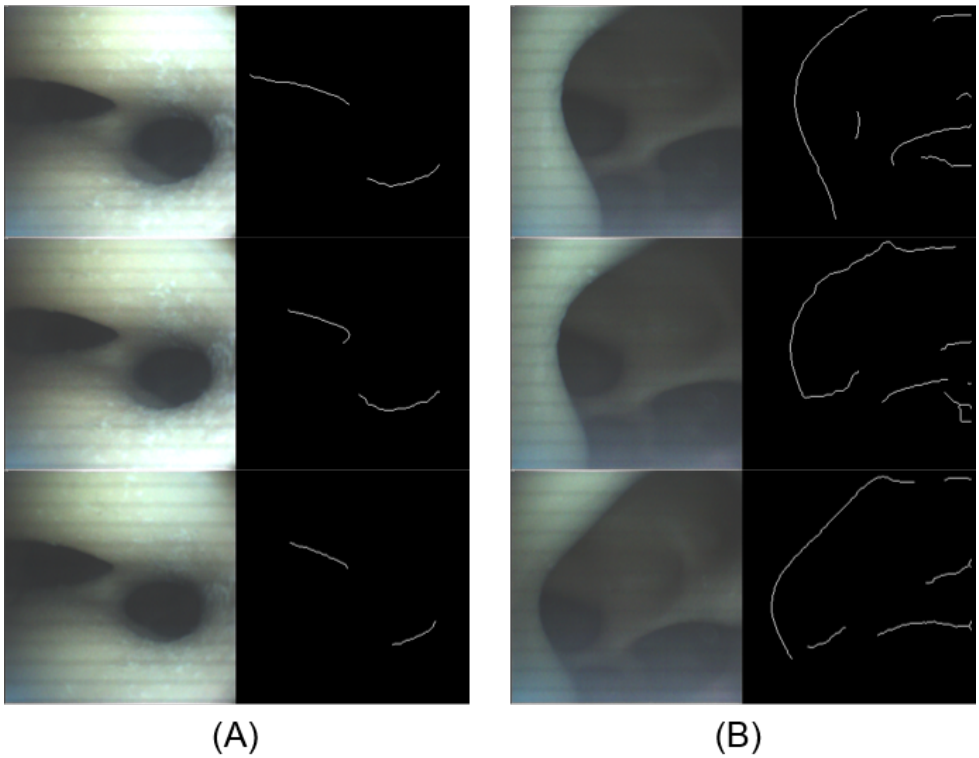

Figure 5.11: The results of edge feature detection A, B calyx image groups. For each image group, the images from the first column are RCIs, while the images from the second column are corresponding edge feature images. In each group, the images from top to bottom, are the first, second, third image of the calyx.

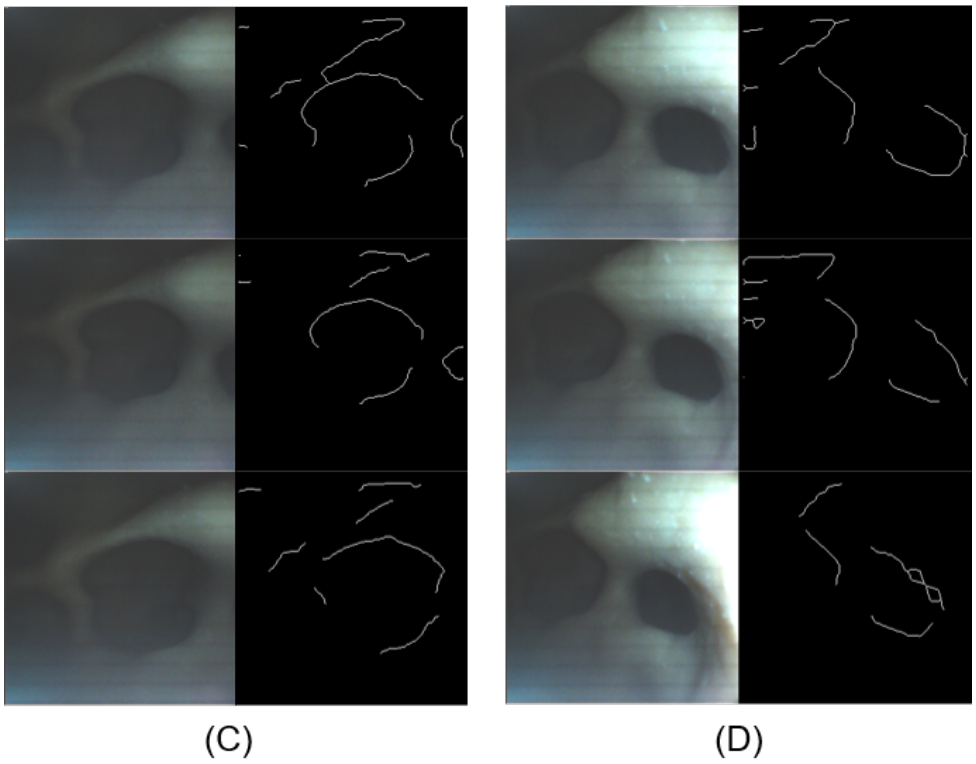

Figure 5.12: The results of edge feature detection C, D calyx image groups. The same structure to Fig. 5.11.

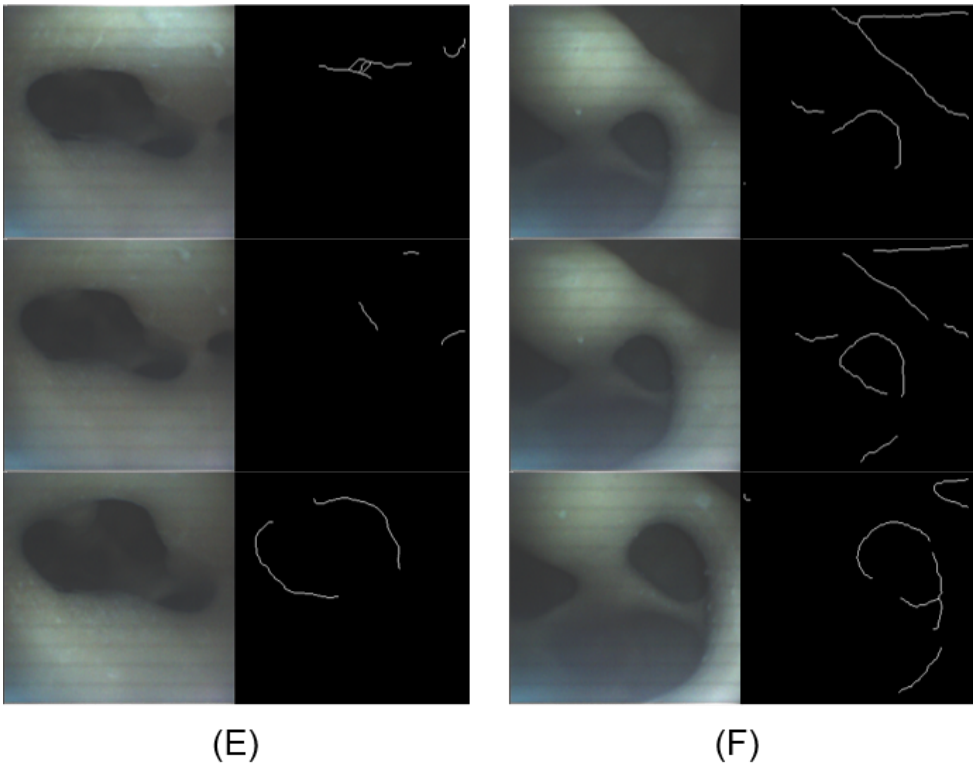

Figure 5.13: The results of edge feature detection E, F calyx image groups. The same structure to Fig. 5.11.

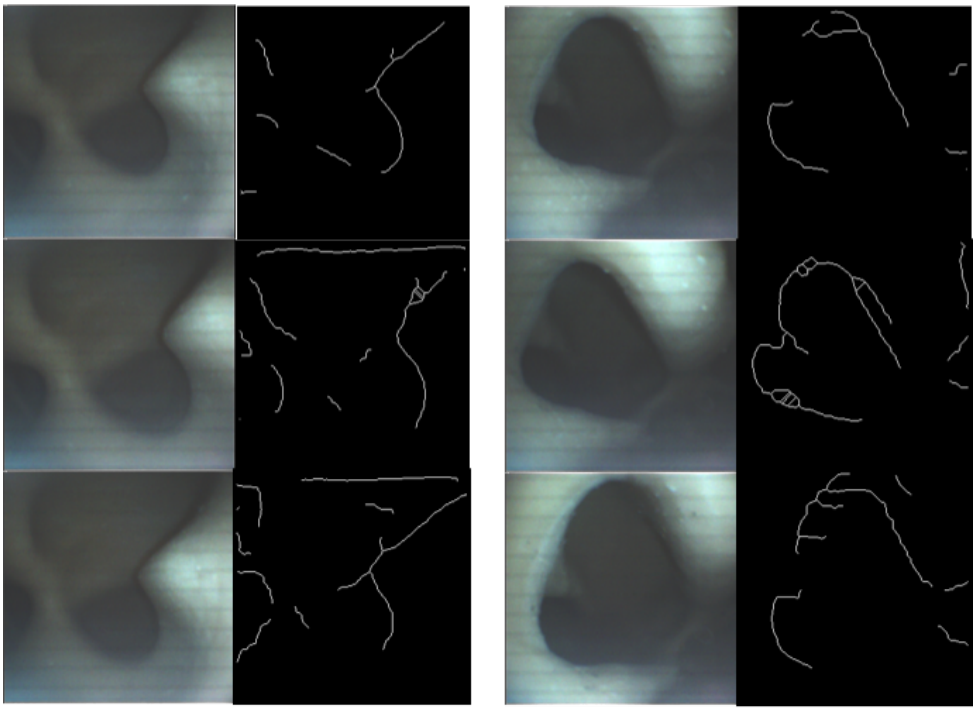

Figure 5.14: The results of edge feature detection G, H calyx image groups. The same structure to Fig. 5.11.

also different, resulting in different numbers of points representing edge features. Fig. 5.15 shows the distribution of the number of edge feature points corresponding to each RCI.

The matching results

After edge feature extraction and conversion into point sets, the point sets corresponding to each image are used to perform one-to-one SC matching with the points corresponding to all images in the VID. The similarity of the two images can be represented by the SC distance between the two point sets. The smaller the value is, the more similar the two images are. The corresponding SC distance is calculated for each VID image. The image corresponding to the minimum value is the best matching result. After matching all RCIs, the results are as shown in Fig. 5.16 to 5.23. Each figure represents only one calyx, which contains three different RCIs. The most similar image matched from the VID is to the right of each RCI. Below those images are their corresponding edge features.

Computation time is an important indicator for the navigation system. If the computation time is not much, then the system has the opportunity to be used in the real-time intervention. The calculation time corresponding to each image is different. Fig. 5.24 shows the time it takes for each image to perform image matching.

Taking the number of edge feature points corresponding to the RCI as the x-axis and the calculation time as the y-axis, a graph of the relationship between the two can be formed (Fig. 5.25).

When the system is used in a real scenario, since it is assumed that the inside of the phantom is unknown, the EMT sensor needs to be used to verify the results. Through the calculated rotation, the translation matrix, the position data of the EMT sensor corresponding to the RCI in the EMT coordinate system can be transformed into the VURS coordinate system. The distance between the transformed data and the position data of the VC corresponding to the matching result picture is obtained and compared with the previously determined threshold so that the matching result can be verified. The distances results for each image are shown in Fig. 5.26, where the yellow horizontal line represents the threshold value.

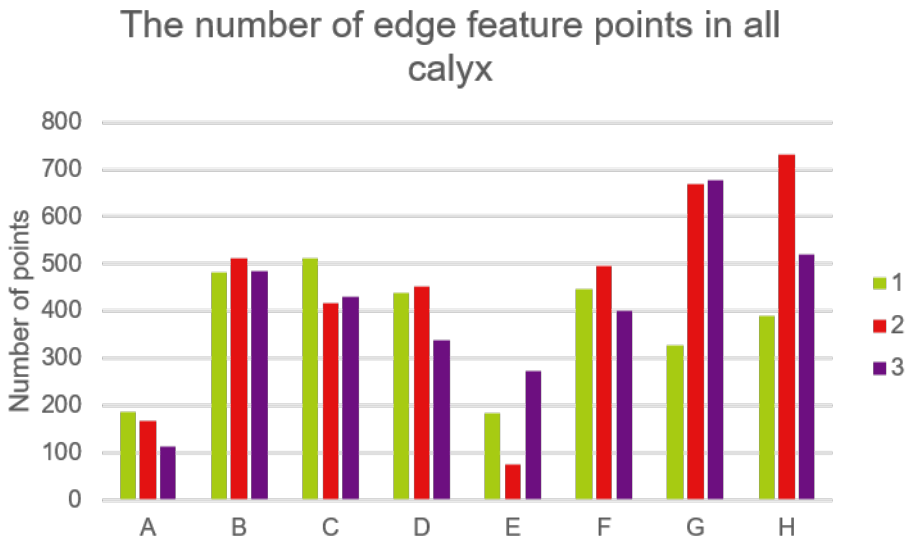

Figure 5.15: The overview of the number of edge feature points of all images for all calyxes. 1, 2, 3: the first, second, third image of each calyx.

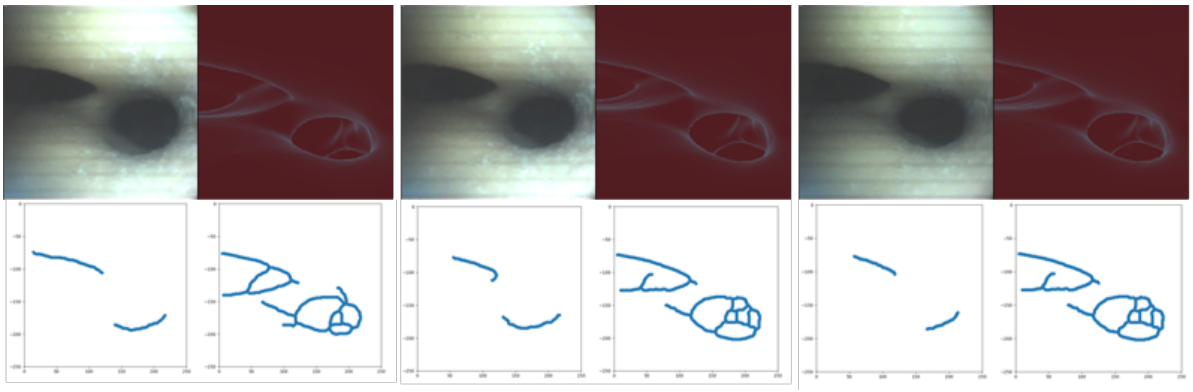

Figure 5.16: The results of calyx A after the image matching. There are three groups for each calyx. In each group, the left-top image is the RCI; the left-bottom image is its edge feature image; the right-top image is the resulting image in VID, and the right-bottom image is its edge feature image.

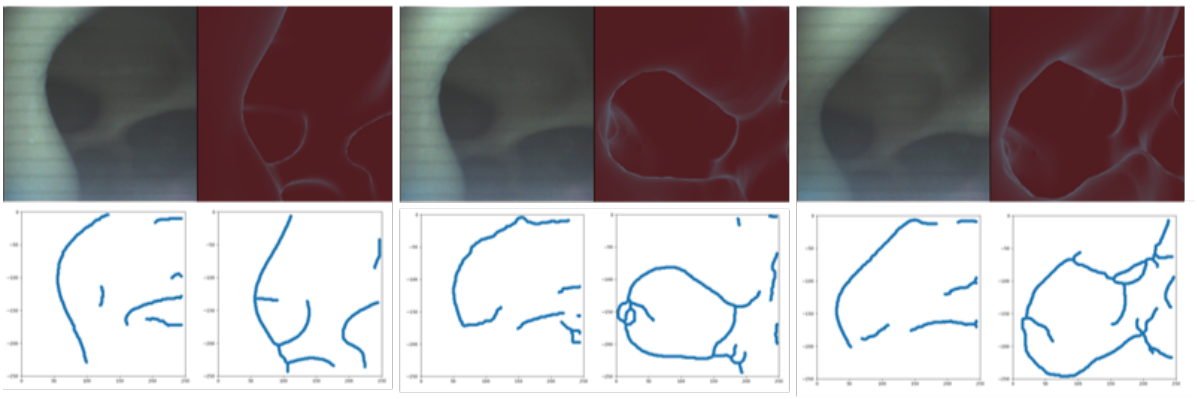

Figure 5.17: The results of calyx B after the image matching. The same structure to Fig. 5.16.

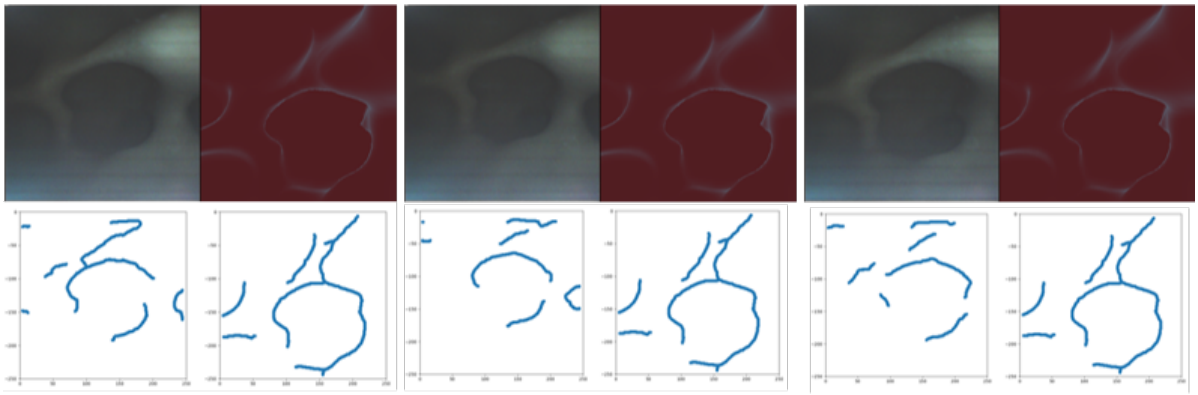

Figure 5.18: The results of calyx C after the image matching. The same structure to Fig. 5.16.

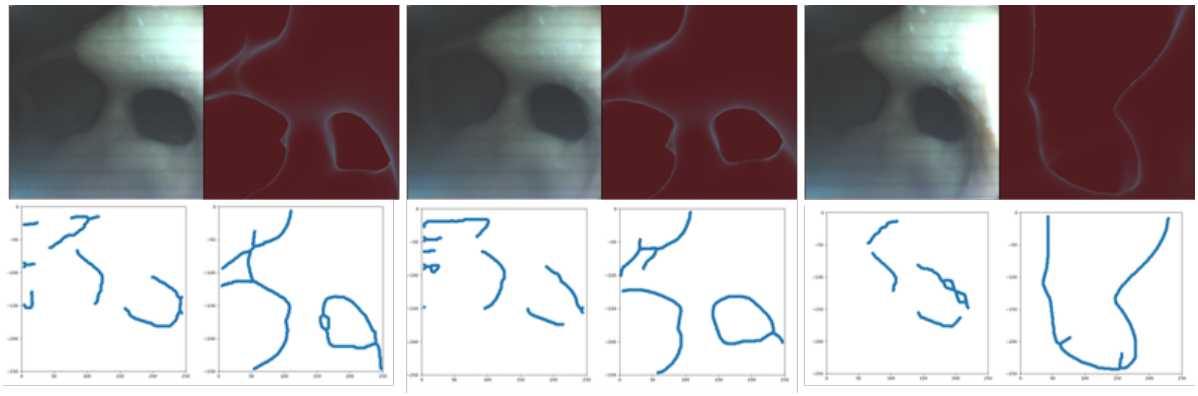

Figure 5.19: The results of calyx D after the image matching. The same structure to Fig. 5.16.

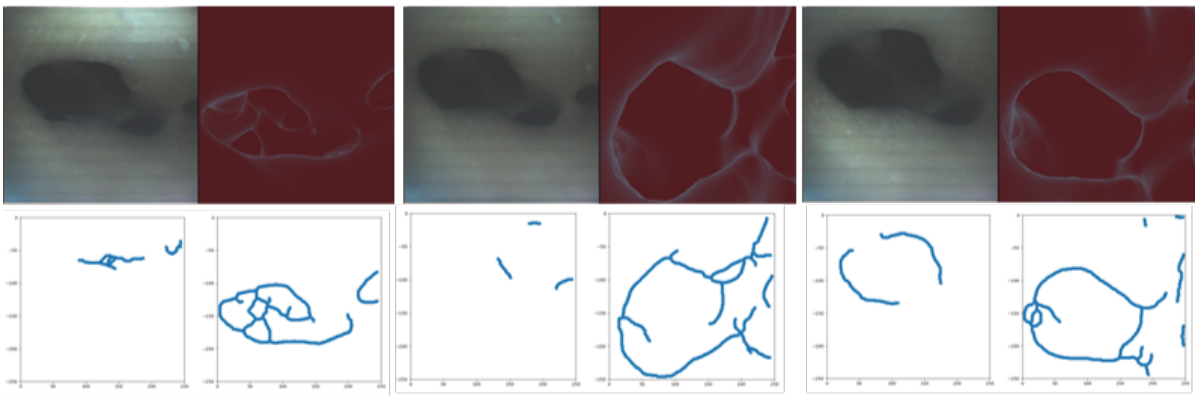

Figure 5.20: The results of calyx E after the image matching. The same structure to Fig. 5.16.

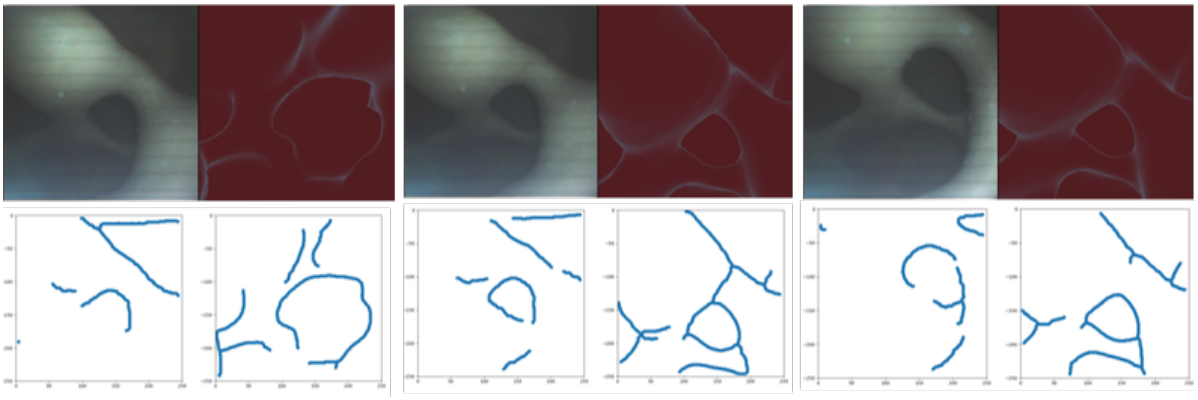

Figure 5.21: The results of calyx F after the image matching. The same structure to Fig. 5.16.

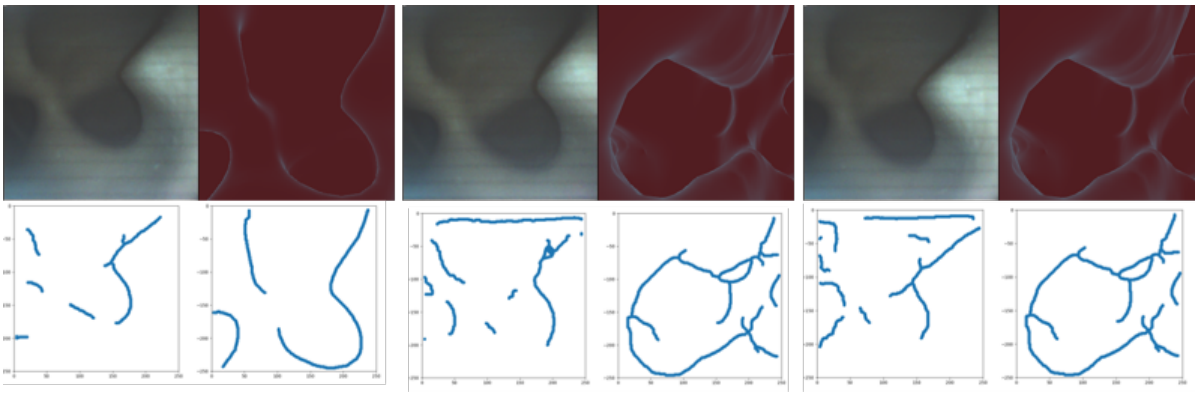

Figure 5.22: The results of calyx G after the image matching. The same structure to Fig. 5.16.

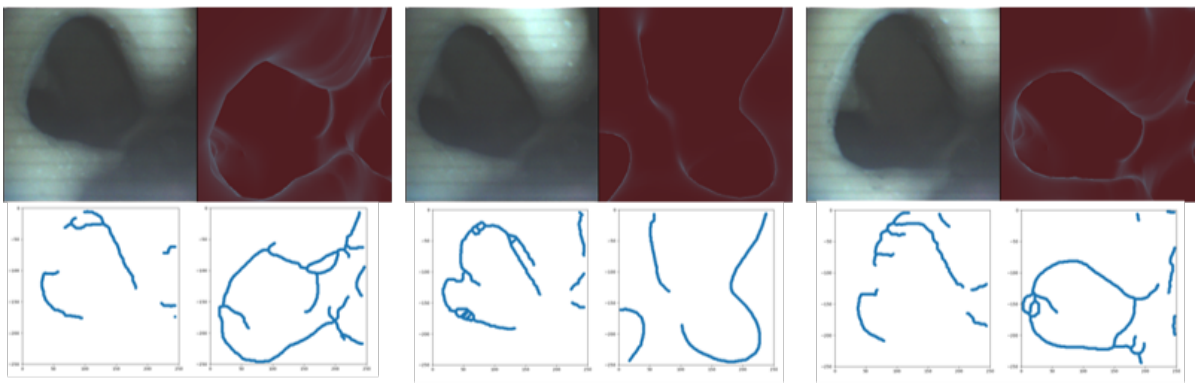

Figure 5.23: The results of calyx H after the image matching. The same structure to Fig. 5.16.

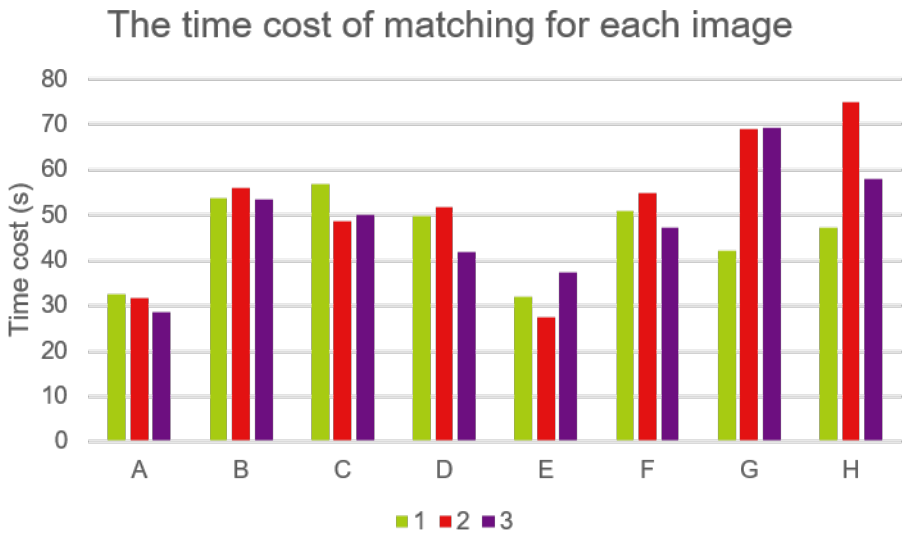

Figure 5.24: The overview of the time cost to perform matching algorithm for all images. 1, 2, 3: the first, second, third image of each calyx.

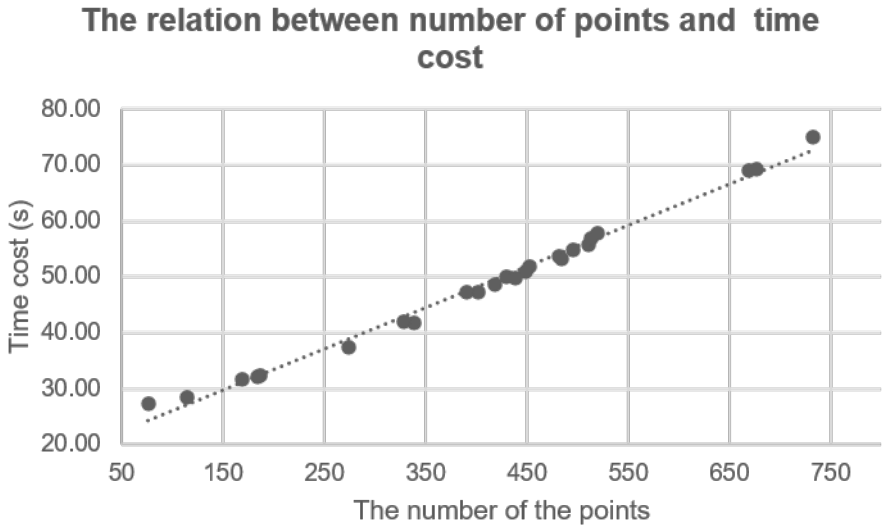

Figure 5.25: The plot of relation between the number of the edge feature points and time cost for all RCIs. The coefficient of determination, R^2, equals to 0.98. The dot line is its trend line after performing the regression.

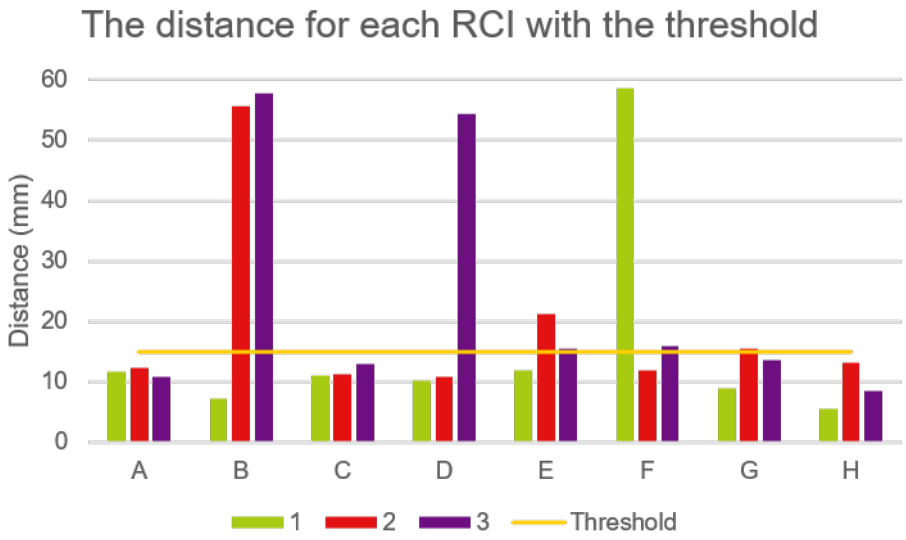

Figure 5.26: The overview for result validation by comparing the distance for each RCI with the threshold. A - H: eight types of RCI. 1, 2, 3: the first, second, third image of each calyx. The yellow line represents the threshold value.

5.3. Visualization in VURS environment

When matching images is done, one of the images in the VID that is considered the best match is the system result. The position and orientation information of the VC corresponding to the image will be extracted and transmitted to the VURS environment. The position in the kidney model will be displayed on the interface. Fig. 5.27 shows the examples of the display result of calyx A and calyx H in the VURS environment after obtaining the correct matching result.

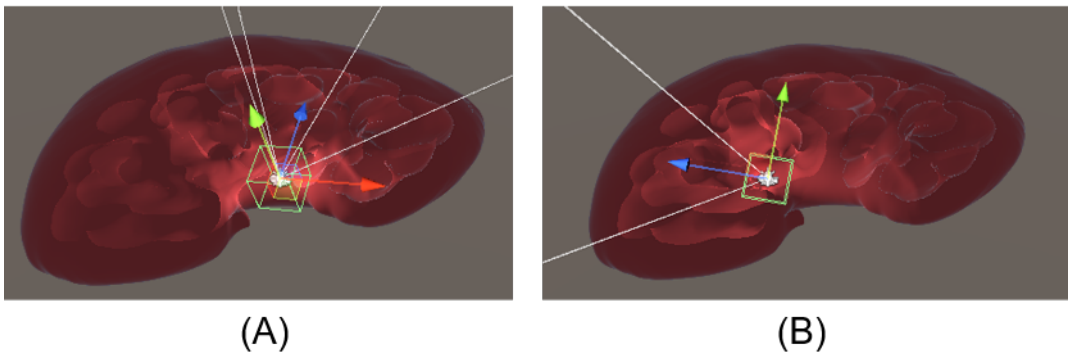

Figure 5.27: The visualization of the location result for calyxes in VURS environment. (A) The example for calyx A; (B) The example for calyx H. The white icons in both images represent the positions of VCs and the coordinate icons represent the orientations of VCs.

As seen from the figures, the kidney model is set to be translucent to allow users to understand the internal structure better. The white point represents the position of the VC, the coordinate system attached to it represents the orientation of the camera, and the divergent white lines identify the FOV of that VC. This global display can better guide the user's understanding of the real situation.

6

Discussion

6.1. Results analysis

This section presents some analysis of the results from the previous chapter.

6.1.1. The relation between time cost and number of points

It can be seen from Fig. 5.25 that the two data types are in a specific linear relationship. Their coefficient of determination, R^2, equals to 0.98, which means the regression predictions of the trend line fit the data well. It can be inferred that when matching the same VID, the calculation time increases linearly with the number of RCI points. This discovery can provide reference support for subsequent attempts to reduce time consumption. That is, the number of edge feature points required to complete the algorithm in a specific time can be obtained through this relationship.

6.1.2. The algorithm performance

From the results of matching simulation images, according to the ground truth, in the 24 simulation images, the number of the images that are correctly matched with VID images is 22. Thus, the performance accuracy is 92% (22/24). In addition, from the matching results of their edge features and the corresponding VID, there are many cases where the edges are not entirely collected due to noise or information that does not belong to the edge features, for example, Fig. 5.9. However, the algorithm still matches the correct calyx image from the VID, which shows that the algorithm can tolerate certain noise and shape deformation and has robustness.

For the real situation, two facts need to focus on. One is that there is a result conflict on judging whether the matching result is correct or not between the existing EMT sensor method and the ground truth. The other is how the system behaves in RCIs and the reason behind this.

Result conflict

There is a conflict between the system accuracy obtained by EMT and the ground truth. According to the matching results corresponding to all distances less than the threshold value as correct, the conclusion is that 16 of the 24 RCIs are correctly matched. However, according to the ground truth, 15 images were matched correctly. Although the numbers of correct results are close, the specific correct images differ. To represent this difference, the confusion matrix [133] can be used based on the ground truth:

$$\begin{aligned} TP &= 14, \quad FN = 1, \\ FP &= 2, \quad TN = 7, \end{aligned} \tag{6.1}$$

where TP, FN, FP, and TN represent true positive, false negative, false positive, and true negative. Among them, TP and FN reflect the number of correct judgments using the threshold. FP shows that two results are recognized as correct, while they are supposed to be incorrect based on the ground truth. FN shows one result is misidentified as incorrect.

The two images of FP are the third image of calyx G (Fig. 5.22), and the second image of calyx H (Fig. 5.23). From the original images, it can be observed that these RCIs have obvious shape

differences from the matching result images. Taking the second picture of calyx H as an example, the matching result image represents calyx G, and according to the standard of EMT judgment, the matching result is correct. The image with FN is the third image of calyx F (Fig. 5.21). There are two possible reasons for this.

First, the RCIs with larger distances than the threshold can still be correctly matched. This could be possible because the SC matching algorithm is scale-invariant and robust as long as the edge features are complete. Therefore, despite the real camera being further away from the object, making the resulting distances larger than the threshold, the calyx images are still correctly matched.

The second reason is the same ureteroscope position and different orientations, resulting in observing two different calyxes. Since the EMT sensor used in this experiment is 5 DOFs, the lack of a rotational degree of freedom makes it impossible to compare the orientation by converting it. Therefore, it may not be enough to judge only by the location information. The misjudgment of the third picture of calyx E is because of one reason. The misjudgment of the remaining three graphs is due to reason two, which is confirmed by looking at the kidney model in the VURS environment.

Performance in RCIs

Based on ground truth, the final image matching result is that among the 24 images, the number of images correctly matched to the corresponding VID is 15, and the correct rate is 62.5%. Compared with the simulation images, the performance accuracy is greatly reduced. By comparing the images of the extracted edge features, it can be found that the edge features extracted from the RCIs have more serious incompleteness and are disturbed by noise. A very intuitive example is calyx E. After image processing, only very few edge features are extracted, which are completely insufficient to describe the shape of this calyx.

Fortunately, even with considerable interference, as long as the image maintains a relatively clear contrast and minimizes the highlights caused by direct light sources, the image can still be correctly matched after image processing. Therefore, the details of image acquisition and image processing may need to be considered more.

6.2. Limitations and Future improvements

6.2.1. Image quality

From the comparison of the results of the RCIs and the simulation images, we can see that the accuracy of the whole system has a great relationship with the image quality. The quality of the images in the real phantom collected in this experiment is not satisfying, and the noise and interference are large, which are caused by the following two aspects: lighting source and camera.

This experiment selected LED light as the light source for illumination. Fibers, which should be more suitable for the size of the robot were discarded because the illumination was not bright enough. However, due to the large size of the LED lamp used in this experiment, there are not many options for its installation position, resulting in diffraction fringes and an obvious halo when the image is acquired. Although these disturbances can be resolved to some extent by subsequent image processing, they still make edge extraction difficult.

As one of the most critical devices in the whole system, the quality of the camera largely determines the quality of the image, which further affects the edge extraction and subsequent image matching. The resolution of the selected camera is only 250 by 250, and it has a 160-degree field of view, which results in images that are not clear and noisy.

Given these two points, in the future, the image quality will be improved by choosing a higher resolution camera within the budget and an illumination light source that is less disturbing but can guarantee the brightness of the image. In addition, how to reasonably arrange the positions of cameras and light sources at the end of the robot to extract complete edge information during image processing is another question that needs to be studied.

6.2.2. Image processing

Another critical step in this system is the image processing problem for RCIs. In the case of monocular vision, due to the different environments of each calyx, the orientation of the robot, and the lighting conditions, the image quality will be different. In this case, it is not a simple task to use a fixed image processing algorithm to deal with all possible images. Different images may show very different results

when the same processing method is used. For example, the main edges of some images become apparent. However, for other images, those points that should not be regarded as edge features also meet the conditions of edge extraction.

Although many studies have focused on how to flexibly adjust the parameters of image processing according to the image, and under certain conditions [79, 134, 105, 104, 106], the robustness of the algorithm can be improved. Nevertheless, there is still much room for improvement. Perhaps in the future, a set of expert systems can be developed by studying the images presented by standard ureteroscope products to solve image noise reduction, contrast adjustment, and edge extraction uniformly.

6.2.3. Image matching

It can be seen from the results that SC matching has certain robustness in dealing with shape deformation. However, using all the points to do the matching, the time consumed by the whole algorithm is very long. Since it is a navigation system, the ultimate goal must be processed in real-time. Nevertheless, the current computing time cannot realize real-time navigation. There may be two directions for improvement in the future: sampling points and screening.

The fewer the points, the smaller the computation and the shorter the computation time for the SC matching algorithm. Sampling the edge feature point set can reduce the number of points. However, how many sampling points to keep is a question that needs further study. When the number of sampling points decreases, edge information is bound to be missing or lost. While retaining more sampling points results in an increase in the amount of calculation. In addition, how to sample is also an important issue. If sampled uniformly, points that are important for some edges may be filtered out, such as sudden changes or points with the most significant curvature. Some sampling methods that preserve the characteristics of the curve have been proposed [135, 136, 137, 138, 139, 140], and these methods can be tried in the future.

This system selects the best matching result from the VID, which means that an RCI will match every image in the VID. This also increases the computation time. Then, preliminary filtering can be applied before matching to reduce images of obviously different VID images. This requires addressing how to define “obviously different”. One possible way is to filter out the calyx that is least likely to fit the distribution by comparing the histogram distributions of various calyx. In this way, fewer VID images need to perform SC matching, and the calculation time is also reduced.

6.2.4. VID collection

After the VURS environment is established, this thesis uses a manual collection of images of the VID corresponding to the calyx, which is very time-consuming and labor-intensive. This is unacceptable for the goal of applying the system to real-world surgery. The entire process, from the establishment of the VURS to the data collection of the VID, is the future goal to achieve full automation. The difficulty lies in how to collect the images of each calyx. To solve this problem, the VC can be guided to collect images utilizing path planning. This path can be obtained by centralizing kidney internal channels and calyxes. Before this, since the internal channels and calyxes are irregular, it may be necessary to perform cylinder matching for these positions in advance, thereby improving the feasibility of centralization. In addition, the movement ability and DOFs of the ureteroscope product also need to be considered to modify the path. This way, an image that is more in line with the realistic calyx angle can be collected to improve the success rate of matching.

6.2.5. Generalization

Since this solution is only verified on a single kidney, and due to robot limitations, all calyx images cannot be obtained, the whole system cannot be considered entirely feasible. On the one hand, in this kidney phantom example, the robot needs to be improved to capture all calyx images, implement the matching algorithm, and then evaluate its performance. On the other hand, more different kidneys are needed to test the performance of the system. This can evaluate whether the system can be generalized and collect more kidney calyx data for further research and the feasibility of combining learning-based methods with this proposed navigation system.

7

Conclusion

An innovative image-matching-based navigation system for robotic URS has been studied in this thesis. As many surgeons find it difficult to navigate the ureteroscope inside the kidney and to know which calyx of the kidney has been observed or is being observed during kidney exploration, this proposed system aims to help surgeons to solve this problem by reducing their burden.

This navigation system uses shape context matching between the RCIs and the VID collected from the VURS environment to find the best matching result. From the result, the real ureteroscope location can be known through its corresponding VC position in the VURS environment.

Both simulations and experiments are used to verify the feasibility and performance of the system. The simulation images are used to verify that the shape context matching-based system can find the correct matching target. Experiments are performed to verify the feasibility of the system under real phantom conditions, as well as to further study the performance of the system.

The system has successfully found a matching image between the highlights and the added noise of the simulation image accompanied by a point light source. For 24 calyx images, the matching accuracy is 90%. This proves the feasibility of the localization system based on image matching, and it is also reasonable to believe that the accuracy of the system is guaranteed when the quality of the RCI is close to the simulation images.

The experimental results using kidney phantom also confirmed the feasibility of the system. However, due to the image quality, the matching success rate for these 24 RCIs dropped to 62.5%.

After simulation and experimentation, some limitations and points that can be improved in the future were found. Higher resolution cameras can be used to improve image quality. The selection of the light source and its relative position to the camera can be further adjusted and studied to reduce the influence of the light source on the image. More stable and flexible image processing algorithms can be studied to ensure the integrity of the extracted edge features. The planned sampling of edge feature points can reduce the number of points used to calculate the shape context, thereby reducing the calculation time and making it possible to achieve real-time positioning.

This study builds a basic system framework, leaving a lot of room for further research and optimization, for example, automatically collecting database images, optimizing the design of the robot, the camera, and the light source, further researching the image matching algorithm to improve accuracy or reduce computing time, considering system performance in non-rigid environments, and applying them in vivo.

References

[1] Michelle López and Bernd Hoppe. “History, epidemiology and regional diversities of urolithiasis”. en. In: *Pediatric Nephrology* 25.1 (Jan. 2010), pp. 49–59. ISSN: 0931-041X, 1432-198X. DOI: 10.1007/s00467-008-0960-5. URL: http://link.springer.com/10.1007/s00467-008-0960-5 (visited on 08/06/2022).

[2] Tilahun Alelign and Beyene Petros. “Kidney Stone Disease: An Update on Current Concepts”. en. In: *Advances in Urology* 2018 (2018), pp. 1–12. ISSN: 1687-6369, 1687-6377. DOI: 10.1155/2018/3068365. URL: https://www.hindawi.com/journals/au/2018/3068365/ (visited on 08/02/2022).

[3] C. K. Chauhan, M. J. Joshi, and A. D. B. Vaidya. “Growth inhibition of Struvite crystals in the presence of herbal extract Commiphora wightii”. en. In: *Journal of Materials Science: Materials in Medicine* 20.S1 (Dec. 2009), pp. 85–92. ISSN: 0957-4530, 1573-4838. DOI: 10.1007/s10856-008-3489-z. URL: http://link.springer.com/10.1007/s10856-008-3489-z (visited on 08/06/2022).

[4] Orson W Moe. “Kidney stones: pathophysiology and medical management”. en. In: *The Lancet* 367.9507 (Jan. 2006), pp. 333–344. ISSN: 01406736. DOI: 10.1016/S0140-6736(06)68071-9. URL: https://linkinghub.elsevier.com/retrieve/pii/S0140673606680719 (visited on 08/06/2022).

[5] Victoriano Romero, Haluk Akpinar, and Dean G Assimos. “Kidney Stones: A Global Picture of Prevalence, Incidence, and Associated Risk Factors”. en. In: (2010), p. 11.

[6] Vidar O Edvardsson et al. “Temporal trends in the incidence of kidney stone disease”. en. In: *Kidney International* 83.1 (Jan. 2013), pp. 146–152. ISSN: 00852538. DOI: 10.1038/ki.2012.320. URL: https://linkinghub.elsevier.com/retrieve/pii/S0085253815556949 (visited on 08/06/2022).

[7] Said Abdallah Al-Mamari. *Urolithiasis in Clinical Practice*. en. In Clinical Practice. Cham: Springer International Publishing, 2017. ISBN: 978-3-319-62437-2. DOI: 10.1007/978-3-319-62437-2. URL: http://link.springer.com/10.1007/978-3-319-62437-2 (visited on 07/04/2021).

[8] Api Chewcharat and Gary Curhan. “Trends in the prevalence of kidney stones in the United States from 2007 to 2016”. en. In: *Urolithiasis* 49.1 (Feb. 2021), pp. 27–39. ISSN: 2194-7228, 2194-7236. DOI: 10.1007/s00240-020-01210-w. URL: https://link.springer.com/10.1007/s00240-020-01210-w (visited on 08/02/2022).

[9] Benjamin W. Turney et al. “Trends in urological stone disease: ANALYSIS OF 10 YEARS OF HES DATA”. en. In: *BJU International* 109.7 (Apr. 2012), pp. 1082–1087. ISSN: 14644096. DOI: 10.1111/j.1464-410X.2011.10495.x. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1464-410X.2011.10495.x (visited on 08/06/2022).

[10] Yu Liu et al. “Epidemiology of urolithiasis in Asia”. en. In: *Asian Journal of Urology* 5.4 (Oct. 2018), pp. 205–214. ISSN: 22143882. DOI: 10.1016/j.ajur.2018.08.007. URL: https://linkinghub.elsevier.com/retrieve/pii/S2214388218300729 (visited on 08/02/2022).

[11] Saeed R Khan et al. “Kidney stones”. In: *Nature reviews Disease primers* 2.1 (2016), pp. 1–23.

[12] Riaz Agha and Gordon Muir. “Does Laparoscopic Surgery Spell the end of the Open Surgeon?” en. In: 96 (2003), p. 3.

[13] Christian Türk et al. “EAU Guidelines on Interventional Treatment for Urolithiasis”. en. In: *European Urology* 69.3 (Mar. 2016), pp. 475–482. ISSN: 03022838. DOI: 10.1016/j.eururo.2015.07.041. URL: https://linkinghub.elsevier.com/retrieve/pii/S0302283815007009 (visited on 08/06/2022).

[14] Dzelaludin Junuzovic et al. “Evaluation of Extracorporeal Shock Wave Lithotripsy (ESWL): Efficacy in Treatment of Urinary System Stones”. en. In: *Acta Informatica Medica* 22.5 (2014), p. 309. ISSN: 0353-8109. DOI: 10.5455/aim.2014.22.309-314. URL: http://www.scopemed.org/fulltextpdf.php?mno=172064 (visited on 08/06/2022).

[15] Prahara Yuri et al. “Meta-analysis of Optimal Management of Lower Pole Stone of 10 - 20 mm: Flexible Ureteroscopy (FURS) versus Extracorporeal Shock Wave Lithotripsy (ESWL) versus Percutaneus Nephrolithotomy (PCNL)”. en. In: *Acta Med Indones* 50.1 (2018), p. 8.

[16] Robert M. Geraghty et al. “Ureteroscopy is more cost effective than shock wave lithotripsy for stone treatment: systematic review and meta-analysis”. en. In: *World Journal of Urology* 36.11 (Nov. 2018), pp. 1783–1793. ISSN: 0724-4983, 1433-8726. DOI: 10.1007/s00345-018-2320-9. URL: http://link.springer.com/10.1007/s00345-018-2320-9 (visited on 08/06/2022).

[17] Tobias M. Goodman. “Ureteroscopy with pediatric cystoscope in adults”. en. In: *Urology* 9.4 (Apr. 1977), p. 394. ISSN: 00904295. DOI: 10.1016/0090-4295(77)90213-8. URL: https://linkinghub.elsevier.com/retrieve/pii/0090429577902138 (visited on 08/06/2022).

[18] Edward S. Lyon, James S. Kyker, and Harry W. Schoenberg. “Transurethral Ureteroscopy in Women: A Ready Addition to the Urological Armamentarium”. en. In: *Journal of Urology* 119.1 (Jan. 1978), pp. 35–36. ISSN: 0022-5347, 1527-3792. DOI: 10.1016/S0022-5347(17)57372-3. URL: http://www.jurology.com/doi/10.1016/S0022-5347%2817%2957372-3 (visited on 08/06/2022).

[19] Demetrius H. Bagley and Brian Calio. “The History of the Development of Ureteral Endoscopy”. en. In: *Ureteroscopy*. Ed. by Bradley F. Schwartz and John D. Denstedt. Cham: Springer International Publishing, 2020, pp. 1–15. ISBN: 978-3-030-26649-3. DOI: 10.1007/978-3-030-26649-3_1. URL: https://link.springer.com/10.1007/978-3-030-26649-3_1 (visited on 08/06/2022).

[20] José Manuel Reis Santos. “Ureteroscopy from the recent past to the near future”. en. In: *Urolithiasis* 46.1 (Feb. 2018), pp. 31–37. ISSN: 2194-7228, 2194-7236. DOI: 10.1007/s00240-017-1016-8. URL: http://link.springer.com/10.1007/s00240-017-1016-8 (visited on 08/06/2022).

[21] Natalie Zelenko et al. “Normal Ureter Size on Unenhanced Helical CT”. en. In: *American Journal of Roentgenology* 182.4 (Apr. 2004), pp. 1039–1041. ISSN: 0361-803X, 1546-3141. DOI: 10.2214/ajr.182.4.1821039. URL: https://www.ajronline.org/doi/10.2214/ajr.182.4.1821039 (visited on 08/06/2022).

[22] Chad Gridley and Bodo Knudsen. “Digital ureteroscopes: technology update”. en. In: *Research and Reports in Urology* Volume 9 (Jan. 2017), pp. 19–25. ISSN: 2253-2447. DOI: 10.2147/RRU.S104229. URL: https://www.dovepress.com/digital-ureteroscopes--technology-update-peer-reviewed-article-RRU (visited on 07/07/2021).

[23] ATMOS MedizinTechnik GmbH & Co. KG. *ATMOS rigid ureteroscope product*. https://atmosmed.com/en/products-solutions/ent/imaging-diagnostics/endoscopes/starre-endoskope. 2020.

[24] KARL STORZ. *The New Flexible Video-Cystoscope C-VIEW*. https://www.karlstorz.com/us/en/highlights-uro.htm#mod-15265. 2022.

[25] OLYMPUS. *Olympus URF-P7/P7R*. https://www.olympus-europa.com/medical/en/Products-and-Solutions/Products/Ureteroscopes.html. 2022.

[26] Brian S. Peters et al. “Review of emerging surgical robotic technology”. en. In: *Surgical Endoscopy* 32.4 (Apr. 2018), pp. 1636–1655. ISSN: 0930-2794, 1432-2218. DOI: 10.1007/s00464-018-6079-2. URL: http://link.springer.com/10.1007/s00464-018-6079-2 (visited on 07/12/2021).

[27] ELMED Medical System. *Avicenna Roboflex - ELMED Medical System*. https://elmed-as.com/products/avicenna-roboflex/. 2020.

[28] European Association of Urology. *Ureteroscopy (URS)*. https://patients.uroweb.org/treatments/ureteroscopy/. 2022.

[29] Fabian Adams et al. “Soft 3D-printed phantom of the human kidney with collecting system”. In: *Annals of biomedical engineering* 45.4 (2017), pp. 963–972.

[30] Cyrill Stachniss. *Robotic mapping and exploration*. Vol. 55. Springer, 2009.

[31] T.D. Soper et al. “*In Vivo* Validation of a Hybrid Tracking System for Navigation of an Ultrathin Bronchoscope Within Peripheral Airways”. In: *IEEE Transactions on Biomedical Engineering* 57.3 (Mar. 2010), pp. 736–745. ISSN: 0018-9294, 1558-2531. DOI: 10.1109/TBME.2009.2034733. URL: http://ieeexplore.ieee.org/document/5290039/ (visited on 04/19/2022).

[32] Tassilo Klein et al. “Fiducial-Free Registration Procedure for Navigated Bronchoscopy”. en. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2007*. Ed. by David Hutchison et al. Vol. 4791. Series Title: Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 475–482. ISBN: 978-3-540-75757-3. DOI: 10.1007/978-3-540-75757-3_58. URL: http://link.springer.com/10.1007/978-3-540-75757-3_58 (visited on 04/19/2022).

[33] Xiongbiao Luo and Kensaku Mori. “Adaptive fiducial-free registration using multiple point selection for real-time electromagnetically navigated endoscopy”. In: ed. by Ziv R. Yaniv and David R. Holmes. San Diego, California, USA, Mar. 2014, 90360A. DOI: 10.1117/12.2043388. URL: http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2043388 (visited on 03/07/2022).

[34] Erlend Fagertun Hofstad et al. “Intraoperative localized constrained registration in navigated bronchoscopy”. en. In: *Medical Physics* 44.8 (Aug. 2017), pp. 4204–4212. ISSN: 0094-2405, 2473-4209. DOI: 10.1002/mp.12361. URL: https://onlinelibrary.wiley.com/doi/10.1002/mp.12361 (visited on 03/13/2022).

[35] Saeedeh Navaei Lavasani et al. “Bronchoscope motion tracking using centerline-guided Gaussian mixture model in navigated bronchoscopy”. In: *Physics in Medicine & Biology* 66.2 (Jan. 2021), p. 025001. ISSN: 0031-9155, 1361-6560. DOI: 10.1088/1361-6560/abca07. URL: https://iopscience.iop.org/article/10.1088/1361-6560/abca07 (visited on 04/19/2022).

[36] Masahiro Oda et al. “Colonoscope navigation system using colonoscope tracking method based on line registration”. In: ed. by Ziv R. Yaniv and David R. Holmes. San Diego, California, USA, Mar. 2014, p. 903626. DOI: 10.1117/12.2043426. URL: http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2043426 (visited on 03/14/2022).

[37] Yehuda Schwarz et al. “Real-Time Electromagnetic Navigation Bronchoscopy to Peripheral Lung Lesions Using Overlaid CT Images”. en. In: *Chest* 129.4 (Apr. 2006), pp. 988–994. ISSN: 00123692. DOI: 10.1378/chest.129.4.988. URL: https://linkinghub.elsevier.com/retrieve/pii/S0012369215388152 (visited on 04/19/2022).

[38] Masahiro Oda et al. “Position-based adjustment of landmark-based correspondence finding in electromagnetic sensor-based colonoscope tracking method”. In: ed. by Robert J. Webster and Ziv R. Yaniv. San Diego, California, United States, Mar. 2016, p. 97861D. DOI: 10.1117/12.2216371. URL: http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2216371 (visited on 03/14/2022).

[39] Masahiro Oda et al. “Machine learning-based colon deformation estimation method for colonoscope tracking”. In: *Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling*. Ed. by Robert J. Webster and Baowei Fei. Houston, United States: SPIE, Mar. 2018, p. 36. ISBN: 978-1-5106-1642-4. DOI: 10.1117/12.2293936. URL: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10576/2293936/Machine-learning-based-colon-deformation-estimation-method-for-colonoscope-tracking/10.1117/12.2293936.full (visited on 03/14/2022).

[40] Rahul Khare, Rebecca Bascom, and William E. Higgins. “Technician-free system for image-guided bronchoscopy”. In: ed. by David R. Holmes and Ziv R. Yaniv. Lake Buena Vista (Orlando Area), Florida, USA, Mar. 2013, p. 867101. DOI: 10.1117/12.2004880. URL: http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2004880 (visited on 03/13/2022).

[41] Patrick D. Byrnes and William E. Higgins. “Construction of a multimodal CT-video chest model”. In: ed. by Ziv R. Yaniv and David R. Holmes. San Diego, California, USA, Mar. 2014, p. 903607. DOI: 10.1117/12.2041609. URL: http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2041609 (visited on 03/13/2022).

[42] Xiaonan Zang et al. “3D endobronchial ultrasound reconstruction and analysis for multimodal image-guided bronchoscopy”. In: ed. by Johan G. Bosch and Marvin M. Doyley. San Diego, California, USA, Mar. 2014, p. 90400X. DOI: 10.1117/12.2036115. URL: http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2036115 (visited on 04/18/2022).

[43] William E. Higgins et al. “Multimodal system for the planning and guidance of bronchoscopy”. In: ed. by Robert J. Webster and Ziv R. Yaniv. Orlando, Florida, United States, Mar. 2015, p. 941508. DOI: 10.1117/12.2077242. URL: http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2077242 (visited on 12/05/2021).

[44] Marco Visentini-Scarzanella et al. “Deep monocular 3D reconstruction for assisted navigation in bronchoscopy”. en. In: *International Journal of Computer Assisted Radiology and Surgery* 12.7 (July 2017), pp. 1089–1099. ISSN: 1861-6410, 1861-6429. DOI: 10.1007/s11548-017-1609-2. URL: http://link.springer.com/10.1007/s11548-017-1609-2 (visited on 04/04/2022).

[45] Mali Shen et al. “Context-Aware Depth and Pose Estimation for Bronchoscopic Navigation”. In: *IEEE Robotics and Automation Letters* 4.2 (Apr. 2019), pp. 732–739. ISSN: 2377-3766, 2377-3774. DOI: 10.1109/LRA.2019.2893419. URL: https://ieeexplore.ieee.org/document/8613897/ (visited on 03/10/2022).

[46] Mali Shen, Stamatia Giannarou, and Guang-Zhong Yang. “Robust camera localisation with depth reconstruction for bronchoscopic navigation”. en. In: *International Journal of Computer Assisted Radiology and Surgery* 10.6 (June 2015), pp. 801–813. ISSN: 1861-6410, 1861-6429. DOI: 10.1007/s11548-015-1197-y. URL: http://link.springer.com/10.1007/s11548-015-1197-y (visited on 04/21/2022).

[47] Cheng Zhao et al. “Generative Localization With Uncertainty Estimation Through Video-CT Data for Bronchoscopic Biopsy”. In: *IEEE Robotics and Automation Letters* 5.1 (Jan. 2020), pp. 258–265. ISSN: 2377-3766, 2377-3774. DOI: 10.1109/LRA.2019.2955941. URL: https://ieeexplore.ieee.org/document/8913461/ (visited on 03/10/2022).

[48] Xinqi Liu et al. “Computer vision-guided bronchoscopic navigation using dual CNN-generated depth images and ICP registration”. In: *Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling*. Ed. by Baowei Fei and Cristian A. Linte. Houston, United States: SPIE, Mar. 2020, p. 77. ISBN: 978-1-5106-3398-8. DOI: 10.1117/12.2549719. URL: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/11315/2549719/Computer-vision-guided-bronchoscopic-navigation-using-dual-CNN-generated-depth/10.1117/12.2549719.full (visited on 03/11/2022).

[49] Peter Mountney et al. “Simultaneous Stereoscope Localization and Soft-Tissue Mapping for Minimal Invasive Surgery”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006*. Ed. by David Hutchison et al. Vol. 4190. Series Title: Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 347–354. ISBN: 978-3-540-44708-5. DOI: 10.1007/11866565_43. URL: http://link.springer.com/10.1007/11866565_43 (visited on 04/21/2022).

[50] Oscar G Grasa et al. “EKF monocular SLAM 3D modeling, measuring and augmented reality from endoscope image sequences”. In: *Medical image computing and computer-assisted intervention (MICCAI)*. Vol. 2. 2009.

[51] S. Speidel et al. “Robust feature tracking for endoscopic pose estimation and structure recovery”. In: ed. by David R. Holmes and Ziv R. Yaniv. Lake Buena Vista (Orlando Area), Florida, USA, Mar. 2013, p. 867102. DOI: 10.1117/12.2007278. URL: http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2007278 (visited on 03/10/2022).

[52] Bruce D Lucas, Takeo Kanade, et al. “An iterative image registration technique with an application to stereo vision”. In: Vancouver. 1981.

[53] Oscar G. Grasa et al. “Visual SLAM for Handheld Monocular Endoscope”. In: *IEEE Transactions on Medical Imaging* 33.1 (Jan. 2014), pp. 135–146. ISSN: 0278-0062, 1558-254X. DOI: 10.1109/TMI.2013.2282997. URL: http://ieeexplore.ieee.org/document/6605585/ (visited on 04/20/2022).

[54] Nader Mahmoud et al. “ORBSLAM-Based Endoscope Tracking and 3D Reconstruction”. In: *Computer-Assisted and Robotic Endoscopy*. Ed. by Terry Peters et al. Vol. 10170. Series Title: Lecture Notes in Computer Science. Cham: Springer International Publishing, 2017, pp. 72–83. ISBN: 978-3-319-54057-3. DOI: 10.1007/978-3-319-54057-3_7. URL: http://link.springer.com/10.1007/978-3-319-54057-3_7 (visited on 02/03/2022).

[55] Nader Mahmoud et al. “SLAM based Quasi Dense Reconstruction For Minimally Invasive Surgery Scenes”. en. In: *arXiv:1705.09107 [cs]* (May 2017). arXiv: 1705.09107. URL: http://arxiv.org/abs/1705.09107 (visited on 04/21/2022).

[56] Long Chen et al. “SLAM-based dense surface reconstruction in monocular Minimally Invasive Surgery and its application to Augmented Reality”. en. In: *Computer Methods and Programs in Biomedicine* 158 (May 2018), pp. 135–146. ISSN: 01692607. DOI: 10.1016/j.cmpb.2018.02.006. URL: https://linkinghub.elsevier.com/retrieve/pii/S0169260717301694 (visited on 02/02/2022).

[57] Wang Cheng et al. “Visual SLAM for bronchoscope tracking and bronchus reconstruction in bronchoscopic navigation”. In: *Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling*. Ed. by Baowei Fei and Cristian A. Linte. San Diego, United States: SPIE, Mar. 2019, p. 9. ISBN: 978-1-5106-2550-1. DOI: 10.1117/12.2512766. URL: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10951/2512766/Visual-SLAM-for-bronchoscope-tracking-and-bronchus-reconstruction-in-bronchoscopic/10.1117/12.2512766.full (visited on 03/11/2022).

[58] Cheng Wang et al. “Improved visual SLAM for bronchoscope tracking and registration with pre-operative CT images”. In: *Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling*. Ed. by Baowei Fei and Cristian A. Linte. Houston, United States: SPIE, Mar. 2020, p. 43. ISBN: 978-1-5106-3398-8. DOI: 10.1117/12.2549949. URL: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/11315/2549949/Improved-visual-SLAM-for-bronchoscope-tracking-and-registration-with-pre/10.1117/12.2549949.full (visited on 03/28/2022).

[59] Nicola Piccinelli et al. “Rigid 3D Registration of Pre-operative Information for Semi-Autonomous Surgery”. In: *2020 International Symposium on Medical Robotics (ISMR)*. Atlanta, GA, USA: IEEE, Nov. 2020, pp. 139–145. ISBN: 978-1-72815-488-6. DOI: 10.1109/ISMR48331.2020.9312949. URL: https://ieeexplore.ieee.org/document/9312949/ (visited on 04/21/2022).

[60] Jose Lamarca et al. “DefSLAM: Tracking and Mapping of Deforming Scenes From Monocular Sequences”. In: *IEEE Transactions on Robotics* 37.1 (Feb. 2021), pp. 291–303. ISSN: 1552-3098, 1941-0468. DOI: 10.1109/TRO.2020.3020739. URL: https://ieeexplore.ieee.org/document/9201190/ (visited on 02/02/2022).

[61] Jiayi Ma et al. “Image Matching from Handcrafted to Deep Features: A Survey”. en. In: *International Journal of Computer Vision* 129.1 (Jan. 2021), pp. 23–79. ISSN: 0920-5691, 1573-1405. DOI: 10.1007/s11263-020-01359-2. URL: https://link.springer.com/10.1007/s11263-020-01359-2 (visited on 07/13/2022).

[62] Xingyu Jiang et al. “A review of multimodal image matching: Methods and applications”. en. In: *Information Fusion* 73 (Sept. 2021), pp. 22–71. ISSN: 15662535. DOI: 10.1016/j.inffus.2021.02.012. URL: https://linkinghub.elsevier.com/retrieve/pii/S156625352100035X (visited on 06/23/2022).

[63] Hans P Moravec. “Techniques towards automatic visual obstacle avoidance”. In: (1977).

[64] C. Harris and M. Stephens. “A Combined Corner and Edge Detector”. en. In: *Procedings of the Alvey Vision Conference 1988*. Manchester: Alvey Vision Club, 1988, pp. 23.1–23.6. DOI: 10.5244/C.2.23. URL: http://www.bmva.org/bmvc/1988/avc-88-023.html (visited on 08/15/2022).

[65] Miroslav Trajković and Mark Hedley. “Fast corner detection”. en. In: *Image and Vision Computing* 16.2 (Feb. 1998), pp. 75–87. ISSN: 02628856. DOI: 10.1016/S0262-8856(97)00056-5. URL: https://linkinghub.elsevier.com/retrieve/pii/S0262885697000565 (visited on 08/22/2022).

[66] Ethan Rublee et al. “ORB: An efficient alternative to SIFT or SURF”. In: *2011 International Conference on Computer Vision*. Barcelona, Spain: IEEE, Nov. 2011, pp. 2564–2571. ISBN: 978-1-4577-1100-8. DOI: 10.1109/ICCV.2011.6126544. URL: http://ieeexplore.ieee.org/document/6126544/ (visited on 08/22/2022).

[67] D.G. Lowe. “Object recognition from local scale-invariant features”. In: *Proceedings of the Seventh IEEE International Conference on Computer Vision*. Kerkyra, Greece: IEEE, 1999, 1150–1157 vol.2. ISBN: 978-0-7695-0164-2. DOI: 10.1109/ICCV.1999.790410. URL: http://ieeexplore.ieee.org/document/790410/ (visited on 08/24/2022).

[68] David G. Lowe. “Distinctive Image Features from Scale-Invariant Keypoints”. en. In: *International Journal of Computer Vision* 60.2 (Nov. 2004), pp. 91–110. ISSN: 0920-5691. DOI: 10.1023/B:VISI.0000029664.99615.94. URL: http://link.springer.com/10.1023/B:VISI.0000029664.99615.94 (visited on 08/24/2022).

[69] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. “SURF: Speeded Up Robust Features”. en. In: *Computer Vision – ECCV 2006*. Ed. by Aleš Leonardis, Horst Bischof, and Axel Pinz. Vol. 3951. Series Title: Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 404–417. ISBN: 978-3-540-33833-8. DOI: 10.1007/11744023_32. URL: http://link.springer.com/10.1007/11744023_32 (visited on 08/24/2022).

[70] Shubhashree Savant. “A Review on Edge Detection Techniques for Image Segmentation”. en. In: 5 (2014), p. 3.

[71] J S Owotogbe, T S Ibiyemi, and B A Adu. “Edge Detection Techniques on Digital Images - A Review”. en. In: 4.11 (2019), p. 4.

[72] Lawrence G Roberts. “Machine perception of three-dimensional solids”. PhD thesis. Massachusetts Institute of Technology, 1963.

[73] Judith MS Prewitt et al. “Object enhancement and extraction”. In: *Picture processing and Psychopictorics* 10.1 (1970), pp. 15–19.

[74] Irwin Sobel, Gary Feldman, et al. “A 3x3 isotropic gradient operator for image processing”. In: *a talk at the Stanford Artificial Project in* (1968), pp. 271–272.

[75] David Marr and Ellen Hildreth. “Theory of edge detection”. In: *Proceedings of the Royal Society of London. Series B. Biological Sciences* 207.1167 (1980), pp. 187–217.

[76] John Canny. “A Computational Approach to Edge Detection”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-8.6 (Nov. 1986), pp. 679–698. ISSN: 0162-8828. DOI: 10.1109/TPAMI.1986.4767851. URL: https://ieeexplore.ieee.org/document/4767851 (visited on 09/03/2022).

[77] J. Kang and R. Doraiswami. “Real-time image processing system for endoscopic applications”. en. In: *CCECE 2003 - Canadian Conference on Electrical and Computer Engineering. Toward a Caring and Humane Technology (Cat. No.03CH37436)*. Vol. 3. Montreal, Que., Canada: IEEE, 2003, pp. 1469–1472. ISBN: 978-0-7803-7781-3. DOI: 10.1109/CCECE.2003.1226181. URL: http://ieeexplore.ieee.org/document/1226181/ (visited on 08/27/2022).

[78] M. Hafner et al. “Endoscopic Image Classification Using Edge-Based Features”. en. In: *2010 20th International Conference on Pattern Recognition*. Istanbul, Turkey: IEEE, Aug. 2010, pp. 2724–2727. ISBN: 978-1-4244-7542-1. DOI: 10.1109/ICPR.2010.667. URL: http://ieeexplore.ieee.org/document/5597011/ (visited on 08/27/2022).

[79] Mahdi Alizadeh et al. “Effects of improved Adaptive Gamma Correction Method on Wireless Capsule Endoscopy images: Illumination compensation and edge detection”. en. In: *20th Iranian Conference on Electrical Engineering (ICEE2012)*. Tehran, Iran: IEEE, May 2012, pp. 1544–1548. ISBN: 978-1-4673-1147-2. DOI: 10.1109/IranianCEE.2012.6292604. URL: http://ieeexplore.ieee.org/document/6292604/ (visited on 08/27/2022).

[80] Eliane Maria Loiola et al. “A survey for the quadratic assignment problem”. en. In: *European Journal of Operational Research* 176.2 (Jan. 2007), pp. 657–690. ISSN: 03772217. DOI: 10.1016/j.ejor.2005.09.032. URL: https://linkinghub.elsevier.com/retrieve/pii/S0377221705008337 (visited on 09/09/2022).

[81] Timothee Cour, Praveen Srinivasan, and Jianbo Shi. “Balanced Graph Matching”. en. In: (2007), p. 8.

[82] H.A. Almohamad and S.O. Duffuaa. “A linear programming approach for the weighted graph matching problem”. en. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15.5 (May 1993), pp. 522–525. ISSN: 01628828. DOI: 10.1109/34.211474. URL: http://ieeexplore.ieee.org/document/211474/ (visited on 09/10/2022).

[83] Qing Zhao et al. “Semidefinite Programming Relaxations for the Quadratic Assignment Problem”. en. In: *Journal of Combinatorial Optimization* (1998), p. 39.

[84] M. Zaslavskiy, F. Bach, and J.-P. Vert. “A Path Following Algorithm for the Graph Matching Problem”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31.12 (Dec. 2009), pp. 2227–2242. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2008.245. URL: http://ieeexplore.ieee.org/document/4641936/ (visited on 09/10/2022).

[85] Zhi-Yong Liu and Hong Qiao. “GNCCP—Graduated NonConvexityand Concavity Procedure”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36.6 (June 2014), pp. 1258–1267. ISSN: 0162-8828, 2160-9292. DOI: 10.1109/TPAMI.2013.223. URL: http://ieeexplore.ieee.org/document/6654127/ (visited on 09/10/2022).

[86] T.S. Caetano et al. “Learning Graph Matching”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31.6 (June 2009), pp. 1048–1058. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2009.28. URL: http://ieeexplore.ieee.org/document/4770108/ (visited on 09/10/2022).

[87] Marius Leordeanu, Rahul Sukthankar, and Martial Hebert. “Unsupervised Learning for Graph Matching”. en. In: *International Journal of Computer Vision* 96.1 (Jan. 2012), pp. 28–45. ISSN: 0920-5691, 1573-1405. DOI: 10.1007/s11263-011-0442-2. URL: http://link.springer.com/10.1007/s11263-011-0442-2 (visited on 09/10/2022).

[88] Paul J. Besl and Neil D. McKay. “Method for registration of 3-D shapes”. In: ed. by Paul S. Schenker. Boston, MA, Apr. 1992, pp. 586–606. DOI: 10.1117/12.57955. URL: http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=981454 (visited on 09/09/2022).

[89] Sébastien Granger and Xavier Pennec. “Multi-scale EM-ICP: A Fast and Robust Approach for Surface Registration”. en. In: *Computer Vision — ECCV 2002*. Ed. by Gerhard Goos et al. Vol. 2353. Series Title: Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, pp. 418–432. ISBN: 978-3-540-47979-6. DOI: 10.1007/3-540-47979-1_28. URL: http://link.springer.com/10.1007/3-540-47979-1_28 (visited on 09/10/2022).

[90] Andrew W Fitzgibbon. “Robust registration of 2D and 3D point sets”. en. In: *Image and Vision Computing* 21.13-14 (Dec. 2003), pp. 1145–1153. ISSN: 02628856. DOI: 10.1016/j.imavis.2003.09.004. URL: https://linkinghub.elsevier.com/retrieve/pii/S0262885603001835 (visited on 09/10/2022).

[91] Steven Gold et al. “NEW ALGORITHMS FOR 2D AND 3D POINT MATCHING: POSE ESTIMATION AND CORRESPONDENCE”. en. In: (1998), p. 13.

[92] S. Belongie, J. Malik, and J. Puzicha. “Shape matching and object recognition using shape contexts”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24.4 (Apr. 2002), pp. 509–522. ISSN: 01628828. DOI: 10.1109/34.993558. URL: http://ieeexplore.ieee.org/document/993558/ (visited on 07/12/2022).

[93] Embodi3D LLC. *Renal Cortex 1.0.0 by michaelmplatt*. https://www.embodi3d.com/files/file/9538-renal-cortex/. 2022.

[94] Formlabs. *Form 3B+, An Advanced Desktop 3D Printer Designed for Healthcare*. https://formlabs.com/3d-printers/form-3b/?utm_source=preform. 2022.

[95] Formlabs. *Model Resin V2*. https://dental.formlabs.com/store/materials/model-resin-v2/. 2022.

[96] Formlabs. *Formlabs Software, 3D Printing Setup, Management, and Monitoring Made Simple*. https://formlabs.com/software/. 2022.

[97] Samuel Axon. *Unity at 10: For better—or worse—game development has never been easier*. https://arstechnica.com/gaming/2016/09/unity-at-10-for-better-or-worse-game-development-has-never-been-easier/. 2016.

[98] Unity Technologies. *Unity for Government & Aerospace*. https://unity.com/solutions/government-aerospace. 2022.

[99] Unity Technologies. *Unity 2021 Gaming Report*. https://create.unity.com/2021-game-report. 2021.

[100] Unity Technologies. *Unity Gaming Report 2022*. https://create.unity.com/gaming-report-2022?elqTrackId=ab6497d32403430b9ce104ffc2415a00&elq=00000000000000000000000000000000&elqaid=3200&elqat=2&elqCampaignId=. 2022.

[101] Tonichi Edeza. *Image Processing with Python — Application of Fourier Transformation*. https://towardsdatascience.com/image-processing-with-python-application-of-fourier-transformation-5a8584dc175b. 2021.

[102] Open Source Computer Vision. *Histogram Equalization*. https://docs.opencv.org/4.5.5/d4/d1b/tutorial_histogram_equalization.html. 2021.

[103] Mei Fang, GuangXue Yue, and QingCang Yu. "The Study on An Application of Otsu Method in Canny Operator". en. In: (2009), p. 4.

[104] Yuan-Kai Huo et al. "An adaptive threshold for the Canny Operator of edge detection". en. In: *2010 International Conference on Image Analysis and Signal Processing*. Zhejiang, China: IEEE, 2010, pp. 371–374. ISBN: 978-1-4244-5554-6. DOI: 10.1109/IASP.2010.5476095. URL: http://ieeexplore.ieee.org/document/5476095/ (visited on 09/07/2022).

[105] Weibin Rong et al. "An Improved Canny Edge Detection Algorithm". en. In: (2014), p. 6.

[106] Ferdous Hossain et al. "Dynamic Thresholding based Adaptive Canny Edge Detection". en. In: *International Journal of Computer Applications* 135.4 (Feb. 2016), pp. 37–41. ISSN: 09758887. DOI: 10.5120/ijca2016908337. URL: http://www.ijcaonline.org/research/volume135/number4/hossain-2016-ijca-908337.pdf (visited on 09/07/2022).

[107] Open Source Computer Vision. *Canny Edge Detector*. https://docs.opencv.org/4.5.5/da/d5c/tutorial_canny_detector.html. 2021.

[108] Nobuyuki Otsu. "A TIreshold Selection Method from Gray-Level Histograms". en. In: (1979), p. 5.

[109] Costanza Culmone et al. "Exploring non-assembly 3D printing for novel compliant surgical devices". en. In: *PLOS ONE* 15.5 (May 2020). Ed. by Tommaso Ranzani, e0232952. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0232952. URL: https://dx.plos.org/10.1371/journal.pone.0232952 (visited on 09/10/2022).

[110] Bio-Inspired Technology Group. *BITE- Bio-Inspired Technology*. https://www.bitegroup.nl/. 2022.

[111] ams-OSRAM AG. *ams OSRAM is a global leader in optical solutions*. https://ams-osram.com/. 2022.

[112] Boston Scientific Corporation. *LithoVue™ Single-Use Digital Flexible Ureteroscope*. https://www.bostonscientific.com/en-US/products/Ureteroscopes/LithoVue.html. 2022.

[113] STEPPERONLINE. *Nema 11 Closed-loop Geared Stepper*. https://www.omc-stepperonline.com/nema-11-closed-loop-geared-stepper-l-31mm-gear-ratio-27-1-encoder-300cpr-11hs12-0674d-pg27-e22-300. 2022.

[114] Amazon.com, Inc. *RATTMMOTOR 400 mm Travel Length Linear Rail*. https://www.amazon.de/-/en/gp/product/B08JYMP6BC/ref=ppx_yo_dt_b_asin_title_o00_s00?ie=UTF8&psc=1. 2022.

[115] Arduino. *Arduino-Due*. https://store.arduino.cc/products/arduino-due. 2021.

[116] Alfred M. Franz et al. “Electromagnetic Tracking in Medicine—A Review of Technology, Validation, and Applications”. en. In: *IEEE Transactions on Medical Imaging* 33.8 (Aug. 2014), pp. 1702–1725. ISSN: 0278-0062, 1558-254X. DOI: 10.1109/TMI.2014.2321777. URL: http://ieeexplore.ieee.org/document/6810177/ (visited on 09/10/2022).

[117] Frederick Raab et al. “Magnetic Position and Orientation Tracking System”. In: *IEEE Transactions on Aerospace and Electronic Systems* AES-15.5 (Sept. 1979), pp. 709–718. ISSN: 0018-9251. DOI: 10.1109/TAES.1979.308860. URL: http://ieeexplore.ieee.org/document/4102227/ (visited on 09/10/2022).

[118] Jack B. Kuipers. “SPASYN-an electromagnetic relative position and orientation tracking system”. In: *IEEE Transactions on Instrumentation and Measurement* 29.4 (Dec. 1980), pp. 462–466. ISSN: 0018-9456. DOI: 10.1109/TIM.1980.4314980. URL: http://ieeexplore.ieee.org/document/4314980/ (visited on 09/10/2022).

[119] Kurt Schicho et al. “Stability of miniature electromagnetic tracking systems”. In: *Physics in Medicine and Biology* 50.9 (May 2005), pp. 2089–2098. ISSN: 0031-9155, 1361-6560. DOI: 10.1088/0031-9155/50/9/011. URL: https://iopscience.iop.org/article/10.1088/0031-9155/50/9/011 (visited on 09/10/2022).

[120] Northern Digital Inc. *Aurora - NDI*. https://www.ndigital.com/electromagnetic-tracking-technology/aurora/. 2022.

[121] Polhemus. *Polhemus Electromagnetic Tracking System*. https://polhemus.com/applications/electromagnetics/. 2022.

[122] Open Source Robotics Foundation. *ROS - Robot Operating System*. https://www.ros.org/. 2021.

[123] Wikipedia. *Robot Operating System - Wikipedia*. https://en.wikipedia.org/wiki/Robot_Operating_System. 2022.

[124] VMware, Inc. *Download VMware Workstation Player*. https://www.vmware.com/products/workstation-player/workstation-player-evaluation.html. 2022.

[125] ams-OSRAM AG. *NanEye, Miniature CMOS image senso*. https://ams.com/naneye#tab/tools. 2022.

[126] OpenCV team. *OpenCV*. https://opencv.org/. 2022.

[127] Scikit-image development team. *Scikit-Image, Image processing in Python*. https://scikit-image.org/. 2022.

[128] Donald Danforth Plant Science Center. *PlantCV*. https://plantcv.readthedocs.io/en/stable/. 2020.

[129] K. S. Arun, T. S. Huang, and S. D. Blostein. “Least-Squares Fitting of Two 3-D Point Sets”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-9.5 (Sept. 1987), pp. 698–700. ISSN: 0162-8828. DOI: 10.1109/TPAMI.1987.4767965. URL: https://ieeexplore.ieee.org/document/4767965 (visited on 09/11/2022).

[130] Olivier D Faugeras and Martial Hebert. “A 3-D recognition and positioning algorithm using geometrical matching between primitive surfaces”. In: *Proceedings of the Eighth international joint conference on Artificial intelligence-Volume 2*. 1983, pp. 996–1002.

[131] TS Huang, SD Blostein, and EA Margerum. “Least-squares estimation of motion parameters from 3-D point correspondences”. In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition*. Vol. 10. IEEE Computer Soc. Press Washington DC. 1986, pp. 112–115.

[132] Nghia. *GitHub - nghiaho12/rigid_transform_3D*. https://github.com/nghiaho12/rigid_transform_3D. 2021.

[133] Stephen V Stehman. “Selecting and interpreting measures of thematic classification accuracy”. In: *Remote sensing of Environment* 62.1 (1997), pp. 77–89.

[134] Guo Zhang et al. “A Medical Endoscope Image Enhancement Method Based on Improved Weighted Guided Filtering”. en. In: *Mathematics* 10.9 (Apr. 2022), p. 1423. ISSN: 2227-7390. DOI: 10.3390/math10091423. URL: https://www.mdpi.com/2227-7390/10/9/1423 (visited on 08/27/2022).

[135] V. Singh and N. Rajpal. “Efficient representation of 2 D contour using non uniform sampling”. In: *2004 International Conference on Signal Processing and Communications, 2004. SPCOM ’04.* Bangalore, India: IEEE, 2004, pp. 37–40. ISBN: 978-0-7803-8674-7. DOI: 10.1109/SPCOM.2004.1458352. URL: http://ieeexplore.ieee.org/document/1458352/ (visited on 07/25/2022).

[136] Kazuya Ose, Kazunori Iwata, and Nobuo Suematsu. “A sampling method for processing contours drawn with an uncertain stroke order and number”. In: *2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA)*. Nagoya, Japan: IEEE, May 2017, pp. 468–471. ISBN: 978-4-901122-16-0. DOI: 10.23919/MVA.2017.7986902. URL: http://ieeexplore.ieee.org/document/7986902/ (visited on 07/25/2022).

[137] Luca Pagani and Paul J. Scott. “Curvature based sampling of curves and surfaces”. en. In: *Computer Aided Geometric Design* 59 (Jan. 2018), pp. 32–48. ISSN: 01678396. DOI: 10.1016/j.cagd.2017.11.004. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167839617301462 (visited on 07/26/2022).

[138] Kazuya Ose, Kazunori Iwata, and Nobuo Suematsu. “Sampling Shape Contours Using Optimization over a Geometric Graph”. en. In: *IEICE Transactions on Information and Systems* E102.D.12 (Dec. 2019), pp. 2547–2556. ISSN: 0916-8532, 1745-1361. DOI: 10.1587/transinf.2018EDP7353. URL: https://www.jstage.jst.go.jp/article/transinf/E102.D/12/E102.D_2018EDP7353/_article (visited on 07/25/2022).

[139] Lizheng Lu and Shiqing Zhao. “High-quality point sampling for B-spline fitting of parametric curves with feature recognition”. en. In: *Journal of Computational and Applied Mathematics* 345 (Jan. 2019), pp. 286–294. ISSN: 03770427. DOI: 10.1016/j.cam.2018.04.008. URL: https://linkinghub.elsevier.com/retrieve/pii/S0377042718301766 (visited on 07/25/2022).

[140] Xi Cheng et al. “Efficient adaptive sampling methods based on deviation analysis for on-machine inspection”. en. In: *Measurement* 188 (Jan. 2022), p. 110497. ISSN: 02632241. DOI: 10.1016/j.measurement.2021.110497. URL: https://linkinghub.elsevier.com/retrieve/pii/S0263224121013804 (visited on 07/25/2022).