Deformable Registration of Sparsely Textured Surfaces for Surgery Assistance

(手術支援のための疎なテクスチャを持つ面の変形レジストレーション)

Su Wai Tun

A Dissertation Submitted to
the Graduate School of Science and Engineering
in Partial Fulfillment of the Requirements for the Degree of
DOCTOR OF ENGINEERING
in
Mathematics, Electronics and Informatics

Supervisor: Professor Takashi Komuro

Saitama University, Japan

September 2022

Contents

List of Figures

List of Tables

Acknowledgements

I would like to express my deepest appreciation to my supervisor Prof. Takashi Komuro for his continuous support, kindness, patience, and understanding of my Ph.D. study. His guidance and immense knowledge helped me in all the time of research and writing of this thesis. Furthermore, I would like to extend my sincere thanks to him for giving me an opportunity as a research assistant in his laboratory. I could not have undertaken this Ph.D. journey without his support.

Special thanks also go to Prof. Hajime Nagahara from Osaka University, for his valuable advice on my research work.

I am also thankful to my dissertation committee, Prof. Tetsuya Shimamura, Prof. Yoshinori Kobayashi, and Prof. Yasuhiro Matsunaga for their valuable suggestions, and insightful comments on this research work.

Many thanks to JASSO, IUCHI, and JGC-S scholarship foundations for their financial support.

Special heartfelt thanks to my parents and my husband for their love and great support to my study.

A very special thanks to my loving father in heaven. This thesis is dedicated to him along my Ph.D. journey. Hope he is watching me from above and smiling for what I have accomplished.

Abstract

In this thesis, we propose two deformable registration methods that can register the sparsely textured surfaces for surgery assistance. The registration processes of the methods are differently performed based on two approaches such as the intensity-based approach and the feature-based approach.

Deformable registration has been one of the challenges in modern medical image analysis due to multimodality imaging, unevenness of anatomical structures, and deformation of the body or organs. In the medical Augmented Reality (AR) field, this technology and its accuracy plays important role when preoperative images such as tumors or vessels are superimposed with the alignment onto the surgery scene to enhance the visualization of the internal structures during the laparoscopic/endoscopic surgery. To track the transformation of the deformable surface in the registration process, the texture of the surface is essential because it can give the useful information such as spatial arrangement of color or intensity of an image to define the objects or interest of the region. Therefore, if the surface has no sufficient texture or minimal texture, it becomes more challenging to perform the registration. Although many approaches have been proposed to address deformation registration, most of them can fail if the surface has poorly texture.

In this thesis, firstly we propose a deformable registration method to register both texture and shape of the deformable surface of densely and sparsely textured objects based on an intensity-based approach. We use Time-of-Flight (ToF) camera to obtain the texture and shape information from the objects. To find the transformation of the deformable surfaces, firstly we set the 2D vertex position of the mesh on the initial image. The registration of this approach is performed by finding the corresponding 2D vertex position. For texture registration, we define the cost function to measure the texture similarity within the meshes between the initial image and the input images. For shape registration, the cost function is defined to compare the inter-vertex length of the meshes between the initial image and input images since we assume that the deformed surfaces should not stretch or shrink. To minimize the cost functions, the gradient descent method is performed with respect to the vertices of the meshes and updated the 2D vertices of the meshes in terms of the gradient calculation. Since the first approach is based on the intensity-based approach, the process is iteratively performed until the cost function is approximately converged. We confirmed the performance of our approach by comparing the methods using shape or texture information on densely

and sparsely textured papers and endoscopic stereo video dataset and our approach can reconstruct 3D shape of sparsely textured surfaces. Therefore, our proposed method can register the objects with sparse texture and little concavity/convexity.

Then, we propose a deformable registration method to register the sparsely texture of the deformable surfaces such as human organs based on the feature-based approach. In this approach, the most distinctive image features are used to track the transformation of the sparsely textured surfaces. Firstly, we apply an accelerated-KAZE (AKAZE) feature detector onto the initial image and input images. All detected features in the initial image are divided into blocks based on their coordinates. To define the most distinctive features among the detected feature points, response values which are one of the properties of the AKAZE features are applied. Therefore, the best feature point in each block of the initial image is fetched by their response values. Then, the correspondent feature points between the blockwise features of the initial image,and all detected features of the input images are detected to track the transformation of the deformable surfaces. Therefore, feature matching is performed by a brute-force matching algorithm to find the correspondence between the initial image and input images. The registration process of this approach is finding the correspondence between blockwise features of the initial image and all detected features of the current input image. Based on this approach, we also demonstrate the surgery assistance by overlaying a tumor image onto the deformable sparsely textured surfaces using the bilinear interpolation method. We show that the effectiveness of our approach using three endoscopic stereo video datasets that have sparsely texture.

As the results of the experiments, our approaches can track the transformation of the sparsely textured surfaces and can be applied to human organs for surgery assistance. In the first approach, the peak signal to noise ratio (PSNR) was calculated to evaluate the performance of registration in term of texture information. As the results, our proposed method was as good as the method using texture only information and the method using texture only information was better than the method using shape-only information. Then, Mean Square Error (MSE) was calculated to evaluate the performance of the registration in term of the shape information. The results showed that our proposed method was better than the method using texture-only information although it was not as good as the method using shape-only information. Therefore, our first proposed method has good registration capability in term of texture and shape of the sparsely textured surfaces.

In the second approach, PSNR was calculated for the evaluation of the registra-

tion performance in term of the texture information. Since we demonstrated the effectiveness of our proposed method using three endoscopic stereo video datasets, PSNR results of all datasets existed in the range of typical PSNR values. In the qualitative results of superimposing a tumor image onto the organs as the demonstration of surgery assistance, we can see that the tumor image can move along with the movement of the organs. Therefore, the second approach has good registration performance on sparsely textured surfaces of the deformable medical images.

Chapter 1

Introduction

1.1 Background

Laparoscopic / endoscopic abdominal surgery remains a challenging environment for computer vision tasks. According to these surgical techniques, the surgical tools with a tiny camera is inserted into the abdomen and these allow the surgeon to see and operate the abdominal cavity that display on the monitor screen acquired from the camera. From the surgical point of view, the visualization of the surgical scene is complex and poor due to these facts such as large occlusion occurring when interacting with surgical tools, illumination variations, organ bleeding, and organ deformation due to respiration motion and heart beating. In recent years, augmented reality (AR) technology has been proposed to enhance the visualization of the internal structure of the organ during the surgery by superimposing intraoperative images such as tumors or vessels that are aligned onto the surgical scene. The relationship between intraoperative image and interoperative image can be established by using registration technique.

Image registration, the processing of establishing the correspondences between images at different times, and from different perceptive, is one of the key technologies in many research fields such as computer vision, medical imaging, and remote sensing. Image registration is very challenging task due to complexity of image similarity measurement and spatial transformation or deformation. The most image registration methods consist of four basic steps.

1. Feature Detection The salient features like regions, line features, edges, and corners are identified in both the initial image and input images. These features are used for further processing and represented by their point representative

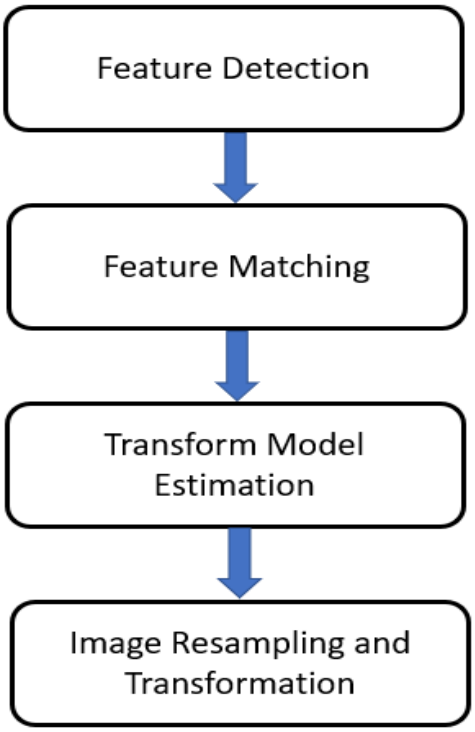


Figure 1.1: Basic Steps of Image Registration Process

(line endings, and distinctive points) that are called control points.

2. Feature Matching The correspondence between the features in the initial image and input image is established. Matching method is based on image content or on the symbolic description of control point-set.

3. Transform Model Estimation The parameters of the mapping functions aligning the input image with the initial image are estimated.

4. Image resampling and Transformation The input image is transformed by means of the mapping functions.

1.1.1 Classification of Registration Techniques

There are many number of registration algorithms that are designed and developed based on the application requirements of image registrations techniques [18, 31]. These are all presented in this section.

1. Intensity-based Methods Intensity-based methods are iterative process. In this approach, entire images or sub-images are registered. The intensity patterns in each image are matched and the measurement of the intensity similarity between the initial image and input images is defined in terms of the modality of the

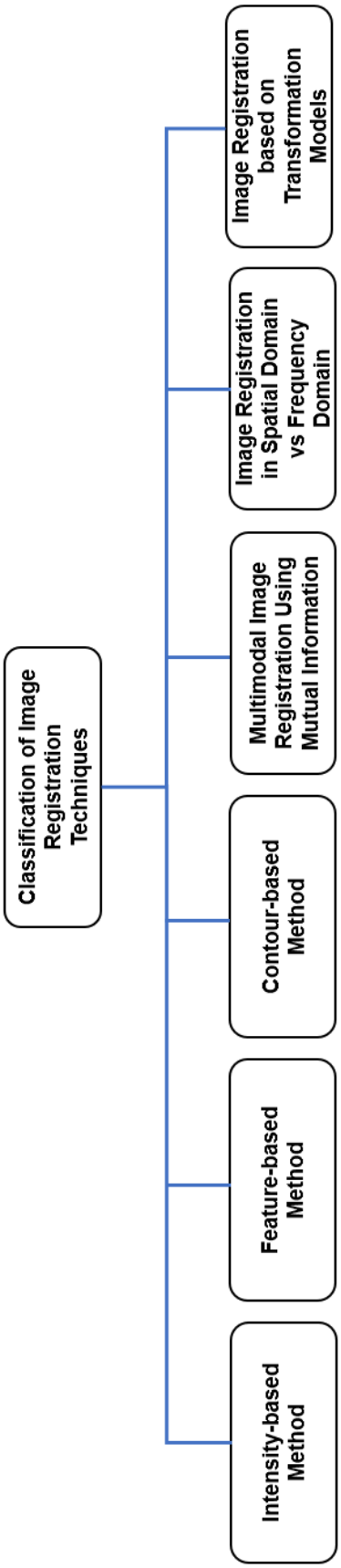


Figure 1.2: Classification of Registration Techniques

images to be aligned, and the parameters of transformation field are adjusted until the similarity measurement is maximized or minimized in terms of similarity measurement approaches.

Advantages When significant features are lacking in images and distinguishing information is given by grey levels/colors rather than local forms and structure, intensity-based methods are applied.

Disadvantages The flatness of the similarity measure maxima (due to the self-similarity of the images) and high computational complexity are the limitations of the intensity-based methods [18].

2. Feature-based Methods In feature-based approach, image features such as points, lines and contours are detected for both initial image and input images by using various feature detection approaches. After detecting the image features in each image, the correspondence between the detected features of the initial image and those of the current input image is detected. By means of establishing the correspondence between the initial image and the input images, the geometrical transformation can be determined to map the initial image to the input images.

Advantages The feature-based methods have low computational complexity, briefness and simplicity and may be robust to illumination changes.They can perform multimodal registration.

Disadvantages Depending on the nature of the images, there may be only a few feature correspondences that make it difficult to track the transformation, especially for poorly textured images.The mismatching features (outliers) may also grow quickly when the images contain repetitive and undistinguishable patterns. The common drawback of the feature-based methods is that the respective features might be hard to detect and/or unstable in time.

3. Contour-based Method Image feature point's high statistical features are matched by using this method. To extract the regions of interest in an image, colour image segmentation is used. To determine the contour of an image, the mean for the collection of colours are calculated. Each RGB pixel in an image is classified as having a colour in a specific range. Then, the Euclidean distance is used to determine similarity. The set of points is the sphere radius and point that exists within the range or on the surface of the sphere meets the specified colour requirement. A binary, segmented image is produced by coding these two set of points in the image with black and white. To eliminate noise, a gaussian filer is applied and the image is blurred with thresholds, and then obtain the contour of an image [18].

Advantages In this method, the gray values are not applied for matching process and overcomes the limitation of the intensity-based methods. The method filters out the redundant information.

Disadvantages Although the accuracy of the contour-based method is more than intensity-based methods, their processing is manual and slow.

4. Multimodal Image Registration Using Mutual Information Multimodal image registration can be defined as the process of the integrating information form mulitple sources. Multimodal image registration is the difficult task, especially in the medical imaging. The mutual information methods are a kind of the intensity-based methods and is a measure of statistical dependency between two data sets, and it is suitable to register the images from different modalities. The entropy of the image does not change even though the histograms changes.

Advantages This method is well suitable for routine clinical use in the variety of application because the maximization of mutual information of corresponding intensities allows for fully automated registration of multimodal images without need for segmentation or user intervention [16].

Disadvantages It can fail the registration process if the image has poor resolution or contains less information.

5. Image Registration in Spatial Domain vs Frequency Domain Most of the registration methods perform in the spatial domain, matching texture patterns or features as matching criteria. Some of the feature matching algorithms are outgrowths of traditional techniques for performing manual image registration, in which the corresponding control points (CP) in images are chosen. When the number of control points exceeds the minimum requirements to define the appropriate transformation model, iterative algorithms like RANSAC can be used to robustly estimate the parameters of a particular transformation type for registration of the images.

Frequency-domain methods find the transformation parameters for registration of the images while working in the transform domain [31].If an acceleration of the computation need to speed up or the images are corrupted by frequency-dependent noise, then these methods are preferred rather than other methods.

Advantages The frequency based methods are more accurate than correlation methods.

Disadvantages In spatial based methods, the performance of the registration can fail when there is poorly image information. Some type of interpolation must be used in the frequency domain.

6. Image Registration based on Transformation Models Image registration process has been classified into rigid registration and non-rigid (deformable) registration in terms of the transformation model of an object. The rigid registration allows the mapping between the objects that need to be uniformly rotated and translated, but it cannot change the size or shape of the objects.

By contrast, the non-rigid registration allows non-uniform transformation and can map the correspondences between the objects that change their size and shape. Since human organs are non-rigid, changing their size and shape with respiration corresponding, the accuracy of deformable registration plays a crucial role in medical AR. Although many approaches have been to address the deformable registration, most of them performed on CT/MRI images and they did not evaluate on sparsely textured surfaces.

1.2 Literature Review

Many approaches has been proposed to address the deformable registration in various perspectives. According to the registration process, these approaches are roughly categorized into intensity-based methods and feature-based methods.

Intensity-based Approaches Qiu et al. proposed the deformable registration algorithm based on the thin-plate spline algorithm to deal with the whole image [20]. In [28], the image deformation on Free-Form Deformations (FFD) based Nonuniform Rational B Spline (NURBS) in hierarchical optimization was modeled to speed up the registration and avoid local minima. [24] presented an improved adaptive bases non-rigid registration algorithm to improve the local optimization. All of them performed on CT/MRI images.

Feature-based Approaches Zhang et al. [29] proposed the non-rigid registration of lung CT images based on hybrid of Harris and SIFT feature detector. In their approach, matched feature points using their hybrid approach were less than SIFT. [5,30] proposed non-rigid registration methods using SURF feature detector. Although Haouchine et al. [5] presented good surface registration results, their approach may fail due to the outlier in tracking process. In [30], the features were mismatched due to SURF algorithm and their registration process performed on CT/MR images. Kim et al. proposed a system for tracking and augmenting a deformable surgical site by detection of image features in laparoscopic surgery [10]. In their approach, they used SIFT feature detector. Their approach can fail due to image blurring and mismatched features. Kajihara et al. [9] presented a registra-

tion method using accelerated-KAZE (AKAZE) feature detector. Their approach can performed the registration using small control points, but it can fail when there is no sufficient number of feature points.

1.3 Objectives and Contribution of the Thesis

In this thesis, we propose two deformable registration methods to register the objects with sparsely textured surfaces and apply to human organs for surgery assistance.

We propose our proposed methods with the intensity-based approach and feature-based approach among the variety approaches to perform the registration. The reason why the intensity-based approach is chosen for the first proposed method is that the intensity-based approach should be used when the surface has lack of the significant information according to prior knowledge because we intend to perform the registration on sparsely textured surfaces such as human organs and reconstruct the deformable surfaces. Since some other approaches such as feature-based approaches rely on high number of correct matches, they can fail when reconstruct the sparsely textured surfaces or repetitively textured surfaces.

In the first method, we use depth and infrared (IR) images that captured with Time-of-Flight (ToF) camera to obtain the shape information and texture information since our proposed method intends to register the surfaces with sparse texture and little concavity or convexity. Since a non-flat surface is used as an initial image, our proposed method can be applied to human organs for surgery assistance. The first method performs the registration based on intensity-based approach because an intensity-based approach should be used to recover the 3D shape of the sparsely textured surface since the feature-based approaches rely on highly correct matched feature points.

We prove that the performance of our approach by comparing the methods using only texture or shape information with videos of textured paper sheets and endoscopic stereo video. The performance of registration in term of texture information is verified by calculating peak signal to noise ratio (PSNR). The performance of registration in term of shape information is evaluated by Mean Square Error (MSE). According to the results, our approach has good registration capability in term of both texture and shape of the deformable objects.

As we intend to assist the surgeon during the surgery in real-time, the computational time of the registration process is needed to reduce. Therefore, we

propose the second method based on the feature-based approach because feature-based approaches have briefness of their process and low computational time over intensity-based approaches. To perform the registration on sparsely textured surfaces such as human organs by using feature-based approach, the tracking the most distinctive feature points is an effective way.

In the second approach, we apply accelerated-KAZE (AKAZE) feature detector on the initial image and the input images for feature detection. The detected features of the initial image are divided into the blocks based on their coordinates and pick up the best feature in each block by using the response values of the detected features. The registration is performed by finding the correspondences between blockwise features of the initial image and all detected features of the input images.

We verify the performance of the second approach by using three endoscopic stereo videos, which have sparse texture. The registration performance in term of texture is evaluated by PSNR. According to the results, the blockwise approach can easily perform the registration process even on sparsely textured surfaces. Moreover, we also demonstrate for surgery assistance based on this approach by superimposing a tumor image onto the sparse textured surfaces.

1.4 Organization of the Thesis

The background,problem formulation, objectives and contribution of this thesis are describe in this section. Chapter 2 describes the detail about 3D registration of deformable objects using a ToF camera. Chapter 3 presents the detail of blockwise feature-based registration for deformable medical images and Chapter 4 describes the comparison of two proposed methods. Finally, the conclusion and future work of the thesis are presented in chapter 5.

Chapter 2

3D Registration of Deformable Objects Using a Time-of-Flight Camera

2.1 Background

3D deformable object registration is an active research topic in computer vision and its application fields. The deformable registration is to find the geometrical transformation to align an image with a reference image. If this technology could be applied to living human organs, it would be possible to assist the surgery by overlaying the vessels and tumors, which were measured in advance, onto the organs in real time.

In the medical field, non-rigid registration is mostly applied on CT or MRI images [15, 28] and it can be used for the preoperative treatment planning. On the other hand, in the robotic research field, deformable registration is often performed on 3D point clouds for complex-shaped objects [7].

There are some studies to recover the 3D shape of the deformable objects from monocular images using deformable registration. Some of them use prior deformation model to recover the shape of deformable surfaces [4, 21]. In [21], they constructed the textured 3D model of the object to compute correspondences between 3D surface locations and 2D image features. Salzmann et al. [22] presented the closed-form solution to recover the shape of non-rigid inelastic surface without any initial shape estimation. In this work, they generated the synthetic textured image correspond to the reference image by using the recovered shape.

It becomes difficult to perform registration when the deformable surface has sparse texture since sparse texture on the surface does not provide useful infor-

mation for registration. Ngo et al. proposed a method for monocular 3D reconstruction of sparsely textured and occluded surfaces [17]. They introduced gradient-based pixel descriptors for robust template matching and the isometric deformation constraints enforcing that the surface should not stretch or shrink. Although they used a Kinect camera for creating the dataset, the depth information was used only for generating the ground truth surfaces and used only image information for 3D reconstruction. Therefore, their method has a limitation that template images need to be obtained from a flat surface.

Our approach is based on [17] but we use depth and IR images captured with a ToF camera to obtain shape and texture information. The main difference is that our proposed method can use a non-flat surface as the reference surface, which allows our proposed method to be applied to human organs for surgery assistance. Our proposed method can register objects with sparse texture and with little concavity or convexity. We demonstrate the effectiveness of our approach using videos of textured paper and an endoscopic stereo video by comparing with the methods using only texture or shape information.

2.2 Related Work

3D deformable object registration is a challenging task in computer vision. Many approaches have been proposed for deformable registration. Existing approaches can be roughly categorized into shape-based registration and texture-based registration.

Shape-based Registration Many studies have been conducted on registration of deformable objects using shape information. Salzmann et al. [21] introduced a deformation model for deformable 3D surfaces. They used a small subset of the angles between facets to parameterize the shape of triangulated mesh and created a representative sample of possible shapes. To produce the low-dimensional models, they performed dimension reduction by using principal component analysis (PCA). To recover the shape of deformable surfaces, their approach needs to create the samples. They assume all shapes are equally probable and so similarly influence the model deriving from these samples.

Dyke et al. [3] proposes two distinct changes to a typical non-rigid Iterative Closet Point (N-ICP) registration pipeline for large-scale and non-isometric deformations. Firstly, they describe a method using the principal scaling factor to estimate anisotropic deformations on a discrete mesh and incorporate it into N-ICP

pipeline. Secondly, they introduce correspondence generation in non-isometrically deformation regions and inconsistent correspondence pruning method based on local geodesics. To effectively handle large deformations, they incorporate r-ring as-rigid-as-possible (ARAP) formulation into their N-ICP registration pipeline. However, if the initial correspondences are largely wrong, the performance of their method is poor. Therefore, their method relies on the initial correspondences.

To track deformation from image to image, some existing approaches need to estimate the initial shape. Salzmann et al. [22] proposed a close-form solution to detect and reconstruct the 3D shape for a non-rigid inelastic surface from 3D to 2D correspondences. Their approach does not need the initial shape estimation. Haung et al. [7] presents pairwise non-rigid registration algorithm for partially overlapped point cloud surface. They define non-rigid registration as an optimization problem. Therefore, deformation optimization is solved by alternating correspondences computation. In their approach, geodesic distance between a set of correspondences is preserved to be stable correspondences. Due to the topology changing, geodesic consistency preservation is invalid. Therefore, their method will fail in this condition.

In medical field, tracking organs deformation due to our respiration system is a challenging task. Lu et al. [14] proposed a non-rigid registration method based on linear elastic model. In their approach, the image region of interest was divided in triangular grid. Extracted image feature point are used to form the irregular triangular grid. For similarity measure, the minimum potential energy was used to achieve their registration. They validated the robustness of their method on 2D CT heart image time series dataset. Due to the large number of triangular and small shape of the triangular, their method still needs many iterations to converge. Kajihara et al. [9] proposed a feature-based non-rigid registration method that establishes the transformation field and estimates the rigid transform in local region and then blends them to interpolate the transform at every pixel. Since their approach is feature-based, their method may not work well if sufficient number of feature points are not extracted.

Texture-based Registration Some studies uses texture information for registration of deformable objects. Sidorov et al. [25] proposed groupwise non-rigid registration for textured surfaces such as human faces that were obtained using a 3D scanner. Their method finds correspondences between meshes and build high quality 3D appearance models.

Savran et al. [23] presented an automated non-linear elasticity surface regis-

tration method by using both attraction forces originating from geometrical and textural similarities. To avoid the loss of information during mapping between 2D and 3D, they use mesh parameterization approach instead of using projection. Although their approach handles the large deformation, if human surface topology changes such as open mouth expression, their approach cannot solve such kind of changes.

In texture-based registration, less texture and occlusions on surface are more challenging problems. Ngo et al. [17] addressed these problems by proposing a template-matching approach. Moreover, they added the additional constraint for 3D shape of the surface that does not stretch or shrink. In their work, they used gradient-based pixel descriptors for robust template matching and compute the relevancy score for each pixel to handle surface occlusions. Their approach can track well-textured and sparsely textured deforming surfaces in monocular video dataset with presence of occlusion or without occlusion. However, their method has a limitation that template images need to be obtained from a flat surface.

2.3 Proposed Framework

We propose a framework that uses a ToF camera for 3D registration of deformable objects. We use IR images for texture registration and depth images for shape registration.

2.3.1 Time-of-Flight (ToF) Camera

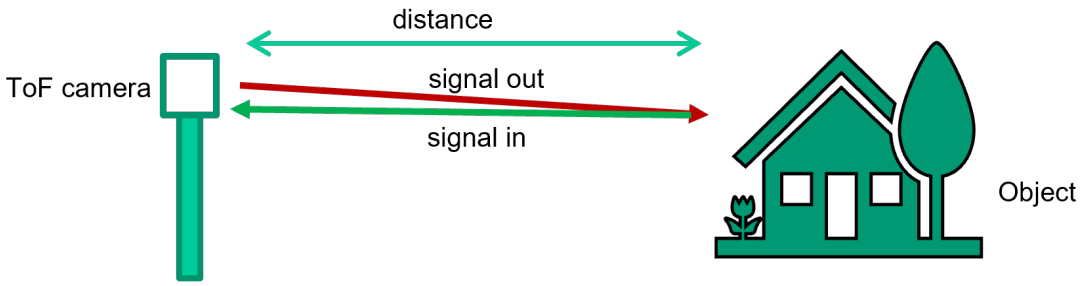


Figure 2.1: Time-of-Flight Principle

Time-of-Flight (ToF) camera is used to produce the range images that are also referred to as depth images, depth maps, xyz maps, surface profiles and 2.5D

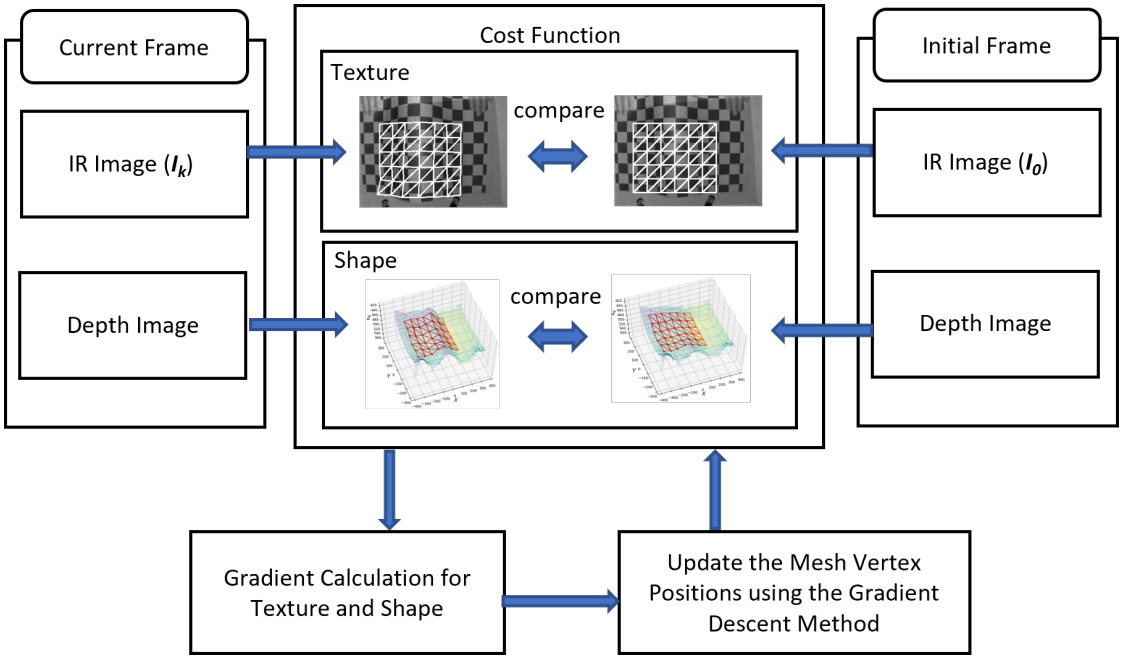


Figure 2.2: Overview of our proposed method. The cost function is defined using texture and shape differences between the initial frame and the input frame. The gradient of the cost function is calculated to minimize the function with respect to the 2D positions of the mesh vertices.

images. It emits the infrared signal that cannot be seen by human eyes to determine the depth information.

ToF is a method that measure the distance between an object and the sensor using the time that it takes that the sensor emits the light signal to the object and bounce it back to the sensor. Then, it produces the depth information.

ToF camera can be used to measure distance and volume, as well as for object scanning, indoor navigation, obstacle avoidance, gesture recognition, and object tracking. It can measure both intensity and distance for each pixel of the object simultaneously. It can also help with 3D imaging and improving augmented reality experiences.

2.3.2 Methodology

Assume that the IR image in the initial frame $I_0(\mathbf{x})$ and the IR image of the current frame $I_k(\mathbf{x})$ are given, where $\mathbf{x} = (x, y)$ is the pixel coordinates. We set the 2D vertex positions of the triangular meshes for the initial frame, $\mathbf{v}_{i,j}^0$, on a regular grid. We perform registration by finding the corresponding 2D vertex positions, $\mathbf{v}_{i,j}$.

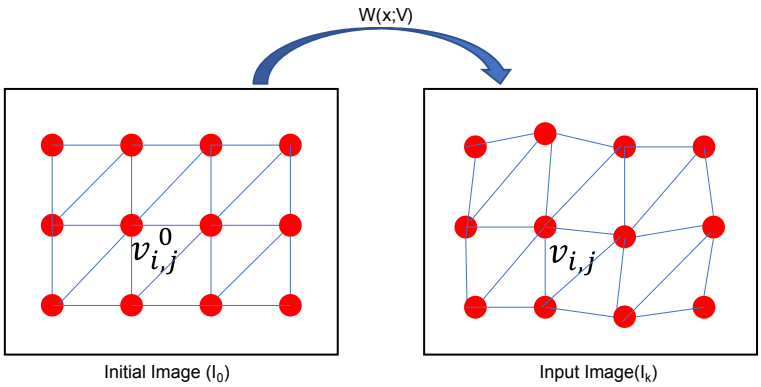


Figure 2.3: The warping function $\mathbf{W}(\mathbf{x}; V)$ is determined from the 2D positions of the mesh vertices $\mathbf{v}_{i,j}$. Using this function, a pixel in the initial frame image is mapped to a pixel in the current frame image.

For texture registration, the texture patterns in each mesh between the initial IR image and the current IR image are compared. The warping function $\mathbf{W}(\mathbf{x}; V)$ is determined from the 2D positions of the mesh vertices $\mathbf{V} = \{\mathbf{v}_{0,0}, \mathbf{v}_{1,0}, \dots\}$. Using the warping function, a pixel in the initial frame image is mapped to a pixel in the current frame image as illustrated in Fig. 2.3. We define the cost for texture registration:

$$E_{img} = \sum_{x} \{I(\mathbf{W}(\mathbf{x}; V)) - I_0(\mathbf{x})\}^2 \tag{2.1}$$

For shape registration, we assume that the deformed surface should not stretch or shrink as in [16]. Therefore, the lengths in 3D space between two adjacent vertices before and after deformation are compared. The 3D positions of the vertices are obtained using the depth images. We define the cost for shape registration:

$$\begin{aligned} E_{len} = \sum_{i,j} \{ &(||\mathbf{X}_{i,j} - \mathbf{X}_{i+1,j}|| - ||\mathbf{X}^0_{i,j} - \mathbf{X}^0_{i+1,j}||)^2 \\ &+ (||\mathbf{X}_{i,j} - \mathbf{X}_{i,j+1}|| - ||\mathbf{X}^0_{i,j} - \mathbf{X}^0_{i,j+1}||)^2 \} \end{aligned} \tag{2.2}$$

where $\mathbf{X}^0_{i,j}$ are the 3D positions of the vertices in the initial frame, $\mathbf{v}^0_{i,j}$, and $\mathbf{X}_{i,j}$ are the 3D positions of the vertices in the current frame, $\mathbf{v}_{i,j}$. By combining these costs, we obtain the total cost function considering both texture and shape:

$$E = \lambda_1 E_{img} + \lambda_2 E_{len} \tag{2.3}$$

where λ_1 and λ_2 are the weights for texture and shape costs, respectively. To minimize the cost function for texture and shape, we use the gradient descent

method with respect to $\mathbf{V}$ [2]. The gradient of the texture cost is calculated as

$$\frac{\partial E_{img}}{\partial v_{i,j}} = 2\sum_{x}\{\nabla I\frac{\partial W}{\partial p}\}^{T}\{I(\mathbf{W}(\mathbf{x};V)) - I_0(\mathbf{x})\} \tag{2.4}$$

The gradient of the shape cost is calculated as

$$\begin{aligned}\frac{\partial E_{len}}{\partial v_{i,j}} = 2\sum_{i,j}\{(||\mathbf{X}_{i,j} - X_{i+1,j}|| - ||\mathbf{X}^0_{i,j} - \mathbf{X}^0_{i+1,j}||)\frac{\partial}{\partial p}||\mathbf{X}_{i,j} - \mathbf{X}_{i+1,j}|| \\ + (||\mathbf{X}_{i,j} - \mathbf{X}_{i,j+1}|| - ||\mathbf{X}^0_{i,j} - \mathbf{X}^0_{i,j+1}||)\frac{\partial}{\partial p}||\mathbf{X}_{i,j} - \mathbf{X}_{i,j+1}||\}\end{aligned} \tag{2.5}$$

The 2D vertex positions of the meshes are updated according to the gradient calculation as below:

$$\mathbf{v}_{i,j} = \mathbf{v}_{i,j} - w\left(\lambda_1\frac{\partial E_{img}}{\partial v_{i,j}} + \lambda_2\frac{\partial E_{len}}{\partial v_{i,j}}\right) \tag{2.6}$$

where w is a weight for the gradient descent method.

This is iteratively performed until the cost function is approximately converged. The overview of our proposed method is shown in Fig. 2.2.

2.4 Experimental Results

We implemented our proposed method and conducted experiments using two types of datasets. Our proposed method was compared with the methods using only texture and only shape information.

2.4.1 Densely and Sparsely Textured Paper Sheets

The videos of densely and sparsely textured paper sheets were used in the experiment. These videos were captured by DepthSense 325 ToF camera. These paper sheets were pushed and pulled from both sides by hand to be deformed. The camera was about 0.3 m away from the sheets. The number of frames were 32 for each video. We made shading correction to compensate unequal lighting on the sheets. The resolution of both IR images and depth images was 320×240 pixels.

We set the number of mesh vertices to 7×6. We used the following parameter values: λ_1= 1, λ_2= 10^5 and w= 10^{-7}. We set λ_1 or λ_2 to be zero for the methods using only shape or texture information.

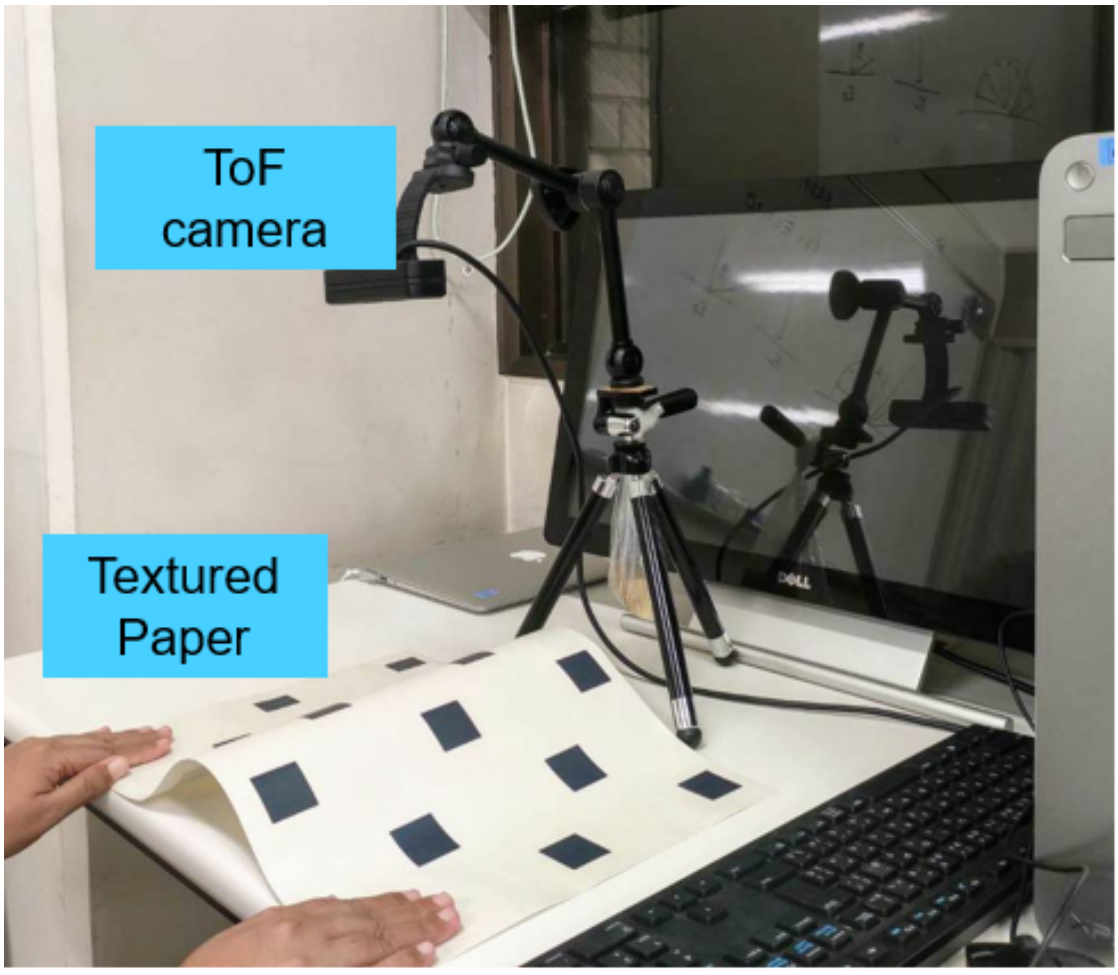


Figure 2.4: Apparatus of Our approach

To evaluate the quality of registration, the input images were warped back into those in the coordinate system of the initial frame image, which we call restored images. The registration and restoration results for densely and sparsely textured paper sheets are shown in Fig. 2.5 and Fig. 2.6, respectively. For both densely and sparsely textured paper sheets, we can see the texture in the restored images is similar to that of the initial frame image. Although the registration using texture only was able to register the texture well, the meshes were distorted a little because the shape of the meshes is not considered in the texture-based registration. The 3D shapes of the densely and sparsely textured paper sheets and meshes are shown in Fig. 2.7 and Fig. 2.8, respectively. In the registration using shape, we can see that the shape of the meshes were more stationary than registration using shape only.

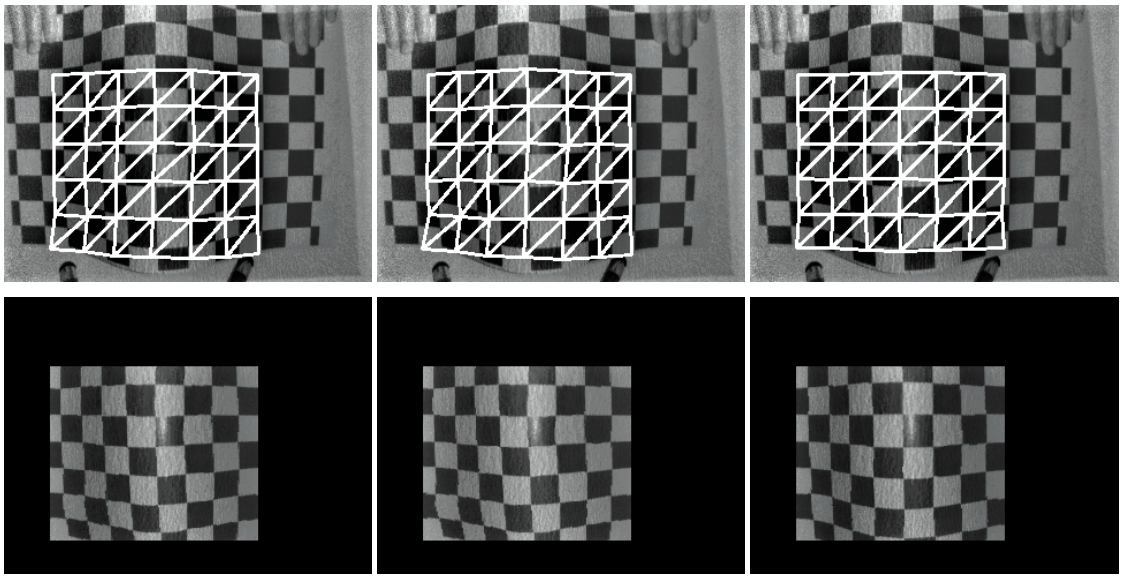
To evaluate the registration performance in terms of texture, we calculated peak signal to noise ratio (PSNR) between the restored images and the initial frame image. To evaluate the registration performance in terms of shape, we calculated Mean Square Error (MSE) between the inter-vertex lengths of the initial mesh and those of the registered mesh. The PSNR results for densely and sparsely textured paper sheets are shown in Fig. 2.9.The MSE results for densely and sparsely textured paper sheets are shown in Fig. 2.10.The results showed that the proposed method was as good as the texture-only method and was better than the shape-only method. MSE results showed that the proposed method was better than the texture-only method but was worse than the shape-only method. These results showed that our proposed method has good registration capability in terms of both texture and shape.

2.4.2 Endoscopic Stereo Videos

We also applied our proposed method to publicly available endoscopic stereo videos that contained deforming heart-1, heart-2 and liver from the Hamlyn Center Laparoscopic / Endoscopic Dataset [13]. To reconstruct the 3D shapes from the stereo video, we estimated the disparity maps by using the semi-global block matching (SGBM) algorithm. The estimated disparity maps are converted into depth images by using the following equation:

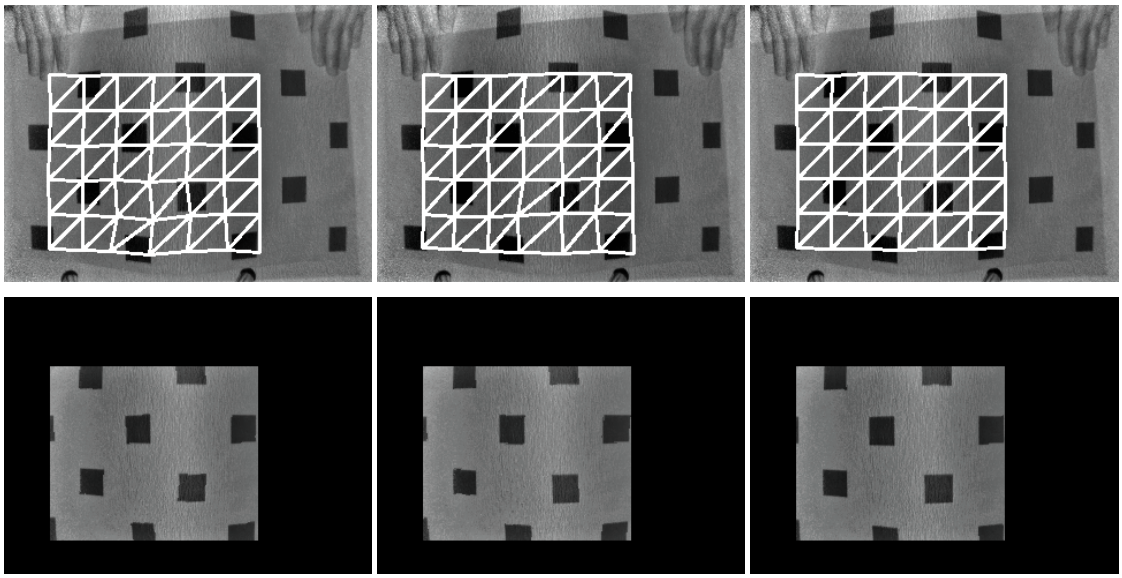
$$z = f\frac{b}{d} \tag{2.7}$$

where z is the depth, f is the focal length, b is the baseline, d is the disparity.



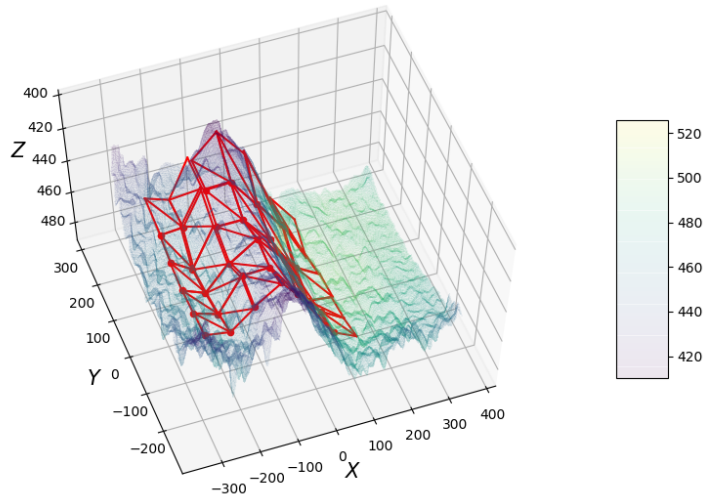
(a) (b) (c)

Figure 2.5: The registration and restoration results for densely textured paper sheet: (a) registration using both texture and shape, (b) registration using texture only, and (c) registration using shape only.

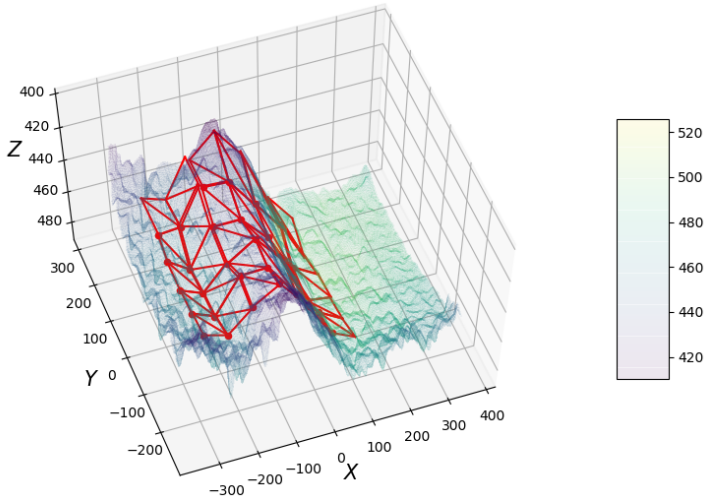


(a) (b) (c)

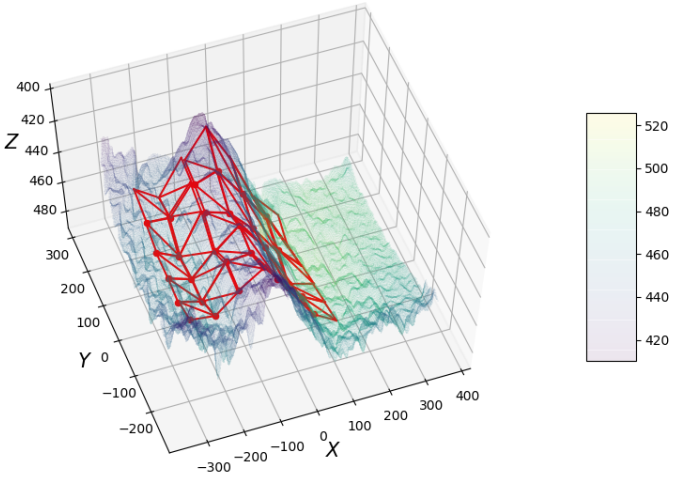
Figure 2.6: The registration and restoration results for sparsely textured paper sheet: (a) registration using both texture and shape, (b) registration using texture only, and (c) registration using shape only.



(a)

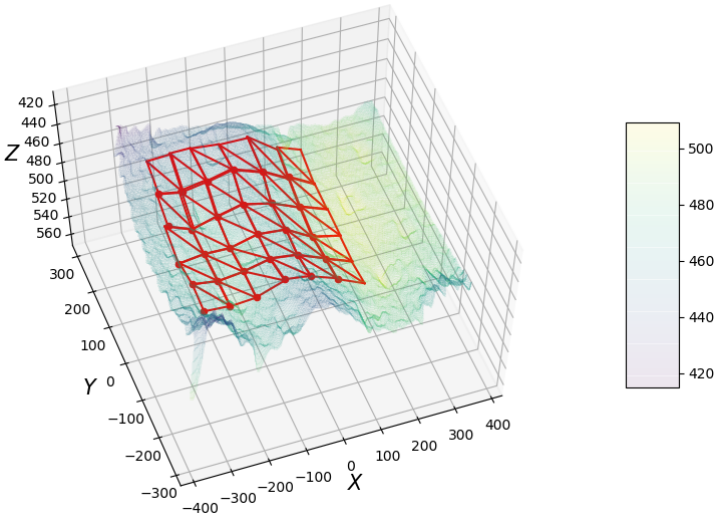


(b)

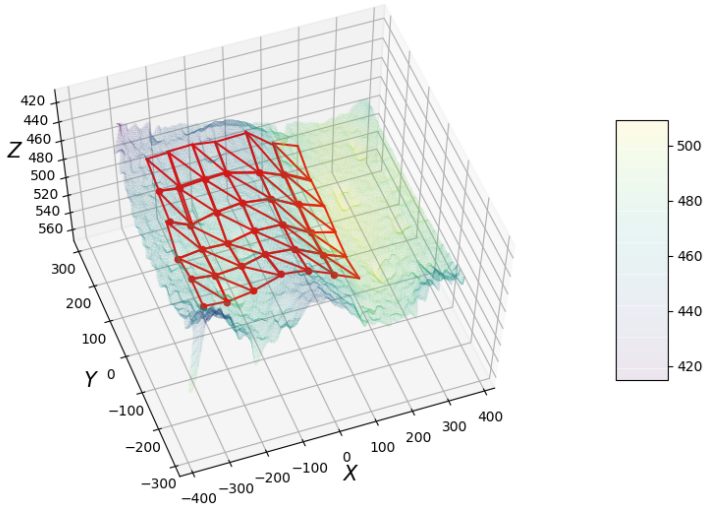


(c)

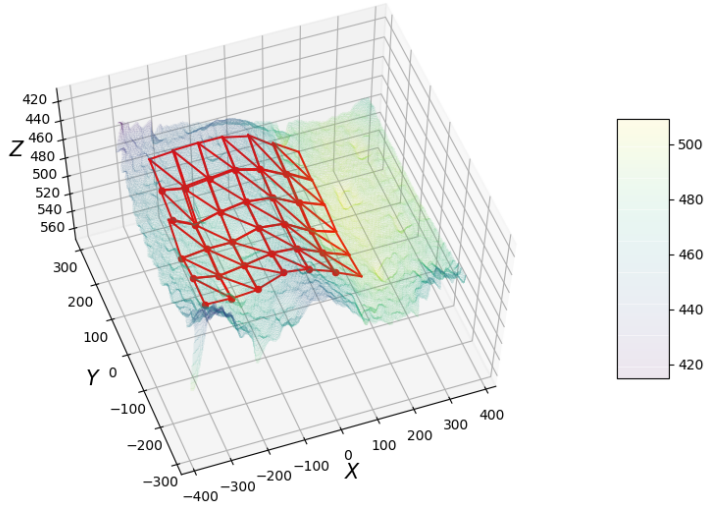
Figure 2.7: The registration of the deformed surface of the densely textured paper sheet: (a) registration using both texture and shape, (b) registration using texture only and (c) registration using shape only.



(a)

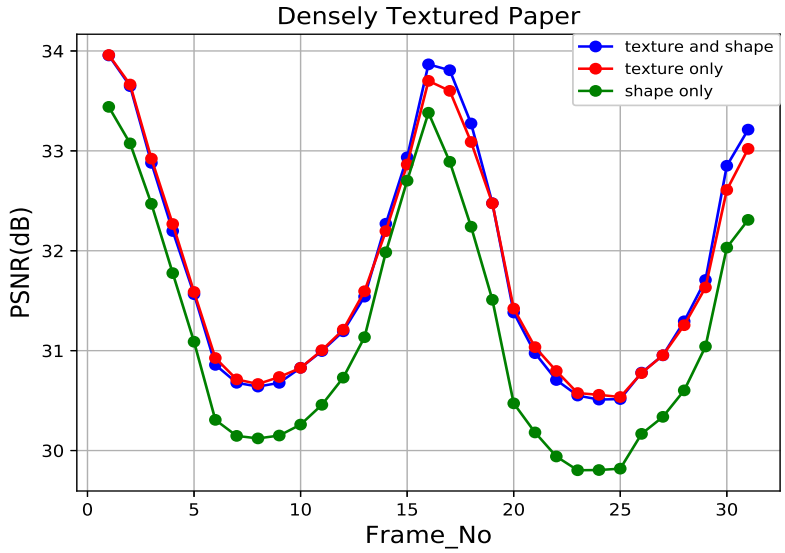


(b)

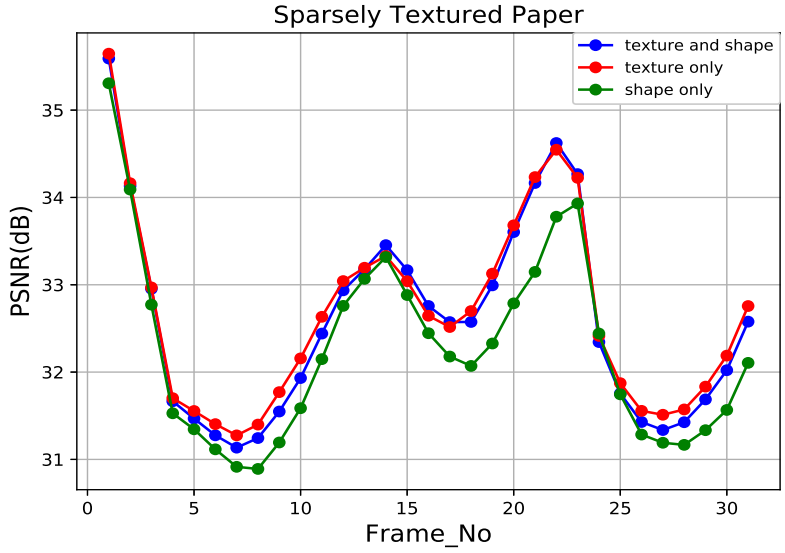


(c)

Figure 2.8: The registration of the deformed surface of the sparsely textured paper sheet: (a) registration using both texture and shape, (b) registration using texture only, and (c) registration using shape only.

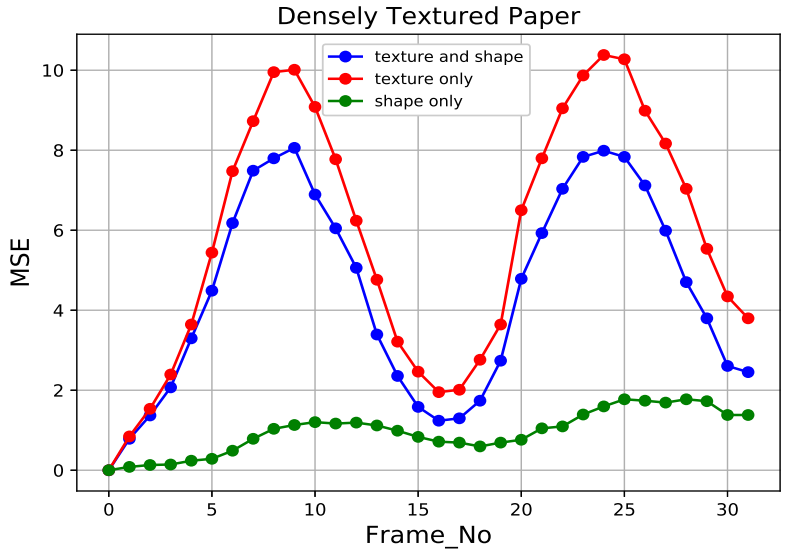


(a)

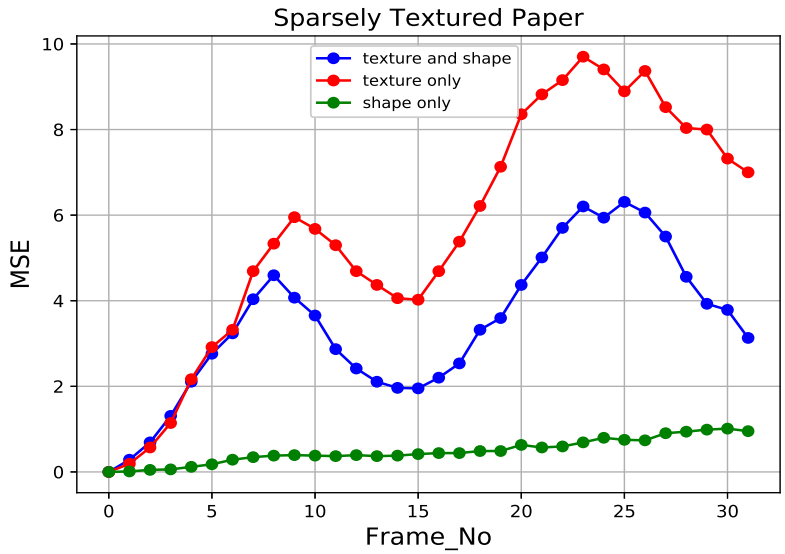


(b)

Figure 2.9: PSNR results for densely and sparsely textured papers per frame number: (a) PSNR results for densely textured paper, (b) PSNR results for sparsely textured paper.

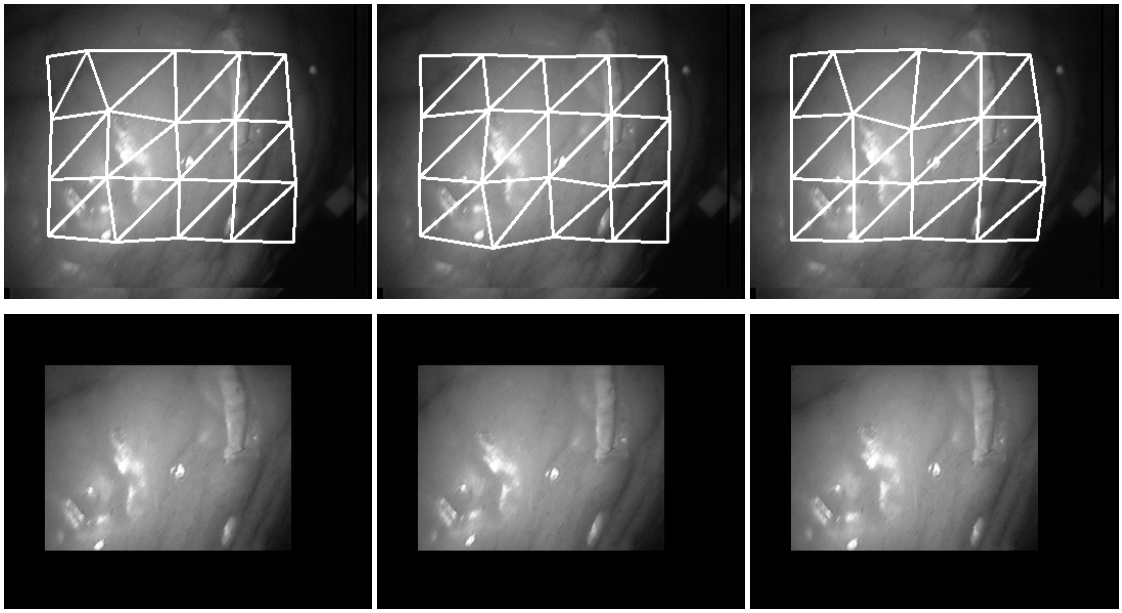


(a)



(b)

Figure 2.10: MSE results for densely and sparsely textured papers per frame number: (a) MSE results for densely textured paper, (b) MSE results for sparsely textured paper.

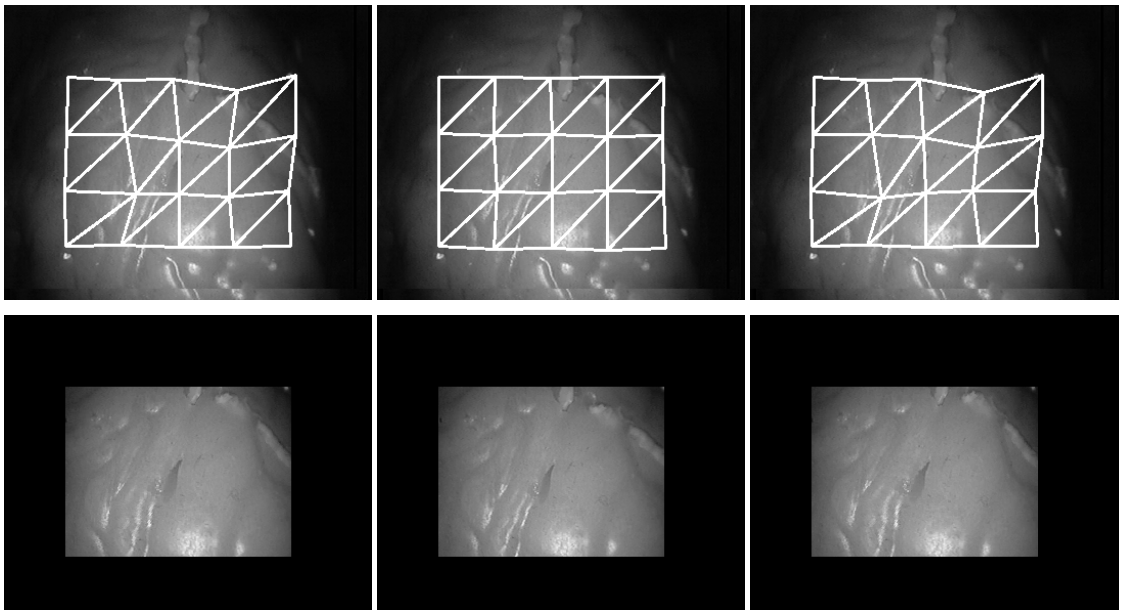


(a) (b) (c)

Figure 2.11: The registration and restoration results for the endoscopic stereo video: (a) registration using both texture and shape, (b) registration using texture only, and (c) registration using shape only.

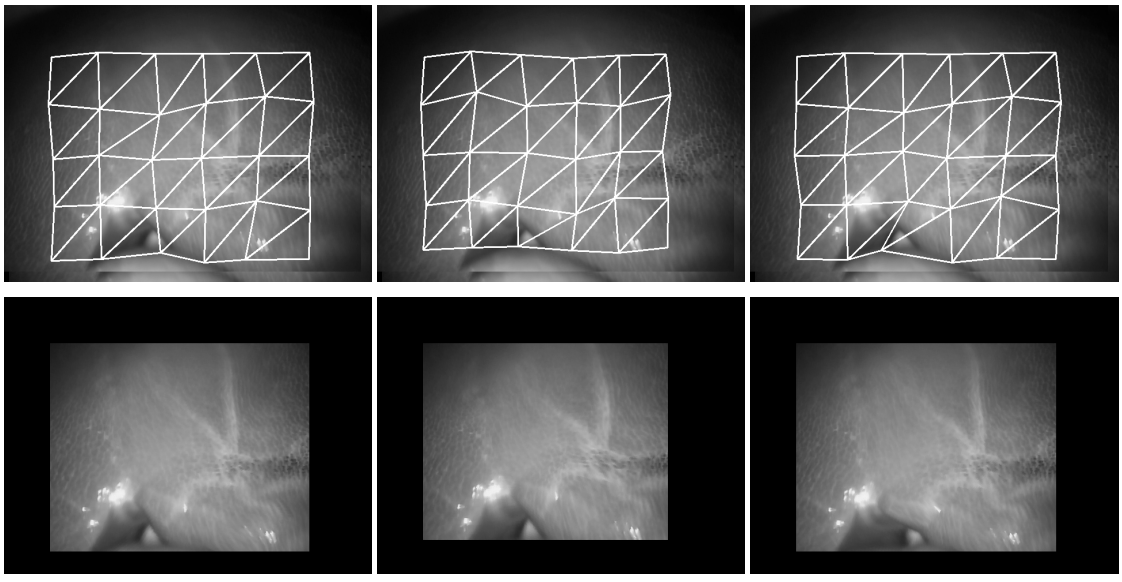
We used 194 frame images in the heart-1 video. The number of mesh vertices was 5×4. For this dataset, we used the parameters λ_1=4 and λ_2= 10^3 for both texture and shape information, λ_1= 1, λ_2= 0 for texture only information, and λ_1= 0, λ_2= 10^5 for shape only information. The qualitative results for stereo video dataset are shown in Fig. 2.11.

We used 100 frame images in the heart-2 video and 49 frame images in liver video. The number of mesh vertices was 5×4 for heart-2 and 6×5. For the heart-2 dataset, we used the parameters λ_1=3 and λ_2= 10^5 for both texture and shape information, λ_1= 2, λ_2= 0 for texture only information, and λ_1= 0, λ_2= 10^5 for shape only information. The qualitative results for stereo video dataset are shown in Fig. 2.12. For the liver dataset, we used the parameters λ_1=1 and λ_2= 10^5 for both texture and shape information, λ_1= 1, λ_2= 0 for texture only information, and λ_1= 0, λ_2= 10^5 for shape only information. The qualitative results for stereo video dataset are shown in Fig. 2.13. We can see the texture in the restored images were similar to that in the initial image for all datasets. Since all videos have less texture, they were more challenging to register.



(a) (b) (c)

Figure 2.12: The registration and restoration results for Heart-2: (a) registration using both texture and shape, (b) registration using texture only, and (c) registration using shape only.



(a) (b) (c)

Figure 2.13: The registration and restoration results for Liver: (a) registration using both texture and shape, (b) registration using texture only, and (c) registration using shape only.

2.5 Discussion

We proved that the effectiveness of our approach by comparing the methods using texture or shape information.

For the method using texture information, only the comparison between the texture patterns in the mesh between the initial image and the current input image was considered. Therefore, the PSNR result of the method using texture information was better than shape only method.

For the method using the shape information, we compared only the lengths in 3D space between two adjacent vertices before and after deformation. The texture patterns in the mesh were not considered. Therefore, the shape of the meshes before and after deformation was more stationary than the other two methods in terms of the results of MSE.

For our approach, we consider both conditions. Therefore, PSNR result was as good as the texture-only method and was better than the shape-only method. According to MSE results, our approach was not as good as the shape-only method, but it was better than the texture-only method.

2.6 Summary

We proposed a framework for 3D deformable objects registration using both texture and shape information obtained using a ToF camera. We demonstrated the effectiveness of our proposed method by using videos of textured paper sheets captured with a ToF camera and an endoscopic stereo video. The experimental results showed that our proposed method has good registration capability in terms of both texture and shape.

Our approach does not need to estimate the initial shape for the shape reconstruction of the objects. Since the non-flat surface was used as the reference surface, our proposed method can be applied to human organs for surgery assistance. Our approach can register both texture and shape of the deformable surfaces simultaneously. Moreover, it can reconstruct the 3D shape of the deformable surfaces. There is a limitation that the performance of our approach depends on the initial parameters.

Chapter 3

Blockwise Feature-based Registration of Deformable Medical Images

3.1 Background

Registration methods can be divided into two categories: intensity-based methods and feature-based methods. Since feature-based methods are simpler and have lower computational complexity than intensity-based methods, feature-based methods are widely used in medical image registration. On the other hand, the performance of registration can fail when a sufficient number of feature points are not detected [9] or when the number of matched features is small [29, 30].

Some feature-based registration approaches have been proposed to reduce the mismatched features, but they do not evaluate their approaches on the sparse texture surfaces [6,11]. Some feature matching approaches have been presented [19, 26] to improve the existing methods by finding a large number of correct matched features on deformable surfaces at an increased speed and accuracy. In [19], their method can recover and track the features that were lost due to occlusion and eliminate the mismatched features, but their approach has a limitation on poorly textured surfaces and there is a computational limitation in [26], which also cannot retrieve a sufficient set of accurate matches.

In this thesis, we propose a blockwise approach for feature-based registration of deformable medical images. We apply the accelerated-KAZE (AKAZE) feature detector on both of initial image frame and input image frames. In the initial image frame, the detected feature points are divided into blocks based on their coordinates and fetch the best feature of each block in terms of the response values

of the detected features. Tracking of feature points is performed by finding the correspondences between the blockwise features of the initial image and all the detected features of the input image. Our approach can register the sparsely textured surfaces including human organs. We show the effectiveness of our approach using three stereo endoscopic videos.

3.2 Related Work

Feature-based Registration on CT/MR and Histological Serial Section Images Kajihara et al. [9] described a feature-based non-rigid registration method with a small number of control points for histological serial section images. In their approach, feature detection is performed by accelerated-KAZE (AKAZE), and brute-force matching using the Hamming distance is adopted for the feature matching process. The keypoints are clustered by using their coordinates to determine the local region. Rigid transformation in each cluster is estimated using RANSAC. Then, rigid transformations are blended to interpolate the transformations at each pixel. Although their method can represent the complex deformation with a small number of control points, it could not perform registration in the image without a sufficient number of feature points. Zhang et al. [29] proposed a hybrid feature detection method for non-rigid registration of lung CT images based on tissue features. The vessel crossing points, vascular endpoints, and tissue boundary points which have high gradients were enough to track the motion of the lung and can be detected by Harris. In this approach, they combined Harris and SIFT to detect blob features since they also used those feature points. Although detected features points by using their hybrid method were more than those by SIFT, matched feature point pairs by using their method were less than those by SIFT.

Lu et al. [14] improved a linear elastic model for non-rigid medical image registration using the elasticity of the minimum energy as a similarity measure, establishing partial differential equations to describe the image deformation, and using the finite element method to solve partial differential equations. Their approach is based on the global registration and extracted the feature points of global image registration to generate the irregular triangle grid for defining the region of interest. They showed the robustness of their approach using the 2D CT heart image time series dataset. Although their method can improve the accuracy of registration and enhance image robustness, their method still needs many iterations to

converge due to the small shape of the triangle and a large number of triangles.

Zheng et al. [30] proposed a coarse-to-fine registration method based on progressive images and SURF algorithm (PI-SURF). For generating multiple progressive images, the reference image and the floating image are fused. These two images are registered for coarse registration results based on the SURF algorithm and the coarse registration result and the reference image are registered to get the fine image registration. They demonstrated their approach using MR-MR and CT/MR images. In their approach, there are some limitations such as there is time-consuming when the intermediate progressive images are generated and there are mismatching features due to the SURF algorithm.

Feature-based Registration for Surgery Assistance Puerto et al. [19] presented a feature matching algorithm called hierarchical multi-affine (HMA) which finds similarities between laparoscopic views. In their method, the detected features are iteratively partitioned into clusters and estimated an affine transformation for each cluster to eliminate incorrect matches from the initial matches. Although the tracked features that were lost are recovered by using affine mapping due to a completed or continuous occlusion or fast camera motion, there are some limitations in their method when the organ or object has poorly textured or when the number of correct matches is very small.

Stoyanov et al. [27] proposed a framework for tissue deformation tracking using a monocular endoscope. Their method was performed by sparse salient features combined with geometric surface parameterization and applied to a phantom heart video sequence to track the motion of an observed fiducial. To compare with ground truth data, the estimated motion of video is compared with the coordinates of the fiducial trajectory of the CT image that was reprojected back into camera space. Their proposed method can estimate the motion of the fiducial, but it cannot handle that their mesh coordinates form with misalignment due to occurring feature matching errors and noise.

Kim et al. [10] proposed a framework for tracking and augmenting a deformable surgical site using shape from shading to recover the 3D shape of the surface and the shape was flattened by using conformal mapping. Feature detection was performed by using SIFT on that flatten shape and matched the features using Pizarro and Bartoli's feature matching algorithm.

For outlier removal, RANSAC was used. Then, pseudo-huber norm cost function was used for optimization. Their aim for augmentation is to determine and visualize the boundary regions in the current input frame by matching the features

between planer surfaces. Therefore, the boundary of the organ in the planer surface was estimated and the vertices of the boundary are mapped into 3D space by using barycentric interpolation. By using the camera's projection function, these 3D vertices were projected onto the current laparoscopic image for the boundary of the organ. To determine the surgical target, the inlier matches were transformed from planer surface to laparoscopic image, and affine-MLS was used for smoothly warping the target positions located on the reference image to the current image by feature correspondences. Then, the locations of the surgical target were marked and overlaid on the top of the original input frame. According to their approach, they can retrieve the surface deformation, but their tracking can fail sometimes due to image blurring and matched features were not found.

Haouchine et al. [5] presented a method for the real-time augmented reality of internal liver structures during surgery. Their approach can locate the in-depth position of the tumors based on partial three-dimensional liver tissue motion using a real-time biomechanical model. To recover the 3D information from the liver surface, they used a stereo endoscope. In their work, Speed-Up Robust Features (SURF) was used to detect salient landmarks in each image pair and tracked by using Lucas-Kanade optical flow since their registration method was a point-to-point registration method. Their method showed good results in terms of surface registration and internal tumors localization. However, their tracking process may fail due to the outliers.

3.3 Proposed Method

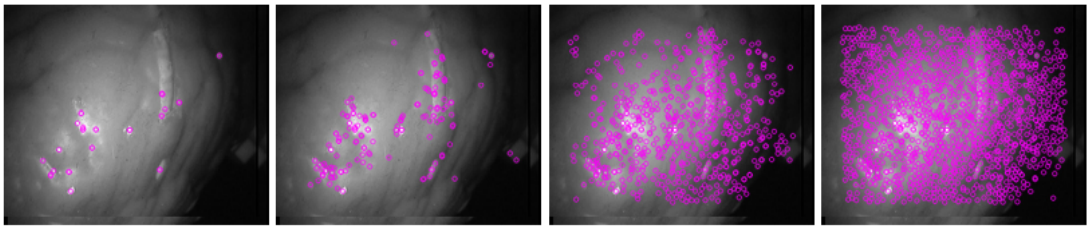
In our approach, we apply the AKAZE feature detector for detecting the feature points in the initial image frame and current input image frame. The best feature points among the detected features of the initial image are fetched by using the blockwise approach. Our deformable registration process is performed by tracking the corresponding points between the initial image and the current input image.

3.3.1 Feature Detection and Blockwise Features

Although there are many methods for feature detection and description, we adopt an accelerated-KAZE(AKAZE) feature detector. Feature detection and description in non-linear scale space are time-consuming due to the high computational load to build the non-linear scale spaces. In the AKAZE feature detector, Fast

Explicit Diffusion (FED) scheme embedded in a pyramidal approach can speed up non-linear scale-space construction and its Modified-Local Difference Binary(M-LDB) descriptor that is invariant in rotation and scale can utilize gradient and intensity information from non-linear scale spaces [1]. Therefore, AKAZE features have low computational and descriptor storage demand. The structure of the AKAZE feature contains 2D coordinate positions, a response that describes the strength of the feature, size, class-id, octave that describes the level of scale space, and angle. We use AKAZE in OpenCV with default parameters except for the threshold value. Threshold, one of AKAZE feature detector parameters, allows accepting the feature points. The less the threshold value we set, the more feature points we can get as shown in Fig. 3.1.

Assume that the initial image frame I_0 and the current image frame I_k are given. Firstly, we apply AKAZE on the initial image frame and divide the detected features into blocks $B = \{b_0, b_1, b_2, \dots b_{n-1}\}$ in terms of their coordinates. Then, the advantage of the response field of the AKAZE feature is taken to define the best feature among the detected features in each block of the initial image frame, which we call blockwise features as illustrated in Fig. 3.2. The higher value of the response field of a feature, the stronger feature is. We also apply AKAZE on the current image frame for feature detection and description.

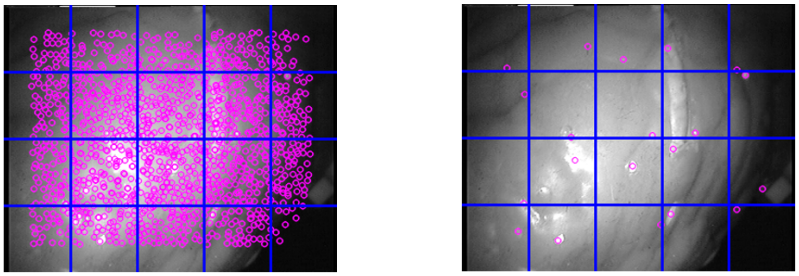


(a) 0.001 (b) 0.0001 (c) 0.00001 (d) 0.000001

Figure 3.1: Detected features with various threshold values.

3.3.2 Feature Matching and Registration

To track the detected feature points in each input image frame, the corresponding points are detected between the initial image frame and input image frames. Therefore, feature matching is performed by brute-force matching algorithm using Hamming distance to find the correspondence between blockwise features of the initial frame and the detected features of the current image frame.



(a) (b)

Figure 3.2: Blockwise Features Approach: (a) features in the initial image I_0 are detected by using AKAZE with threshold=0.000001. The detected features are divided into blocks based on their coordinates and find the best feature in each block using response value of the feature, (b) the best feature in each block of the initial image I_0.

To eliminate the outlier features from matched results, brute-force matching is combined with the k-Nearest Neighbors (KNN) algorithm and Lowe's ratio test [12]. Lowe's ratio test has simple criteria that determine a good match if the distance ratio of the first closest one and the second closest one is smaller than the threshold value (the default value is 0.5). For outlier removal, we set the threshold value=0.7 for our cases. The good match that satisfies the threshold value for each image pair contains a feature in a block of the initial image and a corresponding feature of the current frame.

The coordinates of blockwise features are set as 2D vertex positions of the triangular meshes for the initial image frame while defining the coordinates of the features that correspond to blockwise features as 2D vertex positions of the meshes for the current image frame. We perform the registration by finding features corresponding to blockwise features as shown in Fig. 3.3. If there is no corresponding feature, we use the corresponding feature of the previous frame image.

To superimpose an image onto the deformable surface in AR, each pixel of the superimposing image in terms of the coordinate system of the initial image frame is mapped onto the current deformed surface by using the bilinear interpolation method.

3.4 Experimental Results

We conducted an experiment using three endoscopic stereo videos that contained deforming heart and liver from the Hamlyn Centre Laparoscopic / Endoscopic

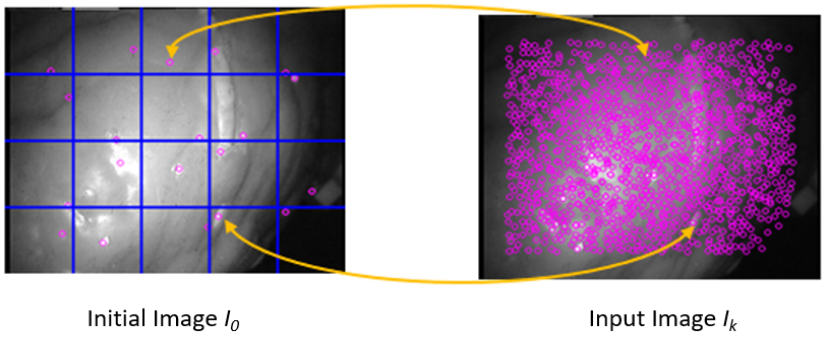


Figure 3.3: Finding the corresponding points in the current input image I_k in terms of blockwise features of the initial image I_0.

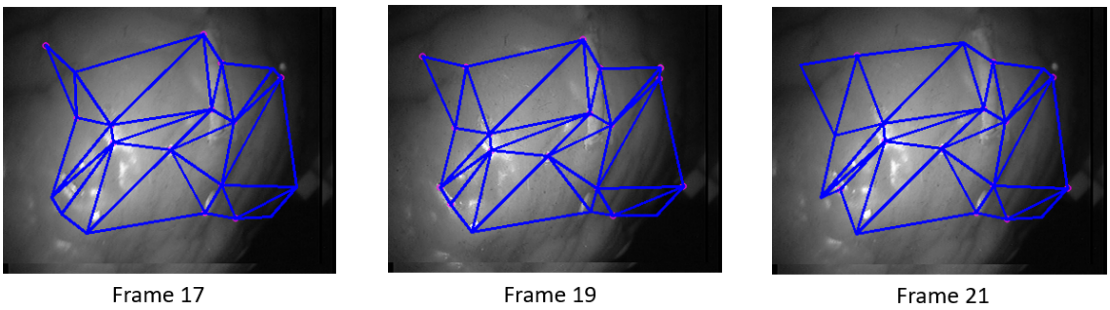
Dataset [13].

We used 194 frame images of Heart-1 video, 100 frames of Heart-2 video, and 49 frames of Liver video. For all videos, ratio test=0.7 was used to remove the outlier after matching the features from the initial frame image to the input frame images. For the evaluation of registration, the input images were localized onto those in the coordinates of the blockwise features of the initial frame, which we call restored images. The registration and restoration results of Heart-1, Heart-2, and Liver are shown in Fig. 3.4, Fig. 3.5, and Fig. 3.6, respectively. For all videos, we can see the texture in the restored images is similar to that of the initial image.

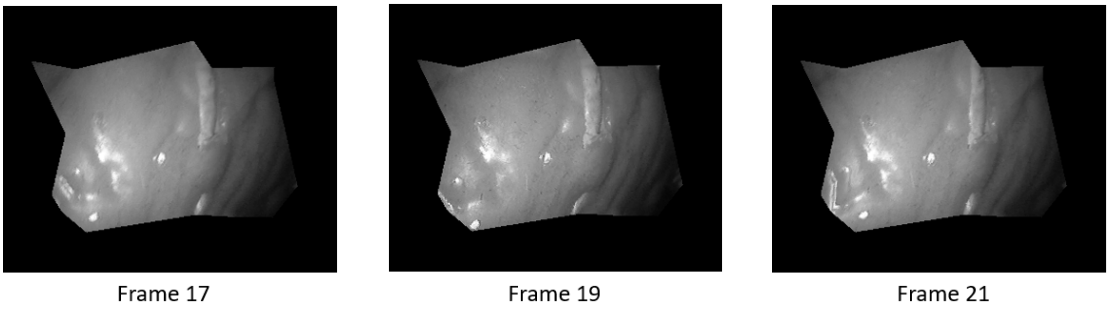
To evaluate the registration performance in term of texture, we calculated the peak signal to noise ratio (PSNR) between the restored images and the initial image. The PSNR results for all videos are shown in Fig. 3.7. PSNR results of all videos exist in the good performance range. The mean PSNR values of Heart-1, Heart-2, and Liver are 36.6, 36.8, and 36.9, respectively.

As a demonstration for surgery assistance, a hand-drawn tumor image is superimposed onto the deformable surface of each image frame. Fig. 3.8 shows the results of superimposing a tumor image onto an organ. Since our approach can perform on sparsely textured surfaces, we can successfully superimpose a tumor image onto the organs. We can see that the tumor moves along with the movement of the organs.

Our proposed method includes five processes. The processing time of each process for 194 image frames is described in Table 3.1. The registration process is from process 1 to process 3. The process 4, superimposing process, describes only for demonstration of surgery assistance. The process 5, restoration process, is performed for the evaluation of registration performance. Therefore, the processing time of our registration process for 194 frames takes only 4.918 seconds.

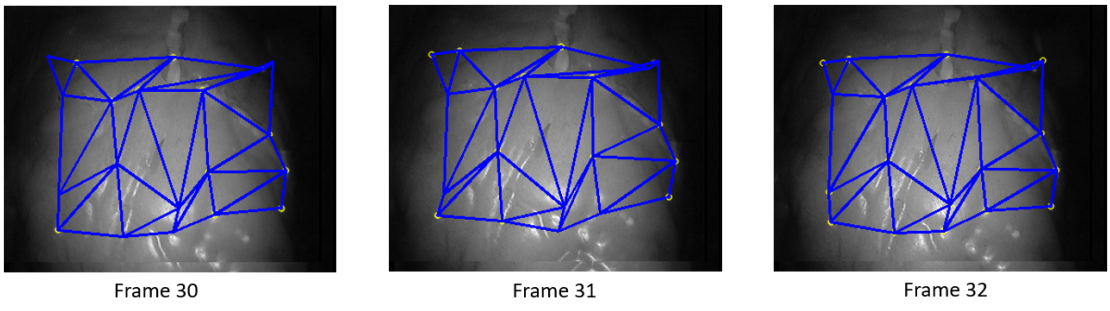


(a)

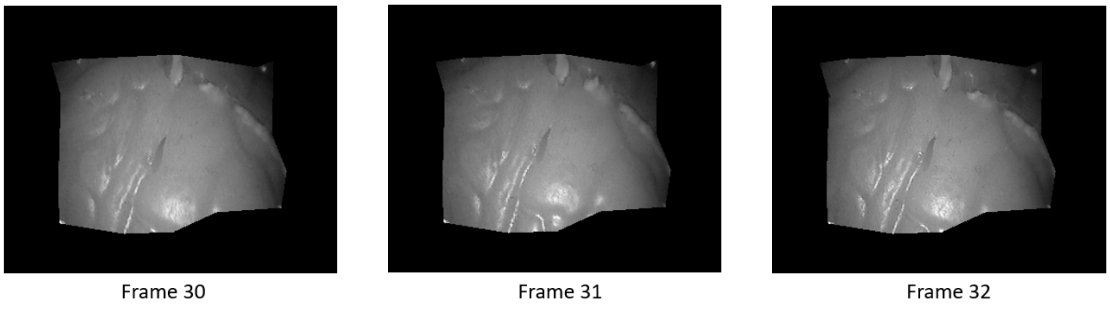


(b)

Figure 3.4: The registration and restoration results in each frame of Heart-1: (a) the registration results, (b) the restoration results.

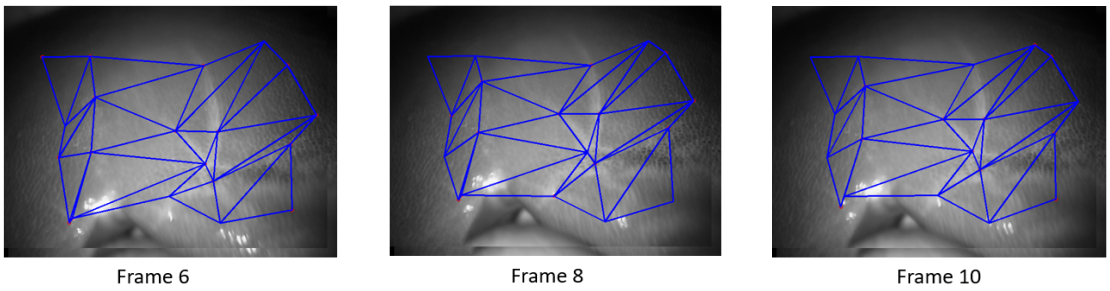


(a)

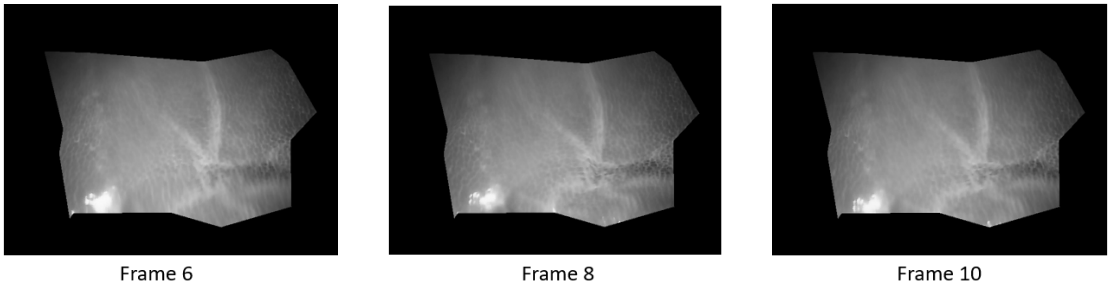


(b)

Figure 3.5: The registration and restoration results in each frame of Heart-2: (a) the registration results, (b) the restoration results.



(a)

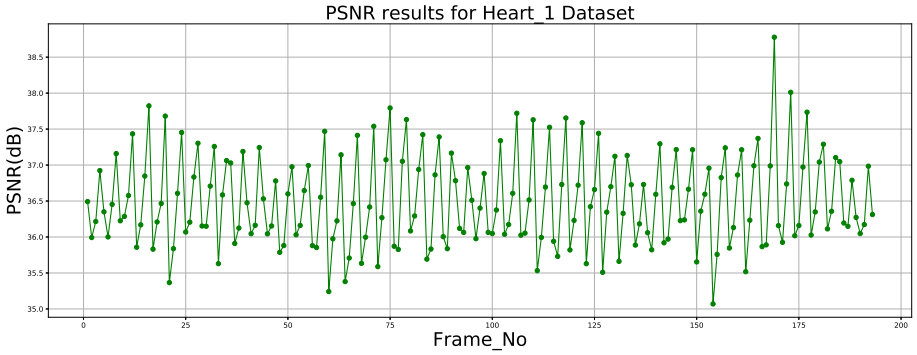


(b)

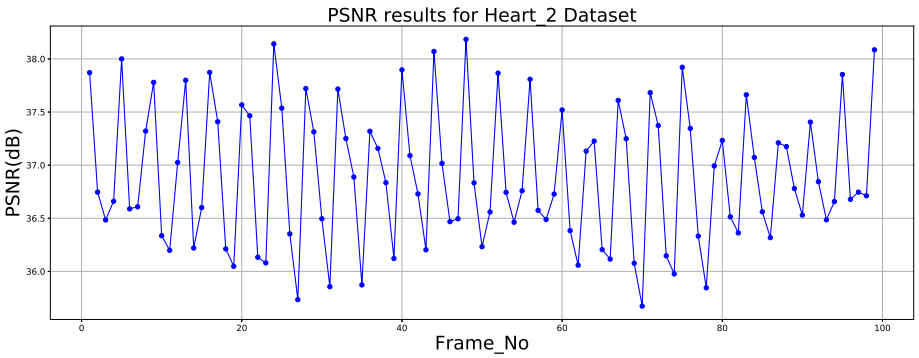
Figure 3.6: The registration and restoration results in each frame of Liver: (a) the registration results, (b) the restoration results.

Table 3.1: Processing Time of Our Proposed Method

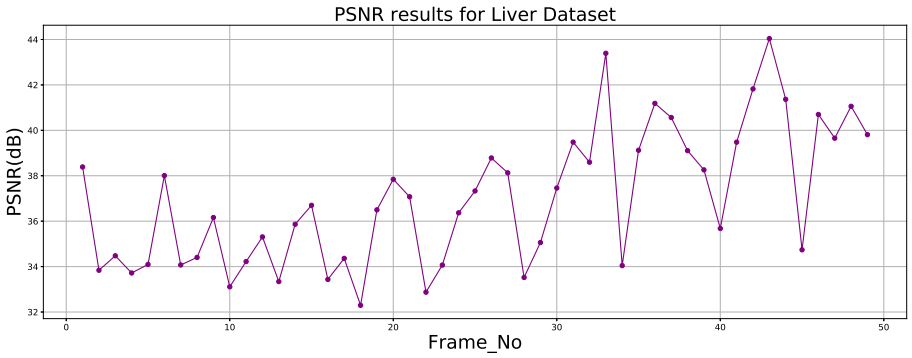
Process (for 194 frames)	Processing Time for Each Process (second)	Processing Time by Binding Each Process (second)
1. Feature Detection and Feature Matching	3.123	3.123
2. Taking the previous point if there is no corresponding point with block-wise features after feature matching and draw key-points on each image	0.0073	(1+2) 3.196
3. Overlaying Mesh and Saving Original images and Results	1.722	(1+2+3) 4.918
4. Superimposing	856.734	(1+2+3+4) 861.652
5. Restoration and Saving restoration results	410.988	(1+2+3+4+5) 1272.640



(a)

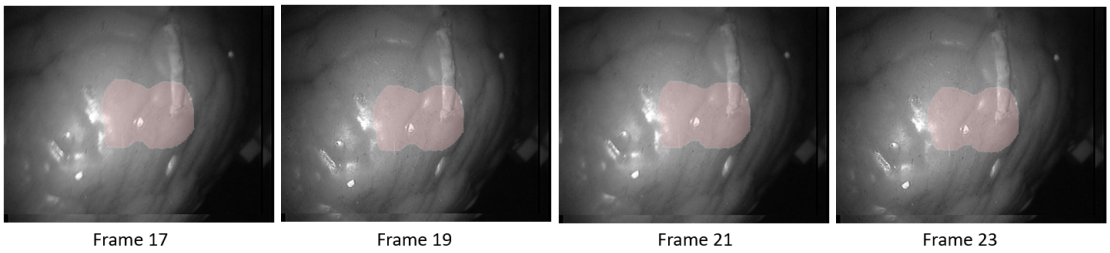


(b)

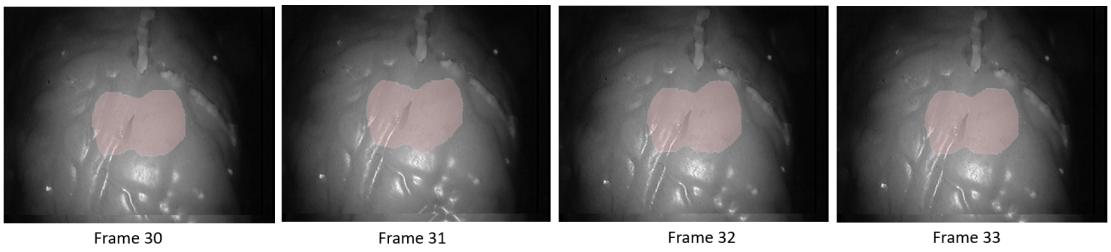


(c)

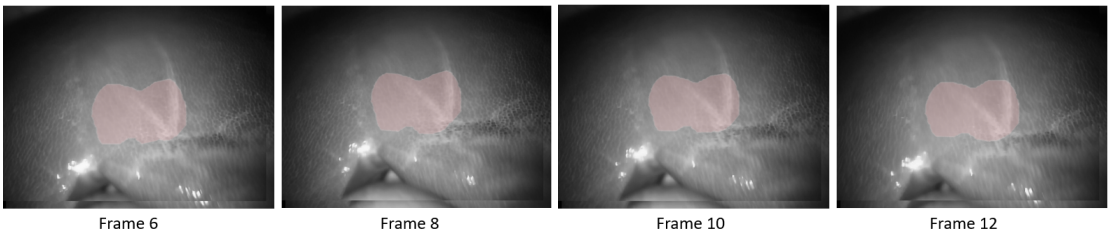
Figure 3.7: PSNR results for stereo endoscopic videos in each frame: (a) PSNR results for Heart-1, (b) PSNR results for Heart-2, and (c) PSNR results for Liver.



(a)



(b)



(c)

Figure 3.8: Superimposing a tumor image on an organ: (a) superimposing a tumor image on each frame of Heart-1, (b) superimposing a tumor image on each frame of Heart-2, and (c) superimposing a tumor image on each frame of Liver

3.5 Summary

We proposed a blockwise feature-based registration framework for deformable medical images. In our approach, the detected feature points in the initial image frame are divided into blocks based on their coordinates and the blockwise feature points are picked up by using response values of the detected feature points. Tracking of feature points is performed by finding the corresponding feature points between the blockwise feature points of the initial image frame and all the detected feature points of the current image frame. We presented the effectiveness of our framework by using sparsely textured endoscopic stereo video datasets. Since the best feature points are used to find the corresponding points between the initial image and input images, the registration can be effectively performed on sparsely textured surfaces. The experimental result showed that our framework can be applied even to sparsely textured medical images.

Although sometimes there are no corresponding feature points in some frames, it is compensated by using the corresponding feature points of the previous frame. We consider only for the 2D registration of the deformable surfaces.

Chapter 4

Comparison of the Two Proposed Methods

We proposed two methods for the deformable registration of sparsely textured surfaces.

4.1 The First Proposed Method

In the first method, the registration is performed by using an intensity-based approach. In this approach, the registration is performed by finding the corresponding 2D vertex position of the mesh between the initial image and the input images.

Since we performed the registration of both texture and shape of the deformable surfaces, we define two cost functions for texture registration and shape registration. Therefore, we define the weights, λ_1 and λ_2 for these cost functions. Then, the total cost function was defined by combining these costs. Since the total cost function is approximately minimized by using the gradient descent method, the weight, w is defined for gradient calculation. Therefore, all weights are important for good registration performance and the balance between λ_1 and λ_2 is also important for the registration of both texture and shape.

Advantages

Our first proposed method can perform the registration on densely and sparsely textured surfaces and register for both texture and shape of deformable surfaces, and reconstruct the 3D shape of the deformable surfaces. It can be also applied on sparsely textured surfaces such as human organs.

Limitations

The registration performance depends on all initial weight values and the balance between λ_1 and λ_2 for both texture and shape registration. Assuming the deformable surface should not stretch or shrink is also a limitation of our approach. It cannot well perform the registration on deformable surfaces that have very less texture and higher illumination.

4.2 The Second Proposed Method

In second method, the registration process is performed by a feature-based approach. By means of AKAZE feature detector, the feature points are detected on both initial image and input images. Since the second proposed method is a blockwise approach, it is important to exist the detected feature points throughout the surfaces to divide all detected features into blocks and to pick up the best feature in each block. Therefore, the threshold parameter of the AKAZE feature detector is also important for this condition. The registration is performed by finding the correspondence between blockwise features of the initial image and all detected features of the input images. For outlier removal, KNN and Lowe's ratio test are applied.

Advantages

Since the processing time of the registration is 4.918 seconds for 194 image frames, this approach can be used for real-time registration. The second proposed method can be applied on very less textured surfaces.

Limitations

For outlier removal, since KNN and Lowe's ratio test are applied, good matches are determined if the distance ratio of the first closest one and the second closest one is smaller than the threshold value. Theoretically, the smaller the threshold value, the fewer matches we get. If the image provides rich feature points, the smaller threshold value of the ratio test can provide a sufficient number of reliable matches. Although a larger threshold value of the ratio test can provide sufficient matches if the images have poor feature points, the outliers can contain among the matches. On the other hand, good matches can be excluded due to the smaller threshold value. Therefore, the performance of the registration also depends on the threshold value of the ratio test.

4.3 Comparison of Two Proposed Methods

The two proposed methods are compared based on the registration performance in terms of texture information. Therefore, by mean of PSNR results, the two method are evaluated.

4.3.1 Heart-1

The comparison of PSNR results between the first approach and the second approach is shown in Fig. 4.1. The result shows that the second approach is better than the first approach because some of the areas of the Heart-1 dataset have very less texture.

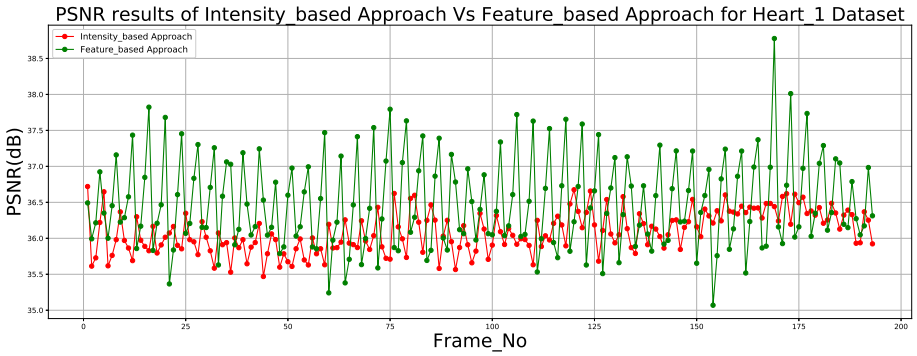


Figure 4.1: PSNR Results of Intensity-based Approach vs Feature-based Approach for Heart-1 Dataset

4.3.2 Heart-2

Fig. 4.2 shows the PSNR results comparison of two proposed methods. Although most of the results of the first proposed are better than those of the second proposed method while some of the results of the second proposed are better than the first method according to the graph. The mesh cannot well track the deformation of the organ because the movement of the organ is upward and downward movement.

4.3.3 Liver

According to the Liver dataset, most of the PSNR results of the second proposed method is better than the first proposed method as shown in Fig. 4.3. Since the

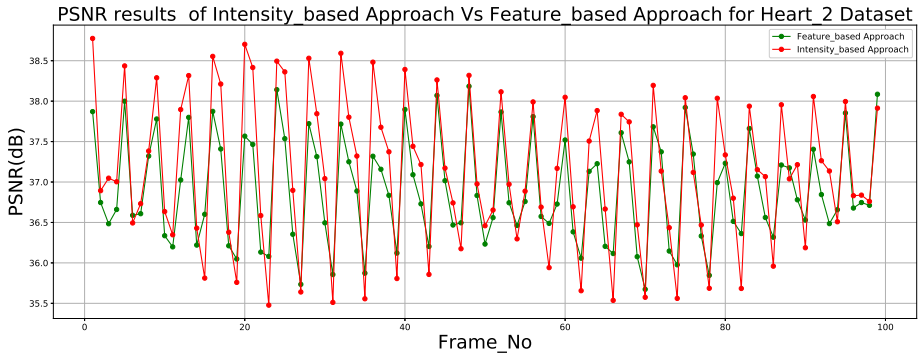


Figure 4.2: PSNR Results of Intensity-based Approach vs Feature-based Approach for Heart-2 Dataset

texture in the Liver video has very smooth, with less texture, and illumination, the first proposed method cannot track the movement of the organ.

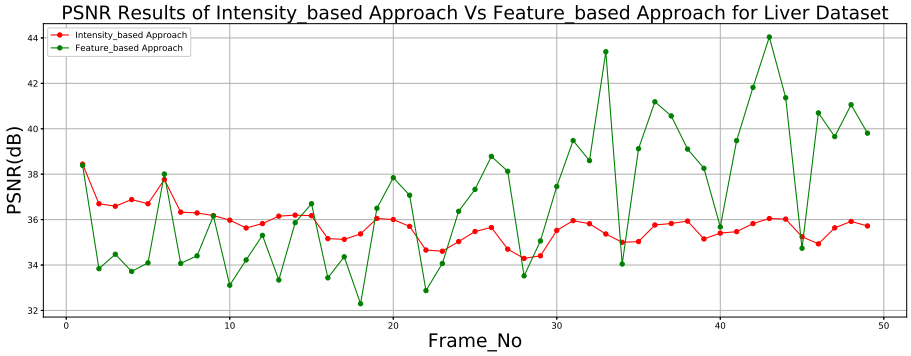


Figure 4.3: PSNR Results of Intensity-based Approach vs Feature-based Approach for Liver Dataset

Chapter 5

Conclusion and Future Work

This chapter concludes our thesis works and shows the future work of our thesis.

5.1 Conclusion

In this thesis, we addressed the problem of registration on sparse texture surfaces of the deformable objects.

We proposed two registration methods, which can be used to register on sparsely textured surfaces such as human organs. The first approach can perform the registration in terms of both texture and shape of the deformable objects with sparse texture. It was confirmed that the first approach has good capability of the registration in term of both texture and shape of densely and sparsely textured objects by comparing three methods using texture or shape information.

In the second approach, the registration can perform even on sparsely textured surfaces such as human organs by using the small set of the best feature points. Moreover, we showed that the demonstration of the surgery assistance by superimposing the tumor onto the deformable surfaces. For demonstration of our approaches, we conducted the experiments using videos of densely and sparsely textured papers, and endoscopic stereo video datasets.

5.2 Future Work

Although our approaches are intended for surgery assistance,there are some limitation that we do not consider occlusion due to the surgical instruments, the organs bleeding, and large deformation due to respiration motion. In future, we will con-

sider superimposing in depth (3D virtual objects) to enhance the visualization of the internal organs more.

Bibliography

[1] Alcantarilla, P.F. and Solutions, T.: Fast explicit diffusion for accelerated features in nonlinear scale spaces. IEEE Trans. Patt. Anal. Mach. Intell, 34(7), pp.1281-1298, 2011.

[2] Baker, S., Matthews, I.: Lucas-Kanade 20 Years On: A Unifying Framework.

[3] Dyke, R. M., Lai, Y. K., Rosin, P. L., Tam, G. K.: Non-rigid registration under anisotropic deformations. Computer Aided Geometric Design, 71, pp.142-156, 2019.

[4] Ellingsen, L. M., Chintalapani, G., Taylor, R. H., Prince, J. L.: Robust deformable image registration using prior shape information for atlas to patient registration. Computerized Medical Imaging and Graphics, 34(1), pp.79-90, 2010.

[5] Haouchine, N., Cotin, S., Peterlik, I., Dequidt, J., Lopez, M.S., Kerrien, E. and Berger, M.O.: Impact of soft tissue heterogeneity on augmented reality for liver surgery. IEEE transactions on visualization and computer graphics, 21(5), pp.584-597, 2014.

[6] Hossein-nejad, Z. and Nasri, M.: Image registration based on SIFT features and adaptive RANSAC transform. International Conference on Communication and Signal Processing (ICCSP), pp. 1087-1091, 2016.

[7] Huang, Q. X., Adams, B., Wicke, M., Guibas, L. J.: Non-rigid registration under isometric deformations. Computer Graphics Forum. 27(5), pp. 1449-1457, 2008.

[8] Jakubović, A. and Velagić, J.: Image feature matching and object detection using brute-force matchers. International Symposium ELMAR. IEEE, pp. 83-86, 2018.

[9] Kajihara, T., Funatomi, T., Makishima, H., Aoto, T., Kubo, H., Yamada, S., Mukaigawa, Y.: Non-rigid registration of serial section images by blending transforms for 3D reconstruction. Pattern Recognition, 96, p.106956, 2019.

[10] Kim, J.H., Bartoli, A., Collins, T. and Hartley, R.: Tracking by detection for interactive image augmentation in laparoscopy. In International Workshop on Biomedical Image Registration . Springer, Berlin, Heidelberg, pp. 246-255, 2012.

[11] Liao, F., Chen, Y., Chen, Y. and Lu, Y.: SAR Image Registration Based on Optimized Ransac Algorithm with Mixed Feature Extraction. In IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium, pp. 1153-1156, 2020.

[12] Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. International journal of computer vision, 60(2), pp.91-110, 2004.

[13] London, I.: Hamlyn Centre laparoscopic/endoscopic video datasets, 2019. URL http://hamlyn.doc.ic.ac.uk/vision/. Accessed 15 Jan 2019

[14] Lu, X., Ma, H., Zhang, B.: A non-rigid medical image registration method based on improved linear elastic model. Optik, 123(20), pp.1867-1873, 2012.

[15] Lu, X., Zhao, Y., Zhang, B., Wu, J., Li, N., Jia, W.: A non-rigid cardiac image registration method based on an optical flow model. Optik, 124(20), pp.4266-4273, 2013.

[16] Maes, F., Loeckx, D., Vandermeulen, D., Suetens, P.:Image registration using mutual information. In: Paragios, N., Duncan, J., Ayache, N. (eds) Handbook of Biomedical Imaging. Springer, Boston, MA, 2015.

[17] Ngo, D. T., Park, S., Jorstad, A., Crivellaro, A., Yoo, C., Fua, P.: Dense Image Registration and Deformable Surface Reconstruction in Presence of Occlusions and Minimal Texture. IEEE International Conference on Computer Vision, pp. 2273-2281, 2015.

[18] Phogat, R.S., Dhamecha, H., Pandya, M., Chaudhary, B. and Potdar, M.: Different image registration methods—an overview. Int J Sci Eng Res, 5, pp.44-9, 2014.

[19] Puerto-Souza, G.A. and Mariottini, G.L.: A fast and accurate feature-matching algorithm for minimally-invasive endoscopic images. IEEE transactions on medical imaging, 32(7), pp.1201-1214, 2013.

[20] Qiu, Z., Tang, H. and Tian, D.: Non-rigid medical image registration based on the thin-plate spline algorithm. In WRI World Congress on Computer Science and Information Engineering ,Vol. 2, pp. 522-527, 2009.

[21] Salzmann, M., Pilet, J., Ilic, S., Fua, P.: Surface deformation models for non-rigid 3D shape recovery. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(8), pp.1481-1487, 2007.

[22] Salzmann, M., Moreno-Noguer, F., Lepetit, V., Fua, P.: Closed-form solution to non-rigid 3D surface registration. European Conference on Computer Vision, pp. 581-594, 2008.

[23] Savran, A., Sankur, B.: Non-rigid registration of 3D surfaces by deformable 2D triangular meshes. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp. 1-6, 2008.

[24] Shen, T.W. and Cheung, K.W.:Improved local optimization for adaptive bases non-rigid image registration. In IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1526-1529, 2016.

[25] Sidorov, K. A., Richmond, S., Marshall, D.: Efficient Groupwise Non-Rigid Registration of Textured Surfaces. IEEE Conference on Computer Vision and Pattern Recognition, 2011.

[26] Souza, G.A.P., Adibi, M., Cadeddu, J.A. and Mariottini, G.L.: Adaptive multi-affine (ama) feature-matching algorithm and its application to minimally-invasive surgery images. In IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2371-2376, 2011.

[27] Stoyanov, D. and Yang, G.Z.: Soft tissue deformation tracking for robotic assisted minimally invasive surgery. In annual international conference of the IEEE engineering in medicine and biology society, pp. 254-257, 2009.

[28] Wang, J., Jiang, T.: Nonrigid registration of brain MRI using NURBS. Pattern Recognition Letters, 28(2), pp.214-223, 2007.

[29] Zhang, R., Zhou, W., Li, Y., Yu, S. and Xie, Y.: Nonrigid registration of lung CT images based on tissue features. Computational and mathematical methods in medicine, 2013.

[30] Zheng, Q., Wang, Q., Ba, X., Liu, S., Nan, J. and Zhang, S.: A Medical Image Registration Method Based on Progressive Images. Computational and Mathematical Methods in Medicine, 2021.

[31] https://en.wikipedia.org/wiki/Image registration

LIST OF PUBLICATIONS

Journal Articles

1. Su Wai Tun, Takashi Komuro, and Hajime Nagahara, " 3D Registration of Deformable Objects Using a Time-of-Flight Camera " *Lecture Notes in Computer Science (LNCS), volume 13017, Springer, 2021*, 11-pages, pp.455-465.

2. Su Wai Tun, Takashi Komuro, and Hajime Nagahara, " Blockwise Feature-based Registration of Deformable Medical Images " *Lecture Notes in Computer Science (LNCS), volume 13393, Springer, 2022*, 12-pages, pp.472-482.