Robust and Fast Schemes for Generation of Matched Features in MIS Images

Muhammad Reza Pourshahabi

A Thesis

in

The Department

of

Electrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy (Electrical and Computer Engineering) at

Concordia University

Montréal, Québec, Canada

August 2024

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Muhammad Reza Pourshahabi

Entitled: Robust and Fast Schemes for Generation of Matched Features in MIS Images

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Electrical and Computer Engineering)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____ Chair
Dr. Rajamohan Ganesan

_____ External Examiner
Dr. Luc Duong

_____ Arm's length Examiner
Dr. William E. Lynch

_____ Examiner
Dr. Hassan Rivaz

_____ Examiner
Dr. Wei-Ping Zhu

_____ Thesis Supervisor
Dr. M. Omair Ahmad

_____ Thesis Supervisor
Dr. M.N.S. Swamy

Approved by _____
Dr. Jun Cai, Graduate Program Director

September 04, 2024 _____
Dr. Mourad Debbabi, Dean, Gina Cody School of Engineering and Computer Science

Abstract

Robust and Fast Schemes for Generation of Matched Features in MIS Images

Muhammad Reza Pourshahabi, Ph.D.

Concordia University, 2024

Robotic-assisted minimally invasive surgery (MIS) offers numerous benefits including smaller incisions, faster recovery, enhanced precision, and remote operations. Image processing operations such as 3D visualization, augmented reality, and image registration, which are often feature-based, are used in MIS. Feature detection, extraction, and matching (FDEM) and feature matching refinement (FMR) constitute the cornerstone of these operations. MIS images are affected by deformation, occlusions, and specular reflection, which hinder the processes of FDEM and FMR, severely affecting the number of matched features.

FDEM is a process in which, given a pair of images, certain distinctive features are detected from the pair, then suitably represented as feature vectors, and finally, the corresponding feature vectors are compared and matched leading to a set of matched features known as a putative set for the pair. On the other hand, FMR is a process in which

the falsely matched pairs of features are, as much as possible, removed from a putative set. The existing FDEM and FMR schemes are computationally expensive or lead to a set of matched features that are not well dispersed over the region of interest and suffer from having an insufficient number of true matches.

The overall objective of this thesis is to propose robust and fast schemes for generation of matched features in MIS images. In the first part of the thesis, a very fast and accurate FMR scheme is proposed. The main idea used in developing this scheme is in determining the size of local neighborhoods so that the smoothness of deformation field can be effectively applied to check the feature topology preservation between the corresponding regions of the pair of images to identify the true matches in the putative set of the pair. In the second part, a fast and accurate FDEM scheme that combines the strong attributes of three well-known FDEM schemes, SIFT, SURF and ORB, is proposed. The focus is on producing putative sets of matched features that have a good spatial quality in addition to a good matching quality. Extensive experiments are conducted to demonstrate the effectiveness of the proposed FMR and FDEM schemes.

Acknowledgments

I would like to express my deepest gratitude to everyone who has supported me throughout the journey of completing this thesis.

First and foremost, I am profoundly grateful to my two supervisors, Professor M. Omair Ahmad, and Professor M.N.S. Swamy. Their invaluable guidance, encouragement, and expertise have been instrumental in shaping this work. Their patience and insightful feedback have pushed me to achieve more than I thought possible.

I owe an enormous debt of gratitude to my loving wife, Hedieh. Her unwavering support, understanding, and encouragement have been my rock throughout this process. Her belief in me has been a constant source of motivation.

I am deeply thankful to my parents, my sister, my wife's parents, and my friends for their love, support, and encouragement. Their belief in me has been invaluable and has made this journey possible.

This thesis would not have been possible without the support and encouragement of all these wonderful people. Thank you.

Contents

List of Figures

List of Tables

List of Symbols

A_f	The area of the bounding box enclosing the ROI in the fixed image.
A_m	The area of the bounding box enclosing the ROI in the moving image.
B_f	The perimeter of the bounding box enclosing the ROI in the fixed image.
B_m	The perimeter of the bounding box enclosing the ROI in the moving image.
$D_{ORfixed}$	The set of detected ORB features from the fixed image.
$D_{ORmoving}$	The set of detected ORB features from the moving image.
$D_{SUfixed}$	The set of detected SURF features from the fixed image.
$D_{SUmoving}$	The set of detected SURF features from the moving image.
D_{th}	The similarity threshold between the displacement vectors of two feature points.
E	Target registration error.
$E_{ORfixed}$	The set of ORB binary descriptors for the ORB detected features in $D_{ORfixed}$.
$E_{ORmoving}$	The set of ORB binary descriptors for the ORB detected features in $D_{ORmoving}$.
$E_{SIfixed}$	The set of SIFT descriptors for the detected SURF features in $D_{SUfixed}$.
$E_{SImoving}$	The set of SIFT descriptors for the detected SURF features in $D_{SUmoving}$.
$\boldsymbol{L}$	A one-dimensional array of labels in which the label in the i^{th} entry indicates

whether the match (P_i, P_i') is *True (T), False (F),* or *Unknown (U)*.

N_{DBF} The number of detected blob features.

N_{DCF} The number of detected corner features.

N_{DF} The number of detected corner and blob features.

N_M The number of matches in the generated putative set.

Q Spatial quality (used in the proposed FDEM scheme).

Q_1 Density metric (used in the proposed spatial quality metric).

Q_2 Dispersion metric (used in the proposed spatial quality metric).

R Clark-Evans aggregation index.

R_{max} The maximum value of the Clark-Evans aggregation index.

$R_{N_{fixed}}$ The aggregation index for the fixed image.

$R_{N_{moving}}$ The aggregation index for the moving image.

$S_{d\mu_i}$ A set of displacement vectors for all the features in the neighborhood μ_i.

$S_{d\rho_i}^t$ A set of displacement vectors of the *True* matches of the features in the neighborhood ρ_i.

$S_{D\mu_i}$ A set of magnitudes of the differences between $\boldsymbol{d}_i$ and each of the displacement vectors in $S_{d\mu_i}$.

S_M A set of matched features.

S_{ORB} The set of matched features using the descriptors in $E_{ORfixed}$ and $E_{ORmoving}$.

S_p The final putative set of matched features.

$S_{SURFSIFT}$ The set of matched features using the descriptors in $E_{SIfixed}$ and $E_{SImoving}$.

$S^t_{SURFSIFT}$	The resulting refined set of matched features obtained by applying VSLD-FMR to $S_{SURFSIFT}$.
S_{μ_i}	A set of matches of all the feature points within the neighborhood μ_i.
$S^s_{\mu_i}$	A subset of the feature points in μ_i whose displacement vectors are similar to $\boldsymbol{d}_i$.
$S^t_{\rho_i}$	A set of all *True* matches of the feature points within the neighborhood ρ_i.
T_{FDEM}	Time to produce a putative set of matched features.
$T_{reg.}$	Time required to perform the registration.
$\boldsymbol{V}$	A vote vector containing the values v_i.
$\boldsymbol{d}^t_{GW}$	A Gaussian weighted average vector of all the displacement vectors in the set $S^t_{d\rho_i}$.
$\boldsymbol{d}_i$	The displacement vector of feature point P_i.
$\boldsymbol{d}^t_{ij}$	The displacement vector of the j^{th} *True* matched feature within the neighborhood ρ_i.
n_i	The number of neighboring feature points of P_i within the neighborhood μ_i.
n_{si}	The number of similar displacement vectors of the feature points in the neighborhood μ_i.
n_{th}	A threshold on n_i.
n^s_{th}	A threshold on n_{si}.
n^t_{th}	A threshold on n_{ti}.
n^v_{th}	A threshold on v_i.
n_{ti}	The number of *True* matches of the feature points within the neighborhood ρ_i.
th_H	A threshold used for determining high spatial quality Q.
th_M	A threshold used for determining medium spatial quality Q ($th_M < th_H$).

w_1	The weight of the density metric Q_1.
w_2	The weight of the density metric Q_2.
w_{ij}	A Gaussian weight for the j^{th} *True* matched feature in the neighborhood ρ_i.
δ_1	A threshold on the low values of the hue component of a pixel point in the ROI.
δ_2	A threshold on the high values of the hue component of a pixel point in the ROI ($\delta_2 > \delta_1$).
δ_3	A threshold on the saturation component of a pixel in the ROI.
δ_4	A threshold on the value component of a pixel in the ROI.
δ_5	A threshold on the number of pixels in the isolated components of the ROI.
δ_6	A threshold on the number of ORB detected features.
δ_7	The size of the square blocks used for partitioning the bounding box enclosing the ROI.
μ_i	A circular neighborhood of radius R_1 centered at feature point P_i.
v_i	The number of votes received by the feature point P_i towards its being a *True* match.
ρ_{fixed}	The density of the features in the fixed image.
ρ_i	A circular neighborhood of radius $R_2 > R_1$ centered at feature point P_i.
ρ_{max}	The maximum density of features in an image, determined empirically.
ρ_{moving}	The density of the features in the moving image.
τ_d	A matching distance threshold.
τ_r	A matching ratio threshold.
τ_E	A Euclidean distance threshold between two spatial locations.

List of Abbreviations

ASpanFormer	Adaptive span transformer
BRIEF	Binary robust independent elementary features
DBSCAN	Density-based spatial clustering of applications with noise
EMDQ	Expectation maximization and dual quaternion
FAST	Features from accelerated segment test
FDEM	Feature detection, extraction, and matching
FMR	Feature matching refinement
FP	False positives
FN	False negatives
GLOF	Guided local outlier factor
HMA	Hierarchical multi-affine
LMR	Learning a two-class classifier for mismatch removal
LoFTR	Local feature matching with transformers
LPM	Locality preserving matching
LWM	Local weighted mean
MIS	Minimally invasive surgery

NSB FMR	Neighborhood structure-based FMR
ORB	Oriented FAST and rotated BRIEF
PATS	Patch area transportation with subdivision for local feature matching
RANSAC	Random sample consensus
RFM	Robust feature matching
ROI	Region of interest
SLAM	Simultaneous localization and mapping
SIFOR	SURF combined with SIFT and ORB
SIFT	Scale invariant feature transform
SGMNet	Seeded graph matching network
SuperPoint	Self-supervised interest point detection and description
SURF	Speeded up robust features
TB FMR	Transformation-based FMR
TN	True negatives
TP	True positives
TPS	Thin plate spline
TRE	Target registration error
VFC	Vector field consensus
VSLD-FMR	Voting based on similarity of local displacement vectors FMR scheme

Chapter 1
Introduction

1.1 General

Simultaneous localization and mapping (SLAM), 3D visualization, augmented reality, image registration and mosaicking are some of the image processing operations, which are often feature-based. Feature detection, extraction, and matching (FDEM) are the essential components of these image processing operations. Feature point detection in an FDEM process is the identification of distinctive keypoints, such as corners and blobs in the images. The purpose of the feature extraction is to construct a description vector, the signature of the feature, for each detected feature. This description vector is desired to be invariant to the geometric and radiometric variations of the feature. Geometric variations include rotational, scaling, affine, projective, and non-linear variations, whereas radiometric variations are due to the sensitivity of the camera sensors to the changes in the lighting conditions [1]. The feature matching part is carried out to match the extracted features between a pair of two images of the same scene. Depending on how the images are captured, each of the two images in the pair is named differently. For example, in stereo vision, we use the terms left and right images to refer to these two images, whereas in monocular vision, we use the terms fixed or reference image to refer to the first image and the moving, test or sensed image to refer to the second image. In this thesis, we refer to the two images as fixed (or reference) and moving images, regardless of the way they are

captured. Each extracted feature from the moving image is compared with all the extracted features of the fixed image. For a given extracted feature of the moving image, a feature in the fixed image is generally considered to be matched [2] if (i) the distance between the descriptor of the feature in the fixed image and that of the extracted feature in the moving image is the lowest among all the distances between the extracted feature in the moving image and all the other features in the fixed image, (ii) this lowest distance is smaller than a given threshold, referred to as the matching distance threshold, and (iii) the ratio of this lowest distance and the distance between the extracted feature in the moving image and the second closest feature in the fixed image is less than a given threshold, referred to as the matching ratio threshold. The performance of an FDEM based application is very much dependent on the feature detection and extraction capability of the FDEM scheme used and its capability of matching the extracted features from the two images. The process of FDEM results in a set, known as putative set, of all the matched features between the fixed and moving images. A putative set obtained from an FDEM scheme is denoted as $\{(P_i, P_i'), i = 1, \dots, N\}$, where P_i and P_i' represent the spatial positions of two matched feature points in the fixed image and moving image, respectively, and N represents the total number of putative matches. The putative set of matched features generally includes some false matches (mismatches or outliers).

1.2 Challenges in Generating Matched Features in MIS Images

Robotic-assisted minimally invasive surgery (MIS) by providing a range of benefits such

as smaller incisions, reduced chance for infection, faster recovery time, lower risk of complications, enhanced precision, surgeon ergonomics, and remote surgery capability [3]-[5], has a very important place in the landscape of modern surgical practices. SLAM, 3D visualization, augmented reality, image registration and mosaicking, which are often feature-based, are frequently used in robotic-assisted MIS surgery [6]-[18].

MIS images undergo significant deformation because of the patient motion, breathing, heartbeat, and interaction with the surgical instruments [16]. Such images also have occlusions caused by the surgical instruments and specular reflection resulting from the shiny tissue surfaces. Furthermore, the tissue surfaces have repetitive textures, and therefore, do not have a rich set of distinctive features. This hinders the process of feature detection and matching between the fixed and moving images. Therefore, the number of matched features resulting from an FDEM scheme is usually low. In view of these special characteristics of MIS images, the process of FDEM may be severely affected. For example, in view of the occlusion and specular reflection, it may not be possible to detect some of the features in one or both of the images. Moreover, the lack of a rich set of distinctive features may result in providing similar feature descriptions. This, along with other reasons mentioned above that lead to a small number of matches, is responsible for the low inliers ratio in MIS images. The inliers ratio is defined as $\Gamma = N_t/N$, where N is the number of matches in the putative set and N_t is the number of true matches. This low value of the inliers ratio underscores the need for an effective feature matching refinement (FMR) scheme.

1.3 A Brief Literature Review of Schemes for Generation of a Set of Matched Features

There are two important processes involved with the generation of a set of matched features, feature detection, extraction, and matching (FDEM) and feature matching refinement (FMR).

SIFT [2], SURF [19] and ORB [20] are the three most popular FDEM schemes in the literature. SIFT and SURF are known for their robustness in detecting and extracting distinctive feature points that are of blob type in images even in the presence of various deformation, scaling, rotation, and partial occlusion. However, SURF is much faster than SIFT in detecting features, whereas SIFT provides descriptions of the extracted features that are generally more distinctive for MIS images, even though it has a slightly larger time for feature extraction. ORB is an FDEM scheme that detects corner features and is known for its very fast processing speed. However, the descriptors of the features extracted by ORB are not as scale invariant as SIFT and SURF descriptors are, even though they are orientation invariant. Moreover, even though the processing time per feature of ORB is quite small, the total processing time is not small in view of the fact that the total number of detected features that need to be processed is generally larger than in SIFT and SURF. It should also be noted that the existing FDEM schemes have focused on achieving good quality putative sets of matched features from the viewpoint of the matching accuracy of the extracted features, but have ignored the density and the spread of the matched features, which are desirable characteristics in applications of putative sets. There are several deep learning-based schemes for FDEM in the literature, of which LoFTR [21], MatchFormer

[22], ASpanFormer [23], and PATS [24] are some of the state-of-the-arts known for providing dense sets of matched features. However, these deep learning techniques despite being run on GPU machines, their run-times are still very large. SuperPoint [25] is one of the most popular deep learning-based methods designed for only detecting and extracting features. SuperGlue [26] and its successors, SGMNet [27] and LightGlue [28] are state-of-the-art deep learning-based feature matching schemes that have obtained the results for the FDEM task by using SuperPoint for detection and extraction. Specifically, the matching times of SuperGlue, as well as those of its two successors, as reported by the authors of LightGlue [28], are still not small, specially when the number of feature points per image is large, say larger than 2K. As for the performance, these three schemes, when combined with SuperPoint, provide similar Recall values that can be considered to be good, but the Precision values are not as large as desirable.

The putative sets resulting from an FDEM scheme have a number of matches that are false, and thus result in lowering their inliers ratios. This is especially so in the case of MIS images. The idea of FMR is to remove the false matches from a putative set of matched features generated by an FDEM scheme. The existing FMR schemes can be classified into two categories. The first category consists of methods that estimate a transformation model to identify the inliers among the matches of a putative set [29]-[32], while the second category includes methods that rely on the local neighborhood structures of the matched features to identify the inliers [33]-[36]. We refer to these two categories as transformation-based (TB) and neighborhood structure-based (NSB) FMR schemes, respectively. The TB FMR schemes (RANSAC-affine [29], HMA [30], EMDQ [31], and VFC [32]) impose a geometric constraint, such as an affine constraint, on the spatial

positions of the putative matches, $\{(P_i, P_i'),\ i = 1, \ldots, N\}$, to estimate the best transformation model f that fits the inliers. The transformation model f is typically estimated through an iterative process. Ultimately, matches that do not conform to the resulting estimated transformation model f are identified as mismatches or outliers and discarded. The NSB FMR schemes (LPM [33], LMR [34], RFM [35], GLOF [36]) refine the matched features by examining the preservation of the topology between the corresponding neighborhoods in the fixed and moving images. In feature matching refinement, topology refers to the spatial relationships or arrangements of features within a local neighborhood of an image. It involves considering the relative positions, orientations, and connections of features in that neighborhood. Under the assumption that the deformation field is smooth, the topology (spatial relationships) within a local neighborhood of an inlier in the fixed image will be preserved in the corresponding neighborhood of the moving image. The preservation of topology means that the relative positions and arrangements of features in the neighborhood remain consistent between the two images despite the deformation. The preservation of the topology between the corresponding neighborhoods in the fixed and moving images is the backbone of all the FMR schemes in this category. A consequence of preservation of neighborhood topology between the corresponding neighborhoods of the two images implies similarity of the displacement vectors of neighboring features, which has been used in the state-of-the-art neighborhood structure-based FMR schemes (LPM [33], LMR [34], RFM [35], GLOF [36]). It is worth noting that both TB FMR and NSB FMR schemes face critical challenges when the percentage of true matches is very low, which is generally the case for MIS

images.

1.4 Motivation and Objective

Among the TB FMR schemes, only two of the methods, namely, HMA and EMDQ, are robust to the inliers ratio of the original putative set. However, the processing time of HMA is large and especially so, when the number of matches in the putative set is large, while the processing time of EMDQ is large when the inliers ratio of the original putative set is low. On the other hand, the existing NSB FMR schemes generally have much faster processing times, but are not as robust to the inliers ratio of the original putative set as HMA and EMDQ are. The NSB FMR schemes highlight the importance of considering local structures and neighborhood topology preservation of the inliers between the fixed and moving images in refining matched features. These schemes [33]-[36] consider the K nearest neighbors (K-NN) in fixed and moving images for each match. When K is small, the number of inliers in the set of K nearest neighbors may be small or even zero. However, it needs to be pointed out that even in the case where K is chosen to be very small, for example for K = 1, the neighborhood produced cannot be guaranteed to be a local neighborhood. When K is large, some of the inliers in the neighborhood may not be local. Therefore, the optimal value of K varies for each match. Although some of the schemes consider multiple values for K, this strategy works well only when the inliers ratio of the original putative set of matches is high. This potentially gives rise to difficulties in successfully applying the preservation of neighborhood topology for classifying the matches as inliers or outliers, when the inliers ratio of the original putative set is low,

which generally is the case for MIS images. Consequently, the robustness of these schemes is compromised, resulting in an inaccurate identification of the inliers.

SIFT, SURF and ORB offer distinct advantages and disadvantages, especially in their feature detection, and extraction parts. It would be worth undertaking an investigation to develop an FDEM scheme that takes advantage of the strong attributes of SIFT, SURF and ORB. It is also worth mentioning that metrics, such as the number of detected features, number of matched features, number of true matched features (which requires a knowledge of the ground truth matches) are used to measure the quality of the putative sets of matched features generated by an FDEM scheme. However, these metrics do not adequately reflect the quality of the resulting putative set of matched features. For an FDEM scheme to produce a good-quality putative set of matched features, there are other characteristics of the putative set, such as the density and dispersion of the matches across the regions of interest in the pair of fixed and moving images, that must be taken into consideration in designing and assessing the performance of an FDEM scheme. We refer to these other characteristics as the spatial quality of the putative set. To the best of our knowledge, there does not exist a formal metric to measure the spatial quality of a putative set. In recent years, the deep learning techniques have provided powerful tool for achieving high accuracy performance in many applications including for the task of FDEM, however, at the expense of requiring large datasets for training of network models and expensive computational resources. As seen from the review of deep learning-based FDEM schemes, the performance level of such schemes is still not very commensurate to that required for targeted applications, such as MIS.

In view of the above limitations of existing FMR and FDEM schemes, the overall

objective of this thesis is to propose robust and fast schemes for generation of matched features in MIS images. For this purpose, in the first part of the thesis, we propose a very fast and accurate FMR scheme for MIS images that is robust to inliers ratios of the original putative set. Then, in the second part of the thesis, we propose a fast and accurate FDEM scheme for MIS images that takes advantage of the existing FDEM schemes to generate a set of putative matched features that has a good overall quality, spatially as well as in terms of the matching accuracy.

In the first part of the thesis, a novel two-stage NSB FMR algorithm which addresses the problems of existing FMR schemes is proposed. In the first stage, a conservative approach is adopted by choosing circular neighborhoods of the size that is a very small fraction of the size of the image. This conservative approach results in forming neighborhoods that better qualify to be local neighborhoods than those formed by using the K-NN based approach. Then, a voting mechanism is devised for the matches within a neighborhood based on the number and similarity of the displacement vectors of the matches within the neighborhood. Casting a vote for a feature point and its neighboring features could be considered more than one time, once when the neighborhood of the feature point in question is considered and again when the neighborhoods of the neighboring feature points are considered. After the voting process, those matches that receive a large number of votes are identified to be true matches. Using the knowledge of true matches found in the first stage, a mechanism is developed in the second stage to determine the status of those matches that still remain unknown, but in a neighborhood larger than that chosen in the first stage. The process carried out in the second stage results in identifying some of the true matches that could not be so identified in the first stage.

In the second part of the thesis, a fast and accurate FDEM scheme is proposed for detection, extraction, and matching of features in MIS images by making a strategic use of the strong attributes of the SIFT, SURF and ORB FDEM schemes. Our strategy in proposing a new FDEM scheme is that we start the detection process by choosing a low-complexity method that results in features that are robust to distortions and at the same time carry out the extraction of the detected features by choosing a method that provides highly scale and rotation invariant feature representations. The next strategy in our scheme is to match the features of the fixed and moving images using these representations and to test the spatial quality of the matches by employing a suitable metric for assessing the spatial quality of a putative set of matched features. For this purpose, we propose a new metric to measure the spatial quality of a set of matched features for a given pair of images. Our third strategy is to decide whether the set of matched features is of sufficient spatial quality based on the outcome of the spatial quality test, or it needs to be supplemented by additional matches obtained from the already detected features or from the detection of other types of features.

1.5 Organization of the Thesis

The thesis is organized as follows. In Chapter 2, the state-of-the-art schemes for feature detection, extraction, and matching, as well as those for feature matching refinement are reviewed. In Chapter 3, a new scheme for FMR is proposed. The underlying principles and ideas on which the proposed scheme is based are first described in detail, and then the two-stage algorithm for the proposed schemes is developed. In Chapter 4, the proposed FDEM

scheme for generating a putative set of matched features for a given pair of images is developed first by giving a top-level description of the scheme and then by providing the incorporation and implementation of the various strategies and by making use of a new metric, which is also developed in this chapter, to measure the spatial quality of the putative set of matched features. Chapter 5 concludes the thesis by providing a summary of the proposed FMR and FDEM schemes and their key novelties. A brief discussion on the scope for further investigation based on the work carried out in this thesis is also included in this chapter.

Chapter 2
Background Material

2.1 Introduction

As stated in Chapter 1, the objective of the thesis is to propose new FMR and FDEM schemes for MIS images. Therefore, it is important to understand the basic ideas and concepts used in the existing FMR and FDEM schemes and their advantages and limitations, before proposing our own methods. In this chapter, we first explore three well-known FDEM schemes, SIFT [2], SURF [19] and ORB [20] including a comparison of the performance and complexity of the various parts of these schemes. We then review the state-of-the-art schemes for FMR, after categorizing them into the transformation-based (TB) and neighborhood structure-based (NSB) FMR schemes. Specifically, we review RANSAC-affine [29], HMA [30], EMDQ [31], and VFC [32] in the TB FMR category, and LPM [33], LMR [34], RFM [35], GLOF [36] in the NSB FMR category, including a comprehensive analysis of these methods, highlighting their strengths and limitations in refining a putative set of matched features.

2.2 Related Work on FDEM Schemes

In this section, we conduct brief reviews of three well-known FDEM schemes, namely, SIFT [2], SURF [19] and ORB [20] on which the FDEM scheme proposed in Chapter 4 is founded.

SIFT is a general-purpose scheme used for detecting, extracting, and matching blob-type features. The process begins by constructing a pyramid of Gaussian filtered images obtained by applying Gaussian filters with increasing values of the filter's standard deviation (referred to as scale) to form an octave. The set of images in the pyramid is known as scale-space representation of the original image. Next, a set of difference of Gaussian (DOG) images are obtained by computing the difference of pair of all adjacent Gaussian filtered images in this octave. Using the set of DOG images, 3 x 3 x 3 windows are formed. If the value of the central pixel in a window has a maximum or minimum value in the entire window, then there exists a blob feature in the original image at the spatial location and the scale corresponding to the central pixel in the window. The spatial location and scale of each blob feature are refined using the interpolation method proposed in [37]. Those blobs that have low contrast or are poorly localized along an edge are eliminated. The above process is repeated by down sampling and applying Gaussian filters in the succeeding octaves. Since the objective in SIFT is to obtain a scale and orientation invariant description for each detected blob feature, next the orientation of each detected feature is determined. For this purpose, the Gaussian-filtered image with the scale closest to the scale of the detected blob feature is selected. The orientation of the blob is estimated from the histogram formed by using the magnitude and orientation of the gradients of all the sample points within a local region surrounding the blob in the selected Gaussian-filtered image. Then, a region of 16 by 16 sample points surrounding the blob (in the selected Gaussian-filtered image) is formed. This 16 by 16 region for a detected blob feature is further sub-divided into 4 by 4 sub-regions. For each 4 by 4 sub-regions, an orientation histogram, which has 8 different directions (0, 45, 90, 135, 180, 225, 270, and

315), is constructed using the magnitude and orientation of the gradients of its 16 sample points. Hence, corresponding to each blob feature, there are 128 orientations, each of which is rotated by an angle associated with the blob orientation already determined to construct a 128-dimensional description vector, which is both scale and orientation invariant. Finally, using the feature descriptors of the blob features in the fixed and moving images, the feature points in the two images are matched by employing the Euclidean distance and the matching scheme described earlier in Chapter 1.

SURF [19] is another general-purpose FDEM scheme that also detects, extracts, and matches blob-type features. For this purpose, SURF first determines an approximate value of the determinant of a Hessian matrix corresponding to each pixel in the original image. Given a specific standard deviation (σ), the Hessian matrix in SURF is defined as a 2 x 2 square matrix consisting of the second order partial derivatives of the Gaussian function in x and y. These partial derivatives are replaced by box filters, whose size is dependent on the standard deviation parameter of the Gaussian function, as the discrete versions of the operators of the second order derivatives. The original image at the given scale is filtered by each of the four box filters giving rise to four filtered images. Now the approximate value of the determinant of the Hessian matrix at a given location (i, j) of the image at a given scale is obtained by using the four-pixel values from the four filtered images each at the location (i, j). We refer to the 2D array of the determinant values at all the (i, j) locations at the given scale as the determinant image corresponding to the original image for that scale. The use of box filters not only provides a very good approximation of the determinant of Hessian matrices, but it also allows very fast filtering by using integral images. This process of filtering is repeated by using box filters of larger filter sizes (larger

scales) in order to form an octave. Then, using the determinant images related to the three consecutive scales in this octave, 3 x 3 x 3 search windows are formed. If the central pixel in the window has a maximum value among all the pixels in the window, then there exists a blob feature at the corresponding scale and the corresponding spatial location of the original image. The spatial location and scale of each blob feature are then refined using the interpolation method proposed in [37]. The above process is repeated using larger size filters (larger scales) in the succeeding octaves. After detecting all the blob features, the orientation of each of the detected blob features is computed by forming a circular neighborhood of radius 6σ (σ being the scale of the detected blob feature) around the detected feature and obtaining the Haar wavelet responses in the x and y directions for the samples within the neighborhood. After detecting the blob features and determining their orientations, a feature descriptor is constructed for each detected blob feature. For this purpose, a square region of size 20σ, centered at a location of the detected blob, is formed and rotated to align it with the orientation of the blob. This square region is divided into 4 x 4 square sub-regions, and Haar wavelet responses in the x and y directions are computed for the samples in each of these sixteen sub-regions. The summation of the responses in the x and y directions and that of their absolute values are computed for each of the sub-regions, which are then put together as a feature descriptor of size 64 for the detected blob. Finally, the matching of the detected features using their descriptors is carried out in the same way as done in the FDEM scheme of SIFT.

ORB [20] is an FDEM scheme that detects, extracts, and matches corner-type features. However, in this method, FAST [38], [39] is used for detecting the corner features, and

BRIEF [40] for constructing the descriptors of the detected features in such a way that the descriptors are orientation invariant. If the intensity of a pixel is significantly greater than or less than the intensities of a specific number of contiguous pixels of a 16-pixel circle formed around that pixel, FAST marks this central pixel as a potential corner feature. FAST then develops a non-maximum suppression technique and applies it to remove the non-distinct feature points from the set of potential features. ORB employs a scale pyramid of the image and detects FAST features at each level in the pyramid. For each of the FAST detected corner features, ORB determines its orientation as the angle of the vector formed from the corner's center and the centroid of all the intensity values within a circular region around the corner. For constructing the feature description vector for each of the detected features, ORB employs the technique of BRIEF [40]. BRIEF forms a binary feature description vector of a given size for each of the features. In ORB, rectangular patches around all the detected features are formed, and 256 pairs of pixels are chosen from all the pixels in a patch, using a predetermined pattern. Each pair’s location is rotated by an angle, which is equal to the orientation of the respective feature point. Each entry in the feature vector corresponds to a specific pair of pixels in the rectangular patch. Therefore, the size of each description vector becomes 256. The value of the entry corresponding to a pair in the description vector of a feature is set to 0 or 1 depending on whether the pixel value of the first pixel is smaller or larger than that of the second one in the pair. Although the scale invariance property of feature descriptors has not been adequately addressed by ORB, since, as mentioned earlier, this method employs a scale pyramid of the image and detects FAST features at each level in the pyramid, ORB is able to produce feature descriptors that are scale invariant to some extent. Just as in the case of SIFT and SURF, in the case of

ORB as well, after extracting the features of the moving image, they are matched with those of the fixed image. However, the Hamming distance instead of Euclidean distance is used to take computational advantage furnished by the binary nature of the feature vectors extracted by ORB.

SIFT, SURF and ORB offer distinct advantages and disadvantages, especially in their feature detection, and extraction parts. For feature detection, SIFT utilizes Gaussian filters and image pyramids for robust scale-invariant feature detection, but it is a computationally expensive detection method. SURF achieves a good balance between speed and accuracy by employing integral images for efficient Hessian matrix calculations in its detection part. On the other hand, ORB employs a scale pyramid of the image and is powered by FAST-based corner detection at each level in the pyramid, which improves its detection speed a lot, at the cost of making it less scale-invariance compared to the detection of SIFT and SURF. In feature extraction, SIFT computes the gradients at sample points around the detected blobs to form orientation histograms, leading to highly distinctive descriptors. On the other hand, for feature extraction, SURF employs integral images, to compute the Haar responses in the x and y directions for the sample points within the neighborhood of the detected blobs, resulting in a fast but less distinctive descriptors compared to that obtained by using SIFT. The feature extraction part of ORB is much faster than those of SIFT and SURF in view of its use of BREIF for this purpose. To summarize the feature extraction parts of the three FDEM schemes, the descriptors resulting from any of these schemes are orientation invariant, whereas the descriptors resulting from SIFT and SURF are scale-invariant but those resulting from ORB are not so to the same extent. Finally, for feature matching, SIFT and SURF have similar speeds provided they both use the same length

vectors for the feature descriptors; however, the former provides a higher matching accuracy in view of its more accurate feature descriptors. On the other hand, the speed of feature matching of ORB benefits from its binary feature descriptors. However, the matching time in ORB could still be quite substantial if the number of detected features in the pairs of images is large, which is generally the case in ORB in comparison to that in SIFT and SURF.

2.3 Related Work on FMR Schemes

As mentioned in Chapter 1, the existing FMR schemes can be classified into two categories, which are transformation-based (TB) and neighborhood structure-based (NSB), respectively. In the following, we provide a brief overview of the existing FMR schemes in both the categories.

2.3.1 Transformation-based FMR Schemes

In [29], an FMR scheme is introduced to identify inlier matches, from the set of putative matches obtained using SIFT, which are then utilized to register the moving image to the fixed image in the presence of nonrigid deformation. For the purpose of feature matching refinement, RANSAC [41] is employed in an iterative manner to estimate the optimal affine transformation model. Matches that do not conform to the estimated optimal affine transformation are considered as mismatches and eliminated. When dealing with nonrigid deformation, imposing a single global geometric constraint on the matches cannot result in a good transformation model. In such a case, it becomes essential to consider the locality

of the matched features when estimating the transformation model. The HMA refinement scheme of [30], which is specifically designed for MIS images, has attempted to overcome the problem of using a single model to refine the matching of the features between the fixed and moving images by using multiple affine models that are estimated in a hierarchical manner to carry out the refinement of the matched features. The authors have shown that the use of the spatially distributed multiple affine models in the HMA scheme can effectively eliminate those pairs of features that have been falsely considered as matched. Even though HMA is an accurate robust scheme for the refinement of the matched features, the time consumed is large and especially so, when the number of putative matched features is large.

In [31], a novel two-part method for matching refinement is developed. In the first part, a scheme called R1P-RNSC is developed within the RANSAC framework to obtain a number of rigid transformation models, each applicable only to a specific subset of all the matches. In the second part, a scheme called EMDQ algorithm has been proposed to generate iteratively smoother deformation fields. Corresponding to each match, a modified transformation is obtained by performing an interpolation among all the transformations of all the points obtained in the previous iteration in an expectation maximization framework, starting from the original transformations of all the points obtained using the R1P-RNSC scheme. EMDQ demonstrates high accuracy and robustness to the inliers ratio. The authors of EMDQ have mentioned that for the purpose of the initialization of transformations, if instead of using R1P-RNSC, one uses algorithms that do not consider rotation then their algorithm can be expedited considerably. But in such a case, the robustness of EMDQ is reduced. However, EMDQ in view of using R1P-RNSC is very slow when the inliers ratio

is very small. In this scenario, R1P-RNSC would require a large number of iterations to find valid rigid transformations.

VFC [32] is an FMR scheme that aims to fit a vector field f, which interpolates the putative matches, thus enabling the inliers to be distinguished from the outliers. The authors formulate the problem of estimating f as an optimization problem with a Tikhonov [42] regularization term. However, in view of the presence of outliers in the putative matches, they adopt the maximum a posteriori (MAP) approach to consider the problem as a mixture model and then employ expectation maximization (EM) algorithm to obtain the vector field f. Authors have pointed out that the effectiveness of VFC diminishes in scenarios where there is a small number of inliers, as is the case in MIS images. In such scenarios, the VFC is no longer robust against inliers ratio and it fails to accurately identify the inliers.

2.3.2 Neighborhood Structure-based FMR Schemes

The authors of the locality preserving matching (LPM) scheme [33] propose a mathematical model formulating it as an optimization problem that incorporates the idea of topology preservation to identify true matches. The cost function consists of two parts. A closed form solution to the optimization problem is obtained by minimizing the first part of the cost function through the consensus of the neighborhood elements and minimizing the second part through the consensus of the neighborhood topology. In this scheme, neighborhoods of a specific feature are obtained by using a multiple K-NN strategy and varying the values of K. Matches with costs below a specified threshold are identified as

inliers. LPM provides good performance with a computational time that is acceptable but not very small in view of using multiple neighborhoods. However, when the inliers ratio in the set of putative matches is small, the number of inliers cannot be expected to be sufficient in those neighborhoods that satisfy the locality condition. As a consequence, the difference between the cost values associated with the inliers and outliers is not very significant, and hence the choice of a threshold value to make a clear distinction between an inlier and outlier becomes extremely difficult.

Learning for mismatch removal (LMR) [34] is another scheme in this category in which a learning-based model is presented for identifying the inliers. This scheme employs a backpropagation neural network, which is trained using a set of match representations and the corresponding set of ground truth labels. The representation of each matched pair is a vector consisting of triplets, each corresponding to one of the multiple K-NN neighborhoods. The first item in a triplet represents the consensus of the neighborhood elements and the other two items represent the consensus of the neighborhood topology. The trained network is then used to classify each match as an inlier or outlier using the representation of the matched pair as the input of the network. When the inliers ratio in the set of putative matches is high, the number of inliers present in most of the K-NN neighborhoods corresponding to a given matched pair is large enough resulting in large values of the elements in the pair's representation if it is an inlier. Consequently, the difference between the match representation vectors for inliers and outliers is significantly high, indicating high between-class (inter-class) variances. However, if the inliers ratio is very low, the number of inliers in each K-NN is expected to be sufficient only for a few neighborhoods of the matched pair resulting in small values for most of the elements in the

pair's representation if it is an inlier; this decreases the between-class variances, thus greatly reducing the performance of LMR. It is also noted that adopting the scheme to a large number of K-NN neighborhoods results in a considerable increase in the processing time.

The RFM-SCAN scheme of [35] is another neighborhood structure-based FMR scheme that customizes the DBSCAN clustering algorithm [43] to cluster putative matches, based on their displacement vectors into motion-consistent clusters. Outliers are identified as matches that do not belong to any of the clusters. RFM-SCAN forms a data point for each matched pair which consists of the spatial locations of the two feature points in the pair along with the displacement vector associated with the pair. Then, the algorithm adaptively determines the two parameters, the neighborhood radius and the minimum cluster size, required by DBSCAN, which clusters the data points to yield a set of putative inliers. Finally, the algorithm updates the values of the two parameters using this putative inlier set and reruns the DBSCAN to obtain the final set of inliers. In RFM-SCAN the neighborhood radius given by an expression, which is defined as a function of the minimum and maximum values of the K-th nearest neighbor (referred to as K-Dist) for each data point and a coefficient μ, with the optimal values of K and μ determined empirically. RFM-SCAN exhibits good performance with a fast computational time. However, when the inliers ratio is very low, there is a higher likelihood for the K-th nearest neighbor for a data point to be an outlier, leading to an increased K-Dist value for that data point. Consequently, this leads to an increased maximum K-Dist value, and therefore, to a larger neighborhood radius to be used by DBSCAN. This, in turn, significantly diminishes the performance of the RFM-SCAN algorithm, since in this case,

the locality of the neighborhoods may be violated.

In [36], an iterative FMR scheme called guided local outlier factor (GLOF) is proposed. It utilizes three types of multiple K-NN. For each matched pair (P_i, P_i') the three types of K-NN neighborhoods are formed with respect to the feature point P_i, the feature point P_i', and the displacement vector associated with the pair, respectively. In each iteration, a score referred to as the local correspondence score ($\widehat{LCS}$), representing the degree of preservation of the neighborhood correspondence, is calculated for each matched pair. At the end of the iterative process, the average value of $\widehat{LCS}$ is obtained for each pair of matched features. A threshold is then applied to classify the pair as an inlier or an outlier. GLOF has been demonstrated to provide good performance. However, when the inliers ratio is very low, the calculated local correspondence scores $\widehat{LCS}$, in most iterations may not be reliable due to an insufficient number of inliers in the K-NN neighborhoods. In such a case, the performance of the algorithm is considerably affected. It is also worth noting that GLOF requires computing three types of K-NN neighborhoods in each iteration, thus resulting in an increased processing time.

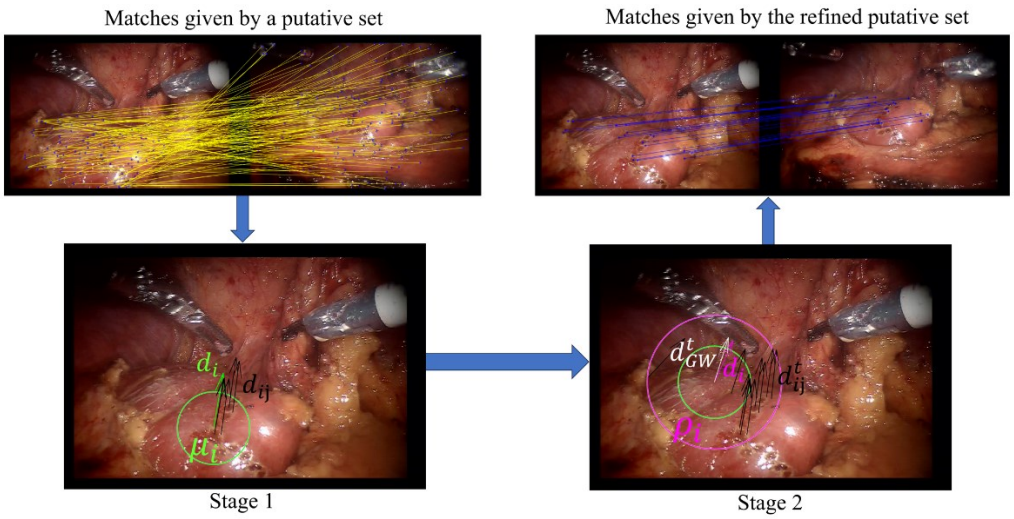
2.4 Summary

In this chapter, we provided a comprehensive review of the state-of-the-art schemes for feature detection, extraction, and matching (FDEM), as well as feature matching refinement (FMR). We first explored in-depth three well-known FDEM schemes, SIFT, SURF, and ORB. We compared their performance and computational complexity in detecting, extracting, and matching features. We then reviewed state-of-the-art FMR

schemes, categorizing them into transformation-based FMR (TB FMR) schemes and neighborhood-structure-based FMR (NSB FMR) schemes. For each scheme, we provided a thorough analysis, highlighting their strengths and limitations in refining putative sets of matched features, with a particular focus on low inlier ratios, which is a common challenge in MIS images. The insights gained from this review are used in developing the proposed FMR and FDEM schemes presented in Chapters 3 and 4.

Chapter 3

A Very Fast and Robust Method for Refinement of Putative Matches of Features in MIS Images



3.1 Introduction

In this chapter, we propose a novel two-stage neighborhood structure-based feature matching refinement (NSB FMR) scheme to remove falsely matched pairs in a given putative set [44]. For the first stage, we follow a conservative approach for choosing small circular neighborhoods of the same size in the fixed image for each feature of this image that belongs to the putative set. This approach of forming neighborhoods enables them to

be better local neighborhoods than those formed by using the K-NN based approach [33]-[36], reviewed in Chapter 2. Then, a voting mechanism is devised for the matches within a neighborhood based on the number and similarity of the displacement vectors of the matches within the neighborhood. Those matches that have displacement vectors similar to those of their neighboring matches get a larger number of votes. After the voting process, matches that receive a large number of votes are identified as true matches in the first stage. Using the knowledge of true matches gained in the first stage, a mechanism is developed in the second stage to determine the status of those matches in the putative set whose status have not yet been determined in the first stage. In the second stage, larger neighborhoods of the same size are formed around each feature point in the fixed image corresponding to the pairs whose status is still unknown. Then, a Gaussian-weighted average of all the displacement vectors of the feature points in the larger neighborhood, whose statuses are known to be true from the first stage, is computed. If this Gaussian-weighted vector is similar to the displacement vector of the matched feature point in question, the status of this matched feature is changed to be true; otherwise, its status is recorded as false. A number of experiments is performed on the proposed FMR scheme using datasets involving real, synthetic, and phantom MIS images, so as to demonstrate the effectiveness of the proposed FMR scheme under various challenging scenarios such as camera occlusion, camera retraction and reinsertion, sudden camera motion, specular reflections, different numbers of matches and inliers ratios, and stereo vision environment. The proposed scheme is also compared with a number of state-of-the-art FMR schemes belonging to the transformation-based and neighborhood structure-based categories.

3.2 Proposed Feature Matching Refinement Scheme

Our objective in developing an FMR scheme, given the putative set of matches, $\{(P_i, P_i'), i = 1, \dots, N\}$, is to produce a one dimensional array, $\boldsymbol{L} = [l_1, l_2, \dots, l_N]$, of labels in which the label in the i^{th} entry indicates whether the match (P_i, P_i') is *True* or *False* indicated by the symbols T and F, respectively. The entries in this vector are initialized with a label U denoting that the status of the match (P_i, P_i') at a given time is *Unknown*.

The proposed FMR scheme is a novel two-stage FMR. In stage 1, for a feature point P_i, $i = 1, \dots, N$, in the fixed image, we form a circular neighborhood μ_i of radius R_1 centered at P_i. By following a voting process, some of the matches whose displacement vectors are consistent with (similar to) the displacement vectors of their neighboring features are labeled as *True* matches. However, there are other matches in the set $\{(P_i, P_i'), i = 1, \dots, N\}$ whose status still remain *Unknown*. In stage 2 of our FMR algorithm, we revisit all the matched feature points whose status after stage 1 still remain *Unknown*. We examine the similarity of the displacement vector of each such feature point with a Gaussian weighted average of the displacement vectors of the neighboring feature points whose status was established to be *True* in stage 1 of the proposed FMR in a neighborhood larger than that given by the circle of radius R_1. Based on the result of this examination, the *Unknown* status of some of the matches may get changed to be *True*. These two stages are explained in detail in the following subsections.

3.2.1 FMR Algorithm – Stage 1

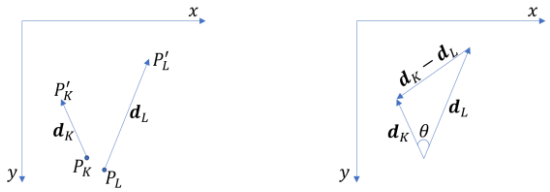
Let P_K and P_L be the locations of any two features, say, the K^{th} and L^{th} features, which are

in close proximity with each other in the fixed image. Let P'_K and P'_L be the locations of the corresponding matched features in the moving image, that is, these are the new locations of the two features resulting from the deformation of the tissue. Therefore, we have $\boldsymbol{d}_K = P'_K - P_K$ and $\boldsymbol{d}_L = P'_L - P_L$ as the displacement vectors of the K^{th} and L^{th} features, respectively, as shown in Figure 3.1(a). Now, let us construct a difference vector $\boldsymbol{d}_K - \boldsymbol{d}_L$ using these two displacement vectors, as shown in Figure 3.1(b). It is seen that, as $|\boldsymbol{d}_K - \boldsymbol{d}_L| \to 0$, $|\boldsymbol{d}_L| \to |\boldsymbol{d}_K|$ and the angle θ between the two displacement vectors tends to zero ($|\cdot|$ denotes the magnitude of the vector). In other words, as the magnitude of the difference vector between the two displacement vectors becomes smaller and smaller, the two vectors $\boldsymbol{d}_K$ and $\boldsymbol{d}_L$ become increasingly more similar. Therefore, we can choose a threshold value, D_{th}, to determine the similarity between the displacement vectors of two feature points. In other words, if the following condition

$$|\boldsymbol{d}_K - \boldsymbol{d}_L| \leq D_{th} \tag{3.1}$$

is satisfied, then we can consider the displacement vectors $\boldsymbol{d}_K$ and $\boldsymbol{d}_L$ to be similar.

In stage 1 of the proposed scheme, using all the feature points $P_i,\ i = 1, \ldots, N$, in the fixed image, we form a k-d tree [45]. We also form a circular neighborhood μ_i of radius R_1 centered at P_i for each of the feature points. Then, this k-d tree is searched to find n_i feature points, P_{i1}, P_{i2}, …, P_{in_i}, within the neighborhood μ_i, iteratively for all the feature points.



(a) (b)

Figure 3.1: Difference vector between the displacement vectors of two feature points. (a) Displacement vectors of K^{th} and L^{th} features. (b) Difference vector between the two displacement vectors $\boldsymbol{d}_K$ and $\boldsymbol{d}_L$.

If the number n_i of neighboring feature points of P_i is less than a threshold n_{th}, we consider this feature point P_i to be an isolated feature point, that is, this feature point does not have sufficient information to determine the similarity of its displacement vector with other displacement vectors in the neighborhood. Therefore, for the present, we refrain from making a decision on the correctness of the matching of (P_i, P_i') of the i^{th} feature point. On the other hand, if $n_i \geq n_{th}$, we form a set of matches $S_{\mu_i} = \{(P_{i1}, P'_{i1}), (P_{i2}, P'_{i2}), \ \ldots, (P_{in_i}, P'_{in_i})\}$ of all the feature points within the neighborhood μ_i, where P'_{ij} is the match of the features P_{ij}, $j = 1, \ldots, n_i$. We next compute the set of displacement vectors for all the features in the neighborhood μ_i, $S_{d\mu_i} = \{\boldsymbol{d}_{i1} = P'_{i1} - P_{i1}, \boldsymbol{d}_{i2} = P'_{i2} - P_{i2}, \ \ldots, \boldsymbol{d}_{in_i} = P'_{in_i} - P_{in_i}\}$, as well as the displacement vector $\boldsymbol{d}_i = P'_i - P_i$ for the feature point P_i. Now, a set $S_{D\mu_i} = \{|\boldsymbol{d}_i - \boldsymbol{d}_{i1}|, |\boldsymbol{d}_i - \boldsymbol{d}_{i2}|, \ \ldots, |\boldsymbol{d}_i - \boldsymbol{d}_{in_i}|\}$ of the magnitudes of the differences between the displacement vector $\boldsymbol{d}_i$ and each of the displacement vectors $\boldsymbol{d}_{ij}, j = 1, \ldots, n_i$, is computed. Next, on each of the items in the set, S_{μ_i}, the condition given by (3.1) is applied to determine the similarity of each of the displacement vectors of the feature points in μ_i with that of the feature point P_i. We form a

subset, $S_{\mu_i}^{s} = \{P_{si1}, P_{si2}, \ \ldots, P_{sin_{si}}\}, n_{si} \leq n_i$, of the feature points in μ_i for which the magnitudes of the corresponding difference vectors in $S_{D\mu_i}$ satisfy the condition given by (3.1). According to our assumption, the number n_{si} of similar displacement vectors of the feature points in the neighborhood μ_i needs to have at least a certain minimum value n_{th}^{s}. Let us now assume that we have a vote vector $\boldsymbol{V} = \{v_1, \ v_2, \ldots, v_N\}$, with all its entries initialized to zero. Then, the value of v_i in $\boldsymbol{V}$, corresponding to the feature point i, is incremented by 2, and all of the entries in $\boldsymbol{V}$ corresponding to the feature points in $S_{\mu_i}^{s}$ are incremented by unity, that is,

$$v_i = \ v_i + 2 \tag{3.2}$$

and

$$v_j = \ v_j + 1, \quad j = si1, si2, \ldots, sin_{si} \tag{3.3}$$

This process is repeated for all the $P_i's, \ i = 1, \ldots, N$. It is to be noted that receiving a higher vote by the match (P_i, P_i') for the i^{th} feature is an indication that not only the displacement vectors of the features in the neighborhood of P_i are similar, but the displacement vectors of the features within the neighboring neighborhoods are also similar. We now define a threshold n_{th}^{v} for the minimum number of votes to be received by the match (P_i, P_i') for it to be considered a $True$ match:

$$l_i \ = T, \ if \ v_i \geq \ n_{th}^{v}, \qquad i = 1, \ldots, N \tag{3.4}$$

At this point, it is important to analyze as to how the size of the neighborhood (that is the value chosen for R_1) has an impact on the accuracy of those matches on which a decision of being *True* (*T*) has been taken. By choosing a small neighborhood via selecting a small value for R_1, we are sure that only the features that are in close proximity to P_i, if

they satisfy condition (3.1), may help the match (P_i, P_i') as well as the matches of the features within μ_i in making them to receive a greater vote of confidence towards their being *True* matches. In this case, at the end of the voting process, we are more confident in labeling those matches that satisfy (3.4) as *True* matches. However, there are other matches in the set $\{(P_i, P_i'),\ i = 1, \dots, N\}$ whose status still remain *Unknown*. At this stage, there are two cases that cause a match to remain *Unknown*. First, if the number of features in the neighborhood μ_i of P_i is less than the threshold n_{th}, which means that P_i is an isolated feature point, we do not make a decision at this point as to whether the match (P_i, P_i') is *True* or *False*. Second, even if P_i is not an isolated feature point, but the number of its votes, v_i, is smaller than n_{th}^v, then also we leave the status of the match as *Unknown*. Therefore, our objective from this point on is to determine whether each of the remaining matches whose status is still *Unknown* is *True* or *False*. As the size of the neighborhood is increased by increasing the value of R_1, we may be able to include in the larger neighborhood more matches that are *True*, but at the risk of also including some of the matches that are really *False,* even if they also satisfy the similarity condition. Therefore, in view of labeling such *False* matches as *True* matches by increasing the value of R_1 for the neighborhood μ_i and simply following the same procedure as we have been following so far is not a viable solution to change the status of a match from *Unknown* to *True*.

3.2.2 FMR Algorithm – Stage 2

We now explain stage 2 of the proposed FMR scheme. For each feature point P_i , $i = 1, \dots, N$, in the fixed image, if $l_i = U$, then using its xy coordinates, we find all the feature

points P_{i1}, P_{i2}, …, P_{iv_i} in the neighborhood ρ_i consisting of a circle of radius $R_2 > R_1$, and centered at P_i employing a k-d tree data structure. Then, we form a set $S_{\rho_i}^t = \{(Q_{i1}, Q'_{i1}), (Q_{i2}, Q'_{i2}), \dots, (Q_{in_{ti}}, Q'_{in_{ti}})\}, n_{ti} \leq v_i$, where Q_{ij} is the feature point in ρ_i and Q'_{ij} is the corresponding feature point in the moving image such that (Q_{ij}, Q'_{ij}) has a label $True$. If the number of *True* matches n_{ti} is larger than or equal to a prespecified threshold n_{th}^t, we compute Gaussian weight for each of the feature points $Q_{ij}, j = 1, \dots, n_{ti}$, using its Euclidean distance from the feature point P_i:

$$w_{ij} = \exp\left(-\frac{|P_i - Q_{ij}|^2}{2\sigma^2}\right), j = 1, \dots, n_{ti} \tag{3.5}$$

where σ^2 is a parameter specifying the variance of the Euclidean distance of the feature point Q_{ij} within the neighborhood ρ_i from its central feature point P_i. We now form a set of displacement vectors of the *True* matches of the features in the neighborhood ρ_i, as $S_{d\rho_i}^t = \{\boldsymbol{d}_{i1}^t = Q'_{i1} - Q_{i1}, \boldsymbol{d}_{i2}^t = Q'_{i2} - Q_{i2}, \dots, \boldsymbol{d}_{in_{ti}}^t = Q'_{in_{ti}} - Q_{in_{ti}}\}$. Using the Gaussian weights, w_{ij}, given by (3.5) and the set $S_{d\rho_i}^t$, we next compute a Gaussian weighted average vector of all the displacement vectors in the set $S_{d\rho_i}^t$

$$\boldsymbol{d}_{GW}^t = \frac{\sum_{j=1}^{n_{ti}} w_{ij} \boldsymbol{d}_{ij}^t}{\sum_{j=1}^{n_{ti}} w_{ij}} \tag{3.6}$$

as well as the displacement vector for the feature point P_i, $\boldsymbol{d}_i = P'_i - P_i$. The label of the match (P_i, P'_i) is changed from *Unknown* to *True*, if the similarity between $\boldsymbol{d}_i$ and $\boldsymbol{d}_{GW}^t$ is satisfied using the condition given by (3.1).

The entire process described above is repeated for all i's ($i = 1, \dots, N$). It is to be noted that at the beginning of stage 2, a temporary array $\boldsymbol{K} = [k_1, k_2, \dots, k_N]$, is initialized as a

copy of array ***L***. During the process in stage 2, which depends on array ***L***, any newly identified *True* labels are stored in array ***K*** while ***L*** remains unchanged. At the end of stage 2, array ***K*** is copied to array ***L***. Finally, all the entries in the array ***L*** are examined and those whose values are still *Unknown* are changed to be *False*.

The proposed FMR scheme consisting of the two stages is henceforth referred to as the voting based on similarity of local displacement vectors FMR (VSLD-FMR) scheme. This scheme is summarized as Algorithm 1.

Algorithm 1: VSLD-FMR Algorithm

Stage 1

Input: The pairs of the locations of all the matched feature points $(P_i,\ P_i'),\ i = 1,\ \ldots,\ N$.

Initialization:

$l_i = U, v_i = 0,$ for $i = 1, \ldots, N$.

for $i = 1{:}N$ **do**

Find all the feature points P_{i1}, P_{i2}, …, P_{in_i} in the neighborhood μ_i of the fixed image consisting of a circle of radius R_1 and centered at P_i employing a k-d tree data structure.

if $n_i \geq n_{th}$ **then**

Form the set $S_{\mu_i} = \{(P_{i1},\ P_{i1}'), (P_{i2},\ P_{i2}')\ \ldots, (P_{in_i},\ P_{in_i}')\}$.

Compute the displacement vector $\boldsymbol{d}_i = P_i' - P_i$.

Compute the set $S_{d\mu_i} = \{\boldsymbol{d}_{i1} = P_{i1}' - P_{i1}, \boldsymbol{d}_{i2} = P_{i2}' - P_{i2}\ \ldots, \boldsymbol{d}_{in_i} = P_{in_i}' - P_{in_i}\}$.

Compute the set $S_{D\mu_i} = \{|\boldsymbol{d}_i - \boldsymbol{d}_{i1}|, |\boldsymbol{d}_i - \boldsymbol{d}_{i2}|,\ \ldots, |\boldsymbol{d}_i - \boldsymbol{d}_{in_i}|\}$.

Form a subset, $S_{\mu_i}^s = \{P_{si1}, P_{si2},\ \ldots, P_{sin_{si}}\}, n_{si} \leq n_i$, of the feature points in μ_i for which the magnitudes of the corresponding difference vectors in $S_{D\mu_i}$ satisfy the condition given by (3.1) with a threshold of D_{th}.

if $n_{si} \geq n_{th}^s$ **then** $v_i = v_i + 2$ and $v_j = v_j + 1,\ j = si1, si2, \ldots, sin_{si}$.

end if

end if

end for

for $i = 1{:}N$ **do**

if $v_i \geq n_{th}^v$ **then** $l_i = T$

end if

end for

Stage 2

$k_i = l_i$, for $i = 1, \ldots, N$.

for $i = 1{:}N$ **do**

if $l_i = U$ **then**

Find all the feature points P_{i1}, P_{i2}, …, P_{iv_i} in the neighborhood ρ_i of the fixed image consisting of a circle of radius $R_2 > R_1$, and centered at P_i employing a k-d tree data structure.

Form a set $S_{\rho_i}^t = \{(Q_{i1},\ Q_{i1}'),\ (Q_{i2},\ Q_{i2}'), \ldots, (Q_{in_{ti}},\ Q_{in_{ti}}')\}, n_{ti} \leq v_i$,of all the matches of the feature points in the neighborhood ρ_i, whose labels are $True$.

if $n_{ti} \geq n_{th}^t$ **then**

Compute the Gaussian weights, w_{ij}, for each of the feature point $Q_{ij}, j = 1, \ldots, n_{ti}$, using its Euclidean distance from the feature point P_i

$$w_{ij} = \exp\left(-\frac{|P_i - Q_{ij}|^2}{2\sigma^2}\right), j = 1, \ldots, n_{ti}$$

Form a set of displacement vectors of the *True* matches in the set $S_{\rho_i}^t$

$$S_{d\rho_i}^t = \{\boldsymbol{d}_{i1}^t = Q_{i1}' - Q_{i1},\ \boldsymbol{d}_{i2}^t = Q_{i2}' - Q_{i2},\ \ldots,\ \boldsymbol{d}_{in_{ti}}^t = Q_{in_{ti}}' - Q_{in_{ti}}\}$$

Compute a Gaussian weighted average vector of all the displacement vectors in the set $S_{d\rho_i}^t$

$$\boldsymbol{d}_{GW}^t = \frac{\sum_{j=1}^{n_{ti}} w_{ij} \boldsymbol{d}_{ij}^t}{\sum_{j=1}^{n_{ti}} w_{ij}}$$

Compute the displacement vector $\boldsymbol{d}_i = P_i' - P_i$.

if $|\boldsymbol{d}_i - \boldsymbol{d}_{GW}^t| \leq D_{th}$ **then** $k_i = T$

end if

end if

end if

end for

$l_i = k_i$, for $i = 1, \ldots, N$.

if $l_i = U$ **then** $l_i = F,$ for $i = 1, \ldots, N$.

3.3 Experimental Results

In this section, we compare the performance and the processing time of our proposed feature matching refinement scheme (VSLD-FMR) with that of nine other state-of-the-art schemes, RANSAC-affine [29], HMA [30], fast VFC (VFC) [32], sparse VFC (sVFC) [32], LMR [34], LPM [33], RFM [35], GLOF [36], and EMDQ [31]. The proposed FMR scheme is implemented in MATLAB and executed on a computer with an AMD Ryzen 7 3800X 8-Core 3.89 GHz processor. The MATLAB source code of the proposed VSLD-FMR scheme is publicly available at https://github.com/Pourshahabi/VSLD-FMR. The MATLAB source codes of the nine schemes with which we compare our scheme are publicly available, and hence, we run them on the same hardware platform as we do our own scheme. It is important to note that for a fair comparison, we evaluate all methods within the same software and hardware platforms. This approach ensures that any performance differences can be attributed to the methods themselves rather than the programming language or hardware platform used. Although the methods we have compared with, as well as our own method, are implemented in MATLAB, it is worth mentioning that by implementing these algorithms in C or C++ or other low-level languages, one could achieve higher speeds.

Three datasets representing various challenging situations are used to test the effectiveness of the proposed FMR scheme. These datasets are the laparoscopic image dataset [46], a synthetic image dataset created by us using the laparoscopic image dataset, and the heart phantom image dataset [47]. The laparoscopic image dataset is a publicly available dataset acquired from real nephrectomy interventions. The dataset contains

images encompassing various challenging scenarios, including instances of camera occlusion, camera retraction and reinsertion, sudden camera motion, and specular reflections. This dataset includes 100 color laparoscopic-surgery image pairs with the resolution of 704 × 480 acquired from six real videos of partial nephrectomy interventions. The authors of [30] have applied SIFT [2] to each pair of images in this dataset, to obtain a set of putative matches for the pair. The matches in each set of putative matches were then manually labeled as 'correct' or 'incorrect' by four experts, forming ground-truth matching data. By selecting and matching a certain number of corners in each pair of the images in this dataset, a set of ground-truth mapping data was also formed for the pair. These two ground-truth data are also available on the site of the laparoscopic image dataset [46].

The second dataset is a synthetic image dataset constructed by us using the 100 pairs of images of the laparoscopic image dataset and the corresponding ground-truth mapping data [46]. We refer to this dataset as the synthetic-laparoscopic image dataset I*. To construct this dataset, we first select the mapping data corresponding to the first pair of the images in the laparoscopic image dataset, and obtain a transformation function using the local weighted mean (LWM) method [48] that maps the points in the mapping points in the fixed image to the corresponding mapping points in the moving image. This transformation is then applied to all the 100 fixed images in the laparoscopic dataset, one by one, resulting in 100 synthetic moving images. This process of obtaining a transformation and applying it to obtain synthetic images is repeated for all the 100 images. Thus, the set of synthetic-laparoscopic image dataset I has a total of 10,000 pairs of images, in which all the 10,000

* In Chapter 4, we will introduce another synthetic-laparoscopic image dataset, which will be referred to as synthetic-laparoscopic image dataset II.

moving images are distinct whereas the set of 10,000 fixed images has only 100 distinct images, which are actually the original 100 fixed images in the laparoscopic image dataset. The reason for using LWM transformation for obtaining synthetic moving images from the fixed images is in view of its ability of introducing local deformation in the former images with respect to the latter ones. In order to introduce other types of differences between a fixed image and a corresponding synthetic moving image, we corrupt randomly each of the 10,000 moving images by one of the six cases of corruption. Each case of corruption consists of a level of blurriness introduced by a 2D Gaussian smoothing kernel with standard deviation of 1 or $\sqrt{2}$, and a level of Gaussian white noise with a mean of 0 and a variance of 0.01, 0.02, or 0.03. In order to obtain a putative set of matches for each pair in this dataset, we first apply SIFT scheme to the fixed and moving images, and then use the resulting matches to form a putative set. Since we already have a transformation function for mapping each pair of the images in this synthetic dataset, we apply this same transformation to obtain the spatial position in the moving image corresponding to a feature point in the fixed image. The error distance defined as the Euclidean distance between the spatial positions of the feature points in the moving image, obtained by SIFT and the transformation, is computed. Matches with a distance error less than or equal to seven pixels are considered true matches, while those with a distance error exceeding seven pixels are classified as false matches. Availability of a large number of pairs of images with different characteristics in this dataset allows us to perform a study on the effectiveness of the various FMR schemes with respect to different inliers ratios and different numbers of matches in the putative sets.

The third dataset is a publicly available dataset of two videos of a heart phantom subjected to cardiac motion, each video comprising 2,427 frames of images of resolution 360 × 288 [47]. The corresponding frames of the two videos are the images captured by the left and right cameras in a stereo vision setup. The intrinsic, extrinsic, and distortion parameters are provided through the process of stereo camera calibration and available in the dataset. The depth maps (3D locations) of the heart phantom at various phases of the cardiac motion are also obtained by using a CT scanner [14], [49]. The same dataset contains 20 files, each containing a depth map corresponding to a particular phase of the cardiac motion. Other than this information, this dataset does not contain any other information, such as a putative set and the corresponding ground-truth matching data, which is the information necessary to run an FMR scheme. To generate a putative set and the corresponding ground-truth matching data, we first apply the SIFT scheme for the detection, extraction, and matching of the feature points in the left and right images. Next, by utilizing the depth maps and the intrinsic, extrinsic, and distortion parameters, we re-project the associated 3D points of the matched features to the images. Matches with a re-projection error less than or equal to five pixels are considered true matches, while those with a re-projection error exceeding five pixels are classified as false matches.

In order to measure the performance of each of the FMR schemes, we employ the metrics of *Accuracy*, *Precision*, *Recall*, *Specificity* and *F-score*. These metrics help in assessing the correctness of the matches and identifying the strengths and weaknesses of the feature matching refinement scheme. To explain these five metrics, we first need to know the definitions of TP, TN, FP, and FN in the context of FMR. TP (True Positives) indicates the number of matched pairs that are correctly identified as true matches by the

refinement scheme. TN (True Negatives) indicates the number of matched pairs that are correctly identified as false matches by the refinement scheme. FP (False Positives) indicates the number of matched pairs that are incorrectly identified as true matches by the refinement scheme. FN (False Negatives) indicates the number of matched pairs that are incorrectly identified as false matches by the refinement scheme. Below are the definitions of each performance metric in the context of feature matching refinement:

$$Accuracy = \frac{TP + TN}{total\ number\ of\ matched\ pairs} \tag{3.7}$$

$$Precision = \frac{TP}{TP + FP} \tag{3.8}$$

$$Recall = \frac{TP}{TP + FN} \tag{3.9}$$

$$F - Score = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{3.10}$$

$$Specificity = \frac{TN}{TN + FP} \tag{3.11}$$

The proposed FMR algorithm has a number of parameters. We now describe as to how the values of these parameters are set. Assuming a reference resolution of 704×480, for images of size of *n* columns and *m* rows, the scale factor is defined as

$$s = \frac{(n / 704 + m / 480)}{2} \tag{3.12}$$

Now using this scale factor, we select $R_1 = \alpha \times s$ and $R_2 = \gamma \times s$, where α and γ are multiplicative factors, whose values are empirically determined as 70 and 130, respectively. The similarity threshold, D_{th}, is chosen as $D_{th} = s \times 13$. The value of each of the thresholds n_{th}, n_{th}^{s}, and n_{th}^{t} is chosen as 2, while the parameter σ, is empirically

determined to be $\sigma = s \times 14$. The parameter n_{th}^{v} is chosen as $n_{th}^{v} = min\ (6, avg(\boldsymbol{V} \geq 3))$, where $avg(\boldsymbol{V} \geq 3)$ is the average of the values of those entries in the vote vector $\boldsymbol{V}$ that are greater than or equal to three. If there is no such entry in the vector $\boldsymbol{V}$, n_{th}^{v} is set to 6 by default. It needs to be pointed out that once the values of the parameters are determined, they do not need to be re-tuned as new frames arrive during a real-life application of our algorithm. Hence, the speed of our proposed algorithm is not affected because of the use of these hyperparameters. The values of the parameters of the other methods, with which we compare our proposed FMR scheme, are the same as that used in the respective schemes.

In the following, we present the results of the various FMR schemes on each of the three datasets, one by one. The purpose of performing experiments on the three datasets is that the analysis of the results obtained using each dataset focuses on illustrating the effectiveness of the various schemes from different points of view.

3.3.1 Results on the Laparoscopic Image Dataset

We begin with the presentation of experimental results by first giving the results of an ablation study showing the effectiveness of the second stage of the proposed FMR algorithm. Table 3.1 shows the performance results of the complete proposed algorithm in terms of the *Accuracy*, *Precision*, *Recall*, *Specificity*, and *F-score*, and the processing time as well as those of the algorithm from which stage 2 is removed. It is seen from this table that stage 2 has a significant impact on the metrics *Recall* and *F-score* at the expense of 1.1 ms additional processing time. Similar results have been observed in the case of the other two datasets.

Table 3.1: Results of the effectiveness of stage 2 on the performance and the processing time of the proposed algorithm using the laparoscopic image dataset.

Method	Acc.	Prec.	Rec.	Spec.	F-score	Time (ms)
Proposed VSLD-FMR Algorithm	0.97	0.92	0.96	0.97	0.94	2.9
VSLD-FMR (Without Stage 2)	0.96	0.92	0.91	0.97	0.91	1.8

In the evaluation of different FMR schemes on the laparoscopic image dataset [46], the primary objective is to examine the performance of the schemes in a natural setup involving instances of camera occlusion, camera retraction and reinsertion, sudden camera motion, and specular reflections.

Table 3.2 shows, for the various schemes, the values of the five metrics as well as the processing time, averaged over all the pairs of images in the laparoscopic image dataset. It is to be noted that the proposed scheme is the only one that achieves a value of 0.92 or higher for all the five metrics. In order to compare the various schemes in this table uniformly, we adopt a procedure, in which we assign 3, 2, and 1 points for the best, second-best, and third-best values, respectively, for a performance metric. According to this procedure, it is seen that the proposed VSLD-FMR, EMDQ, and HMA schemes achieve 12, 10, and 8 points, respectively. Based on these points, even though these three schemes can be considered as the best, second-best, and third-best performing schemes, respectively, conservatively speaking, one can regard them to have similar performance. Regardless the way, one ranks these three schemes, the proposed scheme, has the best and an extremely low computational time of 2.9 ms, which is 8.89% of that of HMA and 5.94% of that of EMDQ.

Table 3.2: Average values of accuracy, precision, recall, specificity, F-SCORE, and processing time on the images of the laparoscopic dataset for the various FMR schemes.

Method	Acc.	Prec.	Rec.	Spec.	F-score	Time (ms)
RANSAC-affine [29]	*0.94*	0.81	*0.93*	0.92	0.86	30.30
HMA [30]	***0.95***	**0.95**	0.83	**0.99**	0.88	32.60
VFC [32]	***0.95***	0.88	0.92	*0.96*	*0.89*	9.82
sVFC [32]	***0.95***	0.87	0.91	*0.96*	0.88	***4.01***
LMR [34]	0.92	0.80	0.89	0.92	0.84	24.60
LPM [33]	0.93	0.82	0.90	0.93	0.85	7.80
RFM [35]	***0.95***	0.85	***0.96***	0.93	*0.89*	*4.10*
GLOF [36]	0.91	0.78	0.90	0.88	0.82	40.70
EMDQ [31]	**0.97**	*0.89*	**0.98**	*0.96*	***0.93***	48.80
Proposed VSLD-FMR	**0.97**	***0.92***	***0.96***	***0.97***	**0.94**	**2.90**

The best, second best, and the third best results are indicated in bold red, bold italic blue, and italic green fonts, respectively.

Figure 3.2 shows the performance results of the various schemes in terms of all the five metrics as well as the processing time, as a function of the inliers ratio. Regardless of the inliers ratio, the performance of the proposed VSLD-FMR algorithm in terms of the various metrics is generally among the top two schemes, except for Recall where its performance is still generally among the top three schemes. It is especially to be noted that, when the inliers ratio is low (less than or equal to 35%), the performance of VSLD-FMR, EMDQ, and HMA are among the top schemes, but the processing time of the VSLD-FMR is smaller than that of HMA by a factor of eight and at least an order of magnitude smaller than that of EMDQ. Hence, it can be concluded that the proposed scheme is the best among all the feature matching refinement schemes, when both the performance and time complexity are simultaneously taken into consideration.

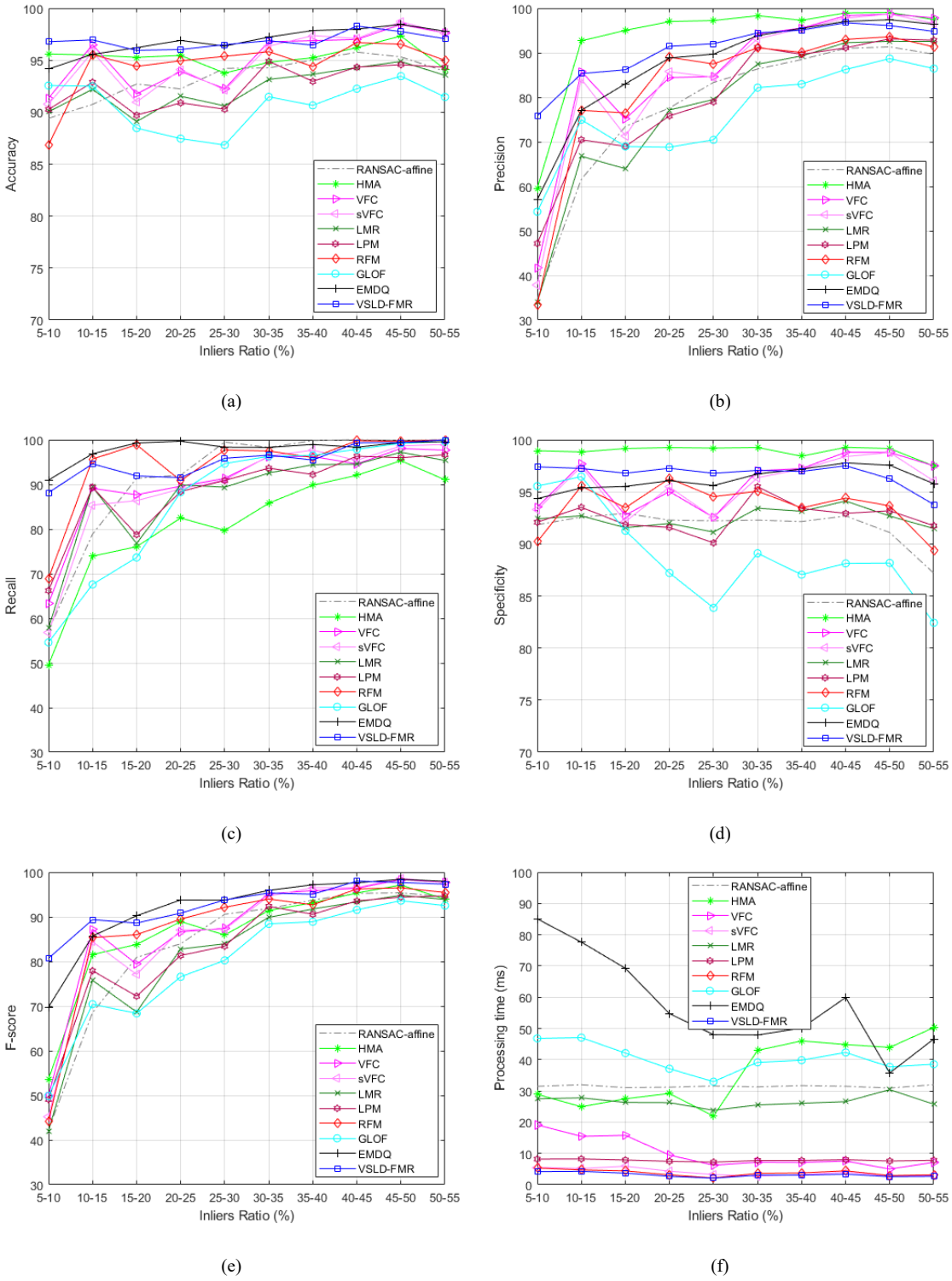


Figure 3.2: Performance and processing time comparison of various FMR schemes on the laparoscopic image dataset. Plots of (a) Accuracy, (b) Precision, (c) Recall, (d) Specificity, (e) F-score, and (f) the processing time as functions of inliers ratio of the various FMR schemes on the laparoscopic image dataset. The average number of matches per pair of images before refinement is about 250.

For visual illustration of feature matching refinement of the proposed scheme, we have selected a pair of images from the laparoscopic image dataset, as shown in Figure 3.3(a). This is a challenging pair of images in which there is a significant deformation between the two images of the pair, and has a remarkably low inliers ratio of 7.63% resulting from the application of SIFT. The total number of matches is 367, in which 339 matches are ground-truth outliers (Figure 3.3(b)) and 28 matches are ground-truth inliers (Figure 3.3(c)). For the purpose of comparison, we have chosen two other FMR schemes, namely, EMDQ and sVFC, of which the former has a very good performance in terms of the five metrics and the latter has a very good processing time. The sVFC scheme (Figure 3.3(d)) identifies 30 matches as inliers, all of which are actually false, whereas the EMDQ FMR scheme (Figure 3.3(e)) identifies 40 matches as inliers, of which 26 matches are indeed true, while 14 matches are actually false. On the other hand, the proposed VSLD-FMR scheme (Figure 3.3(f)) identifies 30 matches as inliers, of which 24 matches are indeed true, while 6 matches are actually false. The performances of the three schemes need to be seen in the light of their processing times. The processing time of the good performing EMDQ scheme is about 22 times of that of the proposed scheme which has an excellent performance. On the other hand, the performance of the sVFC scheme with a processing time of about two times of that of our scheme is very poor.

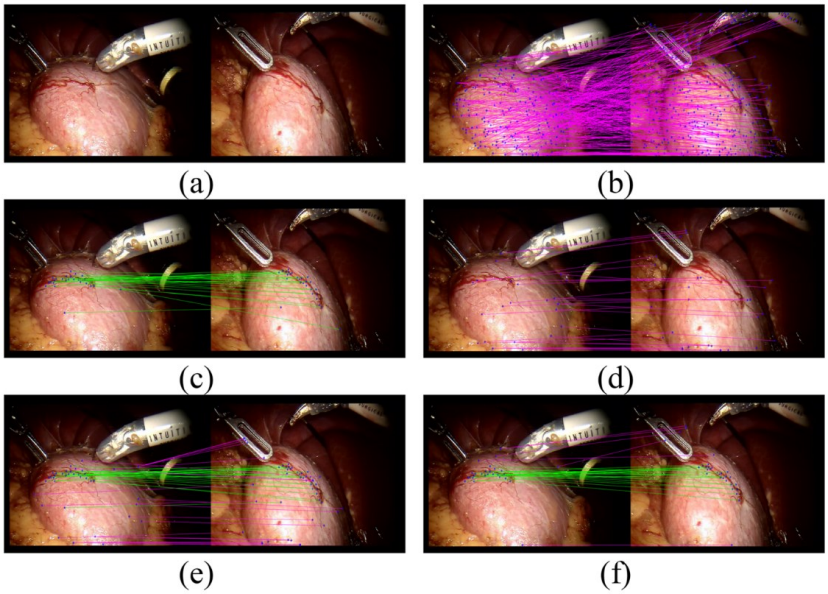


Figure 3.3: Visual illustration of sVFC, EMDQ, and VSLD-FMR schemes on a pair of images from the laparoscopic image dataset. (a) A pair of images from the laparoscopic image dataset. (b) Ground-truth outliers. (c) Ground-truth inliers. Inliers identified by (d) sVFC (processing time: 7.4 ms), (e) EMDQ (processing time: 77.6 ms), and (f) the proposed VSLD-FMR scheme (processing time: 3.6 ms). In (d), (e), and (f), the green lines indicate matches that are correctly identified as true matches, and the lines in magenta color indicate matches that are incorrectly identified as true matches.

3.3.2 Results on the Synthetic-Laparoscopic Image Dataset I

We compare the performance and the time complexity of the various FMR schemes on our synthetic-laparoscopic image dataset I. Even though the differences between the two images of a pair in this dataset are characterized mainly by deformation, blur and noise, the main purpose of using this dataset for evaluating the various FMR schemes is that all the results obtained in our experiments are based on a large set of image pairs with ground-truth that are provided by this dataset. Availability of this large dataset allows us to conduct experiments involving large sets of pairs with specific characteristics such as a given average number of matches per pair, or inliers ratio.

Table 3.3 gives the average performance of the various schemes in terms of the five metrics as well as the processing time for the synthetic-laparoscopic image dataset I. It is seen that for this dataset, the proposed VSLD-FMR and EMDQ schemes are the two top performing schemes. However, the processing time of the former is still about one-fifteenth of that of the latter.

Table 3.3: Average values of accuracy, precision, recall, specificity, F-SCORE, and processing time on the images of the synthetic-laparoscopic image dataset I for the various FMR schemes.

Method	**Acc.**	**Prec.**	**Rec.**	**Spec.**	**F-score**	**Time (ms)**
RANSAC-affine [29]	0.93	0.80	0.88	0.93	0.83	31.40
HMA [30]	*0.96*	**0.99**	0.82	**1.00**	0.89	115.81
VFC [32]	0.88	0.86	***0.98***	0.85	0.88	16.30
sVFC [32]	0.88	0.86	***0.98***	0.85	0.88	***6.64***
LMR [34]	0.95	0.85	0.92	0.95	0.88	30.56
LPM [33]	0.94	0.84	0.91	0.94	0.86	9.17
RFM [35]	0.95	0.86	0.92	0.95	0.89	*7.44*
GLOF [36]	0.91	0.88	0.68	*0.97*	0.76	62.92
EMDQ [31]	**0.98**	*0.93*	**0.99**	***0.98***	**0.96**	90.80
Proposed VSLD-FMR	***0.97***	***0.95***	*0.93*	***0.98***	***0.94***	**5.89**

Figure 3.4 shows the results obtained by applying the proposed FMR scheme (VSLD-FMR) as well as the other schemes on the synthetic-laparoscopic image dataset I. The results of this figure correspond to the case when the average number of matches per pair is about 500. For this purpose, we form a subset of the synthetic-laparoscopic image dataset I in which the numbers of matches range between 460 and 560, so that the average number of matches is approximately 500. This subset contains 2542 such pairs from the original set. It is seen from the plots of this figure that the proposed FMR scheme is the only scheme whose performance in terms of all the metrics, with the exception of *Recall*

values, is always in excess of 80% and often 90% or higher, irrespective of the values of the inliers ratio. Even though the performance of the proposed FMR scheme in terms of *Recall* metric is not the best one, it is still very good by providing *Recall* values that are always greater than 70% and often in excess of 80%. It is to be noted that even three of the schemes, namely, RFM, VFC, and sVFC, which provide *Recall* values better than that provided by our proposed scheme, the performance of the RFM scheme is always lower than that of our scheme in all other metrics irrespective of the value of the inliers ratio, and in the case of VFC and sVFC, it is much poorer than that of our scheme for all the metrics when the inliers ratio is low. EMDQ is another FMR scheme whose *Recall* performance is superior to that of ours, but its performance in terms of $F - score$ is not as good as that of ours when the inliers ratio is very very low (i.e. between 5-10 %). In terms of other two metrics, namely, *Accuracy* and *Specificity*, both the schemes provide values that are larger than 96%, and in terms of *Precision*, our scheme is much superior to EMDQ for low inliers ratio, and both have similar performance in the range of high inliers ratio. We have also performed experiments in which the average numbers of matches per pair are about 400 and 600. It should be mentioned that conclusions similar to that for the case of average number of matches of about 500 per pair can be drawn.

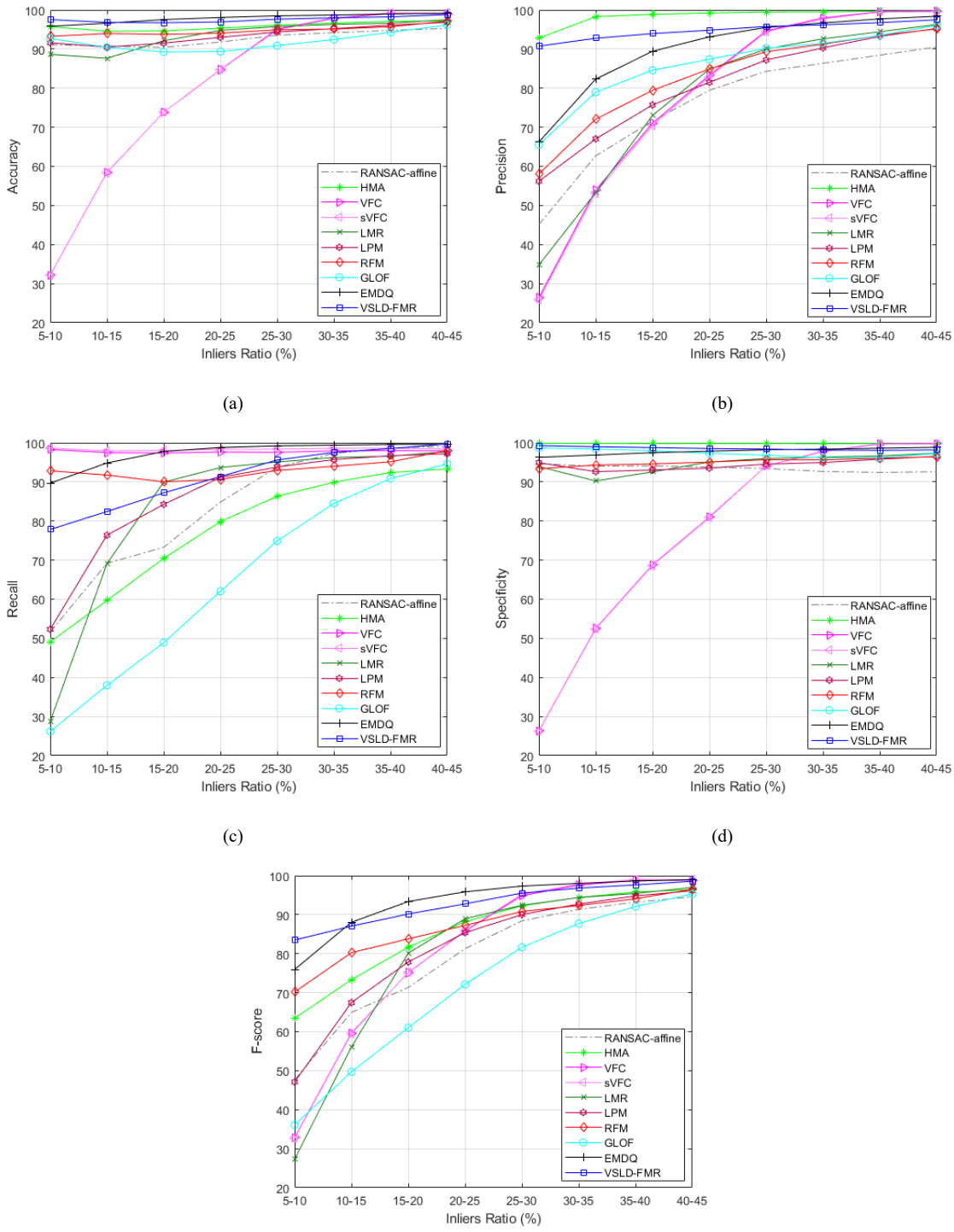


Figure 3.4: Performance comparison of various FMR schemes on the synthetic-laparoscopic image dataset I. Plots of (a) *Accuracy*, (b) *Precision*, (c) *Recall*, (d) *Specificity*, (e) *F-score*, as functions of inliers ratio of the various FMR schemes on the synthetic-laparoscopic image dataset I. The average number of matches per pair of images is about 500.

Figure 3.5(a) shows the average processing times as a function of the inliers ratio for all the FMR schemes on the synthetic-laparoscopic image dataset I for the case when the average number of matches per pair is about 500. It is seen from this figure that the average processing times of five of the schemes, namely, HMA, EMDQ, LMR, GLOF, and RANSAC-affine, are much larger in comparison to that of the other five schemes, making them less attractive for tasks such as MIS. This is unfortunate especially in the cases of HMA and EMDQ that provide very good performance in terms of the five metrics. It is seen from this figure that EMDQ has an exceptional characteristic of its processing time becoming significantly lower as the inliers ratio increases. The reason for this is that in contrast to other methods in this method, R1P-RNSC which is used to find initial rigid transformations requires a large number of iterations when the inliers ratio is low and relatively smaller number of iterations when the inliers ratio is large. In order to have a better comparison of the processing time of the five schemes providing the lowest processing times, namely, VFC, sVFC, LPM, RFM, and VSLD-FMR, in Figure 3.5(b) we show a zoomed version of Figure 3.5(a) for these five schemes. It is seen that the proposed and sVFC schemes provide the best average processing time, and VFC the worst average processing time, among all the FMR schemes irrespective of the inliers ratio. It is noted that even though LPM is the fourth best in terms of the average processing time, its processing time is the least sensitive to variations in the inliers ratio. Finally, from this figure it is also noted that in comparison to sVFC, the processing time of the proposed scheme is less sensitive to the inliers ratio.

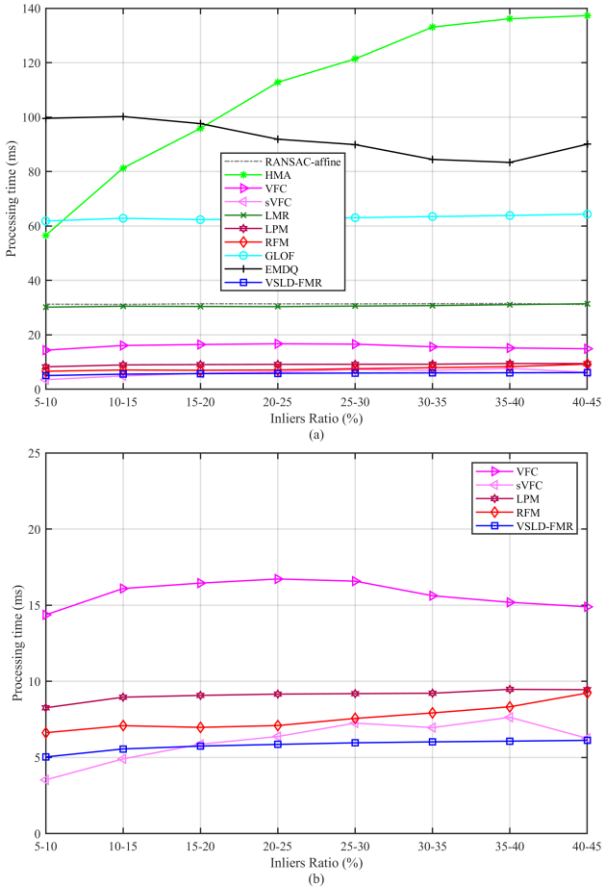


Figure 3.5: Average processing time of various FMR schemes as a function of the inliers ratio on the synthetic-laparoscopic image dataset I. (a) Plots of the average processing time of the various schemes as functions of the inliers ratio on the synthetic-laparoscopic image dataset I when the average number of matches per pair is about 500. (b) Zoomed version of the five schemes with lowest processing times.

Figure 3.6(a) shows the average times of processing of the pairs of images from the synthetic-laparoscopic image dataset I whose numbers of matches lie in the ranges 100-200, 200-300, …, 700-800, for feature matching refinement using the various schemes. In general, the processing time should tend to increase as the number of matches in the pair of images increases. This is what is seen from the plots of Figure 3.6(a) in which the processing times of the various schemes increase with varying amounts as the number of matches increases, with the exception of EMDQ when the number of matches per pair

exceeds 600. The reason for this exception is the fact that the increase in the processing time due to larger number of matches is countered by the decrease in the processing time due to the larger number of pairs of images with larger inliers ratio. In order to have a better comparison of the five schemes with the lowest processing times, namely, VFC, sVFC, LPM, RFM, and VSLD-FMR, in Figure 3.6(b) we show a zoomed version of Figure 3.6(a) for these five schemes. From this figure it is seen that when the number of matches is less than 400, the proposed FMR scheme provides the second lowest processing time with slightly larger time than that required by the RFM scheme. However, beyond 400 matches, the proposed scheme provides the lowest processing time, whereas the processing time of RFM becomes the third or the fourth lowest.

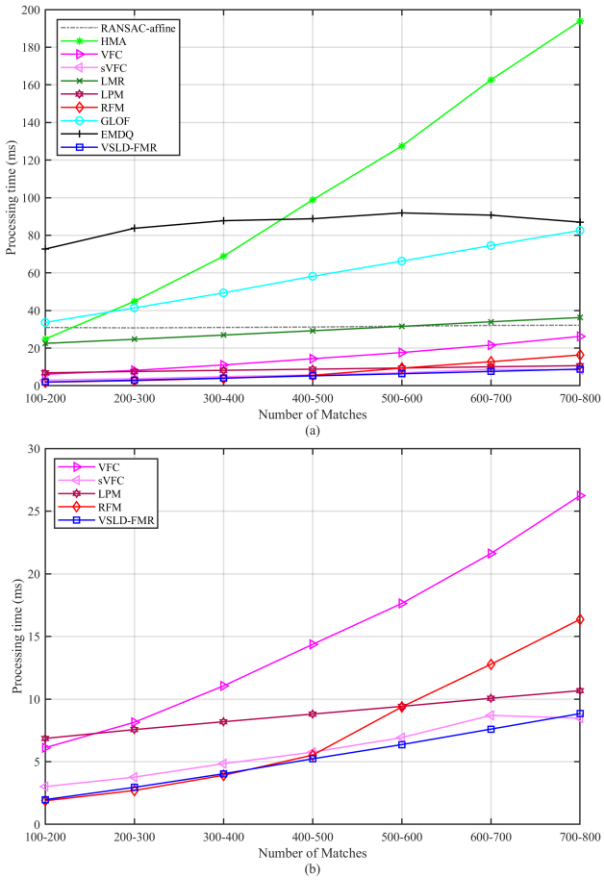


Figure 3.6: Average processing time of various FMR schemes as a function of the number of matches on the synthetic-laparoscopic image dataset I. (a) Plots of the average processing time of the various schemes as functions of the number of matches on the synthetic-laparoscopic image dataset I. (b) Zoomed version of the five schemes with lowest processing times.

3.3.3 Results on the Heart Phantom Image Dataset

We compare the average performance of the various schemes in terms of the five metrics and the processing time for the heart phantom image dataset. The purpose of using this dataset is to examine the impact of viewpoint changes and illumination changes between the left and right images in the pair, which do not have deformation variations, on the performance of the various FMR schemes. Note that in the heart phantom image dataset, the pairs of the images have large values for the inliers ratio, since the left and right images

in the pairs differ only in terms of illumination and viewpoints, and not in terms of the other factors such as deformation.

Table 3.4 gives the average performance of the various schemes in terms of the five metrics as well as the processing time for the heart phantom image dataset. It is seen that for this dataset, the proposed VSLD-FMR and EMDQ schemes are again the top two performing schemes, as in the case of synthetic-laparoscopic image dataset I. However, the processing time of the former is still about one-fifteenth of that of the latter.

Table 3.4: Average values of accuracy, precision, recall, specificity, F-SCORE, and processing time on the images of the heart phantom image dataset for the various FMR schemes.

Method	Acc.	Prec.	Rec.	Spec.	F-score	Time (ms)
RANSAC-affine [29]	0.95	0.92	**1.00**	0.87	0.96	30.41
HMA [30]	0.96	**0.99**	0.94	**0.99**	*0.97*	75.04
VFC [32]	***0.98***	**0.99**	0.97	**0.99**	***0.98***	4.14
sVFC [32]	***0.98***	**0.99**	0.97	**0.99**	***0.98***	*2.62*
LMR [34]	*0.97*	*0.97*	0.98	0.95	*0.97*	24.19
LPM [33]	***0.98***	0.96	**1.00**	0.94	***0.98***	7.47
RFM [35]	*0.97*	0.95	**1.00**	0.92	*0.97*	***2.35***
GLOF [36]	***0.98***	***0.98***	*0.98*	*0.97*	***0.98***	36.95
EMDQ [31]	**0.99**	**0.99**	***0.99***	***0.98***	**0.99**	31.16
Proposed VSLD-FMR	***0.98***	**0.99**	*0.98*	**0.99**	**0.99**	**2.08**

Figure 3.7 (a) – Figure 3.7 (e) show the plots of the five metrics as functions of the inliers ratio resulting from the application of various FMR schemes on the heart phantom image dataset. Note that in the heart phantom image dataset the pairs of the images have large values for the inliers ratio, since the left and right images in the pairs differ only in terms of illumination and viewpoints and not in terms of the other factors such as deformation. It is seen from this figure that the performance of the proposed VSLD-FMR scheme in terms of the *Precision, Specificity, Accuracy* and *F-score* metrics is always greater than 98% for all the values of the inliers ratio with the ranks of the third best and second best in the first two and the last two metrics, respectively. Even though the *Recall* performance of the proposed scheme is not as high as that achieved by some of the other schemes, its *Recall* values nonetheless are always larger than 97%. Figure 3.7 (f) shows the processing time of all the schemes as a function of the inliers ratio. It is seen from this figure that sVFC, RFM, and VSLD-FMR are the three schemes with the lowest processing times, which are almost the same and extremely low.

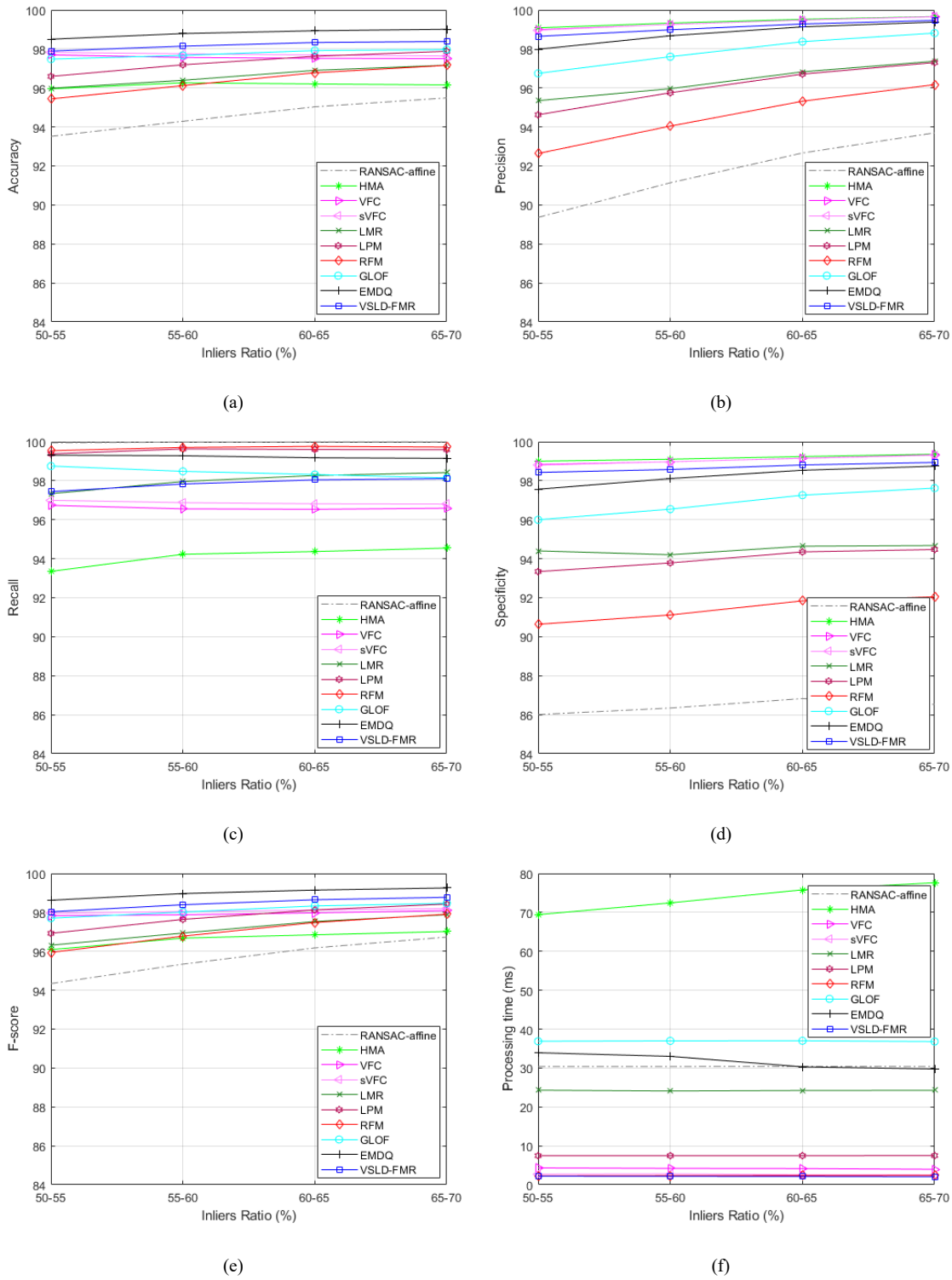


Figure 3.7: Performance and processing time comparison of various FMR schemes on the heart phantom image dataset. Plots of (a) Accuracy, (b) Precision, (c) Recall, (d) Specificity, (e) F-score, and (f) the processing time as functions of inliers ratio of the various FMR schemes on the heart phantom image dataset. The average number of matches per pair of images is about 215.

3.3.4 Results on a Couple of Images from Colonoscopy and Gastrointestinal Image Datasets

In Figure 3.3, we have provided visual illustration of the performance of the three FMR schemes, namely, sVFC, EMDQ, and the proposed VSLD-FMR schemes by considering a pair of images from the laparoscopic image dataset. We now consider a couple of pairs of images from two other datasets, one pair from the SUN colonoscopy video database [50], [51] and the other from the HyperKvasir gastrointestinal dataset [52], [53] to further show the effectiveness of the proposed method. It is important to point out our limitation in using datasets other than those we have already used to carry out a comprehensive evaluation and comparison of the performance of the various FMR schemes. The limitation in using the various other datasets, such as [51], [53] is that they do not provide the ground-truth information on the feature matching between the two images in the pairs of images in the dataset.

For the two pairs of images that we have chosen from the SUN colonoscopy video and HyperKvasir gastrointestinal datasets, we have ourselves generated the ground-truth of the matched features for the purpose of both visual illustration and objective evaluation of the performance of the three FMR schemes by carrying out the following procedure. We first generate putative sets of matches for these two pairs by applying the SIFT scheme for the detection, description, and matching of the feature points in the left and right images. Each match is then examined for its correctness by four judges independently. A particular match in the putative set is considered to be true only if there is a complete unanimity among the four judges for its correctness. Figure 3.8 (Figure 3.9) shows the ground-truth

outliers and inliers of feature matches between the images of a pair of images taken from the SUN colonoscopy video database (the HyperKvasir gastrointestinal dataset), as well as true positive matches shown by green color lines and false positive matches shown by magenta color lines in the refined sets obtained by using the three FMR schemes. It is seen from these figures that the proposed scheme results in refined sets of matches without including any false matches for both pairs of images selected, of which the first pair has a very small value of the inliers ratio (0.17) and the second one has relatively a larger value (0.43).

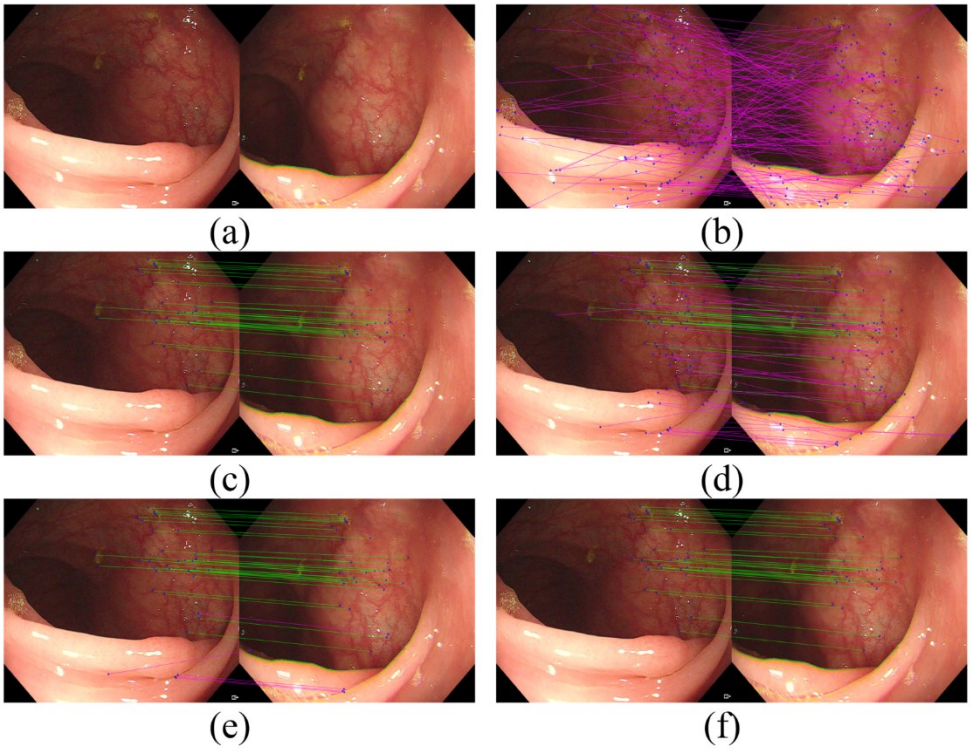


Figure 3.8: Visual illustration of sVFC, EMDQ, and VSLD-FMR schemes on a pair of images from the SUN colonoscopy video database. (a) A pair of images from the SUN colonoscopy video database. (b) Ground-truth outliers. (c) Ground-truth inliers. Inliers identified by (d) sVFC, (e) EMDQ, and (f) the proposed VSLD-FMR scheme. In (d), (e), and (f), the green lines indicate matches that are correctly identified as true matches, and the lines in magenta color indicate matches that are incorrectly identified as true matches.

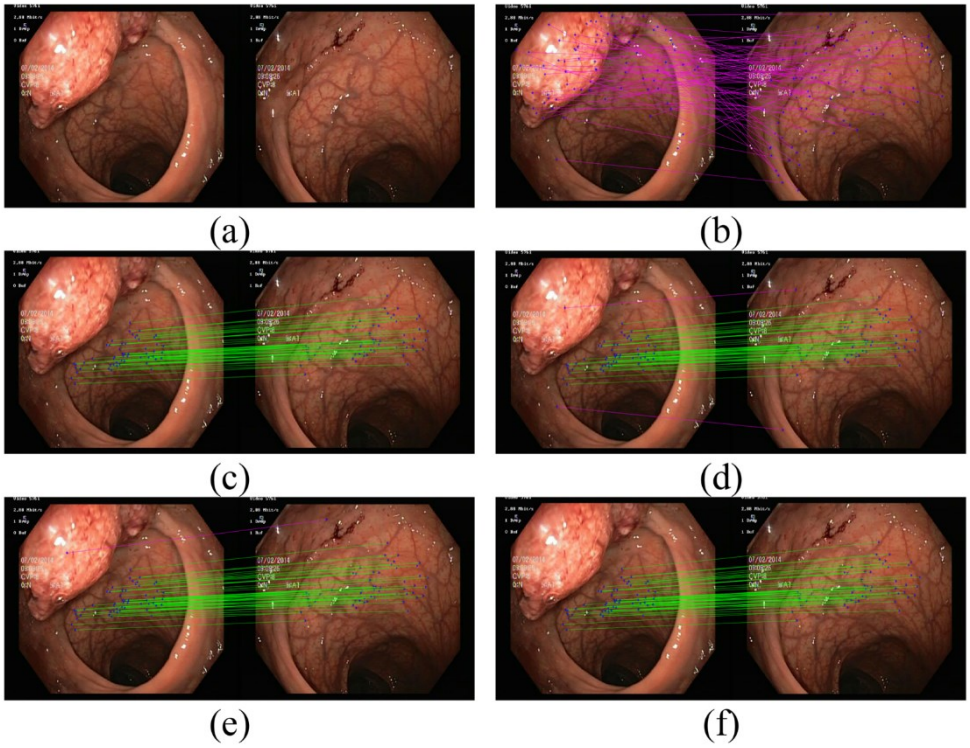


Figure 3.9: Visual illustration of sVFC, EMDQ, and VSLD-FMR schemes on a pair of images from the HyperKvasir gastrointestinal dataset. (a) A pair of images from the HyperKvasir gastrointestinal dataset. (b) Ground-truth outliers. (c) Ground-truth inliers. Inliers identified by (d) sVFC, (e) EMDQ, and (f) the proposed VSLD-FMR scheme. In (d), (e), and (f), the green lines indicate matches that are correctly identified as true matches, and the lines in magenta color indicate matches that are incorrectly identified as true matches.

Table 3.5 gives the performance of the three FMR schemes in terms of TP, FP, TN, FN, and the five metrics as well as the processing time separately for the same two pairs of images considered in Figure 3.8 and Figure 3.9. It is seen from this table that for both pairs, the proposed scheme exhibits a somewhat better performance compared to that of EMDQ, but a performance significantly superior to that of sVFC scheme. Further, the processing time of the proposed scheme is generally a small fraction of that of the other two schemes.

Table 3.5: Values of TP, FP, TN, FN, accuracy, precision, recall, specificity, F-score, and processing time on two pairs of images taken from the SUN colonoscopy video and the HyperKvasir gastrointestinal datasets.

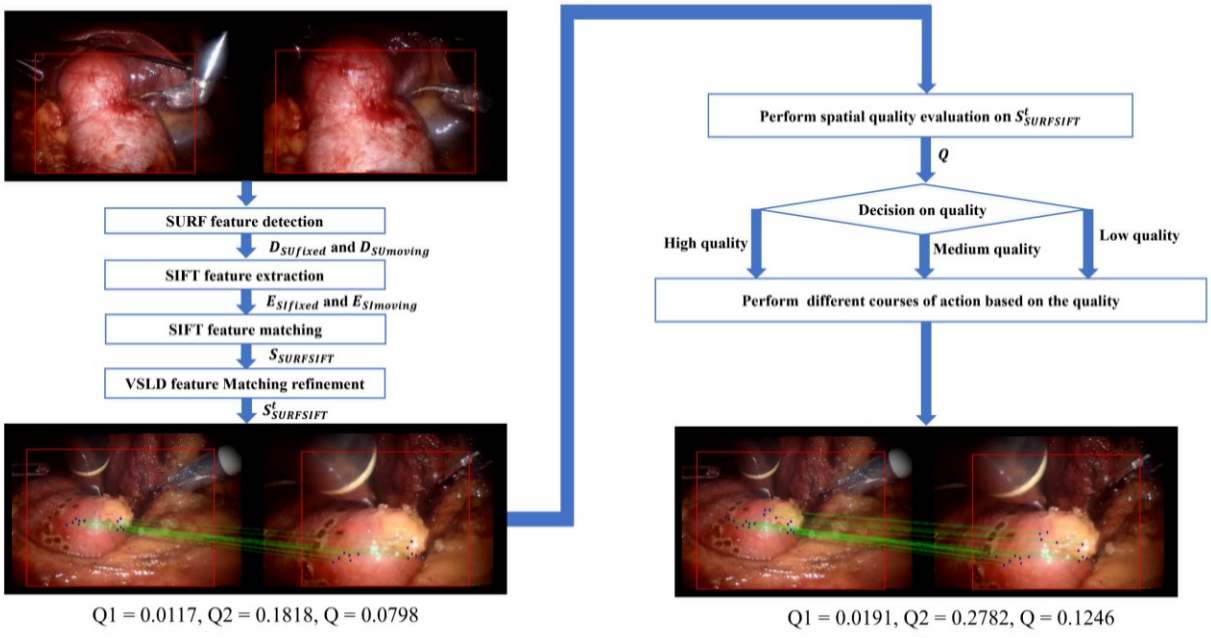
Method	**TP**	**FP**	**TN**	**FN**	**Acc.**	**Prec.**	**Rec.**	**Spec.**	**F-score**	**Time (ms)**
Pair of images with inliers ratio of 0.17 taken from the SUN colonoscopy video database [51]										
sVFC [23]	31	38	112	0	0.79	0.45	1	0.75	0.62	6.6
EMDQ [22]	31	5	145	0	0.97	0.86	1	0.97	0.93	62.2
VSLD-FMR	29	0	150	2	0.99	1	0.94	1	0.97	2.1
Pair of images with inliers ratio of 0.43 taken from the HyperKvasir gastrointestinal dataset [53]										
sVFC [23]	57	2	74	0	0.99	0.97	1	0.97	0.98	3.3
EMDQ [22]	57	1	75	0	0.99	0.98	1	0.99	0.99	27.2
VSLD-FMR	57	0	76	0	1	1	1	1	1	1.7

3.4 Summary

In this chapter, we have proposed an extremely low complexity and accurate FMR scheme referred to as VSLD-FMR, particularly when the pair of images have a low inliers ratio. The main idea used in the proposed scheme while refining a putative set of matched features is in deciding the correctness of the matches in the set based on the number of votes received by its individual matches. A vote to a match towards its being a true match is cast to all those matches within a small circular neighborhood around a feature, which have similar displacement vectors with that of the feature at the center of the neighborhood. Extensive experiments on feature matching refinement using the proposed and nine state-of-the-art schemes on different datasets have been performed. The experimental results have shown that the proposed scheme provides a performance, which is second to none, at an extremely low computational cost, regardless of the inliers ratio.

Chapter 4

A Robust Scheme for Detection, Extraction, and Matching of Features in MIS Images



4.1 Introduction

In this chapter, we propose a fast and accurate FDEM scheme that combines the strong attributes of three well-known FDEM schemes, SIFT, SURF and ORB for generating a putative set of matched features for a given pair of images [54]. We first present a top-level description of the scheme and then provide the incorporation and implementation of the

various strategies used in developing the proposed FDEM scheme. In developing the proposed FDEM scheme, we develop and use a novel metric to measure the spatial quality of the set of matched features. A number of experiments are performed on the proposed FDEM scheme using MIS datasets, including a real laparoscopic image dataset and a synthetic-laparoscopic image dataset, to demonstrate the effectiveness of the proposed FDEM scheme. The performance of the proposed scheme is compared with that of ORB, SURF and SIFT FDEM schemes. Furthermore, in this chapter, image registration is considered as an example of the application of the putative set of matched features generated by the proposed FDEM scheme to study the impact of the spatial quality of the generated putative set on the quality of the application.

4.2 Proposed FDEM Scheme

As mentioned earlier, an FDEM scheme has three parts, namely, feature detection, feature extraction, and feature matching. It has been noted from the review of the related FDEM schemes that the methods used for each of the three parts of SIFT, SURF and ORB FDEM schemes, have distinct advantages and disadvantages. In this chapter, we develop a new FDEM scheme leveraging the advantages of the methods for the three parts used in the schemes of SIFT, SURF and ORB, so as to improve the quality of the generated putative sets of matched features at a reasonably low computational complexity. The proposed FDEM scheme has three stages that are used to generate the putative set of matched features for a pair of images. Depending on the spatial quality of the putative set of matched features (whose matching quality has been enhanced by using a feature matching

refinement scheme) for the pair of images generated by the first stage, an additional stage (stage 2 or 3) is or is not used to improve the spatial quality of the set. Since the decision of using only one or more than one stage to generate the final putative set of matched features depends on the spatial quality of the putative set of the first stage, in this chapter we also develop a good metric to measure the spatial quality of a set of matched features. Now, in this section, we present our proposed FDEM scheme based on the spatial quality of a feature matching set, assuming that a metric for the spatial quality does exist. Then, in the next section, we propose our new metric to measure the spatial quality of a putative set of matched features.

In MIS images, the parts of the images containing surgical instruments and specular reflections are not useful for detection and matching of the features. Hence, it is important to identify a region of interest (ROI) in an MIS image that does not include such parts of the images and use it for the detection of the features. By focusing exclusively on the ROIs of the fixed and moving images, the detection and matching of irrelevant features by an FDEM scheme can be avoided, and thus, the overall accuracy can be enhanced and the processing time reduced. Therefore, in the proposed FDEM scheme, we first identify the ROIs of the fixed and moving images. Since the ROIs in MIS images are dominated mainly by the tissue's red color, in order to identify the ROI of an image, we first transform the input RGB image into an HSV (hue, saturation, and value) color space and use its three components to segment the ROI from the image. The ROI in an MIS image generally has the following properties in terms of the three components of HSV. The red color occupies the lower and upper parts of the hue spectrum. Also, the organ part of an MIS image must have reasonably high values for S and V components. We can specify the

ranges of the three components of HSV color space by using four threshold parameters, δ_1, δ_2, δ_3, and δ_4, and require that the hue component of a pixel point of ROI lies either in the range $[0, \delta_1]$ or $[\delta_2, 1]$, where $\delta_2 > \delta_1$, its saturation component lies in the range $[\delta_3, 1]$, and its value component in the range $[\delta_4, 1]$. Finally, all the isolated components of the ROI that have only small numbers of pixels (less than δ_5) are removed to obtain the final ROI. As the input to an FDEM scheme needs to have a regular shape, we find a rectangular bounding box enclosing the ROI of the image and use the part of the image within the bounding box for developing the proposed FDEM scheme.

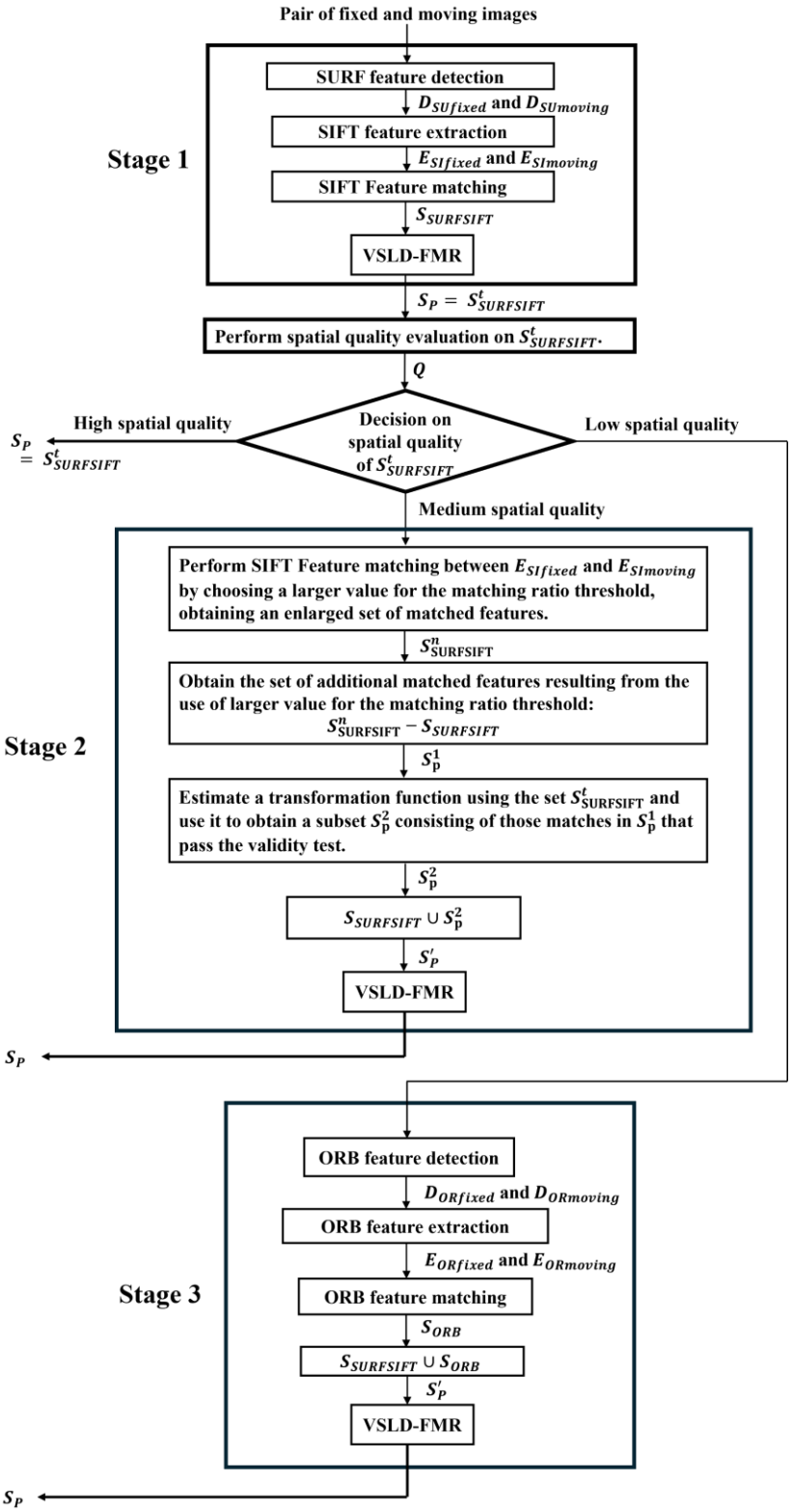


Figure 4.1: Block diagram of the proposed FDEM scheme, SIFOR.

For a given pair of fixed and moving images, the proposed FDEM scheme starts with stage 1, in which the SURF detection method is used to detect the set of features, $D_{SUfixed}$ and $D_{SUmoving}$, from the fixed and moving images, respectively. Even though both SURF and SIFT detect the same type of features, i.e., blob features, we choose SURF in view of its much lower feature detection complexity. The reason for choosing SURF over ORB for detection is that the former is more robust to distortions. Then, for the feature extraction process, the proposed FDEM scheme employs SIFT method to construct the set of descriptors, $E_{SIfixed}$ and $E_{SImoving}$, for the detected SURF features. The reason for choosing SIFT over SURF for extraction is that the former seems to provide feature representations for MIS images that are more distinctive, even though it requires a slightly larger time for extraction. Next, the descriptors in $E_{SIfixed}$ and $E_{SImoving}$ are matched using the SIFT scheme for matching, which is the same as that used by SURF, resulting in a set of matched features, $S_{SURFSIFT}$. In this scheme, the matching distance threshold (τ_d) and the matching ratio threshold (τ_r) are used to find the matches of the features in $E_{SIfixed}$ and $E_{SImoving}$. Finally in stage 1, we perform a feature matching refinement operation on $S_{SURFSIFT}$ using VSLD-FMR [44] to ensure that the resulting refined set $S^t_{SURFSIFT}$ has a good matching quality.

Next, we measure the spatial quality of the set $S^t_{SURFSIFT}$ of matched features produced by stage 1, by employing a suitable spatial quality metric. For this purpose, we develop in the next section, an efficient and effective scheme for determining the spatial quality Q of a putative set of matched features and apply it to $S^t_{SURFSIFT}$. In developing such a scheme, we take into consideration the density of the matches as well as their dispersion over the

region of interest in the pair of images used to obtain the putative set of matched features. It is important to note that the putative set produced by stage 1 has a good matching quality; however, it may or may not have, at the same time, a good spatial quality. Hence, it is important to determine the spatial quality of the generated $S^t_{SURFSIFT}$. If $S^t_{SURFSIFT}$ is of sufficient high spatial quality, then it has a good overall quality, both spatially as well as in terms of the matching accuracy. At this point, based on the spatial quality test we determine whether the set $S^t_{SURFSIFT}$ is of high, medium, or low spatial quality, and depending on the level of the spatial quality we decide as to whether or not $S_{SURFSIFT}$ needs to be supplemented with additional matches. If $S_{SURFSIFT}$ needs to be supplemented with additional matches, then again based on the level of the spatial quality, we decide whether it needs to be supplemented with additional blob matched features or with ORB matched features.

(i) High spatial quality $S^t_{SURFSIFT}$

Using a high threshold th_H we decide whether or not $S^t_{SURFSIFT}$ is of sufficient high spatial quality, that is, if $Q \geq th_H$, then that set is considered to be of high spatial quality. In this case, the final putative set of matched features for the pair of images, S_P, is simply $S^t_{SURFSIFT}$.

(ii) Medium spatial quality $S^t_{SURFSIFT}$

If $Q < th_H$, then by using a threshold $th_M < th_H$, we decide whether or not the set is of medium spatial quality, that is, if $th_M \leq Q < th_H$, we consider that set to be of medium spatial quality. In this case, we generate additional blob matched features by performing the following operations in stage 2:

1. Apply again the SIFT matching scheme on $E_{SIfixed}$ and $E_{SImoving}$ by using the same value for the matching distance threshold (τ_d) as used earlier but by using a larger value for the matching ratio threshold (τ_r) leading to a new set of matched features $S^n_{SURFSIFT}$. Note that $S^n_{SURFSIFT}$ is a set larger than $S_{SURFSIFT}$ and it contains all the matches of the latter set. Remove the matches in $S_{SURFSIFT}$ from the set $S^n_{SURFSIFT}$ forming the set S^1_p.
2. Let n_t be the number of pairs in $S^t_{SURFSIFT}$, and let $S^t_{SURFSIFT} = \{(P^t_i, P^{t\prime}_i), i = 1, \dots, n_t\}$, where P^t_i and $P^{t\prime}_i$ represent the locations of the i^{th} pair of matched feature in the set $S^t_{SURFSIFT}$. Estimate a transformation function that maps the spatial locations, $P^t_i, i = 1, \dots, n_t$ of the feature points in the fixed image, to the corresponding spatial locations, $P^{t\prime}_i$, of the moving image. For estimating the transformation, thin plate spline (TPS) [55], a well-known technique for estimating non-rigid transformation, is used.
3. Let P_f and P_m denote the locations, respectively, in the fixed and moving images of the features of a new match in the set S^1_p. Apply the estimated transformation on P_f to find the corresponding location P^e_m in the moving image. If the Euclidean distance between P_m and P^e_m is less than a distance threshold τ_E then this match is considered to be valid, otherwise it is removed from the set S^1_p. The remaining matches in S^1_p form the set S^2_p. After this process, a set S'_P is formed as the union of $S_{SURFSIFT}$ and S^2_p.
4. Perform VSLD-FMR [44] on S'_P leading to a final putative set S_p produced by stage 2

for the pair of images belonging to the medium category.

(*iii*) *Low spatial quality* $S^t_{SURFSIFT}$

If $Q < th_M$, then the spatial quality of $S^t_{SURFSIFT}$ for a pair of fixed and moving images is considered to be low. In this case, $S_{SURFSIFT}$ is supplemented with ORB matched features in stage 3. The reason for not including additional matches of SURF features in $S_{SURFSIFT}$ by increasing the value of τ_r even higher than that used in the medium case is as follows. In the low-quality case, the number of matches in $S^t_{SURFSIFT}$ would be too small or the dispersion of the matches too poor to reliably estimate a transformation for it to be used for generating additional SURF matched features. The reason for not using SIFT either, for generating additional matched features, is that the detection of features using SIFT is computationally expensive, as noted earlier in the development of our algorithm.

In the originally proposed ORB scheme [20], a measure of cornerness is calculated for each candidate corners, using the Harris corner measure [56], and M strongest corners are selected as the set of detected features in the feature detection part of the ORB and this set is used in the feature extraction and matching parts of ORB. In view of the possibility that M corner features so chosen by ORB may not necessarily be well dispersed over the region of interest, we select M corner features in a way that assures their good dispersion over the region of interest. For this purpose, if the number of ORB detected features in the ROI is less than a threshold δ_6, all the detected features are used for the purpose of feature extraction and matching using ORB. However, if the number of ORB detected features is greater than or equal to δ_6, the bounding box enclosing ROI is partitioned into square blocks each of size $\delta_7 \times \delta_7$. Corresponding to the fixed and moving images of a pair of

images, the sets $D_{ORfixed}$ and $D_{ORmoving}$ are formed by including in these sets the strongest feature from each of the blocks in the two images. The detected features in the sets, $D_{ORfixed}$ and $D_{ORmoving}$ are used to obtain the set of binary descriptors, $E_{ORfixed}$ and $E_{ORmoving}$, and then to obtain the set S_{ORB} of matched features using the feature descriptors contained therein by employing the feature extraction and matching schemes of ORB. The set $S_{SURFSIFT}$ is supplemented with the set S_{ORB} and the resulting set S_P' is finally refined using the scheme of VSLD-FMR [44] presented in Chapter 3, to obtain the final putative set S_P for the pair of images belonging to the low-quality category.

We refer to the proposed FDEM scheme as SIFOR (**S**URF combined with S**IF**T and **OR**B). Figure 4.1 shows a block diagram of SIFOR.

In the following section, we propose a metric, which has been used in the development of the proposed FDEM scheme, SIFOR, to measure the spatial quality of a putative set of matched features.

4.3 Spatial Quality Evaluation of a Set of Matched Features

In all the FDEM schemes developed in the literature, the focus has been on the accuracy of matches of the putative set, that is, FDEM schemes have been designed to provide a putative set with a good matching quality. However, there is another desirable characteristic that a putative set must observe, that is, the matches in the putative set must have good density as well as a good dispersion over the regions of interest in the pair of images, which is very important in almost all applications of putative sets of matched

features. We refer to this characteristic of a putative set as the spatial quality of the set, which to the best of our knowledge, has been neglected in developing a scheme for generating putative sets of matched features. In this section, we propose a new metric to measure the spatial quality of the matches in a putative set.

Let $S_M = \{(P_i, P_i'), i = 1, \ldots, N\}$ be the set of the locations of the matched pairs of features in a pair of fixed and moving images, where P_i and P_i' denote the coordinates of the i^{th} matched features in the two images, respectively.

For formulating a metric to determine the spatial quality of S_M, we consider two characteristics of S_M: (i) a suitable density of the matches, and (ii) a satisfactory dispersion (spread) of the matches across the two images. Therefore, first we define two metrics denoted by Q_1 and Q_2, where the former measures the density, and the latter the dispersion of the features in the fixed and moving images forming the matches in S_M.

We define the first metric Q_1 as

$$Q_1 = \min\left(\frac{\rho_{fixed}}{\rho_{max}}, \frac{\rho_{moving}}{\rho_{max}}\right) \tag{4.1}$$

where $\rho_{fixed} = \frac{N}{A_f}$, and $\rho_{moving} = \frac{N}{A_m}$ are, respectively, the densities of the features in the fixed and moving images, N being the number of matched features in the set S_M, A_f and A_m being the areas of the bounding boxes enclosing the regions of interest in the fixed and moving images, respectively, and ρ_{max} is the maximum density determined empirically.

In [57] an index called Clark-Evans aggregation index (R) has been proposed to measure the degree of clustering (or dispersion) of a point pattern of a population of plants

or of animals using the average nearest-neighbor distance. The authors defined this index as

$$R = \frac{d_o}{d_e} \tag{4.2}$$

with

$$d_o = \frac{\sum_{j=1}^{\eta} d_j}{\eta} \tag{4.3}$$

and

$$d_e = 0.5\sqrt{\frac{A}{\eta}} \tag{4.4}$$

where d_j is the Euclidean distance between the j^{th} individual and its nearest neighbor, η is the total number of individuals in the population, and A is the study area. Since the points near the edges of the study area have fewer neighbors compared to points in the interior area, the author of [58] proposed an edge corrected version for d_e as

$$d_e = 0.5\sqrt{\frac{A}{\eta}} + \left(0.0514 + \frac{0.041}{\sqrt{\eta}}\right)\frac{B}{\eta} \tag{4.5}$$

where B is the perimeter of the study area, and introduced a modified version of Clark-Evans aggregation index obtained by using d_e given by (4.5) in (4.2). The value of R exhibits varying dispersion patterns and it ranges from 0 to 2.149. If the value of R is 0, all the points in the pattern are merged into a single cluster. If it is equal to 1, the distribution of the points is completely random. As R approaches the maximum value of $R_{max} = 2.149$, the distribution of the points tends to become uniform [59].

We normalize the modified Clark-Evans aggregation index as

$$R_N = \frac{1}{R_{max}} \times \frac{\frac{\sum_{j=1}^{\eta} d_j}{\eta}}{0.5\sqrt{\frac{A}{\eta}} + \left(0.0514 + \frac{0.041}{\sqrt{\eta}}\right)\frac{B}{\eta}} \tag{4.6}$$

and adopt this normalized version of the index as a measure of the degree of dispersion of the feature points in the fixed and moving images. This results in obtaining two aggregation indices as given below, one for the fixed image and the other for the moving image,

$$R_{N_{fixed}} = \frac{1}{R_{max}} \times \frac{\frac{\sum_{j=1}^{N} d_j^f}{N}}{0.5\sqrt{\frac{A_f}{N}} + \left(0.0514 + \frac{0.041}{\sqrt{N}}\right)\frac{B_f}{N}} \tag{4.7}$$

$$R_{N_{moving}} = \frac{1}{R_{max}} \times \frac{\frac{\sum_{j=1}^{N} d_j^m}{N}}{0.5\sqrt{\frac{A_m}{N}} + \left(0.0514 + \frac{0.041}{\sqrt{N}}\right)\frac{B_m}{N}} \tag{4.8}$$

where d_j^f and d_j^m are the Euclidean distances between the j^{th} feature and its nearest neighbor in the fixed and moving images, respectively, N is the total number of matched features in the set S_M, A_f and A_m are the areas of the bounding boxes enclosing the regions of interest in the fixed and moving images, respectively, and B_f and B_m are the perimeters of the bounding boxes in the fixed and moving images, respectively. In our case A_f and A_m, and B_f and B_m are determined using the lengths, in terms of the number of pixels, of the sides of the bounding boxes. As mentioned above the value of R_{max} is 2.149. We have

observed that in MIS images the actual values of the modified Clark-Evans aggregation index are significantly lower than this maximum value. Hence, the normalization of the aggregation index by this maximum value will make the values of the normalized modified Clark-Evans aggregation index unnecessarily much smaller. By applying the four FDEM schemes, SIFT, SURF, ORB and SIFOR, on the 100 pairs of laparoscopic image dataset, we have observed that the maximum values of R are 0.7328, 0.8387, 0.4478, and 0.7299, respectively. Hence, we select the value of 0.8387 for R_{max} for normalizing the indices in (4.7) and (4.8). We define the second metric Q_2 as

$$Q_2 = \min\left(R_{N_{fixed}}, R_{N_{moving}}\right) \tag{4.9}$$

We choose the weighted geometric mean of Q_1 and Q_2 given by

$$Q = \left(\prod_{i=1}^{2} {Q_i}^{w_i}\right)^{1/\sum_{i=1}^{2} w_i} \tag{4.10}$$

as a metric to measure the overall spatial quality of a putative set of matched features. In our case, the values of both Q_1 and Q_2 as well as those of the weights w_1 and w_2 are between 0 and 1, and $w_1 = 1 - w_2$. Equation (4.10) can be rewritten as

$$Q = {Q_1}^{1-w_2} {Q_2}^{w_2} \tag{4.11}$$

The reason for choosing the geometric mean of Q_1 and Q_2 as a measure of the spatial quality of a putative set is two-fold. First, this geometric mean as compared to the arithmetic mean has the property that its value tends to be biased towards the lower value. Hence, a higher value of Q is an indication that both Q_1 and Q_2 have reasonably large values. The second reason for our choice of the geometric mean can be explained as follows. It is seen from (4.11) that given the values of Q_1 and Q_2 (irrespective of whether

Q_1 is less than Q_2 or vice-versa), Q is a monotonic function of w_2 having values ranging between Q_1 and Q_2. In applications of putative sets such as registration, even though both the density of the matches (Q_1) and their spread (Q_2) are important, comparatively the spread of the matches is more important. Therefore, the choice of the geometric mean as a metric to measure the quality of a putative set is a suitable choice, since in this metric we can attach more importance to the spread of matches by assigning a larger weight to Q_2 by giving a value to w_2 larger than that of w_1. In our use of the metric Q for the spatial quality of a putative set of matched features, we set $w_1 = 0.3$ and $w_2 = 0.7$. A putative set of matched features generated by an FDEM scheme is considered to be of high, medium, or low quality, depending on its value of the metric Q lies in the region [0.5 1], [0.2 0.5) or [0 0.2). Thus, in SIFOR, we use the values of th_H and th_M as 0.5 and 0.2, respectively. The scheme for spatial quality evaluation of a set of matched features is summarized as Algorithm 1.

Algorithm 1: Spatial quality evaluation of the set S_M

Input:

The pairs of the locations of all the matched feature points in S_M, $(P_i,\ P_i'),\ i = 1,\ \ldots,\ N$, $A_f, A_m, B_f, B_m, \rho_{max}, R_{max}$, and w_2.

1. Calculate $\rho_{fixed} = \frac{N}{A_f}, \rho_{moving} = \frac{N}{A_m}$.
2. Calculate $Q_1 = \min\left(\frac{\rho_{fixed}}{\rho_{max}}, \frac{\rho_{moving}}{\rho_{max}}\right)$.
3. Calculate $R_{N_{fixed}} = \frac{1}{R_{max}} \times \frac{\frac{\sum_{j=1}^{N} d_j^f}{N}}{0.5\sqrt{\frac{A_f}{N}} + \left(0.0514 + \frac{0.041}{\sqrt{N}}\right)\frac{B_f}{N}}$.
4. Calculate $R_{N_{moving}} = \frac{1}{R_{max}} \times \frac{\frac{\sum_{j=1}^{N} d_j^m}{N}}{0.5\sqrt{\frac{A_m}{N}} + \left(0.0514 + \frac{0.041}{\sqrt{N}}\right)\frac{B_m}{N}}$.
5. Calculate $Q_2 = \min\left(R_{N_{fixed}}, R_{N_{moving}}\right)$.
6. Calculate $Q = {Q_1}^{1-w_2} {Q_2}^{w_2}$.
7. **if** $Q \geq th_H$ then the quality of the set $S_M\ is\ High$.
   else if $Q \geq th_M$ then the quality of the set $S_M\ is\ Medium$.
   else the quality of the set $S_M\ is\ Low$.
   end if

4.4 Experimental Results

In this section, we present the performance results and the processing time of our proposed FDEM scheme, SIFOR, and compare these results with that of SIFT [2], SURF [19] and ORB [20] FDEM schemes. The proposed FDEM scheme is implemented in MATLAB and executed on a computer with an AMD Ryzen 7 3800X 8-Core 3.89 GHz processor. The MATLAB has built-in functions for the three schemes with which we compare our scheme. We run these three schemes on the same hardware platform as we do our own scheme.

To evaluate the efficacy of the four FDEM schemes, these schemes are run on two datasets having different challenging conditions. These datasets are the laparoscopic image dataset of [46] (a publicly available dataset) and a synthetic image dataset generated by us using the laparoscopic image dataset. The laparoscopic image dataset consists of 100 pairs

of color laparoscopic-surgery images, each of 704 × 480 resolution, extracted from six real videos of partial nephrectomy interventions. This dataset is specifically designed to represent the complexities of minimally invasive surgery (MIS) environments. It includes challenging conditions, such as a sparse number of detectable features due to large, texture-less areas and visual ambiguities arising from the nature of surgical images. Additionally, the dataset includes cases of substantial image distortion caused by endoscope lenses and uneven lighting. These factors contribute to a high rate of incorrect matches after the initial appearance-based matching phase, making the dataset particularly demanding for feature-matching algorithms. This dataset includes various challenging situations, such as camera occlusion, camera retraction and reinsertion, sudden camera motion, and specular reflections [30]. Therefore, using this dataset for the evaluation of the proposed FDEM scheme should demonstrate the clinical relevance of the scheme. This dataset also has a partial ground-truth mapping data for each of the 100 pairs of images. Since each of these sets is only a partial set of ground-truth matches, this ground-truth mapping data cannot be used for evaluating an FDEM scheme.

The second dataset, referred to as synthetic-laparoscopic image dataset II containing 18,000 pairs of images (9 groups of 2,000 pairs), constructed by us using the 100 pairs of fixed and moving images of the laparoscopic image dataset and the above-mentioned partial ground-truth mapping data. In order to construct this synthetic dataset, we choose the first pair of the images from the laparoscopic image dataset and estimate a transformation function that maps the fixed image in the pair to the moving image using the ground-truth mapping data. For this purpose, we use the local weighted mean (LWM) transformation method [48] due to its capability in capturing the local deformation that

exists between the fixed and moving images. This estimated transformation is then applied to 20 fixed images selected randomly from the laparoscopic image dataset to generate 20 different moving images having deformation with respect to the corresponding fixed images. The above process is repeated for each of the 100 pairs of the laparoscopic image dataset resulting in a total of 2,000 pairs of images.

It is to be noted that the number of distinct fixed images in this set of 2,000 pairs of images is bounded by the number 100 of the fixed images in the laparoscopic image dataset, whereas all the moving images in the 2,000 pairs are distinct.

Note that the LWM transformation used for constructing these 2,000 pairs of images is successful in mapping the deformation between the fixed and moving images in a pair of images. The other differences such as noise and blurring between the fixed and moving images in the pairs of the laparoscopic dataset are not carried over between the images in the pairs in this set of 2,000 pairs of images. In order to make the differences between the fixed and moving images of a pair more realistic, we corrupt the moving images in the set of 2,000 pairs of images by an additive white Gaussian noise or motion blurring. For this purpose, we introduce different corruption models involving only motion blurring (with motion length $M_L = 10$ and 20 pixels, and motion direction $M_\theta = 0, 45$ and 90 degrees), or only additive white Gaussian noise (with a zero mean and standard deviation $\sigma = 0.1$ and 0.2). Thus, we have a total of 9 groups, each containing 2,000 pairs of images. The moving images of groups 1 to 9 are, respectively, corrupted using the corruption models $(\sigma = 0, M_L = 0, M_\theta = 0)$, $(\sigma = 0.1, M_L = 0, M_\theta = 0)$, $(\sigma = 0.2, M_L = 0, M_\theta = 0)$, $(\sigma = 0, M_L = 10, M_\theta = 0)$, $(\sigma = 0, M_L = 10, M_\theta = 45)$, $(\sigma = 0, M_L = 10, M_\theta = 90)$, $(\sigma =$

$0, M_L = 20, M_\theta = 0)$, $(\sigma = 0, M_L = 20, M_\theta = 45)$, and $(\sigma = 0, M_L = 20, M_\theta = 90)$. Finally, it is important to state that each of the 9 groups of our synthetic dataset enjoys the advantage of having the ground-truth information on the matching of the features between the fixed and moving images, and that this dataset can be used to study the robustness of different FDEM schemes to noise or blurring in the MIS images as well as to the deformation between the fixed and moving images.

The values of the parameters in determining the region of interest are set as $\delta_1 = 0.1, \delta_2 = 0.9, \delta_3 = 0.2$, $\delta_4 = 0.5$, and $\delta_5 = 40$. The parameter settings for SIFOR are carried out as follows. For stage 1 of our algorithm, for the detection of SURF features, we set the parameters as 50 for the strongest Hessian threshold (i.e., a threshold on the determinant of the Hessian matrix.), as 3 for the number of octaves, and as 4 for the number of scale levels per octave. In this stage, for SIFT extraction of the detected SURF features, no parameters have been used, but for the matching of SIFT extracted features, we set the parameter values as, τ_r = 0.77, and $\tau_d = 25$. In stage 2 of SIFOR, the value of the matching parameter τ_d is kept the same as used in stage 1, while the value of the parameter τ_r is increased to 0.99, and the distance threshold parameter τ_E is set to 30. In stage 3, for the ORB feature detection, we set the parameter values as 1.2 for the scale factor and 4 for the number of decomposition levels. For matching the ORB extracted features in this stage, we use the same parameter settings as those used in stage 1. In this stage, the value of the parameter δ_6 used for deciding as to whether the ORB detected features need to be sampled is set to 1,400, and the parameter δ_7 used for partitioning the bounding box of ROI into square blocks is set to 3. Since we compare our proposed FDEM

scheme with SIFT, SURF and ORB schemes, it is important to point out the parameter settings used for these three schemes. For the detection part of SIFT, we set the parameter values as 0.0001 for the contrast threshold (i.e., the threshold to filter out low-contrast feature points), 20 for edge threshold (i.e., the threshold to filter out feature points that are located along edges), 3 for the number of layers in each octave, and 1.6 for sigma of Gaussian. For the detection parts of SURF and ORB, as well as for the matching parts of all these three schemes, we use the same parameter values as those used for SIFOR.

We compare the various FDEM schemes using the average values per pair of fixed and moving images of the following metrics, quality index Q_1 given by (4.1), quality index Q_2 given by (4.9), spatial quality index Q given by (4.11), number of detected corner features (N_{DCF}), number of detected blob features (N_{DBF}), number of detected corner and blob features together (N_{DF}), number of matches in the putative set generated (N_M), and time to produce a putative set (T_{FDEM}). The average value of the metric x is denoted by m_x. It is to be noted that SIFOR is designed to produce a putative set that is refined using the VSLD-FMR [44] feature matching refinement scheme presented in Chapter 3. Therefore, for comparing SIFOR with SIFT, SURF and ORB, we first refine their putative sets by using the same feature matching refinement scheme as used by SIFOR, in order to be fair to these FDEM schemes.

4.4.1 Results on the Laparoscopic Image Dataset

In this subsection, we present the results of the proposed scheme, SIFOR, on the laparoscopic image dataset [46] and compare these results with those of three well-known

FDEM schemes, SIFT [2], SURF [19] and ORB [20]. While the laparoscopic image dataset is one of the few datasets publicly available containing natural MIS images, this dataset, at the same time, includes images with realistic scenarios such as deformation, camera occlusion, camera retraction and reinsertion, sudden camera motion, and specular reflections.

Table 4.1: Average number of detected corner features ($\boldsymbol{m_{N_{DCF}}}$), average number of detected blob features ($\boldsymbol{m_{N_{DBF}}}$), average numbers of detected features ($\boldsymbol{m_{N_{DCF}}}$), average number of matches ($\boldsymbol{m_{N_M}}$), the average values of the quality indices Q_1 ($\boldsymbol{m_{Q_1}}$), Q_2 ($\boldsymbol{m_{Q_2}}$), and Q ($\boldsymbol{m_Q}$) of a putative set of matched features, and the average time ($\boldsymbol{m_{T_{FDEM}}}$) in milliseconds taken to produce it by the various FDEM schemes per pair of fixed and moving images, over 100 pairs of images in the laparoscopic image dataset.

	SIFT [2]	**SURF** [19]	**ORB** [20]	**SIFOR (Proposed)**
$\boldsymbol{m_{N_{DCF}}}$	0	0	4367.4	445.0
$\boldsymbol{m_{N_{DBF}}}$	2257.8	1354.9	0	1354.9
$\boldsymbol{m_{N_{DF}}}$	2257.8	1354.9	4367.4	1799.9
$\boldsymbol{m_{N_M}}$	105.0	57.5	145.7	129.1
$\boldsymbol{m_{Q_1}}$	0.1172	0.0650	0.1662	0.1450
$\boldsymbol{m_{Q_2}}$	0.5268	0.5143	0.3010	0.5495
$\boldsymbol{m_Q}$	0.3244	0.2654	0.2327	0.3599
$\boldsymbol{m_{T_{FDEM}}}$	154.7	50.4	66.0	80.14

Table 4.1 provides the average values for N_{DCF}, N_{DBF}, N_{DF}, N_M, quality indices Q_1, Q_2, and Q, and the average processing time T_{FDEM} per pair of fixed and moving images over the 100 pairs of images of the laparoscopic image dataset resulting from the FDEM schemes, SIFT, SURF, ORB and SIFOR. It is seen from this table that the values of m_{Q_1} (i.e., the average value of the density of the matches in the region of interest) obtained from using SIFT, SURF, ORB and SIFOR are 0.1172, 0.0650, 0.1662 and 0.1450, respectively. The values of m_{Q_2} (i.e., the average value of dispersion of the matches over the region of interest) are increasing in magnitudes in the order of ORB, SURF, SIFT, and

SIFOR. It is to be noted that although ORB has the largest value of Q_1, its value of Q_2 is the lowest, indicating that the matched features found by using ORB are not well distributed over the region of interest; in other words, the features in the matched set provided by ORB in the fixed and/or moving images are clustered. Finally, it is seen that the average value of the overall spatial quality index m_Q provided by SIFOR has the largest value, indicating that the proposed FDEM scheme, SIFOR, provides a putative set, which is superior to that provided by the other three schemes, that is, the features included in the putative set provided by SIFOR have both a good density as well as good dispersion over the region of interest.

Figure 4.2 gives the value of the spatial quality index Q of the putative sets of matched features for each of the pairs of the images of the laparoscopic dataset individually resulting from the four FDEM schemes. In this figure, the values of Q have been plotted in terms of the pairs of images arranged with the increasing values of the spatial quality of their putative sets obtained using SIFOR. It is seen from this figure that for an overwhelming number of pairs of images, the spatial quality of the putative sets generated by SIFOR is higher than that obtained by using the other three FDEM schemes.

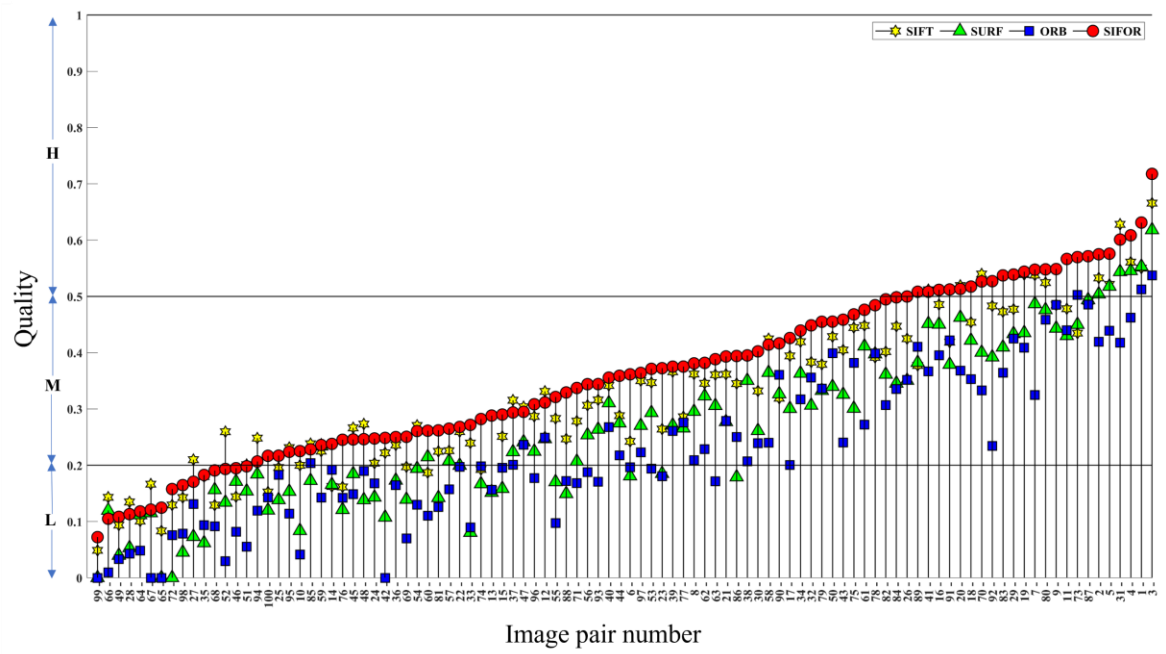
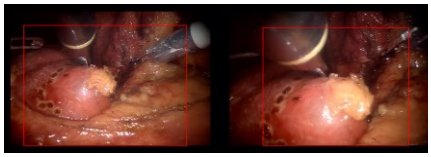


Figure 4.2: Spatial quality Q of the putative sets of matched features generated by the FDEM schemes, SIFT, SURF, ORB and SIFOR for each of the 100 pairs of images of the laparoscopic dataset. L, M, and H represent the ranges of the low, medium, and high spatial quality, respectively, of the putative sets generated by the various schemes.

Table 4.2: Numbers of low, medium, and high spatial quality putative sets of matched features produced by the SIFT, SURF, ORB and SIFOR FDEM schemes using the 100 pairs of images of the laparoscopic dataset.

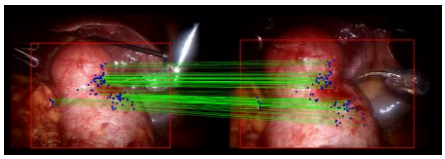
Spatial quality Q	**SIFT** [2]	**SURF** [19]	**ORB** [20]	**SIFOR (Proposed)**
Low	21	41	48	15
Medium	67	53	49	62
High	12	6	3	23

Image pair 65 | Image pair 6 | Image pair 92



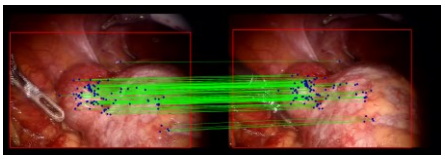
Q1 = 0, Q2 = 0, Q = 0

(Low spatial quality)



Q1 = 0.1047, Q2 = 0.2571, Q = 0.1964

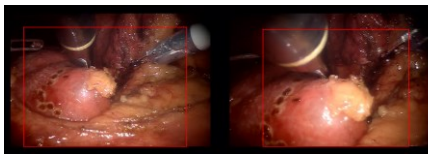
(Low spatial quality)



Q1 = 0.1101, Q2 = 0.3236, Q = 0.2342

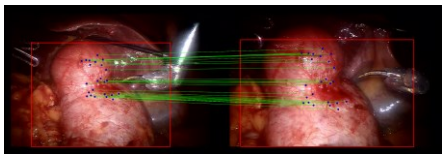
(Medium spatial quality)

(a)



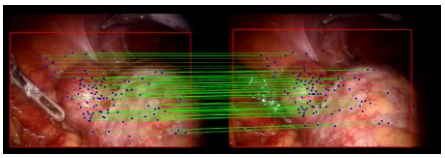
Q1 = 0, Q2 = 0, Q = 0

(Low spatial quality)



Q1 = 0.0267, Q2 = 0.4099, Q = 0.1807

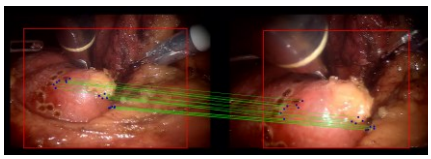
(Low spatial quality)



Q1 = 0.0922, Q2 = 0.7291, Q = 0.3921

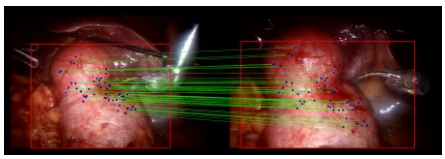
(Medium spatial quality)

(b)



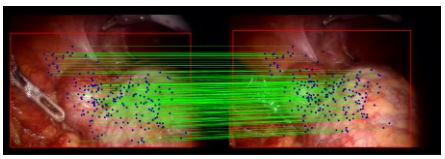
Q1 = 0.0181, Q2 = 0.1606, Q = 0.0834

(Low-quality)



Q1 = 0.0557, Q2 = 0.4554, Q = 0.2424

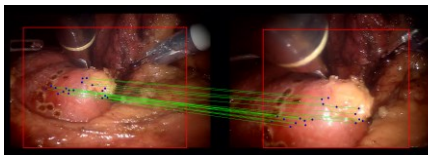
(Medium spatial quality)



Q1 = 0.1964, Q2 = 0.7104, Q = 0.4831

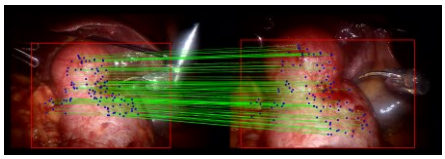
(Medium spatial quality)

(c)



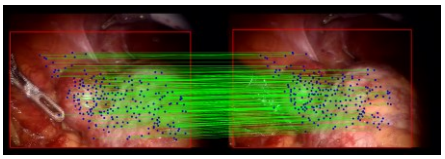
Q1 = 0.0191, Q2 = 0.2782, Q = 0.1246

(Low spatial quality)



Q1 = 0.1180, Q2 = 0.5841, Q = 0.3616

(Medium spatial quality)



Q1 = 0.2182, Q2 = 0.7692, Q = 0.5271

(High spatial quality)

(d)

Figure 4.3: Visual illustration of the true matches in the putative sets provided by the various FDEM schemes for three pairs of images from the laparoscopic image dataset. (a) ORB. (b) SURF. (c) SIFT. (d) SIFOR.

Considering that a putative set to be of low, medium, or high spatial quality if the quality factor lies in the range $[0, 0.2)$, $[0.2, 0.5)$, or $[0.5, 1]$, respectively, it is seen from Figure 4.2 that there are significant number of pairs of images for which the spatial quality of the putative sets generated by SURF and ORB are low. In Table 4.2, we have provided the statistics on the numbers of pairs of images in the laparoscopic dataset whose putative sets resulting from the four FDEM schemes are of low, medium or high spatial quality. It is seen from this table that among the four FDEM schemes, SIFOR provides putative sets with low spatial quality for the lowest number of pairs of images and at the same time it provides the putative sets with high spatial quality for the largest number of pairs of images. In this respect, SIFT provides the next best performance and ORB the lowest performance.

In Figure 4.3, we provide visual illustrations of matching of the features in the putative sets of matched features of some of the selected pairs of the images. For the purpose of this illustration, we have selected one pair of images from each of the categories of the pairs of images whose putative set spatial quality belongs to the low, medium, and high spatial quality categories by using the SIFOR FDEM scheme and used the same pairs of the images to compare the qualities of the putative sets obtained by the four FDEM schemes. It is seen from this figure that for each of these three pairs of images, the spatial quality of the putative set provided by SIFOR is higher than that provided by the other three FDEM schemes. It can be specifically noted from this figure that for the pair of images regarded to be of low spatial quality in the first column, the putative sets obtained by using the SURF and ORB schemes provide no match of the features between the fixed and moving images. On the other hand, the putative set obtained by using the SIFT scheme is able to provide a

certain number of matches of the features, but both the density and the spread of the features are poorer than that of the matched features obtained by SIFOR. It is seen from the pair of images in the middle column that the spatial quality of the putative set provided by SIFOR is far superior to those provided by ORB and SURF, both in terms of the density and the spread of the matches. Even though the spatial quality of the putative set provided by SIFT is superior to those provided by ORB and SURF, this scheme in comparison to SIFOR is not able to provide matches of the features in certain regions of the ROIs in the pair of images. The pair of images chosen in the third column is chosen so as to provide high spatial quality putative set by SIFOR. However, the putative sets generated by the other three schemes for the same pair of images are all of medium quality. Finally, it is seen from the second and third columns of Figure 4.3 that ORB has a larger density of the matches than SURF has, but a much poorer dispersion of the matches in view of the clustering of its matches.

To compare the processing times taken to produce a putative set for a pair of images by the various methods, we return to Table 4.1. It is seen from this table that the average time taken to process a pair of images by SIFOR is slightly more than one-half of that taken by SIFT, the second best performing FDEM scheme, about 1.2 times that taken by ORB, and 1.6 times that taken by SURF. However, in absolute terms, the processing time of 80.14 ms taken by SIFOR is still very small. Moreover, in a real-life situation, the fixed image which is considered to be the reference image is changed only infrequently with the arrival of a new moving frame, therefore the time taken to identify the ROI, feature detection, and feature extraction of the fixed image (reference image) does not need to be included in the processing time of a pair of images. Therefore, in such a scenario, the

processing time of SIFOR can be reduced to 46.4 ms.

4.4.2 Results on the Synthetic-Laparoscopic Image Dataset II

In this subsection, we assess the performance and computational complexity of the FDEM schemes, SIFT, SURF, ORB, and SIFOR, on the synthetic laparoscopic image dataset II, constructed and described earlier in this section. Recall that this dataset has 2,000 distinct synthetic pairs of images constructed from the laparoscopic image dataset. The moving images of these 2,000 pairs have been distorted using 8 corruption models, each using different levels of white Gaussian noise (σ) and motion blurring (M_L, M_θ). In Figure 4.4, we provide the average values of N_M, Q, and T_{FDEM} resulting from the four FDEM schemes per pair of fixed and moving images over 2,000 pairs of the synthetic-laparoscopic image dataset II. The results in Figure 4.4 have been given for 9 different groups, each consisting of 2,000 pairs of fixed and moving images. The moving images in the pairs of images in group 1 are not distorted by noise nor by motion blurring. The moving images in groups 2 and 3 are distorted only by noise with $\sigma = 0.1$ and $\sigma = 0.2$, respectively. The moving images in groups 4, 5, and 6 are distorted only with motion blurring with $M_L = 10$, and M_θ = 0, M_θ = 45, and M_θ = 90, respectively. The moving images in groups 7, 8, and 9 are distorted only with motion blurring with $M_L = 20$, and M_θ = 0, M_θ = 45, and M_θ = 90, respectively.

It is seen from the results for the pairs of images in groups 1, 2, and 3 in Figure 4.4 that even though the average values of N_M (m_{N_M}) resulting from each of the four FDEM schemes are sensitive to noise, SIFOR and ORB are relatively more robust to the levels of noise. From the results for the pairs of images in groups 4 to 9, it is seen that the values of

m_{N_M} resulting from each of the four schemes are again sensitive to motion blurring, with ORB showing the highest sensitivity and SIFOR the least sensitivity among the four FDEM schemes. It is also seen that m_{N_M} is more sensitive to the motion length, M_L, than to the motion angle, M_θ.

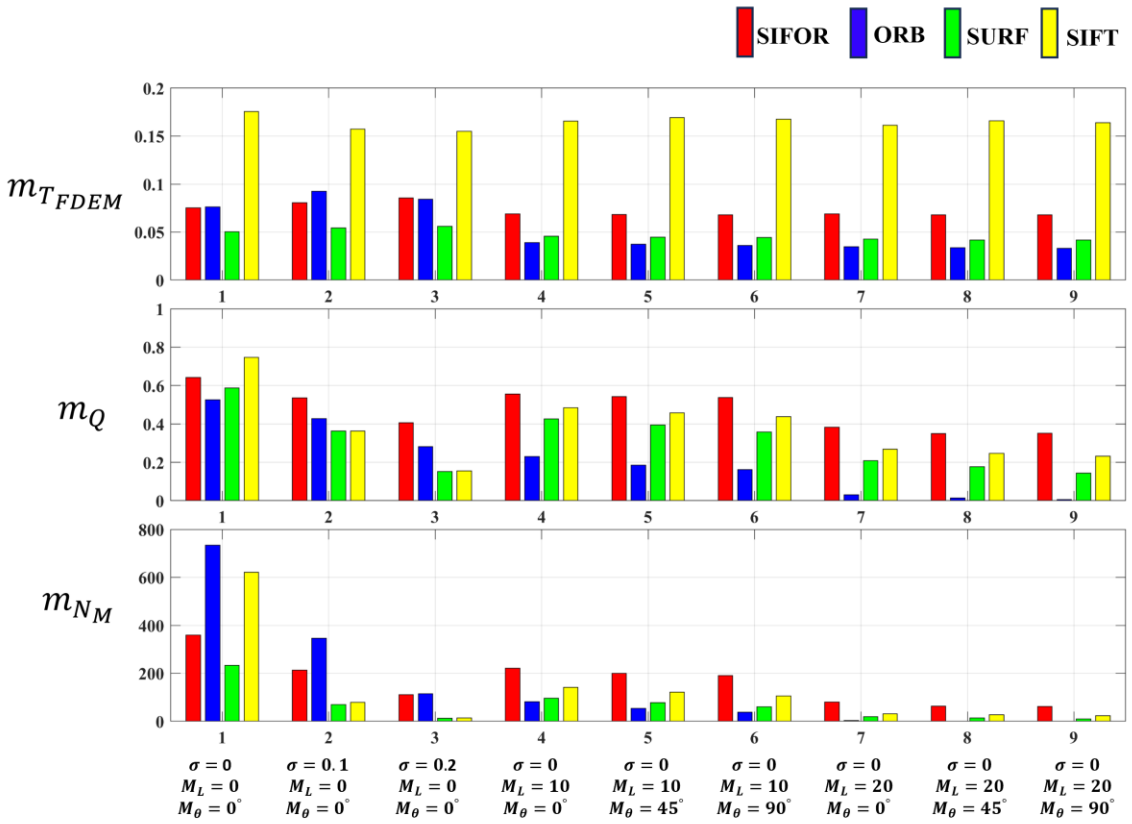


Figure 4.4: Average number of matched features (m_{N_M}), average values of the spatial quality of a putative set (m_Q), and the average time ($m_{T_{FDEM}}$) taken to produce it by the various FDEM schemes per pair of fixed and moving images in the synthetic-laparoscopic image dataset II of 2,000 pairs of images in which the moving images are distorted with 9 different corruption models.

It is to be noted that SIFOR, even for a large motion length of 20, provides much higher value for m_{N_M}, in comparison to the other three FDEM schemes. With regard to the spatial quality of the putative sets generated, it is seen from the results of group 1 (i.e., for the pairs of images in which the moving images are neither noise corrupted, nor motion blurred), the putative sets generated by SIFT, SURF, SIFOR and ORB result in high values for m_Q, that is, all the four schemes are able to generate good spatial quality putative sets. However, it should be noted that the only distortion between the fixed and moving images

of the pairs of images in group 1 is non-linear deformation, and as soon as other distortions (noise and motion blurring) are introduced, it is seen from the results of groups 2 to 9 that SIFOR provides the best values for m_Q among the four FDEM schemes. It is also to be noted that in the presence of noise distortion (groups 2 and 3), ORB generates putative sets having the second-best average spatial quality, and in the presence of motion blurring (groups 4 to 9), SIFT generates putative sets having the second-best average spatial quality, where the average spatial quality of the putative sets generated by ORB is the lowest.

It is seen from Figure 4.4 that in terms of the average processing time, $m_{T_{FDEM}}$, SIFT is computationally the most expensive FDEM scheme. The average processing time of SIFOR is generally less than one half of that of SIFT, and it is generally larger than that of the other two schemes in situations when the noise or motion blurring are present. However, it is important to note that in such situations, the average spatial quality of the putative sets generated by SIFOR is the best. Thus, taking both the spatial quality of the putative sets and the computational time into consideration, SIFOR stands out to be an attractive FDEM scheme.

4.5 MIS Image Registration Using the Putative Set of Matched Features Obtained from the Proposed FDEM Scheme

To show the efficacy of the proposed FDEM scheme in generating high quality putative sets, in this section, we apply the putative sets of matched features so generated for the task

of registration of MIS images. In image registration, given a pair of fixed and moving images, the geometry of the moving image is transformed to that of the fixed image. To accomplish this task, an image transformation T is obtained, which when applied to the pixel locations of the moving image, produces a registered image whose geometry is as close to that of the fixed image as possible.

In MIS, tissue surfaces are always affected by significant deformation due to the patient motion, breathing, heartbeat, and interaction with the surgical instruments [16]. Due to the non-rigid deformation between the fixed and moving images in MIS, for the purpose of the registering the moving image to the fixed image, a transformation that is able to capture the non-rigid deformation between the pair of images is employed. The term deformable image registration (DIR) is generally used in the literature to refer to this type of registration. Among the DIR techniques, feature-based registration schemes are vey attractive, in view of their being generally invariant to geometric and radiometric differences between the reference and moving images and being not sensitive to the initial conditions on the transformation parameters and to large deformation. For a feature-based DIR, first, a putative set of matched features between the reference and moving images is obtained using an FDEM scheme, and then, this set is used to estimate the transformation parameters and the resulting transformation is applied to the moving image to obtain the corresponding registered image. Given a pair of fixed and moving images, the accuracy of a feature-based registration is very much dependent on the matching and the spatial qualities of the putative sets generated. There are several well-known feature-based registration schemes that exist in the literature [60]. In this section, we chose the thin-plate spline (TPS) [55] method for conducting non-rigid registration.

Let $S_M = \{(P_i, P_i'),\ i = 1,\ \ldots,\ N\}$ be the putative set, that is, the set of locations of the matched pairs of features between a pair of fixed and moving images, resulting from the application of an FDEM scheme, where P_i and P_i' denote the coordinates of the i^{th} matched features in the two images, respectively. The parameters of the transformation T are estimated using TPS, such that $T(P_i') \approx P_i,\ i = 1,\ \ldots,\ N$. The estimated transformation is then applied to all the pixel points in the ROI of the moving image to obtain the corresponding locations in the registered image. In order to measure the quality of a registered image, we need to measure the error between the corresponding pixel points of the registered and reference (fixed) images. As mentioned in Section 4.4, the laparoscopic image dataset provides partial ground-truth mapping data for each of the 100 pairs of fixed and moving images. Let $S_G = \{(G_i, G_i'),\ i = 1,\ \ldots, N_G\}$ be the set of the locations of the N_G ground-truth mapping data for a pair of fixed and moving images, where G_i and G_i' denote the coordinates of the i^{th} ground-truth matched points in the two images of a pair, respectively. Let $G_{ri}, i = 1,\ \ldots, N_G$ be the pixel locations in the registered image corresponding to ground truth data for a pair of images. In order to measure the accuracy of the registration, we compute the target registration error (TRE), denoted by, E, given by

$$E = \frac{1}{N_G} \|G_i - G_{ri}\|_2 \tag{4.12}$$

where $\|\boldsymbol{x}\|_2$ denotes the L_2-norm of the vector $\boldsymbol{x}$, as a measure of the registration accuracy.

Table 4.3: The registration results using putative sets for the pairs of images of the laparoscopic dataset generated by applying the FDEM schemes, SIFT, SURF, ORB and SIFOR.

	SIFT [2]	**SURF** [19]	**ORB** [20]	**SIFOR (Proposed)**
Results with original refined putative sets				
$\boldsymbol{m_{N_M}}$	105.0	57.5	145.7	129.1
$\boldsymbol{m_Q}$	0.3244 (2, 3)	0.2654 (3, 5)	0.2327 (4, 6)	0.3599 (1, 1)
$\boldsymbol{N_s}$	100	95	95	100
$\boldsymbol{m_E}$	3.6260 (2, 3)	4.6830 (3, 5)	4.8234 (4, 6)	3.2440 (1, 1)
$\boldsymbol{m_{T_{reg.}}}$**(ms)**	176.1	97.5	235.1	217.9
Results with sampled putative sets				
$\boldsymbol{\delta_{N_M}}$ **,** $\boldsymbol{d}$	99, 28	NA	95, 24	72, 29
$\boldsymbol{m_{N_M}}$	55.6	57.5	54.9	57.0
$\boldsymbol{m_Q}$	0.3086 (2, 4)	0.2654 (3, 5)	0.2300 (4, 7)	0.3258 (1, 2)
$\boldsymbol{m_E}$	3.6987 (2, 4)	4.6830 (3, 5)	4.88 (4, 7)	3.3751 (1, 2)
$\boldsymbol{m_{T_{reg.}}}$**(ms)**	97.4	97.5	97.3	97.5

In Table 4.3, we provide the average registration results over all the 100 pairs of the laparoscopic dataset, in which the value of N_G for its various pairs of fixed and moving

images ranges between 12 and 27. The first row of the results in this table gives the average number of matches in the original refined putative sets generated by each of the four FDEM schemes, SIFT, SURF, ORB and SIFOR. The second row of the results gives the values of the average spatial quality (m_Q) of the putative sets of matched features generated by the various schemes. The third row contains the number (N_s) of pairs of images in the dataset for which, the task of registration could be performed successfully by employing the putative sets obtained from each of the four FDEM schemes. It is to be noted that in order to carry out registration successfully using TPS, there must be at least three non-colinear matched feature points in the putative set. It is seen from the third row of the table that both SURF and ORB have five pairs of images whose putative sets do not satisfy this condition, and hence, their corresponding moving images of these pairs cannot be registered, whereas the moving images of all the 100 pairs can be successfully registered using the putative sets resulting from SIFT and SIFOR. The fourth and fifth rows in this table give the average values of the target registration error (m_E), and the average values of the times ($m_{T_{reg.}}$) required to perform the registration by employing the putative sets obtained by using the four FDEM schemes. It is seen from the results of the fourth row of the table that SIFOR has the lowest registration error. By comparing the registration errors obtained by using the putative sets of the various FDEM schemes, it is seen that there is a clear correlation between these errors and the corresponding values of the spatial qualities of the putative sets, as provided in the second row of the table. Therefore, the putative set of SIFOR, which has the highest average spatial quality results in the lowest average target registration error, whereas the putative set of ORB with the

lowest average spatial quality has the highest average target registration error. By comparing the results in the first and fifth rows of Table 4.3, it is seen that the average registration time $m_{T_{reg.}}$ (i.e., the average time required to both estimate the transformation function and apply it to all the pixel locations in the ROI of a moving image) is dependent on the number of matches. Hence, the registration time using the putative set of an FDEM scheme can be decreased by using a subset of the original putative set with reduced number of matches. However, as the number of matches in the set is reduced to decrease this registration time, it will generally have a negative impact on the registration accuracy.

We now present a simple scheme for sampling the matches in a putative set and construct a subset of the original putative set consisting of only a certain number of sampled matches and use this subset to perform the registration. For this purpose, if the number of matches in the putative set of a pair of images is larger than a threshold $\delta_{N_{TM}}$, then we partition the rectangular bounding box of the ROI in the moving image into square blocks each of size $d \times d$, and if a square block has only one match, we include that particular match in the sampled putative set, and if it has more than one match, then we select from that square block only one single match that has the largest similarity between its features. The resulting reduced-sized sampled putative set of the pair of images is then used for registration of the moving image to the fixed image of the pair. Note that this method of sampling the matches may result in improving the spread of the matches, if the original putative set has clusters of matches. The reduction in the registration time is achieved at the expense of increasing the registration error, TRE, by using a reduced-spatial-quality sampled putative set, which depends on the values of the parameters $\delta_{N_{TM}}$

and d chosen for sampling the original putative set obtained from a given FDEM scheme. It is seen from the fifth row of the results in Table 4.3 that SURF provides the lowest registration time of 97.5 ms by using its original unsampled putative sets of 95 pairs of images of the laparoscopic dataset. In order to study the impact of using a reduced-sized putative set of an FDEM scheme on the registration error, we choose the values of the sampling parameters $\delta_{N_{TM}}$ and d optimally so that the reduced putative set resulting from the original putative set obtained from the FDEM scheme, SIFT, ORB, or SIFOR, would allow us to use a maximum possible average registration time not exceeding of 97.5 ms, which is the lowest average registration time using unsampled putative sets among all the FDEM schemes, that is, the average registration time using the original unsampled putative set of SURF. The sixth row of the results in Table 4.3 gives the optimal values of the sampling parameters $\delta_{N_{TM}}$ and d, to obtain the reduced-sized putative sets corresponding to the original putative sets resulting from SIFT, ORB and SIFOR. The seventh row of the results in this table provides the average number of matches in the reduced-sized putative sets. It is seen from this row that the variance between the average numbers of the matches in the sampled putative sets is much smaller than that of the average number of matches, as given in the first row, for the unsampled putative sets. This is not surprising in view of the fact that the sizes of all the original putative sets have been optimally reduced with a constraint to consume approximately the same registration time of 97.5 ms. The eighth row of the results in this table gives the average spatial quality of the sampled putative sets. In rows 2 and 8 of the results in Table 4.3, we have ranked the average qualities of the different putative sets resulting from the various schemes by using a pair of numbers (a, b),

where a is a number that represents the rank of the spatial quality of an unsampled (sampled) putative set among the four unsampled (sampled) sets, whereas b is a number representing the rank of the spatial quality of a putative set among all the seven putative sets, sampled or unsampled. The value of $a = 1,2,3,$ or 4 in rows 2 and 8 represents the rank of the average spatial quality of a putative set among the four unsampled (sampled) putative sets resulting from the four FDEM schemes, with $a = 1$ indicating the highest spatial quality and $a = 4$ the lowest spatial quality. The value of the second number in the pair $b = 1, 2, 3, ...,$ or 7 in rows 2 and 8 represents the rank of the average spatial quality of a putative set among all the seven putative sets, sampled or unsampled, with $b = 1$ indicating the highest spatial quality and $b = 7$ the lowest spatial quality. It is seen that, as expected, the average spatial quality of the sampled putative sets is lower than that of the corresponding unsampled putative sets. However, the rank of the average spatial quality of the sampled putative sets resulting from an FDEM scheme remains unchanged from that of the corresponding unsampled putative sets. It is seen from the rankings of the putative sets that the average spatial quality of unsampled putative sets as well as that of the sampled putative sets resulting from SIFOR is the highest (i.e., $a = 1$). It is also important to note that the rank of the average spatial quality of the sampled putative sets resulting from SIFOR is $b = 2$, indicating that the average spatial quality of the sampled putative sets resulting from SIFOR is higher than the average spatial quality of even the unsampled putative sets resulting from any of the FDEM schemes, SIFT, SURF or ORB. The results in rows 4 and 9 of Table 4.3 provide the average target registration errors using the unsampled and sampled putative sets, respectively, resulting from the four FDEM

schemes along with their rankings using a pair of numbers (c, d) with the same meaning attached to the values of c and d as that attached to a and b for the ranking of the spatial quality of the putative sets. It is seen from the results of rows 4 and 9 that the average target registration errors using the sampled putative sets resulting from each of the four FDEM schemes have, as expected, increased, but their values are still correlated to the average qualities of the respective sampled putative sets, with the lowest average registration error provided by the sampled putative set generated by SIFOR. Finally, it is to be noted that the average target registration error using the sampled putative sets resulting from SIFOR is lower than that when the registration is performed using even the unsampled putative sets provided by any of the three schemes SIFT, SURF or ORB.

4.6 Summary

In this chapter, we have proposed a low complexity FDEM scheme utilizing the good properties of three well known FDEM schemes, namely, SIFT, SURF and ORB, so as to provide putative sets that have good matching quality as well as good spatial quality. We have proposed a novel metric to measure the spatial quality of a set of matched features and used the same for designing the proposed FDEM scheme, SIFOR. The proposed FDEM scheme has been extensively experimented with using two datasets. The first one is a real laparoscopic dataset, and the second one is a synthetic-laparoscopic dataset constructed from the real laparoscopic dataset. The purpose of constructing the second dataset is to generate pairs of images in which the moving images differ from the corresponding fixed (or reference) images in terms of different amounts of noise and

motion blurring, and to use them to study the impact of such differences between the two images in the pairs of images on the performance of the proposed FDEM scheme. It has been shown that the proposed FDEM scheme generates putative sets of matches using laparoscopic dataset with a quality superior to that of the putative sets produced by SIFT, whose quality is higher than those of SURF and ORB, at a computational time that is approximately one half of that of SIFT. The results on the synthetic dataset have shown that SIFOR performs remarkably better than the other three schemes do, under the condition that the fixed and moving images differ in terms of noise and motion blurring. Finally, to demonstrate the efficacy of our proposed SIFOR scheme, we have used the generated putative sets in the application of MIS image registration. Experimental results have shown that the registered images obtained by using the putative sets generated by SIFOR have the lowest average target registration error.

Chapter 5
Conclusion

5.1 Concluding Remarks

This thesis has been concerned with the problems of feature detection, extraction, and matching (FDEM) and feature matching refinement (FMR). FDEM is a process in which, given a pair of fixed and moving images, certain distinctive features are detected from the pair, then they are suitably represented as feature vectors, and finally, the corresponding feature vectors are compared and matched leading to a set of matched features known as a putative set for the pair. On the other hand, FMR is a process in which the falsely matched pairs of features are, as much as possible, removed from a putative set. FDEM and FMR schemes are the cornerstone of many medical and non-medical applications, such as robotic-assisted minimally invasive surgery, disease diagnosis, autonomous driving, and surveillance and security. MIS images undergo deformation, occlusions, specular reflection, and have non-distinctive features due to the repetitive textures in the tissues. All these factors make the tasks of FDEM and FMR very challenging. The existing FDEM and FMR schemes are computationally intensive and/or have poor performance, especially for MIS images in which the above challenges are more difficult to overcome.

In this thesis, robust and fast schemes for generation of matched features in MIS images have been developed. For this purpose, in the first part of the thesis, a very fast and accurate FMR scheme for MIS images has been proposed that is robust to inliers ratios of

the original putative set. Then, in the second part of the thesis, a fast and accurate FDEM scheme for MIS images has been proposed that takes advantage of the existing FDEM schemes to generate a set of putative matched features that has a good overall quality, spatially as well as in terms of the matching accuracy.

In the first part of the thesis, a novel two-stage NSB FMR scheme for MIS images, referred to as VSLD-FMR, which addresses the problems of existing FMR schemes, has been proposed. In the first stage of the scheme, a conservative approach has been adopted by choosing small circular neighborhoods in the fixed image for each feature of this image that belongs to the putative set. This approach of forming neighborhoods enables them to be better local neighborhoods than those formed by using the K-NN based approach. Then, a voting mechanism has been devised for the matches within a neighborhood based on the number and similarity of the displacement vectors of the matches within the neighborhood. After the voting process, those matches that receive a large number of votes are identified to be true matches. Using the knowledge of true matches gained in the first stage of the scheme, a mechanism has been developed in the second stage of the scheme to determine the status of those matches in the putative set whose status have not yet been determined in the first stage. In the second stage, larger neighborhoods of the same size are formed around each feature point in the fixed image corresponding to the pairs whose status is still unknown, and the displacement vector of the feature point in the question is compared with those of the feature points in this larger neighborhood whose status are known to be true from the first stage. Extensive experiments on feature matching refinement using the proposed and nine state-of-the-art schemes on three different datasets including real, synthetic, and phantom MIS images have been performed. The experimental results have

shown that the proposed VSLD-FMR scheme provides a performance, which is second to none, at an extremely low computational cost, regardless of the inliers ratio of the original putative set.

In the second part of the thesis, a fast and accurate FDEM scheme, referred to as SIFOR, has been proposed for detection, extraction, and matching of features in MIS images by making a strategic use of the strong attributes of the SIFT, SURF and ORB FDEM schemes, so as to provide a putative set of matched features for a pair of images that is of both a good matching quality and a good spatial quality. For developing the proposed SIFOR FDEM scheme, we have proposed and used a novel metric to measure the spatial quality of a set of matched features. We believe that the spatial quality of a putative set is as critical as its matching quality. The spatial quality of the matched features refers to the density of the matched features in the regions of interest as well as to how well they are dispersed over these regions. Therefore, a good spatial quality is vital for tasks such as 3D reconstruction and registration. To the best of our knowledge, there is no metric in the literature that measures the spatial quality of a putative set. We have developed such a metric to measure the spatial quality of a putative set of matched features. The proposed SIFOR FDEM scheme has been extensively experimented using real and synthetic laparoscopic image datasets. The experimental results on the real laparoscopic image dataset have shown that the proposed FDEM scheme generates putative sets of matches with a quality superior to that of the putative sets produced by SIFT, whose quality is higher than those of SURF and ORB, at a computational time that is approximately one half of that of SIFT. The results on the synthetic dataset have shown that, under the condition that the fixed and moving images differ in terms of noise and motion blurring,

SIFOR performs remarkably better than the other three schemes. Finally, we have used the putative sets generated by the various FDEM schemes in the application of MIS image registration. Experimental results have demonstrated the efficacy of the proposed SIFOR scheme, as the registered images obtained using the putative sets generated by SIFOR exhibit the lowest average target registration error.

5.2 Scope for Future Work

While new deep learning-based schemes for finding the matched features are recognized for their potential advantages in accuracy and robustness, this thesis has focused on exploring and improving established techniques. These methods are known for their efficient real-time performance and well-established reliability, which are critical in applications using minimally invasive surgery (MIS) images. The decision of excluding the deep learning-based approaches in developing our FDEM schemes has been because of the requirement of large training datasets and extensive computational resources by such approaches. Yet, in our proposed FDEM scheme, by strategically combining the approaches of well-established non-data-driven and introducing of the idea of spatial quality of the matched features, we have proposed an FDEM scheme with a very high performance at a low processing time. For future work, it would be worth exploring FDEM approaches that combine the strengths of deep learning-based approaches, integrating the proposed feature matching refinement (FMR) technique and the proposed spatial quality metric. The goal should be to guide a lightweight deep network model to provide a performance higher than that provided by the existing deep leaning-based techniques or even higher than that provided by our proposed FDEM scheme, SIFOR. This thesis has

provided two large-scale synthetic MIS image datasets in which images in each pair differ by varying amounts of noise, and blurriness in addition to deformations. These datasets could be used for the training of the deep learning models for the task of FDEM.

References

[1] A. Goshtasby, *Theory and Applications of Image Registration*. Hoboken, NJ, USA: John Wiley & Sons, 2017.

[2] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis*., vol. 50, no. 2, pp. 91-110, 2004.

[3] T. Haidegger, "Autonomy for surgical robots: concepts and paradigms," *IEEE Trans. Med. Robot. Bionics*, vol. 1, no. 2, pp. 65-76, May 2019.

[4] A. Leporini *et al*., "Technical and functional validation of a teleoperated multirobots platform for minimally invasive surgery," *IEEE Trans. Med. Robot. Bionics*, vol. 2, no. 2, pp. 148-156, May 2020.

[5] L. Qian, J. Y. Wu, S. P. DiMaio, N. Navab and P. Kazanzides, "A review of augmented reality in robotic-assisted surgery," *IEEE Trans. Med. Robot. Bionics*, vol. 2, no. 1, pp. 1-16, Feb. 2020.

[6] C. Girerd, A. V. Kudryavtsev, P. Rougeot, P. Renaud, K. Rabenorosoa and B. Tamadazte, "Automatic tip-steering of concentric tube robots in the trachea based on visual SLAM," *IEEE Trans. Med. Robot. Bionics*, vol. 2, no. 4, pp. 582-585, Nov. 2020.

[7] Y. Liu *et al*., "Real-time robust stereo visual SLAM system based on bionic eyes," *IEEE Trans. Med. Robot. Bionics*, vol. 2, no. 3, pp. 391-398, Aug. 2020.

[8] J. Song, J. Wang, L. Zhao, S. Huang and G. Dissanayake, "MIS-SLAM: Real-time large-scale dense deformable SLAM system in minimal invasive surgery based on heterogeneous computing," *IEEE Robot. Autom. Lett*, vol. 3, no. 4, pp. 4068-4075, Oct. 2018.

[9] M. N. Cheema *et al*., "Image-aligned dynamic liver reconstruction using intra-operative field of views for Minimal Invasive Surgery," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 8, pp. 2163-2173, Aug. 2019.

[10] H. Zhou and J. Jagadeesan, "Real-time dense reconstruction of tissue surface from stereo optical video," *IEEE Trans. Med. Imag.*, vol. 39, no. 2, pp. 400-412, Feb. 2020.

[11] P. Vagdargi *et al*., "Real-time 3-D video reconstruction for guidance of transventricular neurosurgery," *IEEE Trans. Med. Robot. Bionics*, vol. 5, no. 3, pp. 669-682, Aug. 2023.

[12] V. Penza, Z. Cheng, M. Koskinopoulou, A. Acemoglu, D. G. Caldwell and L. S. Mattos, "Vision-guided autonomous robotic electrical bio-impedance scanning system for abnormal tissue detection," *IEEE Trans. Med. Robot. Bionics*, vol. 3, no. 4, pp. 866-877, Nov. 2021.

[13] S. Zhang, L. Zhao, S. Huang, M. Ye and Q. Hao, "A template-based 3D reconstruction of colon structures and textures from stereo colonoscopic images," *IEEE Trans. Med. Robot. Bionics*, vol. 3, no. 1, pp. 85-95, Feb. 2021.

[14] D. Stoyanov, M. Visentini-Scarzanella, P. Pratt, and G. Z. Yang, “Real-time stereo reconstruction in robotic assisted minimally invasive surgery,” in *Proc. 10th Int. Conf. Med. Image Comp. Comp-Asst. Intervent.*, 2010.

[15] G. A. Puerto-Souza, J. A. Cadeddu and G. Mariottini, “Toward long-term and accurate augmented-reality for monocular endoscopic videos,” *IEEE Trans. Biomed. Eng.*, vol. 61, no. 10, pp. 2609–2620, Oct. 2014.

[16] M. C. Yip, D. G. Lowe, S. E. Salcudean, R. N. Rohling and C. Y. Nguan, “Tissue tracking and registration for image-guided surgery,” *IEEE Trans. Med. Imag.*, vol. 31, no. 11, pp. 2169–2182, Nov. 2012.

[17] T. Bergen and T. Wittenberg, “Stitching and surface reconstruction from endoscopic image sequences: A review of applications and methods,” *IEEE Journal Biomed. Health Informatics*, vol. 20, no. 1, pp. 304–321, Jan. 2016.

[18] H. Zhou and J. Jayender, “Real-time nonrigid mosaicking of laparoscopy images,” *IEEE Trans. Med. Imag.*, vol. 40, no. 6, pp. 1726–1736, June 2021.

[19] H. Bay, A. Ess, T. Tuytelaars and L. Gool, “Speeded-up robust features (SURF),” *Comput. Vis. Image Understand.*, vol.110, No.3, 2008, pp. 346–359, 2008.

[20] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “ORB: An efficient alternative to SIFT or SURF,” in *Proc. ICCV*, 2011, pp. 2564–2571.

[21] J. Sun, Z. Shen, Y. Wang, H. Bao and X. Zhou, “LoFTR: Detector-free local feature matching with transformers,” in *Proc. CVPR*, 2021, pp. 8918–8927.

[22] Q. Wang, J. Zhang, K. Yang, K. Peng and R. Stiefelhagen, “Matchformer: Interleaving attention in transformers for feature matching,” in *Proc. ACCV*, 2022, pp. 256–273.

[23] H. Chen, Z. Luo, L. Zhou, Y. Tian, M. Zhen, T. Fang, D. McKinnon, Y. Tsin and L. Quan, “Aspanformer: Detector-free image matching with adaptive span transformer,” in *Proc. ECCV*, 2022, pp. 20–36.

[24] J. Ni, Y. Li, Z. Huang, H. Li, H. Bao, Z. Cui and G. Zhang, “PATS: Patch area transportation with subdivision for local feature matching,” in *Proc. CVPR*, 2023, pp. 17776–17786.

[25] D. DeTone, T. Malisiewicz and A. Rabinovich, “SuperPoint: Self-supervised interest point detection and description,” in *Proc. CVPRW*, 2018, pp. 337–349.

[26] P. -E. Sarlin, D. DeTone, T. Malisiewicz and A. Rabinovich, “SuperGlue: Learning feature matching with graph neural networks,” in *Proc. CVPR*, 2020, pp. 4937–4946.

[27] H. Chen, Z. Luo, J. Zhang, L. Zhou, X. Bai, Z. Hu, C.-L. Tai and L. Quan, “Learning to match features with seeded graph matching network,” in *Proc. ICCV*, 2021, pp. 6281–6290.

[28] P. Lindenberger, P. -E. Sarlin and M. Pollefeys, “LightGlue: Local feature matching at light speed,” in *Proc. ICCV*, 2023, pp. 17581–17592.

[29] Q.H. Tran, T.J. Chin, G. Carneiro, M.S. Brown, and D. Suter, “In defence of RANSAC for outlier rejection in deformable registration,” in *Proc. ECCV*, 2012, pp. 274–287.

[30] G. A. Puerto-Souza and G. Mariottini, “A fast and accurate feature-matching algorithm for minimally-invasive endoscopic images,” IEEE Trans. Med. Imag., vol. 32, no. 7, pp. 1201–1214, July 2013.

[31] H. Zhou and J. Jayender, "EMDQ: Removal of image feature mismatches in real-time," *IEEE Trans. Image Process.*, vol. 31, pp. 706-720, 2022.

[32] J. Ma, J. Zhao, J. Tian, A. L. Yuille, and Z. Tu, "Robust point matching via vector field consensus," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1706–1721, Apr. 2014.

[33] J. Ma, J. Zhao, J. Jiang, H. Zhou, and X. Guo, "Locality preserving matching," *Int. J. Comput. Vis.*, vol. 127, no. 5, pp. 512–531, 2019.

[34] J. Ma, X. Jiang, J. Jiang, J. Zhao, and X. Guo, "LMR: Learning a two-class classifier for mismatch removal," *IEEE Trans. Image Process.*, vol. 28, pp. 4045–4059, Aug. 2019.

[35] X. Jiang, J. Ma, J. Jiang and X. Guo, "Robust feature matching using spatial clustering with heavy outliers," *IEEE Trans. Image Process.*, vol. 29, pp. 736–746, 2020.

[36] G. Wang, and Y. Chen, "Robust feature matching using guided local outlier factor," *Pattern Recognit.*, vol. 117, 2021, 107986.

[37] M. Brown and D. G. Lowe, "Invariant features from interest point groups," in *British Machine Vision Conference*, Cardiff, Wales, 2002, pp. 656–665.

[38] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Proc. ECCV*, 2006, pp. 430–443.

[39] E. Rosten, R. Porter, and T. Drummond, "Faster and better: a machine learning approach to corner detection," *IEEE Trans. Pat. Ana. Mach. Intel.*, vol. 32, no. 1, pp. 105–119, Jan 2010.

[40] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary robust independent elementary features," in *Proc. ECCV*, 2010, pp. 778–792.

[41] M. A. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[42] A. N. Tikhonov, and V. Y. Arsenin, *Solutions of Ill-posed Problems*. Washington, DC, USA: Winston, 1977.

[43] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. KDD*, 1996, pp. 226–231.

[44] M. R. Pourshahabi, M. O. Ahmad and M. N. S. Swamy, "A Very Fast and Robust Method for Refinement of Putative Matches of Features in MIS Images for Robotic-Assisted Surgery," *IEEE Trans. Med. Robot. Bionics*, vol. 6, no. 2, pp. 419–432, May 2024.

[45] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Commun. ACM*, vol. 18, No. 9, 1975, pp 509–517.

[46] G. A. Puerto-Souza and G. Mariottini, Laparoscopic dataset [Online] Accessed: 2018, http://ranger.uta.edu/gianluca/feature_matching (2014)

[47] Hamlyn Centre laparoscopic/endoscopic video datasets [Online] Accessed: 2018, http://hamlyn.doc.ic.ac.uk/vision (2012)

[48] A. Goshtasby, "Image registration by local approximation methods," *Image Vis. Computing* 6, no. 4, November 1988, pp 255–61.

[49] P. Pratt, D. Stoyanov, M. Visentini-Scarzanella, and G. Z. Yang, "Dynamic guidance for robotic surgery using image-constrained biomechanical models," in *Proc. 10th Int. Conf. Med. Image Comp. Comp-Asst. Intervent.*, 2010.

[50] M. Misawa, *et al*., "Development of a computer-aided detection system for colonoscopy and a publicly accessible large colonoscopy video database (with video)," Gastrointestinal Endoscopy, Vol. 93, Issue 4, pp. 960–967.e3, 2021.

[51] H. Itoh et al., 2020, "SUN colonoscopy video database," Dataset, sundatabase, Accessed: 2023. [Online]. Available: http://amed8k.sundatabase.org/

[52] H. Borgli, *et al.,* "*HyperKvasir*, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy," *Sci Data* **7**, 283, 2020.

[53] H. Borgli et al., 2020, "Hyperkvasir—The largest gastrointestinal dataset," Dataset, simula. Accessed: 2023. [Online]. Available: https://datasets.simula.no/hyper-kvasir/

[54] M. R. Pourshahabi, M. O. Ahmad and M.N.S. Swamy, "SIFOR: A Robust Scheme for Detection, Extraction, and Matching of Features in MIS Images for Robotic-Assisted Surgery," currently under review.

[55] F. L. Bookstein, "Principal warps: thin-plate splines and the decomposition of deformations," *IEEE Trans. Pat. Ana. Mach. Intel.*, vol. 11, no. 6, pp. 567–585, June 1989.

[56] C. Harris and M. Stephens, "A combined corner and edge detector," in Alvey Vision Conference, 1988, pp. 147–151.

[57] P. J. Clark and F.C. Evans, "Distance to nearest neighbor as a measure of spatial relationships in populations," *Ecology*, vol. 35, no. 4, pp. 445–453, 1954.

[58] K. P. Donnelly, "Simulations to determine the variance and edge effect of total nearest neighborhood distance," in *Simulation studies in archeology,* ed. I. Hodder, Cambridge, NY, USA: Cambridge University Press, 1978, pp. 91–95.

[59] T. J. Oyana and F. M. Margai, *Spatial analysis: statistics, visualization, and computational methods*. Boca Raton, FL, USA: Taylor & Francis Group, 2016.

[60] L. Zagorchev and A. Goshtasby, "A comparative study of transformation functions for nonrigid image registration," *IEEE Trans. Image Process.*, vol. 15, no. 3, pp. 529–538, March 2006.