Study of Computational Image Matching Techniques: Improving Our View of Biomedical Image Data

by

Ahmadreza Baghaie

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Engineering

at

The University of Wisconsin–Milwaukee

December 2016

Abstract

Study of Computational Image Matching Techniques: Improving Our View of Biomedical Image Data

by

Ahmadreza Baghaie

The University of Wisconsin–Milwaukee, 2016
Under the Supervision of Professor Zeyun Yu and Professor Roshan M. D'Souza

Image matching techniques are proven to be necessary in various fields of science and engineering, with many new methods and applications introduced over the years. In this PhD thesis, several computational image matching methods are introduced and investigated for improving the analysis of various biomedical image data. These improvements include the use of matching techniques for enhancing visualization of cross-sectional imaging modalities such as Computed Tomography (CT) and Magnetic Resonance Imaging (MRI), denoising of retinal Optical Coherence Tomography (OCT), and high quality 3D reconstruction of surfaces from Scanning Electron Microscope (SEM) images. This work greatly improves the process of data interpretation of image data with far reaching consequences for basic sciences research. The thesis starts with a general notion of the problem of image matching followed by an overview of the topics covered in the thesis. This is followed by introduction and investigation of several applications of image matching/registration in biomdecial image processing: a) registration-based slice interpolation, b) fast mesh-based deformable image registration and c) use of simultaneous rigid registration and Robust Principal Component Analysis (RPCA) for speckle noise reduction of retinal OCT images. Moving towards a different notion of image matching/correspondence, the problem of view synthesis and 3D reconstruction, with a focus on 3D reconstruction of microscopic samples from 2D images captured by SEM, is considered next. Starting from sparse feature-based matching techniques, an extensive analysis is provided for using several well-known feature detector/descriptor techniques, namely ORB, BRIEF, SURF and SIFT, for the problem of multi-view 3D reconstruction. This chapter con-

tains qualitative and quantitative comparisons in order to reveal the shortcomings of the sparse feature-based techniques. This is followed by introduction of a novel framework using sparse-dense matching/correspondence for high quality 3D reconstruction of SEM images. As will be shown, the proposed framework results in better reconstructions when compared with state-of-the-art sparse-feature based techniques. Even though the proposed framework produces satisfactory results, there is room for improvements. These improvements become more necessary when dealing with higher complexity microscopic samples imaged by SEM as well as in cases with large displacements between corresponding points in micrographs. Therefore, based on the proposed framework, a new approach is proposed for high quality 3D reconstruction of microscopic samples. While in case of having simpler microscopic samples the performance of the two proposed techniques are comparable, the new technique results in more truthful reconstruction of highly complex samples. The thesis is concluded with an overview of the thesis and also pointers regarding future directions of the research using both multi-view and photometric techniques for 3D reconstruction of SEM images.[1]

[1]Because of the vast extent of topics covered in this dissertation, each chapter is written in a self-contained manner for a more rigorous presentation of the works.

To

my parents,

siblings,

and my beloved wife, Maral.

Table of Contents

1 Introduction **1**

2 Image Registration in Biomedical Image Processing with Some Applications **9**

2.1 Curvature-Based Registration for Slice Interpolation of Medical Images . 10
2.1.1 Method . . . 12
2.1.2 Results and Discussion . . . 16
2.1.3 Conclusion . . . 22
2.2 Sparse and Low Rank Decomposition Based Batch Image Alignment for Speckle Reduction of Retinal OCT Images . . . 22
2.2.1 Method . . . 25
2.2.2 Results and Discussion . . . 28
2.2.3 Conclusion . . . 29
2.3 Fast Mesh-Based Medical Image Registration . . . 31
2.3.1 Method . . . 32
2.3.2 Results And Discussion . . . 36
2.3.3 Conclusion . . . 39

3 Sparse Feature-Based Matching for 3D Surface Reconstruction of Multi-View Scanning Electron Microscope Images **41**

3.1 Introduction . . . 41
3.2 Methods . . . 45
3.2.1 SIFT . . . 45
3.2.2 SURF . . . 47
3.2.3 BRIEF . . . 48
3.2.4 ORB . . . 50
3.3 3D Surface Reconstruction of Electron Microscopy Images . . . 51
3.3.1 Scanning Electron Microscope Imaging . . . 51
3.3.2 Feature Detectors/Descriptors for 3D SEM Surface Reconstruction 52
3.4 Experimental Results . . . 56
3.4.1 Experimental Setup . . . 56
3.4.2 Qualitative 3D Visualization . . . 58
3.4.3 SEM Extrinsic Calibration . . . 58
3.5 Discussion . . . 61
3.6 Conclusion . . . 63

4 Sparse-Dense Correspondence for High Quality 3D Reconstruction of Microscopic Samples **65**

4.1 Introduction . . . 65
4.2 Methods . . . 69
4.2.1 SEM Imaging Protocol . . . 70
4.2.2 Scale Invariant Feature Transform (SIFT) . . . 71
4.2.3 Epipolar Rectification . . . 73
4.2.4 SIFT-Flow for Dense Correspondence . . . 75

4.2.5 Disparity Refinement: Bilateral Filtering . . . 79
4.2.6 Depth Estimation . . . 80
4.3 Results & Discussions . . . 81
4.4 Conclusion . . . 88

5 3D Reconstruction of Highly Complex Microscopic Samples Using Non-Local Optical Flow Estimation 95
5.1 Introduction . . . 95
5.2 Methods . . . 99
5.2.1 Overview . . . 99
5.2.2 SEM Imaging Protocol . . . 100
5.2.3 SIFT Feature Detection/Matching and Epipolar Rectification . . . 100
5.2.4 Dense Matching by Optical Flow Estimation . . . 103
5.2.5 Disparity Refinement by Weighted Median Filter . . . 107
5.2.6 Depth Estimation . . . 108
5.3 Results & Discussions . . . 109
5.4 Conclusions . . . 118

6 Conclusions and Future Works 122

Bibliography 128

Appendices 152

Appendix A: Publications 153

Appendix B: Honors and Awards 156

Curriculum Vitae 157

List of Figures

2.1 An illustration of linear displacements between corresponding points utilized in this section. The main goal is to use the top and bottom slices to reconstruct the in-between slice. 13
2.2 Top row: results of slice interpolation with proposed method. Bottom row: results of slice interpolation with linear interpolation method. 16
2.3 (a) Three consecutive slices. The first and third images are used for interpolation with parameters set as: $\tau = 0.03, \alpha = 100$ (b) Top row: results of interpolation for linear and proposed methods respectively. Bottom row: difference images for the results. (c) Close-up of the results of linear and proposed method. (d) Optimized displacement fields in horizontal and vertical directions. 17
2.4 Results of placing 3 slices between two input slices (first and third slices from Fig.2 (a)) using proposed method (Top row) and linear interpolation method (Bottom row) 18
2.5 (a) Three consecutive slices. The first and third images are used for interpolation with parameters set as: $\tau = 0.05, \alpha = 100$ (b) Top row: results of interpolation for linear, non-modified and proposed methods respectively. Bottom row: difference images for the results. (c) Close-up of the results of linear, non-modified and proposed method respectively. 19
2.6 (a) Sample retinal OCT image degraded by speckle noise, (b) Final result of the proposed method using 50 misaligned noisy retinal OCT images . 24
2.7 SNR improvement for different number of input images (5-50) 29
2.8 CNR improvement for different number of input images (5-50) 30
2.9 A portion of one of the input images (a) and the results of ImageJ translation (b), ImageJ rigid (c) and RASL rigid registration based methods (d). 30
2.10 Example of content adaptive mesh generation 37
2.11 (a) Template image, (b) Reference image, (c) Difference image 38
2.12 (a) Displacement fields in horizontal and vertical directions, (b) Registered image, (c) Difference image after registration 39

3.1 Pipeline of 3D surface reconstruction from Electron Microscopy images. At first, multiple images of the specimen from different perspectives are taken. Then, we estimate the relative position including rotation and translation based on the matching points in the image set. After estimating the image motions, the 3D position of all corresponding points would be reconstructed using linear triangulation. The final step is doing an optimization process to find the best match for SEM's extrinsic parameters and all of the initial 3D points. 55
3.2 (A) Shows a set of 2D images from *tapetal cell* of *Arabidopsis thaliana*, which is obtained by tilting the specimen stage 9 degrees from one to the next in the image set. (B) Shows five images from TEM *copper grid*. These images were obtained by tilting the specimen stage 7 degrees from one to the next in the image sequence. The white circle specifies a part of the specimen which will be 3D reconstructed by the proposed framework. . . 56

3.3 Qualitative visualization of 3D SEM surface reconstruction of *tapetal cell* of *Arabidopsis thaliana* using different image feature detector algorithms including SIFT, SURF, BRIEF, and ORB. 58
3.4 Qualitative visualization of 3D SEM surface reconstruction of TEM *copper grid* using different image feature detector algorithms including SIFT, SURF, BRIEF, and ORB. 59

4.1 Overview of the proposed method for high fidelity 3D reconstruction of microscopic samples using a pair of stereo SEM images captured by tilting the specimen stage: (a) a set of stereo SEM images of a *Tapetal Cell* with known tilting angle (9 degrees), (b) result of sparse SIFT feature detection/description with *a contrario* RANSAC approach for outlier removal, (c) set of rectified images with horizontal epipolar lines being demonstrated, (d) bilateral filtered horizontal disparity map, (e) a magnified view of the high quality surface mesh generated using the dense point cloud, and (f) a magnified view of the high fidelity surface model. The proposed method is able to reconstruct the 3D geometry of the microscopic sample with high accuracy. 69
4.2 SEM imaging procedure used for this study. 71
4.3 Factor graph representation of the energy minimization functional with de-coupled horizontal and vertical components 76
4.4 Effects of various spatial ($\sigma_s = \{1, 3, 5\}$, from left to right) and range ($\sigma_r = \{1, 3, 5\}$, from top to bottom) variance for bilateral filtering of the disparity of *Pollen Grain*. The difference map between the initial disparity map and the refined map are also presented. 80
4.5 Relationship between the estimated height (h) and the computed horizontal disparity (d) using the pixel size in sample units (p) and the total tilt angle (θ). 81
4.6 Dense matching results for the rectified image sets: *Tapetal Cell* (column 1), *Copper Bar* (column 2), *Copper Grid* (column 3), *Hexagonal Grid* (column 4) and *Pollen Grain* (column 5). The first row shows the initial difference map. The second row shows the minimization trend for the optimization process defined using dense SIFT descriptors, factor graph representation of the energy functional to be optimized (Figure 4.3) and loopy belief propagation as means of optimization. The third row displays the difference maps after the optimization process. 84
4.7 Qualitative visualization of the proposed 3D SEM reconstruction framework for the *Copper Bar* sample images, acquired by tilting the sample stage by 11 degrees. The set of two-view images can be seen in Table 5.1. Second row displays several views of the reconstructed dense point cloud. The initial cloud contains 196608 points which is sub-sampled here for better visualization. Third row shows the constructed triangular surface mesh. Fourth row depicts a magnified view of the constructed triangular surface mesh. Considering the proposed dense correspondence framework, a highly uniform reconstruction of the curved surface is achieved. 89

4.8 Qualitative visualization of the proposed 3D SEM reconstruction framework for the *Copper Grid* sample images, acquired by tilting the sample stage by 7 degrees. The set of two-view images can be seen in Table 5.1. Second row displays several views of the reconstructed dense point cloud. The initial cloud contains 1228800 points which is sub-sampled here for better visualization. Third row shows the constructed triangular surface mesh. Fourth row depicts a magnified view of the constructed triangular surface mesh. Considering the proposed dense correspondence framework, a highly uniform reconstruction of the curved surface is achieved. 90
4.9 Qualitative visualization of the proposed 3D SEM reconstruction framework for the *Hexagonal Grid* sample images, acquired by tilting the sample stage by 10 degrees. The set of two-view images can be seen in Table 5.1. Second row displays several views of the reconstructed dense point cloud. The initial cloud contains 1228800 points which is sub-sampled here for better visualization. Third row shows the constructed triangular surface mesh. Fourth row depicts a magnified view of the constructed triangular surface mesh. Considering the proposed dense correspondence framework, a highly uniform reconstruction of the curved surface is achieved. 91
4.10 Qualitative visualization of the proposed 3D SEM reconstruction framework for the *Pollen Grain* sample images, acquired by tilting the sample stage by 3 degrees. The set of two-view images can be seen in Table 5.1. Second row displays several views of the reconstructed dense point cloud. The initial cloud contains 447665 points which is sub-sampled here for better visualization. Third row shows the constructed triangular surface mesh. Fourth row depicts a magnified view of the constructed triangular surface mesh. Considering the proposed dense correspondence framework, a highly uniform reconstruction of the curved surface is achieved. 92
4.11 Qualitative comparison between state-of-the-art sparse feature based reconstruction [188, 187] and the proposed sparse-dense reconstruction approach for *Tapetal Cell* (column 1), *Copper Bar* (column 2), a portion of *Copper Grid* (column 3), *Hexagonal Grid* (column 4) and *Pollen Grain* (column 5) micrograph sets. Close inspection of the results reveals superiority of the proposed sparse-dense reconstruction methodology: uniformly distributed mesh nodes, ability to truthfully recover edges and very fine details, consistency over the domain of images. 93

5.1 SEM imaging procedure used for this study. 101
5.2 Relationship between the estimated height (h) and the computed horizontal disparity (d) using the pixel size in sample units (p) and the total tilt angle (θ). 109
5.3 Optical flow estimation results for (a) *Arabidopsis Anther 1*, (b) *Arabidopsis Anther 2*, (c) *Graphene*, (d) *Pseudoscorpion* and (e) *Ash* sample sets. The first row shows the initial difference maps. The second row shows the computed optical flow estimate. Using the optical flow estimate, the first image in each pair is warped and then used for generating the final difference maps as depicted in the third row. It should be noted that the images for *Pseudoscorpion* set are rotated by 90° for visualization purposes. 113

5.4 Effects of weighted median filtering on the horizontal disparity map: the first input micrograph that is used as guidance (left column), before (center column) and after (right column) disparity refinement. Despite inclusion of non-local term in the optical flow energy functional, the outcome can be improved greatly by adding an additional weighted median filtering step. 114
5.5 Comparison of the results for *Graphene*: The first row is the overall as well as a zoomed region of the computed disparity map using the state-of-the-art method of [187] which uses sparse feature-based matching approach and *a contrario* RANSAC for outlier removal. The dense disparity map is created by scattered data interpolation of the sparse disparity values. The second row shows the result of Horn/Schunck optical flow estimation ([85]), which provides a better estimation of the disparity map than that of [187]. The third row shows the result of dense feature matching proposed in [116] which uses dense SIFT features as well as factor graph representation of the matching energy functional optimized by loopy belief propagation. Even though relatively better than [85], the result still suffers from blurred edges. The result of the proposed method is presented in the fourth row. In comparison to the state-of-the-art, the proposed approach generates a sharper and more accurate disparity map. 115
5.6 Comparison of the results for *Ash*: The first row is the overall as well as a zoomed region of the computed disparity map using the state-of-the-art method of [187] which uses sparse feature-based matching approach and *a contrario* RANSAC for outlier removal. The dense disparity map is created by scattered data interpolation of the sparse disparity values. The second row shows the result of Horn/Schunck optical flow estimation ([85]), which provides a better estimation of the disparity map than that of [187]. The third row shows the result of dense feature matching proposed in [116] which uses dense SIFT features as well as factor graph representation of the matching energy functional optimized by loopy belief propagation. Even though relatively better than [85], the result still suffers from blurred edges. The result of the proposed method is presented in the fourth row. In comparison to the state-of-the-art, the proposed approach generates a sharper and more accurate disparity map. 116
5.7 3D red-cyan anaglyphs generated by combining the two rectified stereo views of the microscopic samples as well as the solid 3D models created using Meshmixer ([171]) for (a) *Arabidopsis Anther 1*, (b) *Arabidopsis Anther 2*, (c) *Graphene*, (d) *Pseudoscorpion* and (e) *Ash* sample sets 119
5.8 From start to finish: (a) first image from the *Ash* sample set, (b) 3D solid model generated using the computed disparity estimates and modified using MeshMixer ([171]), (c) 3D printed model. Using the proposed approach, highly complex structure of the sample was captured and reconstructed in the printed model. 120

6.1 Stereo matching results for *Graphene* (first row) and *Ash* (second row) micrograph sets using Semi-Global stereo (left) and Patch-Match stereo (right) with occlusion detection. 127

6.2 Shape, illumination, and reflectance estimation from shading using only one image from the *Copper Bar* set by using the method proposed in [22]. From left to right, the initial image as well as shape, normals, reflectance, shading and illumination. 127

List of Tables

2.1 Comparison of the average MSDs for brain image database (as well as the improvement rates) for the linear, non-modified and the proposed slice interpolation approaches. 21
2.2 Computational time and mean MSD error for pixel-based and mesh-based registration methods 39

3.1 Experimental setup including image sets, and SEM configuration. 57
3.2 Accuracy and reliability validation of the different image feature detector algorithms by employing the proposed 3D SEM surface reconstruction framework. ΔR is given as $R_{real} - R_{estimated}$, indicating error for estimating the 3D rotation. Rotation angles show the ground truth 3D SEM rotations(R_{real}). In each row we used only two images in the set. 62

4.1 Summary of the dataset used in this work. The micrographs are acquired from *Tapetal Cell*, *Copper Bar*, *Copper Grid*, *Hexagonal Grid* and *Pollen Grain* using a Hitachi S-4800 field emission scanning electron microscope (FE-SEM) with sizes ranging from 512×384 to 1280×960 and tilt angles in the range $3 - 11°$. 71
4.2 Summary of rectification results using sparse SIFT matching and the subsequent epipolar rectification. The first and second row in the table indicate the number of SIFT features found in the first and second micrographs of each set, while the third row is the result of *a contrario* methodology for matching the SIFT features according to a homography transform. Fourth and fifth row show the computed homography transformation matrices as results of epipolar rectification by minimizing the Sampson's error. Finally, initial and final Sampson rectification errors are presented. 83
4.3 Summary of dense correspondence results using dense SIFT features, the factor graph representation of the objective function and loopy belief propagation as meas for optimization. The first and second row represent the initial and final root mean squared error (RMSE) of the two input micrographs. The residual errors can be attributed to the noise contained in the micrographs as well as the differences in brightness due to edge effects caused by imaging in the secondary electron (SE) mode. third and fourth rows show the initial and final values of the objective function (note the coefficients $\times 10^9$ and $\times 10^7$). The fifth row shows the ratio between the energy contained in the vertical disparity map and the energy contained in the horizontal disparity map. This provides additional proof for the efficiency of the rectification process as well as the depth estimation step. The last row displays the computational time needed for finding the dense correspondence between input micrographs. 85

5.1 Dataset acquired using a Hitachi S-4800 field emission scanning electron microscope (FE-SEM) by tilting the specimen stage by 7°. The images for *Pseudoscorpion* set are rotated by 90° for visualization purposes. 101

5.2 Rectification results: number of SIFT points found in each input image (rows 1 and 2), number of matching points after *a contrario* RANSAC (row 3), initial and final rectification errors from before and after the quasi-Euclidean rectification (rows 4 and 5). As can be seen, despite careful image acquisition, the initial rectification errors are large. 111

Acknowledgements

First of all I would like to express my gratitude to my advisers, Prof. Zeyun Yu of Departments of Electrical Engineering and Computer Science, University of Wisconsin-Milwaukee and Prof. Roshan M. D'Souza of Department of Mechanical Engineering, Univeristy of Wisconsin-Milwaukee for supporting me during the completion of my PhD study and research. Their insightful thoughts and advises guided me during my study at University of Wisconsin-Milwaukee and the writing of this PhD dissertation.

I would like to thank Prof. Heather A. Owen of Department of Biological Sciences, University of Wisconsin-Milwaukee for her great helps and encouragements during the completion of my thesis. Also, many thanks to Dr. Pahlavan Tafti of Marshfield Clinic Research Foundation for his great help and insight.

I would like to acknowledge Dr. Joseph Carroll of Advanced Ocular Imaging Program (AOIP), Medical College of Wisconsin (MCW) for providing the data and insight for retinal OCT image processing.

Also, I would also like to thank my thesis committee members, Prof. Ethan V. Munson of Department of Computer Science, University of Wisconsin-Milwaukee , Prof. Jun Zhang of Department of Electrical Engineering, University of Wisconsin-Milwaukee and Prof. Ramin Pashaie of Department of Electrical Engineering, University of Wisconsin-Milwaukee. I am gratefully indebted to them for the very valuable comments on this thesis.

Finally, I must express my very profound gratitude to my parents, my brother and sister and my beloved wife, Maral, for their unending support and encouragement. This accomplishment would not have been possible without them. Thank you.

Chapter 1

Introduction

Image matching techniques have been a crucial component in modern general-purpose and biomedical image processing problems. Generally speaking, the matching can be defined as the *attempt* to find a *meaningful correspondence* between *pixels/features* of an image with another. As one would think, this definition is very general, however, it can be used as means to dissect the problem of image matching and determine the various components involved. In another words, various image matching can be differentiated by the answers that they give to these simple questions: a) what constitutes an *attempt*? b) what is considered as a *meaningful correspondence*? c) matching of pixels or features?

Image matching techniques can be categorized into four major classes [219]: 1) multiview analysis techniques, 2) multitemporal analysis techniques, 3) multimodal analysis techniques and 4) scene to model matching techniques. In the first class, several images of the same scene are captured, although from different viewpoints. Matching techniques in this class aim to create a larger two dimensional view of the scene or a three dimensional representation of the scene given the underlying projective transformations involved. Examples of such class include image mosaicing and also Shape from Stereo (SfS) and Shape from Motion (SfM). On the other hand, in multitemporal analysis, images of the same scene are captured at different times in order to track and evaluate changes in the scene between consecutive acquisitions. Examples may include landscape planning and global land usage in remote sensing, motion tracking and optical flow estimation [1] in computer vision applications, and monitoring of tumor evolution in medical image processing. In the third class, even though the images are captured from the same scene, however, different sensors are employed for achieving a more complex and detailed representation.

[1] Optical flow estimation does not necessarily requires fixed viewpoint. However, given its general definition which involves estimating motions of intensity patterns, it can be considered partly as a multitemporal analysis problem.

Fusion of the information from these different sensors usually requires pre-processing steps for accurate alignment of the images. Examples can be found in medical image applications for combining anatomical imaging modalities (Magnetic Resonance Imaging (MRI) or Computer Tomography (CT)) with functional/metabolic activity images captured by Positron Emission Tomography (PET), Single Photon Emission Computed Tomography (SPECT) etc. In the last class, images of a scene and a model of the scene are matched. Examples are target template matching in computer vision applications or comparison of anatomical atlases with patient's images or specimen classification.

Image matching is implemented using two major classes of methods:: a) *Parametric* and b) *Non-parametric* [134]. Parametric matching methods are based on a finite set of parameters or image features. These methods include rigid, affine, land-mark based, principal axes-based, FFT-based, optimal linear and spline-based matching approaches, to name a few. Unlike parametric matching methods, the non-parametric methods are not limited to a finite set of parameters. Diffusion registration, fluid registration, curvature registration and elastic registration [58] as well as optical flow estimation techniques [62] and stereo matching approaches [169] are a few examples of this class of methods.

In general, image matching is considered as an *ill-posed inverse problem*. Therefore the process of solving the problem may consist three components [178]: 1) a deformation model, 2) an objective function to be optimized and 3) an optimization method. A general objective function for matching a template image T to a reference image R can be defined as:

$$E[\mathbf{u}] = D[R, T \circ \mathbf{u}] + \alpha S[\mathbf{u}] \tag{1.1}$$

The left hand side of the equation is the energy or objective function which needs to be optimized; $\mathbf{u}$ is the displacement field. On the right, we have two terms: D and S. The first term is called *(dis)similarity measure* or *distance measure* which acts as the matching criterion between the reference image and the deformed template image. Depending on the choice of this term, the objective function needs to be either minimized or maximized [151]. The second term, S is the *regularization term* which imposes additional constraints on the deformation. Due to ill-posedness of image registration, this

term is needed to be able to have reasonable transformations. This term dictates the validity of transformation. The parameter α is a weight which indicates the amount of regularization. Even though the above formulation is very general, all of image matching methods can be formulated in a same form, implicitly or explicitly. In this PhD dissertation, several problems are considered which all involve some form of image matching at their very core.

Chapter 2 of the dissertation introduces and discusses several applications of image matching in biomedical image processing. Image interpolation/super-resolution is a well-known topic in image processing. In its basic form image interpolation tries to upsample the image to an image with higher resolution. The upsampling can be done differently along the horizontal and vertical directions of the image, which results is expanding or shrinking of the image along the corresponding direction. This is usually the case for medical image processing since due to physical limitations, the resolutions of the captured images are different along different imaging axes, especially in three dimensional (3D) images. For example in modern imaging modalities like MRI or CT, to build a 3D volume scan, multiple 2D slices are captured and then combined. However, the resolution is not the same along the different axes. Usually the resolution is much higher within the plane of 2D images while it is lower along the third dimension. When building 3D models form these scans, the asymmetry in the resolution leads to step-shaped iso-surfaces and discontinuities. This calls for advanced computational techniques to increase the resolution of the volume scans. Assuming MRI or CT images in which the resolution is excellent within the image planes, only interpolation is required along the third dimension. Therefore the problem is called *slice interpolation* since additional slices are needed to be placed between the slices that are already captured. Even though general purpose interpolation techniques (e.g. nearest neighbor, linear, cubic interpolation etc.) can be employed, the results of such methods suffer from jagged and blurring effects near object boundaries. The remedy can be sought in a class of techniques usually referred to as object-based methods which use the extracted information from objects contained in input images as guidance in the process of interpolation. Image matching/registration

based interpolation techniques are examples of such class. Section 2.1 of the dissertation introduces and investigates a novel registration-based slice interpolation method which uses deformable image registration as means for more accurate slice interpolation.

Recently, Optical Coherence Tomography (OCT) has emerged as a powerful technique for obtaining detailed 3D volumetric images of sub-surface tissue. One of the highly regarded fields which takes advantage of the technology is ophthalmology in which taking μm-resolution volume scans using OCT is commonplace. Therefore, the need for sophisticated image processing approaches for dealing with the high amount of data captured has grown in the past two decades, including image matching/registration techniques. Several different applications of using image registration approaches may include noise reduction, multimodal retinal image registration, image mosaicing for extending the field-of-view and involuntary eye motion reduction. Section 2.2 will focus on the problem of noise reduction in OCT image data by taking advantage of Robust Principal Component Analysis (RPCA). The procedure involves simultaneous image matching as well as sparse/low-rank decomposition of the image data into signal/noise components which is proven to provide high accuracy and noise-free results.

Dealing with deformable image registration problems can be computationally challenging since the aim is to find displacement vectors for all of the pixels contained in the images. One solution can be in the use of computers with higher computational power. This could include use of parallel programming by taking advantage of Graphics Processing Units (GPUs) [118]. However, algorithmic optimizations will result in better efficiency. Multi-resolution implementation of the optimization process can be considered as a useful improvement widely found in the literature. In such implementations, the optimization starts from a very coarse grid for capturing larger deformations at first and then moves to finer resolutions to capture smaller deformations. Another solution can be sought in use of adaptive triangular meshes rather than the regular uniform grids employed in multi-resolution techniques. This will improve the representation of the images since the objects and features are not distributed uniformly within the images. Of course this requires additional implications for optimization of the energy functional needed for

image registration. Section 2.3 discusses the idea of mesh-based registration more and provides further details on the implementations of such technique.

As mentioned before, one major class of image matching methods takes the problem of multiview analysis into consideration. In such techniques multiple images of the same scene from different viewpoints are captured and used for the purpose of pixel/feature matching, projective transformation estimation and finally 3D reconstruction of the scene. However, in this dissertation, the general case of 3D reconstruction from 2D images is not considered. Instead, the use of such techniques in 3D reconstruction of microscopic samples is investigated. For this, the Scanning Electron Microscope (SEM) is used as means of capturing high resolution images of micro-structures for the purpose of 3D reconstruction. SEM and its diverse applications have been a very active research area over the recent decades, and scientific studies have covered the use of SEM in many domains ranging from biomedical applications to materials sciences and nano technologies. The SEM is an advanced microscopy device that produces high quality images of microscopic specimen using a focused beam of electrons which can be then captured by two types of detectors, the secondary electron (SE) and the back-scattered electron (BSE), to provide both compositional and/or geometrical information about the microscopic surface. Despite having high resolution, the SEM micrographs still remain 2D. Therefore there is a need for advanced computational methods for revealing the third dimension. As will be discussed in greater detail in the rest of the dissertation, SEM 3D surface reconstruction techniques can be divided into three major classes: a) single-view approaches, b) multi-view approaches and c) hybrid approaches. In single-view approaches, using varying lighting (electron beam) directions on a single perspective, a group of 2D SEM micrographs are captured and used for 3D SEM surface modeling. In multi-view approaches a set of 2D SEM images from different perspectives assists the 3D SEM surface reconstruction process. The hybrid mechanisms try to combine single-view and multi-view algorithms to restore a 3D shape model from 2D SEM images. In single-view 3D surface reconstruction, creating a full model of the microscopic sample is not possible since the images are limited to only one view-point. Moreover, recreating the SEM mi-

crographs of the sample under different illumination conditions is difficult. On the other hand, multi-view approaches offer a more general and achievable framework for the task. Therefore, here, the focus will be on the use of multi-view approaches and this requires use of specifically designed matching techniques.

Depending on the matching technique used, the methods of multiview SEM 3D surface reconstruction can be categorized into two major classes: a) sparse feature-based approaches and b) dense pixel-based approaches. While methods from the first class are employed to establish a set of robust matches between an image pair or a set of images based on sparsely distributed distinct feature-points, dense multiview techniques try to discover matches for all pixels in the images. These matches along with other computational methods will then be used to accurately estimate the projective geometry and 3D surface models. Chapter 3 tries to explore the sparse feature-based class by employing four well-known feature detector/descriptor widely used in the computer vision literature, namely SIFT, SURF, BRIEF and ORB. In each case, at first, distinct features are detected in the set of multiview SEM images which are later described by considering the features' neighborhoods, in a manner specific to each method. These features are then matched between multiple micrographs by employing several steps of optimization and outlier removal in order to enable accurate estimation of fundamental matrix and extrinsic calibration matrices. This is followed by 3D point cloud generation which can be used for the final goal of 3D surface generation. Even though extensive comparisons are representative of superiority of SIFT feature detector/descriptor for the purpose of 3D reconstruction, however, the final outcome will be highly affected by the level of features contained in the images. In other words, the results are not consistent between different image sets. This is mainly due to the sparse distribution of feature points within the image domain which results in a very smooth reconstruction of the surface with many fine details missed in the process.

Chapter 4 tries to build on the result of Chapter 3, using the SIFT method, but with a major twist as the aim is to use such technique not only for sparse feature points, but instead for all the pixels in the images. In this chapter, a novel framework using sparse-

dense correspondence is introduced and investigated for 3D reconstruction of stereo SEM images. After imaging the microscopic samples, the process begins by sparse feature detection/description using SIFT. Using the detected features and after one step of naive matching using nearest neighbor search, *a contrario* RANSAC approach is employed for eliminating the outliers and fundamental matrix approximation. This is later used for rectifying the input pair. The rectification process will cause the displacements to be more concentrated along the horizontal direction. This step enables a simplified 3D reconstruction, as the depth will be proportional to the disparities of each pixel between the multiview images. This is followed by finding the dense correspondence between individual pixels of the rectified input images using the dense SIFT descriptors created for all of the pixels. However, the implementation of such matching the same way as discussed in the previous chapter is not computationally efficient due to high number of pixels. Therefore, more efficient approaches must be employed. For this, the energy functional defined for matching is represented as a factor graph and loopy belief propagation is used for optimization. Given the fact that the input images are rectified, the horizontal disparities can be used for depth approximation. Employing the proposed framework, a more accurate depth estimation can be achieved when compared with sparse feature-based approaches. The results are more consistent with sharper boundaries and less smoothing effects.

In Chapter 5, the same framework is improved by considering a more accurate formulation of the matching energy functional. Even though both formulations perform similarly when having minimal variations in the microscopic surface, when dealing with samples with higher complexities, the performance of the previous approach degrades. Of course, it is still much better than sparse feature-based approaches as will be discussed more later. However, improving the dense matching process as well as a more guided final post-processing is shown to increase the accuracy of the framework greatly.

Given the above overview, the contributions of the dissertation can be summarized as follows:

- The problem of slice interpolation in biomdecial image processing is introduced and

a new deformable image registration-based approach is proposed.

- Noise reduction of retinal OCT images, as an application of image matching, is investigated in detail and a novel approach for simultaneous alignment and signal/noise decomposition of the OCT data is introduced.

- To tackle the computational demands of deformable image registration approaches, the concept of mesh-based registration is explored and the implications of such formulation are investigated.

- Moving towards the main focus of the thesis, the problem of 3D reconstruction of microscopic samples from sets of multiview SEM micrographs is investigated in great detail. To address the issue, both sparse featured-based approaches and dense pixel/descriptor-based approaches are considered.

- A new framework for sparse-dense correspondence for high quality 3D surface reconstruction of microscopic samples is introduced and compared with the state-of-the-art in the field. Employing the proposed framework, higher accuracy levels are achieved in comparison to the sparse feature-based approaches and more surface details are revealed.

- The proposed framework is improved by considering more accurate dense matching techniques which makes it more suitable for microscopic samples with higher complexity levels. Moreover, a new approach for depth map refinement is introduced and investigated. The proposed framework provides an end-to-end pipeline for the researchers in the field, from image acquisition to pre/post-processing, to quantitative analysis of surface attributes of the microscopic samples, and finally, to 3D printing of high fidelity physical models, as the ultimate tangible means of representation.

Chapter 2

Image Registration in Biomedical Image Processing with Some Applications

Medical Image Registration is an active area in the field of image processing with applications ranging from image mosaicing in retinal images [147] to slice interpolation [16] etc. Image registration problems can be categorized into four major categories: *multi-view analysis, multitemporal analysis, multimodal analysis* and *scene to model registration* [219]. Generally speaking, given two images, reference R and template T, the image registration problem is to find a *valid* and *optimal* spatial or geometrical transformation between the two input images. In the process, the pixel values of template image will not change and only the locations will be altered. The range of applications of image registration is vast. Here, several applications of image registration is investigated for slice interpolation and noise reduction of retinal Optical Coherence Tomography (OCT) data. Since the problem of image registration needs optimization in its core to ensure a valid and robust displacement field, the computational complexity can be an issue. Therefore a new algorithm for mesh-based image registration in also introduced and validated here. Section 2.1 introduces and investigates the problem of slice interpolation using deformable image registration. Section 2.2 shows another example of image registration approaches for noise reduction of retinal OCT images. Finally, Section 2.3 aims to tackle the problem of computational complexity of deformable registration by introducing mesh-based registration.

2.1 Curvature-Based Registration for Slice Interpolation of Medical Images

Image interpolation is a well-known research topic in image processing [17] and there have been many studies in this area, especially, in bio-medical applications. With modern image modalities (CT, MRI, light/electron microscopy, etc.), a sequence of 2D images can be provided and used in building 3D models [20, 49]. However, the resolutions of the images are often not identical in all three directions. Usually the resolution in the Z direction is significantly lower than the resolutions in the X and Y directions. For example, in a generic CT, image-plane (X-Y) spatial resolution is of the order of 0.5-2mm. On the other hand, the resolution normal to the image plane (Z) is of the order of 1-15mm. This asymmetry in the resolution causes problems such as step-shaped iso-surfaces and discontinuity in structures in 3D reconstructed models. Therefore utilizing a slice interpolation algorithm to augment the 3D data into a symmetric one is of high demand.

In general, slice interpolation methods can be divided into two groups: intensity-based interpolation, and object-based interpolation. In the first category, the final result of interpolation is directly computed from the intensity values of input images. Linear and cubic spline interpolation methods are two examples of this group. The major advantages of these methods are their simplicity and low computational complexity, which lead to their wide uses in practice. As the final result is basically a weighted average of input images, these methods suffer from blurring effects on object boundaries, yielding unrealistic and visually unappealing results.

On the other hand in object based methods, the extracted information from objects contained in input images is used to guide the interpolation leading to more accurate results. There are many methods proposed in the literature which take into account additional information of objects in order to provide better results. One of the first attempts for object-based interpolation has been made by Goshtasby et al [71]. Using a gradient magnitude based approach, corresponding points between consecutive slices

are found and then the linear interpolation is applied in order to find the in-between slices. An important assumption of this work is that the difference between consecutive slices is small, so they restrict their search for finding correspondence points to small neighborhoods. It is obvious that this assumption is not true in many cases. To reduce the blurriness of edges, some more recent approaches have been studied, including the column fitting interpolation [83], shape-based method [72], morphology-based method [105], and feature-guided shape interpolation method [104]. A comprehensive summary of common methods (both intensity-based and object-based) for slice interpolation was described in [73, 74].

An increasingly important group of approaches for object-based image interpolation is based on image registration. Using the well-known free form deformation non-rigid registration method by Rueckert et al. [165], Penny et al. [150] proposed a registration based method for slice interpolation. Another registration based method was given by Frakes et al. [63] by using a modified version of control grid interpolation (CGI). More recently, Leng et al. [106] described a multi-resolution registration based method for slice interpolation. In general, registration-based slice interpolation methods are guided by two important assumptions. First, the consecutive slices contain similar anatomical features. Second, the registration method is capable of finding the appropriate transformation map to match these similar features. Violation of any of these assumptions results in false correspondence maps, which leads to incorrect interpolation results.

Here, a novel method is developed for slice interpolation by taking into account the well-known curvature-based registration [58, 134]. With a modified version of the registration method and an assumption of having linear movement between corresponding points in given slices, a displacement field is computed and the in-between slice is interpolated using a simple averaging of the registration results. The detail of the proposed method is given in Section 2.1.1, followed by some experimental results along with quantitative and qualitative evaluations of the method in Section 2.1.2.

2.1.1 Method

Image registration is the process of matching a template image (T) to a reference image (R) by computing a spatial transformation that minimizes a cost function [58]. This spatial transformation can range from a simple translation to a non-rigid free-form deformation. Registration can be accomplished based on matching a set of feature points or can be directly applied to individual pixels.

Most image registration method accomplish the task using a variational formulation. The joint functional that is minimized is as follows [58]:

$$E[\mathbf{u}] = D[R, T; \mathbf{u}] + \alpha S \tag{2.1}$$

where E is the energy functional, D represents a distance/similarity measure and S represents the rate of smoothness of $\mathbf{u}$. The parameter α is used to balance the two terms. In this functional, $\mathbf{u}$ should be found such that the joint functional is minimized. This model is called *single direction model* because the reference image is fixed and only the template image is moving. This causes asymmetry in the results in such a way that if we fix the template image and move the reference image to match the template image (backward registration) the result may not be exactly opposite to that of the forward registration. For this reason, this model is modified to be used in the context of image slice interpolation by changing the formulation to the following:

$$E[\mathbf{u}] = D[R_1(\mathbf{x} - \mathbf{u}), R_2(\mathbf{x} + \mathbf{u})] + \alpha S \tag{2.2}$$

where $R_1, R_2 : \Omega \to \mathbb{R}$ are the two images provided as inputs and $\Omega = [0, 1]^2$ is the domain of images, $\mathbf{x}$ is the grid of image values and $\mathbf{u}$ is the displacement values for each grid point. Please note that in Equation (2.2), it is assumed that the slice to be interpolated, denoted by R, is in the middle of the given images. If R is an arbitrary slice between R_1 and R_2, then we first need to compute the distances from R to R_1 and R_2, denoted by d_1 and d_2 respectively. Then we calculate the ratio $r = d_1/(d_1 + d_2)$, and the following equation should be considered for interpolating R:

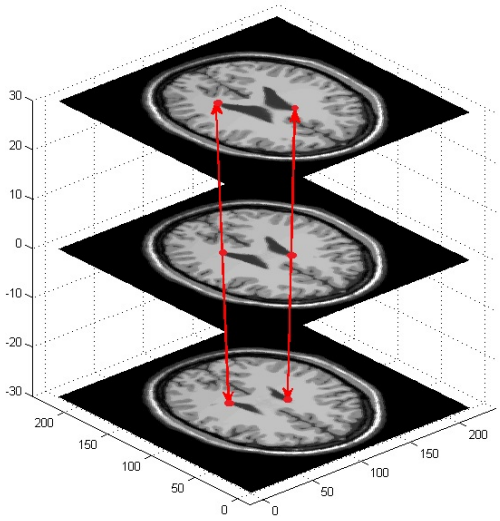


Figure 2.1: An illustration of linear displacements between corresponding points utilized in this section. The main goal is to use the top and bottom slices to reconstruct the in-between slice.

$$E[u] = D[R_1(\mathbf{x} - r\mathbf{u}), R_2(\mathbf{x} + (1-r)\mathbf{u})] + \alpha S \tag{2.3}$$

Without loss of generality, we shall consider Equation (2.2) in the current section for image slice interpolation. In this case $r = 0.5$ but since the coefficient will be the same for both of the images and practically doesn't affect the process of optimization, it is considered to be 1 for simplicity of representation in the rest of the section. Figure 2.1 illustrates the idea behind considering linear displacements between corresponding points that is utilized here for slice interpolation.

Several distance measures for D have been proposed in the literature, including the Sum of Squared Differences (SSD), Mutual Information (MI), Normalized Mutual Information (NMI), Cross Correlation (CC) and Normalized Gradient Fields (NGF) [135]. Here SSD is used as distance measure, and the above formulation can be re-written as:

$$\begin{aligned} D[R_1(\mathbf{x} - \mathbf{u}), R_2(\mathbf{x} + \mathbf{u})] &= \frac{1}{2}|R_1(\mathbf{x} - \mathbf{u}) - R_2(\mathbf{x} + \mathbf{u})|^2_{L_2} \\ &= \frac{1}{2}\int_\Omega (R_1(\mathbf{x} - \mathbf{u}(\mathbf{x})) - R_2(\mathbf{x} + \mathbf{u}(\mathbf{x})))^2 dx \end{aligned} \tag{2.4}$$

For the smoothness term S, several common choices are available, such as elastic,

fluid, demon, diffusion and curvature registration [58]. Here the curvature approach is used, in which the smoothness term is as follows:

$$S[\mathbf{u}] = \frac{1}{2}\sum_{l=1}^{2}\int_{\Omega}(\triangle u_l)^2 \tag{2.5}$$

where $\triangle$ is the curvature operator and the summation is computed over the two dimensions of images and the integral is computed inside the domain of images. As stated in [12], using curvature as the smoothness term, the need for an additional linear affine pre-registration step can be eliminated. Also it should be noted that the curvature operator S^{curv} is defined by $S^{curv} = S^{diff} * S^{diff}$, where $*$ is the convolution operator and S^{diff} is the discrete Laplace operator for 2D images.

In order to minimize the above joint functional in (2.2), the Gateaux derivative of $E[\mathbf{u}]$ is computed. Setting it equal to zero to find the minimum point, an Euler- Lagrange Partial Differential Equation (PDE) can be obtained as:

$$f(\mathbf{x}, \mathbf{u}(\mathbf{x})) + \alpha A^{curv}[\mathbf{u}](\mathbf{x}) = 0 \tag{2.6}$$

where $f(\mathbf{x}, \mathbf{u}(\mathbf{x})) = (R_2(\mathbf{x}+\mathbf{u}) - R_1(\mathbf{x}-\mathbf{u})).(\bigtriangledown R_1(\mathbf{x}-\mathbf{u}) + \bigtriangledown R_2(\mathbf{x}+\mathbf{u}))$ and $A^{curv}[\mathbf{u}] = \triangle^2\mathbf{u}$.

To solve this PDE, a time-stepping iteration method is considered as follows:

$$\partial_t \mathbf{u}^{k+1}(\mathbf{x}, t) = f(\mathbf{x}, \mathbf{u}^k(\mathbf{x}, t)) + \alpha A^{curv}[\mathbf{u}^{k+1}](\mathbf{x}, t), k \geq 0 \tag{2.7}$$

with $\mathbf{u}^0 = \mathbf{0}$. Using a finite difference approximation of the derivative with time step τ and also collecting the grid points with respect to a lexicographical ordering, one can derive a discretized version of (2.7) as follows:

$$(I_n + \alpha\tau A^{curv})\vec{\mathbf{U}}_l^{(k+1)} = \vec{\mathbf{U}}_l^{(k)} + \tau\vec{\mathbf{F}}_l^{(k)}, l = 1, 2 \tag{2.8}$$

where l is the parameter representing the dimension index. Following the same approach as in [58, 134], the optimization process can be done by exploiting Discrete Cosine Trans-

form (DCT). Assuming slices have the size of $m \times n$, the set of coefficients d_{j_1,j_2} are computed as follows:

$$d_{j_1,j_2} = -4 + 2\cos\frac{(j_1-1)\pi}{m} + 2\cos\frac{(j_2-1)\pi}{n} \tag{2.9}$$

where $j_1 = 1, 2, ..., m,\ j_2 = 1, 2, ..., n$. Defining $\mathbf{G} = DCT[\vec{\mathbf{U}}_l^{(k)} + \tau\vec{\mathbf{F}}_l^{(k)}]$ for $l = 1, 2$, we can have:

$$\vec{\mathbf{U}}_l^{(k+1)} = IDCT[\mathbf{V}] \tag{2.10}$$

where $\mathbf{V}_{j_1,j_2} = \mathbf{G}_{j_1,j_2}[1 + \tau\alpha d_{j_1,j_2}^2]^{-1}$ and $IDCT$ is the inverse discrete cosine transform. After finishing the optimization process, a simple averaging of the two transformed input images provides us with the missing in-between slice. Algorithm 1 summarizes the algorithm for curvature registration based slice interpolation method.

Algorithm 1 Curvature registration based slice interpolation algorithm

Initialization: $\tau, \alpha, \mathbf{X}, \mathbf{U}^0 = \mathbf{0}, d_{j_1,j_2}$;

Optimization:

for all $k = 0, 1, ...$ **do**

% Computing forces

$\mathbf{F}_l^{(k)} = (R_2(\mathbf{X} + \mathbf{U}^{(k)}) - R_1(\mathbf{X} - \mathbf{U}^{(k)})).(\bigtriangledown R_1(\mathbf{X} - \mathbf{U}^{(k)}) + \bigtriangledown R_2(\mathbf{X} + \mathbf{U}^{(k)}))$

% Solving the linear system

for all $l = 1, 2$ **do**

$\mathbf{G} = DCT[\vec{\mathbf{U}}_l^{(k)} + \tau\vec{\mathbf{F}}_l^{(k)}]$

for all $j_1 = 1, ..., m,\ j_2 = 1, ..., n$ **do**

$\mathbf{V}_{j_1,j_2} = \mathbf{G}_{j_1,j_2}[1 + \tau\alpha d_{j_1,j_2}^2]^{-1}$

end for

$\vec{\mathbf{U}}_l^{(k+1)} = IDCT[\mathbf{V}]$

end for

end for

Interpolation: $R = (R_1(\mathbf{X} - \mathbf{U}^{final}) + R_2(\mathbf{X} + \mathbf{U}^{final}))/2$

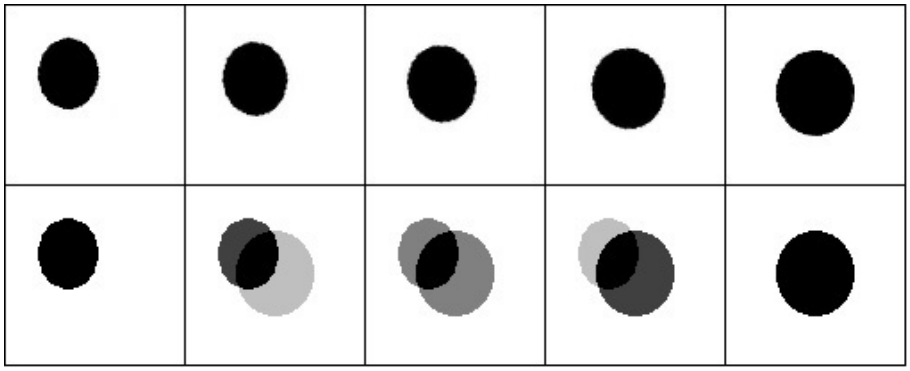


Figure 2.2: Top row: results of slice interpolation with proposed method. Bottom row: results of slice interpolation with linear interpolation method.

2.1.2 Results and Discussion

To validate the proposed method for slice interpolation in medical images, several tests have been conducted. The results of the proposed method are compared with two other methods, in both subjective and objective aspects. As a metric, Mean Squared Difference (MSD) is used for comparison. Assuming I_{org} and I_{rec} as original image and reconstructed image respectively, with the size of $m \times n$, MSD is defined as follows:

$$MSD = \frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} (I_{org}(i,j) - I_{rec}(i,j))^2 \tag{2.11}$$

In the first test, a pair of synthetic images of two circles is used. In the first and last column of Figure 2.2 the input images to the algorithms can be seen. Not only the location but also the size of the circle has changed. The goal is to place 3 in-between slices to show the gradual changes of the shape and location of the circle. The results of the proposed method and linear interpolation method are presented in the second, third and fourth columns of Figure 2.2. As expected, the proposed registration based method is able to correctly track the movement of the circle between two slices. Also from the images it can be seen that the transformation can be modeled as an affine transformation which using curvature registration based method it is perfectly estimated.

For the second test, three consecutive slices as in Figure 2.3 (a) are used. Taking the first and third slices as inputs, the in-between slice is reconstructed by using both linear interpolation and the proposed method. Figure 2.3 (b) shows the interpolation results

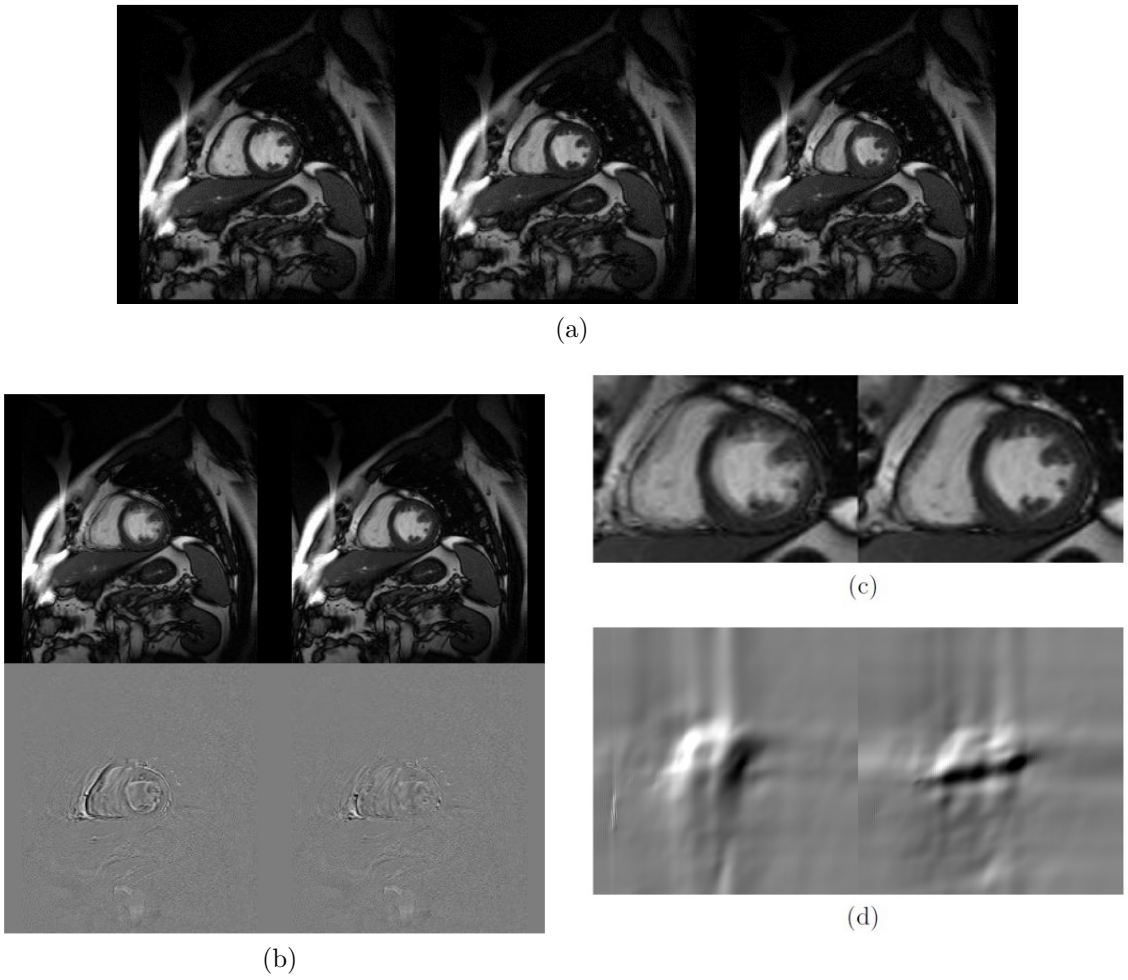


Figure 2.3: (a) Three consecutive slices. The first and third images are used for interpolation with parameters set as: $\tau = 0.03, \alpha = 100$ (b) Top row: results of interpolation for linear and proposed methods respectively. Bottom row: difference images for the results. (c) Close-up of the results of linear and proposed method. (d) Optimized displacement fields in horizontal and vertical directions.

using the two methods (top row) as well as the computed difference images (bottom row) with respect to the original image. The computed MSDs are 84.20 and 52.52 for linear and proposed methods respectively. Figure 2.3 (c) gives a close-up of the results. Figure 2.3 (d) gives the optimized displacement fields for both horizontal and vertical directions. Bright and dark shades represent positive and negative displacement values respectively, while gray shades are for displacements near zero.

As can be seen, the difference between slices is due to the movement of the heart near the center of these images. Using linear interpolation, the movement of heart is not captured, resulting in blurred edges in the interpolated slice. By comparison, the registration-based method captures the movement well and the final result is highly sim-

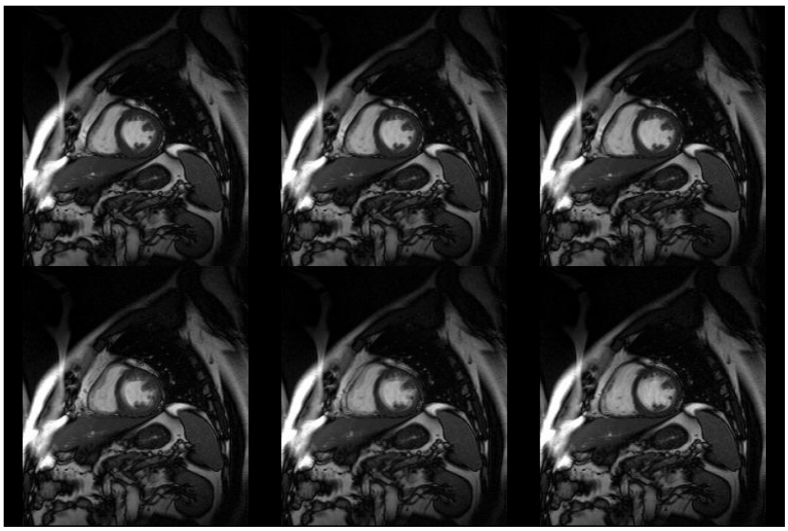
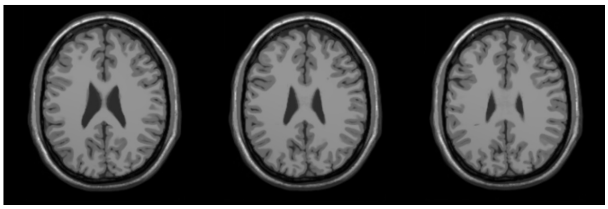


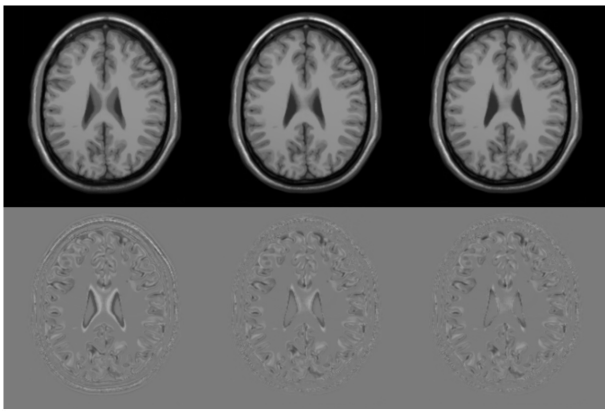
Figure 2.4: Results of placing 3 slices between two input slices (first and third slices from Fig.2 (a)) using proposed method (Top row) and linear interpolation method (Bottom row)

ilar to the original one (middle image in Figure 2.3 (a)). As a result, the MSD error is significantly reduced and the interpolation result is much sharper. Figure 2.4 represents the results of placing 3 in-between slices for the two input images for both proposed method and linear interpolation method. The movement of the heart is perfectly tracked using the proposed registration based method while linear interpolation method cannot capture this movement.

To further demonstrate the strength of the proposed method, the same procedure is applied to another set containing three brain images as shown in Figure 2.5 (a). Using the first and third slice, the interpolation results are produced. Besides the linear interpolation, we also compare the proposed method with a non-modified curvature registration based technique, here called non-modified method. For this method, after registering the reference and moving the template image using curvature registration [58] and finding the optimized displacement fields, linear interpolation along the computed displacement vectors of corresponding points in the reference and moving images is implemented to reconstruct the in-between slice. Figure 2.5 (b) shows the results of interpolation as well as the computed difference images. The MSDs are 71.65, 45.36 and 42.72 for linear, non-modified and proposed methods respectively. As can be seen, the result of linear interpolation has uncertain and highly blurred edges. Result of non-modified method



(a)



(b)



(c)

Figure 2.5: (a) Three consecutive slices. The first and third images are used for interpolation with parameters set as: $\tau = 0.05, \alpha = 100$ (b) Top row: results of interpolation for linear, non-modified and proposed methods respectively. Bottom row: difference images for the results. (c) Close-up of the results of linear, non-modified and proposed method respectively.

is significantly better than linear interpolation, in terms of MSD but due to nonlinear nature of image registration and optimization process, we still have blurred edges. In comparison, the proposed method gives much sharper edges. This becomes more obvious in the difference images where blurred edges cause widened regions of dissimilarity. Also, it should be mentioned that, in the non-modified method, only one of the images is moving. As a result, more iterations of optimization are needed for convergence, and thus more computational time is required. Figure 2.5 (c) gives a close-up of the results for better comparison.

Based on Figure 2.5, the effect of moving both images simultaneously in the proposed method in comparison to moving only one of the images in the non-modified method is

obvious. Moving both images not only reduces the computational time needed for the convergence, but also can prevent the algorithm from getting trapped in local minima which is caused because of large displacements between corresponding points (See Figure 2.5 (c) for the comparison of the methods). Also the integration of linear displacement between corresponding points in the process of optimization reduces the need for additional linear interpolation after registration to a simple averaging between the two deformed images.

A similar test is conducted for the entire brain image database, containing 79 images with the size of 217×181 pixels. For interpolation of each evenly numbered slice, a pair of two consecutive slices with odd numbers is used and the reconstructed images are compared with the corresponding slices from the original database. The average MSDs are presented in Table 2.1. The numbers in parentheses for the non-modified method represents the improvement percentage with respect to linear interpolation method. The numbers in parentheses for the proposed method represent the improvement percentage with respect to linear and non-modified method respectively.

In terms of computational time, excluding the linear interpolation method which obviously takes less time than the other two, the proposed algorithm outperforms the non-modified method. The algorithm is implemented using MATLAB without any specific code optimization procedure. The average time for the proposed method to produce the in-between image is about 40 seconds while for the non-modified method it is about 50 seconds. This is due to the fact that only one of the images is moving which makes it more time consuming for convergence. The optimization process stops when the improvement in SSD is less than 0.01%. Also to produce the results presented here, τ and α are fixed for the whole database ($\tau = 0.05, \alpha = 100$). Of course since τ is the time step of the iterative scheme, the process of finding the best value can be further optimized using a line search method to ensure faster and more robust convergence in fewer iterations.

For the smoothness term, the curvature operator penalizes oscillation in the displacement field [58]. Also it reduces the need for additional affine transformation in the beginning of the process of image registration. The regularization parameter α deter-

Table 2.1: Comparison of the average MSDs for brain image database (as well as the improvement rates) for the linear, non-modified and the proposed slice interpolation approaches.

Method	**Linear**	**Non-modified**	**Proposed**
MSD	118.7652	56.0765 (52.78%)	**54.6450 (53.99%, 2.56%)**

mines the balance between the two terms in the energy functional. Choosing small values for the parameter causes non-smooth displacement in the final results while choosing big values makes the deformation more rigid which is not useful for slice interpolation due to deformation of objects in the consecutive slices. To have a deformable registration between slices, there should be a trade-off between smoothness of the transformation and the rigidity of the movements. Here the value for α is set intuitively and the same for all the tests provided here which may not be appropriate, especially in case of medical images since there might be different objects (organs) with different physical properties within images. For a general discussion on this subject the reader is referred to [95].

It should be noted that the improvement is not significant compared to the results of the non-modified method. But as it is obvious from the results, by integrating the idea of linear movement between corresponding points in the process of optimization better results can be achieved. However, even in the non-modified method, a linear interpolation is needed between the corresponding points in order to reconstruct the in-between slice. Also it should be mentioned that one of the main assumptions of using registration based methods for slice interpolation is that the objects within the input slices can deform or move, but they cannot disappear. In other words, if from one slice to another the object disappears the result of registration based interpolation is unpredictable. This is not an assumption that can be completely preserved when the input data is a stack of medical images. Overall, the proposed registration based method, can manage to improve the results more than 2.5% percent when compared with the non-modified version (See Table 2.1). Of course there is room for improvement both in computational time and quality of the final image.

2.1.3 Conclusion

A new registration-based slice interpolation method is proposed. A modified version of the curvature registration method has been used with the assumption of linear displacements between corresponding points in two input images. The obtained displacement fields for the two input images are utilized to produce the missing in-between slice. In comparison to both linear interpolation and the non-modified registration based method, the proposed method produces lower MSD values and sharper/certain edges. The current implementation was performed in MATLAB without any code optimization procedure. Use of and C/C++ implementation can reduce the computational time drastically.

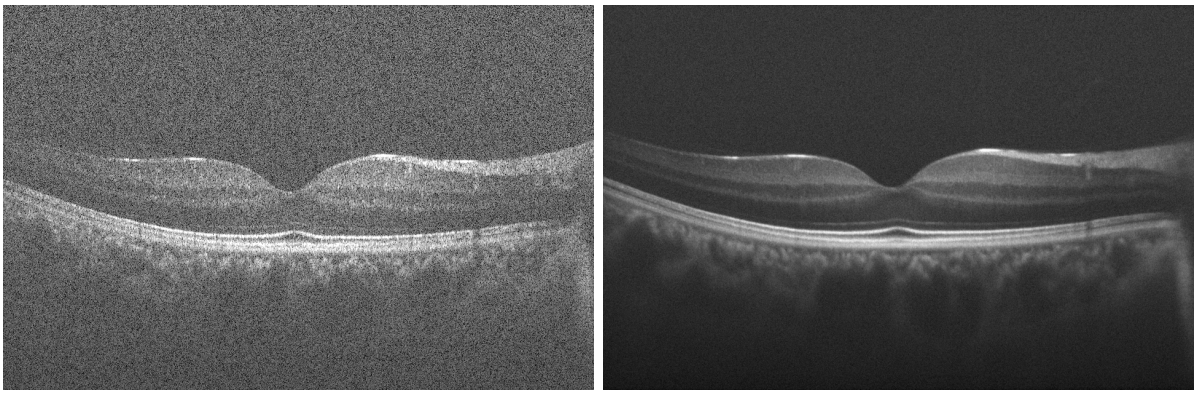
2.2 Sparse and Low Rank Decomposition Based Batch Image Alignment for Speckle Reduction of Retinal OCT Images

Optical Coherence Tomography (OCT) is a powerful non-invasive imaging system for acquiring 3D volumetric images of tissues. OCT as an optical imaging modality, aims to provide cross-sectional images of tissues by measuring the magnitude of back-reflected/backscattered light as well as the echo time delay. The concept resembles that of ultrasound, however, due to high speed of light, direct measurement of the optical echo is impossible. This calls for indirect procedures for measuring the time-of-flight and intensity of the backscattered light which is done by taking advantage of interferometric techniques using ultra-short light pulses or partially coherent light [7, 8, 19, 87, 191]. Throughout the past two decades, new developments in the OCT imaging systems have improved the acquisition time and also the quality of the acquired images. Nowadays taking μm-resolution volume images of the tissues is very common especially in ophthalmology.

Generally speaking, there are three main aspects of retinal OCT image processing: noise reduction, feature segmentation, and image registration. The process of OCT image acquisition results in the formation of irregular granular patterns called speckle. Speckle

is a fundamental property of the signals and images acquired by narrow-band detection systems like Synthetic-Aperture Radar (SAR), ultrasound and OCT. Not only the optical properties of the system, but also the motion of the subject to be imaged, size and temporal coherence of the light source, multiple scattering, phase deviation of the beam and aperture of the detector can affect the speckle [173]. Two types of speckle are present in OCT images: signal-carrying speckle which originates from the sample volume in the focal zone; and signal-degrading speckle, also known as speckle noise, which is created by multiply-scattered out-of-focus light. Figure 2.6 (a) shows a sample retinal OCT image, highly degraded by speckle noise. Delineating micro-structures in the image is of particular importance in OCT image processing for ophthalmology [46, 93] and therefore image segmentation plays a significant role in OCT data analysis, especially retinal layer segmentation. Active contour based techniques [57, 133, 212] and graph-based techniques [41, 42, 68, 92, 179] are very good examples of such approaches. There are several different applications for using image registration approaches in OCT image analysis, such as noise reduction [4, 12, 13, 91], multi-modal retinal image registration [69, 109], image mosaicing [82, 110, 122] and motion correction [101, 100, 154, 160, 205].

Speckle is considered to be multiplicative noise, in contrast to the additive Gaussian noise. Due to limited dynamic range of displays, OCT signals are usually compressed by a logarithmic operator applied to the intensity information which converts the multiplicative speckle noise to additive noise [166]. OCT noise reduction techniques can be divided into two major classes: 1) methods of noise reduction during the acquisition time and 2) post-processing techniques. In the first class, which is usually referred to as *compounding* techniques, multiple uncorrelated recordings are averaged. These include spatial compounding [9], angular compounding [172], polarization compounding [97] and frequency compounding [155]. There are two major classes of post-processing techniques for speckle noise reduction: anisotropic diffusion-based techniques [158, 166] and multi-scale/multi-resolution geometric representation techniques [1, 75, 76, 89, 90, 156, 206]. Use of compressed sensing and sparse representation have also been explored in the past few years [54, 55]. For a more complete review on the different image analysis techniques



(a) (b)

Figure 2.6: (a) Sample retinal OCT image degraded by speckle noise, (b) Final result of the proposed method using 50 misaligned noisy retinal OCT images

in OCT image processing, including noise reduction, the reader is referred to [19] and references therein.

Post-processing averaging/median filtering is also an interesting method for speckle reduction. Usually in such techniques, multiple B-scans of the same location are acquired and then the average/median is taken. It is assumed that the speckle between different recordings is un-correlated. The misalignment between the different B-scans is usually compensated with a parametric image registration technique, such as translation based registration, rigid registration or affine registration. Theoretically, having N B-scans, the Signal-to-Noise-Ratio (SNR) can be improved by a factor of $\sqrt{N}$.

In [91] a dynamic programming based method is used for compensation of the translational movements between several B-scans and reducing the speckle noise. But as it is obvious, translation is not the only possible movement that can happen between different B-scans. A hierarchical model-based motion estimation scheme based on an affine-motion model is used in [4] for registering multiple B-scans to be used for speckle reduction.

Here, another technique for registration-based speckle reduction is proposed. This technique utilizes sparse and low rank decomposition to separate between image features and noise components in each B-scan, while aligning them iteratively. Using this technique, sub-pixel accuracy can be achieved for the alignment process which can further improve the SNR and Contrast-to-Noise-Ratio (CNR) in the final denoised result. Section 2.2.1 contains detailed explanation of the sparse and low rank decomposition based

batch image alignment technique. In Section 2.2.2 the results of this method is presented and compared with translation and rigid registration methods in terms of SNR and CNR.

2.2.1 Method

Robust Principal Component Analysis (RPCA)

Given a large data matrix $D \in \mathbb{R}^{m \times n}$, the RPCA process divides D into two components: $D = L + S$, with L being the low rank component and S being the sparse component [38]. In an l_2 sense, this is the classic Principal Component Analysis (PCA). Even though this technique is widely used in the literature, the inherent sensitivity to outliers makes it less useful in modern applications. This can be remedied by minimizing $||L||_* + \lambda||S||_1$ s.t $L + S = D$ which is proven to have an exact recovery under broad conditions [38].

This concept has been widely used in different branches of computer vision and image processing such as video surveillance, shadow or specularity removal in face recognition, video repairing etc. In [121] this technique is used for single OCT image noise reduction. For post-processing averaging/median filtering, multiple B-scans of the same location are acquired and used for noise reduction. As mentioned before, here the main issue is regarding the misalignment between the B-scans, as well as the differences in the displayed patterns due to eye movement. This requires image alignment prior to averaging/median filtering. One technique is to pre-register the noisy images and use them as inputs for the next stage. Considering the high amount of noise degrading the images, this can cause erroneous alignment. A better way is to combine the image registration task with low rank/sparse decomposition of the data which reduces the effect of noise in the process of registration. Another advantage of this method is its ability to detect the underlying low-rank pattern which results in elimination of retinal features that are not present in all of the slices and only appear due to eye movement. Here, we follow the work of [149] for simultaneous alignment and decomposition of the retinal OCT images.

Robust Alignment by Sparse and Low-Rank Decomposition (RASL)

Assuming a set of n B-scans, the data matrix D can be created by stacking the vectorized images as columns of the matrix. Having completely aligned images, it is expected that D is low-rank, with the possibility of having sparse additional noise: $D = L + S$. In the case of OCT images considered here, additional optimization is needed in order to compensate the misalignment. This is done by assuming a set of parametric transformations, τ, applied to the images. In this formulation we have:

$$\min_{L,S,\tau} rank(L) + \lambda||S||_0 \quad \text{s.t.} \quad D \circ \tau = L + S \tag{2.12}$$

This is a non-convex and NP-hard problem to solve due to the need for minimizing the rank and the l_0 norm. Convex relaxation of the problem as elaborated in [38] results in:

$$\min_{L,S,\tau} ||L||_* + \lambda||S||_1 \quad \text{s.t.} \quad D \circ \tau = L + S \tag{2.13}$$

where $||.||_*$ is the nuclear norm (sum of singular values) and $||.||_1$ is the l_1 norm.

Another difficulty arises from the non-linearity of the constraint $D \circ \tau = L + S$. This can be solved assuming minimal changes in τ in each iteration and linearizing around the current estimate of τ. Therefore:

$$D \circ (\tau + \Delta\tau) \approx D \circ \tau + \sum_{i=1}^{n} J_i \Delta\tau \varepsilon_i \varepsilon_i^T \tag{2.14}$$

where J_i is the Jacobian of the ith image with respect to the transformation parameters and ε_i is the standard basis for $\mathbb{R}^n$. This linearization only holds for small misalignment between the images in the batch. Starting from an initial set of transformations, here the identity transformation, and setting rigid transformation as the desired transformation, at each iteration this linearized convex optimization problem is solved using the normalized images to avoid the trivial solutions until convergence. Algorithm 3 summarizes the process.

The main computational cost of the Algorithm 3 is in the third step: solving the

Algorithm 2 Sparse and Low-rank Based Alignment

Inputs: input images, initial transformation set, $\lambda > 0$

WHILE not converged **DO**

Step 1: compute the Jacobian w.r.t transformations: J_i

Step 2: warp and normalize the images: $D \circ \tau$

Step 3 (inner loop): solve the linearizied convex optimization problem:

$$(L^*, S^*, \Delta\tau^*) \leftarrow \arg\min_{L,S,\Delta\tau} ||L||_* + \lambda||S||_1$$
$$\text{s.t.} \quad D \circ \tau + \sum_{i=1}^{n} J_i \Delta\tau \varepsilon_i \varepsilon_i^T = L + S$$

Step 4: update transformation: $\tau \leftarrow \tau + \Delta\tau$

END WHILE

OUTPUT: solution L^*, S^* and τ^* to problem (2).

linearized convex optimization problem. Since this problem has millions of variables, having a scalable solution is of high importance. Augmented Lagrange Multiplier (ALM) [113] has been proven to have reliable results for such optimization. By defining:

$$h(L, S, \Delta\tau) = D \circ \tau + \sum_{i=1}^{n} J_i \Delta\tau \varepsilon_i \varepsilon_i^T - L - S, \tag{2.15}$$

the augmented Lagrangian function to be optimized is defined as:

$$\mathcal{L}_\mu(L, S, \Delta\tau, Y) = ||L||_* + \lambda||E||_1 + < Y, h(L, S, \Delta\tau) > + \frac{\mu}{2}||h(L, S, \Delta\tau)||_F^2 \tag{2.16}$$

where Y is the Lagrange multiplier matrix, μ is a positive scalar and $||.||_F$ is the Frobenius norm. Choosing an appropriate Y and large enough μ, the augmented Lagrangian function has the same minimizer as the original constrained optimization problem. For further explanations regarding the optimization process the reader is referred to [149, 113].

The final result of the algorithm is the well-aligned stack of images, decomposed into low-rank data set containing image information and sparse component consisting of speckle noise. As investigated in [9], spatial compounding works best with use of median filtering rather than averaging. Here, the final image is created by pixel-wise median filtering of the final low-rank components of the data.

2.2.2 Results and Discussion

For assessing the performance of the proposed method, several metrics are considered. Considering 6 regions of interest (ROI) in the final results, one only containing background noise and the rest containing image features, the metrics can be defined as follows:

$$\begin{aligned} SNR_m &= 20 \times log(\frac{\mu_m}{\sigma_b}) \\ CNR_m &= \frac{\mu_m - \mu_b}{\sqrt{\sigma_m^2 + \sigma_b^2}} \end{aligned} \tag{2.17}$$

where μ_b and σ_b are the mean and standard deviation of the background noise and μ_m and σ_m are the mean and standard deviation of the mth ROI containing image features. The average of these metrics are considered here for comparison.

Different numbers of images of the human retina in the central foveal region are considered for assessing the performance of the proposed algorithm. Figure 3.1(a) shows a sample image of the dataset that is used here. As for other registration based methods, translation and rigid registration techniques available in ImageJ [174] software package are considered for comparison, since they are widely used in the literature and give reasonable performance.

Figure 3.2 represents the improvement in the average SNR of the final image for different techniques, while Figure 3.3 shows the improvement achieved in average CNR for different number of input images. Figure 3.1 (b) shows the final result of the proposed algorithm for speckle noise reduction using 50 misaligned noisy OCT input images.

One critical step in post-processing averaging/median filtering of the OCT images for noise reduction is the pre-selection of the set of images to be registered and averaged. This is due to the presence of μm-level features in the high resolution OCT images. During the imaging session, movement of the eye in 3 dimensions causes these features to appear/disappear between consecutive B-scans. In other words, fine features from very close locations come to focus, while some other features will go out of focus. This makes the pre-selection a necessary step at the beginning of the process. Using RPCA, this can be eliminated. As mentioned before, RPCA tries to divide the input set of data,

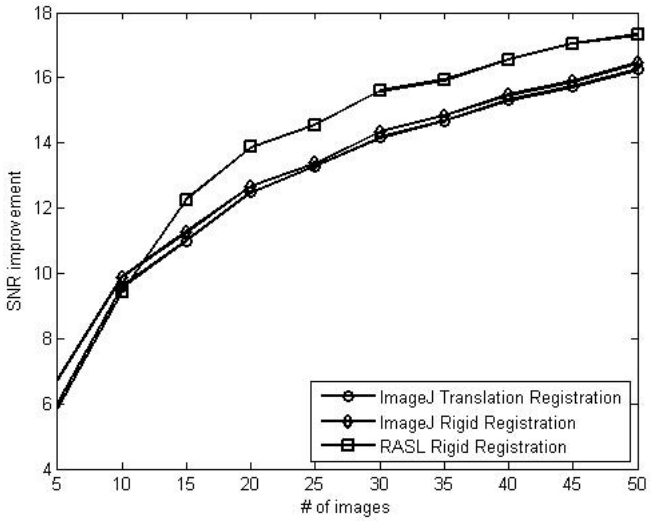


Figure 2.7: SNR improvement for different number of input images (5-50)

here the stack of vectorized B-scans, into two components: one low-rank and one sparse. In this scenario, the algorithm looks for similar patterns in the data that are shared between different B-scans to extract the low-rank component while eliminating different patterns as being noise. In other words, without the need for pre-selection, the RPCA chooses the most common features as the ground truth and neglects the features that only appear in few B-scans and considers them as noise. Even though this can be achieved using simple averaging too given enough number of images, still the main drawback is that naive averaging is indecisive about the common/uncommon features to be averaged causing it to have more blurred features. This is because the uncommon features are diffused to the rest of the data. The same analogy can be applied for elimination of blood vessel shadows between different B-scans. Figure 3.4 displays close-ups of the original and filtered version of the input images using different techniques for comparison.

2.2.3 Conclusion

A new application of the sparse and low rank decomposition based batch image alignment in noise reduction of OCT images is introduced. Having a stack of misaligned, mostly due to eye movements, and noisy retinal OCT images, the process of alignment is done by decomposition of the vectorized image data into low rank and sparse components at each iteration to ensure better final alignment and noise/signal separation. Using SNR and

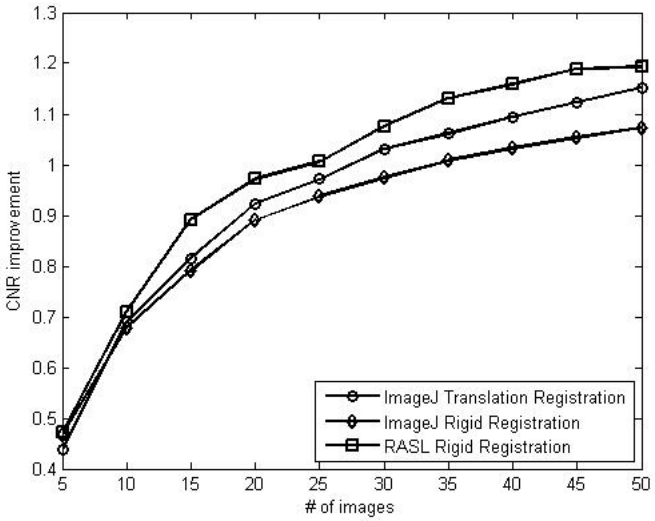
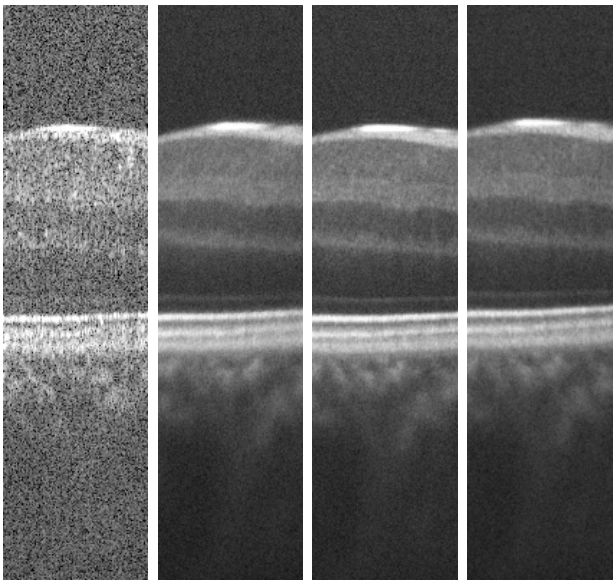


Figure 2.8: CNR improvement for different number of input images (5-50)



(a) (b) (c) (d)

Figure 2.9: A portion of one of the input images (a) and the results of ImageJ translation (b), ImageJ rigid (c) and RASL rigid registration based methods (d).

CNR as metrics, the performance of the method is compared with some other registration based techniques for speckle noise reduction. Our approach gives better performance when bench marked against other techniques with respect to measures such as SNR and CNR while incurring larger computational cost. Using GPU implementations, higher speeds can be achieved, which is not the focus of our work. Also, from an algorithmic point of view, newer techniques have been proposed in the literature for sparse and low rank decomposition, which can be considered for future research.

2.3 Fast Mesh-Based Medical Image Registration

Other than the obvious way of assessing an image registration method which is the similarity of the final deformed template image to the reference image, the computational complexity of the method is also a major factor. This becomes more important in case of non-rigid image registration methods in which the deformation is local rather than global. The main reason for this is due to the size of the images, which results in a very big number of degrees of freedom in the optimization process. This is more obvious in case of non-parametric image registration techniques. One obvious solution to remedy this problem is using parallel computing using specialized hardware such as Graphics Processing Units (GPUs) [61].

From an algorithmic point of view, using multi-resolution techniques can also be considered [43]. These techniques start from a very coarse grid to capture the larger deformations in the images to be registered and then move to much finer grids to capture the smaller deformations. Usually, this is done on a uniform square grid which means that the sampling will take place uniformly. In other words, even though the objects and image features are scattered randomly in the images to be registered, multi-resolution based methods do not take this fact into consideration. Adaptive grid generation using Octrees has been used to address this issue [77]. However, square grids cannot accurately match feature boundaries that are typically curvilinear. Use of triangular meshes enables better representation of curvilinear image feature boundaries [117]. This representation also allows for easy reconstruction of final results as well as the computation of the final optimized displacement. Furthermore, content-based adaptive meshing greatly reduces the computational complexity of the registration process. Of course, this will have its own implications for optimizing the energy functional of the method which will be discussed more in later sections.

There have been a few previous works in this area, usually considering the problem of image registration as a Finite Element Method (FEM) problem [157, 15]. But they usually need additional information about the physical properties of the underlying structure. Also, the process is not completely mesh based. In [157] each update of the displacement

field in each iteration is followed by a re-sampling to the regular image grid, computational procedures to find the update terms for the next update and then re-sampling again to the triangular grid.

Here, a new mesh based image registration technique will be introduced which takes advantage of mesh-based operators without the need to switch between regular image grid and triangular mesh grid. Use of such techniques has previously been investigated for image restoration and segmentation [117, 209, 210, 214, 215]. Taking the reference and template images as the inputs and some needed parameters, the proposed method can achieve excellent accuracy and higher speed compared to regular pixel-based registration methods. Even though a uniform initial mesh can be considered, a content adaptive initial mesh is created for the template image which perfectly matches the image's edges and features. The energy functional is minimized and the final displacement field is reconstructed. Finally, MSD will be used to assess the performance of our registration technique. Section 2.3.1 contains comprehensive details about the proposed methods. Results and discussions are provided in Section 2.3.2.

2.3.1 Method

Taking the general notation as most of the papers in this field, assuming template (T) and reference (R) images as inputs, the goal of image registration is to find a valid and optimal geometrical transformation to be applied to T to become more similar to R, according to some similarity measure. Therefore, the process can be formulated as an optimization problem which tries to optimize some energy functional that can be defined as in equation (2.1). Due to the random distribution of image features, we propose to use a sparser representation of both input images using a content adaptive mesh generator such as the one described in [208].

Formulation of The Mesh-Based Image Registration Method

Assume T and R as input images with the same size, and a set of triangles defined on the template image represented by (V, F), where V is a $n_V \times 2$ matrix containing the

coordinates of n_V mesh nodes or vertices and F is a $n_F \times 3$ matrix, each row containing the indices of nodes creating each one of the n_F triangles/faces. Note that $T(V)$ represents a vector with constant values which only the locations of its values change in the process of optimization. R represents a continuous domain on $X \in \Omega$, hence $R(V) = R(X)|_{X=V}$. Also, it should be mentioned that the set of triangles covers the template image's domain Ω. In the optimization process to find the displacement field, the smoothness term (regularizer) will be applied after each iteration using a diffusion process described later in more detail. The energy functional is therefore define as follows:

$$E[\mathbf{u}(V)] = D[R(X)|_{X=V+\mathbf{u}}, T(V) \circ \mathbf{u}(V)] \tag{2.18}$$

where E and D represent the energy functional and the distance measure respectively. Also, the $\circ$ operator is defined as:

$$T(V) \circ \mathbf{u}(V) = T(V + \mathbf{u}(V)) \tag{2.19}$$

For simplicity of representation, and since it is obvious that the method is applied to mesh nodes, from now on, the notation of a function f of variable V which has a general form of $f(V)$ will be reduced to just f. As mentioned before, several distance or similarity measures can be found in the literature, each having its pros and cons and being suitable for different problems encountered in image registration. Here, the Sum of Squared Differences (SSD) is used which can be defined as follows:

$$\begin{aligned} D(R(X)|_{X=V+\mathbf{u}}, T \circ \mathbf{u}) &= \frac{1}{2}||R(X)|_{X=V+\mathbf{u}} - T \circ \mathbf{u}||^2 \\ &= \frac{1}{2} \sum_{i=1:n_V} (T(V_i) \circ \mathbf{u}(V_i) - R(X_i)|_{X_i=V_i+\mathbf{u}_i})^2 \end{aligned} \tag{2.20}$$

where the last summation is computed over all of the mesh nodes. Minimizing the energy functional and updation of the displacement field can be done considering a gradient descent approach:

$$\mathbf{u}_0^{k+1} = \mathbf{u}_1^k - \tau \bigtriangledown_{\mathbf{u}_1^k} E[\mathbf{u}_1^k] \tag{2.21}$$

where τ is the step size (here 0.005) and $\bigtriangledown_{\mathbf{u}_1^k}$ is the gradient operator with respect to variable u_1^k. The gradient of the energy functional is computed by taking the Gateaux derivative of the distance measure which results in:

$$\begin{aligned} &\bigtriangledown_{\mathbf{u}_1^k} E[\mathbf{u}_1^k] = \bigtriangledown_{\mathbf{u}_1^k} D \\ &= (T(V + \mathbf{u}_1^k) - R(X)|_{X=V+\mathbf{u}}). \bigtriangledown_{\mathbf{u}_1^k} T(V + \mathbf{u}_1^k) \end{aligned} \tag{2.22}$$

where $\bigtriangledown_{\mathbf{u}_1^k} T(V + \mathbf{u}_1^k)$ needs to be computed on mesh nodes.

The reason behind using two different subscripts (0 an 1) in (2.15) is because of the fact that this displacement function needs to be smoothed to ensure regularized displacements in the image domain. For smoothing the displacements on the mesh, a diffusion process needs to be solved on the mesh nodes. This diffusion process can be modeled as follows:

$$\frac{\partial \mathbf{u}_0^{k+1}}{\partial t} = \lambda \bigtriangleup \mathbf{u}_0^{k+1} \tag{2.23}$$

where $\bigtriangleup$ represents the Laplacian operator on mesh nodes. This diffusion process is solved using a forward difference time-stepping approach. Without loss of generality and to reduce the confusion with the gradient descent method's step size, here the time step will be considered as 1. Hence from (2.17):

$$\mathbf{u}_1^{k+1} = \mathbf{u}_0^{k+1} + \lambda \bigtriangleup \mathbf{u}_0^{k+1} \tag{2.24}$$

where $0 < \lambda < 1$ is the smoothing parameter defined by the user (here 0.8). Further simplification will be done in the following sections.

Discretization of Gradient on a Triangular Mesh

Consider node V_i and its 1-ring (N_1) neighbor nodes. Approximation of the gradient of a function f on the location of node V_i can be achieved using linear interpolation of the function f over the surface created by this region. Assuming triangle F_j created by nodes $[V_i V_j V_k]$ as one of the triangles surrounding V_i, the approximation of the gradient on F_j

will be:

$$\begin{aligned}\bigtriangledown f_{T_j} = \frac{1}{4A_j^2}\Big(& f_i[(\overrightarrow{V_{ij}}, \overrightarrow{V_{jk}})(V_k - V_i) + (\overrightarrow{V_{ik}}, \overrightarrow{V_{kj}})(V_j - V_i)] \\ & + f_j[(\overrightarrow{V_{ji}}, \overrightarrow{V_{ik}})(V_k - V_j) + (\overrightarrow{V_{jk}}, \overrightarrow{V_{ki}})(V_i - V_j)] \\ & + f_k[(\overrightarrow{V_{kj}}, \overrightarrow{V_{ji}})(V_i - V_k) + (\overrightarrow{V_{ki}}, \overrightarrow{V_{ij}})(V_j - V_k)]\Big)\end{aligned} \tag{2.25}$$

where f_i is the function value on node V_i, A_j is the area of the triangle F_j, $\overrightarrow{V_{ij}}$ is the vector connecting nodes i and j and $(\overrightarrow{a}, \overrightarrow{b})$ gives the *dot* product of vectors $\overrightarrow{a}$ and $\overrightarrow{b}$. Having the approximation of the gradient on surrounding triangles, the approximate gradient for node V_i can be computed as follows:

$$\bigtriangledown f(V_i) = \frac{1}{A(V_i)} \sum_{j \in N_1(V_i)} A_j \bigtriangledown f_{T_j} \tag{2.26}$$

where $A(V_i) = \sum_{j \in N_1(V_i)} A_j$. For a complete analysis on the approximation error the reader is referred to [204]. The areas of triangles should be computed at the beginning of each iteration.

Diffusion-Based Smoothing of Displacement

Taking the same approach as [47], the Laplacian operator on a mesh can be approximated by the so-called *umbrella* operator on each node as follows:

$$\bigtriangleup \mathbf{u}(V_i) = \frac{1}{m_i} \sum_{j \in N_1(V_i)} \mathbf{u}(V_j) - \mathbf{u}(V_i) \tag{2.27}$$

where m_i is the valence (number of 1-ring neighbors) of node V_i. This operator can be defined in a matrix form:

$$\bigtriangleup \mathbf{u} = (A^{Lap} - I)\mathbf{u} \tag{2.28}$$

where I is the identity matrix and A^{Lap} is a sparse $n_V \times n_V$ matrix which its non-zero elements are defined as:

$$A_{ij}^{Lap} = \frac{1}{m_i}, \ for\ all\ j \in N_1(Vi) \tag{2.29}$$

Considering (2.18) and (2.23) together with a few manipulations, the diffusion process can be simplified as a weighted average of the displacements of the 1-ring neighborhood of each node:

$$\mathbf{u}_1^{k+1} = \big((1-\lambda)I + \lambda A^{Lap}\big)\mathbf{u}_0^{k+1} \tag{2.30}$$

The above equation can be applied iteratively for further smoothness of the displacement field on the mesh nodes. Here, only one iteration of smoothing is applied. The overall algorithm for mesh-based registration is illustrated in Algorithm 2.

Algorithm 3 Fast Mesh-Based Image Registration

Inputs: R, T, (V, F) defined on the template image, λ, τ ;
Pre-Computation: N_1 and neighbor triangles for each mesh node, A^{Lap};

For $k = 1 \rightarrow$ *convergence*
{

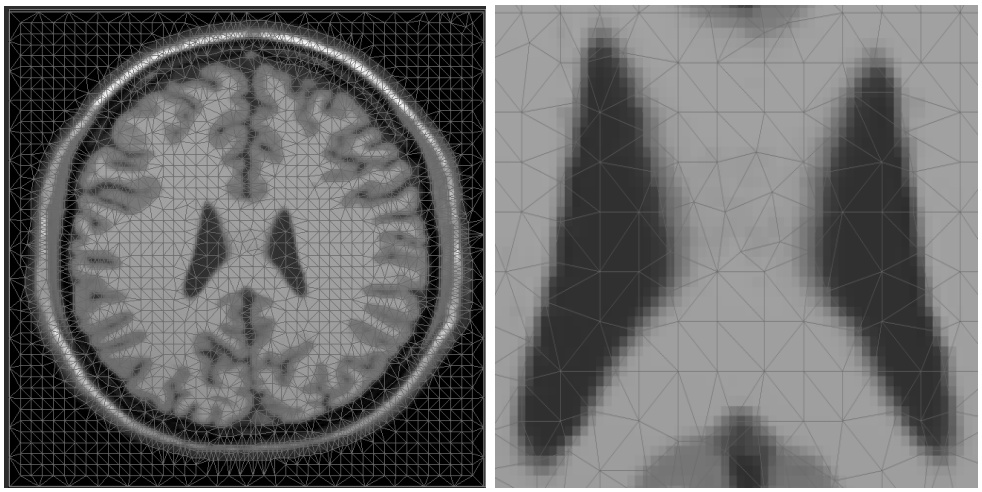
- *Update*:
  - $E[\mathbf{u}] = D(R(X)|_{X=V+\mathbf{u}}, T \circ \mathbf{u})$
  - $\bigtriangledown_{\mathbf{u}_1^k} E[\mathbf{u}_1^k]$
  - $\mathbf{u}_0^{k+1} = \mathbf{u}_1^k - \tau \bigtriangledown_{\mathbf{u}_1^k} E[\mathbf{u}_1^k]$
- *Smoothing*:
  - $\mathbf{u}_1^{k+1} = \big((1-\lambda)I + \lambda A^{Lap}\big)\mathbf{u}_0^{k+1}$

}

2.3.2 Results And Discussion

Content Adaptive Mesh Generation

For generating the content adaptive mesh needed for our algorithm, the method proposed by Ming et al. [208] is used. Based on the discussion given in this paper, the main difference between various mesh generating methods rises from the differences in node placement procedures. In some, the nodes are placed based on feature points in images, while in others, this process is done iteratively, either starting from a coarse mesh and adding new nodes or starting from a dense mesh and removing redundant nodes. The method consists of several steps as follows:



(a) (b)

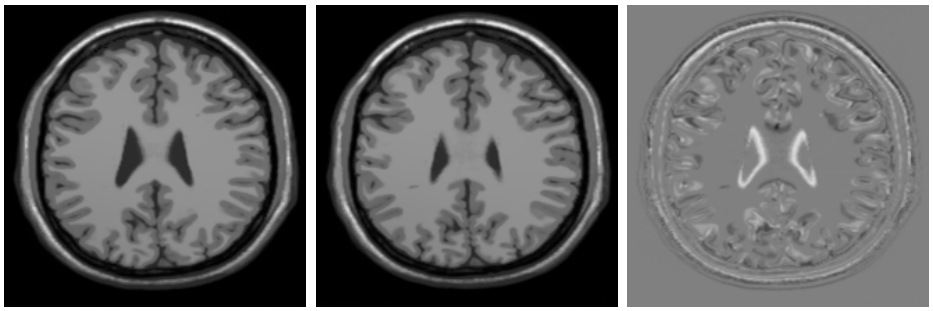
Figure 2.10: Example of content adaptive mesh generation

1. Node generation:
    - Canny sample points;
    - Halftoning sample points;
    - Uniform sample points.
2. Mesh generation via Delaunay triangulation;
3. Image-based mesh smoothing:
    - Image-based Centroid Voronoi Tessellations (CVT) mesh smoothing;
    - Image-based Optimal Delaunay Triangulations (ODT) mesh smoothing;
    - Edge flipping.

The result of this method is a high quality content adapted triangular mesh which is matched accurately with image features and edges. Figure 2.6 displays an example of content adapted mesh generated for a brain cross-section. Image registration results using the proposed method is given in the following sections.

Example 1- Brain CT Images

For the first example, a pair of brain images are considered which are displayed in Figure 2.7. Comparing the reference and template images reveals a rigid transformation as well



(a) (b) (c)

Figure 2.11: (a) Template image, (b) Reference image, (c) Difference image

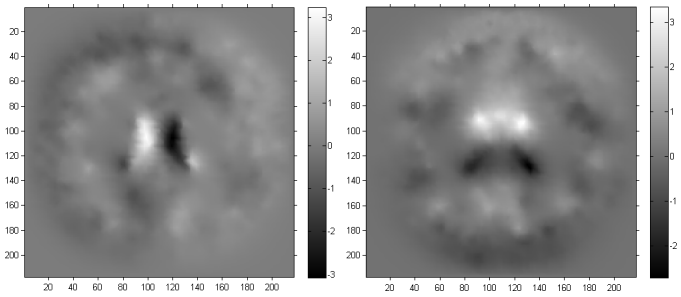
as a non-rigid transformation in the center between the two. Using a content adapted mesh with 5406 nodes and 10744 triangles, the registration is done. The average time for each iteration is about 156 ms for these images. The computed displacement fields as well as the registered image and the difference image after registration can be seen in Figure 2.8. The MSDs before and after registration are 271.8 and 77.3 respectively.

Example 2- Brain CT Database

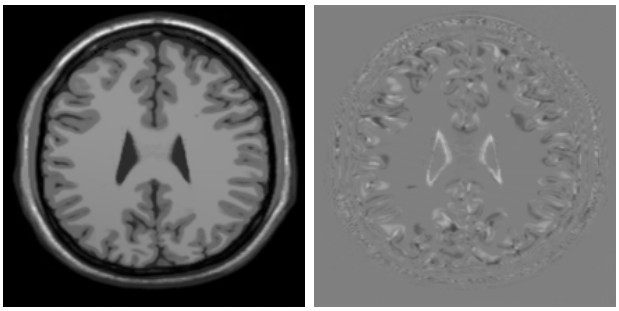
For the second test, a complete database of brain CT images are considered. The database contains 80 images, each of the size 512×512 pixels. Using the content adaptive mesh generation method, a mesh is generated for each image in the database and then used for registration of consecutive slices in the database. Each mesh contains approximately 3300 nodes and 6700 triangles. For better comparison of the speed of the proposed method with pixel-based registration, an implementation of the curvature-based registration method [58] has been used. This implementation takes advantage of a fast Discrete Cosine Transform (DCT) solver. Both of the methods are implemented and tested on MATLAB without any specific optimization and the process of optimization is terminated after 100 iterations. For the pixel-based curvature registration method, the DCT solver is implemented using the embedded DCT function in MATLAB which uses a C implementation, therefore is very fast and optimized while in the implementation of our method, the solver is implemented using MATLAB scripts by the authors. However, the proposed method performs faster. Table 2.2 summarizes the computational time of these two methods, implemented on a desktop computer with an Intel Core i7 3.5 GHz CPU

Table 2.2: Computational time and mean MSD error for pixel-based and mesh-based registration methods

	Pixel-based Method	**Mesh-based Method**
Mean MSD	116.66	108.91
CPU Time	1534 sec	1320 sec



(a)



(b) (c)

Figure 2.12: (a) Displacement fields in horizontal and vertical directions, (b) Registered image, (c) Difference image after registration

and 6 GB of RAM, as well as the mean MSD error of the methods.

2.3.3 Conclusion

A new efficient triangular mesh-based image registration technique is introduced. Table 2.2 illustrates the results of comparison between pixel-based curvature registration method with DCT solver [58] with the proposed mesh-based method. Even though MATLAB is used for both of the methods, one needs to consider that the pixel-based approach is using the internal optimized DCT function (written in C) to solve the linear system at each iteration, unlike the mesh-based technique which uses non-optimized MATLAB functions written by the authors. However the mesh-based technique outperforms the

pixel-based method both in accuracy and speed. Using a MEX or C implementation of the proposed method, higher speeds upto an order of magnitude faster than the MATLAB implementation reported in Table 2.2 are achieved.

Usually image registration techniques, specially the non-parametric ones, work on the pixel level. Multi-resolution techniques do not distinguish between regions that have significant feature content and regions that are featureless/uniform. Octrees are a way to adaptively sub-divide images based on feature content. However, the rectangular boundaries in octrees do no suit feature boundaries that tend to be curvilinear. On the other hand triangular meshes can accurately follow curvilinear feature boundaries. However using triangular mesh has its own implications regarding the definition of the problem of image registration and optimization of the geometric transformation needed to be applied to template image to match the reference image. Here a new technique for fast mesh-based image registration is proposed which can take into account these implications and achieve high accuracy. This method has the dual advantage of a compact representation and fast computation. Furthermore, images at any desired resolution can be considered for registration since we only need to deal with the mesh nodes and not image pixels.

Chapter 3

Sparse Feature-Based Matching for 3D Surface Reconstruction of Multi-View Scanning Electron Microscope Images

3.1 Introduction

3D visualization from a set of 2D images has been an active research area in the last 20 years, with applications in scene reconstruction, movie making, medical visualization, virtual tourism, mobile robot navigation, virtual reality, and computer aided design [2, 188, 189, 200].

Generally speaking, 3D reconstruction techniques can be categorizes into three major classes: 1) single-view, 2) multi-view, and 3) hybrid. In single-view techniques, also known as photometric stereo (PS), 3D surface reconstruction uses a set of 2D images from a single view point but with varying lighting directions. After acquiring the input images, the light directions are determined. This is followed by calculation of the surface normal, albedo, and finally depth estimation. In multi-view class, 3D surface is reconstructed by combining the information gathered from a set of 2D images acquired by changing the imaging view. In such techniques, also known as structure from motion (SFM), at first, feature points are detected in the input images. This is followed by finding the corresponding points in the images (point-matching) and then using the projection geometry theory for estimating the camera projection matrices. Finally, 3D points are generated using linear triangulation. As one can imagine, the hybrid class combines the advantages of the single-view and multi-view techniques for a more accurate 3D reconstruction. Here, the focus is on the multi-view class, especially since for the problem of 3D reconstruction using SEM images, having image set with varying light directions is

difficult, but acquiring images with different titling angles are commonplace. Therefore, having a good understanding of the feature-point detectors and also local descriptors can be beneficial.

Feature point detectors are very popular in many different fields of Computer Vision such as object recognition, 3D reconstruction, image retrieval and camera localization. The found feature points are finally used for finding correspondence between different images. The task of discrete image correspondence consists of three steps: 1) finding interest points, such as corners, blobs and T-junctions, 2) assigning feature vectors to the neighborhood of each interest point, and 3) matching of feature vectors between different images. A wide plethora of detectors and descriptors, as well as their performance comparisons can be found in the literature [115, 120, 94, 131].

The required level of invariance, as well as skew, anisotropic scaling and perspective effects are usually considered when creating new local feature detectors and descriptors. As for the interest point detector, Harris corner detector is the most widely used which is based on the eigenvalues of the second-moment matrix [78]. The work of [115] allowed for automatic scale selection, which is lacking in Harris corner detector, by use of determinant and Laplacian of Hessian matrix, for detection of blob-like structures. This was further refined in the work of [128] and resulted in Harris-Laplace and Hessian-Laplace feature detectors. In terms of speed, the work of [119], replacing Laplacian of Gaussian (LoG) filter with Difference of Gaussian (DoG) filter, not only gives satisfying approximations but also improves the speed. As a conclusion, Hessian-based detectors are more stable and repeatable than their Harris-based counterparts. Also use of determinant rather than the Laplacian can be beneficial. Moreover, higher speeds in detection can be achieved by using approximations like DoG.

For feature descriptors, Gaussian derivatives [60], moment invariants [132], complex features [23] and steerable filters [64] can be mentioned as a few examples. *Scale Invariant Feature Transform* (SIFT) is probably the most well-known descriptor which works based on the histogram of local oriented gradients around the interest point [120]. Refined versions of SIFT like PCA-SIFT [94] and Gradient Location and Orientation His-

togram (GLOH) [130] can also be mentioned. In *Speeded Up Robust Features* (SURF), the detector is based on the Hessian matrix derived from integral images for reducing the computational time. As for the descriptor, a distribution of Haar-wavelet responses in the neighborhood of interest points are used [25].

As a result of increasing use of mobile devices and higher resolution images, the need for a faster and more robust image feature descriptor is of high importance [37]. As for typical SIFT and SURF, at least a 64-vector of 4 byte floating points is needed which adds up to 256 bytes. In general, three classes of approaches are used for reduction of this number [37]:

1. Use of dimensionality reduction techniques such as Principal Component Analysis (PCA) [130] or Linear Discreminant Embedding (LDE) [86];

2. Quantization of floating-point coordinates into integers coded on fewer bits [193, 198, 36]

3. Binarizing the descriptor [175, 192].

All the above mentioned techniques have been satisfactorily used and provided reasonable results. But as it is obvious, computing a long descriptor and then shortening it is not computationally efficient. *Binary Robust Independent Elementary Features* (BRIEF) aims to remedy this problem by building short descriptors directly.

Aiming at lowering the computational costs of common feature detectors and local image descriptors, such as SIFT, ORB is built based on the well-known FAST [162] keypoint detector and BRIEF [37] descriptor, hence called ORB (Oriented FAST and Rotated BRIEF) [164]. As for FAST, even though it is efficient in finding corner keypoints, it lacks the orientation information. So in the first step, orientation operator using a centroid technique [161] is added to FAST. The same problem occurs for BRIEF descriptor which is very sensitive to in-plane rotation. ORB is designed to remedy these problems.

In this chapter, these four well-known methods are compared for surface reconstruction of SEM images. The SEM as a 2D imaging microscope has been widely used in

biological and materials sciences to analyze the surface properties of micro samples. Having 3D surface models from SEM micrographs would provide realistic anatomic shapes of microscopic objects which allows for informative visualization and quantitative measurements of the samples being investigated. The contributions of the chapter can be summarized as follows:

- A new optimized multi-view approach to accurately estimate SEM extrinsic calibration and its 3D surface reconstruction is designed. The method combines multiple view geometry with a global optimization strategy namely Differential Evolutionary (DE) algorithm to reconstruct 3D surfaces from 2D SEM images.
- In the case of multi-view 3D SEM surface reconstruction, the process requires repeated image orientation estimation based on detected and corresponded feature points. The estimation has a crucial impact on the quality of 3D reconstructed shape model. As an important contribution, extensive comparisons are provided to examine the application of SIFT, SURF, BRIEF, and ORB algorithms for both accurate SEM extrinsic calibration and its 3D surface modeling.
- Image feature descriptor algorithms have been widely applied on generic digital images as well as video streams to perform several tasks in computational imaging including image registration, object localization, and object tracking. The current work initiates the study of image feature descriptor algorithms for images obtained by SEM which are naturally different from generic images. This is usually the case since the features are not very distinct in SEM images in comparison to the scenes usually used in computer vision applications which contain sharp edges, multiple objects with distinguishable boundaries and big variations of intensities; not to mention the difference in nature and amount of noise and artifacts within the scenes.

The rest of the chapter is arranged as follows: Section 3.2 contains an overview on the methods. In Section 3.3, the problem of surface reconstruction is discussed in more detail and the use of Differential Evolution (DE) for optimization is discussed. Section

3.4 contains the results of the four methods for 3D surface reconstruction and qualitative and quantitative comparisons will be given. Section 3.5 provides further discussion on the subject and possible enhancements to be considered for future investigations. Section 3.6 concludes the chapter.

3.2 Methods

3.2.1 SIFT

Four stages of SIFT ([120]) are: 1) scale-space extrema detection, 2) keypoint localization, 3) orientation assignment and 4) keypoint descriptors. As the first stage, detection of locations that are invariant to scale changes of the image is of high importance. This can be accomplished by the well received work of [199] on scale space functions. Based on the work of [114], a Gaussian function is considered as the scale-space kernel. The scale-space function of an image, $L(x, y, \sigma)$, can be derived using the convolution of the image $I(x, y)$ with the variable-scale Gaussian function $G(x, y, \sigma)$. Stable keypoint locations can then be detected by finding the scale-space extrema in the difference-of-Gaussian (DoG) function convolved with the image as follows:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \tag{3.1}$$

where k is a constant multiplicative factor. As pointed out by [114], D provides a good approximation for the scale-normalized Laplacian of Gaussian, $\sigma^2 \bigtriangledown^2 G$. Studies of [129] have shown that the extrema of $\sigma^2 \bigtriangledown^2 G$ provide the most stable features. For this, each octave of scale space (doubling of σ) is divided into s intervals, hence $k = 2^{1/s}$.

For detecting the local extrema of $D(x, y, \sigma)$, each sample point is compared with its 26 neighbor points in a $3 \times 3 \times 3$ neighborhood, considering the current image and the above and below images. Frequency of sampling in each scale, as well as, frequency of sampling in the spatial domain are two important parameters which not only affect the repeatability of the algorithm, but also the computational cost. Several detailed tests

in [120] resulted in choosing 3 scale samples per octave and $\sigma = 1.6$ as for the prior smoothing Gaussian.

Accurate localization of the keypoints in the set of candidate keypoints is done by fitting a 3D quadratic function derived from the Taylor expansion of the scale-space function $D(x, y, \sigma)$ as follows:

$$D(X) = D + \frac{\partial D^T}{\partial X}X + \frac{1}{2}X^T\frac{\partial^2 D}{\partial X^2}X \tag{3.2}$$

where $X = (x, y, \sigma)^T$ is the offset from the sample point. Rejecting the unstable points can be done by thresholding the function value at the extremum $D(\hat{X}) = D + \frac{1}{2}\frac{\partial D^T}{\partial X}\hat{X}$. Usually the extrema with $|D(\hat{X})|$ less than 0.03 are removed, which represents the keypoints located in low contrast regions that are highly affected by noise. Thresholding the ratio of principal curvatures can also eliminate poorly defined feature points near the edges. For this, the Hessian matrix H is computed at the location and scale of the keypoint:

$$H(X) = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \tag{3.3}$$

The ratio between the two eigenvalues (r), largest to smallest, can be computed by considering the trace (Tr) and determinant (Det) of the Hessian matrix as follows:

$$\frac{Tr(H)^2}{Det(H)} = \frac{(r+1)^2}{r} \tag{3.4}$$

Usually the keypoints for which the ratio r is less than 10, are eliminated. Rotation invariance can be achieved by assigning proper orientation to each keypoint. Considering the Gaussian smoothed image at each scale, L, the gradient magnitude $m(x, y)$ and orientation $\theta(x, y)$ can be defined by $\sqrt[2]{L_x^2 + L_y^2}$ and $\tan^{-1}\frac{L_y}{L_x}$ respectively, where L_x and L_y are computed using a central difference approximate. Then an orientation histogram is created within a neighborhood of the keypoint. The histogram has 36 bins, with each sample added by a weight computed by its gradient magnitude and also by a Gaussian weighted circular window. Any local peak with 80% value of the highest peak is used for

creating a new keypoint with that orientation.

Final step of the SIFT is creating the local image descriptor. Up until now, location, scale and orientation are determined for each keypoint. The local image descriptor should be computed so it makes the method invariant to differences in illumination and viewpoint. Using the gradients computed at the time of orientation assignment and a Gaussian weighting function with σ equal to one half the width of the descriptor window, the gradient information over each 4×4 subsection in a 16×16 neighborhood around the keypoint are combined into 8 bins histograms, which result in a $4 \times 4 \times 8 = 128$ element feature vector for each keypoint. Normalizing the feature vectors to unit length will reduce the effect of linear illumination change. Thresholding the normalized vector with 0.2 as threshold and re-normalizing it again will reduce the effects of large gradient magnitudes.

3.2.2 SURF

For SURF, scale and rotation invariance are intended in the process of design ([25]). Upright SURF (U-SURF) is the scale-only invariant version of SURF. Starting with the Hessian detector, given a point $X = (x, y)$ in an image I, the Hessian matrix at scale σ is defined as:

$$H(X, \sigma) = \begin{bmatrix} L_{xx}(X, \sigma) & L_{xy}(X, \sigma) \\ L_{xy}(X, \sigma) & L_{yy}(X, \sigma) \end{bmatrix} \tag{3.5}$$

where $L_{xx}(X, \sigma)$ is the convolution of the Gaussian second order derivative with image I at point X. The same goes for $L_{xy}(X, \sigma)$ and $L_{yy}(X, \sigma)$ too. For SURF, a box filter approximation of these Gaussian functions are used. These approximations are represented by D_{xx}, D_{yy} and D_{xy}. Using these approximations, the determinant of the approximate Hessian matrix can be derived as:

$$det(H_{approx}) = D_{xx}D_{yy} - (0.9D_{xy})^2 \tag{3.6}$$

It should be noted that the filter responses are normalized with respect to the size of the mask. Without the need for image pyramids which require Gaussian smoothing and sub-sampling of the images iteratively, here the box filters are up-sampled to achieve the scale information. This leads to masks of sizes 9×9, 15×15, 21×21, 27×27 pixels for different scales. Localizing the interest points in the image and over different scales are done by utilizing a non-maximum suppression in a $3 \times 3 \times 3$ neighborhood. Finally the maxima of the determinant of the Hessian matrices are interpolated in scale and image space with the method proposed in the work of [32].

For defining the descriptor, first the orientation at each interest point needs to be assigned. This is done by combining the results of the Haar wavelets' responses in a circular neighborhood around the interest point, using Gaussian weights. For the case of U-SURF this step is not necessary. Next, a square region centered around the interest point and oriented along the estimated orientation in the previous step is created. The size of this square is $20s$, s being the scale. This region is further divided into smaller 4×4 sub-regions, each of the size 5×5. Considering d_x as the horizontal Haar wavelet response and d_y as the vertical Haar wavelet response, the four-dimensional descriptor V for each sub-region is defined as $V = (\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|)$where the summations are computed over each sub-region. Having a vector of 4 for each of sub-regions (16) leads to a vector of size 64 for each interest point, hence creating the SURF-64 descriptor. Other versions of the descriptor, namely SURF-36 and SURF-128 can be computed in the same manner, using different divisions of the interest region, or computation of the feature vectors in each region.

3.2.3 BRIEF

Classification of image patches can be effectively done by a small number of pairwise intensity comparisons, as previously studied in the work of [142] using Naive Bayesian classifier and also in the work of [107] using randomized classification trees. Based on this assumption, a test τ on patch P of size $S \times S$ can be defined:

$$\tau(P; x, y) := \begin{cases} 1, & \text{if } P(x) < P(y) \\ 0, & \text{otherwise} \end{cases} \tag{3.7}$$

where $P(x)$ is the pixel intensity in a smoothed version of P at $x = (u, v)^T$. A set of binary tests can be defined by choosing a set of n_d (x, y)-location pairs. Therefore the BRIEF descriptor can be defined as an n_d-dimensional bit-string as follows:

$$f_{n_d}(P) := \sum_{1 \leq i \leq n_d} 2^{i-1} \tau(P; x_i, y_i) \tag{3.8}$$

Typically n_d is chosen to be 128, 256 or 512 depending on the desired speed, storage efficiency and rate of recognition. Usually the BRIEF descriptor is referred to as BRIEF-k, with $k = n_d/8$, representing the number of needed bytes for storing the descriptor.

Given that in (7) only the information at single pixels are considered, the need for pre-smoothing on the patches to reduce the sensitivity to noise becomes more apparent. Generally speaking, more difficulty in matching requires more smoothing which means larger variances in Gaussian kernels. Typically the variance is chosen to be 2 and the discrete kernel window is of size 9×9. Another important factor is how to choose n_d test locations (x_i, y_i) to be used in (7) for generating the descriptor. In the work of [142] five different sampling geometries are tested as follows:

1. (X, Y) i.i.d. Uniform$(-\frac{S}{2}, \frac{S}{2})$;
2. (X, Y) i.i.d. Gaussian$(0, \frac{1}{25}S^2)$;
3. X i.i.d. Gaussian$(0, \frac{1}{25}S^2), Y$ i.i.d. Gaussian$(x_i, \frac{1}{100}S^2)$;
4. (x_i, y_i) randomly sampled from discrete locations of a coarse polar grid;
5. $\forall i : x_i = (0, 0)^T$ and y_i all possible values on a coarse polar grid containing n_d points.

Tests conducted reveal that second, third and fourth strategies are superior to the rest, hence usually utilized in BREIF description generation. Last but not least, the

distribution of Hamming distances between the defined descriptors plays a significant role in the amount of recognition rate.

Different tests reveal the superiority of the BRIEF descriptor, in terms of computational speed. In terms of accuracy and recognition rate, since the descriptor lacks rotational invariance, satisfied by SURF for example, BRIEF performs poorly in matching pairs that have big rotation angles.

3.2.4 ORB

ORB is built based on the well-known FAST ([162]) keypoint detector and BRIEF ([37]) descriptor, hence called Oriented FAST and Rotated BREIF (ORB). In FAST, corners are detected by applying an intensity threshold in a circular ring around the center. Usually the radius of the circle is mentioned in the name of the FAST operator, for example FAST-9. In ORB, to make the FAST less responsive along the edges, the Harris corner measure ([78]) is used for ordering the found keypoints in terms of their *cornerness*. Also, a scale pyramid scheme is used for detecting corners in different scales. For the intensity centroid technique for orientation assignment, assuming the general equation for the moments of a patch as:

$$m_{pq} = \sum_{x,y} x^p y^q I(x,y) \tag{3.9}$$

$I(x,y)$ being the image, the centroid can be found as follows:

$$C = (\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}}) \tag{3.10}$$

Constructing a vector from the corner's center, O, to the centroid $\vec{OC}$, the orientation of the patch is $\theta = atan2(m_{01}, m_{10})$, $atan2$ being the quadrant-aware arctangent.

Due to lack of orientation invariance, BRIEF's matching performance is very poor in case of having rotations of bigger than a few degrees. To remedy this, [164] proposed a learning technique over all possible binary tests, to find the ones with high variance and less correlation to ensure higher discriminativity. Creating a set of 300k keypoints from

the PASCAL 2006 database ([53]), as well as all the binary tests drawn from a 31×31 pixel patch is the first step, each test is a pair of 5×5 sub-window of the patch. After running each test against all training patches, the tests are ordered by their distance from a mean of 0.5, creating the vector T. The next step is a greedy search in order to find the 256 most discriminative tests.

3.3 3D Surface Reconstruction of Electron Microscopy Images

3.3.1 Scanning Electron Microscope Imaging

The Scanning Electron Microscope (SEM) utilizes electrons instead of light to determine the surface characteristics of microscopic samples. An SEM includes five principle parts as follows: 1) An electron gun, 2) Scanning system, 3) Detectors, 4) Lens control, and 5) Display monitors ([30]).

Electron beams which are emitted from the electron gun in a vacuum are able to enforce two different signals: Secondary Electrons (SE), and Back-Scattered Electrons (BSE). Each signal produces different type of images. While SE can exhibit greater resolution and topography on the surface, BSE can provide greater contrast and brightness between materials comprising a microscopic sample ([146]).

Since SEM produces 2D images, to effectively visualize and measure the surface properties, we need to reconstruct the 3D surface shape from SEM images. Restoring 3D surface models from SEM micrographs would provide realistic anatomic shapes of micro objects which definitely allow for informative visualization and quantitative measurements of the system being investigated.

3.3.2 Feature Detectors/Descriptors for 3D SEM Surface Reconstruction

The proposed framework described here is based on an optimized multi-view approach. The major part of the framework is doing a refinement process by defining a cost function for any set of parameters (initial position of 3D points and extrinsic SEM parameters which specify rotation and translation from one view point to another) to find the best fitness in the set. Therefore, parameterization of rotation and translation space is the most important part of the work. To this end, estimation of the rotation and translation from the set of corresponding points between two images in the image set should be the first step. This optimization is unlike 2D image registration problems (rigid, deformable etc) in which the motion of the objects happen in an *in-plane* manner ([18, 16]). Here, since the position of the 3D object is changing with respect to the camera view-point, additional constraints are needed to be applied for the estimation of motion which results in a more accurate representation of the 3D object. For this purpose, we employ highly used feature extractor algorithms such as SIFT, SURF, BRIEF, or ORB along with KNN ([5]) and RANSAC ([59]) to find true matching points in the image pair. We next take advantage of epipolar geometry ([79]) to estimate the rotation and translation, and perform linear triangulation to initialize the 3D position of all corresponding points. The last step is a refinement process by defining a cost function for any set of parameters as to whether the set is a good or bad set. Figure 3.1 shows the pipeline of our proposed system.

To perform optimization, the most important part is to parameterize the space of rotation and translation. For the purpose of better flexibility, the quaternion parameterization ([79]) is applied to formulate the 3D rotation.

A quaternion is represented as $\mathbf{z} = a + bi + cj + dk$, where a, b, c, d are real numbers and $i^2 = j^2 = k^2 = -1$. $\mathbf{z}$ is a unit quaternion if and only if:

$$|z| = \sqrt{a^2 + b^2 + c^2 + d^2} = 1 \tag{3.11}$$

The rotation matrix representation is as follows:

$$R(z) = \begin{bmatrix} a^2+b^2-c^2-d^2 & 2bc-2ad & 2bd+2ac \\ 2bc+2ad & a^2-b^2+c^2-d^2 & 2cd-2ab \\ 2bd-2ac & 2cd+2ab & a^2-b^2-c^2+d^2 \end{bmatrix} \tag{3.12}$$

The translation vector of the second position with respect to the first position is defined as $t = (t_x, t_y, t_z)^\top$. Considering the Equation (3.12) for rotation parameterization and t for translation, the parameterization of two projection matrices will be determined by a seven-dimensional vector $\theta = (a, b, c, d, t_x, t_y, t_z)^\top$. Now, SEM extrinsic calibration is equivalent to determining the parameter vector θ^* as Equation (3.13). In this equation, P is the SEM projection matrix which encapsulates rotation and translation (step (4) in Figure 3.1).

$$\theta^* = \arg\min_{\theta}(\sum_{i=1}^{N} \|x_1^i - P(X_i)\|^2 + \|x_2^i - P(\theta, X_i)\|^2) \tag{3.13}$$

In recent and generic 3D surface reconstruction models, the iterative bundle adjustment strategies were employed to solve this kind of equation. The bundle adjustment algorithms are among local minimizer techniques which suffer from different problems. For example, they usually work on differentiable functions only and it is important to have an initial guess close to the real answer to converge. In contrast to the traditional bundle adjustment approaches, Differential Evolutionary (DE) algorithm ([40]) is a global and stochastic optimization approach which does not tolerate these constraints and is known as one the fastest and most reliable evolutionary algorithms to optimize real number functions. DE as a genetic searching based optimization algorithm uses generated populations within the parameter space, then iteratively updates them to find the best possible fitness for an optimization problem. The initial population is modified from one generation to another by using two major operators: 1) Mutation, and 2) Crossover ([40]). The population generation will continue until a termination condition is met (i.e.

number of generations). In the proposed framework, we use a DE based algorithm to find the best values for both 3D position of all matching points as well as the extrinsic parameter of the SEM.

We define $\theta_{k,G}$ as the k-th parameter vector in the G-th generation by:

$$\theta_{k,G} = (a_{k,G}, b_{k,G}, c_{k,G}, d_{k,G}, t_{xk,G}, t_{yk,G}, t_{zk,G}) \tag{3.14}$$

where k=$(1, 2, ..., P_{total})$, and G=$(1, 2, ..., G_{max})$. We assign the size of the population to $POPULATION_{total}$, and the maximum number of generations to G_{max}. We then use the mutation operator $\mathbf{p_{k,G}} = \theta_{p,G} + S \times (\theta_{q,G} - \theta_{r,G})$ to expand deviation from one generation to the next in the population set. S $\in [0, 2]$ and $\theta_{p,G}$, $\theta_{q,G}$, $\theta_{r,G}$ are three individual random agents in the population. The DE algorithm for solving the problem in Equation (3.13) is presented in Algorithm 1. Based on the DE description, the best values of parameters CR $\in [0, 1]$ and S $\in [0, 2]$ would be achieved by executing several experiments on the problem. We started with seven-dimensional parameter vector (θ^*) which is randomly assigned from the uniformly distributed numbers in the range (0,1) at generation G=1. After that, in each generation (G+1), a new parameter vector consisting of rotation and translation will be generated by adding the weighted difference vector between two population members to a third member. After G_{max} iterations (termination condition) ensuring convergence (G_{max}=1000 based on our experiments), the population member θ^* with the highest fitness is evaluated to present the best solution to the 3D SEM surface reconstruction problem.

Algorithm 1. Proposed DE algorithm for 3D SEM surface reconstruction

Input: Matching points, initial SEM extrinsic parameters and 3D locations

Output: The best fitness of SEM extrinsic parameters and 3D points

```
begin
 Doing Initialization:
 set S, CR, POPULATION_Total, G_max;
 Initialize the population {θ_k;(1<=k<=POPULATION_Total)} randomly;
```

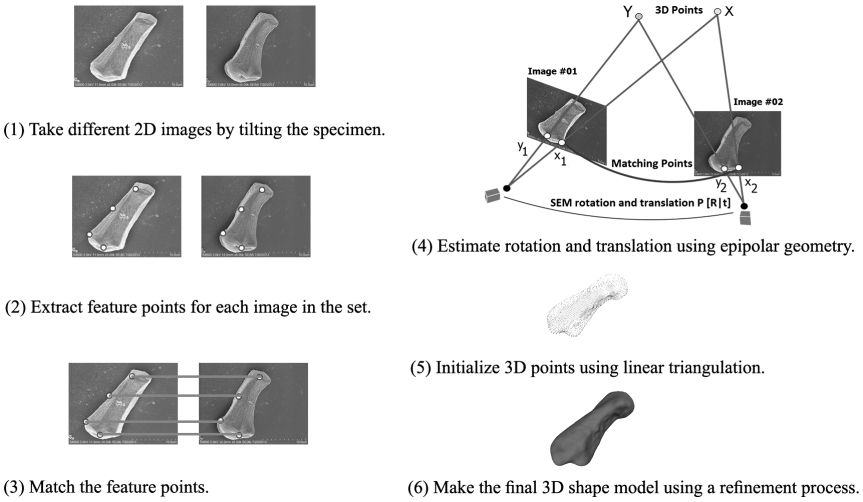


Figure 3.1: Pipeline of 3D surface reconstruction from Electron Microscopy images. At first, multiple images of the specimen from different perspectives are taken. Then, we estimate the relative position including rotation and translation based on the matching points in the image set. After estimating the image motions, the 3D position of all corresponding points would be reconstructed using linear triangulation. The final step is doing an optimization process to find the best match for SEM's extrinsic parameters and all of the initial 3D points.

```
for (G=1; G < G_max; G ++)
  for (k=1; k<=POPULATION_Total; k++)
    Doing Mutation and Crossover operations:
     select three individual agents θ_{p,G}, θ_{q,G}, θ_{r,G} randomly;
       L = U(0,1);
       if L < CR
        p_{k,G} = θ_{p,G} + S × (θ_{q,G} − θ_{r,G})
      else
        p_{k,G} = θ_{k,G};
      if  p_{k,G} < θ_{k,G};
       θ* = p_{k,G};
    end.
  end.
return  θ*
end.
```

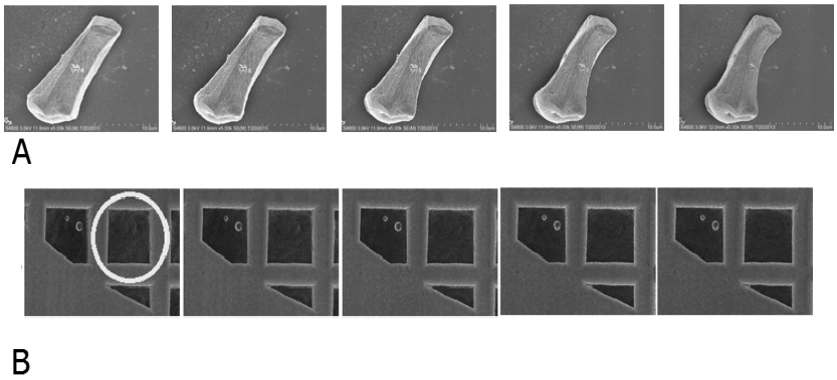


Figure 3.2: (A) Shows a set of 2D images from *tapetal cell* of *Arabidopsis thaliana*, which is obtained by tilting the specimen stage 9 degrees from one to the next in the image set. (B) Shows five images from TEM *copper grid*. These images were obtained by tilting the specimen stage 7 degrees from one to the next in the image sequence. The white circle specifies a part of the specimen which will be 3D reconstructed by the proposed framework.

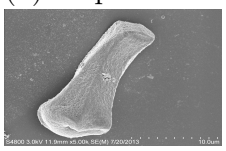
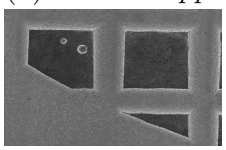
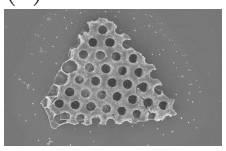
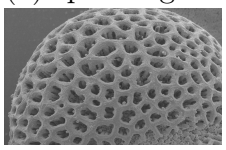
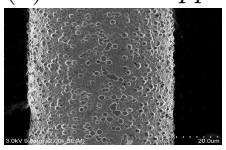
3.4 Experimental Results

To evaluate and compare the performance, reliability, and accuracy of the SIFT, SURF, BRIEF, and ORB algorithms in 3D SEM surface reconstruction, several experiments are carried out in this section. In Section 3.4.1, the experimental setup and images' properties will be explained. Section 3.4.2 contains 3D visualization results obtained using the SIFT, SURF, BRIEF, and ORB algorithms for qualitative comparisons.

3.4.1 Experimental Setup

The major parts of the proposed 3D SEM surface reconstruction framework were developed in Java SE 7 and MATLAB 2012a. We employed 64-bit Windows 7 Operating system on a PC with 3.00 GHz Intel Dual core CPU, 4MB cache, and 4GB of RAM. Image sets and their attributes along with SEM configurations are shown in Table 1. Figure 2 shows only two image sets including *tapetal cell* of *Arabidopsis thaliana* and TEM *copper grid* obtained from different view points. The term TEM stands for Transmission Electron Microscopy.

Table 3.1: Experimental setup including image sets, and SEM configuration.

Image sets	(1): tapetal cell of *Arabidopsis thaliana*  (2): TEM *copper grid*  (3): diatom frustule  (4): pollen grain from *Brassica rapa*  (5): TEM *copper grid* bar 
Images properties	(1): 2560*1920 grayscale, 512 dpi (2): 2560*1920 grayscale, 512 dpi (3): 2560*1920 grayscale, 512 dpi (4): 854*640 grayscale, 512 dpi (5): 512*384 grayscale, 512 dpi
Number of Images	(1): 5 SEM images (2): 5 SEM images (3): 3 SEM images (4): 4 SEM images (5): 5 SEM images
Rotation Angle	(1): 9 degrees (2): 7 degrees (3): 15 degrees (4): 3 degrees (5): 11 degrees
SEM detector	SE (mix)

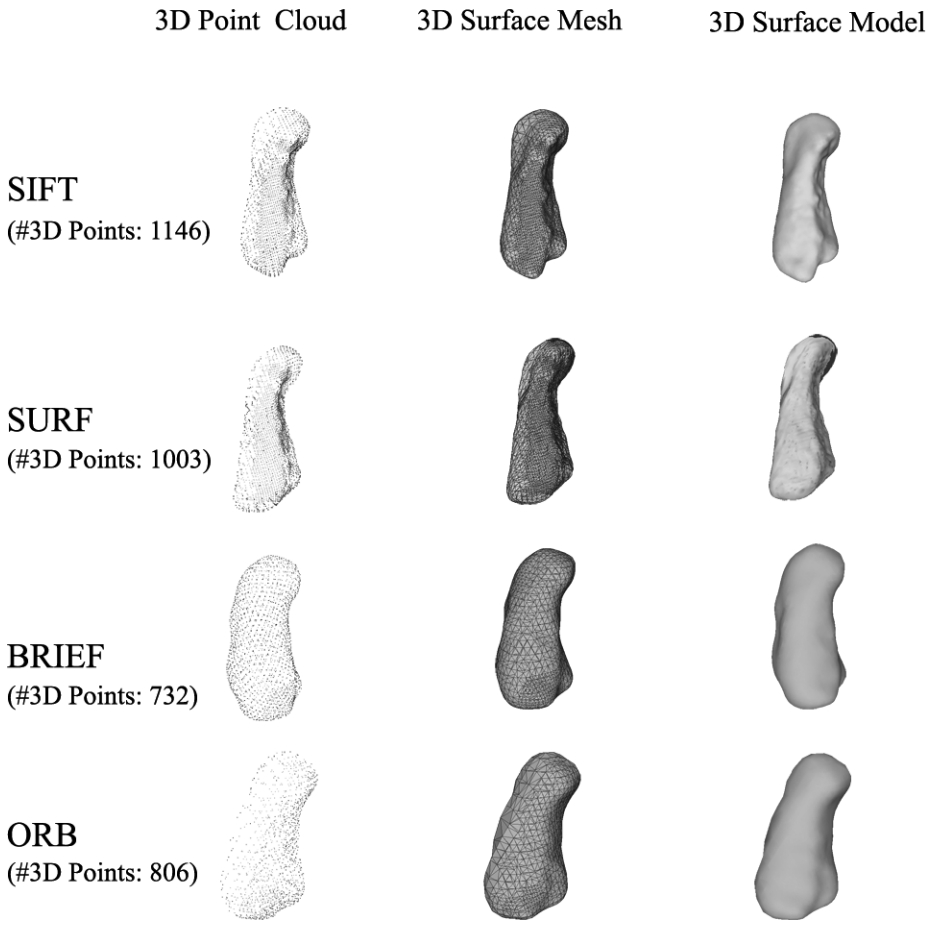


Figure 3.3: Qualitative visualization of 3D SEM surface reconstruction of *tapetal cell* of *Arabidopsis thaliana* using different image feature detector algorithms including SIFT, SURF, BRIEF, and ORB.

3.4.2 Qualitative 3D Visualization

3D point clouds, 3D surface meshes, and 3D shape models of *tapetal cell* of *Arabidopsis thaliana* and TEM *copper grid* which were reconstructed by using the proposed framework using different image feature detector algorithms are shown in Figures 3.3 and 3.4. By considering 2D images, it seems that the SIFT and SURF algorithms would assist creating more realistic 3D surface models than ORB and BRIEF.

3.4.3 SEM Extrinsic Calibration

In this experiment, we examine and compare the accuracy and reliability of the SIFT, SURF, BRIEF, and ORB image feature detector algorithms along with the proposed framework to SEM rotation estimation. We are given the rotation angles, however the

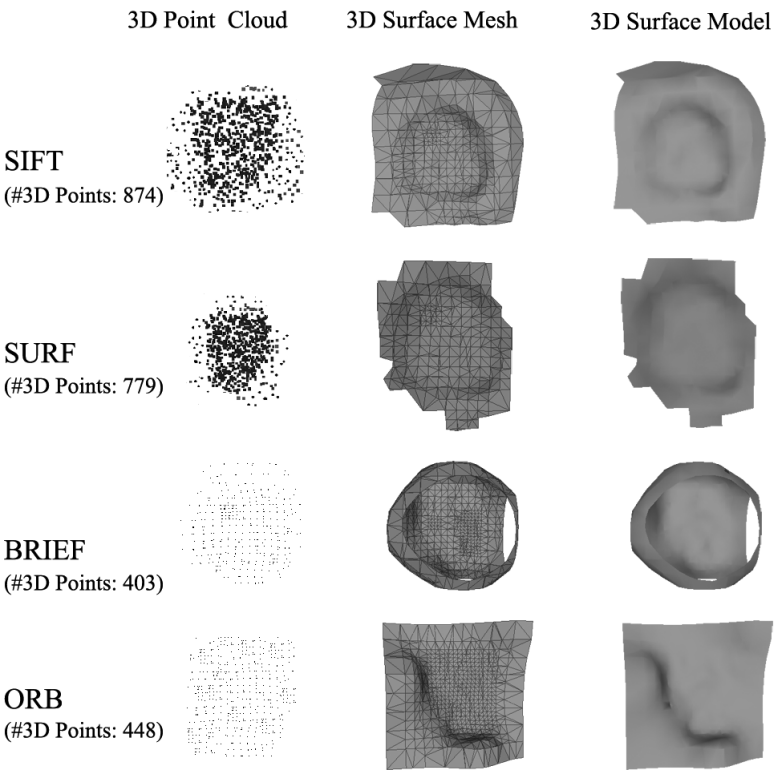


Figure 3.4: Qualitative visualization of 3D SEM surface reconstruction of TEM *copper grid* using different image feature detector algorithms including SIFT, SURF, BRIEF, and ORB.

rotation axis and translation vector are not provided by SEM as a predefined rule. Here, we only use two images in each image set and set the maximum number of DE generations to 1000. We chose 1000 number of DE generations based on our previous experiments illustrated in ([188]).

We eventually got rotation matrices R^t_{SIFT}, R^t_{SURF}, R^t_{BRIEF}, R^t_{ORB} and translation vectors t^t_{SIFT}, t^t_{SURF}, t^t_{BRIEF}, t^t_{ORB} for the *tapetal cell* of *Arabidopsis thaliana* (tilting by 9 degrees) as follows. Each matrix is labeled by its associated image feature detector.

$$R^t_{SIFT} = \begin{bmatrix} 1.0000 & 0.0005 & 0.0015 \\ 0.0007 & 0.9822 & -0.1877 \\ 0.0014 & 0.1877 & 0.9822 \end{bmatrix}, R^t_{SURF} = \begin{bmatrix} 1.0000 & 0.0006 & 0.0019 \\ 0.0009 & 0.9761 & -0.1973 \\ 0.0014 & 0.1973 & 0.9761 \end{bmatrix}$$

$$R^t_{BRIEF} = \begin{bmatrix} 0.9407 & 0.0009 & 0.0031 \\ 0.0003 & 0.9101 & -0.2408 \\ 0.0031 & 0.2408 & 0.9101 \end{bmatrix}, R^t_{ORB} = \begin{bmatrix} 1.0000 & 0.0004 & 0.0029 \\ 0.0007 & 0.9361 & -0.2053 \\ 0.0029 & 0.2053 & 0.9361 \end{bmatrix}$$

$$t^t_{SIFT} = [0.3727 \quad 0.1020 \quad 0.0002], t^t_{SURF} = [0.2347 \quad 0.1015 \quad 0.0017]$$

$$t^t_{BRIEF} = [0.1799 \quad 0.0043 \quad 0.0008], t^t_{ORB} = [0.1915 \quad 0.1060 \quad 0.0012]$$

By doing the same experiments on TEM *copper grid* (tilting by 7 degrees), We obtained a rotation matrices R^c_{SIFT}, R^c_{SURF}, R^c_{BRIEF}, R^c_{ORB} and translation vectors t^c_{SIFT}, t^c_{SURF}, t^c_{BRIEF}, t^c_{ORB} as follows:

$$R^c_{SIFT} = \begin{bmatrix} 1.0000 & 0.0011 & 0.0006 \\ 0.0003 & 0.9928 & -0.1311 \\ 0.0019 & 0.1311 & 0.9928 \end{bmatrix}, R^c_{SURF} = \begin{bmatrix} 1.0000 & 0.0016 & 0.0001 \\ 0.0007 & 0.9703 & -0.1689 \\ 0.0021 & 0.1689 & 0.9703 \end{bmatrix}$$

$$R^c_{BRIEF} = \begin{bmatrix} 1.0000 & 0.0009 & 0.0004 \\ 0.0006 & 0.9033 & -0.2317 \\ 0.0014 & 0.2317 & 0.9033 \end{bmatrix}, R^c_{ORB} = \begin{bmatrix} 0.9817 & 0.0030 & 0.0027 \\ 0.0007 & 0.9451 & -0.2018 \\ 0.0025 & 0.2018 & 0.9451 \end{bmatrix}$$

$$t^c_{SIFT} = [0.1007 \quad 0.0019 \quad 0.0029], t^c_{SURF} = [0.1304 \quad 0.0109 \quad 0.0013]$$

$$t^c_{BRIEF} = [0.8871 \quad 0.0044 \quad 0.0043], t^c_{ORB} = [0.9413 \quad 0.0026 \quad 0.0028]$$

By assuming the basic 3D rotation matrices presented in ([79]), it asserts that the rotation axis using all four image feature detectors are over the X axis, and they offer promising results. However the rotation and translation estimations are a little different from each other.

Since we already had a ground truth rotation in our image set, in the next experiment, we examine the 3D rotation estimation error (ΔR). Results obtained by this experiment are shown in Table 2. The results presented in Table 2 show that the SIFT could assist to estimate the rotation angle in a more accurate fashion than the SURF algorithm, and the SURF detector is better than the BRIEF and ORB.

3.5 Discussion

This chapter presented a comparative study on the performance, accuracy, and reliability of four feature detector algorithms including SIFT, SURF, BRIEF, and ORB on 3D SEM surface reconstruction. In the experiments demonstrated here, promising results on SEM extrinsic calibration and its 3D surface modeling are achieved. The calibration measures obtained from this experiment have been based on only rotation and translation without flipping and scaling. We obtained the 2D SEM micrographs by rotating the SEM stages around different degrees (3-11 degrees), and we know that the SIFT, SURF and ORB are rotation invariant while BRIEF is not. As shown in Table 3.2, while all four algorithms were enable to estimate the rotation axis, employing the BRIEF algorithm provided less precise results in SEM rotation calibration, and in the other side, the SIFT algorithm is the most accurate between these feature point detectors regarding the SEM extrinsic calibration. Since a SEM cannot supply a predefined translation information, We never performed ground truth evaluation for translation vectors, but our estimation for translation calibration have worked for 3D SEM surface modeling. As indicating in Figures 3.3 and 3.4, SIFT and SURF algorithms produced much more convenient 3D surface models by detecting more image features and better estimating the SEM parameters. Overall, from the presented results, it can be seen that SIFT outperforms the rest of the methods in terms of accuracy in the final estimation. Of course this is mainly due to the higher

Table 3.2: Accuracy and reliability validation of the different image feature detector algorithms by employing the proposed 3D SEM surface reconstruction framework. ΔR is given as $R_{real} - R_{estimated}$, indicating error for estimating the 3D rotation. Rotation angles show the ground truth 3D SEM rotations(R_{real}). In each row we used only two images in the set.

Image set	Total matches	Rotation angle	ΔR	Feature detector algorithm
tapetal cell	509	9 degrees	**5.07E-04**	SIFT
	438	9 degrees	5.48E-04	SURF
	219	9 degrees	4.71E-03	BRIEF
	255	9 degrees	3.29E-03	ORB
	441	18 degrees	**7.12E-04**	SIFT
	402	18 degrees	8.05E-04	SURF
	184	18 degrees	6.44E-03	BRIEF
	247	18 degrees	5.91E-03	ORB
TEM copper grid	830	7 degrees	**8.33E-04**	SIFT
	664	7 degrees	8.73E-04	SURF
	283	7 degrees	4.91E-03	BRIEF
	304	7 degrees	3.97E-03	ORB
	722	14 degrees	**8.84E-04**	SIFT
	601	14 degrees	8.93E-04	SURF
	241	14 degrees	5.14E-03	BRIEF
	298	14 degrees	4.19E-03	ORB
diatom frustule	317	15 degrees	**7.13E-04**	SIFT
	291	15 degrees	7.22E-04	SURF
	173	15 degrees	5.03E-03	BRIEF
	229	15 degrees	4.16E-03	ORB
	286	30 degrees	**7.41E-04**	SIFT
	270	30 degrees	8.05E-04	SURF
	114	30 degrees	5.41E-03	BRIEF
	202	30 degrees	4.64E-03	ORB
pollen grain	749	3 degrees	**8.33E-04**	SIFT
	721	3 degrees	8.91E-04	SURF
	610	3 degrees	4.17E-03	BRIEF
	583	3 degrees	4.45E-03	ORB
	673	6 degrees	**8.84E-04**	SIFT
	654	6 degrees	8.97E-04	SURF
	511	6 degrees	4.34E-03	BRIEF
	475	6 degrees	4.81E-03	ORB
TEM copper grid bar	837	11 degrees	**7.11E-04**	SIFT
	796	11 degrees	7.33E-04	SURF
	588	11 degrees	5.02E-03	BRIEF
	649	11 degrees	4.19E-03	ORB
	802	22 degrees	**7.59E-04**	SIFT
	781	22 degrees	7.84E-04	SURF
	546	22 degrees	5.57E-03	BRIEF
	611	22 degrees	4.91E-03	ORB

number of matched points between image pairs under investigation. SURF performs in the same order as SIFT, even with fewer number of matched points. ORB and BRIEF rank next. This is mainly because they are not designed to be rotation invariant. However, a close inspection of the produced results reveal the major shortcomings of sparse feature-based approaches. Even though we deduced that SIFT feature detector/descriptor performs better than the rest in estimating the fundamental matrix and SEM extrinsic calibrations, however, the final 3D surface reconstruction suffers greatly from smoothed edges and boundaries. This is due to limited number of matched pixel between the images. Use of dense descriptors for matching between the image pairs can be considered as means for increasing the quality of the resulted 3D reconstructed surface [11]. In such techniques, the first step of matching which involves keypoint detection is eliminated and the descriptors are computed for all the pixels of the images. This requires a different formulation of the energy functional for optimization, with additional constraints to ensure smooth flow fields from one image to another. This will be the focus of the next two chapters of the dissertation.

3.6 Conclusion

The scanning electron microscope (SEM) utilizes electrons instead of light for imaging of microscopic structures. A typical system usually consists of an electron gun, a scanning system, set of detectors, lens control equipment and displaying monitors. The images acquired from SEM are 2D images of the microscopic structures. As is obvious, visualization and more accurate investigation of the structures requires a 3D model instead of a set of 2D images. In this chapter the use of four well-known feature descriptors, SIFT, SURF, BRIEF and ORB are further investigated for 3D surface reconstruction of SEM imaging data. The common steps in such techniques are usually interest points (corners, blobs, T-junctions) extraction, descriptor vector assigning in the neighborhood of each interest point and finally matching these feature points between different images and forming correspondences to be used for further analysis. Starting from the feature

points and descriptors created by the four above-mentioned methods along with nearest neighbor (NN) search and RANdom SAmple Consensus (RANSAC), the matching points between image pairs are detected. taking advantage of epipolar geometry the rotation and translation matrices are estimated and 3D position of all the corresponding points are initialized using linear triangulation. The final step is the use of Differential Evolution (DE) for finding the best fitness for SEM extrinsic parameters and 3D points.

Multiple tests are conducted and qualitative and quantitative measures are presented for better comparison of the used feature detectors and descriptors for 3D reconstruction of SEM image data. In all of the cases, SIFT performs better than the others, with SURF being the next best method. Of course, number of matched points play a significant role in rotation/translation estimation and subsequently in the final 3D reconstructed surface. The difference is the design process of the said feature detectors/descriptors and the fact that some are not rotation invariant while some are. However, the outcome will help us build a more rigorous understanding of the problem in order to propose a more accurate and robust framework for generating high quality 3D reconstructions. This will be the focus on the next two chapters of the dissertation.

Chapter 4

Sparse-Dense Correspondence for High Quality 3D Reconstruction of Microscopic Samples

4.1 Introduction

Scanning Electron Microscopy (SEM) imaging has been a crucial technique of various studies in biomedical, mechanical, and materials sciences [31, 51, 88]. The SEM has contributed tremendously to observations of surface structure of microscopic samples on a variety of scales down to $1nm$, employing magnification factors of around $\times 100k$. In a SEM, the source of illumination comes from a focused electron beam which scans the surface of the sample, interacting with atoms of the surface. The secondary electron (SE) or back-scattered electron (BSE) detectors are aimed to capture the signals generated by interactions of the beam with the surface. The detection of BSE signal has proven to be beneficial to compositional studies of materials, while SE suits topographical analysis of the samples being examined. However, SEM micrographs still remain two-dimensional (2D). Therefore, having a high fidelity three dimensional reconstruction for a more effective analysis of the surface attributes and topography of the microscopic sample is of high importance. This has attracted many researchers to devise robust and reliable algorithms for 3D reconstruction of microscopic samples captured by SEM [187, 188, 186, 167, 124, 52, 45].

Multiview stereopsis has been an active research area in computer vision community in the recent years, with applications in scene reconstruction, movie making, medical visualization, virtual tourism, mobile robot navigation, virtual reality, and computer aided design [140, 2, 80]. General 3D scene reconstruction techniques can be categorized into three major classes: 1) single-view (also known as Shape from Shading (SfS)), 2)

multi-view (also known as Shape from motion (SfM)), and 3) hybrid [188]. In single-view class, a set of 2D images from a single viewpoint with varying lighting conditions are used for reconstruction. In contrast, in multi-view class, 3D surface is reconstructed by combining the information gathered from a set of 2D images acquired by changing the imaging viewpoint. In such techniques, at first, feature points are detected in the input images. This step is followed by finding the corresponding points in the images and then using the projection geometry theory for estimating the camera projection matrices. The hybrid class combines the advantages of the single-view and multi-view techniques for a more accurate 3D reconstruction [188]. Here, we put the focus on the multi-view class and therefore, having a good understanding of the feature-point detectors and also local descriptors is beneficial.

Feature detection and feature description are among the core components in many computer vision algorithms and a wide range of approaches and techniques have been introduced in the past few decades to address the need for more robust and accurate feature detection/description. Even though there is no universal and exact definition of a *feature* that is independent of the specific application intended, methods of feature detection can be categorized into four major categories [26]: edge detectors, corner detectors, blob detectors and region detectors. The process of feature detection is usually followed by feature description which uses a set of algorithms for describing the neighborhood of the detected features. Generally speaking, the methods of feature description can also be classified into four major classes [26]: shape descriptors, color descriptors, texture descriptors and motion descriptors. By detecting the features and defining the descriptors, one can use the new representation of the input images for a wide range of applications such as wide baseline matching, object and texture recognition, image retrieval, robot localization, video data mining, image mosaicing and recognition of object categories [183, 200].

Direct application of general purpose multi-view sparse/dense 3D reconstruction approaches for reconstruction of microscopic samples can be problematic. In single-view 3D surface reconstruction, creating a full model of the microscopic sample is generally not

possible since the images are limited to only one view-point. Moreover, recreating the SEM micrographs of the sample under different illumination conditions is difficult. On the other hand, multi-view approaches offer a more general and achievable framework for the task. Using such techniques, a more realistic and complete reconstruction can be created. However, the need for more sophisticated matching methods that require higher computational power is inevitable. Use of multi-view techniques for 3D reconstruction of Scanning Electron Microscopy (SEM) images have been explored in the literature in the past few years [187, 186, 167, 52, 220]. Still, there is room for improvement in accuracy and the needed computational resources for 3D reconstruction. Sparse feature based approaches aim to find a set of features in the input images to be represented by an aggregated set of descriptors acquired from their neighborhood. After matching between the features in the images acquired from multiple viewpoints, the projective transformations between the matches are estimated and the set of 3D points are generated. The bottleneck of such techniques is in the first step of the procedure: feature detection. General purpose feature detection techniques are designed for detection of common features that are seen in everyday images and not necessarily the ones that may be present in SEM micrographs. These general features include edges, corners, T-junctions, blobs etc. This is mainly problematic due to presence of noise that can be observed in SEM micrographs. This makes the feature detection unreliable and therefore, there is need for manual adjustment of various parameters involved in the process. Moreover, the microscopic samples/surfaces to be imaged may contain areas with minimal intensity and depth variations in which no features can be detected. This causes the features, and subsequently 3D points, to be distributed non-uniformly. This will greatly affect the subsequent mesh generation and surface reconstruction steps. It should also be noted that several images from different view points are needed for building a realistic reconstruction of the microscopic sample. This increases the computational complexity of the methods. Still, even with the use of multiple views, very fine details are missed when using sparse feature based approaches. In case of micrographs such as those acquired in biological studies, the combination of numerous microscopic objects with varying sizes cause the sparse feature-based approach

to fail. These problems can be remedied by using dense correspondence between multi-view pixels. This can enable realistic topographical reconstruction of the samples and also eliminate the manual effort needed for acquiring more samples.

In this chapter the use of sparse and dense correspondence for high resolution 3D reconstruction of stereo SEM micrographs is introduced and investigated in great depth and detail. Using the proposed approach, we are able to reconstruct high quality uniform meshes of the imaged surfaces which can be later used for further quantitative analysis regarding the topology and surface attributes. The contributions of the current work can be summarized as follows:

1. A new optimized framework for high fidelity 3D surface reconstruction from multi-view SEM micrographs is designed and developed. This is achieved by combining a sparse feature matching approach with high quality dense matching which results in a highly realistic reconstruction of the microscopic sample.

2. An *a contrario* RAndom SAmple Consensus (RANSAC) based sparse matching methodology for eliminating manual specification of parameters along with a quasi-euclidean epipolar rectification for improving depth estimation.

3. A new method combining sparse-dense SIFT feature matching and representation of the energy minimization functional as a factor graph with loopy belief propagation (LBP) optimization for accurate construction of dense vertical/horizontal displacement maps. Given the result of stereo rectification step, the disparities are grossly concentrated along the horizontal direction. This will simplify the process of depth estimation since horizontal disparities are directly proportional to the actual depth. The appropriate coefficient can be estimated by the known titling of the specimen stage during imaging.

The rest of the chapter is organized as follows. Section 4.2 contains detailed explanation on the various steps of the proposed method. It starts with a brief overview of the proposed method followed by subsections on SEM imaging protocol, Scale Invariant Feature Transform (SIFT) feature detection and description, epipolar rectification

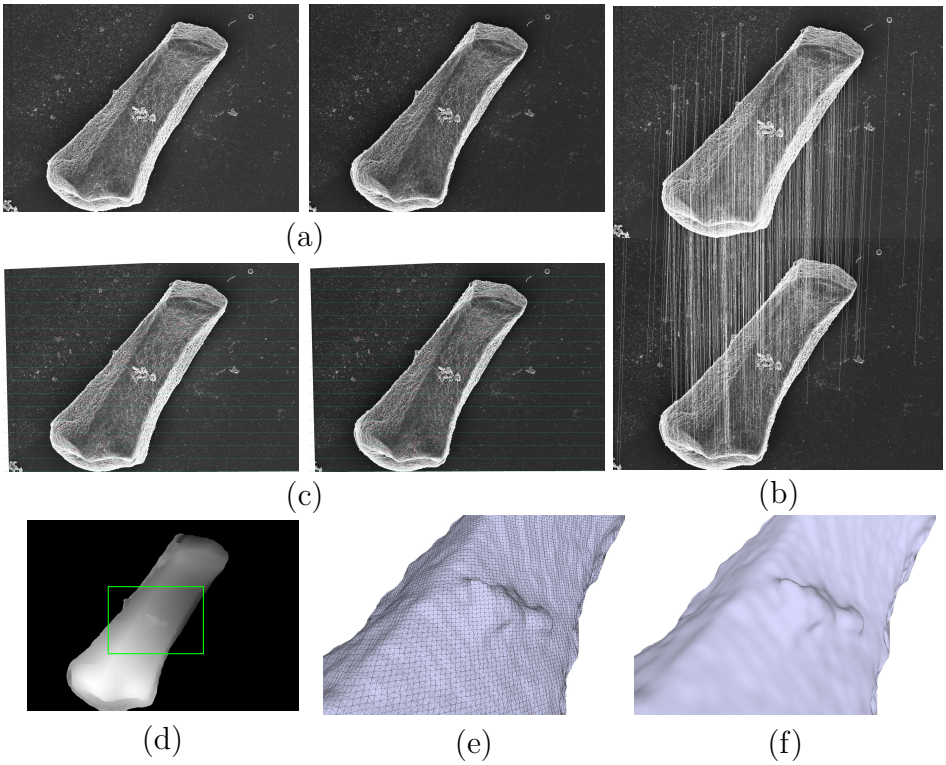


Figure 4.1: Overview of the proposed method for high fidelity 3D reconstruction of microscopic samples using a pair of stereo SEM images captured by tilting the specimen stage: (a) a set of stereo SEM images of a *Tapetal Cell* with known tilting angle (9 degrees), (b) result of sparse SIFT feature detection/description with *a contrario* RANSAC approach for outlier removal, (c) set of rectified images with horizontal epipolar lines being demonstrated, (d) bilateral filtered horizontal disparity map, (e) a magnified view of the high quality surface mesh generated using the dense point cloud, and (f) a magnified view of the high fidelity surface model. The proposed method is able to reconstruct the 3D geometry of the microscopic sample with high accuracy.

using sparse SIFT features and employing *a contrario* RANSAC approach, dense correspondence for vertical/horizontal disparity computation by use of dense SIFT features, disparity refinement by taking advantage of a fast approximation variant of the bilateral filtering and finally depth estimation. In Section 4.3, the results generated by the proposed framework are presented with detailed comparisons with the state-of-the-art sparse feature based approaches. Section 4.4 concludes the chapter.

4.2 Methods

The overview of the proposed method for high fidelity 3D reconstruction of microscopic samples using a pair of stereo SEM images captured by tilting the sample surface by a

known degree is shown in Figure 4.1. Here, stereo pair from a *Tapetal Cell* are used for demonstration. After imaging the microscopic samples, the process begins by sparse feature detection/description using SIFT [120]. Using the detected features and employing *a contrario* RANSAC approach, outliers are eliminated and the fundamental matrix is approximated. This is later used for rectifying the input pair [66, 138]. The rectification process will cause the displacements to be more concentrated along the horizontal direction. Using the SIFT-Flow framework [116], dense correspondence is found for individual pixels between the rectified images. Since the SIFT-Flow can be considered to be a labeling approach with a discrete set of labels, the outcome needs to be refined in order to have smooth transitions between adjacent labels. To achieve this and also to prevent over-smoothing in the regions with large gaps between disparities (e.g. variations in the depth), an approximate fast bilateral filtering technique is employed [144]. The smoothed horizontal disparity map is later used for depth approximation. In the following subsections each of the steps are elaborated in more detail.

4.2.1 SEM Imaging Protocol

In this work, a Hitachi S-4800 field emission scanning electron microscope (FE-SEM) has been utilized to generate the micrographs. This SEM is equipped with a computer controlled 5 axis motorized specimen stage which enables movements in x, y and z directions as well as tilt (-5 to 70°) and rotation (0 to 360°). Specimen manipulations, such as tilt, z-positioning and rotation of the specimen stage, as well as image pre-processing and capture functions were operated through the Hitachi PC-SEM software. The working distance which gives the required depth of focus was determined at the maximum tilt for every single sample at the magnification chosen for image capture. As the specimen was tilted in successive 1° increments until reaching the final value through the software application, the SEM image was centered by moving the stage in the x- and/or y-axes manually. The micrographs were acquired with an accelerating voltage of 3 or 5 kV, utilizing the signals from both the upper and lower SE detectors, as shown in Figure 4.2. The magnification and working distance were held fixed in each captured image of the tilt

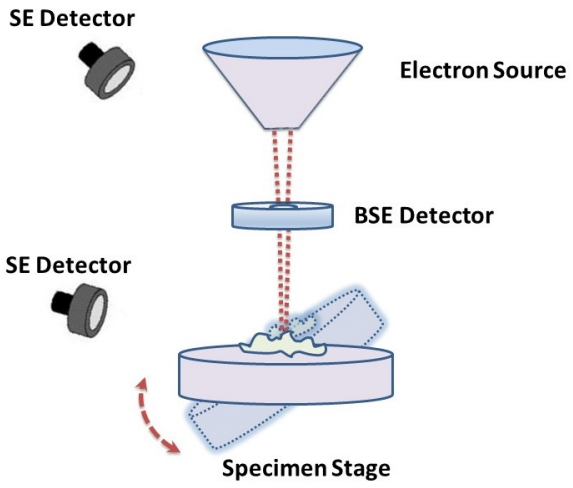
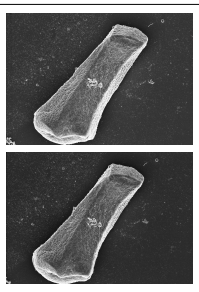
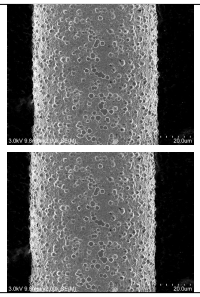
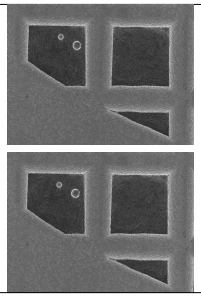
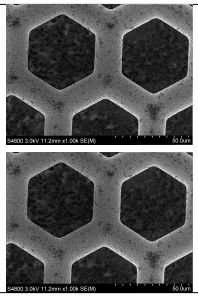
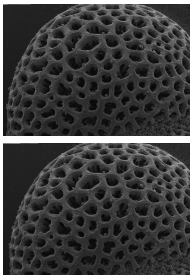


Figure 4.2: SEM imaging procedure used for this study.

Table 4.1: Summary of the dataset used in this work. The micrographs are acquired from *Tapetal Cell*, *Copper Bar*, *Copper Grid*, *Hexagonal Grid* and *Pollen Grain* using a Hitachi S-4800 field emission scanning electron microscope (FE-SEM) with sizes ranging from 512×384 to 1280×960 and tilt angles in the range $3 - 11°$.

	Tapetal Cell	*Copper Bar*	*Copper Grid*	*Hexagonal Grid*	*Pollen Grain*
Images					
Size	1280×960	512×384	1280×960	1280×960	854×640
Tilt Angle	9°	11°	7°	10°	3°

series. Contrast and brightness were adjusted manually to keep consistency between SEM micrographs. Table 4.1 summarizes the data that used in this work. Micrographs from *Tapetal Cell*, *Copper Bar*, *Copper Grid*, *Hexagonal Grid* and *Pollen Grain* are considered for evaluating the performance and accuracy of the proposed approach.

4.2.2 Scale Invariant Feature Transform (SIFT)

Four stages of feature detection/description involved in SIFT method can be summarized as [120]: 1) scale-space extrema detection, 2) keypoint localization, 3) orientation assignment and 4) keypoint descriptors. For the first step, a Gaussian function is considered

as the scale-space kernel based on the work of [114]. By finding the scale-space extrema in the response of the image to difference-of-Gaussian (DoG) masks, not only a good approximation for the scale-normalized Laplacian of Gaussian (LoG) function is provided, but also as pointed out in the work of [129], the detected features are more stable. The local extrema of the response of the image to the DoG masks of different scales is found in a $3 \times 3 \times 3$ neighborhood of the interest point. For accurate localization of the keypoints in the set of candidate keypoints, a three dimensional quadratic function is fitted to the local sample points. By applying a threshold on the value of this fitting function at the extremum, keypoints located in low contrast regions that are highly affected by noise are eliminated. Moreover, thresholding the ratio of principal curvatures can also eliminate poorly defined feature points near the edges. After finalizing the keypoints, orientations can be assigned. This is done by using the gradients computed in the first step of the algorithm when computing DoG responses. Creating a 36-bin histogram for orientations in the keypoint's neighborhood is the next step. Each neighbor contributes to the histogram by a weight computed based on its gradient magnitude and also by a Gaussian weighed circular window around the keypoint.

The final step is the local image descriptor generation. Using the location, scale and orientation determined for each keypoint up until now, the local descriptor is created in a manner which makes it invariant to differences in illumination and viewpoint. This is done by combining the gradients at keypoint locations, as computed in the previous steps, weighted by a Gaussian function over each 4×4 sub-region in a 16×16 neighborhood around the keypoint into 8-bin histograms. This results in a $4 \times 4 \times 8 = 128$ element vector for each keypoint. Normalizing the feature vectors to unit length will reduce the effect of linear illumination changes. This is usually followed by thresholding the normalized vector and re-normalizing it again to reduce the effects of large gradient magnitudes.

In the current work, SIFT is used in two different ways. For the step of sparse image matching required for epipolar rectification, the general SIFT approach is used for locating the feature points and computing the corresponding descriptors. However for the dense matching, feature detection is eliminated and SIFT descriptors are computed

for all the pixels contained in the input images. For more information regarding the detail and implementation of SIFT the reader is referred to [120].

4.2.3 Epipolar Rectification

Given a set of two SEM images of a microscopic sample captured from different viewpoints, the epipolar rectification step attempts to transform the images in such a way that we only have horizontal displacement (disparity) between the corresponding pixels within the images. Assuming a set of sparse naively matched corresponding points generated by SIFT followed by *a contrario* RANSAC (ORSA) outlier removal algorithm [137] and represented as 3-vectors of homogeneous coordinates for the left (X_l) and right (X_r) images, the epipolar constraint can be written as [80]:

$$X_l^T F X_r = 0 \tag{4.1}$$

where F is the fundamental matrix that captures the rigidity constraint of the scene. Having a rectified pair, the fundamental matrix takes the especial form of:

$$F = [e_1]_\times = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \tag{4.2}$$

which means that the epipoles are at infinity in horizontal direction. Therefore, the process of rectification involves finding homographies to be applied to the left and right images to satisfy the epipolar constraint equation when $F = [e_1]_\times$. This can be represented in a mathematical form as:

$$X_l^T F X_r = 0 \equiv (H_l X_l)^T [e_1]_\times (H_r X_r) = 0 \tag{4.3}$$

Having a rotation matrix R for the camera around the focus point, a homography

matrix can be formulated as follows:

$$H = KRK^{-1} \tag{4.4}$$

where K is the camera parameters matrix with (x_c, y_c) as the image center (principal point) and f as the unknown focal length:

$$K = \begin{bmatrix} f & 0 & x_c \\ 0 & f & y_c \\ 0 & 0 & 1 \end{bmatrix} \tag{4.5}$$

Following the formulation proposed in [66, 138] we look for rotation matrices R_l and R_r and focal length which satisfy:

$$E(x_l, y_l, x_r, y_r) = X_l^T K^{-T} R_l^T K^T [e_1]_\times K R_r K^{-1} X_r = 0 \tag{4.6}$$

where $R_r = R_z(\theta_{rz})R_y(\theta_{ry})R_x(\theta_{rx})$, $R_l = R_z(\theta_{lz})R_y(\theta_{ly})$ and $K = K(f = 3^g(w + h))$, with w and h as the width and height of the input images respectively and g in the range $[-1, 1]$. It should also be noted that due to the specific form of $[e_1]_\times$ all of the rotations around the x direction are eliminated since $R_x^t [e_1]_\times R_x = [e_1]_\times$. Assuming the Sampson error as:

$$E_s^2 = E^T (JJ^T)^{-1} E \tag{4.7}$$

where J is the matrix of partial derivatives of E with respect to the 4 variables:

$$J = ((FX_r)_1 \;\; (FX_r)_2 \;\; (F^T X_l)_1 \;\; (F^T X_l)_2) \tag{4.8}$$

we have

$$E_s(X_l, X_r)^2 = \frac{E(X_l, X_r)^2}{||[e_3]_\times F^T X_l||^2 + ||[e_3]_\times F X_r||^2} \tag{4.9}$$

Utilizing Levenberg-Marquardt [141], the method seeks the parameters $(\theta_{ly},\ \theta_{lz},\ \theta_{rx},\ \theta_{ry},\ \theta_{rz},\ g)$ which minimize the sum of Sampson errors over the matching pairs. The optimized pa-

rameters are then used for building the two homographies to be applied to the left and right view images. More elaboration regarding the theory and implementation aspects of the rectification method can be found in [66, 138].

4.2.4 SIFT-Flow for Dense Correspondence

Discontinuity preserving pixel/feature matching is a key component of many computer vision applications. This is unlike the many general purpose image registration approaches in which the computed displacement maps are assumed too be smooth [16, 18]. In such cases, even though the image grid is deformed during the process of registration, the underlying geometry is considered as a whole, without the possibility of folding or overlapping. However, in computer vision applications, the objects that are contained in the images are projective representations of the three dimensional objects in real world. Therefore, the assumption of having discontinuity is necessary as a representation of the difference in the relative distances of the objects to the camera. Here, the problem of matching between image pixels is modeled as a dual-layer factor graph, with decoupled components for horizontal/vertical flow to account for sliding motion. This model is based on the work of [116] which takes advantage of an L_1 truncated norm for achieving discontinuity preservation and higher speeds in matching. Assuming $\mathbf{F}_1$ and $\mathbf{F}_2$ as two dense multi-dimensional SIFT descriptor images, and $\mathbf{p} = (x, y)$ as the grid coordinates of the image, the objective function to be minimized can be written as follows:

$$\begin{aligned} E(\mathbf{w}) = \sum_{\mathbf{p}} min(||\mathbf{F}_1(\mathbf{p}) - \mathbf{F}_2(\mathbf{p} + \mathbf{w}(\mathbf{p}))||, t) + \\ \sum_{\mathbf{p}} \eta(|u(\mathbf{p})| + |v(\mathbf{p})|) + \\ \sum_{(\mathbf{p},\mathbf{q})\in\epsilon} min(\alpha|u(\mathbf{p}) - u(\mathbf{q})|, d) + min(\alpha|v(\mathbf{p}) - v(\mathbf{q})|, d) \end{aligned} \tag{4.10}$$

in which $\mathbf{w(p)} = (u(\mathbf{p}), v(\mathbf{p}))$ is the flow vector at point $\mathbf{p}$. The three summations in this equation are *data*, *small displacement* and *smoothness* terms, respectively. The data term is for minimizing the difference between the feature descriptors along the flow vector,

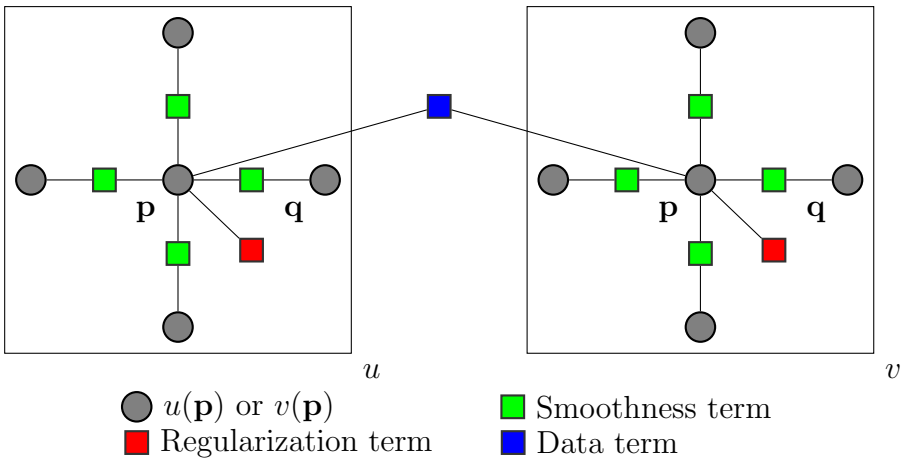


Figure 4.3: Factor graph representation of the energy minimization functional with decoupled horizontal and vertical components

while the small displacement term keeps the displacements as small as possible when no information is available. Finally the smoothness term guaranties that the flow vectors for neighbor pixels are similar. A few parameters that are used in this formulation are: t and d as data and smoothness thresholds and α and η as small displacement and smoothness coefficients, respectively. The values are set to the default values proposed by Liu et al. [116].

As is obvious, in this formulation the horizontal and vertical components are decoupled. This is mainly for reducing the computational complexity. But this gives additional benefit of being able to account for sliding motions during the process of image matching. The objective function is formulated as a factor graph, with ($\mathbf{p}$) and ($\mathbf{q}$) as the variable nodes while the factor nodes represent the data, small displacement and smoothness terms. The flow is then extracted by using a dual-layer loopy belief propagation algorithm. Figure 4.3 shows the factor graph suggested by [116] for optimizing the energy functional of dense matching problem. By using a coarse-to-fine (multi-resolution) matching scheme, one is able to reduce the computational complexity and hence the computation time while achieving lower values for the energy functional.

Belief propagation (BP) is a technique for *exact* inference of marginal probabilities for singly connected distributions [21]. Generally speaking, each node in the graph computes a *belief* based on the messages that it receives from its children and also from its parents. Such a technique is purely local, which means that the updates are unaware of the global

structure of the graph as the graph may contain loops and therefore be multiply connected [148]. In this case, BP cannot compute an exact solution, but at best an approximation which can be surprisingly very accurate [21]. Use of graphical models in image processing tasks usually fall within the category of loopy graphs, which means different variants of BP are used and studied for solving different problems in this area [184, 10].

In the general formulation of BP and subsequently loopy BP (LBP) [139], we assume that node X computes its belief $b(x) = P(X = x|E)$, where E is the observed evidence that is computed by combining the messages from the node's children $\lambda_{Y_j}(x)$ and also its parents $\pi_X(u_k)$. Assuming $\lambda_X(x)$ as the nodes' message to itself representing the evidence, we have:

$$b(x) = \alpha\lambda(x)\pi(x) \tag{4.11}$$

where:

$$\lambda^{(t)}(x) = \lambda_X(x)\prod_j \lambda^{(t)}_{Y_j}(x) \tag{4.12}$$

and:

$$\pi^{(t)}(x) = \sum_u P(X = x|U = u)\prod_k \pi^{(t)}_X(u_k) \tag{4.13}$$

The message that X passes to its parent U_i is given by:

$$\lambda_x(t+1)(u_i) = \alpha\sum_x \lambda^{(t)}(x)\sum_{u_k:k\neq i} P(x|u)\prod_{k\neq i}\pi_X(t)(u_k) \tag{4.14}$$

and the message that X sends to its child Y_j is given by:

$$\pi^{(t+1)}_{Y_j}(x) = \alpha\pi^{(t)}(x)\lambda_X(x)\prod_{k\neq j}\lambda^{(t)}_{Y_k}(x) \tag{4.15}$$

As can be seen, if a message is being generated to pass from node A to B, the contribution of the message from node B to A from the previous iteration is eliminated. Also, normalizing messages at each iteration doesn't have any effect on the final beliefs and has the benefit of preventing numerical underflow [148].

Factor graphs are a means of unifying the directed and undirected graphical models

with the same representation [102]. Such graphs are derived by the main assumption of representing complicated global functions of many variables by factorizing them as a product of several *local* functions of subsets of variables. Generally speaking, a factor graph can be defined as $\mathbb{F} = (\mathbb{G}, \mathbb{P})$ in which $\mathbb{G}$ is the structure of the graph and $\mathbb{P}$ is the parameter of the graph. $\mathbb{G}$ being a *bipartite graph* can be defined as $\mathbb{G} = (\{X, F\}, E)$ where, X and F are variable nodes and factor nodes, respectively, while E is a set of edges connecting a factor f_i and a variable $x \in X_j$. Given evidence as a set of variables with observed values, the process of belief propagation consists of passing local messages between nodes in order to compute the marginal of all nodes. Even though the same concept is used for belief propagation in directed graphs, here, the process can be formulated as passing messages between variable and factor nodes. In this case, two types of messages are passed: 1) message from variable node to factor node ($\mu_{x \to f}$) and 2) message from factor node to variable node ($\mu_{f \to x}$):

$$\mu_{x \to f}(x) \propto \prod_{h \in N_x \setminus \{f\}} \mu_{h \to x}(x) \tag{4.16}$$

$$\mu_{f \to x}(x) \propto \sum_{N_f \setminus \{x\}} \Big(f(X_f) \prod_{y \in N_f \setminus \{x\}} \mu_{y \to f}(y) \Big) \tag{4.17}$$

where x and y are variables, f and h are factors and N_f and N_x are representative of neighbors of the corresponding nodes in the graph. In acyclic graphs, the process of message passing is terminated after two messages are passed on every edge, one in each direction. In such graphs, the process results in an exact inference. Unlike acyclic graphs, belief propagation is done in an iterative manner in cyclic graphs. The process is terminated when having minimal changes in the passed messages according to a predetermined threshold and the result is considered an approximate inference.

Several modifications to the general formulation of the energy minimization procedure is proposed by Liu et al. [116] which are also considered in this work. Different from the general formulation of optical flow, here, the smoothness term is decoupled for allowing separate horizontal and vertical flows. This reduces the computational complexity of the

energy minimization significantly. In this implementation, at first, the *intra*-layer messages are updated for horizontal and vertical flows and then the *inter*-layer messages are updated between horizontal and vertical flows. Moreover, sequential belief propagation (BP-S) [184] is used for better convergence.

4.2.5 Disparity Refinement: Bilateral Filtering

Since the result of the previous step is in general a discrete labeling of the horizontal/vertical disparity maps, neighbor pixels may have different labels. These differences represent themselves as sudden jumps in the final 3D reconstruction results. One should note that this is not always problematic especially in regions where sharp variations of the disparity levels are representatives of edges and different depths. However, in uniform regions these small variations in disparity values should be smoothed for a more visually pleasing reconstruction result.

Bilateral filtering has been shown to provide high capability in noise/variation reduction while preserving edges contained in the images [145]. The general idea is similar to simple Gaussian filtering. However, unlike the Gaussian filtering which only takes the spatial proximity into consideration, bilateral filtering takes both spatial and range information into account. Assuming the noisy image I, the general formulation for the bilateral filtered result $\hat{I}$ at pixel location $\mathbf{p}$ is:

$$\hat{I}_{\mathbf{p}} = \frac{1}{W_{\mathbf{p}}} \sum_{\mathbf{q} \in \mathcal{S}} G_{\sigma_s}(||\mathbf{p} - \mathbf{q}||)\ G_{\sigma_r}(|I_{\mathbf{p}} - I_{\mathbf{q}}|) I_{\mathbf{q}} \tag{4.18}$$

with the normalization factor defined as $W_{\mathbf{p}} = \sum_{\mathbf{p} \in \mathcal{S}} G_{\sigma_s}(||\mathbf{p} - \mathbf{q}||)\ G_{\sigma_r}(|I_{\mathbf{p}} - I_{\mathbf{q}}|)$, $\mathbf{q}$ is the neighbor pixel in the neighborhood $\mathcal{S}$ and G_{σ_s} and G_{σ_r} are the Gaussian weighting functions for spatial and range data, respectively. Direct implementation of the bilateral filter is computationally expensive and therefore, several approximation techniques have been proposed in the literature for speeding up the process [197, 152, 50, 144]. Here, the approximation method proposed by Paris and Durand [144] is used. In their formulation, the image is first converted to a volumetric bilateral grid with homogeneous values. It

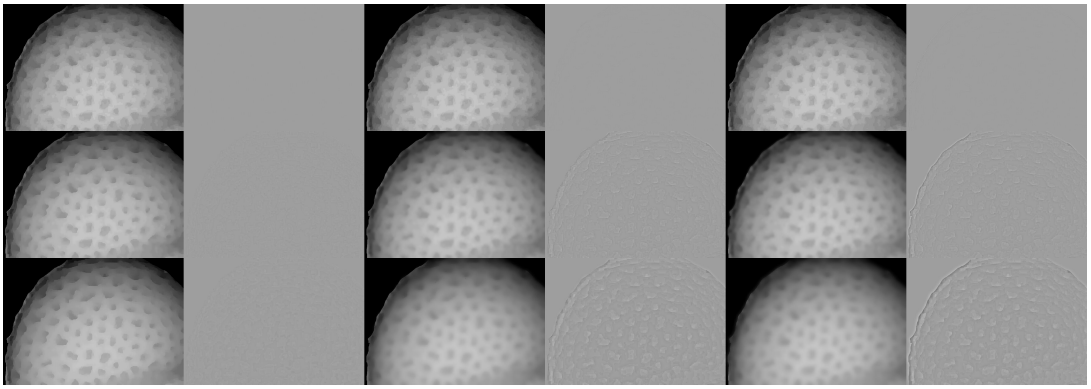


Figure 4.4: Effects of various spatial ($\sigma_s = \{1, 3, 5\}$, from left to right) and range ($\sigma_r = \{1, 3, 5\}$, from top to bottom) variance for bilateral filtering of the disparity of *Pollen Grain*. The difference map between the initial disparity map and the refined map are also presented.

is shown that the bilateral filter can be approximate by Gaussian convolution applied to the grid followed by sampling and normalization of the homogeneous values. The spatial (σ_s) and range (σ_r) variation parameters are chosen experimentally. Figure 4.4 displays the effects of different values for the parameters in the smoothness of the computed horizontal disparity map for the *Pollen Grain*. We aim to smooth the minimal variations in uniform regions while preserving sudden jumps in disparity values associated with bigger differences in the depth. Looking closely at the various spatial and range variances shown in Figure 4.4, it is obvious that bigger values of variances cause the disparity map to be over-smoothed. This can be seen in the corresponding difference maps as more edge details can be seen which indicate that more edge information are smoothed out. To eliminate this and still having a reasonable smoothing effect, the values of σ_s and σ_r are both set to 3, experimentally for creating the results presented here.

4.2.6 Depth Estimation

As mentioned before, stereo rectification transforms the images in a manner in which the displacements will be grossly concentrated in the horizontal direction. This greatly simplifies the process of depth estimation. This is especially useful for the case of 3D reconstruction of SEM images since the tilt angles are very small with high amount of overlap between stereo image pairs. For more general problems like large scale multiple

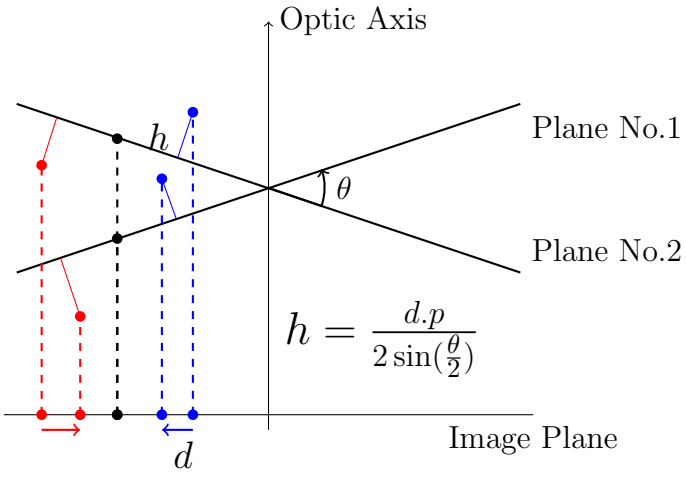


Figure 4.5: Relationship between the estimated height (h) and the computed horizontal disparity (d) using the pixel size in sample units (p) and the total tilt angle (θ).

view stereo (MVS), the proposed technique is not directly applicable and more sophisticated methods are needed [176, 190, 65].

The horizontal disparity computed from the previous step, can be utilized for estimating the depth of the individual pixels contained in the images. This requires that several parameters to be known: tilt angle, magnification and size of each pixel in sample units. Figure 4.5 shows the relationship between the computed horizontal disparity and the height for a few sample points. This can be represented using a simple trigonometric equation [163, 183, 203]:

$$h = \frac{d.p}{2\sin(\frac{\theta}{2})} \tag{4.19}$$

which uses the computed horizontal disparity d, pixel size in sample units (p) and the total tilt angle (θ) to estimate the height (h).

4.3 Results & Discussions

Assessing the performance of proposed framework which consists of several steps is done using the several sets of SEM images (*Tapetal Cell*, *Copper Bar*, *Copper Grid*, *Hexagonal Grid*, *Pollen Grain*) captured by a Hitachi S-4800 field emission scanning electron microscope (FE-SEM), equipped with a computer controlled 5 axis motorized specimen stage which enables movements in x, y and z directions as well as tilt (-5 to 70°) and

rotation (0 to 360°) [189]. As the specimen was tilted in successive 1° increments, the SEM image was centered by moving the stage in the x- and/or y-axes manually. One should note that this does not have any effect on the relative disparity and subsequent estimated depth variations between pixels of the images and may only cause the average disparity to be elevated or lowered. The micrographs were acquired with an accelerating voltage of 3 or 5 kV, utilizing the signals from both the upper and lower SE detectors, as shown in Figure 4.2.

As the first step, sparse SIFT features/descriptors are located following the approach outlined in Section 4.2.2. This step is straightforward and no optimization is taken in order to increase/decrease the number of features. This step is followed by sparse feature matching implemented by employing *a contrario* RANSAC to ensure better outlier removal. Epipolar rectification for finding the appropriate homography transforms for the input micrographs in order to have more horizontally concentrated disparity maps is next. Table 5.2 summarizes the results of sparse SIFT matching and the subsequent epipolar rectification for all of the micrograph sets. The first and second row in the table indicate the number of SIFT features found in the first and second micrographs of each set. As can be seen, the number of detected features is minuscule in comparison to the total number of pixels contained within the images. This number is further reduced after finding the corresponding matches (see third row in Table 5.2). However, it should be noted that these features are not used for 3D reconstruction of the microscopic samples and while having small number of SIFT features can be problematic in the case of sparse feature based reconstruction, here, it does not have a negative impact. In fact, having only eight true matches is enough for estimating the fundamental matrix which captures the rigidity constraint of the scene [81, 136, 66]. The computed homography transforms for the first and second micrographs are displayed in the table as well. This is followed by initial and final Sampson rectification errors. As expected, since the SEM micrographs are captured in a very controlled manner, rectification errors are not very large to begin with. However, epipolar rectification is recommended to ensure minimal operator introduced errors as a result of manual manipulation of the specimen stage. This will guaranty truthful three

Table 4.2: Summary of rectification results using sparse SIFT matching and the subsequent epipolar rectification. The first and second row in the table indicate the number of SIFT features found in the first and second micrographs of each set, while the third row is the result of *a contrario* methodology for matching the SIFT features according to a homography transform. Fourth and fifth row show the computed homography transformation matrices as results of epipolar rectification by minimizing the Sampson's error. Finally, initial and final Sampson rectification errors are presented.

	Tapetal Cell	*Copper Bar*	*Copper Grid*	*Hexagonal Grid*	*Pollen Grain*
im.1 # SIFT keypoints	438	347	183	558	471
im.2 # SIFT keypoints	391	335	177	558	476
ORSA # SIFT matches	**67**	**201**	**32**	**16**	**274**
H_1	$\begin{bmatrix} 0.9739 & 0.0239 & 10.37 \\ -0.0434 & 0.9992 & 55.24 \\ \approx 0 & \approx 0 & 1.026 \end{bmatrix}$	$\begin{bmatrix} 0.9105 & -0.5868 & 127.1 \\ 0.5727 & 0.8205 & -118.4 \\ 0.0001 & -0.0001 & 0.9456 \end{bmatrix}$	$\begin{bmatrix} 0.9991 & -0.00312 & 3.772 \\ 0.0023 & 0.9999 & -3.221 \\ \approx 0 & \approx 0 & 1.000 \end{bmatrix}$	$\begin{bmatrix} 1.003 & -0.0008 & -4.120 \\ 0.0037 & 0.9999 & -4.828 \\ \approx 0 & \approx 0 & 0.9961 \end{bmatrix}$	$\begin{bmatrix} 0.9999 & 0.0027 & -4.803 \\ -0.0094 & 0.9999 & 1.228 \\ \approx 0 & \approx 0 & 1.000 \end{bmatrix}$
H_2	$\begin{bmatrix} 0.9733 & 0.0245 & 10.53 \\ -0.0430 & 1.000 & 48.23 \\ \approx 0 & \approx 0 & 1.025 \end{bmatrix}$	$\begin{bmatrix} 0.9113 & -0.5873 & 126.9 \\ 0.5730 & 0.8204 & -118.6 \\ 0.0001 & -0.0001 & 0.9450 \end{bmatrix}$	$\begin{bmatrix} 0.9987 & -0.0037 & 3.751 \\ 0.0019 & 0.9999 & -2.164 \\ \approx 0 & \approx 0 & 1.001 \end{bmatrix}$	$\begin{bmatrix} 1.002 & -0.0002 & -3.614 \\ 0.002 & 0.9999 & -3.024 \\ \approx 0 & \approx 0 & 0.9970 \end{bmatrix}$	$\begin{bmatrix} 1.000 & 0.0051 & -6.254 \\ -0.01223 & 0.9999 & 2.037 \\ \approx 0 & \approx 0 & 0.9996 \end{bmatrix}$
initial rect. err. (pix)	1.895	1.241	0.9568	0.2826	0.5757
final rect. err. (pix)	**0.3703**	**0.1702**	**0.7198**	**0.1912**	**0.3244**

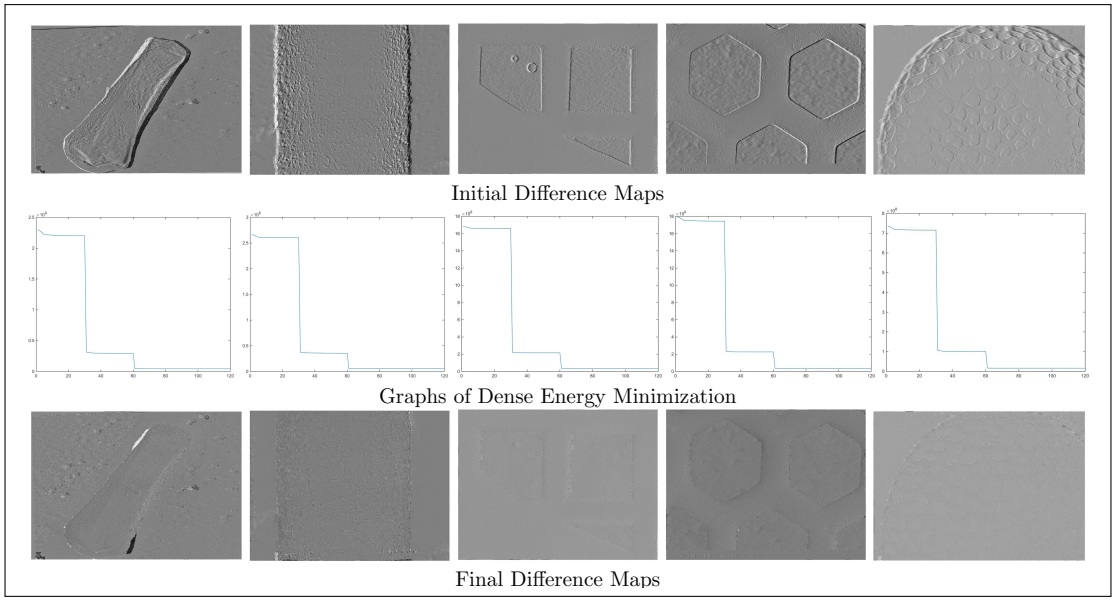


Figure 4.6: Dense matching results for the rectified image sets: *Tapetal Cell* (column 1), *Copper Bar* (column 2), *Copper Grid* (column 3), *Hexagonal Grid* (column 4) and *Pollen Grain* (column 5). The first row shows the initial difference map. The second row shows the minimization trend for the optimization process defined using dense SIFT descriptors, factor graph representation of the energy functional to be optimized (Figure 4.3) and loopy belief propagation as means of optimization. The third row displays the difference maps after the optimization process.

dimensional reconstruction of the underlying microscopic sample.

As for the performance of dense correspondence step, Figure 4.6 summarizes the results for the five used micrograph sets, visually. First row shows the difference maps between the two rectified input images before the process of dense matching, while the third row displays the difference maps after the process using dense SIFT descriptors and loopy belief propagation for minimizing the defined energy functional. The minimization process is implemented in a multi-resolution manner in three distinct stages, as can be seen from the second row of Figure 4.6. In such case, the aim is to recover the larger displacements in a coarser grid while compensating the smaller displacements in a finer grid. The multi-resolution implementation not only can reduce the computational time and complexity significantly. But also, it makes the recovery of true correspondence, in case of having bigger disparities, more achievable. The graphs are representatives of the optimization trend of the dense energy functional.

Table 4.3 summarizes the results of dense correspondence numerically. Here, the root

Table 4.3: Summary of dense correspondence results using dense SIFT features, the factor graph representation of the objective function and loopy belief propagation as meas for optimization. The first and second row represent the initial and final root mean squared error (RMSE) of the two input micrographs. The residual errors can be attributed to the noise contained in the micrographs as well as the differences in brightness due to edge effects caused by imaging in the secondary electron (SE) mode. third and fourth rows show the initial and final values of the objective function (note the coefficients $\times 10^9$ and $\times 10^7$). The fifth row shows the ratio between the energy contained in the vertical disparity map and the energy contained in the horizontal disparity map. This provides additional proof for the efficiency of the rectification process as well as the depth estimation step. The last row displays the computational time needed for finding the dense correspondence between input micrographs.

	Tapetal Cell	*Copper Bar*	*Copper Grid*	*Hexagonal Grid*	*Pollen Grain*
$RMSE_{initial}$	44.67	43.38	19.42	31.42	23.20
$RMSE_{final}$	**25.26**	**17.50**	**8.50**	**12.18**	**7.34**
$E_{initial}$ ($\times 10^9$)	2.29	0.20	1.68	1.79	0.73
E_{final} ($\times 10^7$)	**4.20**	**0.44**	**3.28**	**3.60**	**1.59**
$\frac{\sum \mathbf{v}^2}{\sum \mathbf{u}^2}$ (%)	**1.29**	**0.20**	**0.08**	**0.16**	**0.58**
$\approx$ Elapsed time (s)	44.65	7.50	47.10	47.33	19.60

mean squared error (RMSE) is used as means for assessing the performance of dense matching. The first row in the table is a representative of the initial RMSE while the second row shows the final RMSE. The error is reduced significantly. The residual error can be attributed to the noise contained in the micrographs as well as the differences in brightness in regions due to changing the tilt angel between each image acquisition. Using SEM in the secondary electron (SE) imaging mode, the contrast is mainly dominated by edge effects. This is due to having more secondary electrons that can leave the sample near edges which results in increased brightness. A close inspection of the difference maps provided in the third row of Figure 4.6 reveals that while the difference in non-edge regions is minimal, an increase in the difference can be seen near edges. Fortunately, SIFT is designed in such a way to be able to handle these subtle intensity variations. Therefore this does not have any impacts on the outcome of dense matching process. Initial and optimized values of the energy functional in Equation (4.10) can be seen in the third and fourth rows of the table. Using the factor graph representation of the objective function and loopy belief propagation as means for optimization, the energy is minimized by orders of magnitude. One of the assumptions in simplifying the process of

depth estimation and 3D point cloud generation was based on the observation that energy of vertical disparity map is minuscule in comparison to the energy of horizontal disparity map. This was expected as a result of the sparse feature-based rectification step. More evidence is presented in the fifth row of the table which reveals that the amount of energy contained in the vertical disparity map is in fact very small in comparison to the energy of horizontal disparity map. It should be noted that in case of having larger displacements between corresponding pixels of the initial micrographs, the ratio may increase due to the local nature of belief propagation minimization approach. This is the case of *Tapetal Cell* set where the ratio is larger compared to the rest. Of course, this is still very small to have a major negative impact of the outcome of depth estimation step. Finally, the last row, shows the computation time needed for dense matching between the micrographs in each set. The codes implemented here were a combination of MATLAB and MEX codes executed on a Core i7 CPU @ 3.50 GHz with 12 GB or RAM using MS Windows 7 and MATLAB R2014b. As can be seen, the size of the input micrographs dominate the overall computational need of the proposed dense matching approach. The step is followed by disparity refinement using the approximate bilateral filer discussed in Section 4.2.5.

Having the refined relative disparities and the tilt angle, depth can be estimated using Equation (4.19) and the three dimensional point cloud can be generated. Figures 4.7, 4.8, 4.9 and 4.10 show the results of the proposed method for the *Copper Bar*, *Copper Grid*, *Hexagonal Grid* and *Pollen Grain* image sets, respectively. In each figure, the first row shows several views of the generated dense point cloud (with sub-sampling for better visualization) for each pair of input images. The second row represents several views of the generated high quality surface mesh while the third row shows a magnified view of the generated surface mesh. Using the proposed approach, a high fidelity reconstruction of the microscopic samples is possible. Additionally, a uniform surface mesh can be generated since the distribution of the three dimensional points is uniform within the domain. This is one of the major advantages of using dense correspondence for 3D surface reconstruction in comparison to sparse feature based reconstruction approaches.

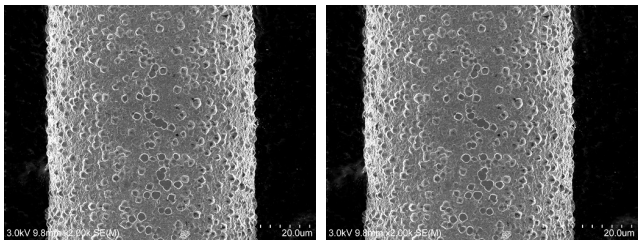
Moreover, higher amount of detail can be reconstructed employing the proposed sparse-dense methodology.

Sparse feature-based techniques rely only on the *features* detected in the images for the purpose of reconstruction [186, 187]. These features are not distributed uniformly within the image domain by default. This is especially more challenging in the regions that lack significant variations in intensity/depth and therefore, are rather flat and uniform. It should also be noted that feature detection techniques, like the ones employed in SIFT, are designed to ignore spatially close or features generated by edges. For the general problem of image matching, this is extremely useful since it helps avoiding computational redundancy and possible effects of noise. However, for surface reconstruction of SEM images, this will lead to erroneous results. For a better representation of the arguments, Figure 4.11 provides a visual comparison between the reconstruction results using the proposed approach and the state-of-the-art sparse feature based approaches presented in [188, 187]. The results are generated for the *Tapetal Cell* (column 1), *Copper Bar* (column 2), *Copper Grid* (column 3), *Hexagonal Grid* (column 4) and *Pollen Grain* (column 5), respectively. As can be seen, sparse feature based approaches suffer from non-uniform surface meshes generated while sharp edges and small features are not truthfully recovered. It should be noted that for the proposed approach, only two micrographs are used while for the sparse feature based reconstruction, five micrographs from different viewpoints are utilized. This is mainly to ensure sufficient matching points between image pairs for being able to build a more truthful reconstruction. Therefore, even though these approaches are proven to be useful in general purpose scene reconstruction applications, since they are not designed for feature detection of SEM images, they achieve limited performance. A closer inspection can provide a more elaborate comparison between the results generated using the two approaches. For the *Tapetal Cell*, as previously shown in Figure 4.1 (e, f), a small dent can be observed in the middle of the cell structure. While the two approaches provide a rather similar geometry for the cell, only the proposed sparse-dense approach can reconstruct the very fine detail. The adverse effect of non-uniform distribution of feature points can be seen in the results produced using the *Copper*

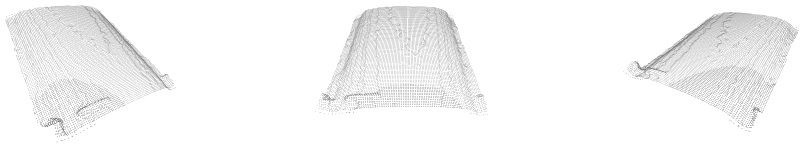
Bar set. Not only did this result in a non-uniform triangulation of the surface, but it also caused the bar's edges to be overly smoothed and deviate from a straight line. Sparsity of the feature points may cause the subsequent mesh general to be prone to over smoothing too. This is the case for the portion of *Copper Grid* shown in the third column. Even though this is the consequence of meshing software, however, it is a direct outcome of the three dimensional point cloud generation using sparse features. On the other hand, in the proposed sparse-dense framework, the edges are reconstructed with high accuracy. Moreover, minimal depth variations in the underlying carbon layer underneath the grid is revealed. For the case of *Hexagonal Grid*, not only the produced mesh is not uniform, but also identical areas are not reconstructed properly. This is obvious in the case of *Pollen Grain* due to higher levels of detail presented in the micrograph set. The sparse feature based approach provides a rather smooth surface, with minimal representation of the porous structure while the proposed approach is able to generate a detail-rich three dimensional surface. Overall, the performance of the proposed approach is superior to the state-of-the-art feature based reconstruction.

4.4 Conclusion

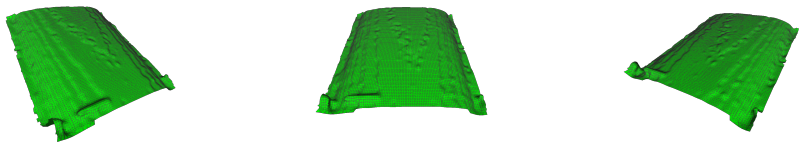
In this chapter, an end-to-end framework for high fidelity 3D reconstruction of microscopic samples from stereo SEM micrographs is proposed. Using a Hitachi S-4800 field emission scanning electron microscope (FE-SEM) which is equipped with a computer controlled 5 axis motorized specimen stage which enables movements in x, y and z directions as well as tilt and rotation, the specimen was tilted in successive 1° increments until reaching the final desired tilt angle with manual movement of the stage in the x- and/or y directions. Even with the most careful acquisition procedure, the acquired images need to be transformed in manner to ensure more accurate 3D reconstruction. In this step, using sparse SIFT features/descriptors and employing *a contrario* RANSAC, matched features are found and outliers that do not satisfy a projective transform are removed. This followed by an stage of rectification for transforming the images to having



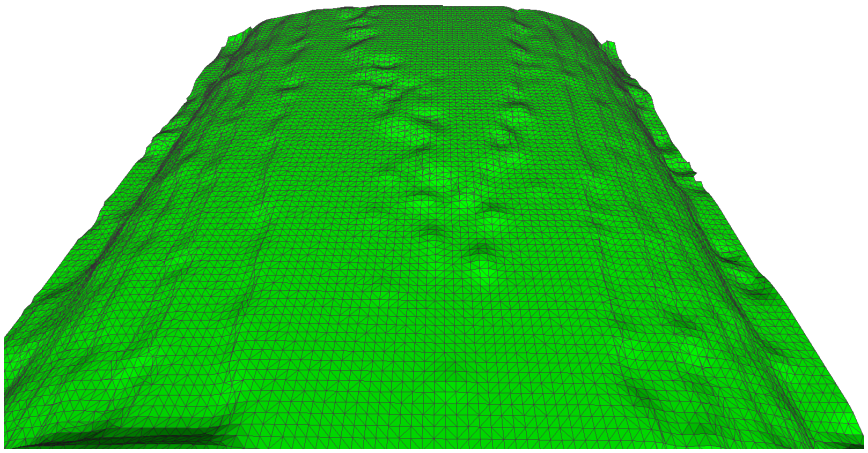
SEM Two-View Images



Dense Point Cloud

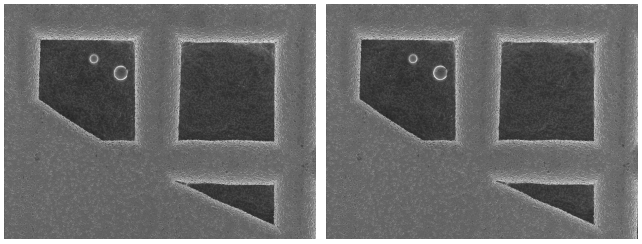


Triangular Surface Mesh

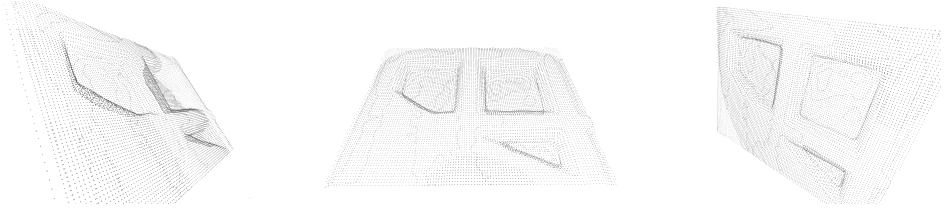


Magnified View

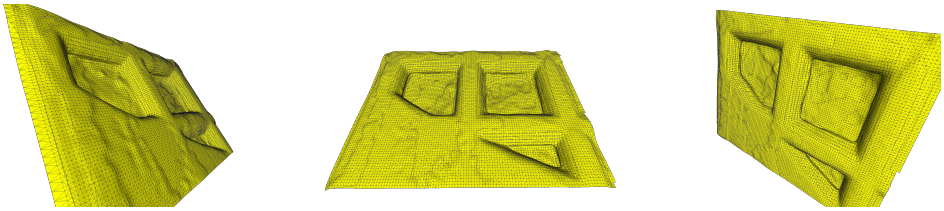
Figure 4.7: Qualitative visualization of the proposed 3D SEM reconstruction framework for the *Copper Bar* sample images, acquired by tilting the sample stage by 11 degrees. The set of two-view images can be seen in Table 5.1. Second row displays several views of the reconstructed dense point cloud. The initial cloud contains 196608 points which is sub-sampled here for better visualization. Third row shows the constructed triangular surface mesh. Fourth row depicts a magnified view of the constructed triangular surface mesh. Considering the proposed dense correspondence framework, a highly uniform reconstruction of the curved surface is achieved.



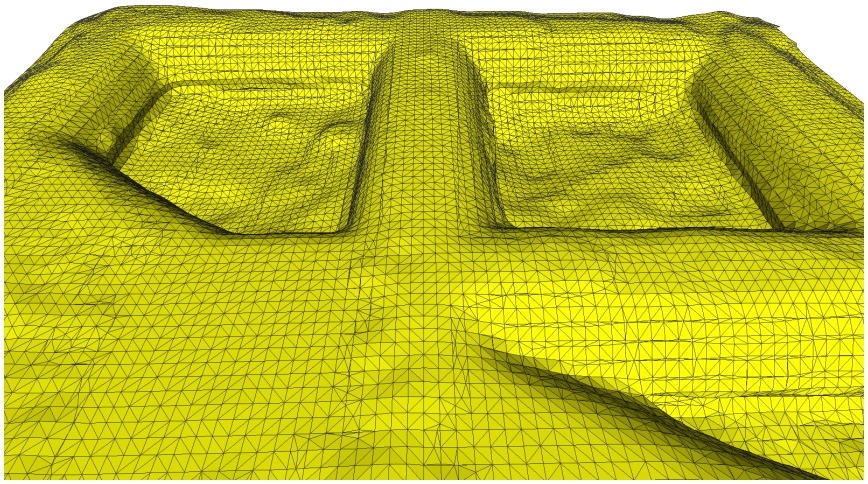
SEM Two-View Images



Dense Point Cloud

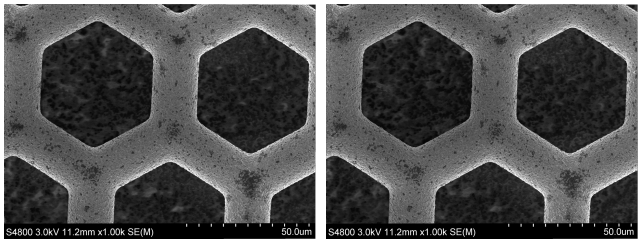


Triangular Surface Mesh

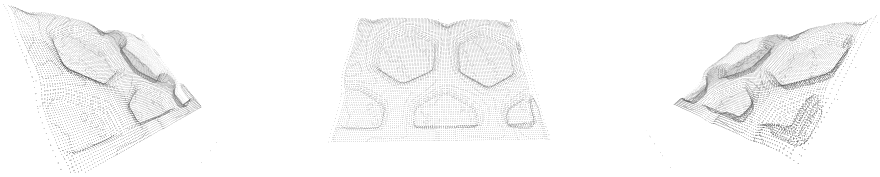


Magnified View

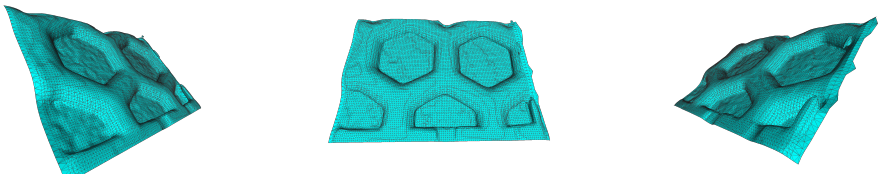
Figure 4.8: Qualitative visualization of the proposed 3D SEM reconstruction framework for the *Copper Grid* sample images, acquired by tilting the sample stage by 7 degrees. The set of two-view images can be seen in Table 5.1. Second row displays several views of the reconstructed dense point cloud. The initial cloud contains 1228800 points which is sub-sampled here for better visualization. Third row shows the constructed triangular surface mesh. Fourth row depicts a magnified view of the constructed triangular surface mesh. Considering the proposed dense correspondence framework, a highly uniform reconstruction of the curved surface is achieved.



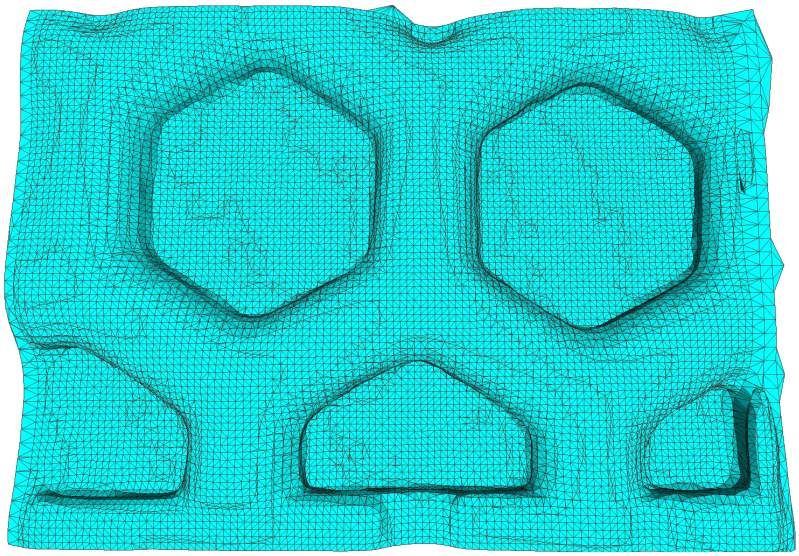
SEM Two-View Images



Dense Point Cloud

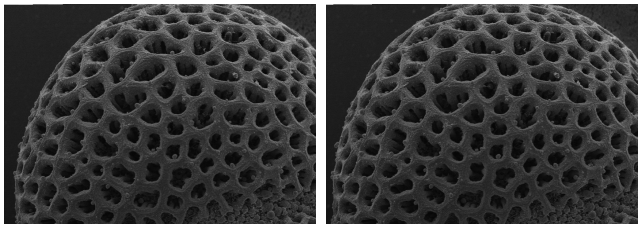


Triangular Surface Mesh

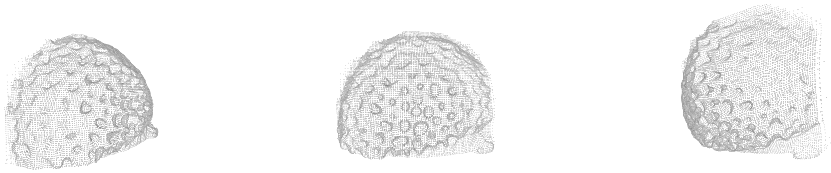


Magnified View

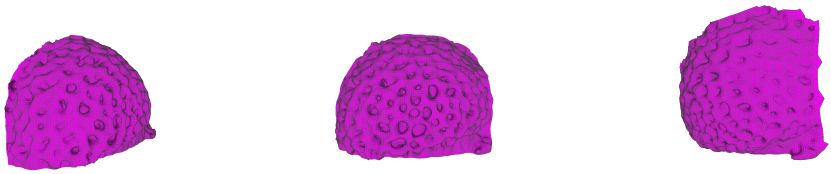
Figure 4.9: Qualitative visualization of the proposed 3D SEM reconstruction framework for the *Hexagonal Grid* sample images, acquired by tilting the sample stage by 10 degrees. The set of two-view images can be seen in Table 5.1. Second row displays several views of the reconstructed dense point cloud. The initial cloud contains 1228800 points which is sub-sampled here for better visualization. Third row shows the constructed triangular surface mesh. Fourth row depicts a magnified view of the constructed triangular surface mesh. Considering the proposed dense correspondence framework, a highly uniform reconstruction of the curved surface is achieved.



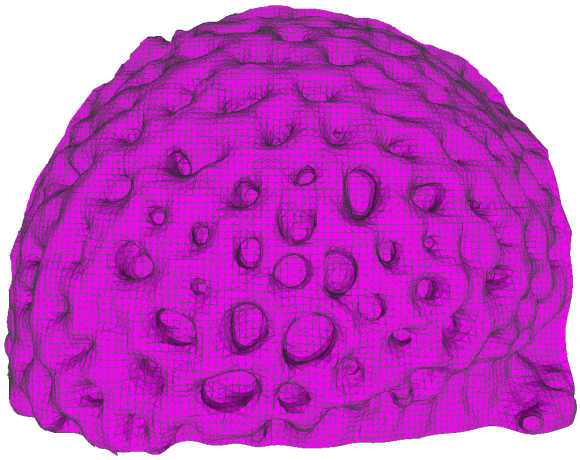
SEM Two-View Images



Dense Point Cloud



Triangular Surface Mesh



Magnified View

Figure 4.10: Qualitative visualization of the proposed 3D SEM reconstruction framework for the *Pollen Grain* sample images, acquired by tilting the sample stage by 3 degrees. The set of two-view images can be seen in Table 5.1. Second row displays several views of the reconstructed dense point cloud. The initial cloud contains 447665 points which is sub-sampled here for better visualization. Third row shows the constructed triangular surface mesh. Fourth row depicts a magnified view of the constructed triangular surface mesh. Considering the proposed dense correspondence framework, a highly uniform reconstruction of the curved surface is achieved.

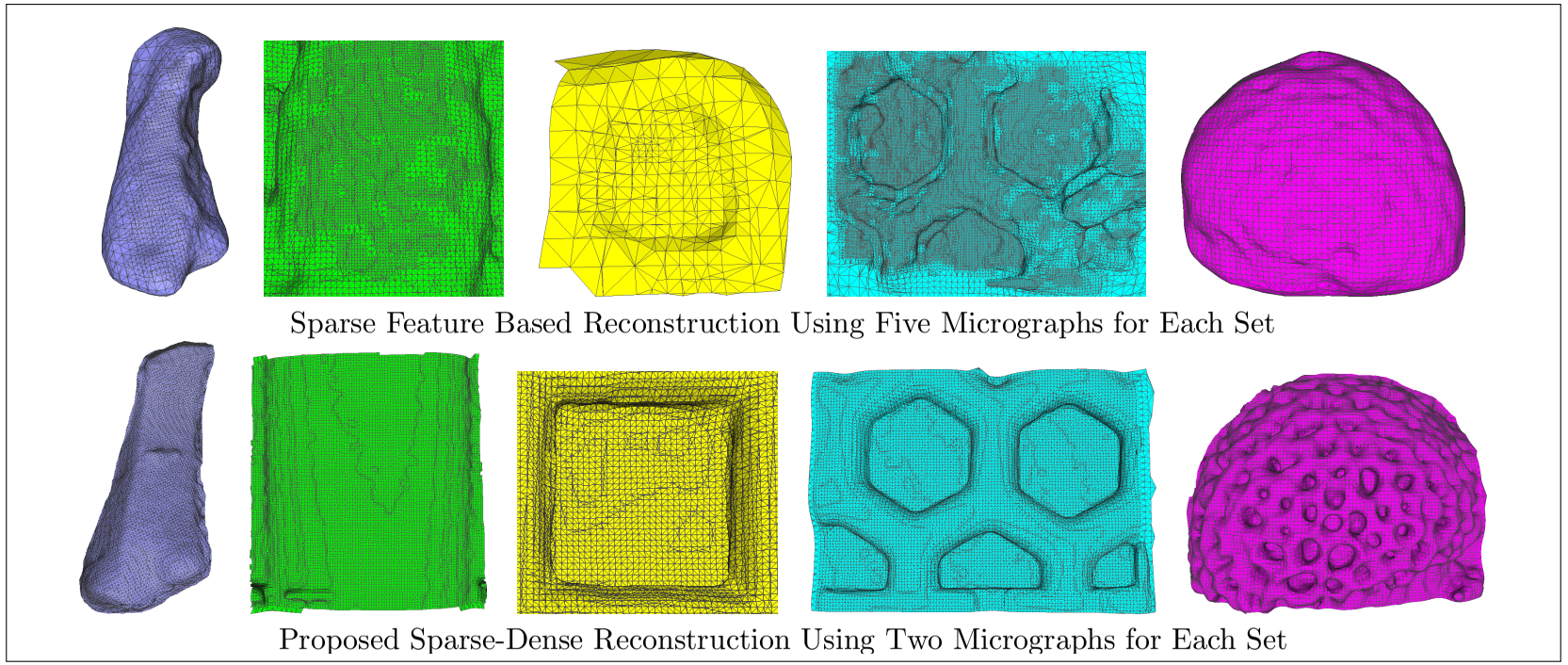


Figure 4.11: Qualitative comparison between state-of-the-art sparse feature based reconstruction [188, 187] and the proposed sparse-dense reconstruction approach for *Tapetal Cell* (column 1), *Copper Bar* (column 2), a portion of *Copper Grid* (column 3), *Hexagonal Grid* (column 4) and *Pollen Grain* (column 5) micrograph sets. Close inspection of the results reveals superiority of the proposed sparse-dense reconstruction methodology: uniformly distributed mesh nodes, ability to truthfully recover edges and very fine details, consistency over the domain of images.

a more horizontally-concentrated disparity. In this manner, given the correct disparity, the process of depth estimation will be simplified greatly since the depth will be directly proportional to the found disparity. For the next step, we take advantage of a constrained optimization procedure using dense SIFT descriptors, factor graph representation of the energy functional to be optimized and loopy belief propagation as means of optimization. Finally, depth is estimated using the bilaterally-filtered horizontal disparity computed from the previous step. Extensive tests and experiments with several sets of SEM micrographs prove the robustness and reliability of the proposed method for high resolution quality 3D reconstruction of microscopic samples. In the next chapter the problem of 3D surface reconstruction for highly complex microscopic samples is discussed and the current framework is improved.

Chapter 5

3D Reconstruction of Highly Complex Microscopic Samples Using Non-Local Optical Flow Estimation

5.1 Introduction

Scanning electron microscope (SEM) is one the principal research and industrial equipment for imaging on the microscopic scale. SEM and its diverse applications have been a very active research over the recent decade, and scientific studies well covered the use of SEM in broad domains ranging from biomedical applications to materials sciences and nano technologies [196, 218, 27, 6, 180, 195, 96]. SEM as an advanced microscopy device produces high quality images of microscopic specimen using a focused beam of electrons which can be then captured by two types of detectors, secondary electron (SE) and back-scattered electron (BSE) detectors, to provide both compositional and/or geometrical information [51]. However, SEM micrographs remain 2D while the need for having a more quantitative knowledge of the 3D shape/surface of the microscopic samples is of high importance. The vast literature of used techniques for this purpose can be categorized into three major classes: a) single-view, b) multi-view, and c) hybrid strategies [188]. In single-view approaches, using varying lighting (electron beam) directions on a single perspective, a group of 2D SEM micrographs are captured and utilized for 3D SEM surface modeling. In multi-view strategies, on the other hand, a set of 2D SEM images from different perspectives assists the 3D SEM surface reconstruction process. While each technology carries its own cons and pros, the hybrid mechanisms try to combine single-view and multi-view algorithms to restore a 3D shape model from 2D SEM images.

The use of single-view algorithms and its application to 3D SEM surface reconstruction

have been well studied in the literature. The Photometric Stereo (PS) [201] as the major strategy in this class tries to estimate the surface normal vectors of the microscopic sample by observing the object being illuminated from different directions. Paluszynski et al. [143] designed a single-view 3D surface modeling approach based on the PS algorithm which also incorporates advanced signal processing algorithms along with both SE and BSE detectors to restore the 3D shape model of SEM images. Pintus et al. [153] developed an automatic alignment strategy for a four-source PS technique for reconstructing the depth map of SEM specimen. Kodama et al. [98] designed a genetic algorithm to tackle the topographical surface reconstruction problem of SEM based on PS method. The proposed genetic algorithm has been applied to the line profile reconstruction from the signals captured by both SE and BSE detectors. Vynnyk et al. [194] proposed a PS based algorithm to 3D SEM surface reconstruction and studied the efficiency of SEM detector system towards a 3D modeling. Slowko et al. [177] designed a PS-based algorithm to reconstruct the 3D surface model of SEM micrographs with the use of angular distribution of back-scattered electron emission to achieve a digital map of surface elevations. This contribution examined different SEM environmental conditions as a high vacuum SEM which was equipped with the BSE detector system utilized for 3D surface reconstruction.

One of the most promising class of methods for 3D surface modeling of SEM images has been the multi-view class which is based on acquisition of multiple images from different perspectives. The Structure from Motion (SfM) [99, 56] and Stereo Vision [3, 126, 125] algorithms are advanced visual computational methods which take into account pixels/feature-points matching to assist for accurate 3D SEM surface reconstruction. The class of multi-view 3D reconstruction approaches can be categorized into two major classes: a) sparse feature-based approaches and b) dense pixel based approaches. While methods from the first class are employed to establish a set of robust matches between an image pair or a set of images based on sparsely placed distinct feature-points, dense multi-view techniques try to discover matches for all points in the images. These matches along with other computational methods will then be used to accurately estimate the projective geometry and 3D surface models [80]. Raspanti et al. [159] presented

a high resolution dense multi-view method for 3D reconstructions of biological samples obtained by a SEM. The work implemented novel solutions including a neural adaptive points matching technique to tackle the problem of dense 3D reconstruction. Samak et al. [167] developed a SfM-based algorithm to restore 3D surface model of SEM micrographs. The proposed method initialized a set of 3D points from 2D corresponding points and then triangulated the obtained 3D points into the 3D surface mesh with a mapped texture on the shape model. Carli et al. [39] evaluated the uncertainty of stereo vision algorithm for the problem of 3D SEM surface modeling. Uncertainty for different cases of tilt and rotation were discussed in the work and a relative uncertainty equal to 5% and 4% was achieved for the case of rotation and tilt respectively. Zolotukhin et al. [220] studied the pros and cons of SfM algorithm focusing on two-view 3D SEM surface reconstruction problem. Tafti et al. [188, 185] reviewed the state of the art 3D SEM surface reconstruction solutions, addressing several enhancements for the research study, and developed a sparse mutli-view algorithm to tackle 3D SEM surface modeling problem. Using machine learning solutions and adaptive strategies, Tafti et al. [187] proposed an improved sparse feature-based multi-view method which outperforms their earlier work in terms of accuracy and computation demands. SEM as an advanced imaging equipment requires careful modification/configuration of internal parameters for 3D reconstruction solutions. Marinello et al. [124] analyzed and studied the 3D reconstruction of SEM images based on different instrumental configurations including calibration, title-angle, magnification and etc. Applications of such sparse/dense matching based techniques can also be found in the works of Mona et al. [52] and Limandri et al. [112], Woo Kim [96] and Gontard et al. [70]. Inspired by the above-mentioned approaches, attempts in devising hybrid approaches to combine single-view and multi-view algorithms for restoring the 3D shape model of a microscopic sample have been attempted [45].

In single-view 3D surface reconstruction, creating a full model of the microscopic sample is not possible since the images are limited to only one view-point. Moreover, recreating the SEM micrographs of the sample under different illumination conditions is difficult. On the other hand, multi-view approaches offer a more general and achievable

framework for the task. However, use of sparse-feature based approach results in blurred edges and smoothed surfaces. This is especially problematic for the very complex microscopic samples, similar to the ones considered here. This requires more advanced matching techniques to capture the very fine details which are missed otherwise, when using sparse feature-based approaches. With the advent of new computer vision-based matching techniques, more accurate and robust approaches can be developed for the problem of 3D surface reconstruction of microscopic samples. In this work, a novel methodology is introduced for high quality 3D reconstruction of microscopic samples using multi-view SEM images. This is to address the growing demand for more accurate reconstruction techniques in fields like biology where the level of complexity of samples is very high. Using the proposed approach, high quality surface meshes of highly complex microscopic samples can be generated which can be used for further quantitative analysis of the surface/shape attributes. The contributions of the current chapter can be summarized as follows:

1. The current chapter introduces and investigates a new optimized and robust approach for dense matching and high quality reconstruction of highly complex microscopic samples from sets of multi-view SEM micrographs. Here, a complete framework is proposed in a step-by-step fashion; from image acquisition to preprocessing to dense matching to depth estimation and finally mesh processing and 3D printing.

2. Taking advantage of non-local nature of median filtering, higher accuracy in finding dense matching points are achieved which results in a more truthful reconstruction of 3D surface. Moreover, additional step of weighted median filtering by use of the corresponding micrographs as guidance is proven to reduce the blurring effects near edges and boundaries of the objects.

3. Having a physical model can be beneficial in order to achieve a more realistic representation of the microscopic samples. Therefore, 3D printing of the reconstructed 3D models are considered here. This is to showcase the superior performance of the

proposed method in recovering very fine details as well as to provide the means for better understanding of the morphology of the samples.

The rest of the chapter is organized as follows. Section 5.2 contains detailed explanations of the techniques proposed in this work. Following a brief overview, the SEM imaging protocol is described. After discussing the pre-processing steps of sparse scale invariant feature transform (SIFT) and epipolar rectification, the method of optical flow estimation with non-local regularization is introduced. As for post-processing of the dense matching results, image guided weighted median filtering is introduced next. The section concludes with true depth estimation using the filtered dense matching results. In Section 5.3, the results generated by the proposed framework are presented with detailed comparisons with the state-of-the-art. Section 5.4 concludes the chapter.

5.2 Methods

5.2.1 Overview

The overview of the proposed work for high quality 3D reconstruction of complex microscopic samples is similar to what is proposed in the previous chapter with few modification to make is more suitable for dealing with microscopic sample with higher complexity. The multi-view micrographs are captured by using a SEM device with computer controlled specimen stage. From each sample, two micrographs are acquired. Epipolar rectification using sparse SIFT features are done next to ensure a more horizontally-concentrated disparity variations between pixels of the micrographs. This is followed by dense correspondence between the pixels using optical flow estimation which provides the one-to-one correspondence between the matching points of the two micrographs. Even though the optical flow estimation method employed here is of high accuracy, high amount of detail contained in the used micrographs may have a negative effect on the outcome. This can be greatly remedied by weighted median filtering which takes advantage of the original micrographs as guidance for filtering the computed disparity map. Finally, the filtered

disparity map is used for high fidelity 3D reconstruction. In the following subsections each of the steps are elaborated in more detail.

5.2.2 SEM Imaging Protocol

In this work, a Hitachi S-4800 field emission scanning electron microscope (FE-SEM) has been utilized to generate the micrographs. This SEM is equipped with a computer controlled 5 axis motorized specimen stage which enables movements in x, y and z directions as well as tilt (-5 to 70°) and rotation (0 to 360°). Specimen manipulations, such as tilt, z-positioning and rotation of the specimen stage, as well as image pre-processing and capture functions were operated through the Hitachi PC-SEM software. The working distance which gives the required depth of focus was determined at the maximum tilt for every single sample at the magnification chosen for image capture. As the specimen was tilted in successive 1° increments until reaching the final value through the software application, the SEM image was centered by moving the stage in the x- and/or y-axes manually. The micrographs were acquired with an accelerating voltage of 3 or 5 kV, utilizing the signals from both the upper and lower SE detectors, as shown in Figure 5.1. The magnification and working distance were held fixed in each captured image of the tilt series. Contrast and brightness were adjusted manually to keep consistency between SEM micrographs. Table 5.1 summarizes the data that used in this work. Micrographs from *Arabidopsis Anther 1*, *Arabidopsis Anther 2*, *Graphene*, *Pseudoscorpion* and *Ash* are considered for evaluating the performance and accuracy of the proposed approach.

5.2.3 SIFT Feature Detection/Matching and Epipolar Rectification

The four stages of feature detection/description involved in SIFT method can be summarized as [120]: 1) scale-space extrema detection, 2) keypoint localization, 3) orientation assignment and 4) keypoint descriptors. For the first step, a Gaussian function is considered as the scale-space kernel. The local extrema of the response of the image to the difference-of-Gaussian (DoG) masks of different scales is found in a $3 \times 3 \times 3$ neighbor-

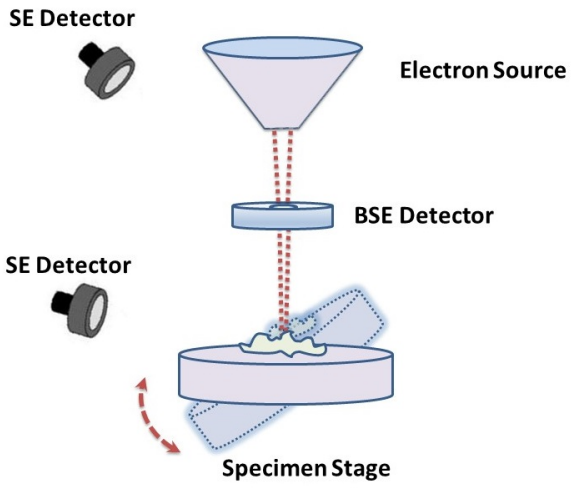
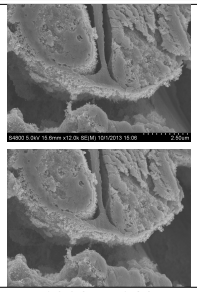
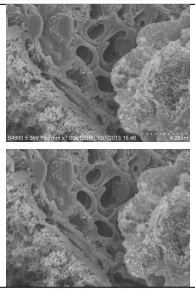
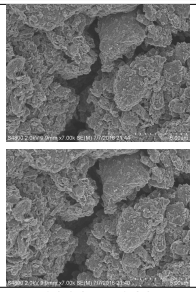
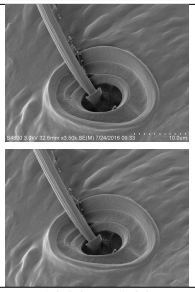
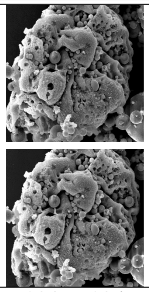


Figure 5.1: SEM imaging procedure used for this study.

Table 5.1: Dataset acquired using a Hitachi S-4800 field emission scanning electron microscope (FE-SEM) by tilting the specimen stage by 7°. The images for *Pseudoscorpion* set are rotated by 90° for visualization purposes.

	Arabidopsis Anther 1	*Arabidopsis Anther 2*	*Graphene*	*Pseudoscorpion*	*Ash*
Images					
Size	1280 × 960	1280 × 960	1280 × 960	1280 × 960	926 × 924
Tilt Angle	7°	7°	7°	7°	7°

hood of the interest point. After several stages of processing for removing the poorly defined keypoints in low contrast regions and near edges using quadratic function fitting and thresholding, the corresponding orientations can be assigned to the keypoints. This is followed by creating a 36-bin histogram for orientations in the keypoint's neighborhood by considering contributions from each neighbor, weighted based on their gradient magnitude and also by a Gaussian-weighed circular window around the keypoint. Using the location, scale and orientation determined for each keypoint up until now, the keypoint's descriptor is computed by combining the gradients at keypoint locations, as computed in the previous steps, weighted by a Gaussian function over each 4×4 sub-region in a 16×16 neighborhood around the keypoint into 8-bin histograms. This results in a $4 \times 4 \times 8 = 128$

element vector for each keypoint.

Given a set of two SEM images of a microscopic sample captured by titling the specimen stage, the epipolar rectification step aims to transform the images to only have horizontal displacements (disparity) between the corresponding pixels within the images. Assuming a set of sparse naively-matched (using nearest neighbors (NN) algorithm) SIFT feature points followed by *a contrario* RANSAC (ORSA) outlier removal algorithm [137] and represented as 3-vectors of homogeneous coordinates for the left (X_l) and right (X_r) images, the epipolar constraint can be written as [80]:

$$X_l^T F X_r = 0 \tag{5.1}$$

where F is the fundamental matrix that captures the rigidity constraint of the scene. Having a rectified pair, the fundamental matrix takes the especial form of:

$$F = [e_1]_\times = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \tag{5.2}$$

which means that the epipoles are at infinity in horizontal direction. Therefore, the process of rectification involves finding homographies to be applied to the left and right images to satisfy the epipolar constraint equation when $F = [e_1]_\times$:

$$X_l^T F X_r = 0 \equiv (H_l X_l)^T [e_1]_\times (H_r X_r) = 0 \tag{5.3}$$

Having a rotation matrix R for the camera around the focus point, a homography matrix can be formulated as $H = KRK^{-1}$ where K is the camera parameters matrix with (x_c, y_c) as the image center (principal point) and f as the unknown focal length: $K = [f\ 0\ x_c; 0\ f\ y_c; 0\ 0\ 1]$. Following the formulation proposed in [66, 138] we look for rotation matrices R_l and R_r and focal length which satisfy:

$$E(x_l, y_l, x_r, y_r) = X_l^T K^{-T} R_l^T K^T [e_1]_\times K R_r K^{-1} X_r = 0 \tag{5.4}$$

where $R_r = R_z(\theta_{rz})R_y(\theta_{ry})R_x(\theta_{rx})$, $R_l = R_z(\theta_{lz})R_y(\theta_{ly})$ and $K = K(f = 3^g(w+h))$, with w and h as the width and height of the input images respectively and g in the range $[-1, 1]$. It should also be noted that due to the specific form of $[e_1]_\times$ all of the rotations around the x direction are eliminated since $R_x^t[e_1]_\times R_x = [e_1]_\times$. Assuming the Sampson's error as:

$$E_s^2 = E^T(JJ^T)^{-1}E \tag{5.5}$$

where J is the matrix of partial derivatives of E with respect to the 4 variables:

$$J = ((FX_r)_1 \;\; (FX_r)_2 \;\; (F^TX_l)_1 \;\; (F^TX_l)_2) \tag{5.6}$$

we have

$$E_s(X_l, X_r)^2 = \frac{E(X_l, X_r)^2}{||[e_3]_\times F^T X_l||^2 + ||[e_3]_\times F X_r||^2} \tag{5.7}$$

Utilizing Levenberg-Marquardt [141], the method seeks the parameters (θ_{ly}, θ_{lz}, θ_{rx}, θ_{ry}, θ_{rz}, g) which minimize the sum of Sampson errors over the matching pairs. The optimized parameters are then used for building the two homographies to be applied to the left and right view images. More elaboration regarding the theory and implementation aspects of the rectification method can be found in [66, 138].

5.2.4 Dense Matching by Optical Flow Estimation

Finding a dense matching map between individual pixels of the input SEM micrographs is of high importance for high quality depth estimation and point cloud generation. One should note that the images are captured of rigid objects, with the only change being in the viewpoint angle. The rigidity of the microscopic samples, then, should be preserved in the dense correspondence maps that are found. This is generally satisfied since the imaged objects are well-textured which makes the process of matching more robust. On the other hand, edges/discontinuities contained in the micrographs should be preserved. This is mainly necessary for distinguishing different regions of more complex microscopic samples, similar to that of considered here (refer to Table 5.1). Being able to preserve the

discontinuities benefits the depth estimation greatly. However, the correspondence maps found should be piece-wise smooth which is usually satisfied in the formulation of energy functional required for matching. For the current work, dense matching is achieved using high quality optical flow estimation.

Optical flow estimation introduced by [85] refers to the estimation of displacements of intensity patterns in image sequences ([62], [11]). Generally speaking, the problem can be formulated as a global energy optimization problem of the form $E_{Global} = E_{Data} + \lambda E_{Prior}$ where the data term, E_{Data}, measures the consistency of the optical flow for the input images and the prior term, E_{Prior}, applies additional constraints for having a specific property for the flow field, for example smoothly varying flow fields. The choice of each term in the global energy functional and also the optimization algorithms varies in different methods for optical flow estimation. Assuming a two-frame (I_1 and I_2) formulation, the objective function can be written as:

$$\begin{aligned} E(\mathbf{u},\mathbf{v}) = \sum_{i,j} \{ & p_D(I_1(i,j) - I_2(i+u_{i,j}, j+v_{i,j})) \\ & + \lambda[p_S(u_{i,j} - u_{i+1,j}) + p_S(u_{i,j} - u_{i,j+1}) + p_S(v_{i,j} - v_{i+1,j}) + p_S(v_{i,j} - v_{i,j+1})]\} \end{aligned} \tag{5.8}$$

with $\mathbf{u}$ and $\mathbf{v}$ as the horizontal and vertical components of the flow field, i, j as the pixel indexes, λ as the regularization parameter and finally, p_D and p_S as the data and spatial prior penalty functions, respectively. In the original work of [85] quadratic functions are used for both the data and spatial penalty functions. But in the literature examples of using Charbonnier ($p(x) = \sqrt{x^2 + \epsilon^2}$)) in the work of [34] and Lorentzian ($p(x) = \log(1 + \frac{x^2}{2\sigma^2})$) in the work of [28] penalty functions and their variants can be found which provide a more robust estimation of the underlying flow fields. To account for large displacements between frames, the above formulation is usually minimized in a multi-resolution manner using incremental pyramid schemes, with steps of Gaussian anti-aliasing and flow outlier removal filters between iterations. This helps the process of linearization of the objective

function manageable and ensures lower chances of being trapped in local optima. As thoroughly discussed in [182], median filtering of the optical flow estimates after each pyramid level has a big impact in the final outcome of the minimization process: while the final energy is higher than what is achieved without median filtering, the optical flow error is minimized. This is due to the *non-local* nature of median filtering which is different from the local pairwise smoothness term. Incorporating the non-local median filtering heuristic as a weighted term in the energy functional can be considered as a means for ensuring minimal over-smoothing across boundaries. This is empirically useful for the problem of dense matching in SEM stereo pairs, especially since very fine details has to be preserved to obtain a more accurate 3D reconstruction.

Explicit formulation of the median filtering in Eq. 5.8 can be approximated by

$$\begin{aligned} E(\mathbf{u},\mathbf{v}) = &\sum_{i,j}\{p_D(I_1(i,j) - I_2(i+u_{i,j}, j+v_{i,j})) \\ &+ \lambda[p_S(u_{i,j}-u_{i+1,j}) + p_S(u_{i,j}-u_{i,j+1}) + p_S(v_{i,j}-v_{i+1,j}) + p_S(v_{i,j}-v_{i,j+1})]\} \\ &+ \lambda_N \sum_{i,j} \sum_{(i',j')\in\mathcal{N}_{i,j}} (|u_{i,j}-u_{i',j'}| + |v_{i,j}-v_{i',j'}|) \end{aligned} \tag{5.9}$$

in which $\mathcal{N}_{i,j}$ is the neighborhood centered at (i,j) and λ_N is the weight determining the contribution of the non-local weighted median term. Due to difficulty of optimization of Eq. 5.9 when having large spatial terms, the objective function can be relaxed using a set of auxiliary horizontal ($\hat{\mathbf{u}}$) and vertical ($\hat{\mathbf{v}}$) flow field components:

$$\begin{aligned} E(\mathbf{u},\mathbf{v},\hat{\mathbf{u}},\hat{\mathbf{v}}) = &\sum_{i,j}\{p_D(I_1(i,j) - I_2(i+u_{i,j}, j+v_{i,j})) \\ &+ \lambda[p_S(u_{i,j}-u_{i+1,j}) + p_S(u_{i,j}-u_{i,j+1}) + p_S(v_{i,j}-v_{i+1,j}) + p_S(v_{i,j}-v_{i,j+1})]\} \\ &+ \lambda_C(||\mathbf{u}-\hat{\mathbf{u}}||^{\mathbf{2}} + ||\mathbf{v}-\hat{\mathbf{v}}||^{\mathbf{2}}) \\ &+ \lambda_N \sum_{i,j} \sum_{(i',j')\in\mathcal{N}_{i,j}} (|\hat{u}_{i,j}-\hat{u}_{i',j'}| + |\hat{v}_{i,j}-\hat{v}_{i',j'}|) \end{aligned} \tag{5.10}$$

where λ_C is a scalar weight which penalizes the contribution of differences between the auxiliary and main flow fields. The current formulation with L_1 minimization is in close accordance with median filtering [181]. Assuming the above explicit representation of median filtering as part of the energy minimization functional, more improvement can be achieved by employing a weighted approach based on the approximate classification of the pixels in the neighborhood. In the non-local term, given a pixel and knowing which pixels in the neighborhood belong to the same surface, higher weights can be assigned while for the other pixels weights are lower [35]. In this manner, the non-local term in Eq. 5.10 is replaced with:

$$\sum_{i,j} \sum_{(i',j') \in \mathcal{N}_{i,j}} w_{i,j}^{i',j'} (|\hat{u}_{i,j} - \hat{u}_{i',j'}| + |\hat{v}_{i,j} - \hat{v}_{i',j'}|) \tag{5.11}$$

The weights $w_{i,j}^{i',j'}$ can be approximated by taking the spatial distance, color-value distance and occlusion states into account ([168, 213, 202]):

$$w_{i,j}^{i',j'} \propto \exp\{-\frac{|i-i'|^2 + |j-j|^2}{2\sigma_1^2} - \frac{|\mathbf{I}(i,j) - \mathbf{I}(i',j')|^2}{2\sigma_2^2 n_c}\} \frac{o(i',j')}{o(i,j)} \tag{5.12}$$

where $\mathbf{I}$ is the color vector in the Lab color space, n_c is the number of color channels, $\sigma_1 = 7$ and $\sigma_2 = 7$. The occlusion variable $o(i,j)$ is defined as:

$$o(i,j) = \exp\{-\frac{d^2(i,j)}{2\sigma_d^2} - \frac{(I(i,j) - I(i + u_{i,j}, j + v_{i,j}))^2}{2\sigma_e^2}\} \tag{5.13}$$

where $d(i,j)$ is the one-sided divergence function (only negative values, and positives considered as zero). This variable is near zero for occluded pixels while close to one in non-occluded regions. The parameters σ_d and σ_e are set to 0.3 and 20, respectively according to [168]. Following the work of [111], an approximate solution for the auxiliary flow filed components, $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$, can be found for all of the pixels.

Full implementation of the above requires high computational power. A simple modification can reduce the computational need immensely. Since the weighted formulation is designed to overcome the negative impacts of over-smoothing boundaries in the process

of optical flow estimation while the estimates in the uniform regions are very accurate, different methodologies can be applied to ensure an accurate solution while demanding less computational power. Using a Sobel edge detector and having the current estimate of optical flow, motion boundaries can be detected and then dilated to determine the flow boundary regions. In these regions the weighted formulation with a 15×15 neighborhood is applied while in non-boundary regions, a 5×5 un-weighted approach is taken. This will reduce the computational time drastically.

Optimizing Eq. 5.10 will result in the flow field representing how the pixels moved between the micrographs. Given that the input micrographs are rectified, the vertical components of the flow fields are negligible in comparison to the horizontal components. In fact, the energy of the vertical disparity map is less than 1% of that of horizontal disparity. Considering this, the vertical disparity map is disregarded for the rest of the steps.

5.2.5 Disparity Refinement by Weighted Median Filter

As can be seen from the micrographs used in the current work, the level of detail can be very high due to presence of many microscopic objects in the samples. This can be mainly problematic since the variation of the size is also large. A great representative is the *Ash* sample in which objects of various sizes as well as regions with different textures are present. This cannot be fully recovered by the previous steps and therefore, our goal of a more truthful 3D reconstruction can be compromised. However, this can be greatly remedied by using the original images for guiding towards a more accurate correspondence. Here, we propose to use *weighted median filtering* as means for error correction. In this manner, the original images serve as guidance for a more accurate filtering of the computed disparity map.

Weighted median filter, as is obvious from the name, aims to replace the image pixels with weighted median of the neighborhood pixels within a local window [33, 123]. Assuming image I and the corresponding feature map $\mathbf{f}$, and pixel p in image I located at the center of a local window $\mathcal{R}(p)$ with radius r, for each pixel $q \in \mathcal{R}(p)$ a weight w_{pq}

will be assigned which is a representative of the affinity of the two pixels in the feature map $\mathbf{f}$. This can be represented as

$$w_{pq} = g(\mathbf{f}(p), \mathbf{f}(q)) \tag{5.14}$$

where g is the influence function (Gaussian, reciprocal, cosine, etc.) [217]. Given $n = (2r+1)^2$ as the number of pixels in the local window $\mathcal{R}(p)$, the value and weight element os all n pixels can be expressed as $\{(I(q), w_{pq})\}$. After sorting values in an ascending order, the weighted median operator returns the new pixel p^* such as:

$$p^* = min\ k \quad s.t. \quad \sum_{q=1}^{k} w_{pq} \geq \frac{1}{2} \sum_{q=1}^{n} w_{pq}. \tag{5.15}$$

which means that the sum of corresponding weights for all pixels before p^* should be almost half the sum of all weights. It should be noted that in this formulation, feature map $\mathbf{f}$ determines the weights.

For our work, use of weighted median filtering is considered for achieving a more accurate correspondence. Given the computed disparity map from the previous step, and also having the first micrograph from each image set that is used for optical flow estimation as the feature map, the disparity map is filtered using the weighted median filter. Even though the straightforward implementation of the method is simple, it can be very time consuming due to spatially varying weights and the median property. [217] proposed the use of joint-histogram with median tracking and necklace table data structure for fast implementation of the weighted median filter. Employing this approach, a more detailed disparity map can be achieved which results in a higher fidelity 3D reconstruction.

5.2.6 Depth Estimation

Stereo rectification transforms the images in a manner in which the displacements will be grossly concentrated in the horizontal direction. This greatly simplifies the process of depth estimation. This is especially useful for the case of 3D reconstruction of SEM images since the tilt angles are very small with high amount of overlap between stereo

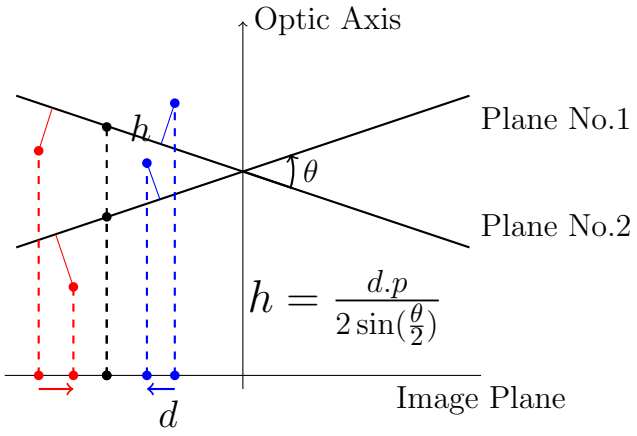


Figure 5.2: Relationship between the estimated height (h) and the computed horizontal disparity (d) using the pixel size in sample units (p) and the total tilt angle (θ).

image pairs. For more general problems like large scale multiple view stereo (MVS), the proposed technique is not directly applicable and more sophisticated methods are needed [176, 190, 65].

The horizontal disparity computed from the previous step, can be utilized for estimating the depth of the individual pixels contained in the images. This requires several parameters to be known: tilt angle, magnification and size of each pixel in sample units. Figure 5.2 shows the relationship between the computed horizontal disparity and the height for a few sample points. This can be represented using a simple trigonometric equation [163, 183, 203]:

$$h = \frac{d.p}{2\sin(\frac{\theta}{2})} \tag{5.16}$$

which uses the computed horizontal disparity d, pixel size in sample units (p) and the total tilt angle (θ) to estimate the height (h).

5.3 Results & Discussions

Assessing the performance of the proposed method is done in several steps both qualitatively and quantitatively. Using a Hitachi S-4800 field emission scanning electron microscope (FE-SEM), the micrographs of the five sample sets (*Arabidopsis Anther 1*, *Arabidopsis Anther 2*, *Graphene*, *Pseudoscorpion* and *Ash*) are captured. The device is equipped with computer controlled 5 axis specimen stage which enables movements along

three coordinate axis as well as tilting and rotation. The process of image acquisition is done in a controlled manner by an expert with manual adjustments of focus and re-centering when needed. Between the two views acquired for each sample, only the tilt angle is changed while the distance between the specimen stage and the detectors as well as zooming factor are kept constant. In order to keep the image acquisition sessions consistent, the tilt angle between micrographs of each set is set to 7°. However, similar tilt angles in close range would produce the same results as evidenced by our previous experiments. One should note that the amount of overlap between images is a key factor in a more accurate 3D reconstruction. Keeping the tilt angle small, as well as re-centering the sample after tilting the specimen stage will ensure a more accurate and robust matching and therefore result in a more truthful reconstruction.

The first step of the proposed approach consists of finding distinctive feature points in the two input micrographs from each set to be used for stereo rectification. Given the initial SIFT feature points, SIFT descriptors are computed as described in Section 2.3. This is followed by putative matching of the SIFT descriptors considering naive nearest neighbor search. Since it is assumed that SIFT descriptors capture information about the neighborhood of each feature point, putative matching produces reasonable number of correct matches. However, it cannot be expected to have a completely accurate matching between feature points due to noise and also similarities in textures contained in the input images. Therefore, one should find a reasonable transform between the matching points that satisfies some error criteria for the majority of matched features. In our work, without going into much detail as this subject is a very well-studied concept in computer vision, a variant of random sample consensus (RANSAC), namely *a contrario* RANSAC (ORSA), is used in order to find correct matches that satisfy a homography transform between the two images. This is followed by formulating the Sampson's error to be used for rectifying the input pair in order to have horizontally concentrated matchings. This step is necessary for the process of dense matching needed for high quality 3D reconstruction. In sparse feature-based approaches used for 3D reconstruction of microscopic samples [186, 187], computation of fundamental matrix and the subsequent projective

Table 5.2: Rectification results: number of SIFT points found in each input image (rows 1 and 2), number of matching points after *a contrario* RANSAC (row 3), initial and final rectification errors from before and after the quasi-Euclidean rectification (rows 4 and 5). As can be seen, despite careful image acquisition, the initial rectification errors are large.

	Arabidopsis Anther 1	*Arabidopsis Anther 2*	*Graphene*	*Pseudoscorpion*	*Ash*
im.1 # SIFT keypoints	783	981	2089	195	1633
im.2 # SIFT keypoints	658	893	2488	65	1652
ORSA # SIFT matches	**214**	**268**	**487**	**18**	**418**
initial rect. err. (pix)	2.393	6.910	14.055	1.223	2.766
final rect. err. (pix)	**0.802**	**0.425**	**0.277**	**0.971**	**0.472**

transformation is computationally efficient. This is due to small number of matching points in comparison to the total number of pixels in the images. However, having the dense matching for all the pixels in the images requires specific configurations. Rectifying the input pair simplifies the problem of 3D point cloud generation. In this case, the need for computing the fundamental matrix and projective transformation using all of the matching points is eliminated. Table 5.2 summarizes the result of the rectification process used for this study for all of the sample sets. The first and second row in the table represents the number of individual SIFT feature points found in the input images. This is followed in the third row by the number of true matches after putative nearest neighbor matching and ORSA outlier elimination. Even though this number consists of a small portion of the initial matches, however, for the purpose of stereo rectification is enough. The number of initial and final matches is lower for the *Pseudoscorpion* set due to lower amount of variations and texture in the images of the set. The table continues with the initial and final rectifications errors obtained using the quasi-Euclidean stereo rectification process. Having a more horizontally-concentrated matching between image pixels will ensure more accurate and robust 3D reconstruction.

The rectification step is followed by optical flow estimation to determine the dense matching between individual image pixels in the image pair. Figure 5.3 shows the results of optical flow estimation. For better visualization of the effects of dense matching, the difference maps are also displayed. The first row shows the initial difference map between the input images of the pair. The second row shows the estimated optical flows for compensating the movements of individual pixels in the two images. The computed

flow is color-coded, with red representing positive values and blue representing negative values. Utilizing the computed optical flow estimates, the first image of the pair can be warped to generate the second image. The difference maps between the warped first image and the second image of each pair are excellent representatives of the performance of the matching procedure. These are shown in the third row of Figure 5.3. Inspecting the computed optical flows reveals very important properties of the image matching that is required here. Image registration as an example of image matching tries to find the correspondence between pixels of two images [16, 18]. However, the general formulation assumes that the matching points are all laid on the same plane. This is not the case for many computer vision problems, optical flow estimation included. In such cases the computed correspondence must be discontinuity preserving. In order words, an image as a projective depiction of a scene may contain several objects which can move independently and therefore, the computed flow patterns should account for that. Having a rectified stereo pair as input to the optical flow estimation approach results in a horizontally-concentrated flow estimate, as expected. Our experiments show that the energy contained in the vertical component of optical flow is less than 1% of the horizontal component which is ideal for an accurate reconstruction. Therefore, for 3D reconstruction, only the horizontal component is used as the disparity map.

Even though the employed optical flow approach produces highly accurate results, due to lack of color in the initial SEM images, the results may suffer from blurred edges. This is mainly problematic in highly complex samples used here, *Ash* for example. To ensure a more accurate estimation, further post-processing is done using weighted median filtering as described in Section 2.5. Using the first image as guidance, because the optical flow is computed from the first to the second image in the pair, the disparity map is filtered taking advantage of weighted median filtering. Figure 5.4 shows the effects of the employed post-processing filtering on a portion of the *Pseudoscorpion* and *Ash* disparity maps. While the initial disparity maps has blurred edges and bumpy appearances, the result of weighted median filtering is more sharp and accurate. Moreover, more detail is preserved in the resulted disparity map as can be seen from the presented images.

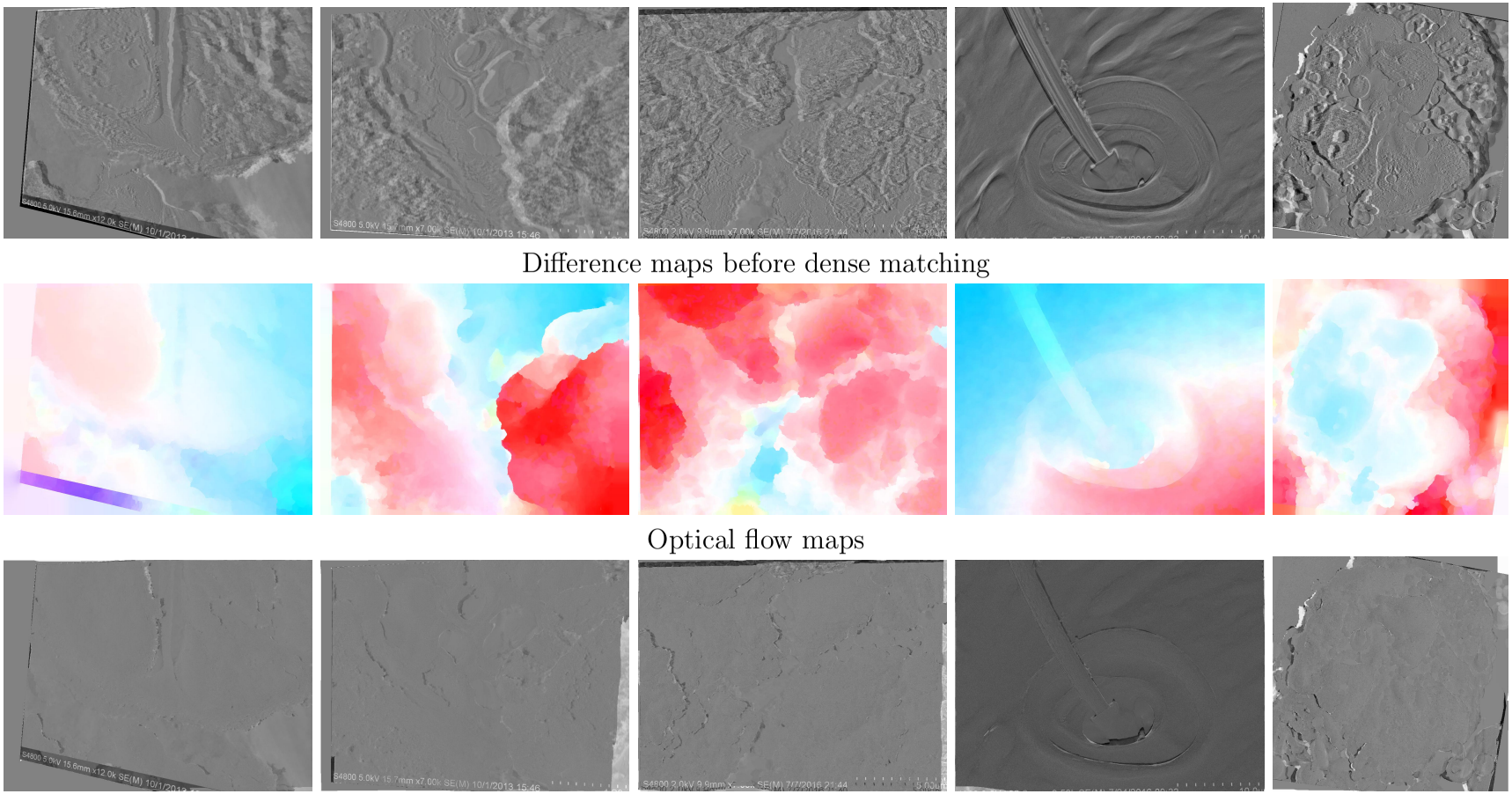


Figure 5.3: Optical flow estimation results for (a) *Arabidopsis Anther 1*, (b) *Arabidopsis Anther 2*, (c) *Graphene*, (d) *Pseudoscorpion* and (e) *Ash* sample sets. The first row shows the initial difference maps. The second row shows the computed optical flow estimate. Using the optical flow estimate, the first image in each pair is warped and then used for generating the final difference maps as depicted in the third row. It should be noted that the images for *Pseudoscorpion* set are rotated by 90° for visualization purposes.

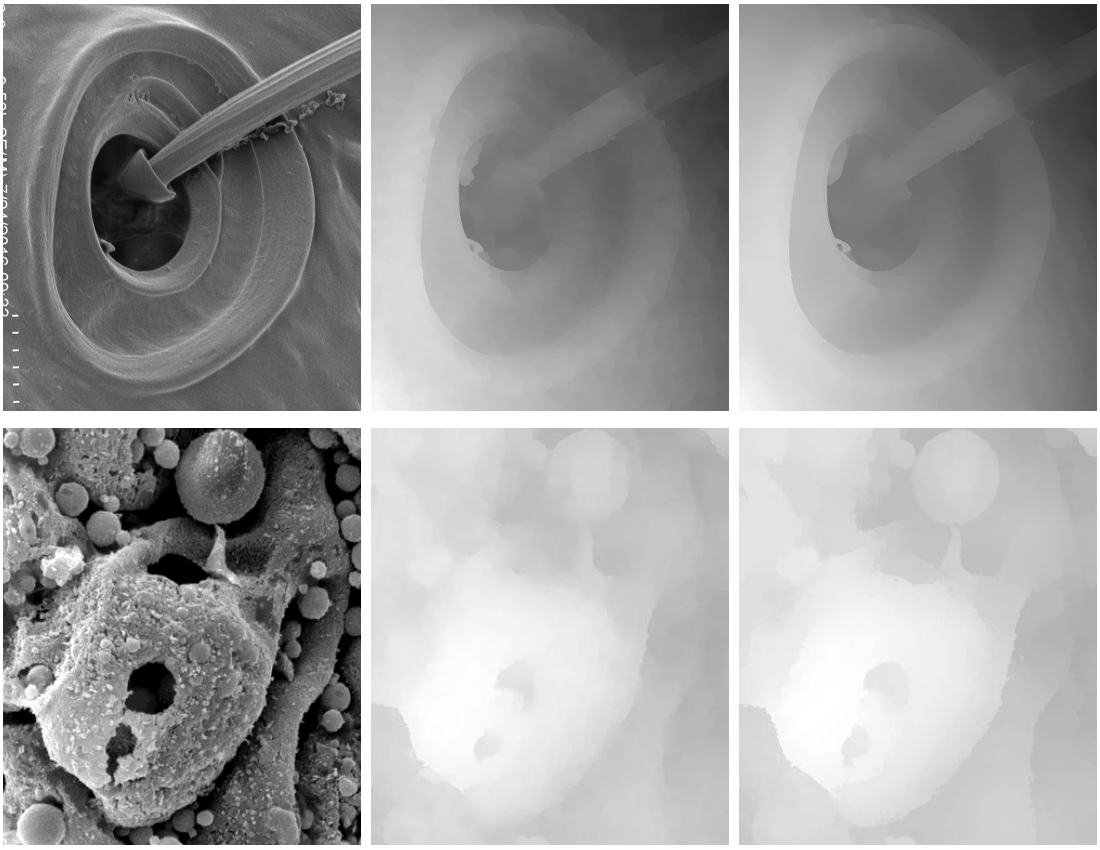
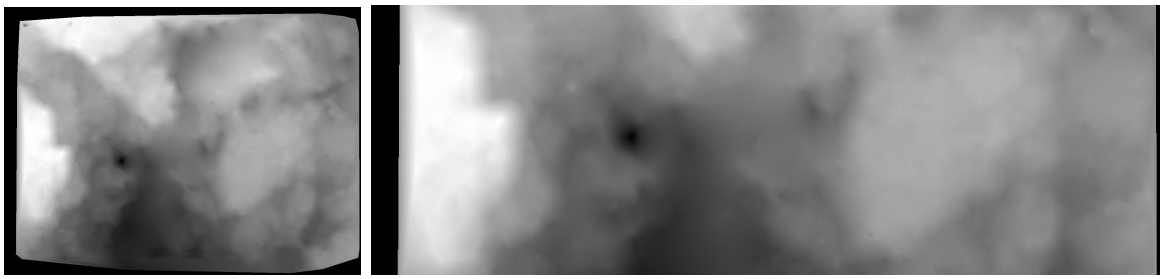


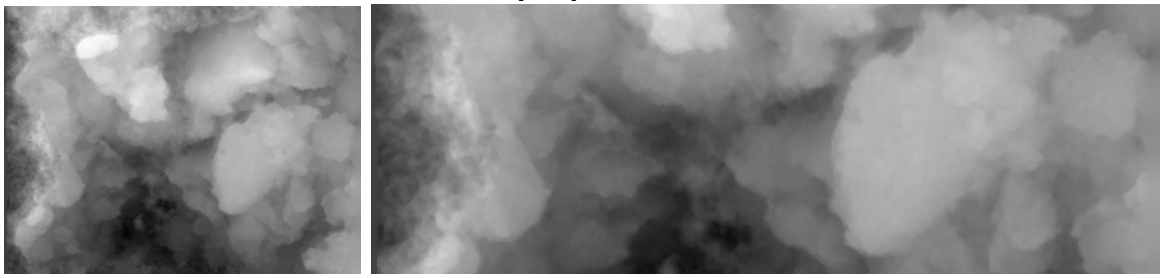
Figure 5.4: Effects of weighted median filtering on the horizontal disparity map: the first input micrograph that is used as guidance (left column), before (center column) and after (right column) disparity refinement. Despite inclusion of non-local term in the optical flow energy functional, the outcome can be improved greatly by adding an additional weighted median filtering step.

For a more comprehensive analysis, the proposed dense matching approach is compared with several other methods previously used in the literature for dense matching and subsequently 3D reconstruction. Sparse feature-based approaches track the movements of distinct feature points in the input images in order to compute the fundamental matrix and projective transformation [186, 187]. To generate a dense disparity map, similar to that of created by our approach for a better comparison of the performance, the sparse disparity values are interpolated employing a Delaunay triangulation-based interpolation method. As for dense matching schemes, the works of [85] and [116] are good examples. While the first one works based on the pixels correspondence, the later extends a similar idea to matching of dense SIFT descriptors.

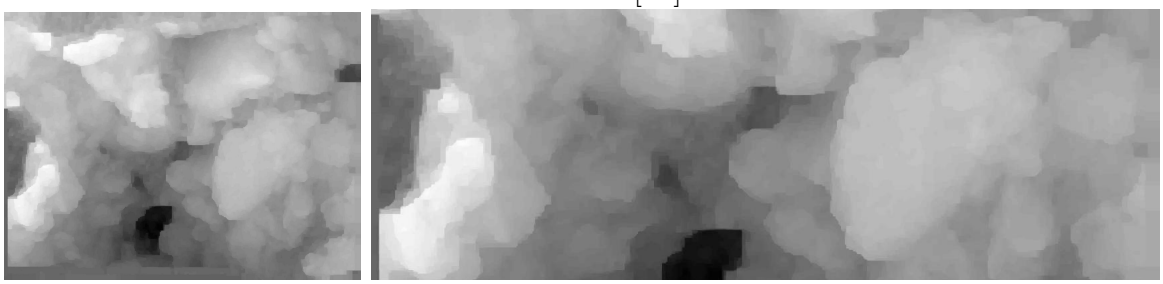
Figures 5.5 and 5.6 display the disparity maps computed using the above-mentioned methods as well as the proposed approach for the *Graphene* and *Ash* sample sets, re-



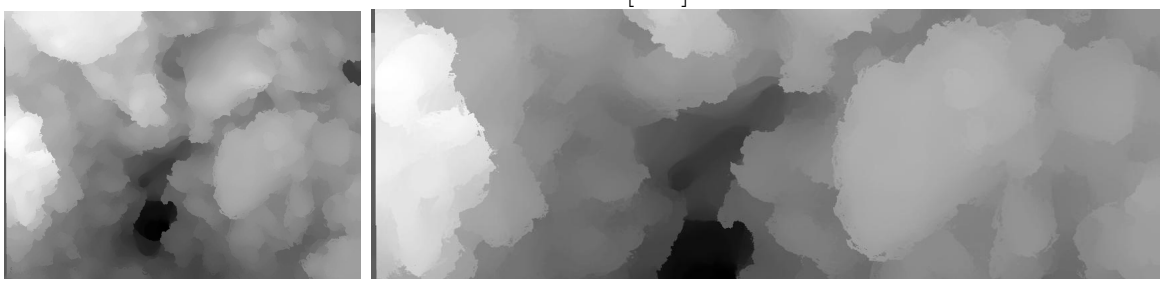
Method of [187] + interpolation



Method of [85]



Method of [116]



Proposed approach

Figure 5.5: Comparison of the results for *Graphene*: The first row is the overall as well as a zoomed region of the computed disparity map using the state-of-the-art method of [187] which uses sparse feature-based matching approach and *a contrario* RANSAC for outlier removal. The dense disparity map is created by scattered data interpolation of the sparse disparity values. The second row shows the result of Horn/Schunck optical flow estimation ([85]), which provides a better estimation of the disparity map than that of [187]. The third row shows the result of dense feature matching proposed in [116] which uses dense SIFT features as well as factor graph representation of the matching energy functional optimized by loopy belief propagation. Even though relatively better than [85], the result still suffers from blurred edges. The result of the proposed method is presented in the fourth row. In comparison to the state-of-the-art, the proposed approach generates a sharper and more accurate disparity map.

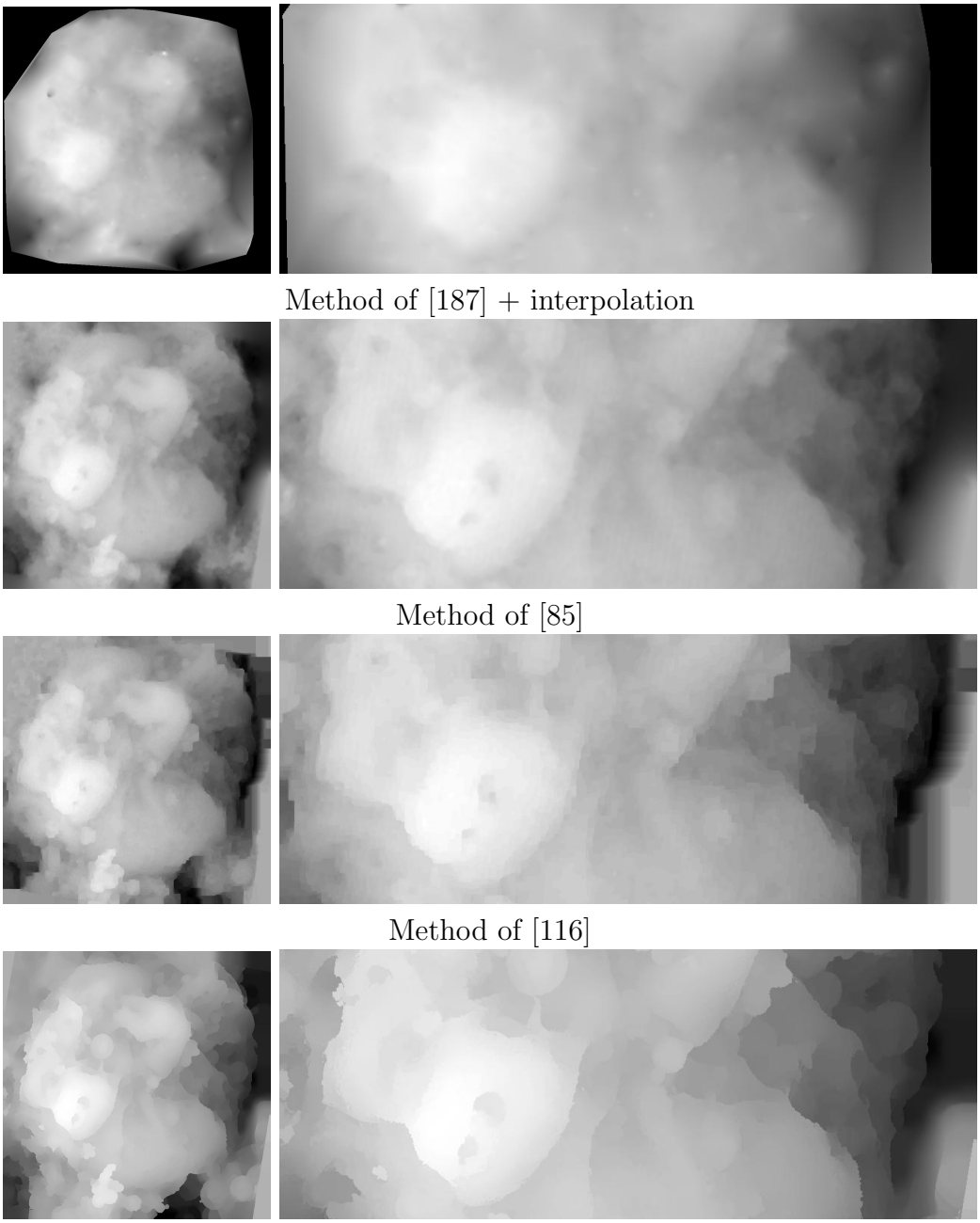


Figure 5.6: Comparison of the results for *Ash*: The first row is the overall as well as a zoomed region of the computed disparity map using the state-of-the-art method of [187] which uses sparse feature-based matching approach and *a contrario* RANSAC for outlier removal. The dense disparity map is created by scattered data interpolation of the sparse disparity values. The second row shows the result of Horn/Schunck optical flow estimation ([85]), which provides a better estimation of the disparity map than that of [187]. The third row shows the result of dense feature matching proposed in [116] which uses dense SIFT features as well as factor graph representation of the matching energy functional optimized by loopy belief propagation. Even though relatively better than [85], the result still suffers from blurred edges. The result of the proposed method is presented in the fourth row. In comparison to the state-of-the-art, the proposed approach generates a sharper and more accurate disparity map.

spectively. In each figure, the left column shows the overall disparity map while the right column is a zoomed view for a better visual comparison of the various techniques. Close inspection of the provided results displays the superiority of the proposed approach. As expected, the outcome of the sparse feature-based approach is highly blurred near edges with significant loss of details presented in the images. Even tough such techniques are mainly used with more than two input images, the performance is the same as evident from the results. In contrast, dense matching approaches produce more accurate results. Using the modern implementation of the method of [85] provided by [182] a more accurate disparity map is generated. In the results of the method, more details are present and discontinuities are better preserved. However, in cases of having larger displacements near the margins of the input images (left side of the *Graphene* results) the estimated optical flow is not as accurate as the sparse feature-based approach. Using the dense descriptor matching scheme in the work of [116], this is mostly resolved. In this technique, at first two 128-dimensional dense SIFT descriptor image of both the first and second image in the pair are created. To compute the matching, a factor graph representation of the specifically defined energy functional is introduced and the process of optimization is done using loopy belief propagation. Employing the dense descriptor matching methodology more accurate results can be achieved. The last row in Figures 5.5 and 5.6 is the disparity result using the proposed approach. Employing the proposed approach, higher levels of details can be reached in the resulted disparity maps. With higher accuracy in preserving the discontinuities a more truthful reconstruction can be made. This is more evident in the samples with higher complexity level, *Ash* sample set for example. As shown in Figure 5.6, the proposed approach can recover disparity values even for smaller objects in the images, while in contrast, the other methods presented here cannot due to high amount of blur around edges and boundaries.

Having the height estimate for each point, a dense 3D point cloud can be generated and further used for 3D surface reconstruction. To eliminate the effects of smoothing introduced by general purpose mesh generating toolsets, similar to that of used in MeshLab [127], Delaunay triangulation is done by utilizing the image grid as the set of mesh

nodes. The triangular mesh nodes are then transformed from the two dimensions of the image plane to the three dimensions of the model using the computed height estimates. This practically eliminates the smoothing effects which generally happen near the edges of the objects and in areas that contain sudden jumps due to sharp changes in the depth estimate. Using simple MATLAB scripts, the generated 3D surface can be transformed and saved as standard STL files which can be later used for further mesh modification and processing using more specialized software. Use of edge aware mesh smoothing procedures can be considered in order to have a more pleasing appearance without loosing details of the edges and sudden changes of depth. Figure 5.7 shows 3D red-cyan anaglyphs generated by combining the two rectified stereo views of the microscopic samples as well as the solid 3D models created using Meshmixer [171]. The generated models can also be sent out for 3D printing as the ultimate means for creating a tangible representation of the complex microscopic structure. Figure 5.8 (a) shows one image from *Ash* sample, while (b) shows another view of the 3D solid model created using the computed disparity estimates and modified using MeshMixer. Finally, (c) is an image captured from the 3D printed model.

5.4 Conclusions

In this chapter, a novel and accurate approach is introduced for high fidelity 3D reconstruction of highly complex microscopic samples. This is an extension to the framework proposed in the previous chapter. In the proposed methodology, multi-view SEM micrographs from two different view-points are captured using a Hitachi S-4800 field emission scanning electron microscope (FE-SEM). The micrographs are acquired with 7° tilt angle differences made possible by the provided computer controlled 5 axis specimen stage. The image acquisition is then followed by one stage of pre-processing which consists of four steps: a) sparse SIFT feature detection/description, b) nearest neighbor search for finding the putative sparse matching, c) *a contrario* RANSAC for outlier removal and finally d) quasi-Euclidean stereo rectification. This step is necessary due to the need for high

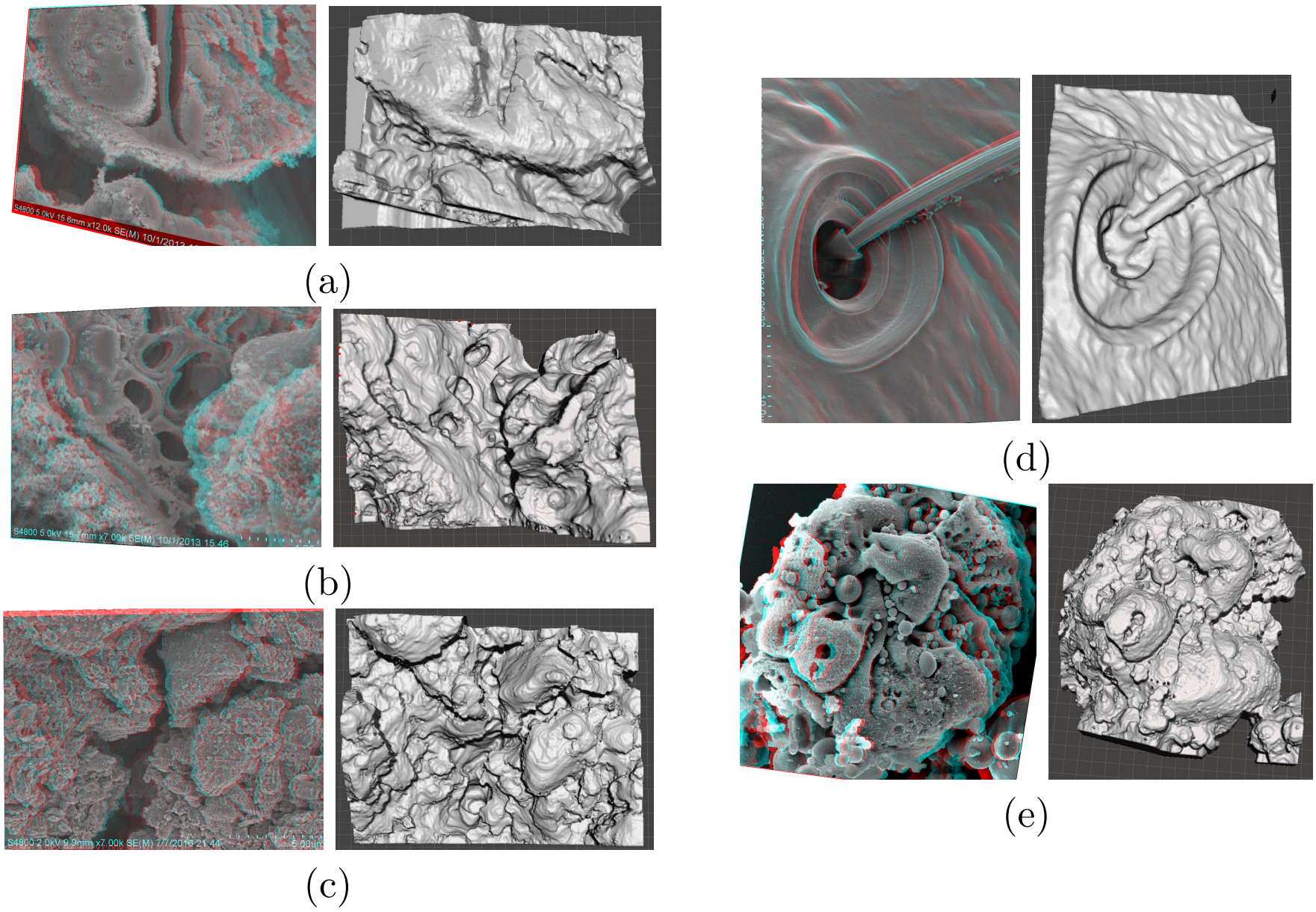


Figure 5.7: 3D red-cyan anaglyphs generated by combining the two rectified stereo views of the microscopic samples as well as the solid 3D models created using Meshmixer ([171]) for (a) *Arabidopsis Anther 1*, (b) *Arabidopsis Anther 2*, (c) *Graphene*, (d) *Pseudoscorpion* and (e) *Ash* sample sets

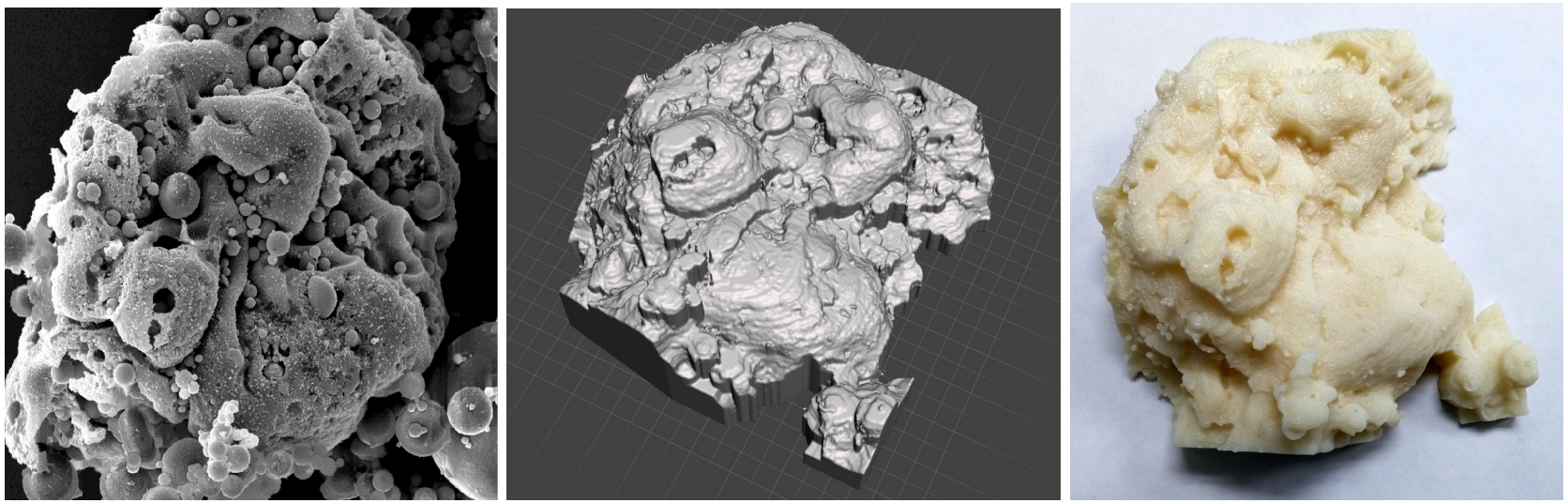


Figure 5.8: From start to finish: (a) first image from the *Ash* sample set, (b) 3D solid model generated using the computed disparity estimates and modified using MeshMixer ([171]), (c) 3D printed model. Using the proposed approach, highly complex structure of the sample was captured and reconstructed in the printed model.

quality dense correspondence required for accurate 3D reconstruction of highly complex samples used here. The pre-processing stage is followed by dense matching by employing non-local based optical flow estimation. Using this technique, a highly accurate estimate of dense correspondence can be achieved. To ensure a more accurate disparity map as well as eliminating blurred edges, a post-processing filtering step using weighted median filtering is done which uses the first image in each pair as the guidance. Finally, the disparity map is used to generate the 3D point cloud of the microscopic sample. The 3D point cloud is later used for high quality surface mesh generation. Quantitative and qualitative comparisons reveal the superiority of the proposed method to the state-of-the-art sparse feature-based techniques used for 3D surface reconstruction of SEM images. Moreover, the produced results are experimentally proven to be extremely accurate and suitable for 3D printing, as evident by the sample produced.

Chapter 6

Conclusions and Future Works

In this dissertation the problem of image matching and a vast range of its applications are discussed and investigated. Starting with Chapter 1, the thesis provides a general definition of the problem and then proceeds to dive deeper into various forms of image matching. Chapter 1 is concluded with brief introduction of the main topics covered in the rest of the thesis. Chapter 2 aims to investigate a few biomedical image processing problem in which the concept of image matching is present in their very core. Starting with the problem of slice interpolation, a new deformable image registration-based approach is proposed for accurate slice interpolation of volumetric MRI and CT scans. This is a necessary step since the resolution of the captured volumetric scans differ along different axes. Employing the proposed slice interpolation approach will enable a more accurate 3D reconstruction of the volumetric scans. Chapter 2 is continued with investigation of another biomedical image processing problem which can be efficiently solved by image matching. Retinal Optical Coherence Tomography (OCT) imaging as a highly regarded field has attracted many researcher around the world. In the last section of Chapter 1, a new image matching-based approach is proposed which takes advantage of Robust Principle Component Analysis (RPCA) as well as rigid image registration for speckle noise reduction on retinal OCT images. One bottleneck encountered when dealing with deformable image registration approaches is in regard to the computational complexity and demand of such techniques since the algorithm tries to estimate displacement vectors for all of the pixels contained in the input images. This will be more problematic when having very high resolution images for matching. To remedy this, a new triangular mesh-based registration approach is proposed in section 3 of Chapter 2.

The main focus in the dissertation is put on the applications of various image matching techniques for the problem of 3D surface reconstruction, especially microscopic surfaces,

with use of Scanning Electron Microscope (SEM) as means for capturing high quality 2D micrographs from simple and complex microscopic samples. The literature of used approaches for this purpose can be categorized into three major classes: 1) single-view, 2) multi-view, and 3) hybrid strategies. In single-view approaches, using varying lighting (electron beam) directions on a single perspective, a group of 2D SEM micrographs are captured and utilized for 3D SEM surface modeling. In multi-view strategies, on the other hand, a set of 2D SEM images from different perspectives assists the 3D SEM surface reconstruction process. The hybrid mechanisms try to combine single-view and multi-view algorithms to restore a 3D shape model from 2D SEM images. The class of multiview methods can be further divided into to categorizes: a) sparse feature-based techniques, and b) dense pixel-based approaches. While methods from the first class are employed to establish a set of robust matches between an image pair or a set of images based on sparsely placed distinct feature-points, dense multi-view techniques try to discover matches for all points in the images. These matches along with other computational methods will then be used to accurately estimate the projective geometry and 3D surface models. Chapter 3 aims to investigate the use of sparse feature-based approaches for the problem of 3D surface reconstruction of microscopic samples. Employing four well-known, widely used feature detector/descriptor approaches, namely SIFT, SURF, BRIEF and ORB, comprehensive comparisons are provided and the performance of the methods are investigated. In all of the cases, SIFT performs better than the others, with SURF being the next best method. However, a close inspection of the final results reveals the significant shortcomings of sparse feature-based approaches. Unlike general purpose 3D reconstruction problems, in SEM, the level of detail contained in the micrographs can vary drastically between different microscopic samples. Therefore a 3D reconstruction approach is expected to be able to handle such variations and produce a reconstruction with very high accuracy. Using sparse feature-based approaches result in a very smoothed reconstruction of the microscopic surface with all the fine details missing. Moreover, the performance is not consistent in all the regions of the input micrographs. This is mainly due to having various levels of details in different regions of the images. This

can be problematic when working with samples that contains very uniform regions. In such cases the first component for the sparse feature-based approaches, feature detection, cannot find enough number of matches for accurate estimation of fundamental matrix and projective transformations. Therefore the need for more robust and accurate approaches is of high importance.

Building on the concept of feature-based surface reconstruction, Chapter 4 introduces a new framework in which not only distinct feature but instead all the pixels of the input images are used for finding the appropriate correspondence and 3D reconstruction. The proposed framework works based on sparse-dense correspondence of features/pixels present in stereo SEM micrographs. The first stage of the framework uses SIFT features/descriptors as well as naive nearest neighbor (NN) search and *a contrario* RANdom SAmple Consensus (RANSAC) for quasi-Euclidean stereo rectification. This is followed by used of dense SIFT descriptors in a specifically designed matching energy functional represented as a factor graph and optimized by taking advantage of loopy belief propagation in order to find dense correspondence between all the pixels in the images. Due to rectification, the disparities are highly concentrated along the horizontal direction which makes the process of 3D point cloud generation very easy. The proposed framework outperforms sparse feature-based approaches greatly both in terms of accuracy and reliability. This framework is furthered improved in Chapter 5 by employing a more robust and accurate matching scheme as well as disparity refinement procedure. This is especially useful when dealing with microscopic samples with higher level of details. In such cases, one should ensure that boundaries of the objects contained in the micrographs are preserved in the process of matching. Not only that, but also, the improved framework will enable a more accurate reconstruction which can lead to higher fidelity 3D models that can be sent out for 3D printing as evident by the sample created as a tangible means of representations.

Future Research

Micrograph Pre-Processing

Here, a highly constrained optimization approach is employed for dense matching with custom regularization terms which minimizes the effects of image distortions. However, in special cases it may be required to pre-process the input micrographs for a better matching. These pre-processing steps range from edge-preserving noise reduction and blur removal to contrast enhancement. This is more necessary when having more than two views.

Different Dense Descriptors for Dense Correspondence

The first assumption in the majority of the methods proposed in the literature for dense matching and optical flow estimation is the brightness constancy during movements of pixels between images of the sequence. However, this is not always the case for SEM micrographs. One solution, as pursued in Chapter 4, is to use structural descriptors rather than pixels for estimating the matching. Use of dense descriptors for dense matching and optical flow estimation has been investigated in our previous works [14, 11] using various dense descriptors, such as Leung-Malik (LM) filter bank [108], Gabor filter bank [67], Schmid filter bank [170], Root Filter Set (RFS) filters, Steerable filters [64], Histogram of Oriented Gradients (HOG) [44] and Speeded Up Robust Features (SURF) [24]. The same approaches can be considered here using the above mentioned dense descriptors with the possibility of newer descriptors such as DAISY [190] which is proven to be useful for high accuracy dense matching.

Occlusion Handling

Occlusion handling as an interesting and challenging problem is a widely studied problem in the computer vision community [211, 202, 207]. This arises as a result of movements of objects in the scene or the change of imaging viewpoint. This is more problematic in case of large displacements of objects between frames in the image sequence which is largely the case for general purpose optical flow estimation or stereo matching. However,

for the problem of 3D reconstruction of microscopic samples using SEM micrographs, the problem is more relaxed. On one hand, the SEM micrograph acquisition is done in a very organized manner with careful sample preparation and controlled imaging procedures. On the other hand, unlike the general optical flow or stereo matching, the amount of displacements can be adjusted by manual manipulation of the specimen sample. This, as mentioned before, does not have a negative impact on the subsequent depth estimation since it will not alter the relative disparity between the matching points and may only elevate or decrease the mean depth of the whole microscopic sample. Moreover it should be noted that we have limitations on the possible tilt angles dictated by the SEM imaging system. However, in case of multiview stereopsis and/or for more complex microscopic samples, by taking occlusion handling procedures into account, a more accurate reconstruction can be achieved. Examples of such techniques can be seen in Figure 6.1 using the well-known Semi-Global stereo [84] and Patch-Match stereo [29] methods. While Semi-Global approach uses a fronto-parallel assumption for the disparity support windows, Patch-Match takes advantage of a slanted support window for a more accurate reconstruction. However, both can be used in order to detect occluded regions in the input micrographs. More extensive evaluations are left for future research. Moreover, incorporating the mesh generation procedures in the process of depth estimation can be considered in the future [216].

Hybrid Approaches: Combining SFM & SFS

In the class of single view 3D reconstruction approaches, images from a single viewpoint but with various lighting conditions are captured and used for the purpose of reconstruction. The methods in this class have been previously used for 3D reconstruction from SEM images [103, 48, 143, 153]. However, due to difficulty in generating SEM micrographs under different illumination directions, they achieved moderate success. Even though several hybrid approaches have been introduced in the literature combining SFS with SFM [45], the advent of modern SFS algorithms can improve the performance of 3D reconstruction approaches. Figure 6.2 shows a sample result produced using only one

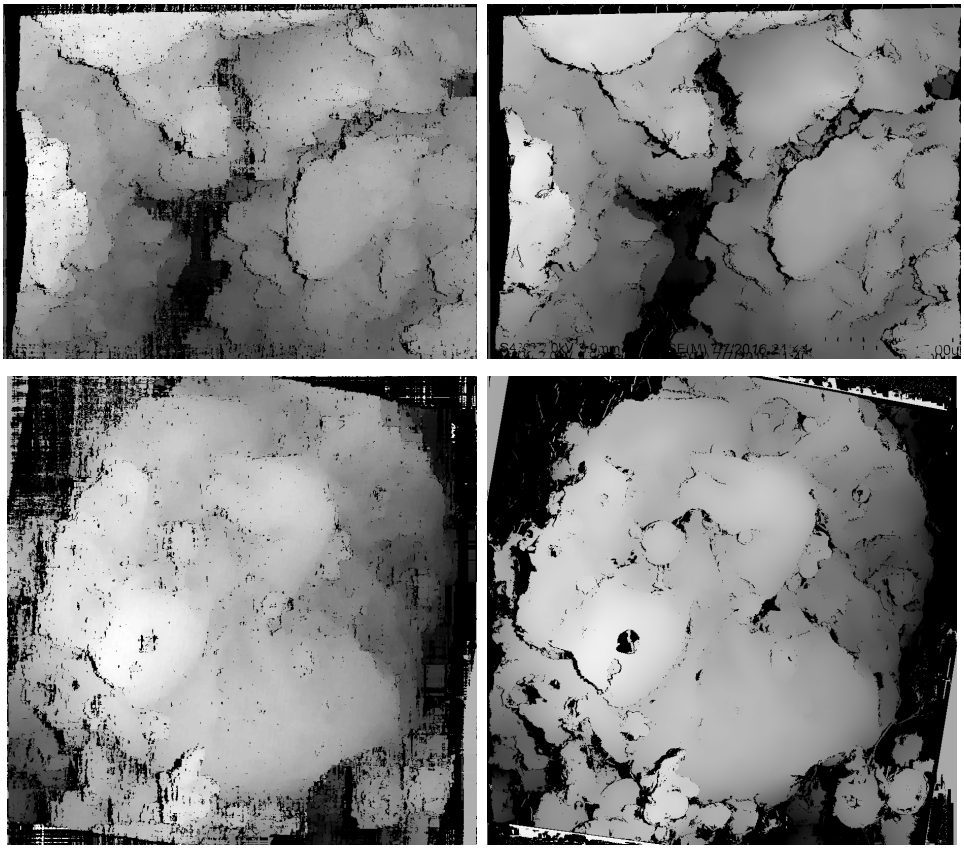


Figure 6.1: Stereo matching results for *Graphene* (first row) and *Ash* (second row) micrograph sets using Semi-Global stereo (left) and Patch-Match stereo (right) with occlusion detection.

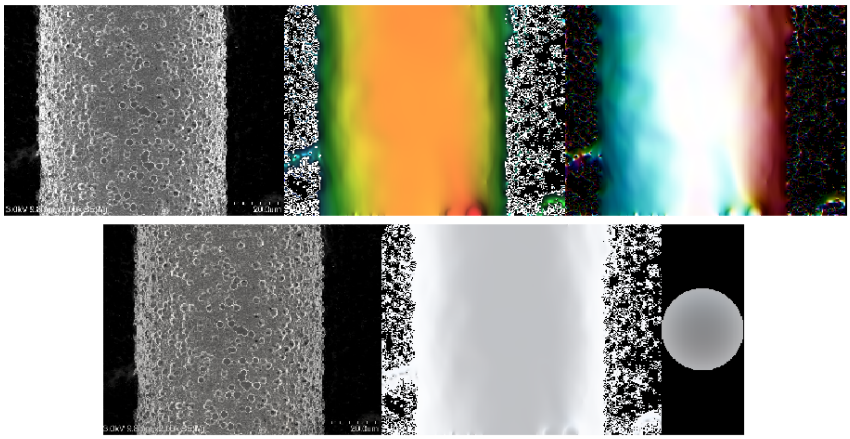


Figure 6.2: Shape, illumination, and reflectance estimation from shading using only one image from the *Copper Bar* set by using the method proposed in [22]. From left to right, the initial image as well as shape, normals, reflectance, shading and illumination.

image from the *Copper Bar* micrograph set by taking advantage of the work of Barron and Malik [22]. More rigorous analysis and investigations on the use of such techniques will benefit the field greatly.

Bibliography

[1] D. C. Adler, T. H. Ko, and J. G. Fujimoto. Speckle reduction in optical coherence tomography images by use of a spatially adaptive wavelet filter. *Optics letters*, 29(24):2878–2880, 2004.

[2] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. M. Seitz, and R. Szeliski. Building rome in a day. *Communications of the ACM*, 54(10):105–112, 2011.

[3] J.-J. Aguilar, F. Torres, and M. Lope. Stereo vision for 3d measurement: accuracy analysis, calibration and industrial applications. *Measurement*, 18(4):193–200, 1996.

[4] D. Alonso-Caneiro, S. A. Read, and M. J. Collins. Speckle reduction in optical coherence tomography imaging by affine-motion image registration. *Journal of biomedical optics*, 16(11):116027–1160275, 2011.

[5] N. S. Altman. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46:175–185, 1992.

[6] M. Anbarasu, M. Anandan, E. Chinnasamy, V. Gopinath, and K. Balamurugan. Synthesis and characterization of polyethylene glycol (peg) coated fe 3 o 4 nanoparticles by chemical co-precipitation method for biomedical applications. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 135:536–539, 2015.

[7] F. Atry, S. Frye, T. J. Richner, S. K. Brodnick, A. Soehartono, J. Williams, and R. Pashaie. Monitoring cerebral hemodynamics following optogenetic stimulation via optical coherence tomography. *IEEE Transactions on Biomedical Engineering*, 62(2):766–773, 2015.

[8] F. Atry and R. Pashaie. Analysis of intermediary scan-lens and tube-lens mechanisms for optical coherence tomography. *Applied optics*, 55(4):646–653, 2016.

[9] M. R. Avanaki, R. Cernat, P. J. Tadrous, T. Tatla, A. G. Podoleanu, and S. A. Hojjatoleslami. Spatial compounding algorithm for speckle reduction of dynamic focus oct images. *Photonics Technology Letters, IEEE*, 25(15):1439–1442, 2013.

[10] A. Baghaie. Markov random field model-based salt and pepper noise removal. *arXiv preprint arXiv:1609.06341*, 2016.

[11] A. Baghaie, R. M. DSouza, and Z. Yu. Dense correspondence and optical flow estimation using gabor, schmid and steerable descriptors. In *Advances in Visual Computing*, pages 406–415. Springer, 2015.

[12] A. Baghaie, R. M. D'souza, and Z. Yu. Sparse and low rank decomposition based batch image alignment for speckle reduction of retinal oct images. In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 226–230. IEEE, 2015.

[13] A. Baghaie, R. M. D'souza, and Z. Yu. Application of independent component analysis techniques in speckle noise reduction of retinal oct images. *Optik-International Journal for Light and Electron Optics*, 127(15):5783–5791, 2016.

[14] A. Baghaie, R. M. DSouza, and Z. Yu. Dense descriptors for optical flow estimation: A comparative study. *submitted to Journal of Imaging*, 2016.

[15] A. Baghaie and H. A. Moghaddam. Finite element method based modeling of cardiac deformation estimation under abnormal ventricular muscle conditions. *arXiv preprint arXiv:1412.4192*, 2014.

[16] A. Baghaie and Z. Yu. Curvature-based registration for slice interpolation of medical images. In *Computational Modeling of Objects Presented in Images. Fundamentals, Methods, and Applications*, pages 69–80. Springer, 2014.

[17] A. Baghaie and Z. Yu. Structure tensor based image interpolation method. *AEU-International Journal of Electronics and Communications*, 69(2):515–522, 2015.

[18] A. Baghaie, Z. Yu, and R. M. D'souza. Fast mesh-based medical image registration. In *Advances in Visual Computing*, pages 1–10. Springer, 2014.

[19] A. Baghaie, Z. Yu, and R. M. DSouza. State-of-the-art in retinal optical coherence tomography image analysis. *Quantitative imaging in medicine and surgery*, 5(4):603, 2015.

[20] A. Bakhshinejad, A. Baghaie, V. L. Rayz, and R. M. D'Souza. A proper orthogonal decomposition approach towards merging cfd and 4d-pcmr flow data. In *The 28th Society for Magnetic Resonance Angiography, Chicago, IL, USA*, 2016.

[21] D. Barber. *Bayesian reasoning and machine learning*. Cambridge University Press, 2012.

[22] J. T. Barron and J. Malik. Shape, illumination, and reflectance from shading. *IEEE transactions on pattern analysis and machine intelligence*, 37(8):1670–1687, 2015.

[23] A. Baumberg. Reliable feature matching across widely separated views. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, volume 1, pages 774–781. IEEE, 2000.

[24] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359, 2008.

[25] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. In *Computer Vision–ECCV 2006*, pages 404–417. Springer, 2006.

[26] J. Bernal, F. Vilarino, and J. Sánchez. Feature detectors and feature descriptors: Where we are now. *Universitat Autonoma de Barcelona, Barcelona*, 2010.

[27] E. Bettini, T. Eriksson, M. Boström, C. Leygraf, and J. Pan. Influence of metal carbides on dissolution behavior of biomedical cocrmo alloy: Sem, tem and afm studies. *Electrochimica Acta*, 56(25):9413–9419, 2011.

[28] M. J. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer vision and image understanding*, 63(1):75–104, 1996.

[29] M. Bleyer, C. Rhemann, and C. Rother. Patchmatch stereo-stereo matching with slanted support windows. In *BMVC*, volume 11, pages 1–11, 2011.

[30] J. J. Bozzola and L. D. Russel. *Electron Microscopy*. Jones and Bartlett, USA, 1992.

[31] J. J. Bozzola and L. D. Russell. *Electron microscopy: principles and techniques for biologists*. Jones & Bartlett Learning, 1999.

[32] M. Brown and D. G. Lowe. Invariant features from interest point groups. In *BMVC*, number s 1, 2002.

[33] D. Brownrigg. The weighted median filter. *Communications of the ACM*, 27(8):807–818, 1984.

[34] A. Bruhn, J. Weickert, and C. Schnörr. Lucas/kanade meets horn/schunck: Combining local and global optic flow methods. *International Journal of Computer Vision*, 61(3):211–231, 2005.

[35] A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 60–65. IEEE, 2005.

[36] M. Calonder, V. Lepetit, P. Fua, K. Konolige, J. Bowman, and P. Mihelich. Compact signatures for high-speed interest point description and matching. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 357–364. IEEE, 2009.

[37] M. Calonder, V. Lepetit, C. Strecha, and P. Fua. Brief: Binary robust independent elementary features. In *Computer Vision–ECCV 2010*, pages 778–792. Springer, 2010.

[38] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011.

[39] L. Carli, G. Genta, A. Cantatore, G. Barbato, L. De Chiffre, and R. Levi. Uncertainty evaluation for three-dimensional scanning electron microscope reconstructions based on the stereo-pair technique. *Measurement Science and Technology*, 22(3):035103, 2011.

[40] U. K. Chakraborty. *Advances in Differential Evolution*. Prentice Hall, USA, 2008.

[41] S. J. Chiu, J. A. Izatt, R. V. OConnell, K. P. Winter, C. A. Toth, and S. Farsiu. Validated automatic segmentation of amd pathology including drusen and geographic atrophy in sd-oct images. *Invest Ophthalmol Vis Sci*, 53(1):53–61, 2012.

[42] S. J. Chiu, X. T. Li, P. Nicholas, C. A. Toth, J. A. Izatt, and S. Farsiu. Automatic segmentation of seven retinal layers in sdoct images congruent with expert manual segmentation. *Optics express*, 18(18):19413–19428, 2010.

[43] M. Corvi and G. Nicchiotti. Multiresolution image registration. In *Image Processing, 1995. Proceedings., International Conference on*, volume 3, pages 224–227. IEEE, 1995.

[44] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.

[45] R. Danzl and S. Scherer. Integrating shape from shading and shape from stereo for variable reflectance surface reconstruction from sem images. In *SEM Images, 26th Workshop of the Austrian Association for Pattern Recognition*, 2003.

[46] D. C. DeBuc. *A review of algorithms for segmentation of retinal image data using optical coherence tomography*. INTECH Open Access Publisher, 2011.

[47] M. Desbrun, M. Meyer, P. Schröder, and A. H. Barr. Implicit fairing of irregular meshes using diffusion and curvature flow. In *Proceedings of the 26th annual*

conference on Computer graphics and interactive techniques, pages 317–324. ACM Press/Addison-Wesley Publishing Co., 1999.

[48] W. Drzazga, J. Paluszynski, and W. Slowko. Three-dimensional characterization of microstructures in a sem. *Measurement Science and Technology*, 17(1):28, 2005.

[49] R. M. D'Souza, A. Bakhshinejad, A. Baghaie, and V. L. Rayz. Reconstructing high fidelity hemodynamic flow fields by merging patient-specific computational fluid dynamics (cfd) and 4d phase contrast magnetic resonance data. In *ISMRM Workshop on Quantitative MR Flow: Innovation & Implementation for Clinical & Physiological Insights, San Francisco, CA, USA*, 2016.

[50] F. Durand and J. Dorsey. Fast bilateral filtering for the display of high-dynamic-range images. In *ACM transactions on graphics (TOG)*, volume 21, pages 257–266. ACM, 2002.

[51] R. F. Egerton. *Physical principles of electron microscopy: an introduction to TEM, SEM, and AEM*. Springer Science & Business Media, 2006.

[52] M. Eulitz and G. Reiss. 3d reconstruction of sem images by use of optical photogrammetry software. *Journal of structural biology*, 191(2):190–196, 2015.

[53] M. Everingham, A. Zisserman, C. Williams, and L. Van Gool. The pascal visual object classes challenge 2006 (voc 2006) results. 2006.

[54] L. Fang, S. Li, R. P. McNabb, Q. Nie, A. N. Kuo, C. Toth, J. Izatt, S. Farsiu, et al. Fast acquisition and reconstruction of optical coherence tomography images via sparse representation. *Medical Imaging, IEEE Transactions on*, 32(11):2034–2049, 2013.

[55] L. Fang, S. Li, Q. Nie, J. A. Izatt, C. A. Toth, and S. Farsiu. Sparsity based denoising of spectral domain optical coherence tomography images. *Biomedical optics express*, 3(5):927–942, 2012.

[56] O. D. Faugeras and F. Lustman. Motion and structure from motion in a piecewise planar environment. *International Journal of Pattern Recognition and Artificial Intelligence*, 2(03):485–508, 1988.

[57] D. C. Fernández, N. Villate, C. Puliafito, and P. Rosenfeld. Comparing total macular volume changes measured by optical coherence tomography with retinal lesion volume estimated by active contours. *Investigative Ophthalmology & Visual Science*, 45(13):3072–3072, 2004.

[58] B. Fischer and J. Modersitzki. A unified approach to fast image registration and a new curvature based registration technique. *Linear Algebra and its applications*, 380:107–124, 2004.

[59] M. A. Fischler and R. C. Bolles. Random sample consesus: a paradigm for model fitting with applications to image analysis and automated cartography. *in Communications of the ACM*, 1981.

[60] L. Florack, B. T. H. Romeny, J. J. Koenderink, and M. A. Viergever. General intensity transformations and differential invariants. *Journal of Mathematical Imaging and Vision*, 4(2):171–187, 1994.

[61] O. Fluck, C. Vetter, W. Wein, A. Kamen, B. Preim, and R. Westermann. A survey of medical image registration on graphics hardware. *Computer methods and programs in biomedicine*, 104(3):e45–e57, 2011.

[62] D. Fortun, P. Bouthemy, and C. Kervrann. Optical flow modeling and computation: a survey. *Computer Vision and Image Understanding*, 134:1–21, 2015.

[63] D. H. Frakes, L. P. Dasi, K. Pekkan, H. D. Kitajima, K. Sundareswaran, A. P. Yoganathan, and M. J. Smith. A new method for registration-based medical image interpolation. *Medical Imaging, IEEE Transactions on*, 27(3):370–377, 2008.

[64] W. T. Freeman and E. H. Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern analysis and machine intelligence*, 13(9):891–906, 1991.

[65] Y. Furukawa and J. Ponce. Accurate, dense, and robust multiview stereopsis. *IEEE transactions on pattern analysis and machine intelligence*, 32(8):1362–1376, 2010.

[66] A. Fusiello and L. Irsara. Quasi-euclidean uncalibrated epipolar rectification. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4. IEEE, 2008.

[67] D. Gabor. Theory of communication. part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, 93(26):429–441, 1946.

[68] M. K. Garvin, M. D. Abràmoff, R. Kardon, S. R. Russell, X. Wu, and M. Sonka. Intraretinal layer segmentation of macular optical coherence tomography images using optimal 3-d graph search. *Medical Imaging, IEEE Transactions on*, 27(10):1495–1505, 2008.

[69] M. Golabbakhsh and H. Rabbani. Vessel-based registration of fundus and optical coherence tomography projection images of retina using a quadratic registration model. *IET Image Processing*, 7(8):768–776, 2013.

[70] L. C. Gontard, R. Schierholz, S. Yu, J. Cintas, and R. E. Dunin-Borkowski. Photogrammetry of the three-dimensional shape and texture of a nanoscale particle using scanning electron microscopy and free software. *Ultramicroscopy*, 169:80–88, 2016.

[71] A. Goshtasby, D. Turner, L. V. Ackerman, et al. Matching of tomographic slices for interpolation. *Medical Imaging, IEEE Transactions on*, 11(4):507–516, 1992.

[72] G. J. Grevera and J. K. Udupa. Shape-based interpolation of multidimensional grey-level images. *Medical Imaging, IEEE Transactions on*, 15(6):881–892, 1996.

[73] G. J. Grevera and J. K. Udupa. An objective comparison of 3-d image interpolation methods. *Medical Imaging, IEEE Transactions on*, 17(4):642–652, 1998.

[74] G. J. Grevera, J. K. Udupa, and Y. Miki. A task-specific evaluation of three-dimensional image interpolation techniques. *Medical Imaging, IEEE Transactions on*, 18(2):137–143, 1999.

[75] Q. Guo, F. Dong, S. Sun, B. Lei, and B. Z. Gao. Image denoising algorithm based on contourlet transform for optical coherence tomography heart tube image. *Image Processing, IET*, 7(5):442–450, 2013.

[76] D. Gupta, R. Anand, and B. Tyagi. Ripplet domain non-linear filtering for speckle reduction in ultrasound medical images. *Biomedical Signal Processing and Control*, 10:79–91, 2014.

[77] E. Haber, S. Heldmann, and J. Modersitzki. Adaptive mesh refinement for nonparametric image registration. *SIAM journal on scientific computing*, 30(6):3012–3027, 2008.

[78] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey vision conference*, volume 15, page 50. Manchester, UK, 1988.

[79] R. Hartely and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge University Pressy, UK, 2004.

[80] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[81] R. I. Hartley. In defense of the eight-point algorithm. *IEEE Transactions on pattern analysis and machine intelligence*, 19(6):580–593, 1997.

[82] H. C. Hendargo, R. Estrada, S. J. Chiu, C. Tomasi, S. Farsiu, and J. A. Izatt. Automated non-rigid registration and mosaicing for robust imaging of distinct retinal capillary beds using speckle variance optical coherence tomography. *Biomedical optics express*, 4(6):803–821, 2013.

[83] W. E. Higgins, C. J. Orlick, and B. E. Ledell. Nonlinear filtering approach to 3-d

gray-scale image interpolation. *Medical Imaging, IEEE Transactions on*, 15(4):580–587, 1996.

[84] H. Hirschmuller. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on pattern analysis and machine intelligence*, 30(2):328–341, 2008.

[85] B. K. Horn and B. G. Schunck. Determining optical flow. In *1981 Technical symposium east*, pages 319–331. International Society for Optics and Photonics, 1981.

[86] G. Hua, M. Brown, and S. Winder. Discriminant embedding for local image descriptors. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, 2007.

[87] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, et al. Optical coherence tomography. *Science*, 254(5035):1178–1181, 1991.

[88] E. Jensen. Types of imaging, part 1: Electron microscopy. *The Anatomical Record*, 295(5):716–721, 2012.

[89] Z. Jian, L. Yu, B. Rao, B. J. Tromberg, and Z. Chen. Three-dimensional speckle suppression in optical coherence tomography based on the curvelet transform. *Optics express*, 18(2):1024–1032, 2010.

[90] Z. Jian, Z. Yu, L. Yu, B. Rao, Z. Chen, and B. J. Tromberg. Speckle attenuation in optical coherence tomography by curvelet shrinkage. *Optics letters*, 34(10):1516–1518, 2009.

[91] T. M. Jørgensen, J. Thomadsen, U. Christensen, W. Soliman, and B. Sander. Enhancing the signal-to-noise ratio in ophthalmic optical coherence tomography by image registration method and clinical examples. *Journal of biomedical optics*, 12(4):041208–041208, 2007.

[92] R. Kafieh, H. Rabbani, M. D. Abramoff, and M. Sonka. Intra-retinal layer segmentation of 3d optical coherence tomography using coarse grained diffusion map. *Medical image analysis*, 17(8):907–928, 2013.

[93] R. Kafieh, H. Rabbani, and S. Kermani. A review of algorithms for segmentation of optical coherence tomography from retina. *Journal of medical signals and sensors*, 3(1):45, 2013.

[94] Y. Ke and R. Sukthankar. Pca-sift: A more distinctive representation for local image descriptors. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–506. IEEE, 2004.

[95] M. E. Kilmer and D. P. O'Leary. Choosing regularization parameters in iterative methods for ill-posed problems. *SIAM Journal on matrix analysis and applications*, 22(4):1204–1221, 2001.

[96] K. W. Kim. Biomedical applications of stereoscopy for three-dimensional surface reconstruction in scanning electron microscopes. *Applied Microscopy*, 46(2):71–75, 2016.

[97] M. Kobayashi, H. Hanafusa, K. Takada, and J. Noda. Polarization-independent interferometric optical-time-domain reflectometer. *Lightwave Technology, Journal of*, 9(5):623–628, 1991.

[98] T. Kodama, X. Li, K. Nakahira, and D. Ito. Evolutionary computation applied to the reconstruction of 3-d surface topography in the sem. *Journal of electron microscopy*, 54(5):429–435, 2005.

[99] J. J. Koenderink and A. J. Van Doorn. Affine structure from motion. *JOSA A*, 8(2):377–385, 1991.

[100] M. F. Kraus, J. J. Liu, J. Schottenhamml, C.-L. Chen, A. Budai, L. Branchini, T. Ko, H. Ishikawa, G. Wollstein, J. Schuman, et al. Quantitative 3d-oct mo-

tion correction with tilt and illumination correction, robust similarity measure and regularization. *Biomedical optics express*, 5(8):2591–2613, 2014.

[101] M. F. Kraus, B. Potsaid, M. A. Mayer, R. Bock, B. Baumann, J. J. Liu, J. Hornegger, and J. G. Fujimoto. Motion correction in optical coherence tomography volumes on a per a-scan basis using orthogonal scan patterns. *Biomedical optics express*, 3(6):1182–1199, 2012.

[102] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger. Factor graphs and the sum-product algorithm. *Information Theory, IEEE Transactions on*, 47(2):498–519, 2001.

[103] K. M. Lee and C.-C. J. Kuo. Surface reconstruction from photometric stereo images. *JOSA A*, 10(5):855–868, 1993.

[104] T.-Y. Lee and C.-H. Lin. Feature-guided shape-based image interpolation. *Medical Imaging, IEEE Transactions on*, 21(12):1479–1489, 2002.

[105] T.-Y. Lee and W.-H. Wang. Morphology-based three-dimensional interpolation. *Medical Imaging, IEEE Transactions on*, 19(7):711–721, 2000.

[106] J. Leng, G. Xu, and Y. Zhang. Medical image interpolation based on multi-resolution registration. *Computers & Mathematics with Applications*, 66(1):1–18, 2013.

[107] V. Lepetit and P. Fua. Keypoint recognition using randomized trees. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(9):1465–1479, 2006.

[108] T. Leung and J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. *International journal of computer vision*, 43(1):29–44, 2001.

[109] Y. Li, G. Gregori, R. W. Knighton, B. J. Lujan, and P. J. Rosenfeld. Registration of oct fundus images with color fundus photographs based on blood vessel ridges. *Optics express*, 19(1):7–16, 2011.

[110] Y. Li, G. Gregori, B. L. Lam, and P. J. Rosenfeld. Automatic montage of sd-oct data sets. *Optics express*, 19(27):26239–26248, 2011.

[111] Y. Li and S. Osher. A new median formula with applications to pde based denoising. *Communications in Mathematical Sciences*, 7(3), 2009.

[112] S. Limandri, V. G. Josa, M. C. Valentinuzzi, M. E. Chena, and G. Castellano. 3d scanning electron microscopy applied to surface characterization of fluorosed dental enamel. *Micron*, 84:54–60, 2016.

[113] Z. Lin, M. Chen, and Y. Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*, 2010.

[114] T. Lindeberg. Scale-space theory: A basic tool for analyzing structures at different scales. *Journal of applied statistics*, 21(1-2):225–270, 1994.

[115] T. Lindeberg. Feature detection with automatic scale selection. *International journal of computer vision*, 30(2):79–116, 1998.

[116] C. Liu, J. Yuen, and A. Torralba. Sift flow: Dense correspondence across scenes and its applications. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(5):978–994, 2011.

[117] K. Liu, M. Xu, and Z. Yu. Feature-preserving image restoration from adaptive triangular meshes. In *Asian Conference on Computer Vision*, pages 31–46. Springer, 2014.

[118] K. Liu, G. Yao, and Z. Yu. Parallel acceleration for modeling of calcium dynamics in cardiac myocytes. *Bio-medical materials and engineering*, 24(1):1417–1424, 2014.

[119] D. G. Lowe. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. Ieee, 1999.

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

[121] F. Luan and Y. Wu. Application of rpca in optical coherence tomography for speckle noise reduction. *Laser Physics Letters*, 10(3):035603, 2013.

[122] K. L. Lurie, R. Angst, and A. K. Ellerbee. Automated mosaicing of feature-poor optical coherence tomography volumes with an integrated white light imaging system. *Biomedical Engineering, IEEE Transactions on*, 61(7):2141–2153, 2014.

[123] Z. Ma, K. He, Y. Wei, J. Sun, and E. Wu. Constant time weighted median filtering for stereo matching and beyond. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 49–56, 2013.

[124] F. Marinello, P. Bariani, E. Savio, A. Horsewell, and L. De Chiffre. Critical factors in sem 3d stereo microscopy. *Measurement Science and Technology*, 19(6):065705, 2008.

[125] D. Marr, T. Poggio, E. C. Hildreth, and W. E. L. Grimson. A computational theory of human stereo vision. In *From the Retina to the Neocortex*, pages 263–295. Springer, 1991.

[126] Y. Matsumoto and A. Zelinsky. An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement. In *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, pages 499–504. IEEE, 2000.

[127] MeshLab. Meshlab, 2005. http://meshlab.sourceforge.net/.

[128] K. Mikolajczyk and C. Schmid. Indexing based on scale invariant interest points. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 1, pages 525–531. IEEE, 2001.

[129] K. Mikolajczyk and C. Schmid. An affine invariant interest point detector. In *Computer VisionECCV 2002*, pages 128–142. Springer, 2002.

[130] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors.

Pattern Analysis and Machine Intelligence, IEEE Transactions on, 27(10):1615–1630, 2005.

[131] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *International journal of computer vision*, 65(1-2):43–72, 2005.

[132] F. Mindru, T. Tuytelaars, L. V. Gool, and T. Moons. Moment invariants for recognition under changing viewpoint and illumination. *Computer Vision and Image Understanding*, 94(1):3–27, 2004.

[133] A. Mishra, A. Wong, K. Bizheva, and D. A. Clausi. Intra-retinal layer segmentation in optical coherence tomography images. *Optics express*, 17(26):23719–23728, 2009.

[134] J. Modersitzki. *Numerical methods for image registration*. Oxford university press, 2003.

[135] J. Modersitzki. *FAIR: flexible algorithms for image registration*, volume 6. SIAM, 2009.

[136] L. Moisan, P. Moulon, and P. Monasse. Fundamental matrix of a stereo pair, with a contrario elimination of outliers. *Image Processing On Line*, 6:89–113, 2016.

[137] L. Moisan and B. Stival. A probabilistic criterion to detect rigid point matches between two images and estimate the fundamental matrix. *International Journal of Computer Vision*, 57(3):201–218, 2004.

[138] P. Monasse. Quasi-euclidean epipolar rectification. *Image Processing On Line*, 1, 2011.

[139] K. P. Murphy, Y. Weiss, and M. I. Jordan. Loopy belief propagation for approximate inference: An empirical study. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pages 467–475. Morgan Kaufmann Publishers Inc., 1999.

[140] P. Musialski, P. Wonka, D. G. Aliaga, M. Wimmer, L. Gool, and W. Purgathofer. A survey of urban reconstruction. In *Computer graphics forum*, volume 32, pages 146–177. Wiley Online Library, 2013.

[141] J. Nocedal and S. Wright. *Numerical optimization*. Springer Science & Business Media, 2006.

[142] M. Ozuysal, M. Calonder, V. Lepetit, and P. Fua. Fast keypoint recognition using random ferns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(3):448–461, 2010.

[143] J. Paluszyński and W. Slowko. Surface reconstruction with the photometric method in sem. *Vacuum*, 78(2):533–537, 2005.

[144] S. Paris and F. Durand. A fast approximation of the bilateral filter using a signal processing approach. *International journal of computer vision*, 81(1):24–52, 2009.

[145] S. Paris, P. Kornprobst, J. Tumblin, and F. Durand. *Bilateral filtering: Theory and applications*. Now Publishers Inc, 2009.

[146] K. D. Parry-Vernon. Scanning electron microscopy: an introduction. *III-Vs Review*, 13:40–44, 2000.

[147] N. Patton, T. M. Aslam, T. MacGillivray, I. J. Deary, B. Dhillon, R. H. Eikelboom, K. Yogesan, and I. J. Constable. Retinal image analysis: concepts, applications and potential. *Progress in retinal and eye research*, 25(1):99–127, 2006.

[148] J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 2014.

[149] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma. Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(11):2233–2246, 2012.

[150] G. P. Penney, J. A. Schnabel, D. Rueckert, M. A. Viergever, and W. J. Niessen. Registration-based interpolation. *Medical Imaging, IEEE Transactions on*, 23(7):922–926, 2004.

[151] G. P. Penney, J. Weese, J. Little, P. Desmedt, D. L. Hill, D. J. Hawkes, et al. A comparison of similarity measures for use in 2-d-3-d medical image registration. *Medical Imaging, IEEE Transactions on*, 17(4):586–595, 1998.

[152] T. Q. Pham and L. J. Van Vliet. Separable bilateral filtering for fast video preprocessing. In *2005 IEEE International Conference on Multimedia and Expo*, pages 4–pp. IEEE, 2005.

[153] R. Pintus, S. Podda, and M. Vanzi. An automatic alignment procedure for a four-source photometric stereo technique applied to scanning electron microscopy. *IEEE Transactions on Instrumentation and Measurement*, 57(5):989–996, 2008.

[154] M. Pircher, B. Baumann, H. Sattmann, C. K. Hitzenberger, et al. Simultaneous slo/oct imaging of the human retina with axial eye motion correction. *Optics express*, 15(25):16922–16932, 2007.

[155] M. Pircher, E. Go, R. Leitgeb, A. F. Fercher, C. K. Hitzenberger, et al. Speckle reduction in optical coherence tomography by frequency compounding. *Journal of Biomedical Optics*, 8(3):565–569, 2003.

[156] A. Pizurica, L. Jovanov, B. Huysmans, V. Zlokolica, P. De Keyser, F. Dhaenens, and W. Philips. Multiresolution denoising for optical coherence tomography: a review and evaluation. *Current Medical Imaging Reviews*, 4(4):270–284, 2008.

[157] K. Popuri, D. Cobzas, and M. Jagersand. Fast fem-based non-rigid registration. In *Computer and Robot Vision (CRV), 2010 Canadian Conference on*, pages 378–385. IEEE, 2010.

[158] P. Puvanathasan and K. Bizheva. Interval type-ii fuzzy anisotropic diffusion algorithm for speckle noise reduction in optical coherence tomography images. *Optics express*, 17(2):733–746, 2009.

[159] M. Raspanti, E. Binaghi, I. Gallo, and A. Manelli. A vision-based, 3d reconstruction technique for scanning electron microscopy: Direct comparison with atomic force microscopy. *Microscopy research and technique*, 67(1):1–7, 2005.

[160] S. Ricco, M. Chen, H. Ishikawa, G. Wollstein, and J. Schuman. Correcting motion artifacts in retinal spectral domain optical coherence tomography via image registration. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009*, pages 100–107. Springer, 2009.

[161] P. L. Rosin. Measuring corner properties. *Computer Vision and Image Understanding*, 73(2):291–307, 1999.

[162] E. Rosten and T. Drummond. Machine learning for high-speed corner detection. In *Computer Vision–ECCV 2006*, pages 430–443. Springer, 2006.

[163] S. Roy, J. Meunier, A. Marian, F. Vidal, I. Brunette, and S. Costantino. Automatic 3d reconstruction of quasi-planar stereo scanning electron microscopy (sem) images. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4361–4364. IEEE, 2012.

[164] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. Orb: an efficient alternative to sift or surf. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2564–2571. IEEE, 2011.

[165] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformations: application to breast mr images. *Medical Imaging, IEEE Transactions on*, 18(8):712–721, 1999.

[166] H. M. Salinas and D. C. Fernández. Comparison of pde-based nonlinear diffusion approaches for image enhancement and denoising in optical coherence tomography. *Medical Imaging, IEEE Transactions on*, 26(6):761–771, 2007.

[167] D. Samak, A. Fischer, and D. Rittel. 3d reconstruction and visualization of microstructure surfaces from 2d images. *CIRP Annals-Manufacturing Technology*, 56(1):149–152, 2007.

[168] P. Sand and S. Teller. Particle video: Long-range motion estimation using point trajectories. *International Journal of Computer Vision*, 80(1):72–91, 2008.

[169] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International journal of computer vision*, 47(1-3):7–42, 2002.

[170] C. Schmid. Constructing models for content-based image retrieval. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 2, pages II–39. IEEE, 2001.

[171] R. Schmidt and K. Singh. Meshmixer: an interface for rapid mesh composition. In *ACM SIGGRAPH 2010 Talks*, page 6. ACM, 2010.

[172] J. Schmitt. Array detection for speckle reduction in optical coherence microscopy. *Physics in Medicine and Biology*, 42(7):1427, 1997.

[173] J. M. Schmitt, S. Xiang, and K. M. Yung. Speckle in optical coherence tomography. *Journal of biomedical optics*, 4(1):95–105, 1999.

[174] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri. Nih image to imagej: 25 years of image analysis. *Nature methods*, 9(7):671–675, 2012.

[175] G. Shakhnarovich. *Learning task-specific similarity*. PhD thesis, Massachusetts Institute of Technology, 2005.

[176] S. Shen. Accurate multiple view 3d reconstruction using patch-based stereo for large-scale scenes. *IEEE transactions on image processing*, 22(5):1901–1914, 2013.

[177] W. Slowko and M. Krysztof. Detector system for three-dimensional imaging in the variable pressure/environmental sem. 2013.

[178] A. Sotiras, C. Davatzikos, and N. Paragios. Deformable medical image registration: A survey. *Medical Imaging, IEEE Transactions on*, 32(7):1153–1190, 2013.

[179] P. P. Srinivasan, S. J. Heflin, J. A. Izatt, V. Y. Arshavsky, and S. Farsiu. Automatic segmentation of up to ten layer boundaries in sd-oct images of the mouse retina with and without missing layers due to pathology. *Biomedical optics express*, 5(2):348–365, 2014.

[180] E. Strelcov, J. Cothren, D. Leonard, A. Y. Borisevich, and A. Kolmakov. In situ sem study of lithium intercalation in individual v 2 o 5 nanowires. *Nanoscale*, 7(7):3022–3027, 2015.

[181] D. Sun, S. Roth, and M. J. Black. Secrets of optical flow estimation and their principles. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2432–2439. IEEE, 2010.

[182] D. Sun, S. Roth, and M. J. Black. A quantitative analysis of current practices in optical flow estimation and the principles behind them. *International Journal of Computer Vision*, 106(2):115–137, 2014.

[183] R. Szeliski. *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010.

[184] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother. A comparative study of energy minimization methods for markov random fields with smoothness-based priors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(6):1068–1080, 2008.

[185] A. P. Tafti. 3d sem surface reconstruction: An optimized, adaptive, and intelligent approach. 2016.

[186] A. P. Tafti, A. Baghaie, A. B. Kirkpatrick, J. D. Holz, H. A. Owen, R. M. DSouza, and Z. Yu. A comparative study on the application of sift, surf, brief and orb for 3d surface reconstruction of electron microscopy images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, pages 1–14, 2016.

[187] A. P. Tafti, J. D. Holz, A. Baghaie, H. A. Owen, M. M. He, and Z. Yu. 3dsem++: Adaptive and intelligent 3d sem surface reconstruction. *Micron*, 87:33–45, 2016.

[188] A. P. Tafti, A. B. Kirkpatrick, Z. Alavi, H. A. Owen, and Z. Yu. Recent advances in 3d sem surface reconstruction. *Micron*, 78:54–66, 2015.

[189] A. P. Tafti, A. B. Kirkpatrick, J. D. Holz, H. A. Owen, and Z. Yu. 3dsem: A 3d microscopy dataset. *Data in Brief*, 6:112–116, 2016.

[190] E. Tola, V. Lepetit, and P. Fua. Daisy: An efficient dense descriptor applied to wide-baseline stereo. *IEEE transactions on pattern analysis and machine intelligence*, 32(5):815–830, 2010.

[191] P. H. Tomlins and R. Wang. Theory, developments and applications of optical coherence tomography. *Journal of Physics D: Applied Physics*, 38(15):2519, 2005.

[192] A. Torralba, R. Fergus, and Y. Weiss. Small codes and large image databases for recognition. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[193] T. Tuytelaars and C. Schmid. Vector quantizing feature space with a regular lattice. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, 2007.

[194] T. Vynnyk, T. Schultheis, T. Fahlbusch, and E. Reithmeier. 3d-measurement with the stereo scanning electron microscope on sub-micrometer structures. *Journal of the European Optical Society-Rapid publications*, 5, 2010.

[195] T. Wang, Q. Hu, M. Zhou, J. Xue, and Y. Luo. Preparation of ultra-fine powders from polysaccharide-coated solid lipid nanoparticles and nanostructured lipid carriers by innovative nano spray drying technology. *International Journal of Pharmaceutics*, 511(1):219–222, 2016.

[196] D. C. Ward, D. H. Ubelaker, J. Stewart, and V. S. Braz. The use of sem/eds

analysis to distinguish dental and osseus tissue from other materials. *Journal of Forensic Science*, 47(5):1–4, 2002.

[197] B. Weiss. Fast median and bilateral filtering. *Acm Transactions on Graphics (TOG)*, 25(3):519–526, 2006.

[198] S. Winder, G. Hua, and M. Brown. Picking the best daisy. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 178–185. IEEE, 2009.

[199] A. P. Witkin. Scale-space filtering: A new approach to multi-scale description. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'84.*, volume 9, pages 150–153. IEEE, 1984.

[200] C. Wöhler. *3D computer vision: efficient methods and applications*. Springer Science & Business Media, 2012.

[201] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical engineering*, 19(1):191139–191139, 1980.

[202] J. Xiao, H. Cheng, H. Sawhney, C. Rao, and M. Isnardi. Bilateral filtering-based optical flow estimation with occlusion detection. In *European Conference on Computer Vision*, pages 211–224. Springer, 2006.

[203] J. Xie. Stereomicroscopy: 3d imaging and the third dimension measurement. *Agilent Technologies, Santa Clara, CA.*, 2011.

[204] G. Xu. Convergent discrete laplace-beltrami operators over triangular surfaces. In *Geometric Modeling and Processing, 2004. Proceedings*, pages 195–204. IEEE, 2004.

[205] J. Xu, H. Ishikawa, G. Wollstein, L. Kagemann, and J. S. Schuman. Alignment of 3-d optical coherence tomography scans to correct eye movement using a particle filtering. *Medical Imaging, IEEE Transactions on*, 31(7):1337–1345, 2012.

[206] J. Xu, H. Ou, E. Y. Lam, P. Chui, and K. K. Wong. Speckle reduction of retinal optical coherence tomography based on contourlet shrinkage. *Optics letters*, 38(15):2900–2903, 2013.

[207] L. Xu, J. Jia, and Y. Matsushita. Motion detail preserving optical flow estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(9):1744–1757, 2012.

[208] M. Xu, Z. Gao, and Z. Yu. Feature-sensitive and adaptive mesh generation of grayscale images. In *Computational Modeling of Objects Presented in Images. Fundamentals, Methods, and Applications*, pages 204–215. Springer, 2014.

[209] M. Xu, J. Wang, and Z. Yu. Image edge enhancement and segmentation via randomized shortest paths. In *Biomedical Engineering and Informatics (BMEI), 2012 5th International Conference on*, pages 290–294. IEEE, 2012.

[210] M. Xu and Z. Yu. 3d image segmentation based on feature-sensitive and adaptive tetrahedral meshes. In *Image Processing (ICIP), 2016 IEEE International Conference on*, pages 854–858. IEEE, 2016.

[211] Q. Yang, L. Wang, R. Yang, H. Stewénius, and D. Nistér. Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(3):492–504, 2009.

[212] A. Yazdanpanah, G. Hamarneh, B. R. Smith, and M. V. Sarunic. Segmentation of intra-retinal layers from optical coherence tomography images using an active contour approach. *Medical Imaging, IEEE Transactions on*, 30(2):484–496, 2011.

[213] K.-J. Yoon and I. S. Kweon. Adaptive support-weight approach for correspondence search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4):650–656, 2006.

[214] Z. Yu, J. Wang, Z. Gao, M. Xu, and M. Hoshijima. New software developments for quality mesh generation and optimization from biomedical imaging data. *Computer methods and programs in biomedicine*, 113(1):226–240, 2014.

[215] Z. Yu, M. Xu, and Z. Gao. Biomedical image segmentation via constrained graph cuts and pre-segmentation. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5714–5717. IEEE, 2011.

[216] C. Zhang, Z. Li, Y. Cheng, R. Cai, H. Chao, and Y. Rui. Meshstereo: A global stereo model with mesh alignment regularization for view interpolation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2057–2065, 2015.

[217] Q. Zhang, L. Xu, and J. Jia. 100+ times faster weighted median filter (wmf). In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2830–2837, 2014.

[218] W. Zhou and Z. L. Wang. *Scanning microscopy for nanotechnology: techniques and applications*. Springer science & business media, 2007.

[219] B. Zitova and J. Flusser. Image registration methods: a survey. *Image and vision computing*, 21(11):977–1000, 2003.

[220] A. Zolotukhin, I. Safonov, and K. Kryzhanovskii. 3d reconstruction for a scanning electron microscope. *Pattern recognition and image analysis*, 23(1):168–174, 2013.

Appendices

Appendix A: Publications

Journal Papers

- Tafti, Ahmad P., Jessica D. Holz, **Ahmadreza Baghaie**, Heather A. Owen, Max M. He, Zeyun Yu, "3DSEM++: adaptive and intelligent 3D SEM surface reconstruction" *Micron*, 2016. doi: 10.1016/ j.micron.2016.05.004

- **Baghaie, Ahmadreza**, Roshan M. D'souza, and Zeyun Yu. "Application of Independent Component Analysis Techniques in Speckle Noise Reduction of Retinal OCT Images", *Optik - International Journal for Light and Electron Optics*, Volume 127, Issue 15, August 2016, Pages 5783-5791. doi:10.1016/j.ijleo.2016.03.078

- Tafti, Ahmad P., **Ahmadreza Baghaie**, A.B. Kirkpatrick, J.D. Holz, H.A.Owen, Zeyun Yu, and Roshan M. D'souza. "A Comparative Study on the Application of SIFT, SURF, BRIEF and ORB for 3D Surface Reconstruction of Electron Microscopy Images" *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 2016. doi: 10.1080/21681163.2016.1152201

- **Baghaie, Ahmadreza**, Roshan M. D'souza, and Zeyun Yu. "State-of-the-Art in Retinal Optical Coherence Tomography Image Analysis." *Quantitative Imaging in Medicine and Surgery* 2015;5(4):603-617. doi: 10.3978/j.issn.2223-4292.2015.07.02.

- **Baghaie, Ahmadreza**, and Zeyun Yu. "Structure tensor based image interpolation method." *AEU-International Journal of Electronics and Communications*, 69.2 (2015): 515-522. doi:10.1016/j.aeue.2014.10.022

Conference Papers

- **Baghaie, Ahmadreza**, Roshan M. DSouza, and Zeyun Yu. "Dense Correspondence and Optical Flow Estimation Using Gabor, Schmid and Steerable Descrip-

tors." *Advances in Visual Computing.* Springer International Publishing, 2015. 406-415.

- **Baghaie, Ahmadreza**, Roshan M. D'souza, and Zeyun Yu. "Sparse And Low Rank Decomposition Based Batch Image Alignment for Speckle Reduction of retinal OCT Images." *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on* , vol., no., pp.226,230, 16-19 April 2015, doi: 10.1109/ISBI.2015.7163855

- **Baghaie, Ahmadreza**, Zeyun Yu, and Roshan M. Dsouza. "Fast Mesh-Based Medical Image Registration." *Advances in Visual Computing.* Springer International Publishing, 2014. 1-10.

- **Baghaie, Ahmadreza**, and Zeyun Yu. "Curvature-Based Registration for Slice Interpolation of Medical Images." *Computational Modeling of Objects Presented in Images. Fundamentals, Methods, and Applications.* Springer International Publishing, 2014. 69-80.

Under Review/Under Preparation

- **Baghaie, Ahmadreza**, Chi Zhang, Ali Bakhshinejad, Heather A. Owen, Hongyang Chao, Roshan M. D'Souza, Zeyun Yu, "Experiments with Classic and Modern Stereo Matching Approaches for 3D Surface Reconstruction of Microscopic Samples", *Under Preparation*

- **Baghaie, Ahmadreza**, Ahmad P. Tafti, Heather A. Owen, Roshan M. D'Souza, Zeyun Yu, "Three-Dimensional Reconstruction of Highly Complex Microscopic Samples Using Scanning Electron Microscopy and Optical Flow Estimation", *Under Review*

- **Baghaie, Ahmadreza**, Ahmad P. Tafti, Heather A. Owen, Roshan M. D'Souza, Zeyun Yu, "SD-SEM: Sparse-Dense Correspondence for 3D Reconstruction of Microscopic Samples", *Under Review*

- **Baghaie, Ahmadreza**, Zeyun Yu, and Roshan M. D'souza. "Involuntary Eye Motion Correction in Retinal Optical Coherence Tomography: Hardware or Software Solution?" *Under Review*

- **Baghaie, Ahmadreza**, Zeyun Yu, and Roshan M. D'souza. "Dense Descriptors for Optical Flow Estimation: A Comparative Study" *Under Review*

Appendix B: Honors and Awards

- **Chancellor's Award**, Electrical Engineering Department, University of Wisconsin-Milwaukee, Fall 2016
- **Chancellor's Award**, Computer Science Department, University of Wisconsin-Milwaukee, Fall 2016
- **3rd Place** with A. P. Tafti, Larry Hause Student Poster Competition, IEEE, 2016
- **2nd Place** with A. Bakhshinejad, Larry Hause Student Poster Competition, IEEE, 2016
- **1st Place** with A. Bakhshinejad, Research Poster Competition, Graduate Students Section, College of Engineering and Applied Science, UWM, 2016
- Major Contribution in Funded Grant Proposal: **GE Healthcare Catalyst Grant**, 2015
- Major Contribution in Funded Grant Proposal: **CTSI Pilot and Collaborative Clinical and Translational Research Grant**, 2015
- **Travel Grant** for 11'th International Symposium on Visual Computing (ISVC), Las Vegas, NV, 2015
- **Travel Grant** for International Symposium on Biomedical Imaging 2015 (ISBI), New York, NY, 2015
- **Travel Grant** for 10'th International Symposium on Visual Computing (ISVC), Las Vegas, NV, 2014
- **Travel Grant** for 4'th Computational Modeling of Objects Presented in Images: Fundamentals, Methods and Applications (CompIMAGE), Pittsburgh, PA, 2014
- **3rd Place**, Research Poster Competition, Graduate Students Section, College of Engineering and Applied Science, UWM, 2013

Curriculum Vitae

Ahmadreza Baghaie

Education

- **University of Wisconsin-Milwaukee (UWM)** — **Milwaukee, WI**
  PhD in Electrical Engineering , GPA: 4/4 — *2012–2016*
  Image Processing, Computer Vision, Pattern Recognition
- **K.N.Toosi University of Technology** — **Tehran, Iran**
  M.Sc in Biomedical Engineering, GPA: 16.39/20 — *2006–2009*
  Presentation of Dynamic Model of the Left Ventricle's Wall Based on Active Mesh Method Using MRI Image Series.
- **Isfahan University** — **Isfahan, Iran**
  B.Sc in Electrical Engineering, GPA: 16.10/20 — *2002–2006*
  Design and Implementation of Intermediate Frequency (IF) Block.

Research Interests and Relevant Courses

- Signal and Image Processing
- Sparse Representation and Compressive Sensing
- Graphical Models
- Statistical Pattern Recognition
- Biomedical Imaging
- Numerical Analysis
- Scientific Computing
- Biomedical Instrumentation

Honors and Awards

- **Chancellor's Award**, Electrical Engineering Department, UWM, Fall 2016
- **Chancellor's Award**, Computer Science Department, UWM, Fall 2016
- **3rd Place** with A. P. Tafti, Larry Hause Student Poster Competition, IEEE, 2016
- **2nd Place** with A. Bakhshinejad, Larry Hause Student Poster Competition, IEEE, 2016
- **1st Place** with A. Bakhshinejad, Research Poster Competition, Graduate Students Section, College of Engineering and Applied Science, UWM, 2016
- Major Contribution in Funded Grant Proposal: **GE Healthcare Catalyst Grant**, 2015
- Major Contribution in Funded Grant Proposal: **CTSI Pilot and Collaborative Clinical and Translational Research Grant**, 2015
- **Travel Grant** for 11t́h International Symposium on Visual Computing (ISVC), Las Vegas, NV, 2015
- **Travel Grant** for International Symposium on Biomedical Imaging 2015 (ISBI), New York, NY, 2015
- **Travel Grant** for 10t́h International Symposium on Visual Computing (ISVC), Las Vegas, NV, 2014
- **Travel Grant** for 4t́h Computational Modeling of Objects Presented in Images: Fundamentals, Methods and Applications (CompIMAGE), Pittsburgh, PA, 2014
- **3rd Place**, Research Poster Competition, Graduate Students Section, College of Engineering and Applied Science, UWM, 2013
- **Ranked Among 5%** of Participants of the National Entrance Examination for Master Degree, 2006
- **Semifinalist**, National Mathematics and Electrical Engineering Olympiad, Isfahan, Iran 2005
- **Ranked Among 1%** of Participants of the National Entrance Examination for Bachelor Degree, Iran, 2002

Research and Teaching Experiences

Teaching

- **Isfahan University, Isfahan, Iran**
  Teaching Assistant for Electronics II-Lab, Communication Circuits-Lab and Pulse Technique-Lab, 2004-2006
- **Azad University, Tehran Shargh, Tehran, Iran**
  Instructor for Ordering Circuits-Lab, 2009-2010
- **University of Wisconsin-Milwaukee, Milwaukee, WI**
  Teaching Assistant for Electronics II-Lab, 2013-2015

Research

- **Biomedical Engineering Lab, K.N. Toosi University of Technology, Tehran, Iran**
  Research Assistant Deformation Model of Cardiovascular Muscle Based On 4D MRI Images, 2007-2009
- **K.N. Toosi University of Technology, Tehran, Iran**
  Research Assistant Web-Pages Classification Based on Textual and Visual Features, Sept 2008- Feb 2009
- **Advanced Ocular Imaging Program (AOIP), Medical College of Wisconsin, Milwaukee, WI**
  Research Assistant OCT speckle noise reduction, OCT eye motion correction, OCT volume registration,2014-2016
- **University of Wisconsin-Milwaukee, Milwaukee, WI**
  Research Assistant Image super-resolution, slice interpolation, OCT speckle noise reduction, OCT volume registration, 3D surface reconstruction, 4D PC-MRI Analysis, 2012-present

Scientific Reviewer

- **Kentucky Science and Engineering Foundation** (Grant)
- **IEEE Transactions on Biomedical Engineering**, IEEE (Journal)
- **Biomedical Signal Processing and Control**, Elsevier (Journal)
- **Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization**, Taylor & Francis (Journal)
- **International Journal of Computer Vision & Signal Processing** (Journal)
- **International Journal of Computer Vision and Image Processing**, (Journal)
- **International Symposium on Biomedical Imaging (ISBI) 2017**, IEEE (Conference)
- **International Symposium on Biomedical Imaging (ISBI) 2016**, IEEE (Conference)
- **International Conference on Informatics, Electronics & Vision (ICIEV) 2016**, IEEE (Conference)
- **International Symposium on Biomedical Imaging (ISBI) 2015**, IEEE (Conference)

Publications

Journal Papers

1. Tafti, Ahmad P., Jessica D. Holz, **Ahmadreza Baghaie**, Heather A. Owen, Max M. He, Zeyun Yu, "3DSEM++: adaptive and intelligent 3D SEM surface reconstruction" *Micron*, 2016. doi: 10.1016/j.micron.2016.05.004
2. **Baghaie, Ahmadreza**, Roshan M. D'souza, and Zeyun Yu. "Application of Independent Component Analysis Techniques in Speckle Noise Reduction of Retinal OCT Images", *Optik - International Journal for Light and Electron Optics*, Volume 127, Issue 15, August 2016, Pages 5783-5791. doi:10.1016/j.ijleo.2016.03.078
3. Tafti, Ahmad P., **Ahmadreza Baghaie**, A.B. Kirkpatrick, J.D. Holz, H.A.Owen, Zeyun Yu, and Roshan M. D'souza. "A Comparative Study on the Application of SIFT, SURF, BRIEF and ORB for 3D Surface Reconstruction of Electron Microscopy Images" *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 2016. doi: 10.1080/21681163.2016.1152201
4. **Baghaie, Ahmadreza**, Roshan M. D'souza, and Zeyun Yu. "State-of-the-Art in Retinal Optical Coherence Tomography Image Analysis." *Quantitative Imaging in Medicine and Surgery* 2015;5(4):603-617. doi:

10.3978/j.issn.2223-4292.2015.07.02.

5. **Baghaie, Ahmadreza**, and Zeyun Yu. "Structure tensor based image interpolation method." *AEU-International Journal of Electronics and Communications*, 69.2 (2015): 515-522. doi:10.1016/j.aeue.2014.10.022

Conference Papers

1. D'Souza, Roshan M., Ali Bakhshinejad, **Ahmadreza Baghaie**, Vitaliy L. Rayz, "Reconstructing High Fidelity Hemodynamic Flow Fields by Merging Patient-Specific Computational Fluid Dynamics (CFD) and 4D Phase Contrast Magnetic Resonance Data", *ISMRM Workshop on Quantitative MR Flow: Innovation & Implementation for Clinical & Physiological Insights*, San Francisco, CA, USA 2016
2. Tafti, Ahmad P., **Ahmadreza Baghaie**, Mehdi Assefi, Anne Nikolai, Zeyun Yu, Hamid R. Arabnia, and Peggy Peissig, "OCR as a Service: an experimental evaluation of Google Docs OCR, Tesseract, ABBYY FineReader, and Transym", *Accepted for Presentation at ISVC'16*
3. Bakhshinejad, Ali, **Ahmadreza Baghaie**, Vitaliy L. Rayz, Roshan M. D'souza, "A proper orthogonal decomposition approach towards merging computational fluid dynamic and time resolved phase contrast magnetic resonance flow data", *28th SMRA Annual International Conference*, 2016
4. **Baghaie, Ahmadreza**, Roshan M. D'Souza, and Zeyun Yu. "Dense Correspondence and Optical Flow Estimation Using Gabor, Schmid and Steerable Descriptors." *Advances in Visual Computing*. Springer International Publishing, 2015. 406-415.
5. **Baghaie, Ahmadreza**, Roshan M. D'souza, and Zeyun Yu. "Sparse And Low Rank Decomposition Based Batch Image Alignment for Speckle Reduction of retinal OCT Images." *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on* , vol., no., pp.226,230, 16-19 April 2015, doi: 10.1109/ISBI.2015.7163855
6. **Baghaie, Ahmadreza**, Zeyun Yu, and Roshan M. D'souza. "Fast Mesh-Based Medical Image Registration." *Advances in Visual Computing*. Springer International Publishing, 2014. 1-10.
7. **Baghaie, Ahmadreza**, and Zeyun Yu. "Curvature-Based Registration for Slice Interpolation of Medical Images." *Computational Modeling of Objects Presented in Images. Fundamentals, Methods, and Applications*. Springer International Publishing, 2014. 69-80.
8. **Baghaie, Ahmadreza**, and H. Abrishami Moghaddam. "A consistent model for cardiac deformation estimation under abnormal ventricular muscle conditions." *World Congress on Medical Physics and Biomedical Engineering*, September 7-12, 2009, Munich, Germany. Springer Berlin Heidelberg, 2010.
9. **Baghaie, Ahmadreza**, H.A. Moghaddam, "Finite Element Method Based Cardiac Infarction Localization", *18th Iranian Conference on Electrical Engineering*, Tehran, Iran,2009 (In Farsi)
10. **Baghaie, Ahmadreza**, H.A. Moghaddam, "General Movement Model Of Left Ventricular Muscle Using Cardiac MRI ", *17th Iranian Conference on Electrical Engineering*, Tehran, Iran ,2009 (In Farsi)
11. **Baghaie, Ahmadreza**, M. Forouzanfar and H.A. Moghaddam, "Automatic diagnosis of diabetes using data dependent kernel ", *15th Conference on Biomedical Engineering*, Mashad Iran , 2008, (In Farsi)

Under Review/Under Preparation

1. **Baghaie, Ahmadreza**, Chi Zhang, Ali Bakhshinejad, Heather A. Owen, Hongyang Chao, Roshan M. D'Souza, Zeyun Yu, "Experiments with Classic and Modern Stereo Matching Approaches for 3D Surface Reconstruction of Microscopic Samples", *Under Preparation*
2. **Baghaie, Ahmadreza**, Ahmad P. Tafti, Heather A. Owen, Roshan M. D'Souza, Zeyun Yu, "Three-Dimensional Reconstruction of Highly Complex Microscopic Samples Using Scanning Electron Microscopy and Optical Flow Estimation", *Under Review*
3. **Baghaie, Ahmadreza**, Ahmad P. Tafti, Heather A. Owen, Roshan M. D'Souza, Zeyun Yu, "SD-SEM: Sparse-Dense Correspondence for 3D Reconstruction of Microscopic Samples", *Under Review*

4. Bakhshinejad, Ali, **Ahmadreza Baghaie**, Alireza Vali, Vitaliy L. Rayz, Roshan M. D'souza, "Reconstructing High Resolution Hemodynamic Velocity Profiles from Low Resolution and Noisy 4D Flow Data Using Proper Orthogonal Decomposition and Ridge Regression", *Under Review*
5. **Baghaie, Ahmadreza**, Zeyun Yu, and Roshan M. D'souza. "Involuntary Eye Motion Correction in Retinal Optical Coherence Tomography: Hardware or Software Solution?" *Under Review*
6. **Baghaie, Ahmadreza**, Zeyun Yu, and Roshan M. D'souza. "Dense Descriptors for Optical Flow Estimation: A Comparative Study" *Under Review*

Books

1. **Baghaie, Ahmadreza**. "MATLAB for everyone!", *Kian Publication*, 590 pages, 2014 (In Farsi)
2. **Baghaie, Ahmadreza**. "Control Engineering in MATLAB and SIMULINK", *Kian Publication*, 304 pages, 2012 (In Farsi)
3. **Baghaie, Ahmadreza**. "Android: Practical Reference", *Kian Publication*, 400 pages, 2012 (In Farsi)
4. **Baghaie, Ahmadreza** and Mehdi Baghaie. "Finite Element method Programming Using MATLAB", *Kian Publication*, 288 pages, 2012 (In Farsi)
5. **Baghaie, Ahmadreza**. "Image Processing With C", *Kian Publication*, 606 pages, 2011 (In Farsi)

Pre-Prints

1. **Baghaie, Ahmadreza**. "Markov Random Field Model-Based Salt and Pepper Noise Removal" *arXiv preprint arXiv:1609.06341* (2016).
2. **Baghaie, Ahmadreza**, and Hamid Abrishami Moghaddam. "Finite Element Method Based Modeling of Cardiac Deformation Estimation under Abnormal Ventricular Muscle Conditions." *arXiv preprint arXiv:1412.4192* (2014).

Computer and Language Skills

- **Programming Languages:** Familiar with: Matlab, C/C++/C#, TeX.
- **Industry Software Skills:** Matlab (Advanced), LTspice (Intermediate), LabVIEW (Intermediate), AVR programming (Intermediate), Ansys (Basic), Most MS Office products (Advanced).
- **Test scores:** TOEFL iBT(97), GRE (V:147, Q:160, W:3)

References

- Available on request