ADVANCED RETINAL IMAGING: FEATURE EXTRACTION, 2-D REGISTRATION, AND 3-D RECONSTRUCTION

By

THITIPORN CHANWIMALUANG

Bachelor of Engineering
Chulalongkorn University
Bangkok, Thailand
1995

Master of Science
Pennsylvania State University
State College, PA
2001

Submitted to the Faculty of the
Graduate College of
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 2006

ADVANCED RETINAL IMAGING: FEATURE EXTRACTION, 2-D REGISTRATION, AND 3-D RECONSTRUCTION

Dissertation Approved:

Guoliang Fan, Ph.D.

Dissertation Advisor

Gary Yen, Ph.D.

Daqing Piao, Ph.D.

Sundar Madihally, Ph.D.

Gordon Emslie, Ph.D.

Dean of the Graduate College

ACKNOWLEDGMENTS

First, I would like to express my sincere thanks to my advisor, Professor Guoliang Fan, for his guidance and support throughout my Ph.D. study. He has taught me not only the technical knowledge, but also attitudes towards research. I am also gratitude to Professor Gary Yen who is my committee chair and a co-PI on this project. I would like to thank Professor Daqing Piao and Professor Sundar Mahidally for serving on my dissertation committee. My thanks also go to Professor Stephen R. Fransen, who is an ophthalmologist at the University of Oklahoma Health Science Center, for providing ETDRS retinal images and his medical advices. I also want to thank the current and previous members of my research group, visual computing and image processing laboratory (VCIPL), for their helps and friendship.

Most important of all, I want to express my deepest gratitude towards my family: my parents, my sisters, and my grandmother, who always have faith in me. I cannot thank them enough for their unconditional love, supports, patience, and understanding. To my grandmother who is 81 years old, always concerns about my health and wishes to see me coming home with my success before her death: grandma, I will come home soon and you are part of my success. To my parents who have taught me honesty, perseverance, and merit: you are my good examples and part of my accomplishment. To my sisters who are willing to listen to every of my silly complaints: you are my encouragement.

This work was supported by an OHRS award for project number HR03-33 from the Oklahoma Center for the Advancement of Science and Technology (OCAST).

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1

Introduction

1.1 Motivations and Objectives

Diabetes is the leading cause of blindness among working-age Americans, and many patients with vision-threatening diabetic retinopathy remain asymptomatic until blindness occurs [10]. The great majority of this blindness can be prevented with proper eye examination and treatment by ophthalmologists who rely on the results of randomized clinical trials, called Early Treatment Diabetic Retinopathy Study (ETDRS), to guide their treatment of patients with diabetes [11]. ETDRS requires sets of retinal images to be captured from different fields of an eye. Because ophthalmologists rely on multiple retinal images for disease diagnosis and evaluation, these images need to cover a required area of the retina. The retinal images need to meet the image quality criteria defined by ETDRS protocol. Each set of fundus photographs should be assessed for quality before the patient leaves the imaging center. A photographer is required to decide whether a particular image set meets the three ETDRS's image quality assessment (IQA) requirements: (1) clarity & focus; (2) field definition; and (3) stereo effect. A software tool to standardize and certify image quality is in demand. Several types of visual models including graphical retinal mapping, 3-D retinal surface reconstruction, and 2-D retinal registration can (1) assist ophthalmologists in diagnosing, analyzing, and evaluating the disease; (2) facilitate clinical studies; and (3) be used as a spatial map during laser surgical procedures. Furthermore, many respectable researches in psychology have said that a person can always put a smile on his/her face but eyes usually reveal his/her true feelings. We agree with the statement completely considering that eyes are the only place where we can, without any invasive medical techniques, directly see inside a human body. We can observe, in real time, a retina's blood vessels which are significant indicators to not only eye-related diseases but also various other diseases. In this work, advanced retinal imaging approaches have been developed according to the aforementioned needs as follows.

- Objective 1: To develop an efficient feature extraction algorithm that can effectively segment blood vessels and to select reliable point correspondences for latter processing.
- Objective 2: To develop a robust 2D retinal image registration algorithm that can register seven ETDRS retinal images together and perform image quality assessment on field-coverage.
- Objective 3: To study an accurate 3D retinal surface reconstruction algorithm that can recover the 3D shape of retina from ETDRS retinal images.

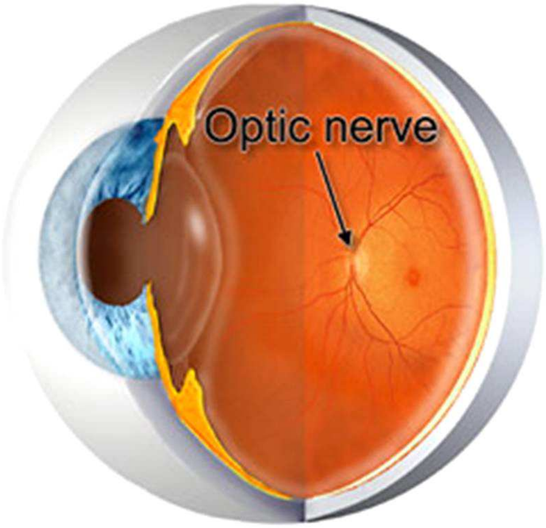


Figure 1.1: Human retina.

1.2 Significance and Background

Diabetic retinopathy is a complication of diabetes [12]. Vision of a person with diabetic retinopathy is shown in Fig. 1.2(b). The disease affects blood vessels inside the retina. The retina is an area lying at the back of the eyeball as shown in Fig. 1.1. In accordance with [12, 11, 13], the earliest stage of the disease, the tiny blood vessels, or capillaries, become thinner, weaker and eventually they leak blood (microaneurysm) as illustrated in Fig. 1.3(b). A patient's sight at this stage is still good but an ophthalmologist can detect and notice the abnormalities in the retina. As the disease progresses, some blood vessels are blocked. These trigger the retina to grow new blood vessels, which are abnormal, fragile, and easily bleed as shown in Fig. 1.3(c). In the later stage of the disease, new blood vessels are grown continuously as well as scar tissue as shown in Fig. 1.3(d). Ultimately, retina will be detached from an eye. National eye institute (NEI) recommends everyone with diabetes to have comprehensive eye exam at least once a year because diabetic retinopathy has no early warning symptoms or signs. According to [14], there are 20.8 million people in the United States, or 7% of the population, who have diabetes. While an estimated 14.6 million have been diagnosed with diabetes, unfortunately, 6.2 million people (or nearly one-third) are unaware that they have the disease. Failure to undergo universally recommended annual eye examinations is the primary cause of this continued loss of sight. If detected early, majority of the severe vision loss from diabetic retinopathy can be prevented with proper examination and treatment by ophthalmologists. Ophthalmologists primarily rely on the results of randomized clinical studies called ETDRS to guide their treatment of patients with diabetes.

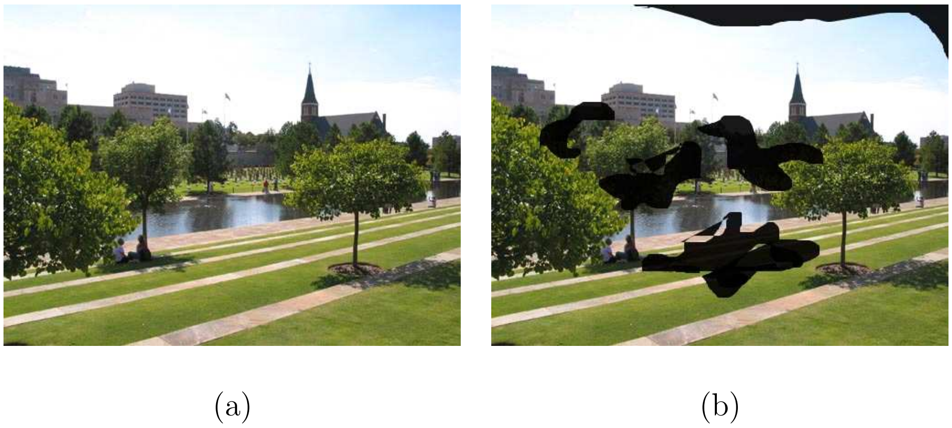


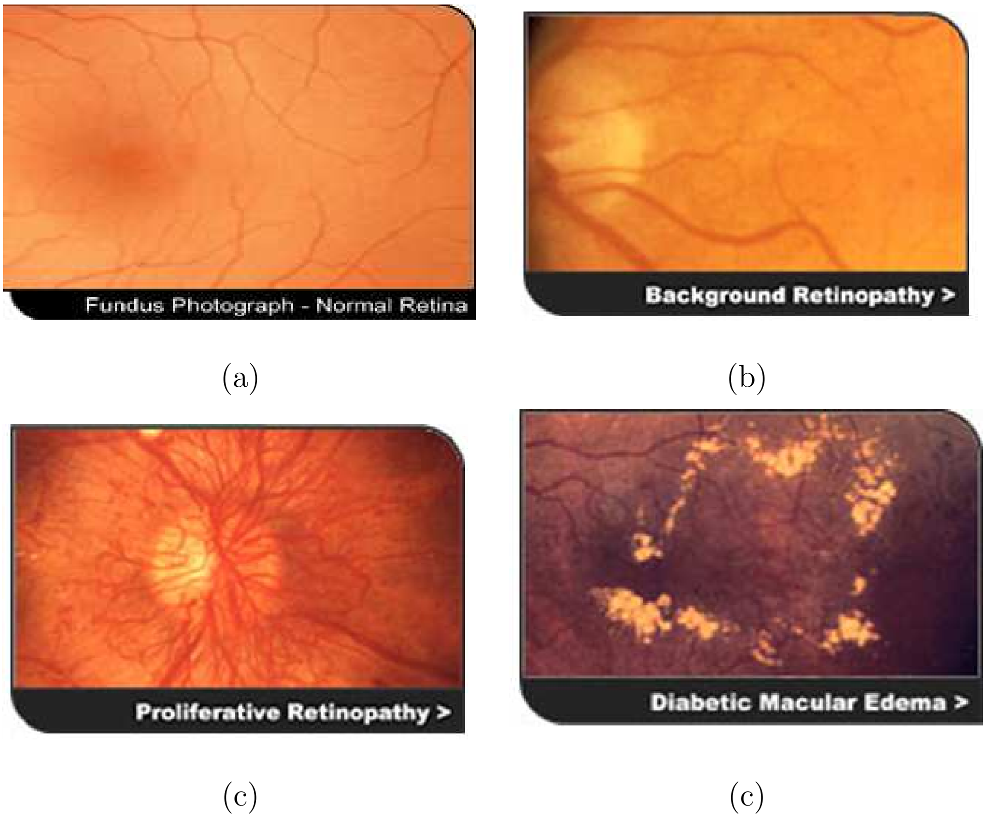
Figure 1.2: Diabetic retinopathy. (a) Normal vision. (b) Same scene viewed by a person with diabetic retinopathy.

1.3 NIH's ETDRS Protocols

The Early Treatment Diabetic Retinopathy Study (ETDRS) implemented standardized retinal imaging, classification and severity staging for diabetic retinopathy as well as proving the therapeutic benefit of laser photocoagulation surgery in preventing vision loss [15]. This multicenter, randomized clinical trial designed to evaluate treatment of patients with nonproliferative or early proliferative diabetic retinopathy. A total of 3,711 patients were recruited to be followed for a minimum of 4 years to provide long-term information on the risks and benefits of the treatments under study. The study demonstrated a statistically significant reduction in severe visual loss for those eyes with early treatment [11]. The ETDRS also developed an internationally recognized disease severity scale indicating the risk for diabetic retinopathy [16]. ETDRS protocols have become the "gold standard" for evaluating diabetic retinopathy [17] and diabetic macular edema [18].

1.4 Technical Challenges

In this work, we have addressed retinal image analysis issues: (1) image quality assessment (IQA) in terms of field coverage which can help in grading, and support


Fundus Photograph - Normal Retina
Background Retinopathy >
Proliferative Retinopathy >
Diabetic Macular Edema >

(a) (b)

(c) (c)

Figure 1.3: Different stages of diabetic retinopathy. (a) An example of normal retinal image. (b) A retinal image with microaneurysms. (c) Proliferated diabetic retinopathy with fragile newly grown blood vessels. (d) A retina image with some types of scar tissues. (http://www.inoveon.com)

the grader training processes; (2) several types of visual models, i.e. graphical retinal map, 2-D retinal registration, and 3-D reconstructed retinal surface, which can greatly assist ophthalmologists in diagnosing and evaluating the disease, significantly facilitate clinical studies, as well as assist in laser surgical procedure by using visual models as spatial maps. The technical challenges associated with retinal image analysis are

- **Feature extraction and correspondence selection** Retinal images may not be well-focused and blurred due to inappropriate image acquisition conditions. Image quality variability is the major challenge for feature extraction retinal images. Specifically, it includes:

  - Poor lighting condition can cause glaring or introduce artificial effects in retinal images.
  - Retinal images are usually dominated by the red homogeneous color with different shades. Background and foreground are difficult to distinguish.
  - Non-uniform illumination will lead to inconsistent intensity of both blood vessel and background within an image and across images.

- **2-D retinal image registration**: In addition to the challenges for feature extraction and correspondence selection, there are also some difficulties for ETDRS-based retinal image registration.

  - Large homogeneous or textureless areas in retinal images complicate the registration process due to insufficient information for area-based approaches and inadequate features in feature-based approaches.
  - The overlaps between multi-field images are relatively small. This presents another major difficulty to the procedure. It is not reliable to estimate the transformation for the whole images based on the small overlap regions.
  - The curved retinal surface and camera motion across ETDRS fields requires high-order non-linear transformations for image registration.

- **3-D retinal surface reconstruction** The 2-D registration results are the perquisite for 3-D retinal surface reconstruction where we are facing the same challenges in 2-D registration as well as some new problems as follows.

  - The fundus camera parameters are unknown. Camera calibration has to be done prior to 3-D reconstruction.
  - The virtual lens effect from human cornea and lens distortion from the fundus camera have to be taken into account for 3-D reconstruction.

1.5 Research Flow

In order to provide readers better understandings of the materials and subjects investigated in this work, we use this section to provide the general ideas covered in this dissertation. An architecture of our system can be decomposed into a three-layer hierarchy as illustrated in Fig.1.4.

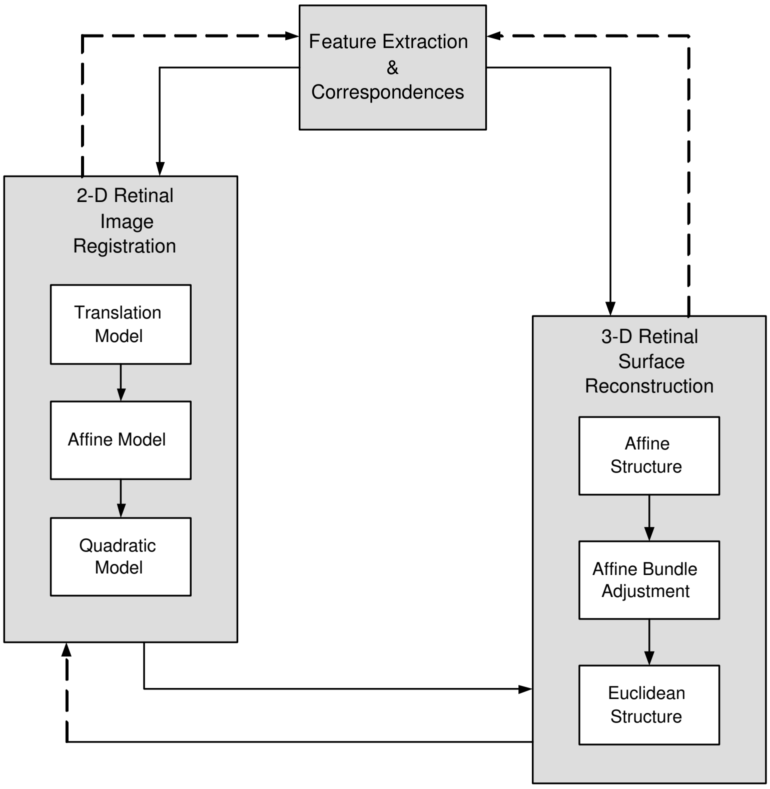


Figure 1.4: Architecture of the proposed research.

Feature correspondences provide input information for both 2-D retinal image registration and 3-D retinal surface reconstruction. The vice-versa direction, knowledge and information regarding displacement, structure and motion will further improve correspondences accuracy. A similar analogy exists between 2-D registration and 3-D surface reconstruction. Displacements from 2-D model supply additional input information for 3-D surface reconstruction. On the other hand, geometric shape can considerably improve 2-D registration results. In 2-D registration layer, three-stratification sublayers are presented in order to improve algorithm's robustness, efficiency and accuracy. The least complex model, translation, is first identified. Then, information from translation model is used as constraints for the affine model. Finally, quadratic model is achieved with constraints from affine model. A similar stratification approach is used in 3-D surface reconstruction. Although the most relaxed solution in projective space is projective structure, we start with affine structure. This is due to the fact that we use affine camera. Once affine structure is obtained, affine bundle adjustment is employed to jointly refine all parameters. Then, extra metric information is used to correct the structure back to Euclidean structure. Brief explanation regarding each layer is given next.

Feature Extraction and Correspondence Selection The first problem can be divided into two subproblems, feature extraction and feature matching/correspondences. Let us start with the first subproblem, feature extraction. We want to extract features because features can reduce the amount of data needed to be processed. Since "Not all information is created equal [19]", what are the most suitable features for matching? Points are used most of the time. Lines or blobs are also good features since they provide more information compared to points. Good features should contain as much distinguishable information as possible. The next question would be what are the best approaches to extract features? And once features are obtained, how to match features across images? Although there are studies of image correlation [20, 21, 22], feature extraction and correspondences are pretty much image-dependent problems. Certain algorithms are good for some specific types of images but perform poorly on other image types.

2-D Image Registration Registration is a problem on how to coincide two or more images. Two images are often taken at different times, viewpoints, modes, or resolutions. Additionally, an image plane, and an world plane are often not parallel. Hence, it is impossible to simply overlay two images together. To register two images, an "optimal" transformation model has to be identified. Numerous algorithms have been proposed regarding this topic. These methods differs in many aspects: (1) feature-based methods versus area-based methods; (2) batch methods versus RANSAC-like methods; (3) low-level methods, e.g. optical flow, autocorrelation, versus shape-based methods, e.g. template matching; (4) spatial domain versus frequency domain.

3-D Surface Reconstruction Visual reconstruction is a process to recover a 3-D scene or model from multiple images. It is usually referred to as a structure from motion, SFM, problem. A process usually recovers objects' 3-D shapes, cameras' poses (positions and orientations), and cameras' internal parameters (focal lengths, principle points, and skew factors). Many possible camera models exist. A perspective projection is the standard. However, other projections, e.g. affine, orthographic, are sometimes prove more useful and practical for a distant camera. The main differences between projections are the required level of calibration. 3-D reconstruction problem has been extensively studied and currently is a very active research topic.

1.6 Original Contributions

Before the outline of each chapter is given, we would like to summarize main contributions we believe we have made for this work. This dissertation is written based on a journal paper, four conference papers. There were certain motivations and contributions at the times each topic was investigated. The specific nature of retinal images leads to novel solutions for each problem.

1.6.1 Feature Extraction and Correspondence Selection

Vascular tree and its bifurcation/crossover points are used as feature and correspondences. This is due to the two following reasons: (1) vascular tree spans the whole retina hence it exists in every retinal images; (2) bifurcation/crossover points offer more distinguishable information than homogeneous area throughout the retina. Because of non-uniform intensity of both background and foreground as well as the low contrast, multi-directional match filters are employed to enhance the contrast of blood vessels from the background. Next, a new description of co-occurrence matrix has been proposed. Then, a new thresholding method based on the idea of local entropy is introduced to extract blood vessels from the background. Because all blood vessels should be connected, length filtering is used to get rid of small separated regions. Finally, bifurcation/crossover points are detected by morphological thinning operation and window-based probing approach. The simulation results suggest that the proposed framework is efficient and robust. Additionally, our false positive rates are lower than other computational expensive techniques while the true positive rates are comparable.

- **A new co-occurrence matrix's definition** The new definition takes into the account of both image spatial structure and noise in an image. Simulation results on several matched filter images demonstrate good performance in terms of robustness and accuracy.

- **A new thresholding algorithm** The proposed equation is slightly resemblance to both local entropy thresholding and relative entropy (or cross entropy) thresholding methods. The simulation results on various matched filter images exhibit better performance in terms of robustness and accuracy.

1.6.2 2-D Retinal Image Registration

Because area-based and feature-based have their own strengths and limitations, in this work, we combine both area-based and feature-based registration methods to get the advantages each method has offered along with other decision-making criteria in order to obtain the best optimal solution. In order to achieve robustness and efficiency, hierarchical technique, translation, affine, and quadratic, is incorporated. Because of non-uniform intensity within an image, binary mutual information is proposed for translation estimation. It demonstrates better performance in terms of robustness compared with traditional gray-scale mutual information. In addition, multi-scale searching strategy is applied to avoid large combinatorial searching space. Sampling point correspondences are introduced when bifurecation/crossover points are inadequate. An iterative closest point algorithm is used to refine feature points as well as transformation models. Furthermore, two parameters characterizing the displacements along vertical and horizontal directions in translation model, suggesting relative positions of each field, can be used for image quality assessment (IQA) regarding field coverage definition.

- **A hybrid registration method** We combine both area-based and feature-based registration methods to get the advantages each method has offered. A translation model is estimated through binary mutual information while higher-models are approximated through feature-based methods.

- **Binary mutual information** A binary mutual information is proposed. It demonstrates better performance in terms of robustness compared with traditional gray-scale mutual information.

- **Empirical conditions** The conditions are effectively combined all the techniques into one flow where the algorithm is able (1) to adaptively select an appropriate transformation model, (2) to determine whether sampling point correspondences have to be involved, and more importantly, (3) to reject invalid registered pairs.

- **Image quality assessment (IQA)** We have addressed an issue of image quality assessment in terms of field coverage by the two parameters characterizing the displacements along vertical and horizontal directions in translation model.

1.6.3 3-D Retinal Surface Reconstruction

In this research, we assume a weak-perspective camera because of the two following reasons: (1) the ETDRS imaging standard specifies a 30° field of view each eye (narrow field of view); (2) each retinal image has small depth variation. We derive an affine camera model and show mathematical proof of an affine condition. Affine structure from motion has been investigated and an affine factorization method is used for initial reconstruction because the approach can accommodate multiple images and utilize the use of all feature points. An affine bundle adjustment based on a nonlinear optimization technique is used to refine an affine shape and affine cameras. Then, a Euclidean constraint is involved to correct both the affine shape and cameras into Euclidean space up to a similarity. Next, we take into an account of an eyeball's geometric constraint in order to generate denser points for surface. We assume that an eyeball is an approximated sphere. A point-based sphere fitting method is introduced.

- **The condition for the affine camera** We have shown a mathematical proof of an affine camera from a standard linear camera as well as its condition.

- **Affine bundle adjustment** Inspired by projective bundle adjustment, we introduce affine bundle adjustment to refine all parameters, affine shape and affine cameras, simultaneously.

- **A point-based spherical fitting method** We introduce a linear approach to solve a nonlinear surface approximation problem.

- **Constrained affine bundle adjustment with lens distortion updates** We introduce an optimization process which incorporates a geometrically meaningful definition and lens distortion into a cost function. We propose a constrained optimization algorithm in an affine space rather than the traditional Euclidean space. The procedure optimizes all of the parameters, camera's parameters, 3-D points, physical shape of a retinal surface, and lens distortion, simultaneously.

1.7 Outline

The organization of this dissertation is illustrated in Fig. 1.5.

- The motivation and significance of this research as well as relevant materials and subjects are presented in this chapter.

- Chapter 2 reviews the-state-of-art research on retinal image analysis and categorizes them into five different areas, i.e., structure segmentation, image registration, 3D reconstruction, image quality assessment (IQA), and image classification.

- Chapter 3, we reviewe the mathematical background of the proposed research, and the materials presented here serve the foundation of latter chapters. Specifically, we will address 2-D retinal image registration in Chapter 5 that requires 2-D/2-D transformations, and 3-D retinal reconstruction is discussed in Chapter 6 that involves a camera projection (3-D/2-D transformations) and 3-D registration (3-D/3-D transformations).

- Chapter 4 deals with the first layer of architecture, feature extraction. Features are significantly important in motion estimation techniques because they are input to the algorithms. However, most works studied often neglect this part and assume features are available. We have proposed a feature extraction algorithm for retinal images. Bifurcations/crossovers are used as features. A new thresholding algorithm based on our definition of co-occurrence matrix is proposed. As a result, vascular tree which is an important structure to indicate many diseases has also been extracted.

- In chapter 5, we consider 2-D retinal image registration which are the problem of the transformations of 2-D/2-D. Both linear and nonlinear models are incorporated that account for motions and distortions. A hybrid method has been introduced in order to take advantages different methods have offered along with other decision-making criteria. Binary mutual information is proposed for translation estimation. Hierarchical technique, translation, affine, and quadratic, is incorporated.

- In chapter 6, a 3-D retinal surface reconstruction issue has been addressed. To generate a 3-D scenes from 2-D images, a camera projection or transformations of 3-D/2-D techniques have been investigated. We choose an affine camera to be a represented model for a fundus camera. We have provide our proof to justify the use of affine camera. An affine bundle adjustment based on non-linear optimization technique is established to refine an affine shape and affine cameras. A point-based spherical approximation is introduced.

- In chapter 7, an objective for this chapter is to solve the problem in an optimal way which means to estimate structure and camera parameters simultaneously by minimizing a physically meaningful cost function. An optimization procedure, called constrained affine bundle adjustment with lens distortion updates, is proposed to improve the algorithm's performance in terms of accuracy and robustness.

- Chapter 8 states future works and concludes the dissertation.

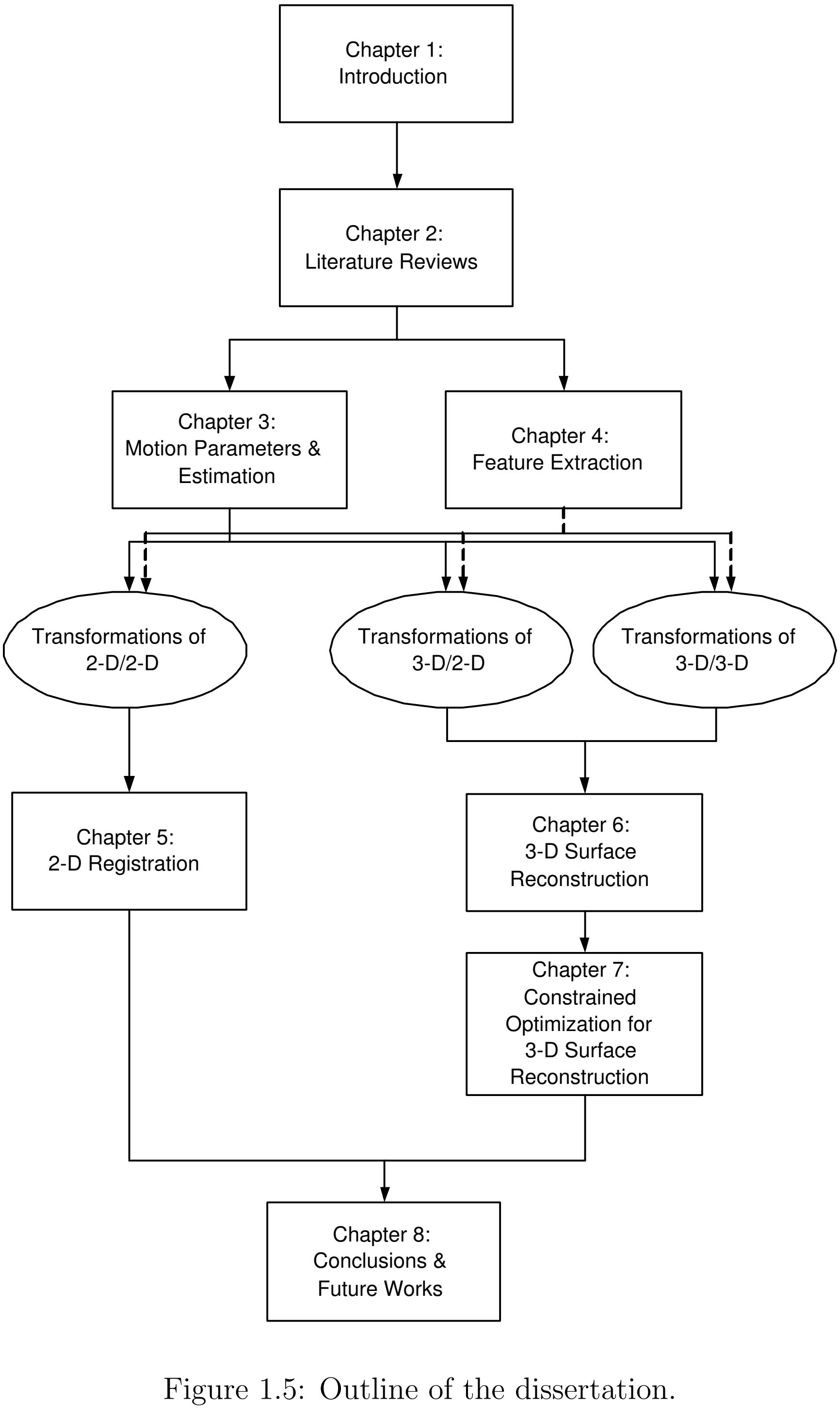


Figure 1.5: Outline of the dissertation.

CHAPTER 2

Literature Reviews

2.1 Overview of Retinal Image Analysis Research

Computer-assisted retinal image analysis can be used for many purposes: (1) helping ophthalmologists diagnosing and evaluating eye-related diseases; (2) patient screening and grading disease severity; (3) facilitating clinical studies; (4) quantifying retinal image quality; and (5) assisting in laser surgical procedure. Numerous techniques have been investigated and developed regrading retinal image analysis researches. The objective of this chapter is to categorize, characterize, and review these algorithms. Before we move further into more detail, we'd like to note here that there are several retinal image types in which two types are widely used: (1) normal retinal (fundus) images; and (2) fluorescein angiogram (FA) images. A FA image is obtained by injecting a special dye, called fluorescein, into patient's vein in the arm. The dye moves quickly to blood vessels inside the eyes. The result is that blood vessels become more prominent from the background (high contrast). In image analysis point of view, therefore, FA images are generally less complicated to be processed in comparison with normal fundus images. Here, we categorize retinal image-related researches into five main groups: (1) structure segmentation; (2) 2-D registration; (3) 3-D reconstruction; (4) image quality assessment (IQA); and (5) image classification. The flowchart is illustrated in Fig. 2.8 where the gray-shaded boxes refer to the works involved in this work.

2.2 Structure Segmentation

Image segmentation is one of the fundamental problems in computer vision and many other research areas. Many subsequent tasks, e.g. feature extraction, pattern recognition, image retrieval, image registration, image compression, image classification, and etc., rely on the quality of image segmentation process. There are no universal theories as to what is the best approach for image segmentation. Segmentation is pretty much an image-dependent problem. However, good segmentation is that segmented region should be uniform with respect to some semantic characteristics. As for the retinal image case, we'd like to classify segmentation into two main categories, (1) anatomical structures which include blood vessel, optic nerve, and fovea; (2) pathological structures, i.e. lesion, which are abnormal structures. The goal is to detect and present the location of important structures in the retina as well as to find correspondences across retinal images. Here, we only focus on blood vessel detection since our main tasks are registration and reconstruction and blood vessels are used as features.

2.2.1 Anatomical Structure Segmentation

Two significant anatomical structures presented in a retina are blood vessels and optic nerve as shown in Fig. 2.1. Besides, all the medical motivations mentioned above, a segmentation of the vascular tree seems to be the most appropriate representation for the image registration applications. This is due to the two following reasons: (1) vascular tree spans the whole retina hence it exists in every retinal images; (2) bifurcation/crossover points offer more distinguishable information than other homogeneous areas throughout the retina.

A variety of approaches have been proposed for vascular segmentation [9, 8, 6, 23, 24, 25, 26, 27, 28, 29, 30, 31]. Here, we arrange them into two main categories, supervised and unsupervised techniques. The major difference is that supervised

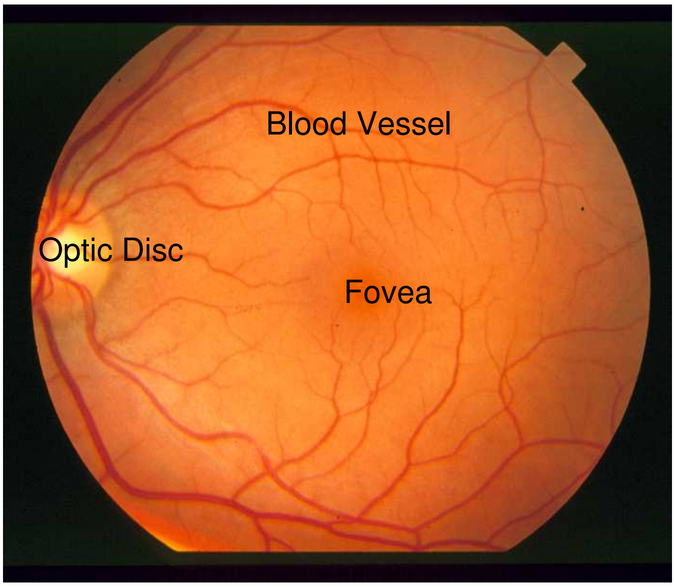


Figure 2.1: Anatomical structures in a human retina.

techniques require training data. Although, generally speaking supervised methods should yield better segmentation results because of additional knowledge, training database, most of the algorithms proposed belong to unsupervised category. This is due to the fact that hand-labeled vessels is a tedious task which takes more than a couple of hours to complete just one retinal image. Therefore, it is not practical for real applications.

- **Supervised Approaches.** Manually labeled images are required for training purpose. To the best of our knowledge, there are only four retinal's blood vessel segmentation papers [25, 32, 6, 7] belonging to this group. Sinthanayothin et.al. [25, 32] used a multi-layer perceptron for blood vessel classification with back-propagation as a training approach. An advantage is its ability to deal with nonlinear classification. The limitations are that it is difficult to see what is going on inside the hidden layers and training process is repeated every time new features are incorporated. They partitioned each retinal image into 10×10-pixels sub-images. There were total 25094 sub-images in which $\frac{5}{6}$ of the data were hand-labeled ground-truth images and used as their training set. The rest was used for validation. Staal et.al. [6, 7] used ridge detection to locate candidate blood vessel segments. Ridges are defined as points where the image has an extremum in the direction of the largest surface gradient [33, 6]. The direction of largest surface gradient is the eigenvector of the Hessian matrix corresponding to the largest absolute eigenvalue. After ridges are defined, several feature sets and decision-making criteria are applied to create convex sets as well as classify pixels to be vessels or not. Then, ridge pixels were groups and convex sets were formed. After that, the image was partitioned into patches based on the convex set. Every pixel was assigned to the convex set to which it was closest. Next, feature sets were formed and kNN-classifier was employed. Twenty hand-labeled ground-truth were used as a training set and twenty images were used for validation. Although, good segmentation results are reported, the algorithm is computationally expensive and hand-labelled ground-truth images are mandatory.

As the title has suggested, approaches belong in this group *need* hand-labelled ground-truth images. Hand-labelled vascular tree is not practical in real application since it consumes a great deal of time. This is a significant drawback of algorithms in this group.

- **Unsupervised Approaches.** No ground-truth information is provided. Variety of unsupervised approaches have been proposed. We further divide them based on the proposed techniques.

  - Window-Based. Window is referred to a pattern that is used to transform an image. The process is usually followed by a binarized technique. Pinz et.al. [26] used local gradient maxima because it occurs at the boundary of the vessels, the significant edges along these boundaries were extracted.

The grouping process searched a partner for each edge which satisfies certain criteria like opposite gradient direction and spatial proximity. Only the vascular centerlines could be detected. FA images were used in their research. Zana et.al. [27] also used FA images. Therefore, they defined a vessel as a bright pattern and linearly piece-wise connected. Opening morphological filters with linear structuring elements were used. Each structuring element is 15-pixels long (every 15^o). The sum of top-hats on the filtered image brightened all blood vessels (linear parts) and reduced small bright noise. In order to remove non-vessel parts, principal curvature was computed by using Laplacian followed by morphological opening.

- Classification-Based. First step is to divide an image into different regions. Then, multiple rules are applied to classify pixels in each region as being vessels or not vessel. Hoover et.al. [9] used twelve 16×16-pixel matched filter proposed in [34] to map the vascular tree. A set of criteria was tested to determine the threshold of the probe region, and ultimately to decide if the area being probed was a blood vessel. Since the MFR image was probed in a spatially adaptive way, different thresholds were applied throughout the image for mapping blood vessels. Jiang et.al. [8] used verification-based multiple threshold probing framework. A retinal image was probed at different threshold values. At a particular threshold, Euclidean distance transform was performed. Then, vessel candidates were pruned by means of the distance map to only retain centerline pixels. Finally blood vessels were reconstructed by the particular threshold.
- Tracking-Based. Zhou et.al. [24], defined each vessel segment by three attributes, direction, width, and center point. The density distribution of cross section of a blood vessel were estimated using Gaussian shaped function. Individual segments were identified using a search procedure which

kept track of the center of the vessel and made some decisions about the future path of the vessel based on certain vessel properties. This method required that beginning and ending search points were manually selected using cursor. FA images were used in their research. Can et.al. [35] used tracing method that based on adaptive exploratory processing of an image. Algorithm explored the image along a grid to seek local gray-level minima by using directional low-pass template. Intersections between grid and local minima were labeled as seed points along with their orientations. Seed points were then used for tracing which was repeated for 16 directions. The tracing followed the strongest edge. Chutatape et.al. [36] used second-order derivative Gaussian matched filters to locate center point, width, and orientations. Then, extended Kalman filter was employed for the optimal linear estimation of the next possible location. The algorithm began from circumference of an optic disc. Wu et.al. [37] divided blood vessels into large and small. Blood vessels were enhanced with matched filter. Gabor standard deviation filter was used to distinguish the large and small vessels. Then, 2-D Gaussian filter was used for tracing. Different rules for forward and backward verification were used for large and small vessels.

For window-based techniques, performance of algorithms largely depend on thresholding techniques. To our surprise, most papers belonged to this group do not put a focus on thresholding algorithm. For classification-base techniques, multiple threshold values are applied to different regions. This leads to several decision-making criteria which in themselves require numerous threshold values to select an appropriate threshold value for a particular region. For tracking-base techniques, an initial point needs to be identified. Moreover, only centerline, not the whole branch of vascular tree, can be extracted.

We have addressed the blood vessel extraction issue and have proposed a new thresholding technique. More detail of our algorithm can be found in Chapter 4.

2.2.2 Pathological Structure Segmentation

Since our research does not focus on extracting lesion and due to limited time, we will not go into detail of this subject. We include the subject here for the sake of completeness. Basically, the algorithms can be divided into the same two main groups, supervised and unsupervised, as in anatomical structure extraction algorithms.

2.3 2-D Image Registration

Image registration is a process trying to coincide two or more images. To register two images, an optimal transformation must be identified. Image registration is used for many purposes, e.g. integrate information from different images, detect changes in images taken at different times, object recognition, and etc. There are two major types of distortions in an image: (1) misalignment between images; (2) distortion caused by camera (world plan and image plane are not parallel). Hence, it is impossible to simply overlay two images together. Several retinal-related registration methods have been proposed [38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51]. Reviews of general registration can be found in [52, 53]. Several reviews of medical image registration can be found in [54, 55, 56, 57]. In the 2-D retinal image registration, we categorized based on two criteria, types and techniques. As for the first criteria, types, we classify registration types into three main categories, (1) view-based or mosaic registration; (2) temporal-based registration; and (3) modal-based registrations. Regarding the second criteria, techniques, we group registration techniques into only two main categories, feature-based and area-based. Feature-based methods require set of feature correspondences, which can be points, lines, or blobs, in order to find the optimal registration model. Area-base methods deal with images without trying to locate salient features. The algorithms are based on pixel intensities and certain objective functions. Typically, there are two major factors that may degrade the performance of area-based methods: (1) non-consistent/non-uniform contrast within an image; and (2) large homogeneous/textureless areas. The performance of feature-based methods largely depends on sufficient and/or reliable correspondences, especially, when the overlapping part of an image pair is very limited or when there are mis-matched correspondences.

We first group the 2-D retinal registration researches based on types: (1) view-based or mosaic registration; (2) temporal-based registration; and (3) modal-based registrations. For each type, it can be further divided into feature-based and area-based techniques.

2.3.1 View-Based Registration

This category concerns retinal images taken at different views and how to integrate these images into one view as shown in Fig. 2.2. There are fourteen images taken per one eye according to ETDRS standard. Seven images from different areas of a retina and their corresponding stereo pairs. Integrate all of the images into one piece can be used as a spatial map during surgical procedure and can assist ophthalmologists in evaluating disease.

Can et.al. [38, 39] proposed hierarchical feature-based approach. The similarity matrix for all possible correspondences was computed based on the orientations of vascular centerlines and the similarity measure was converted to a prior probability. The transformation was estimated in a hierarchical way, from the zeroth-order model to the first-order model and finally to the second-order model. Stewart et.al. [43] proposed a feature-based approach called dual-bootstrap iterative closest point algorithm for registration. The approach started from one or more initial, low-order

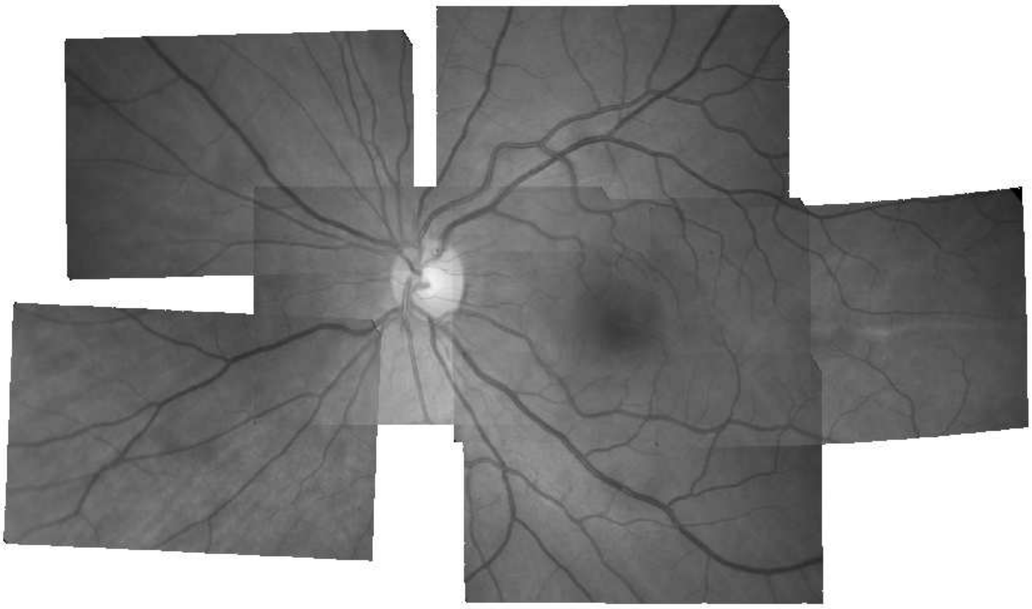


Figure 2.2: View-based registration. See more detail discussion in Chapter 5.

estimates that were only accurate in small image regions called bootstrap regions. In each bootstrap region, the method iteratively refined the transformation estimation, expanded the bootstrap region, and tested to see if a higher-order model could be used. The method required accurate initialization of at least one point correspondence. High success rates were reported in [43]. We have addressed the view-based registration issue [58, 59] More detail of our algorithm can be found in Chapter 5.

2.3.2 Temporal-Based Registration

Retinal images from the same patient are acquired at different times. Temporal-based registration can greatly help ophthalmologists to examine progress of the treatment or disease.

Fang et.al. [60] introduced an area-based affine elastic model for temporal registration. Thinned vascular tree was extracted by using morphological, linear filter, and region growing techniques. An energy function was defined to elastically register two images based on binary vascular tree. Ritter et.al. [1] used area-based technique

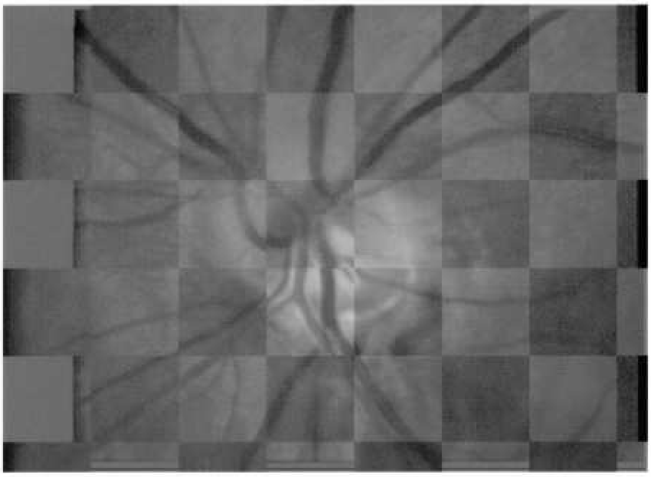


Figure 2.3: Temporal-based registration [1].

by employing mutual information as a similarity criteria. Simulated annealing was used as a searching technique. Simulation result from Ritter et.al. [1] is shown in Fig. 2.3.

2.3.3 Modal-Based Registration

Retinal images are acquired from different sensors, normal mode, FA mode, and other modes. Different modal images offer distinctive information. Ophthalmologists need to combine diverse information from different modes in order to correctly diagnosing and evaluating the diseases.

In [2], registration between two modes, FA and red-free (RF) modes, had been studied. Matsopoulos et.al. [2] used area-based technique. A coarse vessel segmentation was performed in both FA and RF images. The measure of match (MoM), which was similar to a logical operation, was proposed to be used as an objective function and the genetic algorithm was chosen to be the optimization technique. Three transformations, affine, bilinear, andprojective, were included. Simulation result from [2] is shown in Fig. 2.4. Matsopoulos et.al. [61], later, proposed a feature-based method for modal-based registration. Vascular centerlines and their bifurcation points were

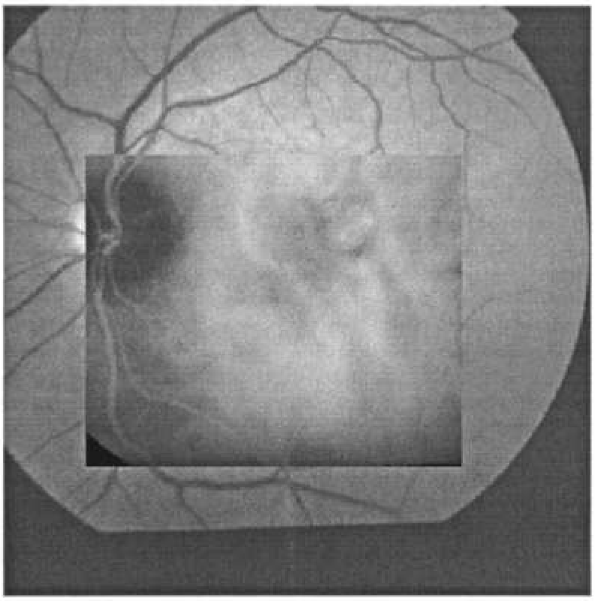


Figure 2.4: Modal-based registration [2].

extracted. Correspondences were identified by using neural network-based approach, self organizing maps (SOM). Affine transformation was estimated in a least mean square sense. Zana et.al. [40] used feature-based technique. Landmark points were extracted and labeled with vessel orintations. An angle-based invariant was computed to give a probability for two points to match. Then, Bayesian Hough transform was used to sort the transformations with their respective likelihood. The most likely transformation was chosen for registration.

2.4 3-D Reconstruction

3-D reconstruction is a process to recover a 3-D scene or model from multiple images. It is usually referred to as a structure from motion, SFM, problem. A process usually recovers objects' 3-D shapes, cameras' poses (positions and orientations), and cameras' internal parameters (focal lengths, principle points, and skew factor). Many possible camera models exist. A perspective projection is the standard. Affine and orthographic projections are widely used in distant cameras. Stereo techniques, e.g. cepstrum, are also commonly used for depth estimation because of their simplicity. They requires only a stereo pair and it does not need camera calibration. Although a general SFM problem has been extensively studied, surprisingly there are not much researches published and dedicated to 3-D retinal reconstruction. We categorize retinal-related 3-D reconstruction into two main categories, 3-D surface reconstruction and local depth reconstruction.

2.4.1 3-D Surface Reconstruction

3-D surface reconstruction refers to global depth reconstruction where we consider the curvature of an eyeball. Deguchi et.al. [3, 62] modelled both fundus camera and eye lens with a single lens. They utilized the fact that a fundus has a spherical shape and image of sphere by the eye lens results in a quadratic surface. They, then calibrated a camera by using two-plane method to get the quadratic surface. Then, eye lens parameters were identified to recover fundus's spherical surface. The simulation result from [3] is shown in Fig. 2.5. Choe et. al. [4] used PCA-based directional filters to extract candidate seed points (Y features). A gradient descent was employed to model Y features and match pairs of features. A plane-and-parallax was employed to estimate the epipolar geometry because a near-planar retinal surface can obstruct a traditional fundamental matrix estimation. The stereo pair is rectified. Then, a Parzen window-based mutual information was used to generate dense disparity map. The simulation result from [4] is shown in Fig. 2.6. In this work, we have also addressed the 3-D surface reconstruction issue using a different approach. More detail of our algorithm can be found in Chapter 6.

2.4.2 Local Depth Reconstruction

Local depth reconstruction refers to recovering the depth of individual objects, e.g. optic nerve and lesions, inside retina. Mitra et.al. [63] proposed the use of power

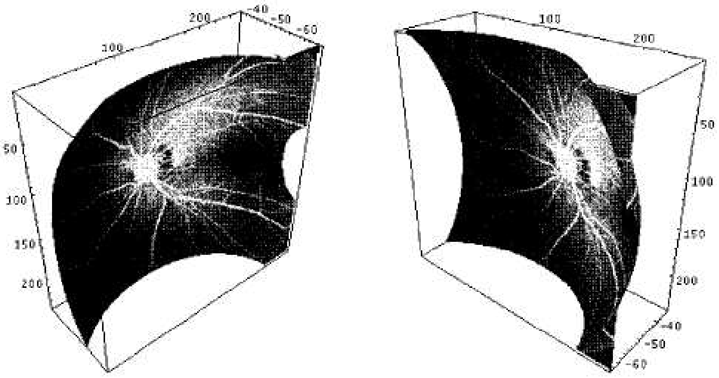


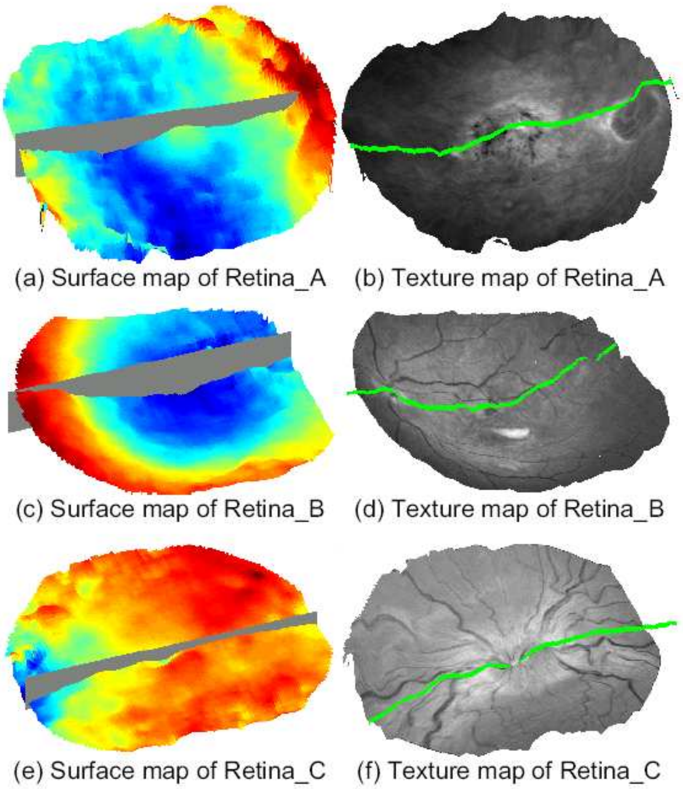
Figure 2.5: 3-D retinal surface reconstruction [3].

cepstrum to find disparity between a stereo pair. Then depth was calculated by a simple triangulation method. Corona et.al. [5] extended the idea from Mitra et.al. [63] and proposed a framework that combined the use of power cepstrum and cross-correlation techniques to extract optic nerve's depth from a stereo pair. Then, b-spline was employed to generate optic nerve smooth surfaces. The simulation result from [5] is shown in Fig. 2.7.

2.5 Image Quality Assessment (IQA)

Image quality assessment plays a significant role in digital image processing research. Many efforts have been made to quantify image quality [64, 65, 66, 67]. In retinal image case, because ophthalmologists rely on retinal images for disease diagnoses and evaluation, the retinal images need to meet the image quality criteria. Each set of fundus photographs should be assessed for quality before the photographs are sent to ophthalmologists. A photographer should be able to decide whether a particular image set meets the three ETDRS requirements:

1. Clarity & focus. This is an obvious requirement in IQA. Focus is defined as sharpness which means the transition between background and foreground is sharp. Clarity is defined as image contrast.



(a) Surface map of Retina_A (b) Texture map of Retina_A

(c) Surface map of Retina_B (d) Texture map of Retina_B

(e) Surface map of Retina_C (f) Texture map of Retina_C

Figure 2.6: 3-D retinal surface reconstruction [4].

2. Field definition. Images need to cover the required areas of retina. The positions of key anatomical structures, optic nerve and fovea, in images are also necessary.
3. Stereo effect. Depth can be perceived only if displacement between images is in an acceptable range.

A software tool to standardize and certify image quality is in demand to assist photographers. Photographers can take a second shot immediately if necessary, rather than calling the patient back for another visit. A software tool should be able to quantify an objective image quality that correlate with perceived quality measurement.

There are limited number of works that have been published and devoted to the issue of retinal image quality assessment. Lee et.al. [68] used template intensity histogram derived from 20 good-quality images. Its base or width was employed as

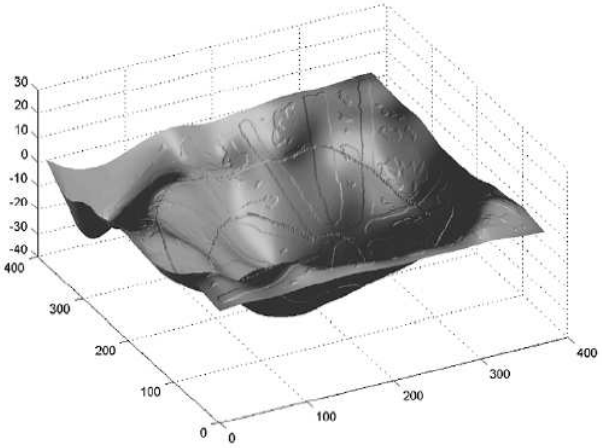


Figure 2.7: Optic nerve's depth reconstruction [5].

an indicator of contrast. The quality of a target image was evaluated by convolving its histogram with the template histogram. Lalonde et.al. [69] suggested the use of two criteria, the distribution of the edge magnitudes and the local distribution of the pixel intensity, to assess images into three groups, good, fair, and poor. Awawdeh et.al. [70] proposed the use of power cepstrum for stereo quality assessment. The stereo pair are added. Then, DCT-based power cepstrum was applied to estimate the displacements which were indicators to stereo quality. In this work, we have also addressed the field definition issue [58, 59] which is a by-product of 2-D image registration. More detail of our method can be found in Chapter 5. To the best of our knowledge, there is no other previous work on the field definition topic.

2.6 Image Classification

The term, image classification, is referred to a process used to assign a specific class to a pixel. In retinal image case, it describes the anatomical and pathological classification. Since our research does not focus on image classification and due to limited time, we will not go into detail of this subject. We include the subject here for the sake of completeness.

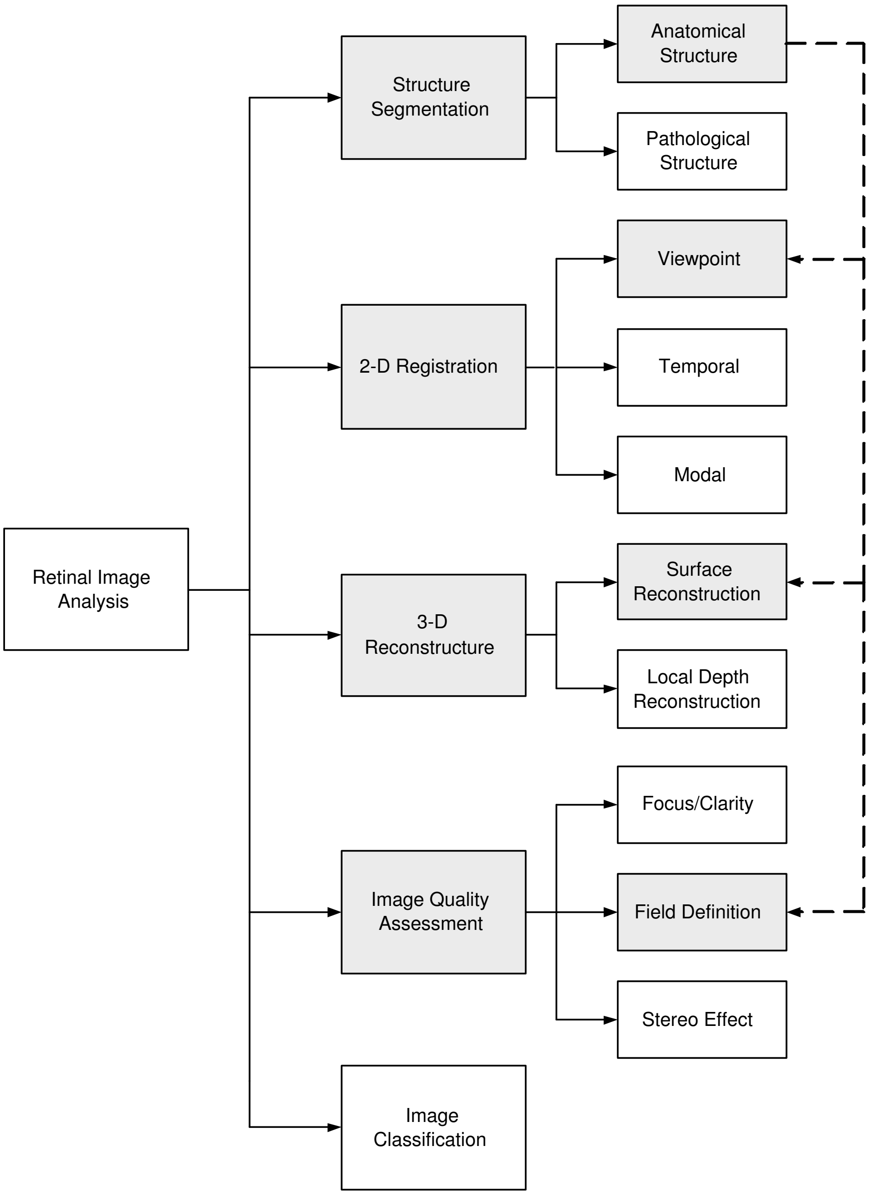


Figure 2.8: Overview of retinal image analysis research. The gray-shaded boxes refer to the topics involved in this work.

CHAPTER 3

Motion Parameters And Motion Estimation

3.1 Introduction

The motion model is needed in order to infer information from multiple images. Suitable motion parameters for representing the possible motions of the camera has to be chosen before motion estimation. The motion parameters are nothing but the parameters in coordinate transformation. There are three major types of motion parameters: (1) 2-D/2-D, (2) 3-D/3-D, and (3) 3-D/2-D. In the 2-D/2-D case, they map an image coordinate, (x, y) to another image coordinate $(\acute{x}, \acute{y})$. 2-D retinal image registration requires 2-D/2-D transformations. In the 3-D/2-D case, they map a world coordinate (X, Y, Z) to an image coordinate (x, y). The 2-D/3-D transformation is generally referred to as camera projection. In order to perform 3-D retinal surface reconstruction, the 3-D/2-D transformations (or camera projection) need to be identified. In the 3-D/3-D case, a world coordinate (X, Y, Z) is mapped to another world coordinate $(\acute{X}, \acute{Y}, \acute{Z})$. After 3-D retial surface is approximated, we need to fuse all of the retinal surfaces into one single view. This task involves the 3-D/3-D transformations. Variety of other tasks, e.g. 3-D reconstruction, 2-D registration, video segmentation, object tracking, site monitoring, and etc, need motion estimation as well. Transformation of 2-D/2-D case and 3-D/3-D case are reviewed in sections 3.2 and 3.3. The 3-D/2-D mapping, camera projection, will be discussed in section 3.4.

3.2 Transformations of 2-D/2-D

The 2-D/2-D transformations are the mapping between image coordinate systems. Several commonly used coordinate transformations are listed in Table 3.1.

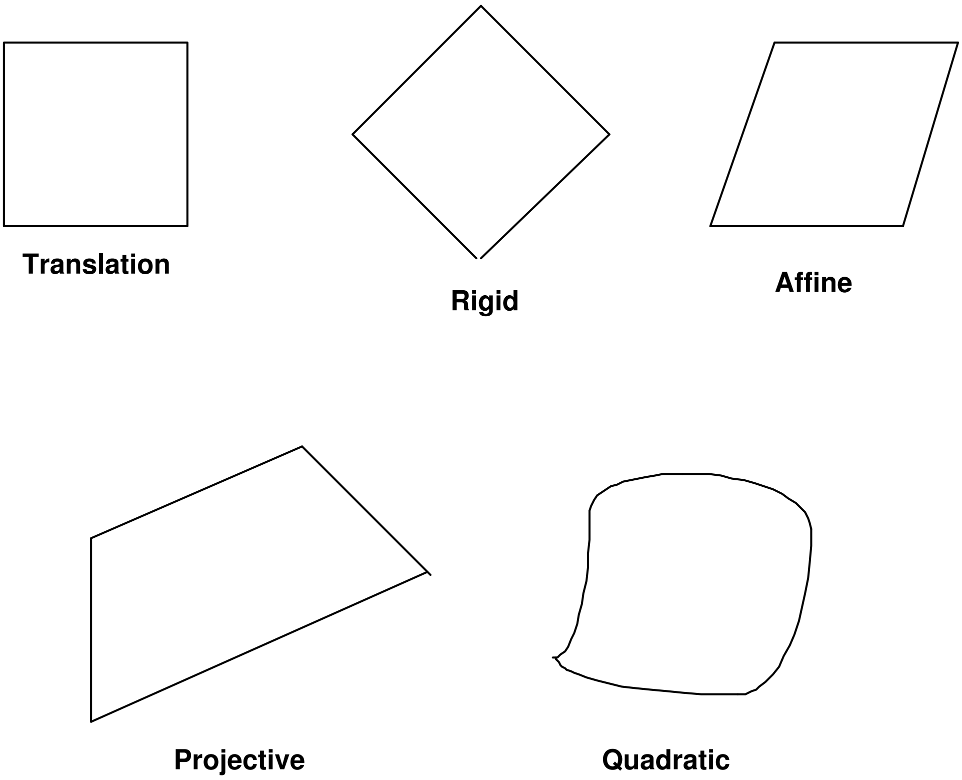


Figure 3.1: Results from 2-D coordinate transformation.

From Table 3.1, t_x and t_y are translation parameters in vertical and horizontal directions respectively. s is a scaling factor. $r_{11}, \ldots, r_{22}$ are rotation parameters. $\theta_{11}, \ldots, \theta_{26}$ are general motion parameters.

A translation model is the simplest one. It can only handle displacements between images. A rigid model can further deal with translation, rotation and scaling between images. Both of them can not handle images with distortions. Absolute distance and area are preserved. They are called rigid invariants. An affine model includes translation, rotation, scaling, and shearing distortion. Parallelism and relative distance along parallel direction are affine invariants. A projective is the general linear transformation describing translation, rotation, scaling, shearing distortion, keystoning distortion, and chirping distortion. The projective invariants are a ratio of ratios (cross ratio) of distances and linearity. Chirping means the effect of increasing or decreasing spatial frequency with respect to spatial location [71]. Keystoning means the effect of convergence lines [72]. The results from each transformation are illustrated in Fig. 3.1. A quadratic model is a nonlinear transformation, accounting for translation, rotation, scaling, shearing, keystoning, chirping, and nonlinear distortions. Nonlinear distortions include several components, e.g. radial distortions, tangential distortions.

3.3 Transformations of 3-D/3-D

3-D/3-D transformations are the mapping between two 3-D world coordinate systems. Several commonly used coordinate transformations are listed in Table 3.2.

From Table 3.2, t_x, t_y and t_z are translation parameters. $r_{11}, \ldots, r_{33}$ are elements in rotation matrices. s is a scaling factor. $\theta_{11}, \ldots, \theta_{44}$ are general motion parameters.

A projective model is the most general case. Every points in the projective space are treated equally. Same as in 2-D/2D transformation, the cross ratio is preserved in the projective space. When the plane at infinity is identified, the projective space becomes an affine space. Hence, points and plane at infinity are affine invariants. Parallelism and relative distance are other affine invariants. Once an absolute conic is identified, the affine space becomes a similarity (or metric) space. Assume (X, Y, Z, T) represents a homogeneous coordinate. Then, $X^2 + Y^2 + Z^2 = 0$; $T = 0$ is known as the absolute conic [73]. Therefore, absolute conic is preserved in the similarity space. Without scaling effect, the similarity space becomes an Euclidean space. Absolute distance and volume are Euclidean invariants. The results from each transformation are illustrated in Fig. 3.2.

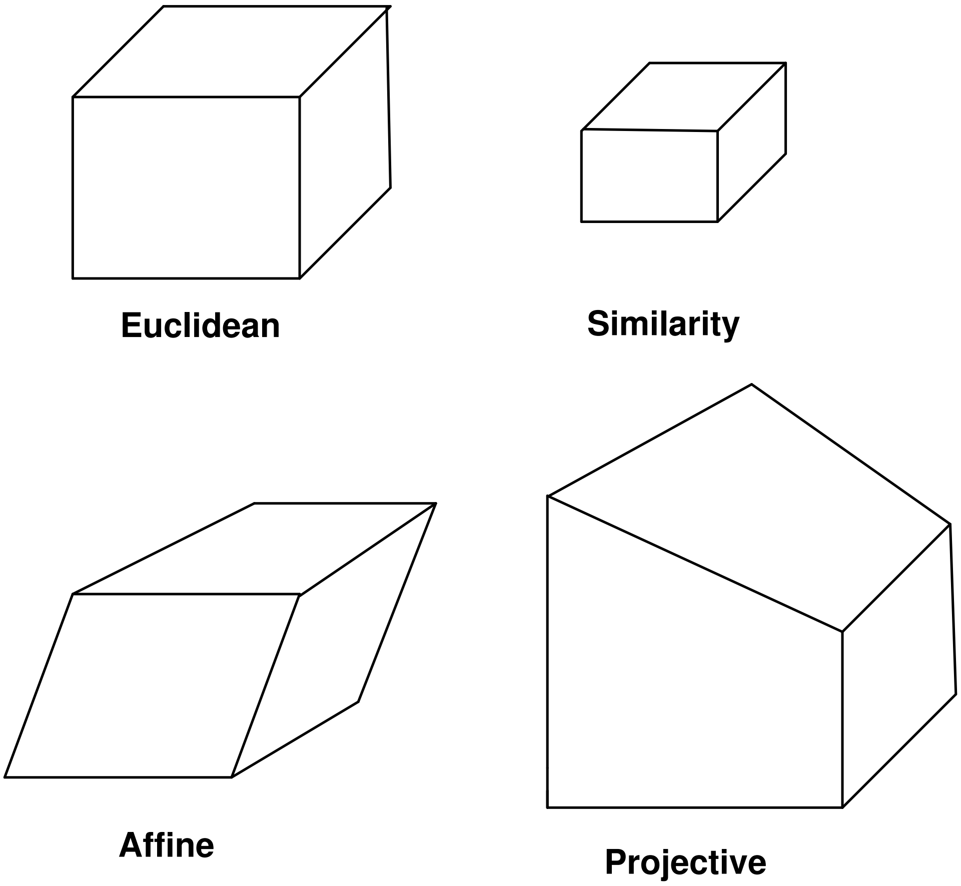


Figure 3.2: Results from 3-D coordinate transformation.

3.4 Transformations of 3-D/2-D

A camera is a mapping between a 3-D world coordinate system and 2-D image coordinate system. Therefore, the 3-D/2-D transformation is generally referred to as a camera projection. Camera motion has been one of the most important subjects in computer vision researches. Motion parameters can be estimated from two or more images depending on types of projection. Before camera motions are described, it is necessary to consider process of image formation.

3.4.1 Image Formation

A pinhole model is a basic camera model for the central projection of points in a scene into an image plane as depicted in Fig. 3.3.

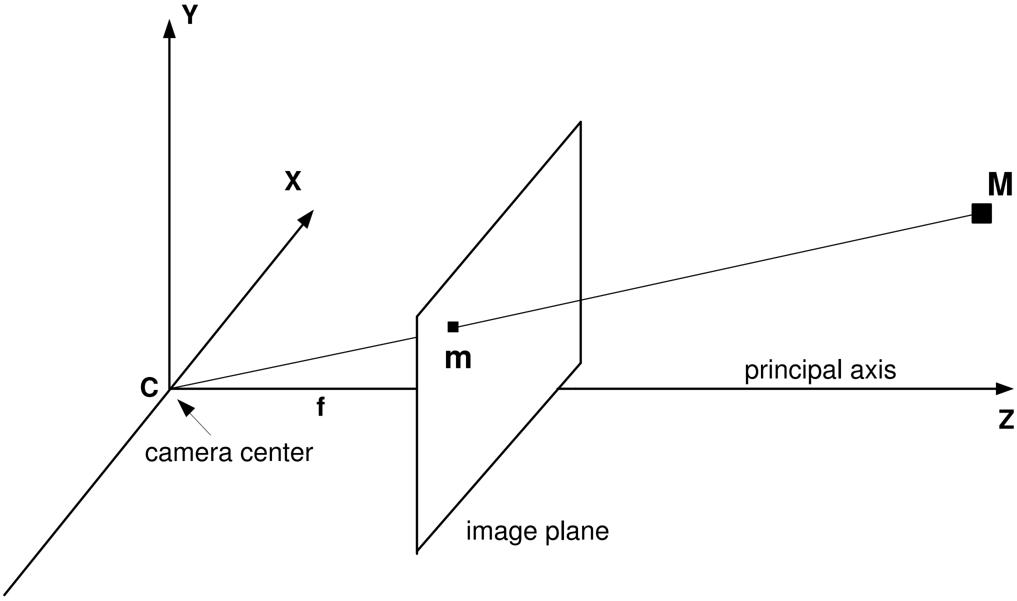


Figure 3.3: Pinhole camera geometry. Camera coordinate system is aligned with world coordinate system.

Under a pinhole camera model, a point in space with coordinate $\hat{\mathbf{M}} = (X, Y, Z)^T$ is mapped to a point on the image plane $\hat{\mathbf{m}} = (x, y)$. Then we have the relationship

$$\begin{aligned} x &= \tfrac{fX}{Z} \\ y &= \tfrac{fY}{Z}, \end{aligned} \tag{3.1}$$

where f represent the focal length of the camera.

Using homogeneous coordinate with principal point offset (c_x, c_y) and skew angle s of image's pixel, the central projection can be represented as

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & s & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \tag{3.2}$$

In general a camera coordinate frame and a world coordinate frame are not coincide. The equation becomes

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R}^T & -\mathbf{R}^T t \\ 0_3^T & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \tag{3.3}$$

where $\mathbf{R}$ is a 3×3 rotation matrix and $t = [t_x, t_y, t_z]^T$ is a translation vector.

The equation can be simplified to

$$\begin{aligned} m &= \mathbf{K}[\mathbf{R}| - \mathbf{R}t]M \\ m &= \mathbf{P}M, \end{aligned} \tag{3.4}$$

where m and M denote image and world homogeneous coordinates respectively. K represents the effect on the projection known as intrinsic parameters. $[\mathbf{R}| - \mathbf{R}t]$ are called extrinsic parameters. P is camera projection matrix.

3.4.2 Camera Models

Three camera models, perspective, weak perspective, and orthographic models, which are widely used in computer vision, and two motion transformations, projective and affine ,associated with the three camera models are discussed in this section. Readers may find the terms are a bit confusing. This is due to the fact that two scientific fields, photogrammetry and mathematics, use different terms to describe the same transformations. For instance, people in photogrammetry use the term perspective projection to describe a general linear camera which happens to be in the same format as a projective transformation used in the mathematical field.

Perspective Projection A projective projection is the most general case. The projective transform or perspective camera can be represented by a mathematical equation as

$$\begin{aligned}\mathbf{T}_{projective} &= \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R_1^T & -R_1^T t_x \\ R_2^T & -R_2^T t_y \\ R_3^T & -R_3^T t_z \end{bmatrix} \\ &= \begin{bmatrix} f_x R_1^T + sR_2^T + c_x R_3^T & f_x D_x + sD_y + c_x D_z \\ f_y R_2^T + c_y R_3^T & +f_y D_y + c_y D_z \\ R_3^T & D_z \end{bmatrix},\end{aligned} \qquad (3.5)$$

where R_i^T is the $i-$th row of the rotation matrix $\mathbf{R}$ and $D_x = -R_1^T t_x$, $Dy = -R_2^T t_y$, $D_z = -R_3^T t_z$.

Image and world coordinates are related by

$$\hat{m} = \begin{bmatrix} \frac{(f_x R_1^T + sR_2^T)M + f_x D_x + sD_y}{R_3^T M + D_z} \\ \frac{f_y R_2^T M + f_y D_y}{R_3^T M + D_z} \end{bmatrix} + \begin{bmatrix} c_x \\ c_y \end{bmatrix}, \qquad (3.6)$$

where $\hat{m}$ denotes non-homogeneous image coordinates.

Weak-Perspective Projection A weak perspective camera can be represented in a mathematical form as

$$\mathbf{T}_{affine} = \begin{bmatrix} f_x R_1^T + sR_2^T & f_x D_x + sD_y + c_x D_z \\ f_y R_2^T & f_y D_y + c_y D_z \\ \mathbf{0}_3^T & D_z \end{bmatrix}. \qquad (3.7)$$

Equation 6.3 has the similar form as a general affine transform. Hence, it is termed affine projection. Some call it affine camera or weak-perspective camera. Image and world coordinates are then related by

$$\hat{m} = \begin{bmatrix} \frac{(f_x R_1^T + s R_2^T)\mathbf{M} + f_x D_x + s D_y}{D_z} \\ \frac{f_y R_2^T \mathbf{M} + f_y D_y}{D_z} \end{bmatrix} + \begin{bmatrix} c_x \\ c_y \end{bmatrix}. \tag{3.8}$$

More detail including the mathematical proof of an affine camera can be found in section 6.2.

Orthographic Projection Orthographic is the simplest case of a camera projection. It projects points along z-axis. The projection can be represented as

$$\mathbf{T}_{orthograpic} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{3.9}$$

The three projection models are illustrated in Fig. 3.4. An orthographic projection has five degree of freedom, three parameters for a rotation matrix and two parameters for the displacements. It is suitable for the case where a world plane and an image plane are parallel. An affine camera has eight degree of freedom corresponding to the eight non-zero element in a matrix. It is appropriate for a distant camera or a large focal length camera. A general projective camera has eleven degree of freedom, defined up to an arbitrary scale. It is a general definition for any linear camera.

3.5 Affine Structure From Motion

An affine structure from motion theorem is first proposed by Koenderink and Van Doorn [74]. They have shown that two distinct views are enough to reconstruct a scene up to an arbitrary affine transformation without a camera calibration. They have suggested the use of local coordinate frame (LCF). Later their algorithm has been refined by Quan et.al. [75], Demy et.al. [76], and Shapiro [77]. Then, Tomasi and Kanade [78] proposed an affine factorization method which eliminates the use of LCF and instead utilize the entire set of points. This section will review the affine

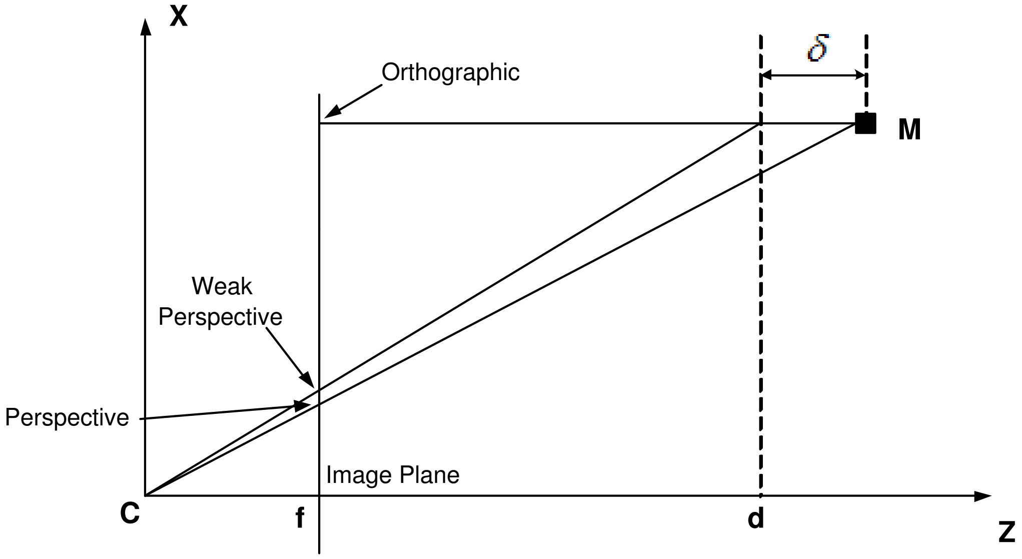


Figure 3.4: One-dimensional image formation. Orthographic is projected along z-axis. Perspective is projected along principal ray direction. Weak perspective is a combination between perspective and orthographic. A point, first, is projected along z-axis to a plane $Z = d$. Then, perspective projection from the plane.

structure from motion. Furthermore, we categorize affine structure from motion into two main categories, two view geometry and multiple view geometry.

3.5.1 Two View Geometry

Geometric Approach: Local Coordinate Frame Assume n general points in a scene. Four non-coplanar scene points, M_i, $i \in 0, \dots 3$, with M_0 being a reference point can be considered as defining a 3-D affine basis. Define axis vectors $E_i = M_i - M_0$, $i \in 0, \dots 3$ which are called the local coordinate frame (LCF). Any points in a scene can be expressed as follow

$$M_i = M_0 + \alpha_i E_1 + \beta_i E_2 + \gamma_i E_3, \quad i \in 1, \dots n-1, \tag{3.10}$$

where α, β, and γ are affine invariant coordinates. The prove is given next.

Under a 3-D affine transformation,

$$\acute{M}_i = \mathbf{R}M_i + T, \tag{3.11}$$

where $\acute{M}$ is the new world position, $\mathbf{R}$ is a 3×3 matrix, and T is a 3-vector.

Then,

$$\begin{aligned} \acute{M}_i - \acute{M}_0 &= \mathbf{A}(M_i - M_0) \\ &= \alpha_i \mathbf{A}E_1 + \beta_i \mathbf{A}E_2 + \gamma_i \mathbf{A}E_3 \\ &= \alpha_i \acute{E}_1 + \beta_i \acute{E}_2 + \gamma_i \acute{E}_3. \end{aligned} \tag{3.12}$$

Equation 3.12 demonstrates that α, β, and γ are indeed affine invariant coordinates. The affine coordinates are independent of frame. Let's extend the idea into image plane under affine projection.

$$\begin{aligned} m &= \mathbf{A}M + d \\ \acute{m} &= \acute{\mathbf{A}}M + \acute{d}, \end{aligned} \tag{3.13}$$

where m and $\acute{m}$ are 2-vector image image coordinates from first and second views respectively. $\mathbf{A}$ and $\acute{\mathbf{A}}$ are general 2×3 matrices. d and $\acute{d}$ are general 2×1vectors. From Equations 3.11, 3.12, 3.13, and the differences of vectors eliminate addition terms, T and d, we get

$$\begin{aligned} m_i - m_0 &= \alpha_i e_1 + \beta_i e_2 + \gamma_i e_3 \\ \acute{m}_i - \acute{m}_0 &= \alpha_i \acute{e}_1 + \beta_i \acute{e}_2 + \gamma_i \acute{e}_3, \end{aligned} \tag{3.14}$$

where $e_i = \mathbf{A}E_i$, $i \in 0, \ldots, 3$ and $\acute{e}_i = \acute{\mathbf{A}}\mathbf{R}E_i$, $i \in 0, \ldots, 3$.

Although it seems redundant to represent image coordinates with three basis, the extra basis allows 3-D affine coordinates, (α, β, γ), to be computed. The solution may be obtained by minimizing a cost function in a least mean square sense.

Algebraic Approach: Fundamental Matrix Given two affine views, the relationship between them can be defined as follows [79] [80]

$$a\acute{x}_i + b\acute{y}_i + cx_i + dy_i + e = 0, \tag{3.15}$$

where $m_i = [x_i, y_i, z_i]^T$, $\acute{m}_i = [\acute{x}_i, \acute{y}_i, \acute{z}_i]$, and a, b, c, d, e are unknown parameters. This equation is termed affine epipolar constraint. Rearrange Equation 3.15 in a matrix format

$$\acute{m}_i^T \mathbf{F}_A m_i = \begin{bmatrix} \acute{x}_i & \acute{y}_i & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & a \\ 0 & 0 & b \\ c & d & e \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = 0. \tag{3.16}$$

F_A is called affine fundamental matrix which is an algebraic representation of affine epipolar geometry. The epipolar geometry is the geometry between two views.

3.5.2 Multiple View Geometry

Local Coordinate Frame Local Coordinate Frame explained in Section 3.5.1 can be extended to accommodate multiple view geometry. Assume n features appear in f distinct views.

$$\mathbf{W} = \mathbf{LS}, \tag{3.17}$$

where $\mathbf{W}$ is a $2f \times (n-1)$ matrix containing the observations, $\mathbf{L}$ is a $2f \times 3$ matrix containing the affine basis, and $\mathbf{S}$ is an unknown $3 \times (n-1)$ matrix containing the affine structure.

Affine Factorization Tomasi and Kanade [78] has proposed an affine factorization method which eliminate the use of LCF and utilize the whole set of point correspondences. Suppose there are f affine views and n point correspondences from each view.

$$\hat{m}_i = \mathbf{A}_i \hat{M} + d_i, \quad i \in 1, \dots f, \tag{3.18}$$

where $\hat{m}$ and $\hat{M}$ denotes image and world non-homogeneous coordinates respectively. $\mathbf{A}$ is an arbitrary 2×3 matrix and d represents any 2×1 vector.

With the similar idea of LCF, we need to select one point as an origin or a center of mass. The Equation 3.18 becomes

$$\begin{aligned} \triangle\hat{m}_i &= \mathbf{A}_i(\triangle\hat{M}), \quad i \in 1, \dots f \\ \mathbf{W} &= \mathbf{LS}, \end{aligned} \tag{3.19}$$

$\mathbf{W}$ denotes a $2f \times n$ matrix containing set of 2D point correspondences with respect to the center of mass. $\mathbf{S}$ denotes a $3 \times n$ matrix containing affine shape. $\mathbf{L}$ denotes $2f \times 3$ matrix.

With rank theorem, $\mathbf{W}$ is at most rank three. Singular value decomposition (SVD) is used to factorized $\mathbf{W} = \mathbf{UWV}^T$. Therefore, $\mathbf{L}$ and $\mathbf{S}$ are the left and right eigenvectors corresponding to the three greatest eigenvalues.

$$\begin{aligned} \mathbf{L} &= U_3 \\ \mathbf{S} &= W_3 V_3^T. \end{aligned} \tag{3.20}$$

Concatenated Image Space Shapiro [77] has provided an insight geometrical meaning of an affine factorization method proposed by Tomasi and Kanade [78] in terms of concatenated image space (CI space). For simplification, assume there are 2 views. If $\mathbf{L}$ in Equation 3.19 is decomposed into three columns $\mathbf{L} = [\mathbf{l}_1|\mathbf{l}_2|\mathbf{l}_3]$, then Equation 3.19 can be rewritten as

$$w_i = \triangle X_i \mathbf{l}_1 + \triangle Y_i \mathbf{l}_2 + \triangle Z_i \mathbf{l}_3. \tag{3.21}$$

Shapiro [77] has indicated that

"the 4-dimensional w_i is a linear combination of the three column vectors, and lies on a hyperplane π in the 4-dimensional CI space. This hyperplane is spanned by the three columns of $\mathbf{L}_i$ and its orientation depends solely on the motion of the camera, while the distribution of points within the hyperplane depends solely on the scene structure (as shown in Fig. 3.5)."

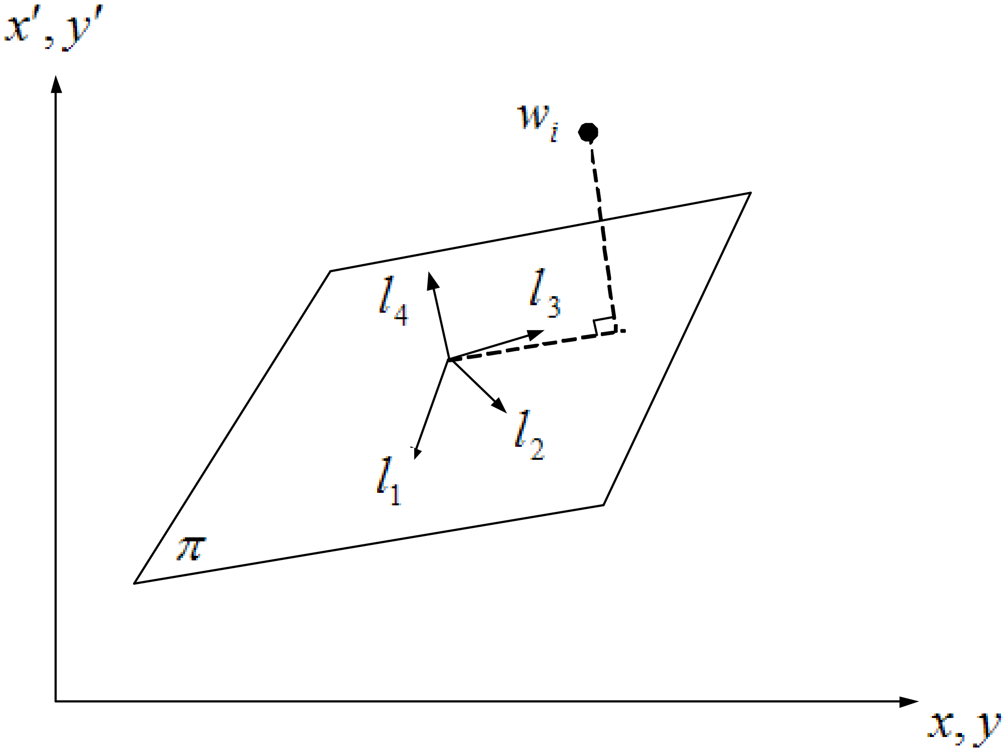


Figure 3.5: The concatenated image (CI) space.

If the images are noise free, then w_i would lie exactly on the hyperplane π. In practice, noise relocates w_i from hyperplane π. Hyperplane π which best fitted the w must be identified. If noise distribution is assumed to be zero mean, isotropic and Gaussian, then the maximum likelihood estimation of optimal hyperplane is found by minimizing

$$\Sigma_{i=0}^{n}(w_i - Ls_i)^T \Lambda_{w_i}^{-1}(w_i - Ls_i). \tag{3.22}$$

Because points are independent of each other, we simply assume $\Lambda_{w_i} = \sigma^2 I$ for all i. In this case, the above equation becomes

$$\min_{L,s} \Sigma_{i=0}^{n}(w_i - Ls_i)^2. \tag{3.23}$$

This is exactly what affine factorization does. The CI space and the affine coordinate frame give insights into the physical concepts of the affine factorization method. The cameras' motion can be represented by a hyperplane and its orientation. While the 3-D points, in an ideal case, should be lain on the hyperplane.

3.6 Conclusions

In this chapter, we have reviewed the mathematical background of the proposed research, and the materials presented here serve the theoretical foundation of latter chapters. Specifically, we will address 2-D retinal image registration in Chapter 5 that requires 2-D/2-D transformations, and 3D retinal reconstruction is discussed in Chapter 6 that involves a camera projection (3-D/2-D transformations) and 3-D registration (3-D/3-D transformations). Also, as the prerequisite of all geometrical transformations discussed above, feature points or correspondences have to be extracted first that is the focus of the next chapter.

Table 3.1: 2-D Transformation Models

Models	Transformation Models	DOF
Linear Transformations		
Translation	$\begin{pmatrix} \acute{x} \\ \acute{y} \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$	2
Rigid	$\begin{pmatrix} \acute{x} \\ \acute{y} \\ 1 \end{pmatrix} = \begin{pmatrix} sr_{11} & sr_{12} & t_x \\ sr_{21} & sr_{22} & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$	4
Affine	$\begin{pmatrix} \acute{x} \\ \acute{y} \\ 1 \end{pmatrix} = \begin{pmatrix} \theta_{11} & \theta_{12} & t_x \\ \theta_{21} & \theta_{22} & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$	6
Projective	$\begin{pmatrix} \acute{x} \\ \acute{y} \\ 1 \end{pmatrix} = \begin{pmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \\ \theta_{31} & \theta_{32} & \theta_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$	8
Nonlinear Transformations		
Quadratic	$\begin{pmatrix} \acute{x} \\ \acute{y} \end{pmatrix} = \begin{pmatrix} \theta_{11} & \theta_{12} & \theta_{13} & \theta_{14} & \theta_{15} & \theta_{16} \\ \theta_{21} & \theta_{22} & \theta_{23} & \theta_{24} & \theta_{25} & \theta_{26} \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \\ x^2 \\ y^2 \\ xy \end{pmatrix}$	12

Table 3.2: 3-D Transformation Models

Models	Transformation Models	DOF
Euclidean	$\begin{pmatrix} \acute{X} \\ \acute{Y} \\ \acute{Z} \\ 1 \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$	6
Similarity	$\begin{pmatrix} \acute{X} \\ \acute{Y} \\ \acute{Z} \\ 1 \end{pmatrix} = \begin{pmatrix} sr_{11} & sr_{12} & sr_{13} & t_x \\ sr_{21} & sr_{22} & sr_{23} & t_y \\ sr_{31} & sr_{32} & sr_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$	7
Affine	$\begin{pmatrix} \acute{X} \\ \acute{Y} \\ \acute{Z} \\ 1 \end{pmatrix} = \begin{pmatrix} \theta_{11} & \theta_{12} & \theta_{13} & t_x \\ \theta_{21} & \theta_{22} & \theta_{23} & t_y \\ \theta_{31} & \theta_{32} & \theta_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$	12
Projective	$\begin{pmatrix} \acute{X} \\ \acute{Y} \\ \acute{Z} \\ 1 \end{pmatrix} = \begin{pmatrix} \theta_{11} & \theta_{12} & \theta_{13} & \theta_{14} \\ \theta_{21} & \theta_{22} & \theta_{23} & \theta_{24} \\ \theta_{31} & \theta_{32} & \theta_{33} & \theta_{34} \\ \theta_{41} & \theta_{42} & \theta_{43} & \theta_{44} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$	15

CHAPTER 4

Feature Extraction

4.1 Introduction

We want to extract features because features can reduce the amount of data needed to be processed. Since "Not all information is created equal [19]", what are the most suitable features for matching? Points are used most of the time. Lines or blobs are also good features since they provide more information compared to points. Good features should contain as much distinguishable information as possible. The next question would be what are the best approaches to extract features? Feature extraction are pretty much image-dependent problems. Certain algorithms are good for some specific types of images but perform poorly on other image types. Several researches have been entirely devoted to this topic.

In this research, vascular tree and its bifurcation/crossover points are used as feature and correspondences. This is due to the two following reasons: (1) vascular tree spans the whole retina hence it exists in every retinal images; (2) bifurcation/crossover points offer more distinguishable information than homogeneous area throughout the retina. The automatic detection of blood vessels in the retinal images can help physicians for the purposes of diagnosing ocular diseases, patient screening, and clinical study, etc. Information about blood vessels in retinal images can be used in grading disease severity or as part of the process of automated diagnosis of diseases. Blood vessel appearance can provide information on pathological changes caused by some diseases including diabetes, hypertension, and arteriosclerosis. The most effective treatment for many eye-related diseases is the early detection through regular screenings.

The organization of this chapter is as follows. The importance of blood vessel detection in medical applications are given in this section. Currently available different methods for the detection of blood vessels have been reviewed in section 4.1.2. Then, Section 4.3 describes the implementation of the proposed algorithm. Simulation results and the performance of our algorithm are presented in section 4.4.

4.1.1 Entropy Thresholding

Pun [81, 82] was the first to adopt Shannon's information theory [83] in image thresholding applications. Pun's thresholding method, entropy was defined by only its gray level histogram. Pun only considered one probability distribution. Kapur et.al. [84], later, extended the idea in Pun's by considering two probabilities distributions, one for object and one for background, to binarize an image into foreground and background. Those two methods, however, did not take spatial relationship or image structure into its entropy. The two images with identical histogram will always result in the same threshold value. Pal et.al. [85] was the first to propose the use of local entropy thresholding for image thresholding applications which takes the spatial distribution or image structure into consideration. Instead of a gray-level histogram, Pal et.al. [85, 86] proposed two-dimensional histogram called a co-occurrence matrix. Elements in co-occurrence matrix represented spatial distribution in an image. Chang et.al. [87] proposed a relative entropy (cross entropy) approach for image thresholding applications. The method involved Kullback-Leiber distance in finding a threshold value that minimize the mismatch between a gray-scale image and a binarized image. In this work, we combine the concepts proposed by Pal [85, 86] and Chang et.al. [87] and propose a new thresholding method. We also define a new definition for a co-occurrence matrix.

4.1.2 Blood Vessel Extraction

There were many previous works on extracting blood vessels in retinal images which can be found in chapter 2: literature review. Here we'll mention a few of these methods that have been considered to be state-of-the-art methods at the time they were proposed and have been cited by almost every retinal vascular tree extraction papers. An efficient piecewise threshold probing technique was proposed by Hoover et.al. [9] where the matched-filter-response (MFR) image was used for mapping the vascular tree. A set of criteria was tested to determine the threshold of the probe region, and ultimately to decide if the area being probed was a blood vessel. Since the MFR image was probed in a spatially adaptive way, different thresholds can be applied throughout the image for mapping blood vessels. Jiang et.al. [8] used verification-based multiple threshold probing framework. A retinal image was probed at different threshold values. At a particular threshold, Euclidean distance transform was performed. Then, vessel candidates were pruned by means of the distance map to only retain centerline pixels. Finally blood vessels were reconstructed by the particular threshold. Staal et.al. [6] used supervised method. Twenty hand-labeled ground-truth were used as a training set and twenty images were used for validation. Ridge detection was employed to locate candidate blood vessel segments. Then, ridge pixels were groups and convex sets are formed. After that, the image was partitioned into patches based on the convex set. Every pixel was assigned to the convex set to which it was closest. Next, feature sets were formed and kNN-classifier was employed.

4.2 Preliminaries

The entropy is defined as a function of the state probability [83].

Theorem: Let $p(s_i)$ be the probability of a sequence s_i of gray levels of length q, where a sequence s_i of length q is defined as a permutation of q gray levels.

$$H^{(q)} = -\frac{1}{q}\sum_i p(s_i)\log_2 p(s_i). \tag{4.1}$$

Where the summation is taken over all gray level sequences of length q. Then $H^{(q)}$ is a monotonic decreasing function of (q) and $lim_{q\to\infty}H^{(q)} = H$, the entropy of the image. For different values of q, we get various orders of entropy.

In the case of an image, the global entropy or $H^{(1)}$ is not an adequate measure since image pixel intensities are not independent of each other. Different images with identical histograms will always yield the same value of entropy since the spatial distribution is not taken into account in the global entropy computation. The local entropy or $H^{(2)}$ [85], on the other hand, can better distinguish two images in terms of their spatial structures, because it considers the dependency of image pixel intensities. The local entropy os defined as

$$H^{(2)} = -\frac{1}{2}\sum_i\sum_j p_{ij}\log_2 p_{ij}, \tag{4.2}$$

where p_{ij} is the probability of co-occurrence of the gray levels i and j. The co-occurrence matrix is an $L \times L$ dimensional matrix (L is the number of intensity levels) [85]. It indicates the transition of pixel intensities between adjacent pixels. The original co-occurrence matrix proposed by Pal [85, 86] is asymmetric by considering the horizontally right and vertically lower transitions. t_{ij} is defined as follows:

$$t_{ij} = \sum_{l=1}^{P}\sum_{k=1}^{Q}\delta, \tag{4.3}$$

where
$$\delta = 1 \quad \text{if} \begin{cases} I(l,k) = i \quad \text{and} \quad I(l+1,k) = j \\ \qquad\qquad\quad \text{or} \\ I(l,k) = i \quad \text{and} \quad I(l+1,k+1) = j \end{cases}$$
$$\delta = 0 \quad \text{otherwise.}$$

Therefore, the local entropy can preserve the structure details of an image. Two images with identical histograms but different spatial distribution will result in different local entropy (also different threshold values).

4.3 Proposed Framework

In this paper, we propose a framework to efficiently locate and extract blood vessels in ocular fundus images. The proposed algorithm is composed of four steps, matched filtering, modified local entropy thresholding, length filtering, and bifurcation/crossover point detection. Compare with the method in [24], our proposed algorithm does not involve human intervention. Since our algorithm can automatically estimate one optimal threshold value, it requires less computational complexity compared with the methods in [9], [8], and [6]. In addition, our method doesn't require additional information, training database as it is mandatory in [6].

4.3.1 Vascular Tree Extraction

Because of non-uniform intensity of both background and foreground as well as the low contrast, multi-directional match filters are employed. We observe three interesting properties of the blood vessels in retinal images. (1) Two edges of a vessel always run parallel to each other. Such objects may be represented by piecewise linear directed segments of finite width. The gradient directions for these two edge elements are 180^o apart and hence they are sometimes referred to as "anti-parallels". (2) The contrast between vessels and other retinal surfaces are very low. Blood vessels appear darker relative to the background. Sample of blood vessel gray-level profile along direction perpendicular to their length is plotted in Fig. 4.1. It was observed that the vessels never have ideal step edges. Although the intensity profile varies by a small amount from vessel to vessel, it can be approximated by a Gaussian curve. (3) Although the width of a vessel decreases as it travels radially outward from the optic disc, such a change in vessel caliber is a gradual one. The widths of the vessels are found to lie within a range $36 - 180\mu m$. For our initial calculation, we assume that all the blood vessels in the images are of equal width.

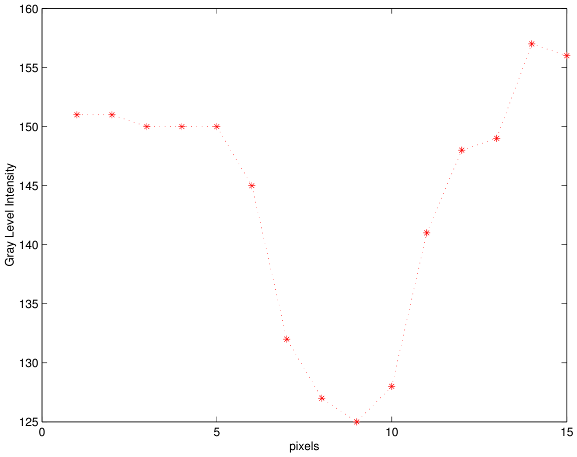


Figure 4.1: The gray-level profile of the cross section of a blood vessel.

Match Filter. In [34], the gray-level profile of the cross section of a blood vessel can be approximated by a Gaussian shaped curve. The concept of matched filter detection is used to detect piecewise linear segments of blood vessels in retinal images. Blood vessels usually have poor local contrast. The two-dimensional matched filter kernel is designed to convolve with the original image in order to enhance the blood vessels. A prototype matched filter kernel is expressed as

$$f(x,y) = -\exp(\tfrac{-x^2}{2\sigma^2}), \text{ for } |y| \leq L/2, \tag{4.4}$$

where L is the length of the segment for which the vessel is assumed to have a fixed orientation. The parameter L is chosen to be equal to 9 pixels. Here the direction of the vessel is assumed to be aligned along the y-axis. Because a vessel may be oriented at any angles, the kernel needs to be rotated for all possible angles. Assuming an angular resolution of 15^o, twelve different kernels have been constructed to span all possible orientations (Fig. 4.2). A set of twelve 16x15 pixel kernels is applied by convolving to a fundus image and at each pixel only the maximum of their responses is retained.

For example, given a retinal image in Fig. 4.5(a) which has low contrast between blood vessels and background , its MFR version is shown in Fig. 4.5(b), where we can see blood vessels are significantly enhanced.

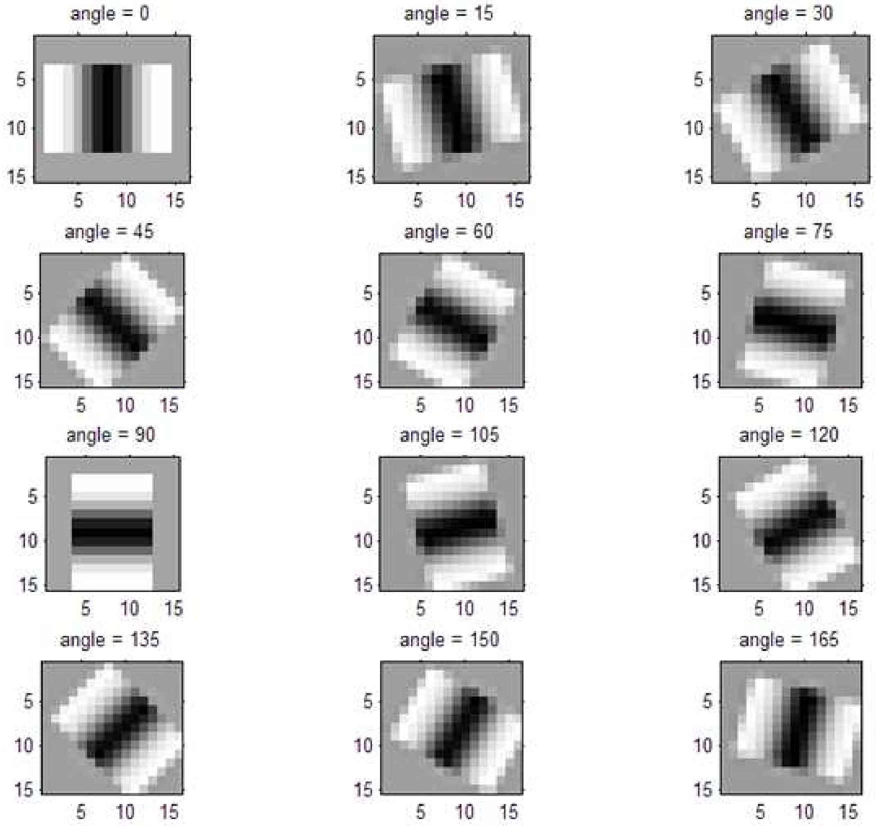


Figure 4.2: Illustration of 12 matched filter kernels along different directions where $\sigma = 2.0$.

Modified Local Entropy Thresholding Algorithm As the second step, the MFR image is processed by a proper thresholding scheme, which can be used to distinguish between enhanced vessel segments and the background. An efficient entropy-based thresholding algorithm, which takes into account the spatial distribution of gray levels, is used because an image pixel intensities are not independent of each other. We, specifically, implement a thresholding technique which is a blend of a local entropy thresholding [85] and a relative (or cross) entropy thresholding [87]. In a match-filtered retinal image, enhanced blood vessels are very sparse compared with the uniform background. This leads to a highly peaky co-occurrence matrix with a low entropy that is not appropriate for local entropy thresholding. Also, the local entropy thresholding aims to maximize the local entropy of foreground and background without considering the unbalanced proportion between them. Therefore, blood vessels extracted by local entropy thresholding are usually not complete, and some detailed structures are missed. We made two modifications to improve the results of blood vessel extraction.

First, we develop a *smoothed* co-occurrence matrix to increase the entropy and to reduce the peak in the co-occurrence. The co-occurrence matrix of an image show the intensity transitions between adjacent pixels. The original co-occurrence matrix is asymmetric by considering the horizontally right and vertically lower transitions. We want to add some jittering effect to the co-occurrence matrix that tends to keep the similar spatial structure but with less variations, i.e., $T = [t_{i,j}]_{N \times N}$ is computed as follow (Fig. 4.3(a)).

```
• For every pixel $(l, k)$ in an image $I$
    - $i = I(l, k)$;
    - $j = I(l, k + 1)$;
    - $d = I(l + 1, k + 1)$;
    - $t_{ij} = t_{id} + 1$;
• End
```

Fig. 4.3(b) and (c) compare the original co-occurrence matrix and the modified one for a typical match-filtered image. Two matrices still share a similar structure that is important for the valid thresholding result. Also, the latter one has more entropy with a much smaller standard deviation, which is more desirable for local entropy

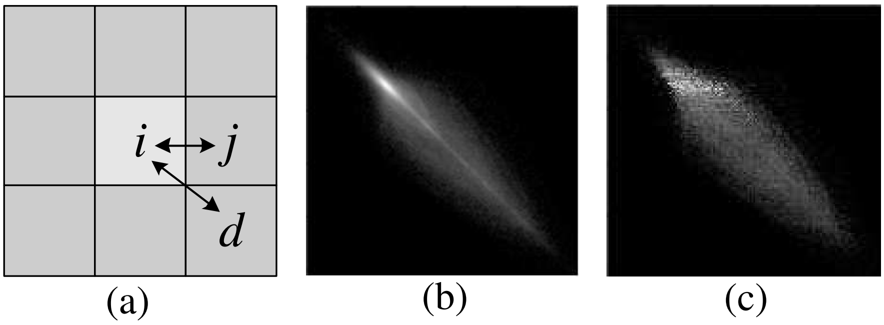


Figure 4.3: (a) The computation of the new co-occurrence matrix. (b) The original co-occurrence matrix in a normalized log scale ($\sigma = 261.63$). (c) The a modified co-occurrence matrix in a normalized log scale $\sigma = 9.00$).

thresholding. One may wonder whether the modified co-occurrence matrix still well represent the original spatial structure. Actually, considering a smooth area where j and d are very close or identical, the computation in Fig. 4.3(a) implicitly introduces certain low-pass filtering effect and some structured noise to the co-occurrence matrix.

Second, we want to preserve the complete vascular tree structure after thresholding by modifying the original threshold selection criterion. The threshold selected by the local entropy aims to maximize the local entropy of foreground and background without considering the small proportion of the foreground compared to the background. Therefore, we propose to select the optimal threshold that maximizes the local entropy of the binarized image that tends to retain an appropriate foreground/background ratio. The larger the local entropy, the more balanced ratio between foreground and background. If s, $0 \leq s \leq L-1$, is a threshold, s can partition a co-occurrence matrix into four quadrants, namely A, B, C, and D (Fig. 4.4).

Then we define the local entropy of the binary image due to the foreground and the background as

$$H_b^{(2)}(s) = -P_A \log_2 P_A - P_C \log_2 P_C, \tag{4.5}$$

where P_A and P_B are the probability sums in quadrants A and C, respectively. s corresponding to the maximum of $H_b^{(2)}(s)$ is used as the optimal threshold for blood

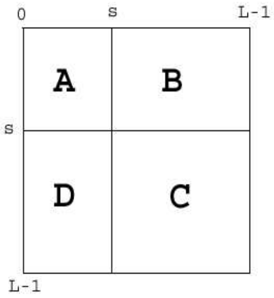
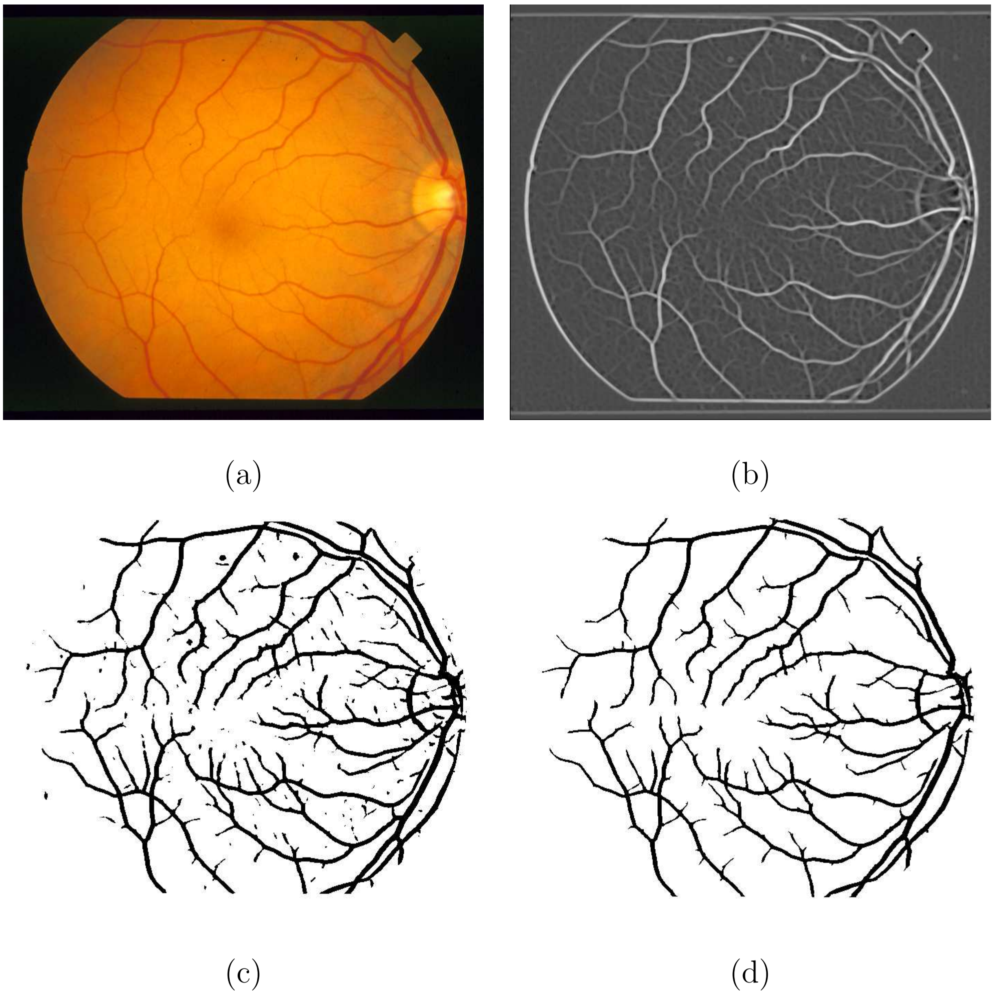


Figure 4.4: Four quadrants of a co-occurrence matrix.

vessel segmentation. As shown in Fig. 4.10, the modified local entropy thresholding algorithm can better preserve detailed blood vessels compared with the original one. For the MFR image shown in Fig. 4.5(b), the entropy-based thresholding result is shown in Fig. 4.5(c) where we can see blood vessels are clearly segmented from the background.

Length Filtering. As seen in Fig. 4.5(c), there are still some misclassified pixels in the image. Here we want to produce a clean and complete vascular tree structure by removing misclassified pixels. Length filtering is used to remove isolated pixels by using the concept of connected pixels labeling described in [88]. Connected regions correspond to individual objects. We first need to identify separate connected regions. The binary image is simply an array of '1's and '0's. The length filtering tries to isolate the individual objects by using the eight-connected neighborhood and label propagation. We assume the binary image contains pixels with value '1' on objects and '0' on background.

1. Set the current label $L = 1$;
2. Scan in a raster order from the top left to the bottom right;
3. If encountering a pixel '1', check its neighbors in the upper-left half of its 8-connected neighborhood(W, NW, N, NE).



(a) (b)

(c) (d)

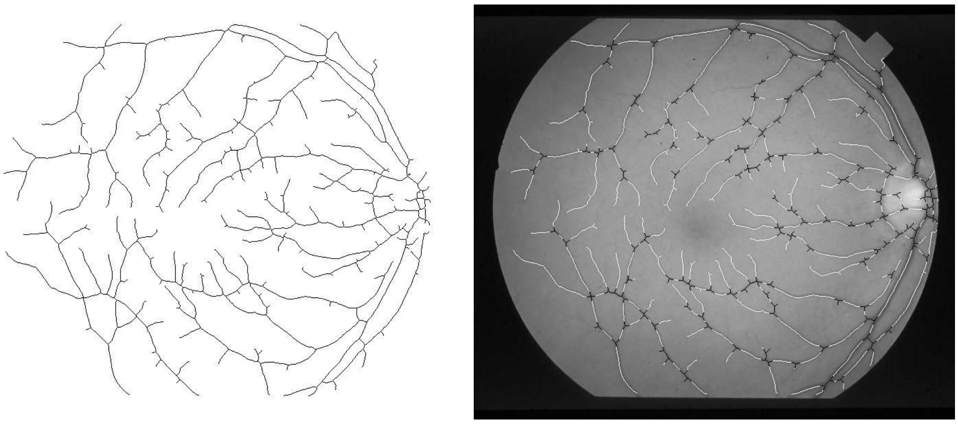
Figure 4.5: (a) An original retinal image. (b) Matched filtering result. (c) Local entropy thresholding result. (d) Vascular tree.

- If one of them have $L > 1$, set the current pixel's label to that value;
- Else If there is more than one label represented in the pixel's half-neighborhood, then these labels should be noted as ”equivalent”;
- Else set the current pixel's label to the current label; increment the current label.

4. Go to step 2;

5. Relabel all “equivalent” labelled pixels to the same value.

Once the algorithm is completed, only the resulting classes exceed a certain number of pixels, e.g., 250, are labeled as blood vessels. Classes, that are not labeled as blood vessels, are eliminated. Fig. 4.5(d) shows the results after length filtering based on Fig. 4.5(c), where a clean vascular tree is presented.



(a) (b)

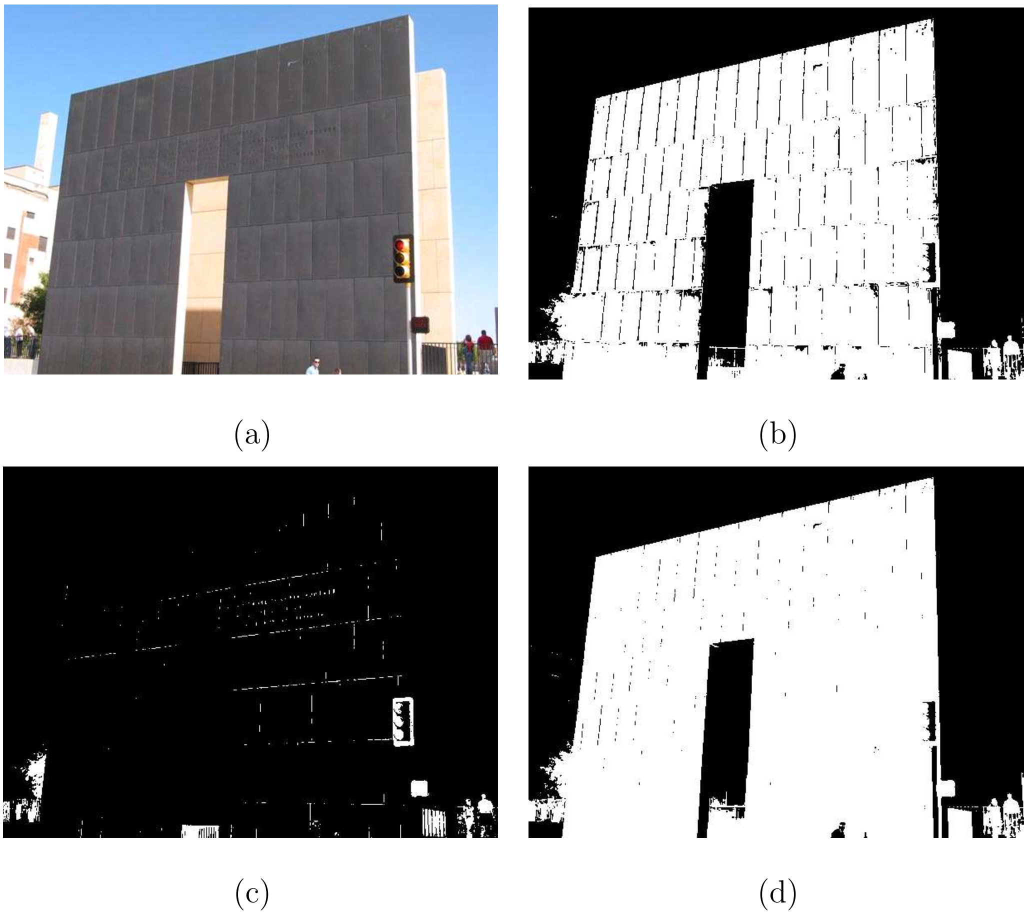
Figure 4.6: (a) One-pixel wide vascular tree. (b) One-pixel wide vascular tree with intersections and crossovers overlaying on gray-scaled image.

4.3.2 Bifurcation/crossover Detection

Vascular intersections and crossovers are the most appropriate representative features because (1) vascular tree spans the whole retina hence it exists in every retinal images; (2) bifurcation/crossover points offer more distinguishable information than other homogeneous areas throughout the retina. If a vascular tree is one-pixel wide, the branching points can be detected and characterized efficiently from the vascular tree. Morphological thinning is applied to the vascular tree in order to get one-pixel-wide vascular tree as shown in Fig. 4.6(a). In order to save computational time, a 3 × 3 neighborhood window is used to probe and find the branching points. If the number of vascular tree in the window is great than 3, it is a branch point. Then a 11 × 11 neighborhood is applied through a detected branching points in order to eliminate the small intersections. We consider only the boundary pixels of a 11 × 11 square. If the number of vascular tree on the boundary is greater than 2, we mark it as an intersection/crossover. Fig. 4.6(b) presents the vascular tree with the intersections and crossovers.

4.4 Experimental Analysis

4.4.1 Thresholding Algorithm



(a)

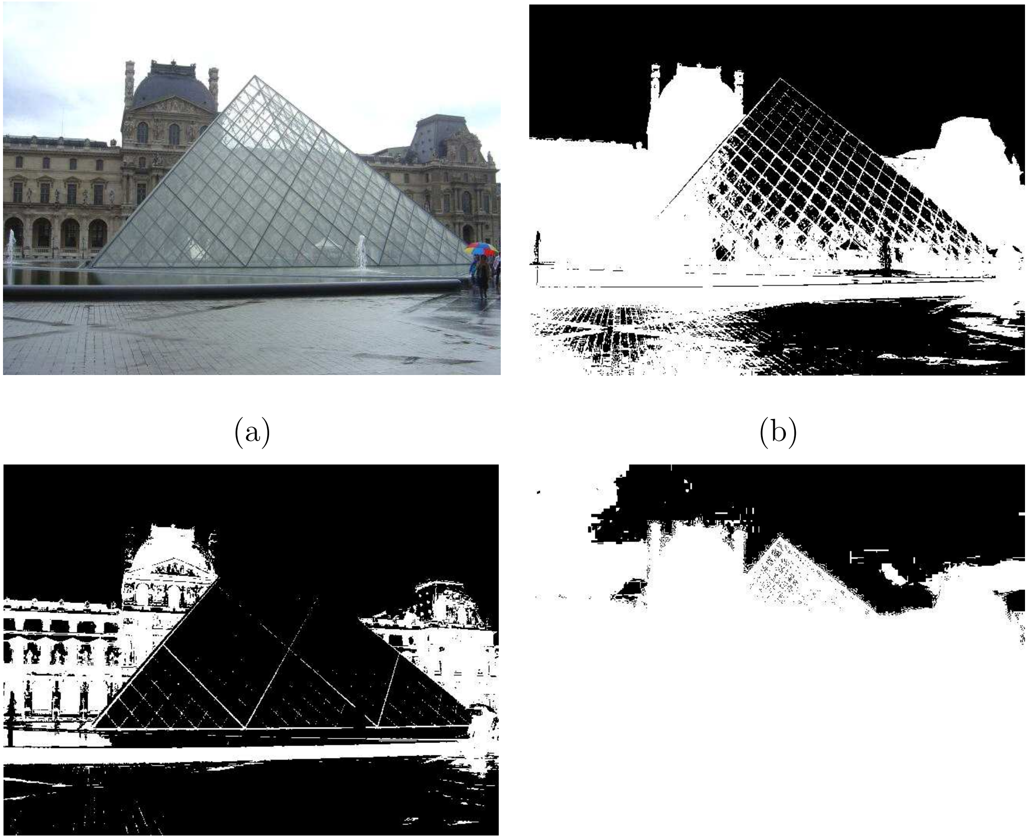
(b)

(c)

(d)

Figure 4.7: (a) An original image. (b) Our thresholding result (threshold = 101). (c) Local entropy thresholding result (threshold = 75). (d) Cross entropy result (threshold = 112).

A co-occurrence matrix is a representation of spatial relationship in an image.



(a) (b)

(c) (d)

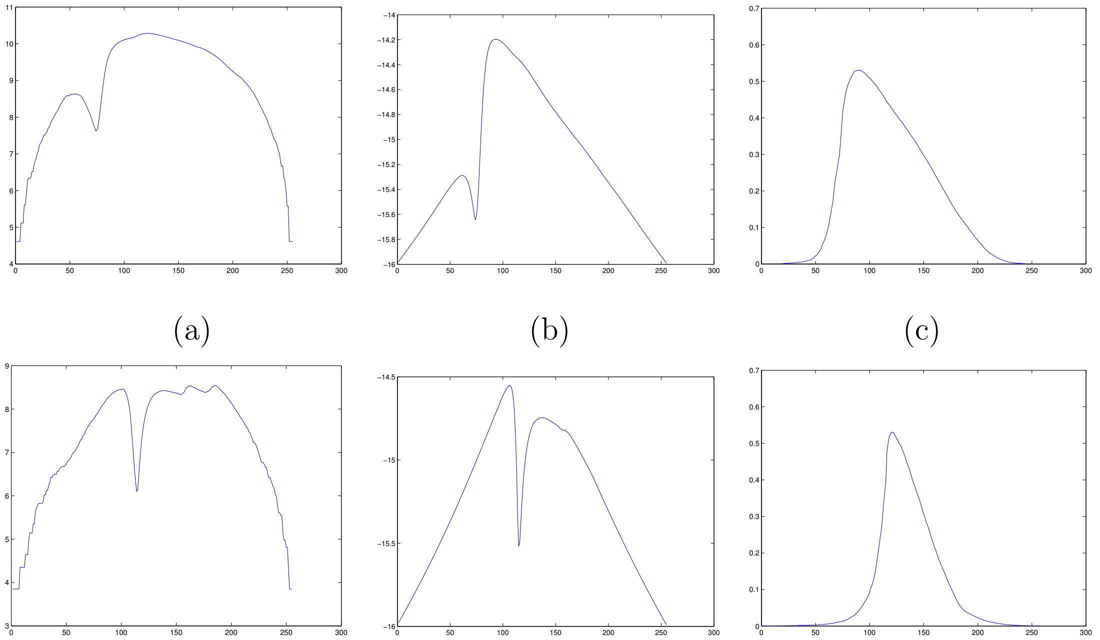
Figure 4.8: (a) An original image. (b) Our thresholding result (threshold = 136). (c) Local entropy thresholding result (threshold = 98). (d) Cross entropy result (threshold = 253).

Each element in a co-occurrence matrix gives an idea about the transition of intensities between adjacent pixels. Our proposed thresholding algorithm works well on any MFR images because a MFR image have a closer range of intensity changes compared with a normal gray-scale image. Therefore, in MFR images, an original co-occurrence matrix's definition produces very distinct peaks in such a narrow which can corrupt the power of a statistical method. On the other hand, our method spreading out the weight and reduce the range of standard deviation which results in a better threshold selection. We also compare our proposed algorithm with the other two entropy-based thresholding methods, namely local entropy thresholding [85, 86] and relative (or cross) entropy thresholding [87], on different types of images including retinal images. Fig. 4.7 and Fig.4.8 illustrate examples of simulation results on normal gray-scale images of our proposed algorithm versus the other two methods. Algorithm's performance depend on images. Table 4.1 demonstrates numerical results of the three approaches on MFR retinal image. Twenty retinal images provided by Hoover [89] are used to test the three entropy-based thresholding algorithms. Plots of three entropy-based thresholding algorithms are illustrated in Fig. 4.9 in which our algorithm always provides a distinct peak regardless of the image type. Examples of simulation results are shown in Fig. 4.10 for a normal retinal image and Fig. 4.11 for a retinal image with lesions. While the performance of cross entropy and our algorithm's performance are comparable in a normal image, our algorithm are more robust to lesions. In the MFR image, it is quite obvious that our algorithm's performance is better than the other two entropy-based methods.

4.4.2 Blood Vessel Extraction Evaluation

We use a set of twenty retinal images provided by Hoover [9] because both fundus and ground-truth images are available online [89] and other state-of-the-art algorithms [9, 8, 6] use Hoover database to evaluate their algorithms' performance. Therefore, not only we can evaluate and analyze the performance of the proposed framework but we also can impartially compare performance of the proposed algorithm with others.

Performance of blood vessel extraction is evaluated by using the true positive and the false positive rates as in [9]. Any pixel which was hand-labeled as vessel and the algorithm labeled as vessel is counted as true positive. Any pixel which was hand-labeled as non-vessel and the algorithm labeled as vessel is counted as false positive. The true positive rate is calculated by normalizing true positive by the total pixel number of the hand-labeled vessel. The false positive rate is calculated by normalizing false positive by the total pixel number of the hand-labeled non-vessel.



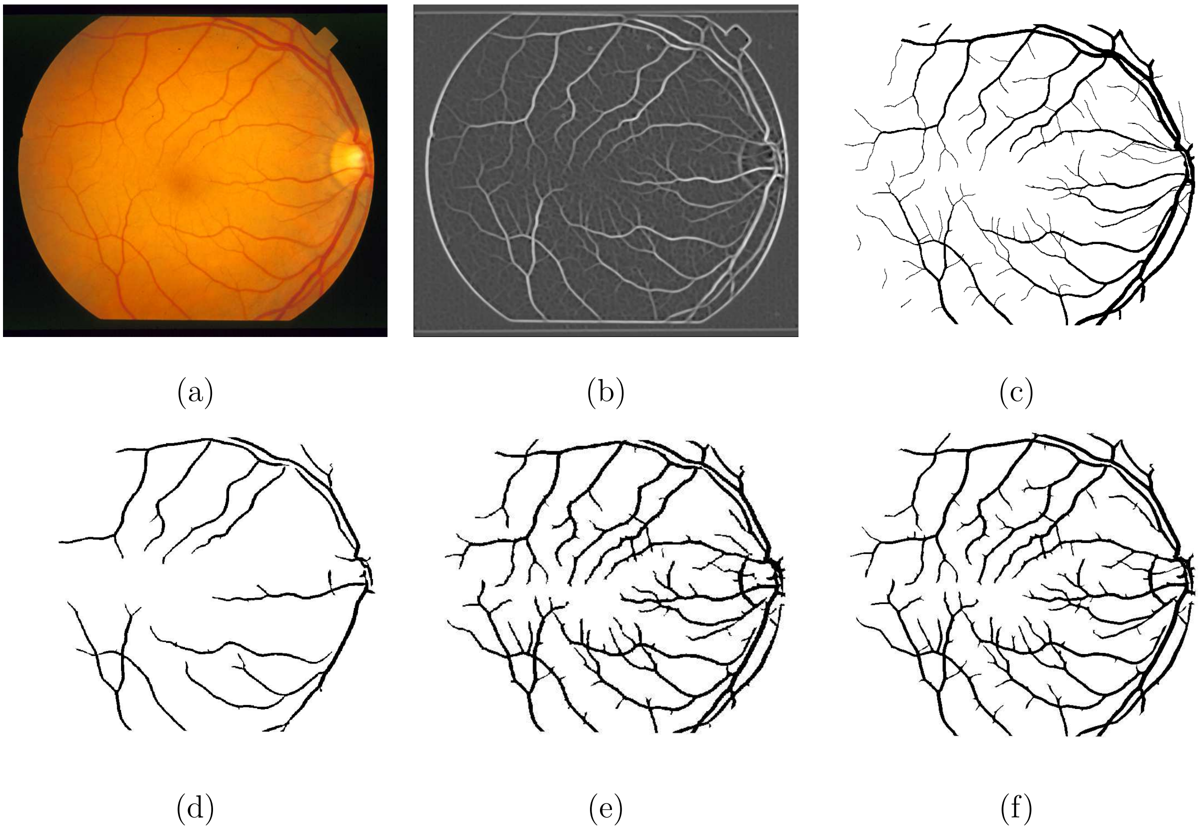
(a) (b) (c)

(d) (e) (f)

Figure 4.9: First row: entropy plots of an image shown in Fig. 4.10. Second row: entropy plots of an image shown in Fig. 4.11. (a),(d) Local entropy. (b),(e) Relative entropy. (c),(f) Modified local entropy.

Specifically, we classify retinal images into three categories, normal retinal images, abnormal retinal images with some lesions, and retinal images with obscure blood vessel appearance. Fig. 4.13 shows an example of simulation result from a normal retinal images. Fig. 4.14 presents an example simulation result from an obscure blood vessel appearance image. An example of simulation result from an abnormal retinal image with some lesions is shown in Fig. 4.15.

In order to evaluate the performance of our algorithm, we compare our simulation results with state-of-the-art results obtained from Hoover et.al. [9], Jiang et.al. [8], Staal et.al. [6, 7], and hand-labeled ground truth segmentations. Numerical performance of the proposed framework and other methods are demonstrated in Fig. 4.12. The best quantitative performance belongs to, as predicted, a normal image category. The numerical performances of a group of obscure blood-vessel appearance and lesion


(a) (b) (c)

(d) (e) (f)

Figure 4.10: (a) An original retinal image. (b) A matched filtering result. (c) A ground-truth segmentation. (d) The local entropy thresholding result. (e) The relative entropy thresholding result. (f) The modified local entropy threshold.

images are comparable. In lesion image category, although the proposed algorithm misdetect lesions as blood vessel, the false positive rate of a lesion image group is lower and/or comparable to other approaches. This is due to the fact that lesions only account for small portions in an image. Although segmentation performance is decreased with presence of lesions, the additional detected lesions are actually good for 2-D registration and 3-D surface reconstruction because those lesions are used as additional candidate features. Overall performance of our algorithm is comparable to other computational expensive algorithms. While the other three methods require multiple threshold values and various decision-making criteria, we believe one single threshold yield better optimal segmentation results. If an image is partitioned into

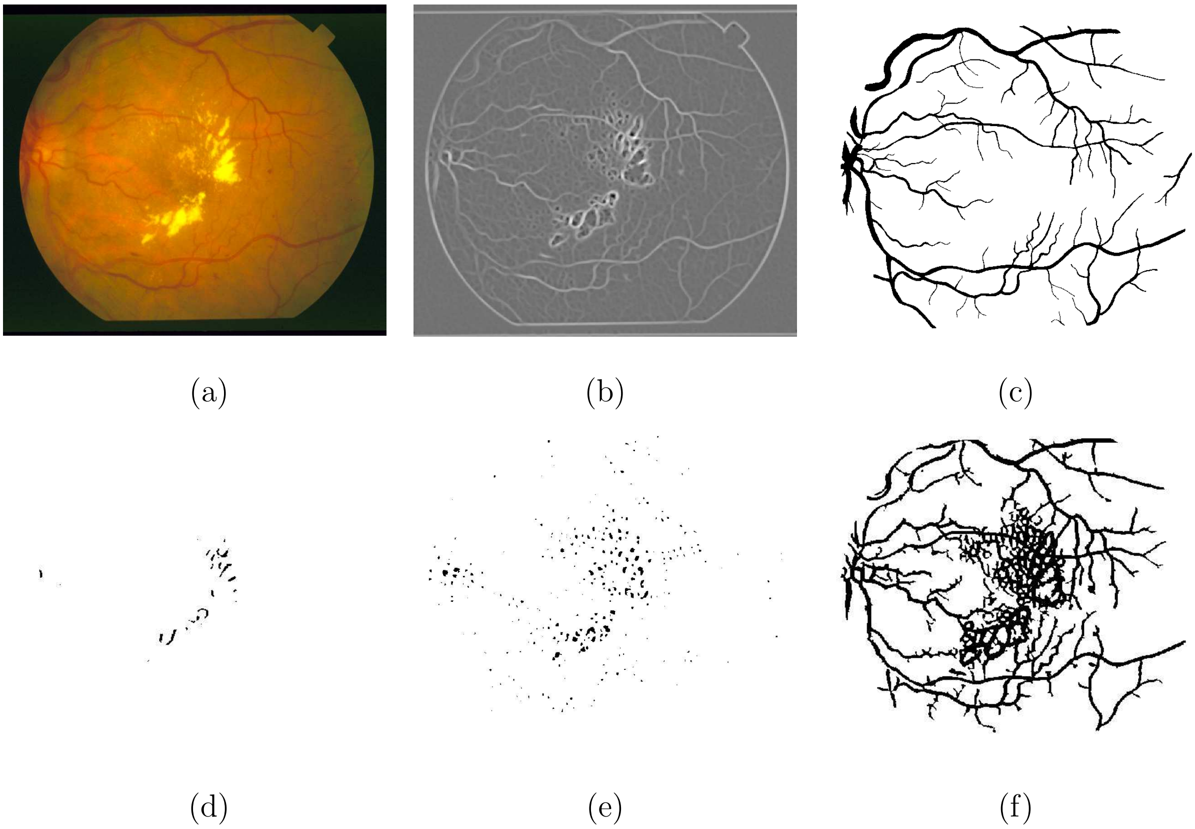


Figure 4.11: (a) An original retinal image. (b) A matched filtering result. (c) A ground-truth segmentation. (d) The local entropy thresholding result. (e) The relative entropy results. (f) The modified local entropy threshold.

multiple small regions and one particular small region is given, even by human eyes it is difficult to distinguish between foreground (blood vessel) and background in that small region. This remark may not be so apparent in Hoover database, but in high-resolution images, e.g. ETDRS database, this observation is quite obvious since there are background patterns which highly resemblance blood vessel structures all over the retinal images.

4.5 Conclusions

We have introduced an efficient and robust algorithm for blood vessel detection in ocular fundus images with a modified co-occurrence matrix used for local entropy

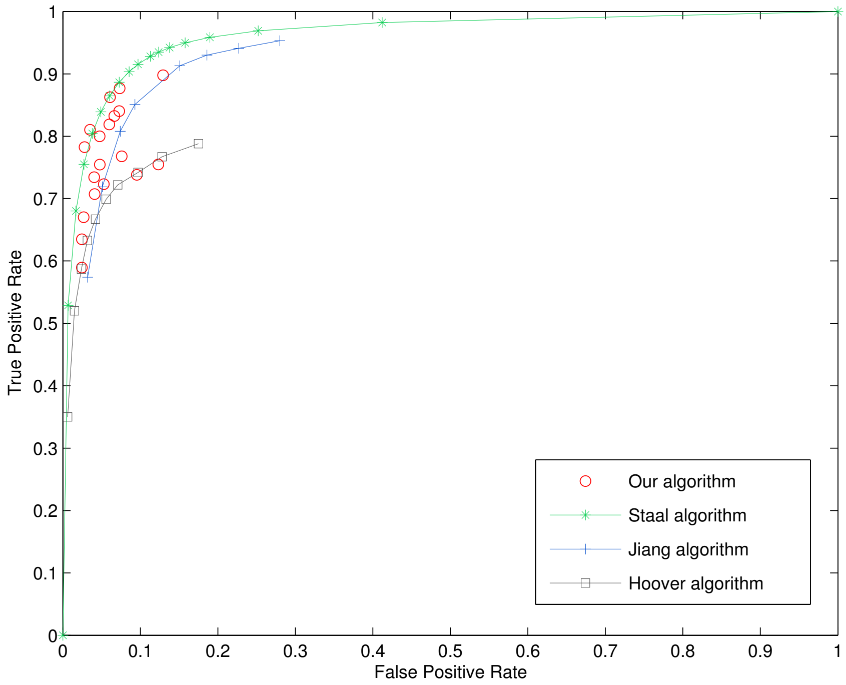
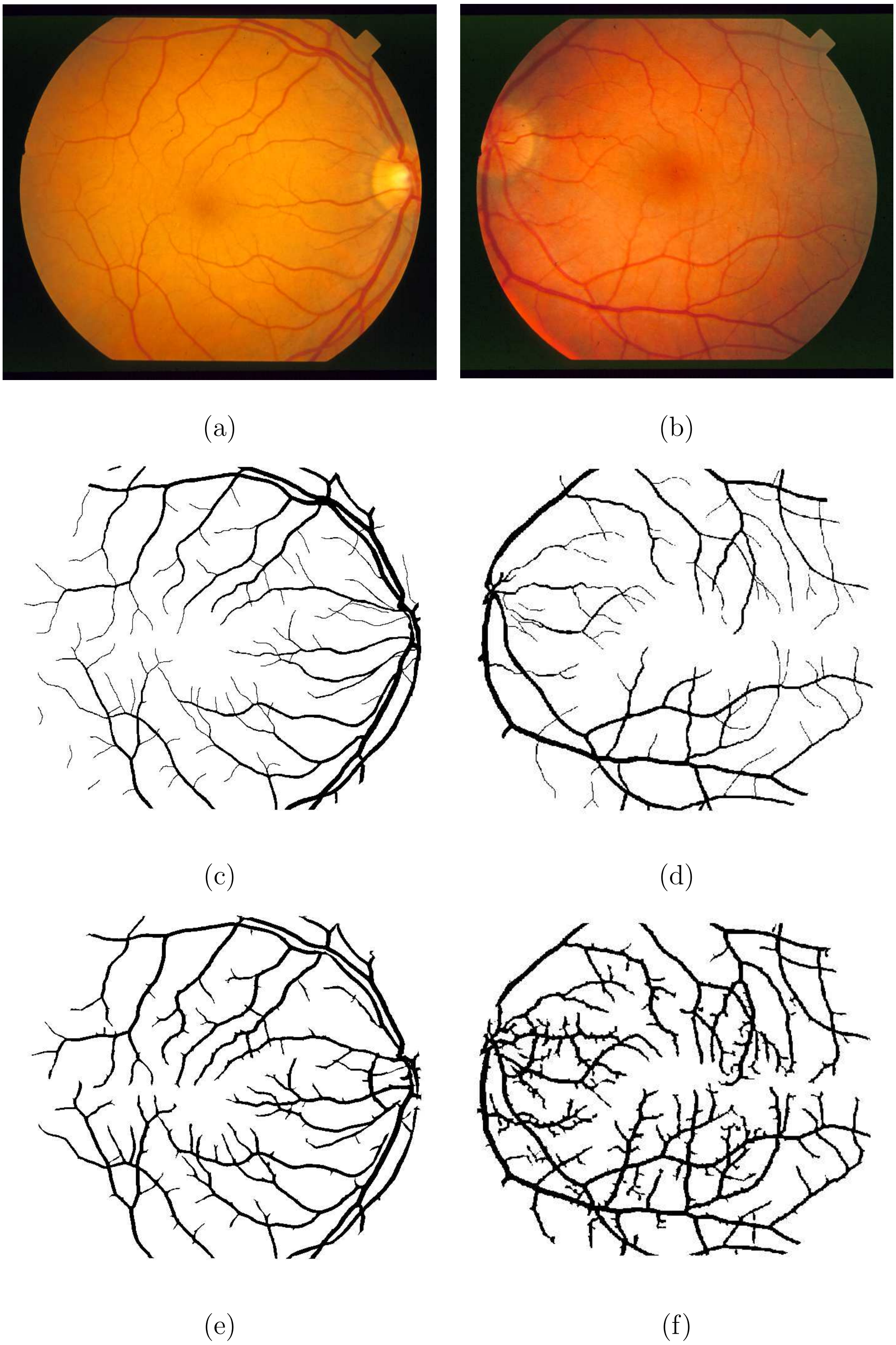


Figure 4.12: Performance of our approach versus other methods, Staal's algorithm [6, 7], Jiang's algorithm [8], and Hoover's algorithm [9].

thresholding. The proposed method retains the computational simplicity, and at the same time, can achieve good simulation results. Additionally, our false positive rates are lower than other computational expensive techniques while the true positive rates are comparable.

Table 4.1: Threshold values of three different approaches for twenty retinal images

Retina	Our Approach	Cross Entropy Thresholding	Local Entropy Thresholding
0001	121	106	184
0002	82	98	123
0003	121	101	150
0004	69	76	139
0005	90	107	128
0044	63	74	109
0077	82	92	114
0081	86	96	138
0082	89	93	122
0139	110	121	161
0162	64	68	86
0163	89	95	111
0235	89	98	117
0236	89	97	123
0239	97	106	134
0240	87	98	132
0255	81	86	125
0291	113	119	145
0319	83	92	116
0324	93	105	122

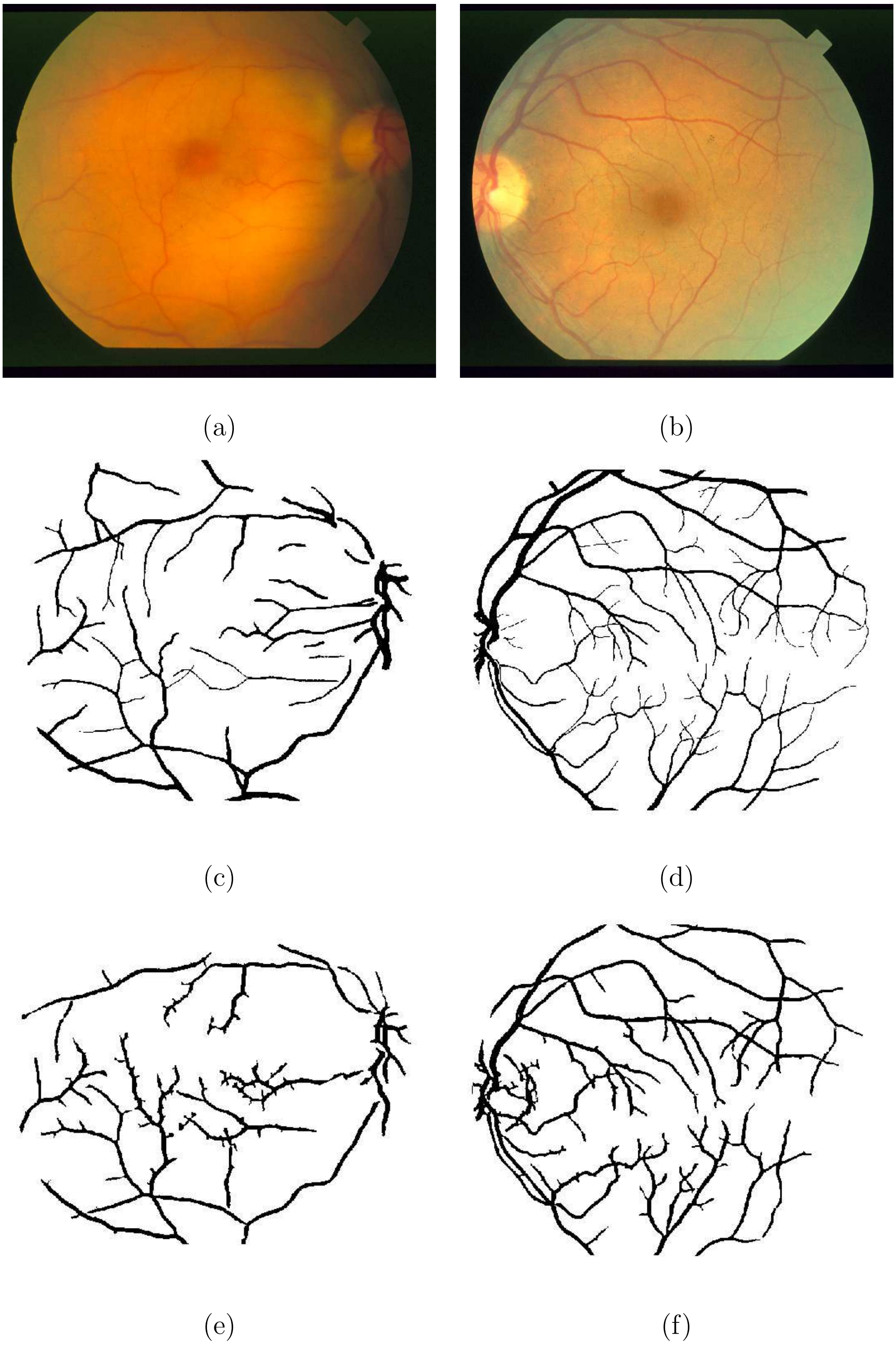


(a) (b)

(c) (d)

(e) (f)

Figure 4.13: (a),(b) Examples of normal retinal images. (c),(d) Hand-labeled ground-truth images. (e),(f) Our segmentation results.



(a)

(b)

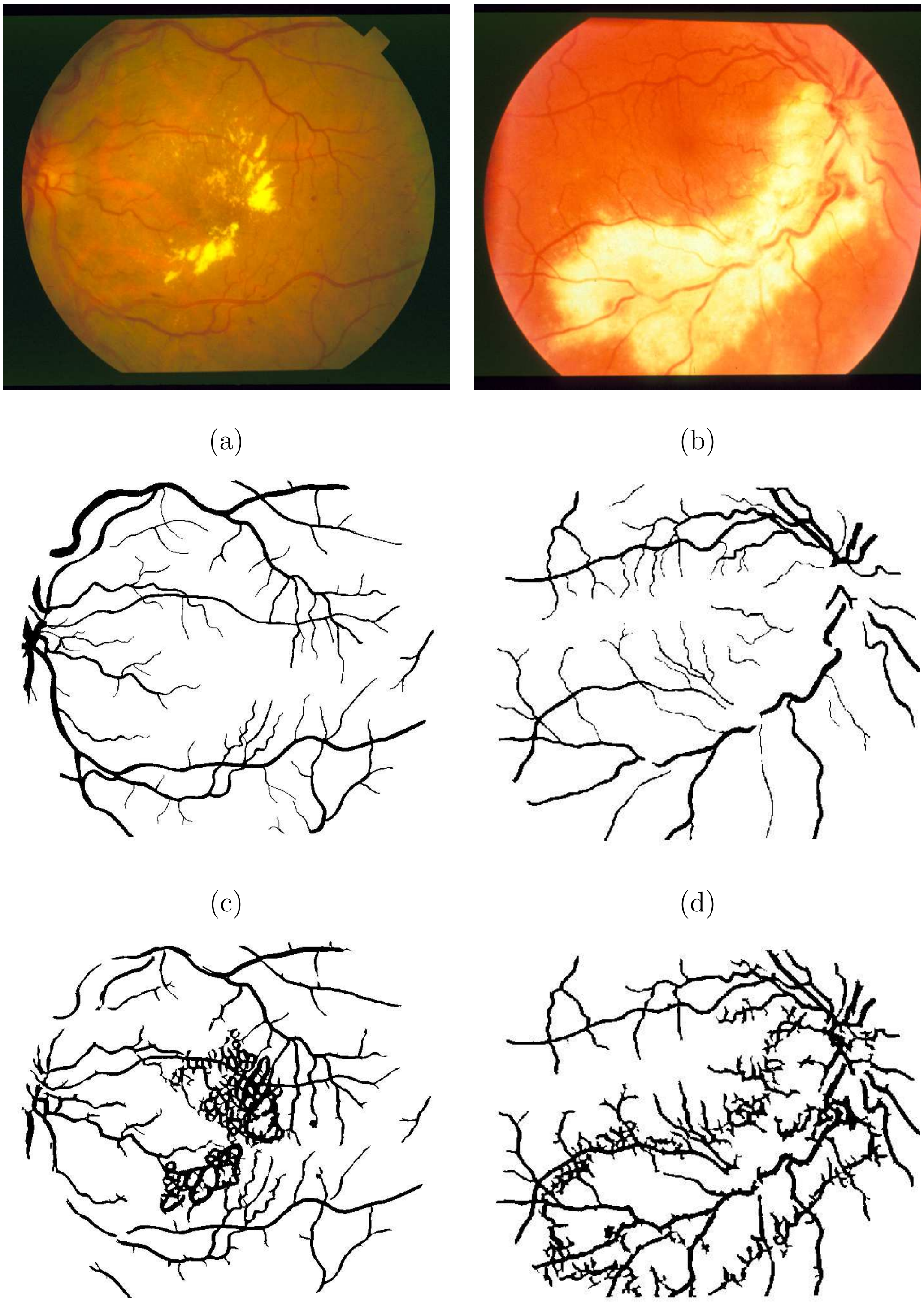
(c)

(d)

(e)

(f)

Figure 4.14: (a),(b) Examples of obscure blood-vessel retinal images. (c),(d) Hand-labeled ground-truth images. (e),(f) Our segmentation results.



(a)

(b)

(c)

(d)

(e)

(f)

Figure 4.15: (a),(b) Examples of retinal images with lesions. (c),(d) Hand-labeled ground-truth images. (e),(f) Our segmentation results.

CHAPTER 5

2-D Retinal Image Registration

5.1 Introduction

Registration is a problem on how to coincide two or more images. Two images are often taken at different times, viewpoints, modes, or resolutions. Additionally, an image plane, and an world plane are often not parallel. Hence, it is impossible to simply overlay two images together. To register two images, an "optimal" transformation model has to be identified. Numerous algorithms have been proposed regarding this topic. These methods differs in many aspects: (1) feature-based methods versus area-based methods; (2) batch methods versus RANSAC-like methods; (3) low-level methods, e.g. optical flow, autocorrelation, versus shape-based methods, e.g. template matching; (4) spatial domain versus frequency domain. In the case of medical imaging, disease diagnosis and treatment planning are often supported by multiple images acquired from the same patient. Image registration techniques, hence, are needed in order to integrate the information gained from several images to obtain a comprehensive understanding.

The organization of this chapter is as follows. The importance of 2-D retinal registration in medical applications are given in this section. Currently available different methods for retinal registration have been reviewed in section 5.1.1. Our methodology is given in Section 5.1.2. Section 5.2 provides information concerning technical background. Then, Section 5.3 describes the implementation of the proposed algorithm. Simulation results and the performance of our algorithm are presented in section 5.4.

5.1.1 Literature Reviews

Image registration is a fundamental problem to several image processing and computer vision applications [52, 53]. A broad range of image registration methods have been proposed for different medical imaging applications including retinal image registration. Various criteria, e.g., modalities, dimensionalities, elasticity of the transformation, have been proposed to categorize registration methods [52, 53, 55, 54]. Typically, retinal image registration techniques are classified as feature-based and area-based methods.

Area-based techniques are generally based on pixel intensities and certain optimized objective functions, such as least mean square error, cross-correlation, phase correlation, or mutual information, [90, 2, 1, 91, 92, 93, 94]. In the case of retinal image registration, area-based approaches are often used in multimodal or temporal image registration applications. In [1], mutual information was used as a similarity measure and simulated annealing was employed as a searching technique. In [2], the measure of match (MOM) was proposed as an objective function and the genetic algorithm was chosen to be the optimization technique. Nevertheless, the searching space of transformation models (affine, bilinear, and projective) is huge. The greater the geometric distortion between the image pair, the more complicated the searching space. Typically, there are two major factors that may degrade the performance of area-based methods: non-consistent/non-uniform contrast within an image and large homogeneous/textureless areas.

Feature-based methods are somewhat similar to manual registration [38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51]. The approach assumes that point correspondences are available in both images, and the registration process is performed by maximizing a similarity measure computed from the correspondences. In [40], the bifurcation points of a vascular tree, also called landmark points, were labeled with surrounding vessel orientations. An angle-based invariant was then computed to give a probability for every two matching points. After that, the Bayesian Hough transform was used to sort the transformations according to their respective likelihoods. In [38], the similarity matrix for all possible correspondences was computed based on the orientations of vascular centerlines and the similarity measure was converted to a prior probability. The transformation was estimated in a hierarchical way, from the zeroth-order model to the first-order model and finally to the second-order model. Nonetheless, sufficient feature points have to be available. In [43], the dual-bootstrap iterative closest point (dual-bootstrap ICP) algorithm was introduced. The approach started from one or more initial, low-order estimates that were only accurate in small image regions called bootstrap regions. In each bootstrap region, the method iteratively refined the transformation estimation, expanded the bootstrap region, and tested to see if a higher-order model can be used. The method required accurate initialization of at least one point correspondence. High success rates were reported in [43]. The performance of feature-based methods largely depends on sufficient and/or reliable correspondences, especially, when the overlapping part of an image pair is very limited or when there are mis-matched correspondences.

5.1.2 Methodology

In this paper, we study retinal image registration in the context of the National Institutes of Health (NIH), Early Treatment Diabetic Retinopathy Study (ETDRS) standard protocol [11]. The ETDRS protocol defines seven 30° fields of each retina with specific field coverage. A robust ETDRS image registration algorithm is required to (1) assess image quality in terms of ETDRS field coverage, and to (2) support ETDRS-based disease staging. Three major challenges are present. First, small overlaps between adjacent fields lead to inadequate landmark points (crossovers and bifurcations) for feature-based methods. Second, the contrast and intensity distributions within an image are not spatially uniform or consistent. This can deteriorate the performance of area-based techniques. Third, high-resolution ETDRS images contain large homogeneous nonvascular/textureless regions which result in difficulties for both feature-based and area-based techniques.

Because area-based and feature-based have their own strengths and limitations, in this work, we combine both area-based and feature-based registration methods to get the advantages each method has offered along with other decision-making criteria in order to obtain the best optimal solution. In order to achieve robustness and efficiency, hierarchical technique, translation, affine, and quadratic, is incorporated. Binary mutual information is proposed for translation estimation. It demonstrates better performance in terms of robustness compared with traditional gray-scale mutual information. In addition, multi-scale searching strategy is applied to avoid large combinatorial searching space. Furthermore, two parameters characterizing the displacements along vertical and horizontal directions in translation model, suggesting relative positions of each field, can be used for IQA regarding field coverage definition.

There are three major steps in the proposed algorithm. First, binary vascular trees are extracted from retinal images using a modified co-occurrence matrix for local entropy-based thresholding method [95, 96]. Next, zeroth-order translation is estimated by maximizing mutual information based on the binary image pair (area-based). Specifically, a local entropy-based peak selection scheme and a multi-resolution searching strategy are developed to improve the accuracy and efficiency of translation estimation. Third, we use two types of features, landmark points and sampling points, for higher-order transformation estimation. Sampling points, which are acquired by imposing a grid onto the thinned vascular tree, are only introduced when landmark points do not meet certain criteria. Simulation results on 504 pairs of ETDRS retinal images show the effectiveness and robustness of the proposed registration algorithm.

5.2 Preliminaries

5.2.1 NIH ETDRS Protocol

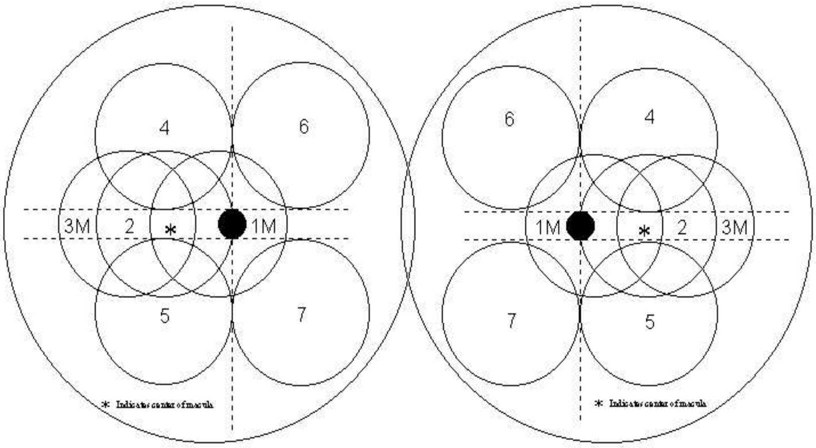


Figure 5.1: ETDRS seven-standard fields (right/left eyes) (http://eyephoto.ophth.wisc.edu/Photographers.html)

The importance of the ETDRS protocols and challenges in their implementation call for the automated software tool for image quality assessment (IQA). The retinal images need to meet the image quality criteria defined by ETDRS protocol. Each set of fundus photographs should be assessed for quality before the photographs are sent to the Coordinating center. A photographer is required to decide whether a particular image set meets the three ETDRS requirements: (1) clarity & focus; (2) field definition; and (3) stereo effect. In this work, we focus on field definition. The evaluation of field definition involves (1) horizontal and vertical displacements between image fields; and (2) verification of relative positions of key features, i.e. optic disc and fovea, in an image. The images are captured sequentially from field 1 and the required coverage defines acceptable overlapping regions. The ETDRS imaging standard specifies seven stereoscopic 30° fields of each eye, is defined in Table 5.1 and illustrated in Fig. 5.1. Field 1 is centered on the optic disc. Field 2 is centered on the center of macula. Overlapping parts of fields 1 and 2 (fields 1/2) as well as fields 2/3 are roughly 50% of the image size. For other fields, the overlapping parts are typically less than 25%. It is worth mentioning that the displacements are not always consistent and depend on patient cooperation and photographer skills.

5.2.2 Image Quality Assessment (IQA): Field Definition

The ETDRS protocol specifies seven stereoscopic 30° fields of each eye, as defined in Table 5.1 and Fig. 5.1. The overlap of field pairs 1 and 2 (or fields 1/2) [1] as well as that of fields 2/3 are roughly 50% of the image size. For other field pairs, the overlaps are typically less than 25%. It is worth mentioning that the field displacements are not always consistent and depend on patient cooperation and photographer's skills. The importance of the ETDRS protocol and the challenges in its practical implementation call for automated software tools for image quality assessment (IQA) that checks the relative positions, i.e., horizontal/vertical displacements, of every image pair according to Table 5.1. By comparing the offset, i.e., T_o, which is the difference between the desired vertical/horizontal displacements and actual ones, with the diameter of optic disc (DD), an image pair is categorized as good ($T_o < 1/2DD$), fair ($1/2DD \leq T_o \leq 1DD$), or poor ($T_o > 1DD$). Therefore, the IQA of ETDRS field definition boils down to a problem of image registration followed by displacement verification. We will briefly review some technical background of retinal image registration in the following.

[1] The notation of "fields 1/2" indicates that field 1 is the fixed image (the model image) and field 2 is the image being mapped to (the distorted image)

5.2.3 Global and Local Entropy

The global entropy, or the entropy, of a n-state system is defined as a function of the state probability [83],

$$H = -\sum_{i=1}^{n} p_i \log_2 p_i, \tag{5.1}$$

where p_i is the probability of state i. In the case of an image, the entropy is not an adequate measure since image pixel intensities are not independent of each other. Different images with identical histograms will always yield the same value of entropy since the spatial distribution is not taken into account in the global entropy computation. The local entropy [85], on the other hand, can better distinguish two images in terms of their spatial structures, because it considers the dependency of image pixel intensities. The local entropy is defined as

$$H^{(2)} = -\sum_{i}\sum_{j} p_{ij} \log_2 p_{ij}, \tag{5.2}$$

where p_{ij} is the probability of co-occurrence of the intensity levels, i and j. The co-occurrence matrix is an $L \times L$ dimensional matrix (L is the number of intensity levels) [85]. It indicates the transition of pixel intensities between adjacent pixels. The local entropy, therefore, can indicate the spatial structure/pattern of an image. The local entropy or second-order entropy will be used in this research for two main purposes: 1) selecting the threshold for vascular tree extraction; and 2) choosing an optimal match out of multiple competitive matches in the translation estimation.

5.2.4 Area-Based Retinal Image Registration

As mentioned before, area-based retinal image registration techniques are usually based on image pixel intensities and certain optimization functions. Specifically, we focus on mutual information (MI) that shows the similarity between one image pair based on the histograms and the joint histogram. The definition of mutual information

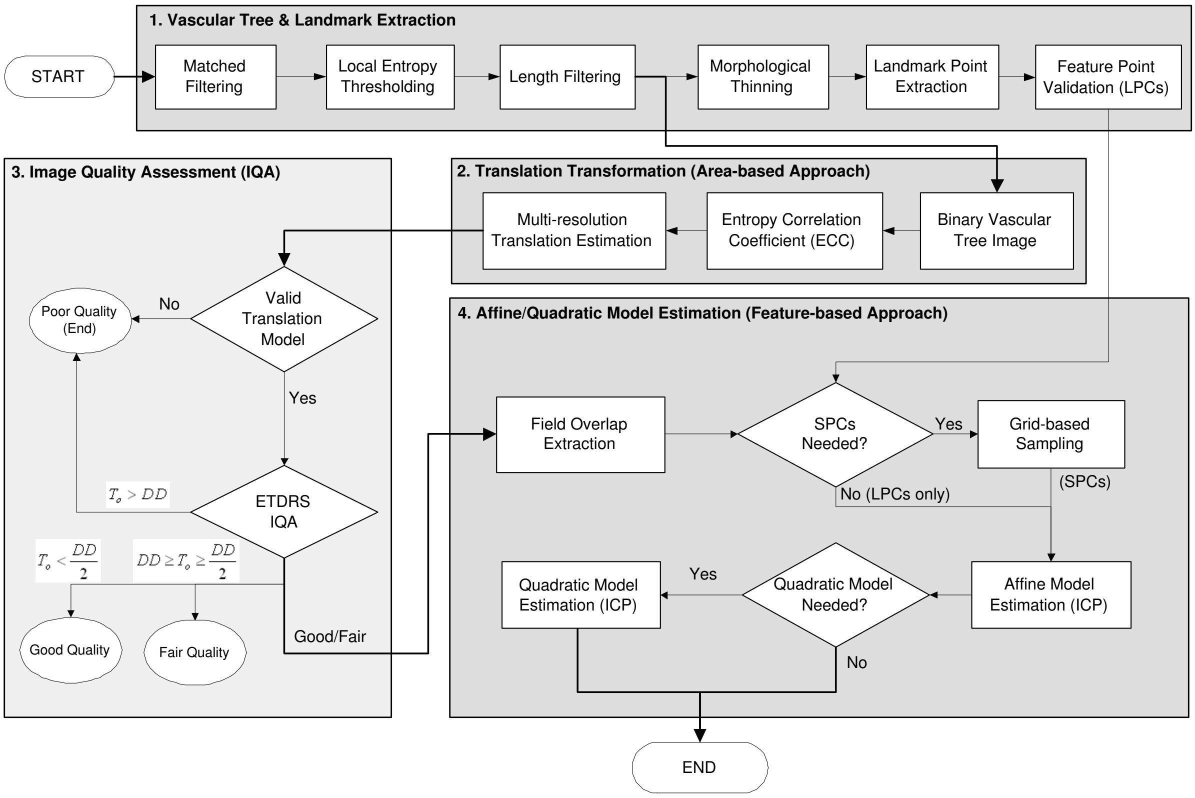

1. Vascular Tree & Landmark Extraction
START
Matched Filtering
Local Entropy Thresholding
Length Filtering
Morphological Thinning
Landmark Point Extraction
Feature Point Validation (LPCs)
2. Translation Transformation (Area-based Approach)
Multi-resolution Translation Estimation
Entropy Correlation Coefficient (ECC)
Binary Vascular Tree Image
3. Image Quality Assessment (IQA)
Poor Quality (End)
No
Valid Translation Model
Yes
$T_o > DD$
ETDRS IQA
$T_o < \frac{DD}{2}$
$DD \ge T_o \ge \frac{DD}{2}$
Good Quality
Fair Quality
Good/Fair
4. Affine/Quadratic Model Estimation (Feature-based Approach)
Field Overlap Extraction
SPCs Needed?
Yes
Grid-based Sampling
No (LPCs only)
(SPCs)
Quadratic Model Estimation (ICP)
Yes
Quadratic Model Needed?
Affine Model Estimation (ICP)
No
END

Figure 5.2: The flowchart of the proposed algorithm (LPCs: landmark point correspondences and SPCs: sampling point correspondences.)

can be presented in various ways [97]. Here, we use the definition as follows,

$$MI(\mathbf{I}_u, \mathbf{I}_v) = H(\mathbf{I}_u) + H(\mathbf{I}_v) - H(\mathbf{I}_u, \mathbf{I}_v), \tag{5.3}$$

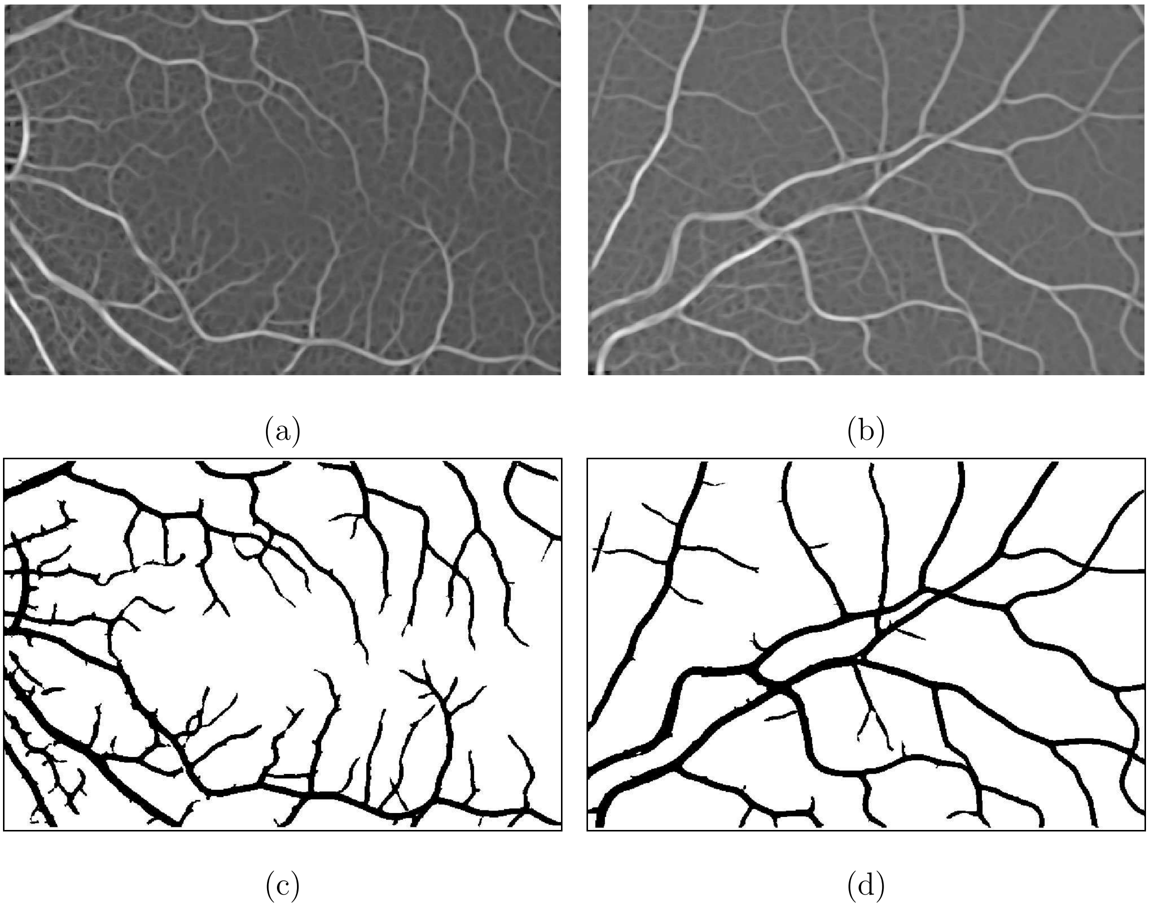
where $H(\mathbf{I}_u)$ is the global entropy, defined in equation (5.1), of image $\mathbf{I}_u$. $H(\mathbf{I}_u, \mathbf{I}_v)$ is the joint entropy, i.e., the entropy of the joint probability distribution of images $\mathbf{I}_u$ and $\mathbf{I}_v$.

For the registration purpose, the MI is computed for the overlap of two images. It is, therefore, sensitive to the size of overlaps. According to [97], the overlap size influences the mutual information measure in two ways. First, a decrease in the overlap size decreases the number of samples, which degrades the estimation of statistical probability distributions. Second, it has been shown in [98] that the MI computation is not robust when the overlaps are too small. The entropy correlation coefficient

(ECC) is a normalized measure of MI [99, 100], which is less sensitive to area changes in the overlaps, as defined in the following:

$$ECC(\mathbf{I}_u, \mathbf{I}_v) = 2 - \frac{2H(\mathbf{I}_u, \mathbf{I}_v)}{H(\mathbf{I}_u) + H(\mathbf{I}_v)}, \tag{5.4}$$

where $H(\mathbf{I}_u)$ and $H(\mathbf{I}_v)$ represent global entropy of images $\mathbf{I}_u$ and $\mathbf{I}_v$ respectively. $H(\mathbf{I}_u, \mathbf{I}_v)$ is the joint entropy between two images. It is shown that the ECC is generally a better option for registering images since it is less susceptible to different sizes of overlaps [99, 100].



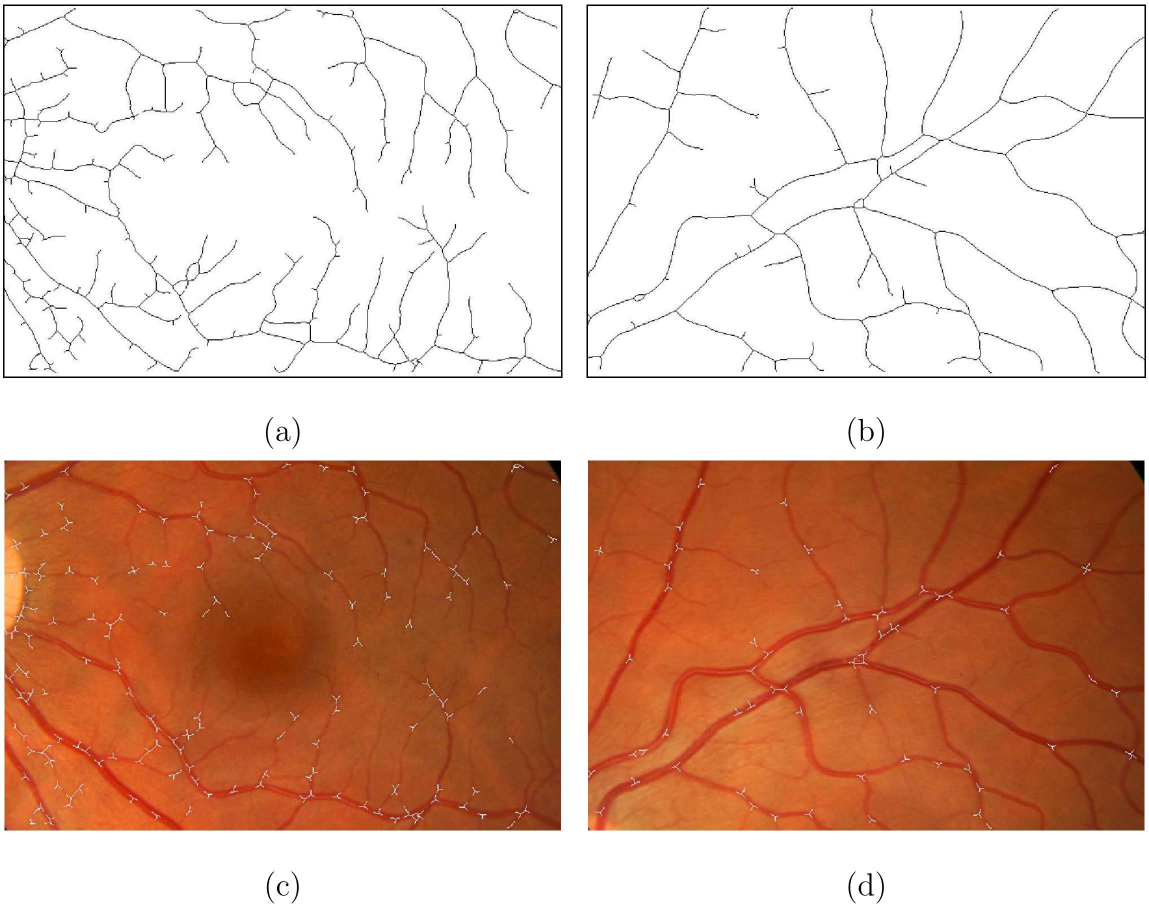
(a)

(b)

(c)

(d)

Figure 5.3: Vascular tree and the centerline extraction of fields 2 and fields 4 of Julie's retinal images. (a) Match-filtered image: field 2; (b) Match-filtered image: field 4; (c) Binary image: field 2; (d) Binary image: field 4.



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

Figure 5.4: Vascular tree and the centerline extraction of fields 2 and fields 4 of Julie's retinal images. (a) Thinned binary image: field 2; (b) Thinned binary image: field 4; (c) Crossover/bifurcation points: field 2; (d) Crossover/bifurcation points: field 4.

5.2.5 Feature-Based Retinal Image Registration

Feature-based methods relies on correspondence points in both images. The matching process identifies reliable correspondences by maximizing an objective function related to features. Then the transformation is estimated by minimizing correspondences' displacement, i.e., the registration error,

$$\hat{\mathbf{M}} = \arg\min_{\mathbf{M}} \mathrm{median}_{\mathbf{p}\in\mathbf{P}} \min_{\mathbf{q}\in\mathbf{Q}} \|\mathbf{p} - T(\mathbf{q};\mathbf{M})\|^2, \tag{5.5}$$

where $\mathbf{P}$ and $\mathbf{Q}$ denote the feature point sets from two images. $T(\mathbf{q};\mathbf{M})$ represents the transformation operation of point $\mathbf{q}$ given model $\mathbf{M}$. The *median* function can also be used in (5.5) that is less sensitive to outliers compared with the *mean* function

[38, 39]. As listed in Table 5.2, the transformation models often used include the translation model, the affine model, and the quadratic model [1, 38, 39, 43]. The translation model consists of two parameters characterizing the displacements along horizontal and vertical directions. The affine model describes translation, rotation, shearing, and scaling. However, the affine model cannot address the non-linearity, i.e., retina's curvature. The quadratic transformation has 12 parameters and it can cope with non-linearity. The estimated transformation model, $\hat{\mathbf{M}}$, can be further adjusted by using the Iterative Closest Point algorithm (ICP) to refine correspondences [101]. The ICP is a procedure for iteratively matching a set of points in two images. Given an initial transformation model, for $\mathbf{p} \in \mathbf{P}$, we need to find the closest point $\mathbf{q} \in \mathbf{Q}$ by following:

$$d(\mathbf{p}, \mathbf{Q}) = \min_{\mathbf{q} \in \mathbf{Q}} \parallel \mathbf{p} - T(\mathbf{q}; \hat{\mathbf{M}}) \parallel, \tag{5.6}$$

where $d(,)$ is a distance metric. Then the model will be re-estimated according to (5.5) after correspondence refinement, and so on. The iteration will be terminated when $d(,)$ is stable.

5.3 Proposed Framework

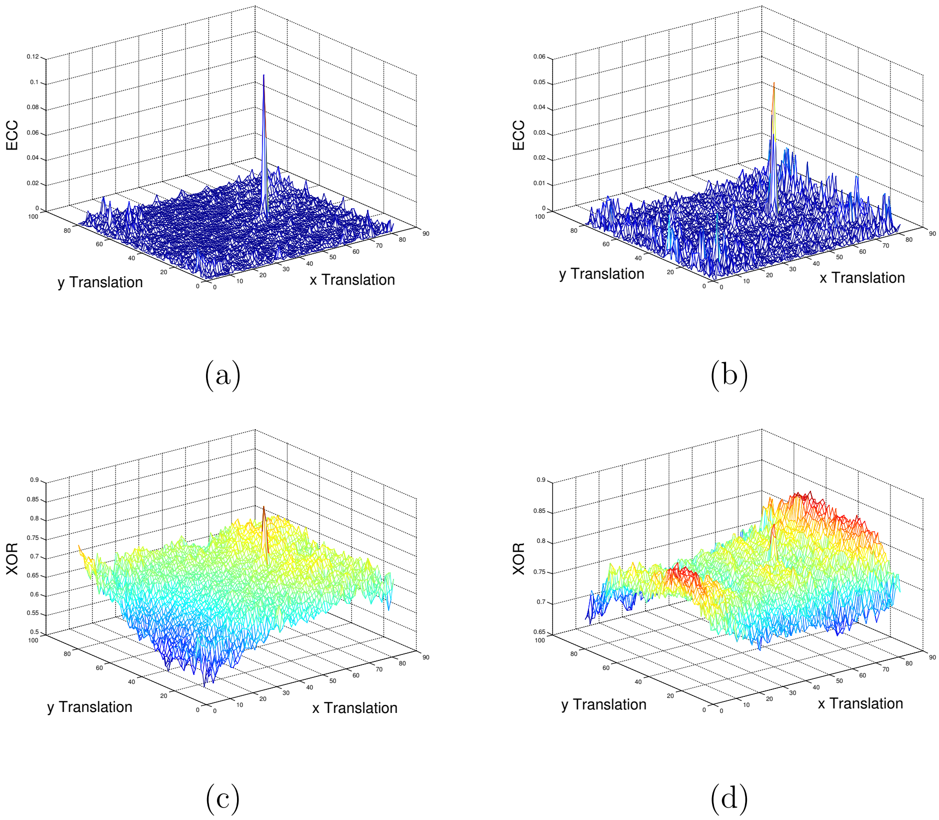
As mentioned before, area-based and feature-based image registration methods have their own strengths and limitations in the context of the ETDRS protocol. In this work, we propose a hybrid registration approach for ETDRS images, which effectively takes the advantages of both area-based and feature-based methods in one flow. As shown in Fig. 5.2, the proposed algorithm is composed of three major steps which are discussed in details below.

5.3.1 Vascular Tree Extraction

First, a match filter is applied to enhance the prominence of blood vessels [34]. Second, a local entropy-based thresholding scheme is used which takes into account the spatial distribution of gray levels and can well preserve the spatial structures in the binarized image. Subsequently, a length filtering technique is used to remove misclassified pixels or insignificant small segments. For the match filtering results are shown in Fig. 5.3(a) and (b), the entropy-based thresholding with length filtering results are shown in Fig. 5.3(c) and (d) where we can see blood vessels are clearly segmented from the background. The binary vascular tree will be used for area-based registration. Then a morphological thinning operation is employed to obtain the centerline of the vascular tree. Finally, vascular crossover/bifurcation points, which will be used for feature-based registration, are located by a two-step window-based probing process. An initial probing process is first conducted to find potential branching points by locating the location where its 3×3 neighborhood contains more than 3 pixels belong to the thinned vascular tree. Then a 11×11 window is applied to all potential branching points to identify true branching points. If there are more than 2 vessel pixels of on the window boundary, it will be marked as a bifurcation/crossover point, as shown in Fig. 5.4 (c) and (c). More detail of vascular tree extraction can be found in Section 4.3.

5.3.2 Translation Estimation

The translation estimation is crucial for the hierarchical model estimation approach, since it will define a constraint for the higher-order model estimation as well as a restriction for correspondence selection. As seen from Fig. 5.3, manually determining a rough translation between fields 2/4 in this image set is difficult because of small overlaps with complex vascular tree structures. In addition, not every pair of retinal images complies exactly with ETDRS field definition standard. Moreover, the point correspondences are not particularly distinguishable from each other. Instead of using crossover/bifurcation points, the translation estimation here is implemented by an area-based method which is based on a binary vascular tree due to the following three reasons. 1) The vascular tree is, undeniably, the most prominent structure and spans all ETDRS seven fields. 2) MI or ECC may not robust when the contrast/intensity distributions are not consistent within each image, and the binary image will greatly enhance the strength of MI and ECC computation. 3) In the ETDRS protocol, geometric distortion in seven-field images usually are not significant with negligible rotation and scaling, resulting in a relatively small searching space.



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

Figure 5.5: Sample plots of binary image-based ECC vs. logical operation, XOR, at every possible translation in the coarsest scale (downsampled by 16). (a) The binary image-based ECC of Julie's fields 1/2; (b) The binary image-based ECC of Julie's fields 2/3 (c) The logical operation XOR of Julie's fields 1/2; (d) The logical operation XOR of Julie's fields 2/3.

Binary ECC. Traditionally, MI or ECC based registration has been used on gray-scale images. However, MI or ECC is not robust when the contrast/intensity distributions within each image field are not consistent, invalidating the statistical dependency across images. Therefore, we, instead, compute the ECC based on binary vascular tree images. Given two binary vascular trees images, $\mathbf{I}_u^{'}$, and $\mathbf{I}_v^{'}$, we estimate the translation model that aligns images $\mathbf{I}_u$ and $\mathbf{I}_v$ by maximizing the ECC between $\mathbf{I}_u^{'}$, and $\mathbf{I}_v^{'}$, as defined in the following:

$$\hat{\mathbf{M}} = \arg\max_{\mathbf{M}} ECC(\mathbf{I}_u^{'}, T_0(\mathbf{I}_v^{'}; \mathbf{M})), \tag{5.7}$$

where $T_0(\mathbf{I}_v^{'}; \mathbf{M})$ translates image $\mathbf{I}_v^{'}$ with model $\mathbf{M}$. The ECC is defined in equation (5.4). Sample plots of binary ECC for every possible translation at the coarsest scale (downsampled by 16) are shown in Fig. 5.5 (a) and (b) where the distinct peaks, indicating the optimal translations, can be easily identified. One might wonder, though, why not just employ the simple logical operation XOR. Here, we also tested XOR by developing an objective function defined as follows:

$$\hat{\mathbf{M}} = \arg\max_{\mathbf{M}} \frac{XOR(\mathbf{I}_u^{'}, T_0(\mathbf{I}_v^{'}; \mathbf{M}))}{H \times W}, \tag{5.8}$$

where $H \times W$ is the size of the overlaps. Example plots of XOR logical operation are shown in Fig. 5.5 (c) and (d). Although XOR can also perform well in an image pair with insignificant distortions, XOR output does not give a distinguishable narrow peak as shown in Fig. 5.5 (c). Moreover, in the case of retinal images with moderate geometric distortions, XOR does not provide the noticeable peaks, or it sometimes fails to locate the accurate translation as shown in Fig. 5.5 (d).

We want to further manifest the advantages of ECC over XOR by defining an energy concentration function to numerically evaluate the peak quality of ECC's and XOR's outputs. The proposed energy concentration function is a simple indicator of peak distinction, which is defined as follow:

$$\Psi_K = \frac{\sum_{k=1}^{K} \phi_k^2}{\sum_{i=1}^{N_r} \sum_{j=1}^{N_c} (\psi_{i,j})^2} \times 100\%, \tag{5.9}$$

where Ψ_K is the energy proportion of K highest peaks, ϕ_k is the k^{th} largest peak value, $\psi_{i,j}$ is the ECC/XOR output under translation (i, j), and N_r and N_c are the numbers of all possible translations in row and column respectively.

We compare ECC and XOR on 72 image pairs in terms of peak distinction by using Ψ_K as shown in Table 5.3. It is shown that ECC is substantially better than XOR in estimating the optimal translation between a binary image pair. This might be due to the statistical capability of ECC which makes ECC's peaks much more distinct than those of XOR.

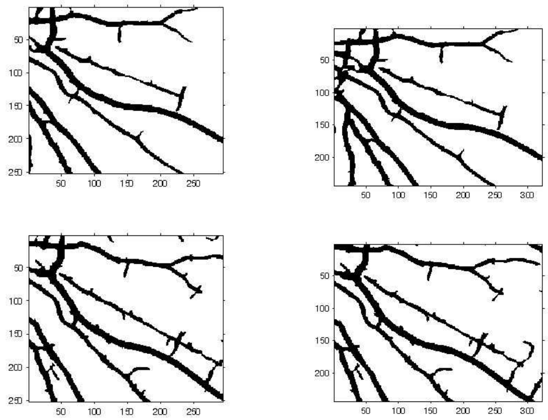


Figure 5.6: The overlaps of two possible translation models (left column and right column) that produce similar values of ECC peaks for fields 1/7.

Binary Local Entropy. It is possible that there are multiple competitive peaks in ECC outputs, indicating several possible translations as shown in 5.6. Sometimes the highest peak does not necessarily represent the optimal translation due to the geometric distortion and the dissimilarity of binary image pairs. Therefore, an auxiliary criterion is needed to select the right peak. Since the local entropy can indicate the spatial structure in an image, we use it to evaluate resemblance between two overlaps. Let $\mathbf{I}^{(u)}$ and $\mathbf{I}^{(v)}$ represent the overlaps from binary vascular tree images $\mathbf{I}_u^{'}$ and $\mathbf{I}_v^{'}$ respectively. The unique optimal translation is obtained by

$$\hat{\mathbf{M}} = \arg\min_{\mathbf{M}\in\Omega} \left| H^{(2)}(\mathbf{I}^{(u)}) - H^{(2)}\left(T_0(\mathbf{I}^{(v)};\mathbf{M})\right)\right|, \tag{5.10}$$

where $\Omega = \{\mathbf{M}_1, \ldots, \mathbf{M}_N\}$ is a set of all possible translations with sufficiently large ECC values according to equation (5.7). $H^{(2)}$ is the local entropy defined in equation (5.2).

Multi-Resolution Searching Strategy. Given the translation model, there are only two parameters in the searching space which still could be very large due to the high-resolution nature of ETDRS images. Instead of going through every possible translation in the original scale, a multi-resolution searching scheme is developed in order to reduce the computational complexity. A binary image is first represented in a pyramid of multiple resolutions from the coarsest scale to the finest scale. Let $\mathbf{I}_u^{'} = \{\mathbf{I}_u^{(0)}, \mathbf{I}_u^{(1)}, \mathbf{I}_u^{(2)}, \ldots, \mathbf{I}_u^{(J)}\}$ and $\mathbf{I}_v^{'} = \{\mathbf{I}_v^{(0)}, \mathbf{I}_v^{(1)}, \mathbf{I}_v^{(2)}, \ldots, \mathbf{I}_v^{(J)}\}$ represent the finest to the coarsest scales of binary images $\mathbf{I}_u^{'} = \mathbf{I}_u^{(0)}$ and $\mathbf{I}_v^{'} = \mathbf{I}_v^{(0)}$, respectively. First, the algorithm finds a binary-ECC peak, an optimal translation, at the coarsest scale, i.e., $\mathbf{I}_u^{(J)}$ and $\mathbf{I}_v^{(J)}$, $\mathbf{M}^{(J)} = \{(r_u^{(J)}, s_u^{(J)}), (r_v^{(J)}, s_v^{(J)})\}$ where $(r_u^{(J)}, s_u^{(J)})$ and $(r_v^{(J)}, s_v^{(J)})$ are two coordinates in two images showing the optimal translation between them. Then $\mathbf{M}^{(J)}$ can specify a constrained searching neighborhood at the finer scale, i.e., $\mathcal{N}(\mathbf{M}^{(J)}) = \{(2r_u^{(J)} + i, 2s_u^{(J)} + j), (2r_v^{(J)} + m, 2s_v^{(J)} + n) | i, j, m, n = -5, ..., 5\}$. where the optimal translation at scale $J-1$ scale can be obtained as follows:

$$\mathbf{M}^{(j-1)} = \arg\max_{\mathcal{N}(\mathbf{M}^{(J)})} ECC(\mathbf{I}_u^{(j-1)}, \mathbf{I}_v^{(j-1)}), \tag{5.11}$$

where $j = J, J-1, ..., 1$. This procedure starts from the coarsest scale, and it is repeated until the finest scale is reached where an optimal translation, $\mathbf{M}^{(0)}$, is achieved at the pixel-level for images $\mathbf{I}_u^{(0)}$ and $\mathbf{I}_v^{(0)}$.

If the translation model $\mathbf{M}^{(0)}$ is reliable, we can move forward to higher-order models, such as affine/quadratic models, to obtain better registration performance. However, if $\mathbf{M}^{(0)}$ is incorrect, when there is significant geometric distortion between an image pair (unlikely to happen for ETDRS images) or there is no obvious vascular tree structure in overlaps (especially when image clarity is poor), the ECC-based translation estimation of binary images fails (this also indicates poor quality of ETDRS field coverage). In this work, we define two criteria based on the ECC output to check the correctness of $\mathbf{M}^{(0)}$. The first is energy concentration which measures the energy percentage of top three peaks as defined in (5.9), i.e., Ψ_3. The other is peak distinction that is the ratio of top two peaks, i.e., $\Phi = \phi_1/\phi_2$ where ϕ_k is the kth largest peak. A valid translation model usually leads to large Ψ_3 and Φ, indicating a sufficiently good match.

5.3.3 Affine/Quadratic Transformations

Usually, feature-based methods are supposed to be more reliable than area-based approaches if sufficient and accurate feature points are available and a proper translation model can be obtained which can greatly facilitate the subsequent higher-order transformation estimation. We here employ a feature-based scheme in order to refine the registration transformation from the zeroth-order to the higher-order model, where two different types of feature points, i.e., landmark points and sampling points, are involved, as illustrated in Fig. 5.2.

Landmark Point Correspondences (LPCs). Landmark points are the crossover/ bifurcation points of vascular tree. The initial translation model achieved from section 5.3.2 and landmark points obtained from section 5.3.1, are employed as the rudimentary guideline to establish the initial set of LPCs, $C^{'}$, for the first-order affine model.

$$C^{'} = \{(\mathbf{p}_i, \mathbf{q}_j) | dist(T(\mathbf{q}_j; \hat{\mathbf{M}}), \mathbf{p}_i) \leq err, \mathbf{p}_i \in \mathbf{P}, \mathbf{q}_j \in \mathbf{Q}\}, \tag{5.12}$$

where $\mathbf{P}$ and $\mathbf{Q}$ are the sets of N_u and N_v landmark points from $\mathbf{I}_u^{'}$ and $\mathbf{I}_v^{'}$, respectively, $\hat{\mathbf{M}}$ is the initial translation model, $dist(,)$ denotes the Euclidean distance, and err is a threshold (e.g., 30 for the affine model and 5 for quadratic model). We need to have one-to-one matchings for all LPCs. However, there is no guarantee that LPCs in C' are one-to-one matching. In fact, a specific landmark point in $\mathbf{P}$ may have multiple matches, a single match, or no match at all in $\mathbf{Q}$. Therefore, we create a similarity matrix, $S = \{s_{i,j} | i = 1, ..., N_u; j = 1, ..., N_v\}$, with the purpose of assuring one-to-one matching for every LPC. The similarity measure, $s_{i,j}$ as defined below, is a coarse measure to quantify the resemblance between $\mathbf{p}_i$ and $\mathbf{q}_j$.

$$s_{i,j} = \begin{cases} \mathbf{x}_i \cdot \mathbf{y}_j, & (\mathbf{p}_i, \mathbf{q}_j) \in C^{'} \\ 0, & \text{otherwise} \end{cases} \tag{5.13}$$

where $\mathbf{x}_i$ and $\mathbf{y}_j$ are obtained by placing a 9×9 window centered at $\mathbf{p}_i$ and $\mathbf{q}_j$ on the thinned images of $\mathbf{I}_u^{'}$ and $\mathbf{I}_v^{'}$ respectively. One-to-one LPC matchings are achieved by

$$C = \{(\mathbf{p}_i, \mathbf{q}_j) | j = \arg \max_{j \in 1, ..., N_v} s_{i,j}, i = 1, ..., N_u\}. \tag{5.14}$$

After C is obtained, we need to examine the reliability of LPCs. Let σ_x^2 and σ_y^2 be the second central moments of vertical/horiztonal coordinates of LPCs in the overlap of size $H_o \times W_o$. We define $\frac{H_o}{\sigma_x}$, $\frac{W_o}{\sigma_y}$ to show how the LPCs spread in the overlap. If they are large (e.g., > 4), LPCs are likely to cluster together in a small area. Then the sampling process is needed to involve more feature points for image registration.

Sampling Point Correspondences (SPCs). Sampling points can be acquired by imposing grid lines on the thinned vascular tree, where the intersections between blood vessels and the grid are marked as sampling points as shown in Fig. 5.7. If LPCs are not sufficient to estimate affine (at least 3) or quadratic (at least 6) models, SPCs are introduced to facilitate feature-based registration by providing some auxiliary information to LPCs. However, SPCs are usually less trustable compared with LPCs, since they are acquired from the thinned vascular tree which often exhibits strong linearity and are likely be linearly dependent if the vascular tree is sparse. Therefore, SPCs are only involved when LPCs do not meet certain criteria, and more details are discussed in the simulation. Given a set of sampling points, SPCs can be achieved in the same way as LPCs defined in equations (5.12) and (5.14). It is worth noting that the similarity metric defined in (5.14) is less effective for the SPCs where the vessel exhibits strong straightness. Thus SPCs are more valuable when the vascular tree has strong non-linearity, indicating more LPCs.

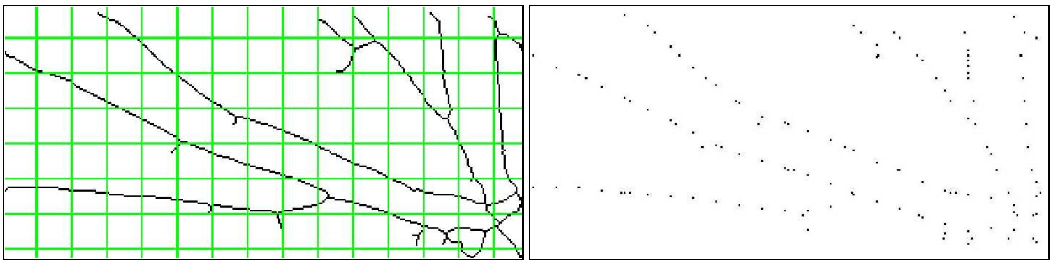


Figure 5.7: An example of the sampling process. (Left) A grid is placed on the thinned binary vascular tree; (b) The sampling points.

Iterative Closest Point (ICP). Both LPCs and SPCs are the input for the ICP algorithm which is a procedure to refine the model estimation by finding the closest point $\mathbf{q} \in \mathbf{Q}$ for every $\mathbf{p} \in \mathbf{P}$ given transformation $\mathbf{M}$. During the ICP iteration, *bad* LPCs/SPCs, i.e., the ones with significant Euclidean distances after the transformation (e.g., 6 pixels), are eliminated. The affine model is re-estimated at each iteration by finding the minimum mean square error solution according to equation (5.5). The iteration is terminated when the model is stable or there is no change to LPCs/SPCs. We will proceed to the quadratic model if the registration error of the affine model is significant and we have at least 6 LPCs. According to [38], a quadratic model is a good approximation to retinal surface since retinal surface approximates a sphere. The quadratic model could be useful when the overlap is significant. However, in the case of ETDRS, we prefer not to proceed to the quadratic model when overlaps between two images are very limited or LPCs are not sufficient (say less than 6). It may not be robust to estimate the transformation of the whole image based only on small overlaps where the spatial information across different fields is too limited. The higher-order model, sometimes, could introduce significant distortions to the image pairs with small overlaps.

5.4 Experimental Analysis

The ETDRS images were provided by Inoveon [2] which is a medical services company delivering solutions to diagnose eye-related diseases, mainly diabetic retinopathy. Since the ETDRS protocol defines seven stereoscopic fields in each eye (two sets of six pairs for each eye), there will 24 image pairs for one patient. Totally, 504 pairs collected from 21 patients were involved for algorithm evaluation. These images were taken with a Kodak DCS520 digital camera coupled with a Zeiss FF450 fundus camera which has the original resolution of 1152×1728. In order to make the size of registered images tractable, we down-sampled all images to 600×900. Still the final registration results are very large, nearly 2000×3000.

5.4.1 Vascular Tree Extraction

We hand-labeled vascular tree structures for several retinal images for the comparison purpose, as shown in Fig. 5.8. Performance of blood vessel extraction is evaluated by using the true positive and the false positive rates as in [9]. Any pixel which was hand-labeled as vessel and the algorithm labeled as vessel is counted as true positive. Any pixel which was hand-labeled as non-vessel and the algorithm labeled as vessel is counted as false positive. The true positive rate is calculated by normalizing true positive by the total pixel number of the hand-labeled vessel. The false positive rate is calculated by normalizing false positive by the total pixel number of the hand-labeled non-vessel. As shown in Fig. 5.8, the true positive rates for fields 1 and 2 are 0.9050 and 0.8854, and the false positive rates for fields 1 and 2 are 0.0798 and 0.0619. In general, our vascular segmentation results are thicker than ground-truths,

[2] http://www.inoveon.com/index.html

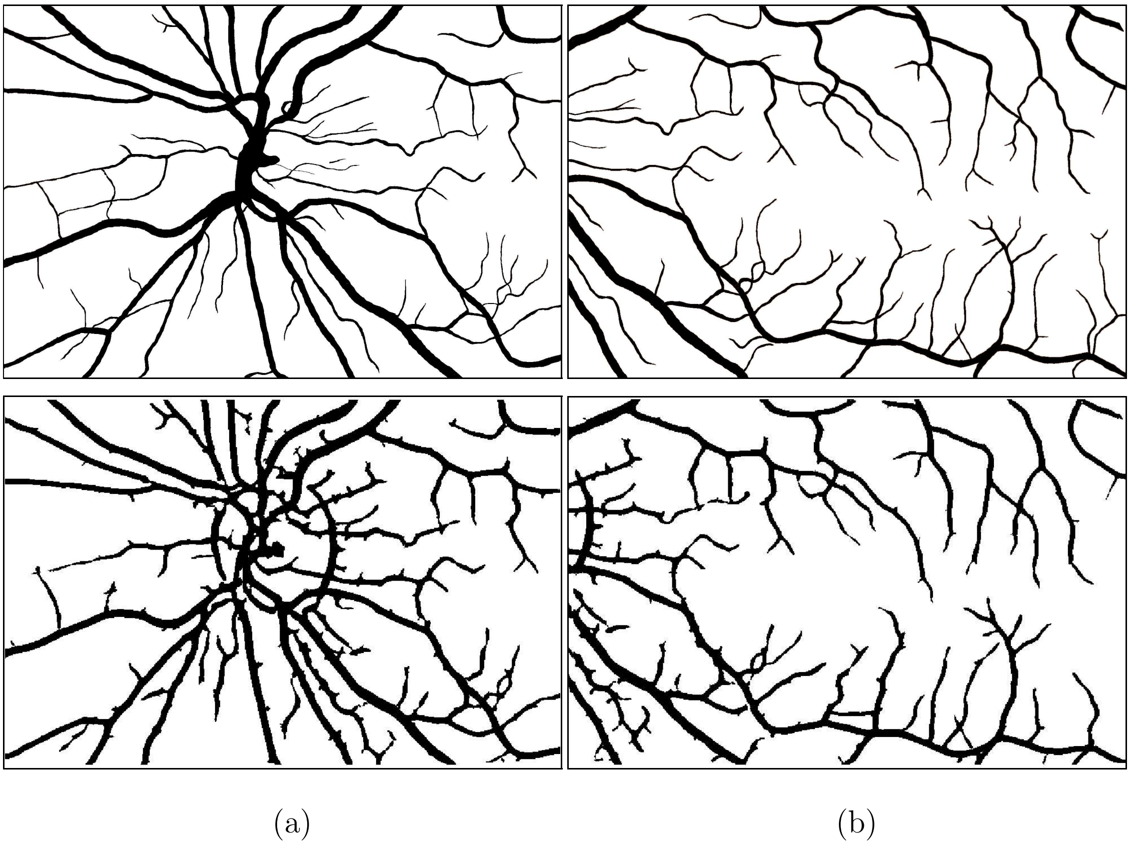
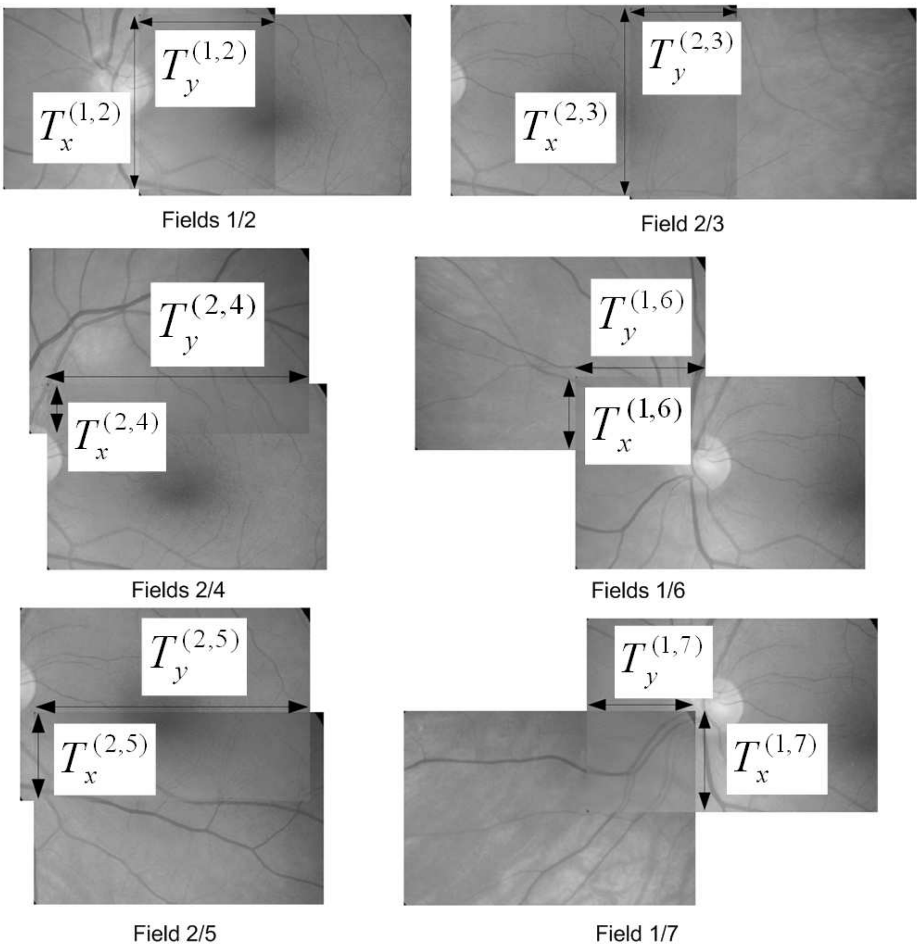


Figure 5.8: Hand-labeled groundtruths (top) and the extracted vascular trees using the proposed algorithm (bottom) of field 1 (a) and field 2 (b).

and this thickening effect is introduced by the match filter which tends to increase the width of vessels. Due to the symmetric property of the match filter, this artifact has negligible effect on the area-based translation estimation. Moreover, the feature-based transformation estimation mainly relies on thinned centerlines of vascular tree which are relatively stable after match filtering.

5.4.2 IQA Simulation

We assume that photographers can indicate whether the positions of optic disc and macular in fields 1 and 2 are accurate and they also can obtain the diameter of optic disc (DD). In order to perform IQA, we define vertical/horizontal displacements for each ETDRS field pair in Fig. 5.9, and Table 5.4 lists all ideal translation displace-



Fields 1/2

Field 2/3

Fields 2/4

Fields 1/6

Field 2/5

Field 1/7

Figure 5.9: Vertical/horizontal displacements of six ETDRS image pairs.

ments for each field pair according to the ETDRS field definition specified in Table 5.1 and Fig. 5.1. Then the IQA of field coverage can be easily implemented by computing the offset, i.e., T_o, which is the difference between the ideal displacements and the actual ones given by the translation model. By comparing T_o with DD, an image pair can be classified as ”good”, ”fair”, or ”poor” quality. For example, all six pairs in Fig. 5.9 are ”good” pairs. Since a correct translation model guarantees the IQA validity, the IQA accuracy is mainly determined by the success rate of translation estimation. Actually, the actual IQA accuracy could be higher, since most cases when the ECC-based translation estimation fails are due to the significant incompliance to the ETDRS field definition which are “poor” pairs.

5.4.3 Registration Performance Evaluation

In this section, we present both quantitative and qualitative analysis of the three major techniques on 504 ETDRS image pairs. It was found that the ICP iteration runs 1 $\sim$ 5 rounds in most cases. We use the registration error defined in (5.5) to evaluate the algorithm performance. The numerical results are summarized in Table 5.5. Given the binary vascular tree, the total computational time for image registration per pair is approximately 20 seconds on a 2.8GHz PC and the Matlab 6.5 platform without algorithm optimization.

Translation Model Estimation. The registration error of the translation model over the whole data set is 21.30 pixels, as shown in Table 5.5. This performance is still acceptable for the IQA purpose, since DD is usually near 200 pixels. Fig. 5.15 shows an example of the translation-based registration result, where ghost lines due to registration errors are clearly visible as seen in the zoomed-in regions. Out of 504 registered pairs, there are 39 pairs are rejected by manual validation, and the success rate is 92.3%. As mentioned before, we have defined two criteria to determine the credibility of the translation model based on the ECC output, i.e., energy concentration of top three peaks (Ψ_3) and peak distinction of top two peaks (Φ). From those bad pairs, we found both Ψ_3 and Φ are relatively small compared with others. Thus we use $\Psi_3 > 13\%$ and $\Phi > 2.0$ as two empirical conditions to accept the estimated translation model. These two criteria are able to identify 36 bad pairs out of 39 pairs manually rejected. The failure rate of translation estimation is shown in Fig. 5.10, where the failure rates of fields 1/6 and 1/7 are relatively higher due to small overlaps (usually less than 20%).

Affine/Quadratic Model Estimation. We have 468 image pairs for which the translation model is accepted and a higher-order model is applied. Fig. 5.11 shows the change of registration errors resulting from consecutive models.

It is clearly seen that the error is dramatically reduced from a lower-order model to

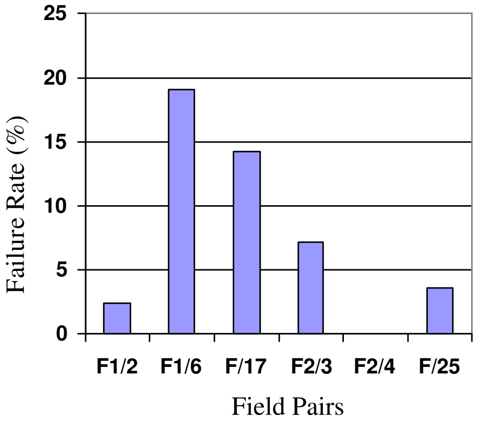


Figure 5.10: The failure rates of translation estimation for six ETDRS field pairs.

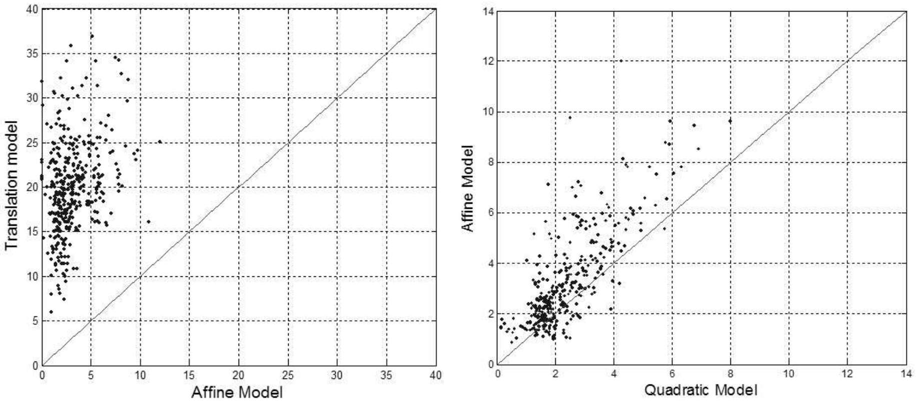
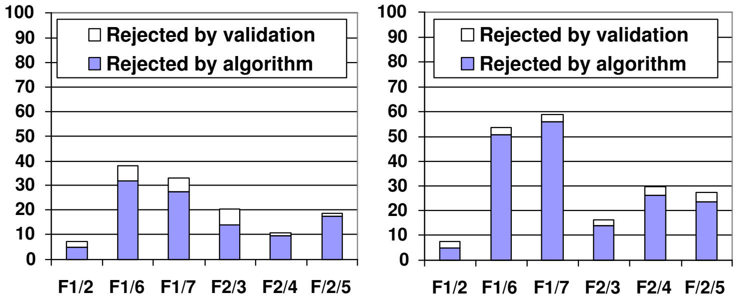


Figure 5.11: Error comparisons for consecutive transformation models.

a higher-order model. From Table 5.5, the error of the affine model is 2.68 pixels which is much less than that of the translation model, and the error is further reduced to 1.92 for the quadratic model. The numerical values demonstrate that the hierarchical strategy can improve the robustness and accuracy of image registration progressively [38]. However, since the affine and quadratic models need at least 3 and 6 LPCs, respectively, many pairs cannot proceed to high-order models due to insufficient LPCs, and also there are some registered pairs are rejected by manual validation, the failure rates are shown in Fig. 5.12.

LPCs and SPCs. Although affine/quadratic models can greatly reduce the registration error, they have high failure rates due to the lack of sufficient LPCs in



(a) Affine Model (b) Quadratic Model

Figure 5.12: The failure rates of 468 image pairs which proceed from the translation model to affine and quadratic models based on LPCs only.

many field pairs, especially when using the quadratic model in fields 1/6 and 1/7 where few LPCs exist. The use of SPCs could reduce failure rates by adding more feature points for feature-based registration. As shown in Table 5.5, after involving SPCs, all 468 pairs proceed to affine/quadratic models, among which 443 pairs and 423 pairs are manually validated for the affine and quadratic models, respectively. The overall failure rates are shown in Fig. 5.13. Moreover, SPCs slightly reduce for the registration error of affine model (from 2.68 to 2.59). It is interesting to note that SPCs significantly improve the accuracy of the quadratic model in fields 1/2 and 2/3 where LPCs are usually sufficient, and SPCs have some negative effect in other field

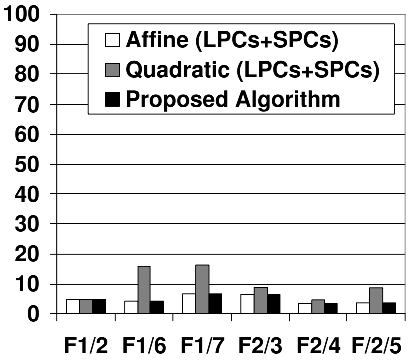


Figure 5.13: The failure rates (after manual validation) of affine/quadratic models (using both LPCs and SPCs) and the proposed algorithm for six field pairs.

pairs where LPCs are often limited. This is because SPCs are less reliable and they can only provide some auxiliary information to LPCs for image registration.

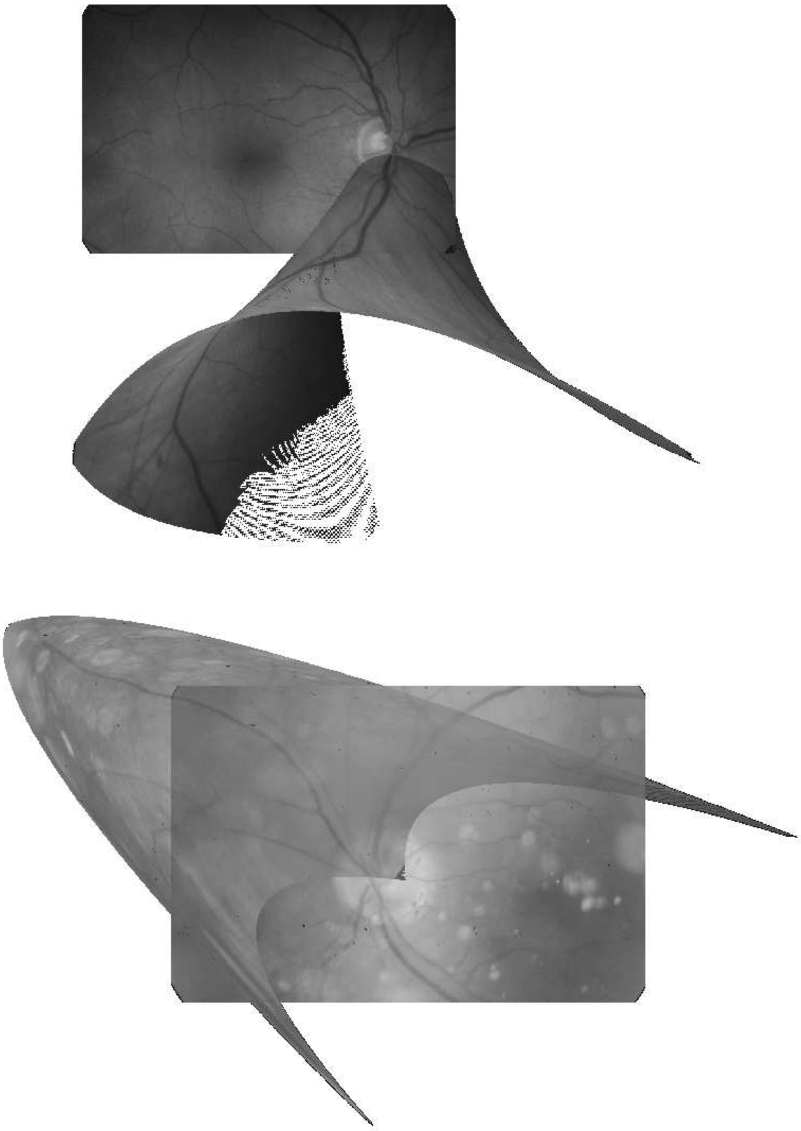


Figure 5.14: Two examples of using the quadratic transformation on fields 2/5 and fields 1/6 with insufficient LPCs, where the registration errors are 0.4 and 1.4.

5.4.4 Discussions of the Proposed Algorithm

Above techniques constitute the major components of the proposed algorithm. It is worth mentioning that the registration error may not be always trustable to evaluate the registration performance, especially when LPCs are limited and SPCs are involved for higher-order model estimation. Fig. 5.14 shows two examples where the registration errors fails to indicate the accuracy of high-order registration results. Therefore, we have to effectively combine above techniques into one flow where the proposed algorithm is able (1) to adaptively select an appropriate transformation model, (2) to determine whether SPCs have to be involved, and more importantly, (3) to reject invalid registered pairs. As mentioned before, we have developed a set of if-then conditions based on the simulation results which allow the algorithm to achieve these objectives. We summarize these conditions as follow.

- Condition 1. Translation Model: The translation model is essential to the whole algorithm, and its correctness is validated by the ECC output. If energy concentration and peak distinction are sufficiently large, i.e., $\Psi_3 > 13\%$ or $\Phi > 2.0$, the translation model is accepted based on which the IQA is performed, and the registration proceeds to higher-order models. Otherwise, the algorithm terminates. This condition can reject 36 pairs out of 39 pairs where the translation estimation fails due to poor image quality and significant geometric distortion.

- Condition 2. Affine/Quadratic Models: For ETDRS images, the affine model is usually sufficient in most field pairs, and going to the quadratic model may be risky or could be wrong due to small overlaps. No matter whether or not SPCs are involved, the quadratic model is only applied when there are at least 6 LPCs. It is found that most pairs undergoing the quadratic model are in the fields 1/2 and 2/3.

- Condition 3. LPCs and SPCs: There are two cases under which SPCs have to be involved: (1) When the number of LPCs is less than 3; (2) Or if LPCs cluster together in a small area, i.e., $\frac{H_o}{\sigma_x} > 4.0$ or $\frac{W_o}{\sigma_y} > 4.0$. We found that the sparsely distributed LPCs usually lead to undesired registration results, even though the reported registration errors are small. LPCs are expected to be uniformly distributed in the overlaps, otherwise SPCs will be involved.

Above empirical conditions are incorporated into the proposed algorithm, as shown in Fig. 5.2, which is able to provide a high success rate and low registration errors, as shown in Fig. 5.13 and Table 5.5. On the on hand, the success rate of the proposed algorithm (445 out of 468, i.e., 95.1%) is higher than that of the affine model without/with sampling (372/443 out of 468, i.e., 79.5%/94.7%). On the other hand, the registration error (2.04) is lower than that the affine model without/with sampling (2.68/2.59), and it is close the that of the quadratic model without/with sampling (1.92/1.94). One registration example of the proposed algorithm is shown in Fig. 5.16, where the quadratic model is applied in fields 2/3 and the SPCs are involved in fields 1/6, 2/3, and 2/4. This six field pairs also pass the IQA with the good quality of ETDRS field coverage according to Table 5.4.

Moreover, the proposed algorithm is able to work well for retinal image pairs that are defocused/blur with very poor contrast as well as image pairs with pathologies, as shown in Fig. 5.17. In Fig. 5.17(a), blood vessels are barely perceptible and there is no LPC. In Fig. 5.17(b), spot lesions are proliferated across both images. The proposed algorithm still can successfully registers these image pairs. A single area-based or feature-based approach alone may not be sufficient for ETDRS image registration. The proposed hybrid registration algorithm shows the promising performance on ETDRS images.

Overall speaking, there are a few major characteristics of the proposed algorithm. (1) It can handle retinal images with relatively small overlaps, and the performance is less dependent on the size of overlaps due to the power of ECC estimation on binary vascular tree images. (2) No LPC is required for the algorithm. (3) The algorithm selects an appropriate model for each field pair and involves SPCs when it is necessary. However, the proposed algorithm may fail when the geometric distortion between image pairs is significant (unlikely to happen in the ETDRS case), where the ECC-based translation estimation may be incorrect.

5.5 Conclusions

This chapter presents an ETDRS retinal image registration algorithm that effectively combines both area-based and feature-based methods into one flow. Three empirical conditions are used (1) to select an appropriate transformation model, (2) to determine whether the sampling process is needed, and (3) to reject invalid registration results, so that we can maximize the success rate and minimize the registration error. The proposed method can be used for the IQA purpose in terms of ETDRS field definition and to facilitate the implementation of ETDRS protocols in clinical trails.

Table 5.1: Field Coverage Specification

Fields	**Specifications**
Field 1	Centered at optic disc.
Field 2	Centered at macula.
Field 3	The center of the macula appears approximately mid-way between the edge and the center of the field.
Field 4	The lower edge of the field is tangent to a horizontal line passing through the upper edge of the optic disc & the nasal edge of the field is tangent to a vertical line passing through the center of the disc.
Field 5	The upper edge of the field is tangent to a horizontal line passing through the lower edge of the optic disc & the nasal edge of the field is tangent to a vertical line passing through the center of the disc.
Field 6	The lower edge of the field is tangent to a horizontal line passing through the upper edge of the optic disc & the temporal edge of the field is tangent to a vertical line passing through the center of the disc.
Field 7	The upper edge of the field is tangent to a horizontal line passing through the lower edge of the optic disc & the temporal edge of the field is tangent to a vertical line passing through the center of the disc.

Table 5.2: The Transformation Models For 2-D Retinal Registration

Model	Transformation Models	DOF
Translation	$\begin{pmatrix} p_x \\ p_y \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & m_2 \\ 0 & 1 & m_8 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} q_x \\ q_y \\ 1 \end{pmatrix}$	2
Affine	$\begin{pmatrix} p_x \\ p_y \\ 1 \end{pmatrix} = \begin{pmatrix} m_0 & m_1 & m_2 \\ m_6 & m_7 & m_8 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} q_x \\ q_y \\ 1 \end{pmatrix}$	6
Quadratic	$\begin{pmatrix} p_x \\ p_y \end{pmatrix} = \begin{pmatrix} m_0 & m_1 & m_2 & m_3 & m_4 & m_5 \\ m_6 & m_7 & m_8 & m_9 & m_{10} & m_{11} \end{pmatrix} \begin{pmatrix} q_x \\ q_y \\ 1 \\ q_x^2 \\ q_y^2 \\ q_x q_y \end{pmatrix}$	12

Table 5.3: The energy concentration of binary image-based ECC vs. logical operation XOR

Fields	ECC: Avg. Ψ		XOR: Avg. Ψ	
	Ψ_1	Ψ_3	Ψ_1	Ψ_3
1/2	41.2%	71.5%	0.06%	0.17%
1/6	17.4%	27.6%	0.06%	0.17%
1/7	18.6%	29.4%	0.08%	0.23%
2/3	42.1%	58.6%	0.05%	0.15%
2/4	48.0%	72.8%	0.09%	0.25%
2/5	33.8%	67.8%	0.05%	0.15%
Avg. Ψ	40.6%	54.6%	0.07%	0.19%

Table 5.4: Ideal vertical/horizontal displacements

Field Pair	Desired vertical/horizontal displacements $(T_x^{'}/T_y^{'})$ $H = 600$ and $W = 900$ are the image size.	
	$T_x^{'}$	$T_y^{'}$
1/2	H	$0.5W$
1/6	$0.5H - 0.5DD$	$0.5W - 0.5DD$
1/7	$0.5H - 0.5DD$	$0.5W - 0.5DD$
2/3	H	$0.5W$
2/4	$0.5H - 0.5DD$	W
2/5	$0.5H - 0.5DD$	W

Table 5.5: The registration error and overlaps between fields.

Fields	Median Overlap	Median Registration Errors (pixels)					
		Translation	LPCs-Only Algorithm		LPCs-SPCs Algorithm		Proposed Algorithm
			Affine	Quadratic	Affine	Quadratic	
1/2	52.97%	$19.21^{(84,82)}$	$3.15^{(78,76)}$	$2.36^{(78,76)}$	$2.86^{(82,78)}$	$2.09^{(82,78)}$	$2.15^{(82,78)}$
1/6	18.91%	$27.78^{(84,68)}$	$2.39^{(47,43)}$	$1.42^{(34,32)}$	$2.33^{(69,66)}$	$1.87^{(69,58)}$	$1.91^{(69,66)}$
1/7	22.50%	$22.58^{(84,72)}$	$2.55^{(53,49)}$	$1.75^{(32,30)}$	$2.33^{(73,68)}$	$2.09^{(73,61)}$	$1.97^{(73,68)}$
2/3	44.24%	$20.80^{(84,78)}$	$3.64^{(68,63)}$	$2.95^{(68,66)}$	$3.43^{(79,74)}$	$2.41^{(79,72)}$	$2.62^{(79,74)}$
2/4	25.39%	$19.10^{(84,84)}$	$1.97^{(76,75)}$	$1.08^{(62,59)}$	$1.96^{(84,81)}$	$1.19^{(84,80)}$	$1.79^{(84,81)}$
2/5	32.17%	$18.38^{(84,81)}$	$2.43^{(67,66)}$	$1.96^{(62,59)}$	$2.66^{(81,78)}$	$2.02^{(81,74)}$	$2.11^{(81,78)}$
Average	32.70%	$21.30^{(504,465)}$	$2.68^{(389,372)}$	$1.92^{(336,322)}$	$2.59^{(468,443)}$	$1.94^{(468,423)}$	$2.04^{(468,445)}$

Totally, there are 84 retinal image pair for each field pair. Some pairs might be unable to perform registration - without the sampling process.

(N_1, N_2) : N_1 indicates the number of image pairs are registered.

N_2 specifies the number of registered pairs after manual validation.

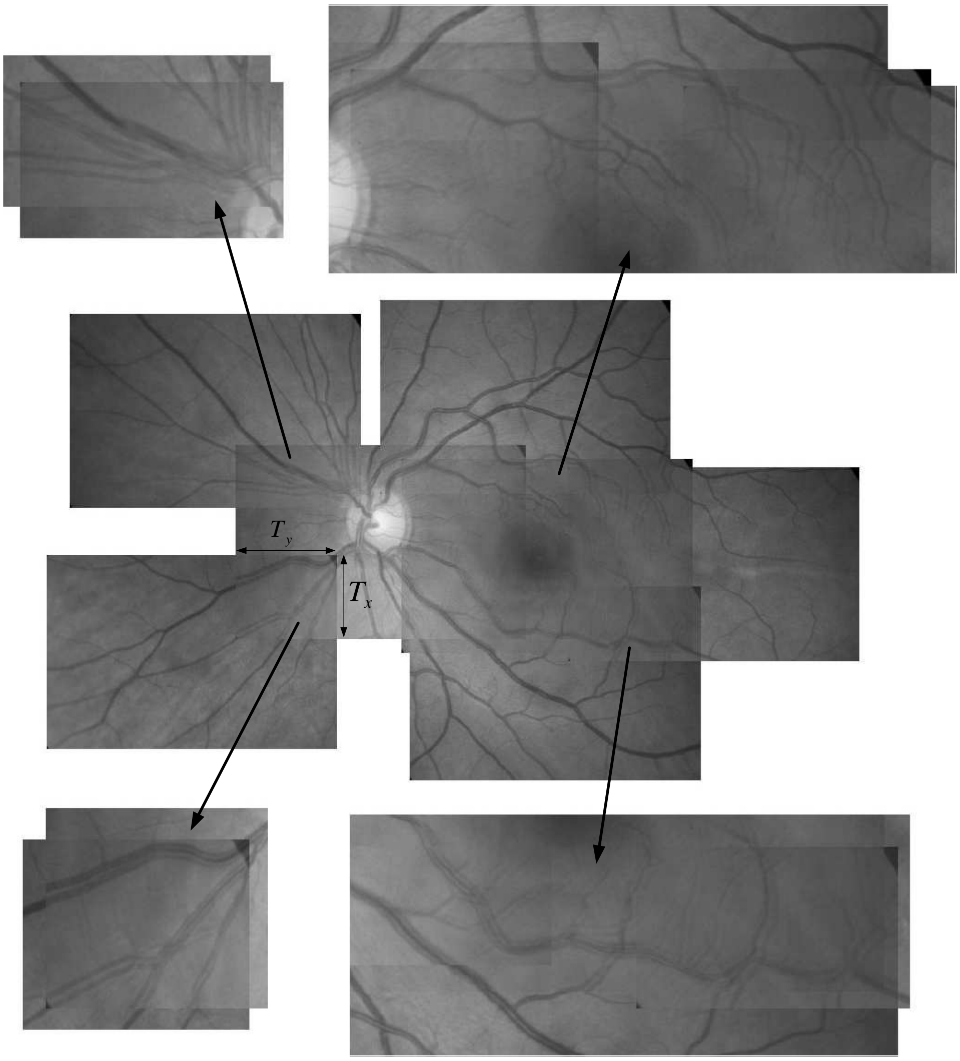


Figure 5.15: A translation model-based registration result where the vertical/horizontal displacements in fields 1/7 are depicted. The registration errors are as follows: fields 1/2 = 15.81, fields 1/6 = 22.06, fields 1/7 = 13.83, fields 2/3 = 12.60, fields 2/4 = 21.56, and fields 2/5 = 18.97. The median error is 17.39.

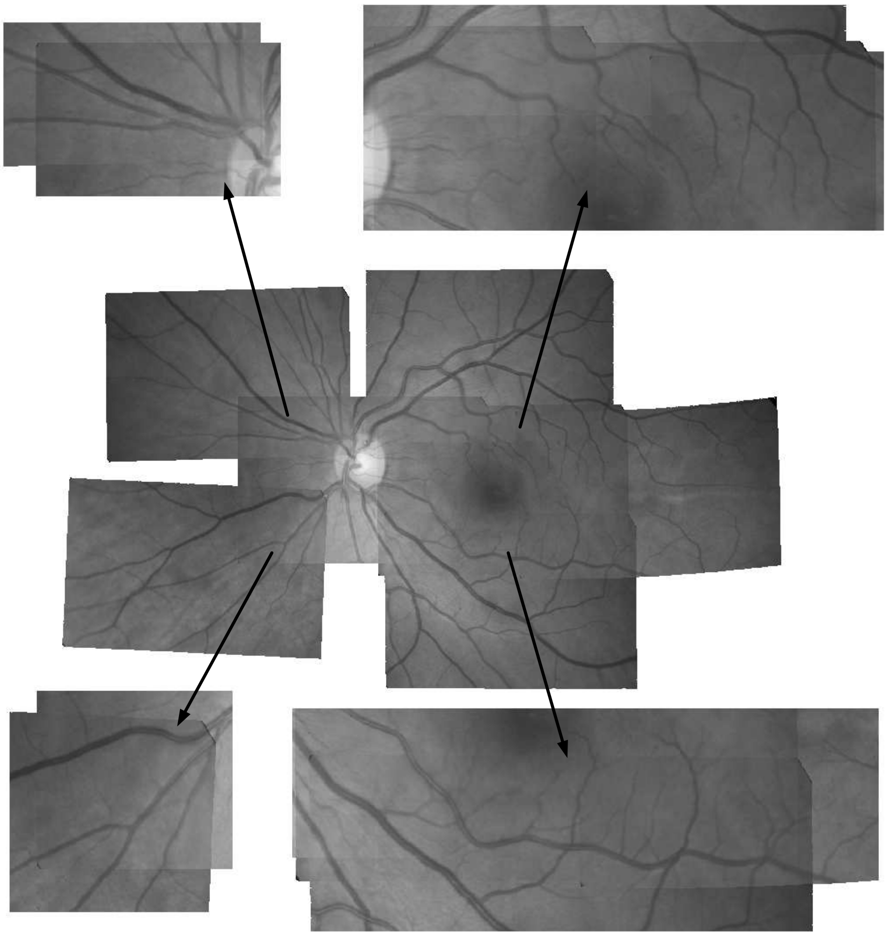
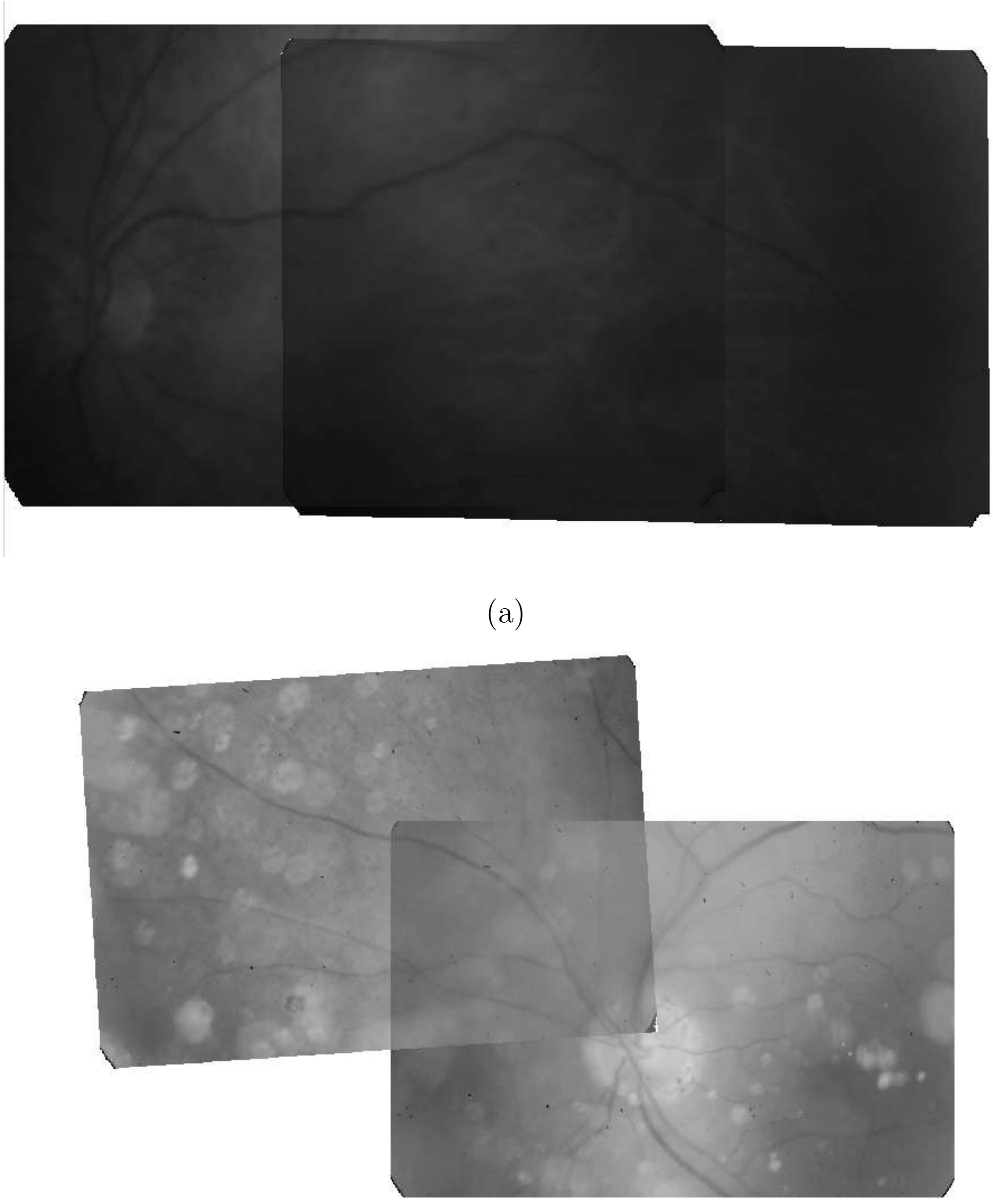


Figure 5.16: The final registration result of an ETDRS image set. The quadratic transformation is applied to fields 2/3. The affine model is applied to fields 1/2, 1/6, 1/7, 2/4, and 2/5. The registration errors are given as as follows: fields 1/2: 2.00, fields 1/6: 0.52, fields 1/7: 1.23, fields 2/3: 1.66, fields 2/4: 2.78, and fields 2/5: 1.49. The median error is 1.58. The sampling points (SPCs) are involved in fields 1/6, 2/3 and fields 2/4.



(a)

(b)

Figure 5.17: (a) An example of a poor quality retinal image pair with no LPC in the overlap. (b) An example of a retinal image pair with pathology.

CHAPTER 6

3-D Retinal Surface Reconstruction

6.1 Introduction

Visual reconstruction is a process to recover a 3-D scene or a model from multiple images. It is usually referred to as the structure from motion (SFM) problem. The process usually recovers objects' 3-D shapes, cameras' poses (positions and orientations), and cameras' internal parameters (focal lengths, principle points, and skew factors) as demonstrated in Fig. 6.1. Many possible camera models exist. A perspective projection is the standard. However, other projections, e.g., affine and orthographic projections, are proved simpler and practical for a distant camera. The main differences between projections are the required level of calibration. Stereo techniques, e.g., cepstrum, are also commonly used for depth estimation because of their simplicity. They requires only a stereo pair and it does not need camera calibration. However, cepstrum-based approaches can only provide qualitative analysis for depth estimation. Although the general SFM problem has been extensively studied, surprisingly there are not much researches published and dedicated to 3-D retinal reconstruction.

In this work [102], 3-D retinal surface reconstruction refers to the global geometric shape recovery where we impose the geometrical constraint of human retina. A simple stereo technique does not work for 3-D retinal surface reconstruction due to the unknown camera parameters and the complex lens distortion of fundus imaging. Deguchi et. al. [3, 62] modelled both the fundus camera and the human cornea with a virtual optical lens. They utilized the fact that a retinal surface has a spherical shape and imaging a sphere through the eye lens results in a quadratic surface. The camera was calibrated by using the two-plane method. Then, eye lens parameters were estimated iteratively to recover fundus's spherical surface. Choe et. al. [4] used PCA-based directional filters to extract candidate seed points (Y features). A gradient descent was employed to model Y features and match pairs of features. A plane-and-parallax was employed to estimate the epipolar geometry because a near-planar retinal surface can obstruct a traditional fundamental matrix estimation. The stereo pair is rectified. Then, a Parzen window-based mutual information was used to generate dense disparity map.

Two additional characteristics in ETDRS images make 3D retinal reconstruction even more challenging. First, the largest area in a retinal image is textureless. Second, ETDRS images have small overlaps between different fields. Both challenges present complexities to 3D surface reconstruction due to insufficient information and inadequate features. We assume an affine camera because of the two following reasons: (1) the ETDRS imaging standard specifies a 30° field of view each eye (narrow field of view); (2) each retinal image has small depth variation. We have derived an affine camera from the standard projective camera and show that an affine camera is an appropriate model for retinal surface reconstruction from ETDRS images which can be found in section 6.2. Retinal images, first, need to be corrected due to lens distortions. An initial affine shape is obtained from a previously proposed factorization method. The affine shape and the camera model are, then, jointly refined with affine bundle adjustment. Later, the geometrical constraint of human retina is imposed to recover a Euclidean structure up to a similarity transform and we introduce an efficient point-based linear approach to approximate the retinal spherical surface. Compared with previous methods, the proposed one is robust, efficient, and less sensitive to noise and lens distortion due to the linear nature of the affine camera. It also does not require a very accurate camera calibration

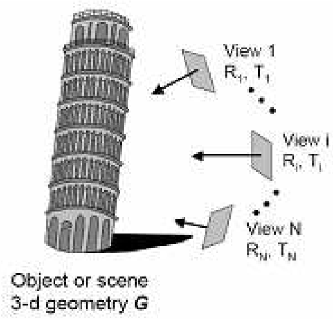


Figure 6.1: The structure from motion (SFM) problem.

The organization of this chapter is as follows. The introduction of 3-D surface reconstruction is given in this section. We have derived an affine camera from the standard projective camera and show mathematical proof of its condition in section 6.2. Then, Section 6.4 describes the implementation of the proposed algorithm. Simulation results and the performance of our algorithm are presented in section 6.5.

6.2 Affine Camera for 3-D Retinal Surface Reconstruction

We provide a mathematical proof of an affine camera from the standard projective camera and derive its condition for good reconstruction performance. Consequently, it also supplies reasons to justify the use of an affine camera representing a fundus camera. Let us start with general projective camera. From chapter 3, Equation 3.6 can be written in homogeneous coordinate system as

$$m = \begin{bmatrix} (f_x R_1^T + s R_2^T)M + f_x D_x + s D_y + c_x \\ f_y R_2^T M + f_y D_y + c_y \\ R_3^T M + D_z \end{bmatrix}, \tag{6.1}$$

where m and M denote image and world homogeneous coordinates respectively. R_i^T is the $i-$th row of the rotation matrix $\mathbf{R}$ and $D_x = -R_1^T t_x$, $Dy = -R_2^T t_y$, $D_z = -R_3^T t_z$ where t_x, t_y and t_z are translation parameters. f represent the focal length of the camera. (c_x, c_y) and s are principal point and skew angle of image's pixel respectively.

Because camera's principal ray direction is R_3^T (see Fig. 3.3), $D_z = -R_3^T t_z$ is the distance between camera center and the world's origin in the direction of camera's principal ray and $R_3^T M_i$ is the distance between 3-D point i and the world's origin in the direction of camera's principal ray (relative depth in principal ray direction). If we assume that the relative depth in principal ray direction $R_3^T M_i$ is small compared to D_z, then Equation 6.1 can be rewritten as

$$m = \begin{bmatrix} (f_x R_1^T + s R_2^T)M + f_x D_x + s D_y + c_x \\ f_y R_2^T M + f_y D_y + c_y \\ D_z \end{bmatrix}. \tag{6.2}$$

Therefore, we get another camera model which can be represented in a mathematical form as

$$\mathbf{T}_{affine} = \begin{bmatrix} f_x R_1^T + s R_2^T & f_x D_x + s D_y + c_x D_z \\ f_y R_2^T & f_y D_y + c_y D_z \\ \mathbf{0}_3^T & D_z \end{bmatrix}. \tag{6.3}$$

Equation 6.3 has the similar form as a general affine transform. Hence, it is termed affine projection. Some call it an affine camera or a weak-perspective camera. Compare Equations 6.1 and 6.2, the difference is only in the last row. Hence, we will rewrite both equations as follows

$$m_{projective} = \begin{bmatrix} \tilde{x} \\ \tilde{y} \\ d + \delta \end{bmatrix} \quad m_{affine} = \begin{bmatrix} \tilde{x} \\ \tilde{y} \\ d \end{bmatrix}, \tag{6.4}$$

where $\tilde{x}$ and $\tilde{y}$ represent $(f_x R_1^T + s R_2^T)M + f_x D_x + s D_y + c_x$ and $f_y R_2^T M + f_y D_y + c_y$ respectively. d denotes D_z and δ denotes $R_3^T M$

If we divide elements in Equation 6.4 with the third element to create non-homogeneous coordinates, we have the following.

$$\hat{m}_{projective} = \begin{bmatrix} \frac{\tilde{x}}{d+\delta} \\ \frac{\tilde{y}}{d+\delta} \end{bmatrix} \quad \hat{m}_{affine} = \begin{bmatrix} \frac{\tilde{x}}{d} \\ \frac{\tilde{y}}{d} \end{bmatrix}. \tag{6.5}$$

Then,

$$\frac{\hat{m}_{projective}}{\hat{m}_{affine}} = \frac{d}{d+\delta}. \tag{6.6}$$

From Equation 6.6, we can conclude that if δ or relative depth in principal ray direction is small compared with D_z then an affine camera is a good choice for camera model. In a retinal image case, the retinal surface depth is relatively small compared with the distance from the retina to a fundus camera. Therefore, the affine camera is a preferred camera model to represent a fundus camera.

6.3 Correspondence Selection

Point correspondences are automatically selected by using our proposed hybrid retinal image registration (Chapter 5) [58]. The algorithm can be summarized as follows. First, binary vascular trees are extracted from retinal images. Second, zeroth-order translation is estimated by maximizing mutual information based on the binary image pair. Specifically, a local entropy-based peak selection scheme and a multi-resolution searching strategy are developed to improve the accuracy and efficiency of translation estimation. Third, with a translation constraint and two types of point correspondences, landmark points and sampling points, a feature-based registration method is used along with other decision-making criteria to estimate higher-order transformation and further refine point correspondences.

6.4 Proposed Framework

The proposed framework comprises multiple steps which related to our other previous works. The flowchart is shown in Fig. 6.2. First, binary vascular trees are extracted from retinal images using a modified local entropy-based thresholding method [95, 96].

Next, a 2-D registration method is perform to generate more feature points as well as to refine feature points [58, 59]. Then, images are corrected by removing lens distortions. An affine factorization approach is employed to recover an initial retina's affine surface. Inspired by projective bundle adjustment, an affine bundle adjustment is proposed to refine 3-D shape and cameras. After that, constraints are imposed to recover Euclidean structure up to a scale factor. Then, geometric constraint is applied to generate denser feature points. Finally, a point-based approach is introduced to approximate a retinal spherical surface.

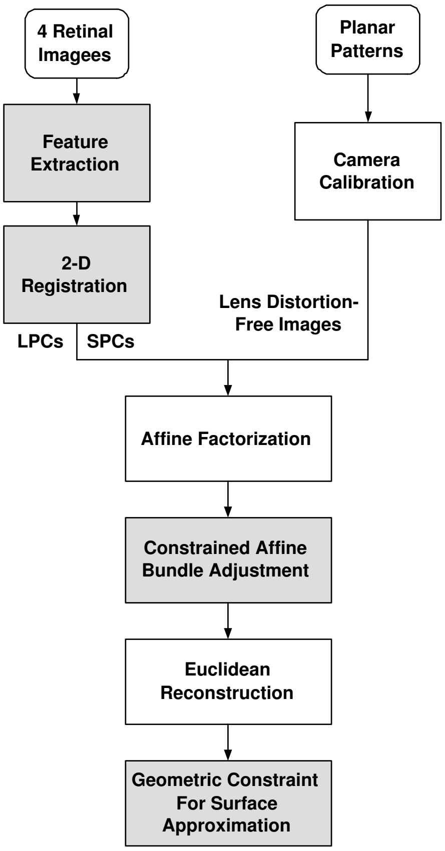


Figure 6.2: The flowchart of the proposed algorithm.

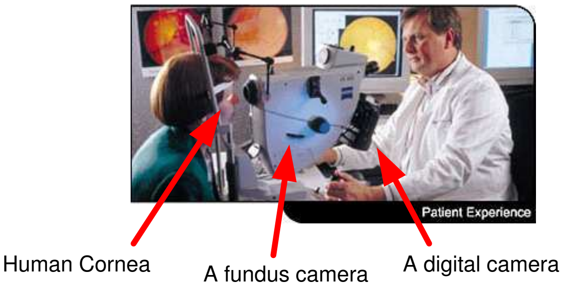


Figure 6.3: Retinal images are obtained from a fundus camera which composes of an actual camera and a digital camera attached to a fundus camera.

6.4.1 Lens Distortion Removal

As shown in Fig. 7.1, there is a series of optics involved in the retinal imaging process, which includes an actual fundus camera, an digital camera, and the human cornea. All of these optics could be modeled as one virtual lens that contributes to certain lens distortion, e.g., radial distortion, in retinal images [3, 62]. The lens distortion has to be removed prior to 3D retinal surface reconstruction. The fundus camera is not a linear camera and an individual eye lens can be considered as an additional camera which causes additional distortions. A relationship between a retinal surface (world points) **M** and its retinal image (image points) **m** is given by

$$m = \mathbf{K}[\mathbf{R} - \mathbf{R}t]M, \tag{6.7}$$

where **K** is the fundus camera's intrinsic parameters. **R** and t represent rotations and translations of a fundus camera.

In this work, we employ the planar pattern calibration method proposed in [103] to remove the lens distortion in retinal images. Zhang [103] has suggested the use of a planar pattern with different orientations. We can assume the model plane is on $Z = 0$. The above equation can be rewritten as

$$
\begin{aligned}
\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} &= \mathbf{K} \begin{bmatrix} R_1 & R_2 & R_3 & d \end{bmatrix} \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} \\
&= \mathbf{K} \begin{bmatrix} R_1 & R_2 & d \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix},
\end{aligned} \tag{6.8}
$$

where R_i denotes $i-$th column of rotation matrix $\mathbf{R}$ and d denotes $-\mathbf{R}t$.

Retinal image coordinates m and retinal surface coordinates M are related by a 3×3 homography $\mathbf{H} = \mathbf{K}[R_1 \ R_2 \ d]$ where $\mathbf{H}$ is defined up to an arbitrary scale factor. Therefore, we have

$$
H_1^T K^{-T} K^{-1} H_2 \quad = 0, \tag{6.9}
$$

where $\mathbf{H} = [H_1 \ H_2 \ H_3]$.

Equation 6.9 gives the fundus camera's intrinsic parameters $\mathbf{K}$. Once $\mathbf{K}$ is known, extrinsic parameters can be computed from Equation 6.8. The solution can be solved through minimizing an algebraic distance then refine it through Levenberg-Marquardt algorithm with a following cost function

$$
\sum_{i=1}^{v} \sum_{j=1}^{n} \| m_{ij} - \check{m}(\mathbf{K}, \mathbf{R}_i, d_i, M_j) \|^2, \tag{6.10}
$$

where we have v views/images and n correspondences. $\check{m}(\mathbf{K}, \mathbf{R}_i, d_i, M_j)$ is the projection of point M_j in image i. R_l denotes $l-$th row of rotation matrix $\mathbf{R}$ and d denotes $-\mathbf{R}t$. We have created a set of chessboard images using the actual fundus camera. Then we obtained all lens parameters by using the camera calibration toolbox [1] which are used to remove the lens distortion in real retinal images.

[1] http://www.vision.caltech.edu/bouguetj/calib_doc/

A pinhole camera assumes world points, image points, and optical center are collinear. However, these assumptions are not hold in a fundus camera. The fundus camera is not a linear camera and an individual eye lens can be considered as an additional camera which causes additional distortions in retinal images. The most common deviation is radial distortion which causes the actual image points to be displaced radially in the image plane [104]. In addition, centers of curvature of lens surfaces are not always strictly collinear which cause tangential distortion [104]. Furthermore, several other distortions also exist, e.g. Linear distortion occurs when image axes are not orthogonal. Prism distortion is caused by imperfection of lens manufacturing . Heikkila et.al. [105] has proposed to use radial and tangential distortions since other distortions are either insignificant or can be included in radial and tangential distortions. The radial distortion can be expressed

$$\begin{aligned} \delta x_{radial} &= \tilde{x}(k_1 r^2 + k_2 r^4 + \ldots), \\ \delta y_{radial} &= \tilde{y}(k_1 r^2 + k_2 r^4 + \ldots), \end{aligned} \tag{6.11}$$

where $(\tilde{x}, \tilde{y})$ are image coordinates in metric unit, $r = \sqrt{\tilde{x}^2 + \tilde{y}^2}$, and k_1, k_2, $\ldots$ are coefficients for radial distortion. The expression for tangential distortion is

$$\begin{aligned} \delta x_{tangential} &= 2p_1\tilde{x}\tilde{y} + p_2(r^2 + 2\tilde{x}^2) \\ \delta y_{tangential} &= p_1(r^2 + 2\tilde{y}^2) + 2p_2\tilde{x}\tilde{y}, \end{aligned} \tag{6.12}$$

where p_1 and p_2 are coefficients for tangential distortion. Image pixel (x, y) and metric $(\tilde{x}, \tilde{y})$ coordinates are related by

$$\begin{aligned} x &= c_x + f_x\tilde{x} + s\tilde{y} \\ y &= c_y + f_y\tilde{y}, \end{aligned} \tag{6.13}$$

where (c_x, c_y) denotes a principal point, f_x and f_y represent focal length, and s denotes skew factor. With lens distortion, Equation 6.10 becomes

$$\sum_{i=1}^{v}\sum_{j=1}^{n} \|m_{ij} - \breve{m}(\mathbf{K}, \mathbf{R}_i, k_1, k_2, p_1, p_2, d_i, M_j)\|^2, \tag{6.14}$$

where we have v views/images and n correspondences. $\check{m}(\mathbf{K}, \mathbf{R}_i, k_1, k_2, p_1, p_2, d_i, M_j)$ is the projection of point M_j in image i. R_l denotes $l-$th row of rotation matrix $\mathbf{R}$ and d denotes $-\mathbf{R}t$. k_1 and k_2 are coefficients for radial distortion. p_1 and p_2 are coefficients for tangential distortion.

Initialization for all of the parameters need to be identified first. Lens distortions parameters can be initialized to be zeros. Intrinsic parameters are initialized by Equation 6.9. Then, extrinsic parameters are initialized by Equation 6.8. Finally, Equation 6.14 is optimized through a Levenberg-Marquardt algorithm. We believe not all lens distortions are removed at his step. The process of camera calibration is meant to remove lens distortion introduced by the fundus camera. The distortion induced by the human cornea is still in the image and it could be removed by the later constrained bundle adjustment algorithm discussed in chapter 7.

6.4.2 Initial Retinal Affine Surface

With lens distortion-free retinal images, we assume an affine projection for a fundus camera because (1) the ETDRS imaging standard specifies a 30° field of view each eye (narrow field of view); (2) each retinal image has small depth variation. We have derived an affine camera from the standard projective camera and provide mathematical proof for its condition in section 6.2. An affine structure from motion problem has been investigated in section 3.5. We use affine factorization [78] method for initial reconstruction because the approach can accommodate multiple images and utilize the use of all feature points. Suppose there are f retinal images and n point correspondences from each image.

$$\hat{m}_i = \mathbf{A}_i \hat{M} + d_i, \quad i \in 1, \dots f, \tag{6.15}$$

where $\hat{m}$ and $\hat{M}$ denotes retinal image and retinal surface non-homogeneous coordinates respectively. $\mathbf{A}$ is an arbitrary 2×3 matrix and d represents any 2×1 vector.

With the similar idea as LCF, select one point as an origin or select a center of mass. The Equation 6.15 becomes

$$\begin{aligned} \triangle \hat{m}_i = \mathbf{A}_i(\triangle \hat{M}), \quad i \in 1, \ldots f \\ \mathbf{W} = \mathbf{P}\mathbf{M}, \end{aligned} \tag{6.16}$$

where $\mathbf{W}$ denotes a $2f \times n$ matrix containing set of 2D point correspondences with respect to the center of mass. $\mathbf{M}$ denotes a $3 \times n$ matrix containing affine shape of the retinal surface. $\mathbf{P}$ denotes $2f \times 3$ matrix comprising f fundus camera model.

With the rank theorem, $\mathbf{W}$ is at most rank three. Singular value decomposition (SVD) is used to factorized $\mathbf{W} = \mathbf{U}\mathbf{W}\mathbf{V}^T$. Therefore, $\mathbf{P}$ and $\mathbf{M}$ are the left and right eigenvectors corresponding to the three greatest eigenvalues.

$$\begin{aligned} \mathbf{P} \quad &= \mathbf{U}_3 \\ \mathbf{M} \quad &= \mathbf{W}_3\mathbf{V}_3^T, \end{aligned} \tag{6.17}$$

where $\mathbf{P} = P_1, P_2, \ldots, P_f$ denotes set of f fundus cameras and M denotes affine retinal surface.

6.4.3 Affine Bundle Adjustment

Inspired by projective bundle adjustment [73], we introduce affine bundle adjustment for the affine camera. Bundle adjustment is an optimization process of refining a visual reconstruction to produce *jointly optimal* structure and viewing parameters [106] [73] [107]. In other words, all the parameters, structure and camera parameters, are optimized simultaneously. It is usually formulated as a nonlinear least square problem.

In our case, we want to estimate and refine affine cameras $\breve{\mathbf{P}}$ and the affine retinal surface $\breve{\mathbf{M}}$ simultaneously. $\breve{m}(\breve{P}_i, \breve{M}_j)$ is a projection of point $\breve{M}_j$ in the ith image. We try to minimize distance between the projected point $\breve{m}(\breve{P}_i, \breve{M}_j)$ and the observed point m_{ij} by optimizing the 3-D point in the affine space and the affine cameras. If the

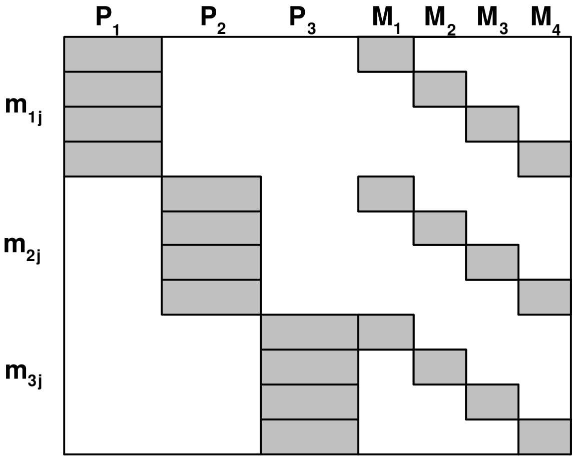


Figure 6.4: The sparse Jacobian matrix for affine bundle adjustment comprises 3 cameras and 4 points.

2D correspondences are noise free, this distance should be zero. If noise distribution is assumed to be zero mean, isotropic and Gaussian with certain variance, then the maximum likelihood estimation is equivalent to the solution to the minimum mean square error problem defined below,

$$\min_{\breve{P}_i,\breve{M}_j} \sum_{i=1}^{v} \sum_{j=1}^{n} \|\breve{m}(\breve{P}_i, \breve{M}_j) - m_{ij}\|^2. \tag{6.18}$$

Since bundle adjustment can become an extremely large minimization problem because of large number of parameters, instead of directly use of the Levenberg-Marquardt algorithm, a sparse Levenberg-Marquardt algorithm is employed to efficiently reduce computational cost. There are two main parameters, affine fundus cameras **P** and the affine retinal surface **M**, to be minimized. We, first, partition a Jacobian matrix according to the two main parameters. We get

$$\begin{aligned} \mathbf{J}\delta &= \epsilon, \\ [\mathbf{A}|\mathbf{B}] \begin{bmatrix} \delta_a \\ \hline \delta_b \end{bmatrix} &= \epsilon, \end{aligned} \tag{6.19}$$

where **J** represents a Jacobian matrix. **A** and **B** denote the first derivative of retinal

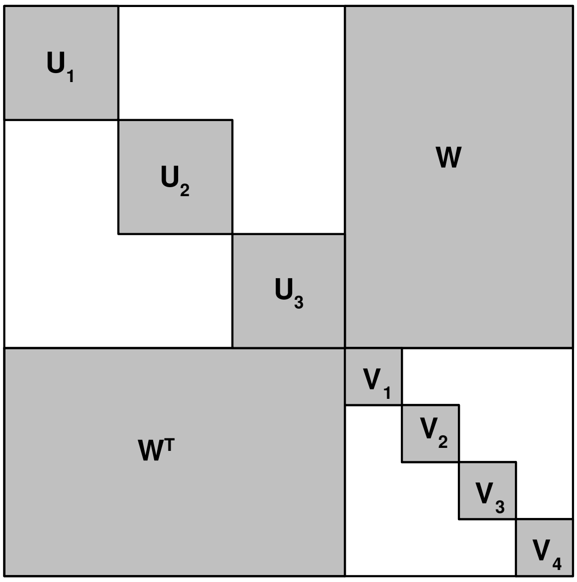


Figure 6.5: The sparse Hessian matrix comprises 3 cameras and 4 points.

image points with respect to affine fundus cameras and affine retinal surface respectively. The normal least mean square equation with zero mean, isotropic and Gaussian noise

$$\begin{aligned} \mathbf{J}^T\mathbf{J}\delta &= \mathbf{J}^T\epsilon \\ \left[\begin{array}{c|c} \mathbf{A}^T\mathbf{A} & \mathbf{A}^T\mathbf{B} \\ \hline \mathbf{B}^T\mathbf{A} & \mathbf{B}^T\mathbf{B} \end{array}\right] \left[\begin{array}{c} \delta_a \\ \hline \delta_b \end{array}\right] &= \left[\begin{array}{c} \mathbf{A}^T\epsilon \\ \hline \mathbf{B}^T\epsilon \end{array}\right] \\ \begin{bmatrix} \mathbf{U} & \mathbf{W} \\ \mathbf{W}^T & \mathbf{V} \end{bmatrix} \begin{bmatrix} \delta_a \\ \delta_b \end{bmatrix} &= \begin{bmatrix} \epsilon_a \\ \epsilon_b \end{bmatrix}. \end{aligned} \tag{6.20}$$

The form of a sparse Jacobian matrix of dimension $fn \times (6f + 3n)$ for affine bundle adjustment is illustrated in Fig. 6.4 and a sparse Hessian matrix is shown in Fig. 6.5. By multiplying both sides of Equation 6.20 with $\begin{bmatrix} \mathbf{I} & -\mathbf{W}\mathbf{V}^{-1} \\ 0 & I \end{bmatrix}$, we get

$$\begin{bmatrix} \mathbf{U} - \mathbf{W}\mathbf{V}^{-1}\mathbf{W}^T & 0 \\ \mathbf{W}^T & \mathbf{V} \end{bmatrix} \begin{bmatrix} \delta_a \\ \delta_b \end{bmatrix} = \begin{bmatrix} \epsilon_a - \mathbf{W}\mathbf{V}^{-1}\epsilon_b \\ \epsilon_b \end{bmatrix}. \tag{6.21}$$

According to Equation 6.21, we first find solution for δ_a. Then δ_a is used as additional input to find solution for δ_b. To solve the problem more efficient, the problem can be exploited by the fact that a specific residual is only dependent on one 3-D point and one camera which yields a very sparse structure (see Fig. 6.4 and Fig. 6.5).

6.4.4 Euclidean Reconstruction of Retinal Surface

To recover a retina's Euclidean surface from an affine surface, a 3 × 3 nonsingular matrix, $\mathbf{D}$ needs to be identified. From Equation 6.16, we get

$$\begin{aligned} \mathbf{W} &= \mathbf{P}\mathbf{D}\mathbf{D}^{-1}\mathbf{M} \\ &= (\mathbf{P}\mathbf{D})(\mathbf{D}^{-1}\mathbf{M}) \\ &= \acute{\mathbf{P}}\acute{\mathbf{M}}, \end{aligned} \tag{6.22}$$

where D is called a metric constraint.

Several different solutions for different affine camera projections were proposed. Tomasi and Kanade [78] proposed a solution for orthographic projection. Weinshall and Tomasi [108], [109] introduced a solution under weak-perspective camera. Poleman and Kanade [110] [111] proposed a solution for paraperspective projection. Quan [112], Kurata et.al. [113] attempted to congregate those solutions into one unified framework for general affine camera without having to calibrate the camera. The constraint formulations [112] are reviewed as follow

$$\begin{aligned} \mathbf{P}_i\mathbf{D} &= \mathbf{K}_i\mathbf{R}_i \\ \mathbf{P}_i\mathbf{D}\mathbf{D}^T\mathbf{P}_i^T &= \mathbf{K}_i\mathbf{R}_i\mathbf{R}_i^T\mathbf{K}_i^T = \mathbf{K}_i\mathbf{K}_i^T. \end{aligned} \tag{6.23}$$

If the images are assumed to be taken by the same affine camera, then the intrinsic parameters $\mathbf{K}$ are the same for every views. The following constraints are obtained

$$\arg\min_{\mathbf{X}} \sum_{i=1}^{f-1}((\frac{u_i^T\mathbf{X}u_i}{v_i^T\mathbf{X}v_i} - \frac{u_{i+1}^T\mathbf{X}u_{i+1}}{v_{i+1}^T\mathbf{X}v_{i+1}})^2 + (\frac{u_i^T\mathbf{X}v_i}{v_i^T\mathbf{X}v_i} - \frac{u_{i+1}^T\mathbf{X}v_{i+1}}{v_{i+1}^T\mathbf{X}v_{i+1}})^2), \tag{6.24}$$

where $\mathbf{P}_i = \begin{bmatrix} u_i^T \\ v_i^T \end{bmatrix}$ and $\mathbf{X} = \mathbf{D}\mathbf{D}^T$. Equation 6.24 can be minimized altogether by a maximum likelihood estimation with an assumption of zero mean, isotropic, and Gaussian noise. Procedure to calculate the metric constraint $\mathbf{D}$ is as follow. First, compute the Cholesky parameters $\mathbf{C}\mathbf{C}^T$ instead of $\mathbf{D}\mathbf{D}^T$ to ensure that the solution is positive-definite. Recover a rotation matrix $\mathbf{R}_1$ from the first view by QR factorization. Finally, $\mathbf{D}$ is achieved by $\mathbf{C}\mathbf{R}_1^T$.

6.4.5 Point-Based Surface Approximation

We take into an account of an eyeball's geometric constraint in order to approximate the 3-D retinal surface. We assume that eyeball is an approximated sphere. We introduce a point-based sphere fitting method. The method is accomplished by first selecting a reference point $M_k = (X_k, Y_k, Z_k)$ from 3-D point cloud. Every points has to satisfy the sphere equation shown below.

$$\begin{aligned} &(X_k - A)^2 + (Y_k - B)^2 + (Z_k - C)^2 = R^2 \\ &(X_j - A)^2 + (Y_j - B)^2 + (Z_j - C)^2 = R^2, \quad j \in 1, ..., n, j \neq k, \end{aligned} \tag{6.25}$$

where (A, B, C) and R are sphere's center point and radius respectively. Subtracting the two equations and rearranging the terms, we get

$$\begin{gathered} (X_k^2 - X_j^2) + (Y_k^2 - Y_j^2) + (Z_k^2 - Z_j^2) \\ = 2(X_k - X_j)A + 2(Y_k - Y_j)B + 2(Z_k - Z_j)C, \\ j \in 1, ..., n, j \neq k. \end{gathered} \tag{6.26}$$

Equation 6.26 is in a linear format. Sphere's center point (A, B, C) can be obtained by solving multiple linear equations. Then a radius R can be computed in a least mean square sense. A searching technique can be incorporated to accelerate the procedure finding the best reference point. Ideally speaking, every point has to satisfy the sphere equation. An error at a particular point j is calculated by the following equation:

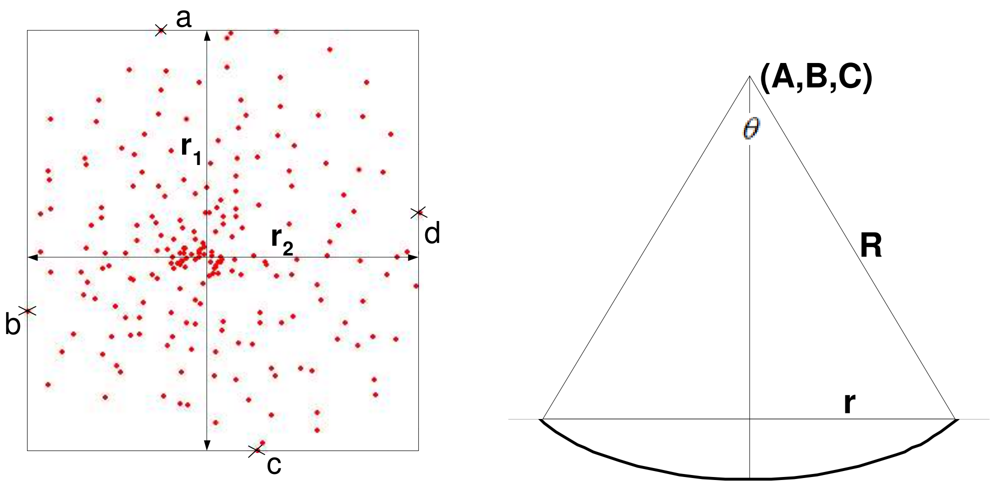
$$E_j^{(k)} = \| M_j^T \mathbf{Q}_k M_j \|, \tag{6.27}$$

where $\mathbf{Q}_k$ is a 4×4 sphere matrix by a reference point k. $E_j^{(k)}$ represents an error at point j by using $\mathbf{Q}_k$. By minimizing the following equation, the optimal sphere surface with best fitness to all points can be achieved.

$$\hat{\mathbf{Q}} = \arg \min_{k \in 1,\ldots,n} \sum_{j=1}^{n} E_j^{(k)}, \tag{6.28}$$

where $E_j^{(k)}$ is defined in Equation (6.27).

By observe the plot of reconstructed surface's errors versus 2-D spatial locations of an initial point selection, we can have a good rough estimation of an initial point's position. More detail discussion can be found in section 6.5.



(a) (b)

Figure 6.6: 3-D point clouds: (a) Top view. (b) Side view.

From Fig. 6.6(a) illustrates the top view of the reconstructed 3-D point clouds. We approximate the 2-D radius (circle's radius) by using four farthest points, a, b, c, and d, shown in Fig. 6.6(a) to calculate r_1 and r_2. In the case of synthesized data or a perfect sphere, a final circle's radius is computed by averaging the two radii $r = (r_1 + r_2)/2$. In the case of real retinal images, radii r_1 and r_2 correspond to the size of overlaps vertically and horizontally respectively. Once the sphere's center (A, B, C), the sphere's radius R, and the circle's radius r are estimated, we can approximate the spreading angle θ, shown in Fig. 6.6(b), of the reconstructed surface as follow

$$\theta = 2\arcsin(\frac{r}{R}). \tag{6.29}$$

After the spreading angles along both directions, θ_1, θ_2, are estimated, we can map a retinal image onto an approximated partial sphere specified by θ_1, θ_2. This 3-D retinal surface will serve as a baseline reference based on which the local depth recovery can be further estimated for the regions of interest.

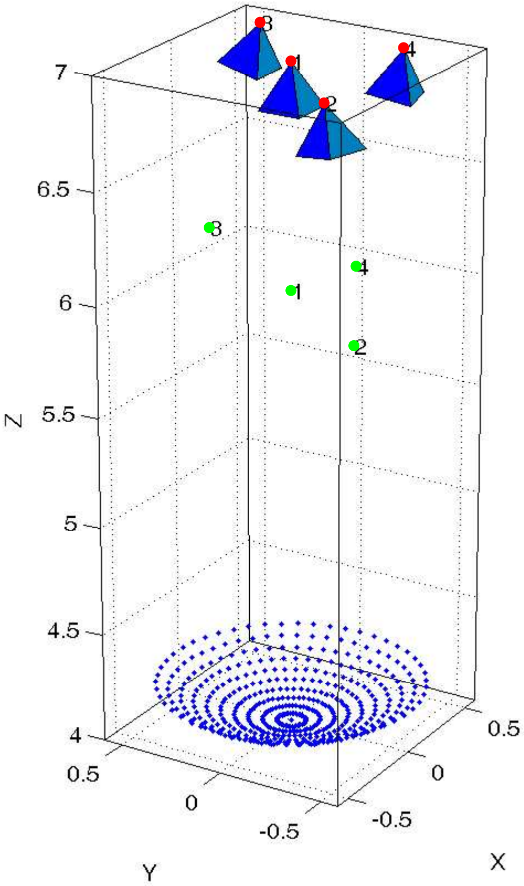


Figure 6.7: The set up of four synthetic cameras. Point cloud is constructed on a spherical surface with spreading angle of 90^o.

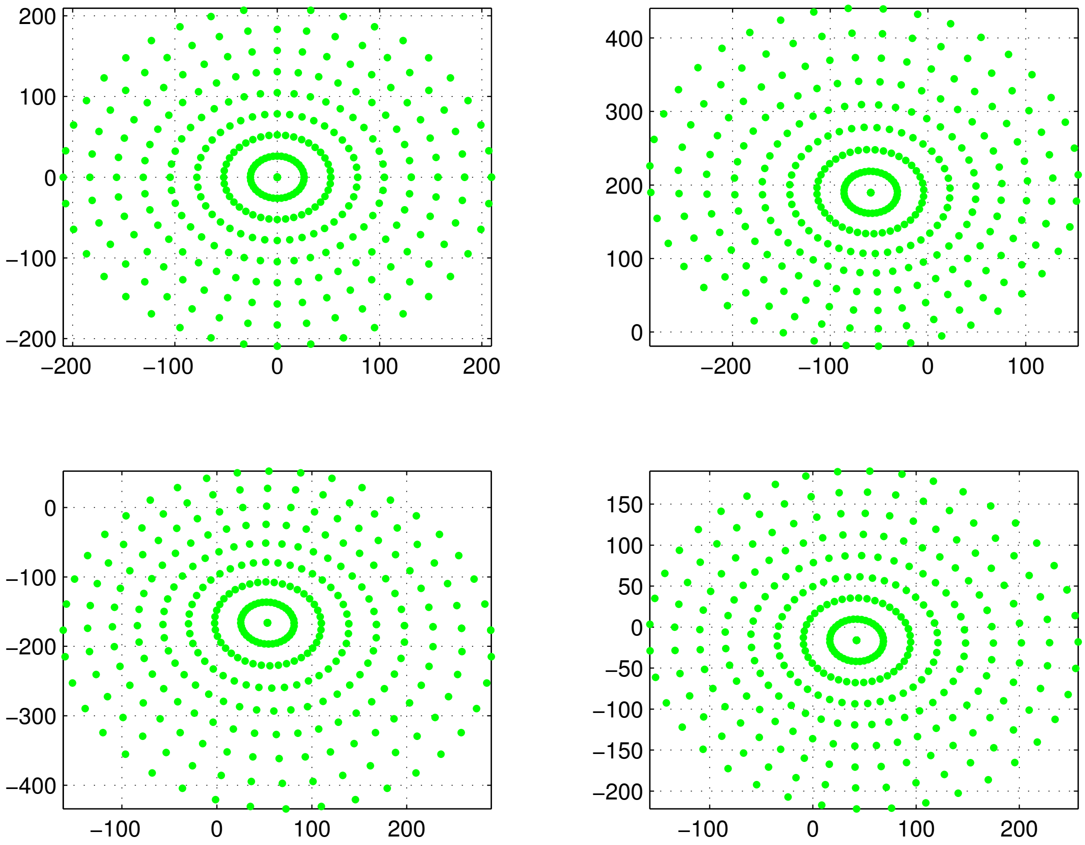


Figure 6.8: Four images generated by the four cameras shown in Fig. 6.7.

6.5 Experimental Analysis

We, first, test our algorithm on synthesized data to ensure that the proposed framework is reasonable, robust, and accurate. Most importantly, the synthetic data allow us to measure the algorithm's performance numerically because quantitative performance is impossible to obtained from real retinal images. We generate 3-D partial sphere point cloud with the spreading angle of 90^o in a world coordinate system as shown in Fig. 6.11(a). Then, four virtual cameras are positioned according to the ETDRS imaging setting, as shown in Fig. 6.7. The four synthetic images captured by these cameras are shown in Fig. 6.8. In the following experiments, we added zero-mean and isotropic Gaussian noise of different levels to correspondence measurements for algorithm evaluation. Four images are the minimum settings for Euclidean reconstruction using an affine camera. Four retinal images used are two from adjacent fields and the other two from their stereo pair.

6.5.1 Surface Approximation on Synthesized Data

Qualitative Analysis The 3D reconstruction results on synthetic data are shown in Fig. 6.11 under noise variance 0.5. The initial affine shape is shown in Fig. 6.11(b) which is more like a quadratic surface instead of a sphere. This shape distortion is probably because the affine camera is only an approximation to the ideal projective camera. After affine bundle adjustment and Euclidean reconstruction, the reconstructed surface can be obtained as shown in Fig. 6.11(c) which is much closer to the original spherical surface. However, without affine bundle adjustment, the Euclidean reconstruction result still keeps its quadratic shape without significant improvement.

Quantitative Analysis w.r.t. Noise Zero-mean, isotropic, Gaussian noises with different variances are added to image measures to test the robustness of affine bundle adjustment. Two numerical criteria are used to evaluate the effectiveness of the 3-D retinal surface reconstruction, i.e., the surface approximation error defined in and the spreading angle (i.e., the curvature). The surface approximation error is defined as follows,

$$\bar{E} = \frac{1}{n}\sum_{k=1}^{n} E_k, \qquad (6.30)$$

and

$$E_k = M_k^T \hat{\mathbf{Q}} M_k, \qquad (6.31)$$

where $\hat{\mathbf{Q}}$ is defined in Equation (6.28). $\bar{E}$ gives the average surface fitness error with respect to the optimal sphere surface $\hat{\mathbf{Q}}$, and E_k is the fitness error of point M_k. Both of them are used for performance evaluation in the following. Additionally, we can compute the spreading angle according to Equation 6.29 as the second criterion for performance evaluation. Fig. 6.9 and Fig. 6.10 show the errors of surface approximation versus noise variances in terms of two criteria. At each noise level, the algorithm is performed ten times to obtain an average error. It is shown that bundle adjustment does improve the reconstruction performance significantly and sustain good performance under strong noises.

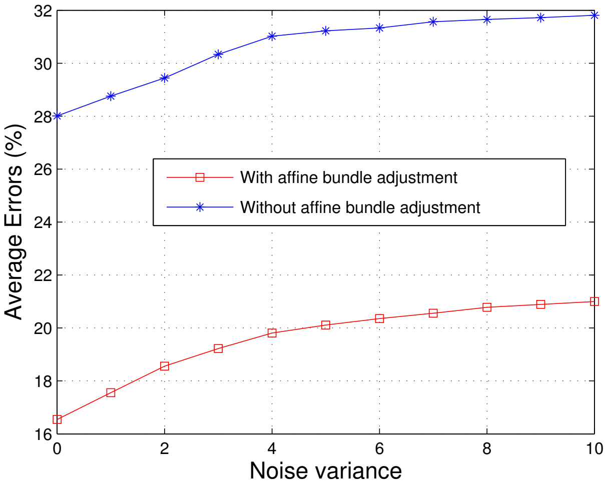


Figure 6.9: The errors of spreading angles (%) versus noise variances. The overall spreading angle of an original synthetic partial sphere is 90^o.

Quantitative Analysis w.r.t. Reference Point Selection We plot the errors of reconstructed surface versus 2-D spatial locations of the reference point in order to understand the relationship between the location of reference point and the surface approximation error. Regardless of the noise level, if affine bundle adjustment is not performed, the error plot always possesses a similar shape as shown in Fig. 6.12(a). We observe that points around an average depth yield the minimum error. If affine bundle adjustment is involved, the error plot always retains a similar shape as shown in Fig. 6.12(b). Points around the bottom produce the minimum error. These observations implicitly convey useful information for selecting a good reference point. Given a quadratic surface (Fig. 6.11(b)), the point-based surface approximation estimates a spherical surface along a quadratic's average depth. If the reconstructed shape is closer to a sphere (Fig. 6.11(c)), then the algorithm would produce a surface

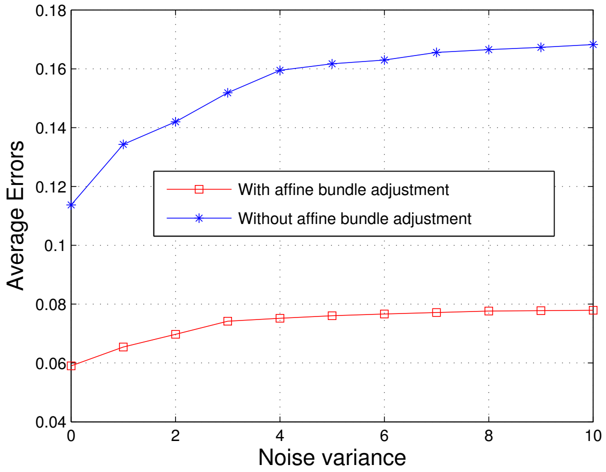


Figure 6.10: The errors between reconstructed points and surface approximation versus noise variances.

that matches the bottom of the point cloud.

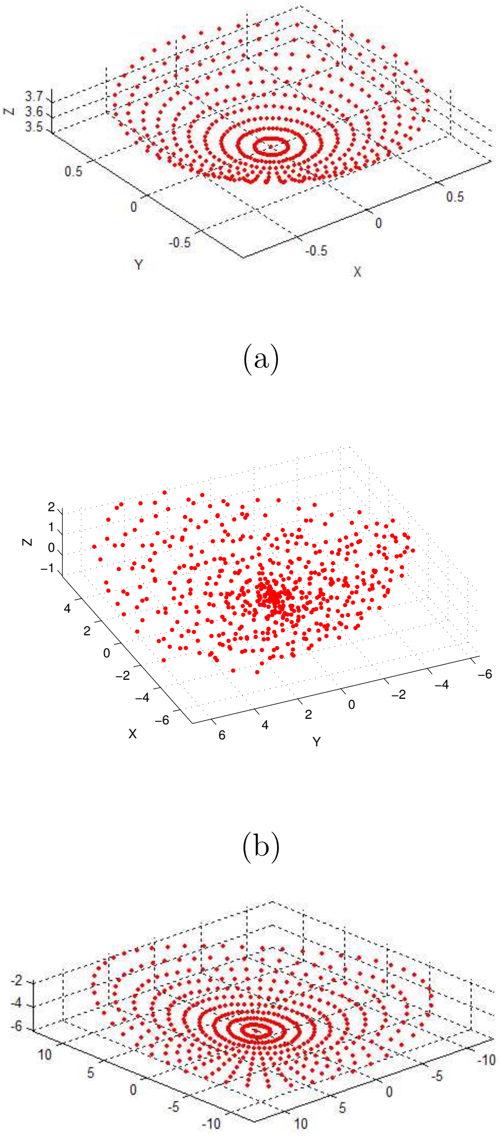
6.5.2 Surface Approximation on Retinal Images

Lens Distortion Removal A grid pattern with different orientations is used for the purpose of removing lens distortions caused by two main elements, a nonlinear fundus camera and an individual eye lens. The fundus camera is not a linear camera and an individual eye lens can be considered as an additional camera which causes additional distortions in retinal images. Examples of a grid pattern and a retinal image with lens distortion removal are illustrated in Fig. 6.13 and 6.14 respectively.

Qualitative and Quantitative Evaluation Two sets of retinal images, illustrated in Fig. 6.15, are used in the experiment. Each column of Fig. 6.15 depicts a set of retinal images which includes two stereo pairs of field 1 and 2. Point correspondences, also shown in Fig. 6.15, have been automatically extracted by our previously proposed algorithm [58]. The experimental results of 3-D retinal surface reconstruction are shown in Fig. 6.16(a),(b). The surface approximation error between a reconstructed 3-D point and the approximated surface is calculated according to Equation (6.27). Fig. 6.17(a) shows the improvement due to lens distortion removal and Fig. 6.17(b) illustrates the further improvement from affine bundle adjustment. The best performance is achieved by removing the lens distortion and by using affine bundle adjustment.

6.6 Conclusions

We have showed 3D retinal surface reconstruction using an affine camera model for ETDRS retinal images. The robustness and effectiveness of the proposed algorithm are rooted in the linear nature of the affine camera and the prior knowledge about the shape of human retinal. Also, the reconstruction performance is significantly improved by lens distortion removal and affine bundle adjustment. In the next chapter, we will incorporate the geometrical constraint and the lens distortion update into affine bundle adjustment to further improve the reconstruction accuracy.

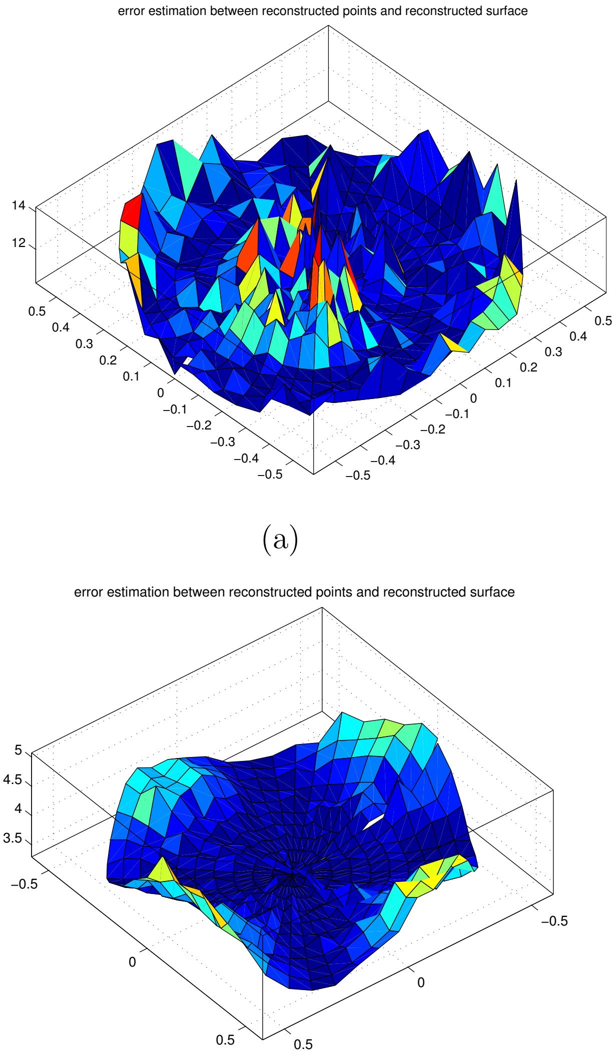


(a)

(b)

(c)

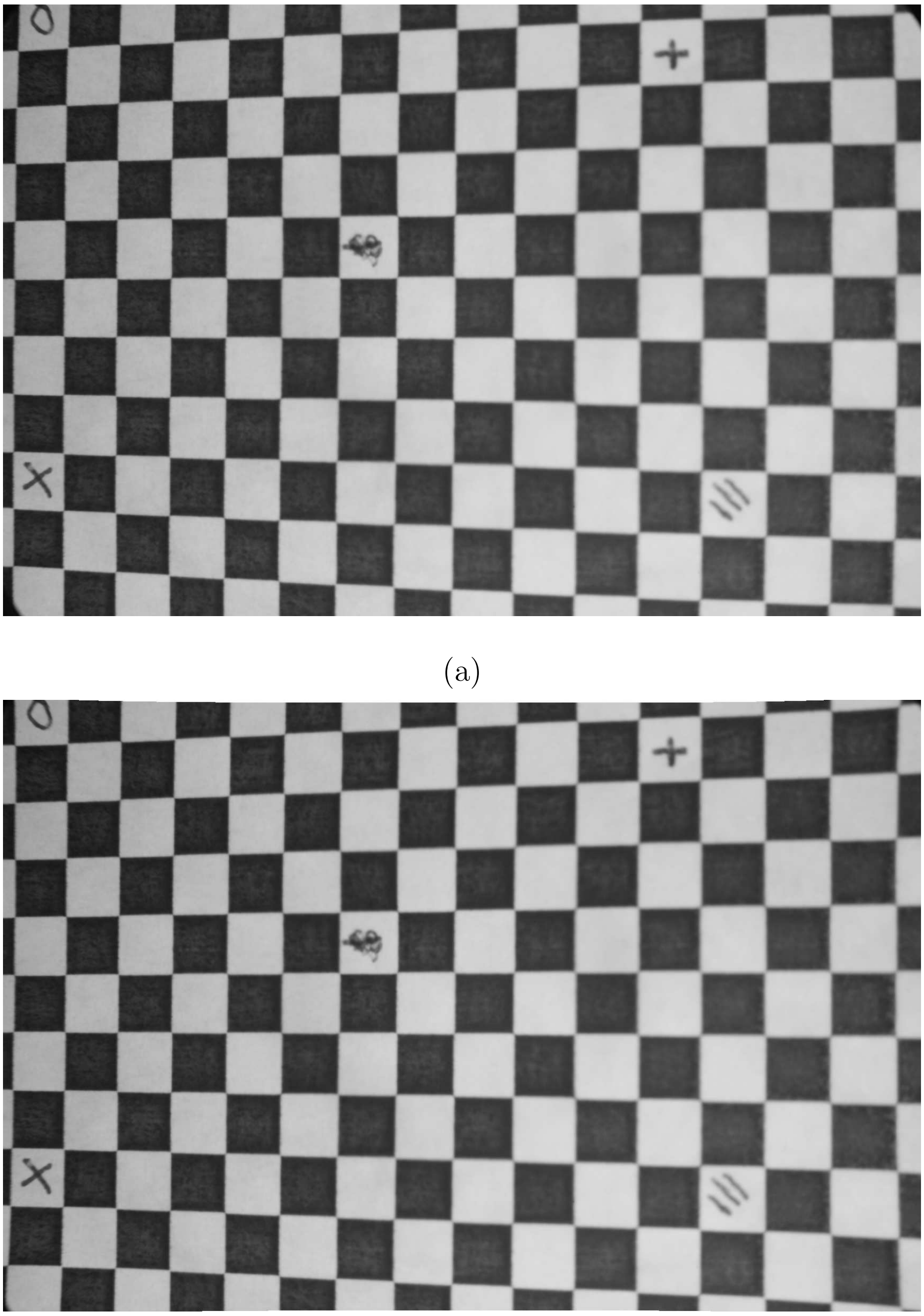
Figure 6.11: 3-D surface reconstruction on synthetic data. (a) A partial synthetic spherical shape. (b) Affine reconstruction of a partial synthetic spherical shape. (c) Euclidean reconstruction of a partial synthetic spherical shape.



(a)

(b)

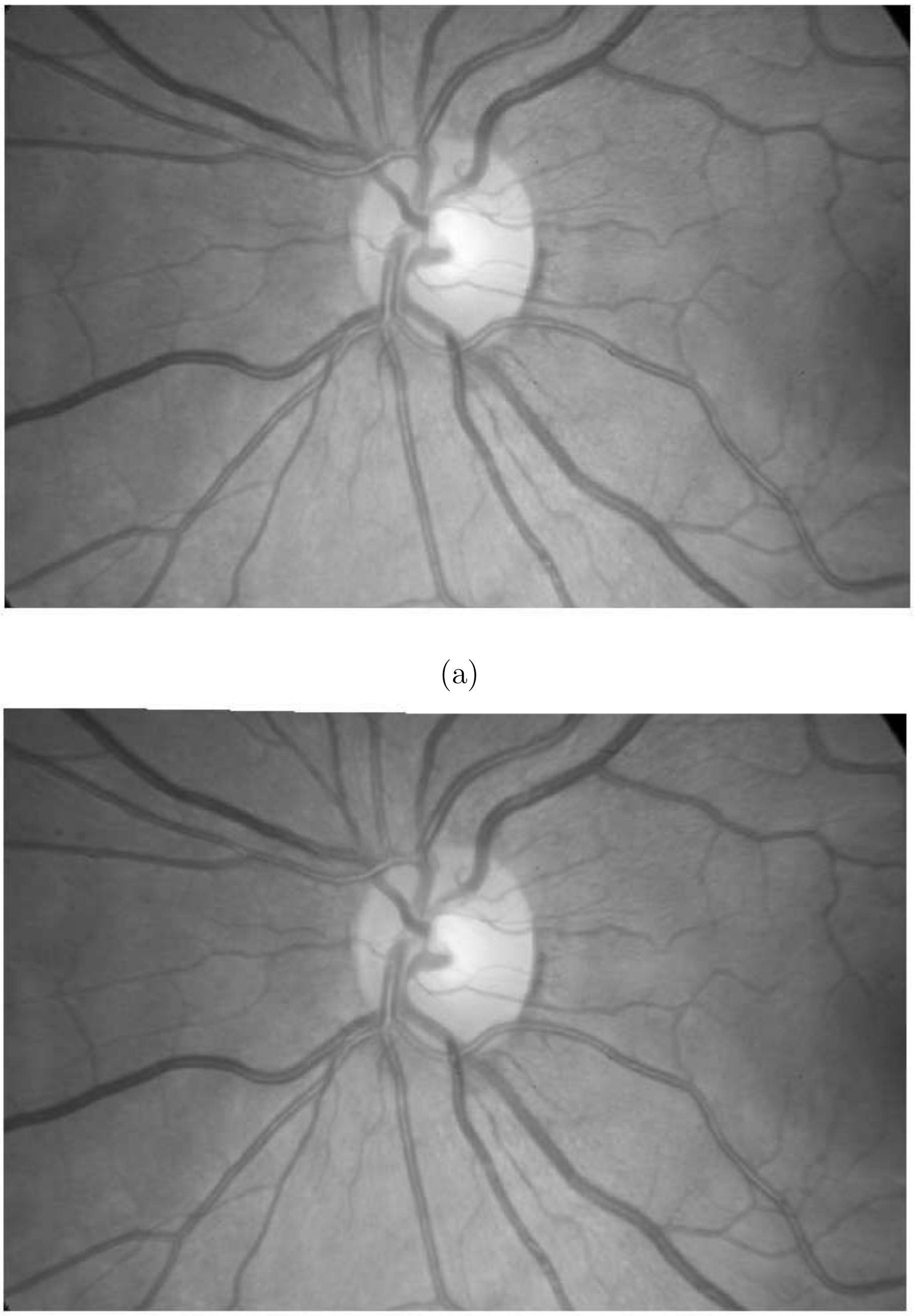
Figure 6.12: The errors versus 2-D spatial locations of the initial point selection. (a) Without affine bundle adjustment. (b) Affine bundle adjustment is involved.



(a)

(b)

Figure 6.13: (a) An example of a grid pattern. (b) A lens distortion-free grid pattern.



(a)

(b)

Figure 6.14: (a) An example of a retinal image. (b) A lens distortion-free image.

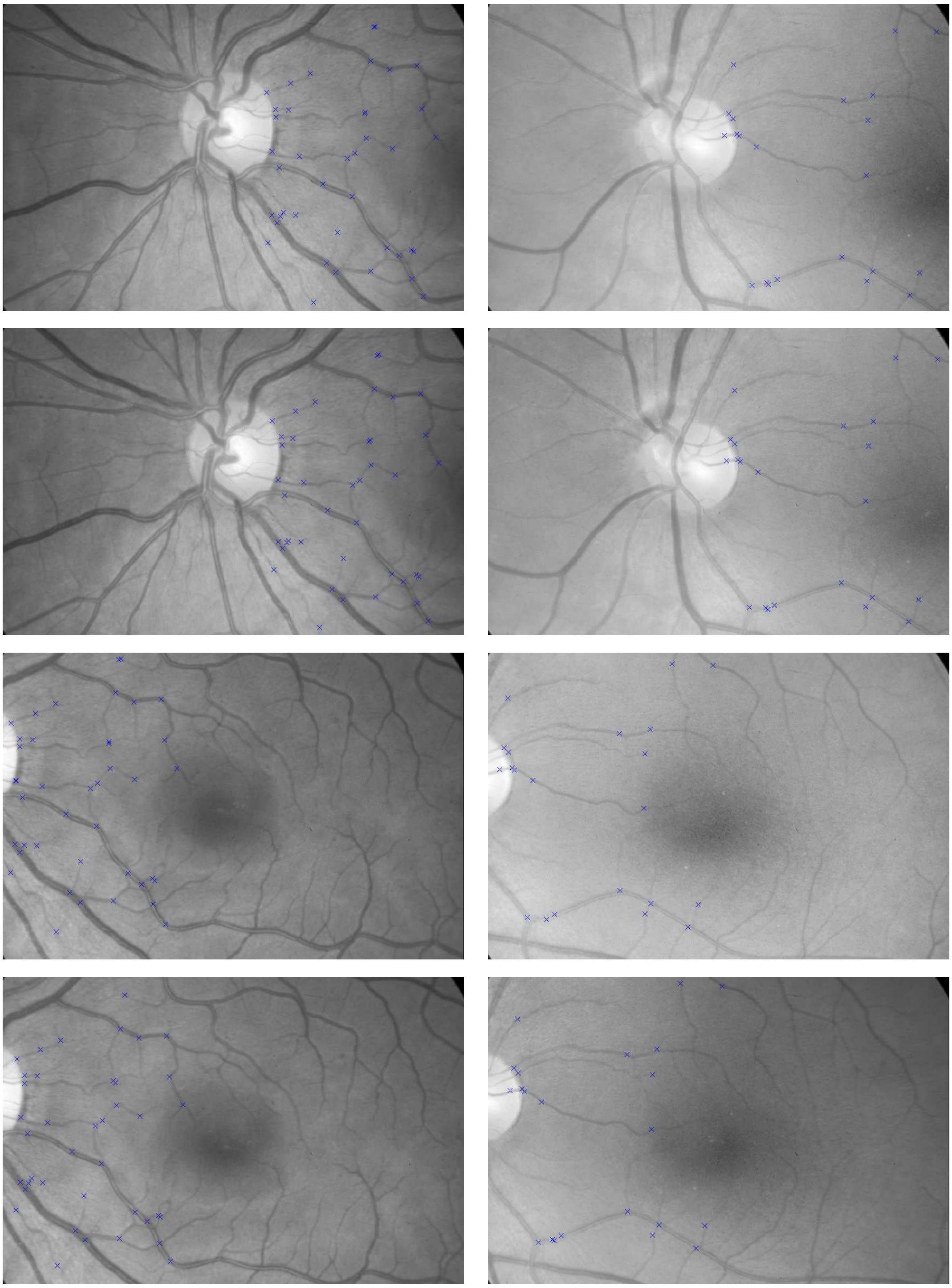
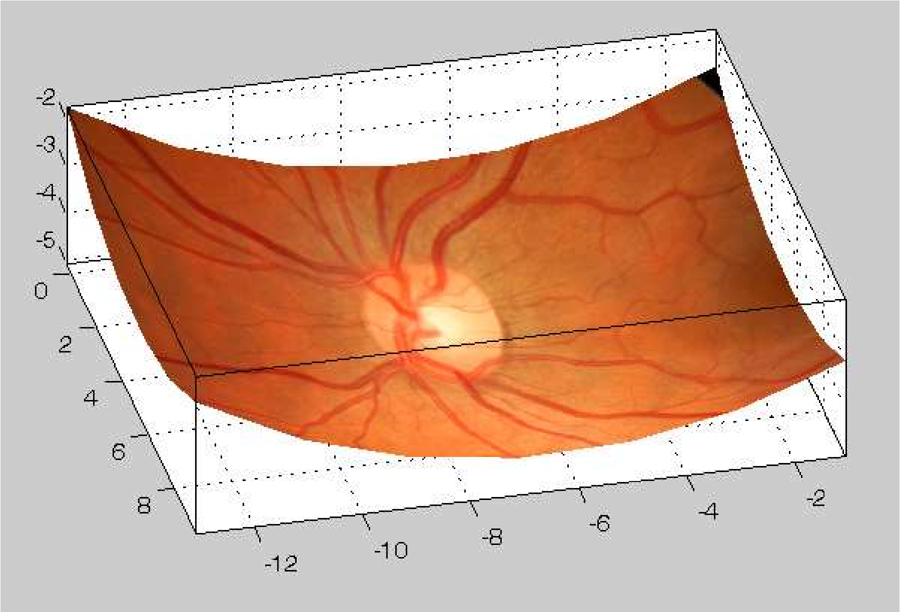
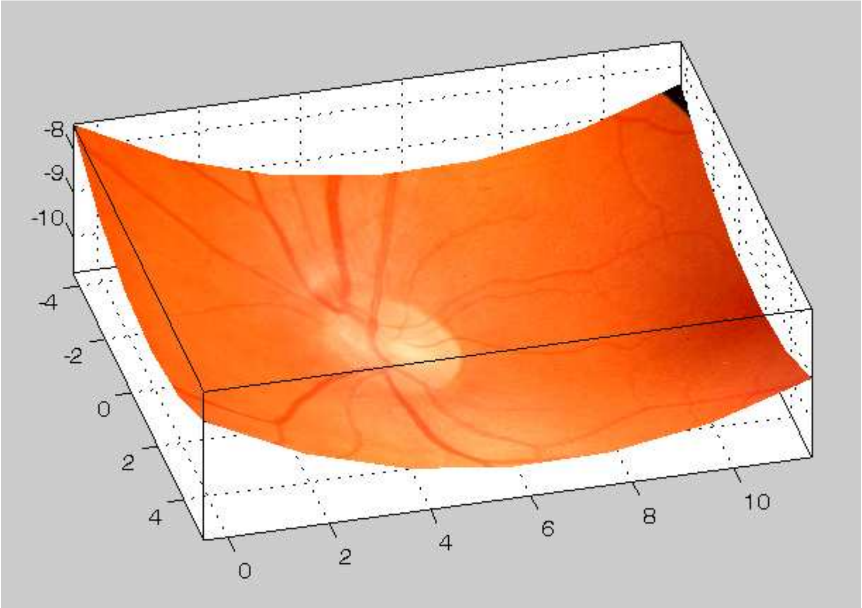


Figure 6.15: Two sets of retinal images with marked point correspondences. First and second rows show stereo pairs of field 1. Third and fourth rows show stereo pairs of field 2.



(a)



(b)

Figure 6.16: The 3-D retinal reconstruction results with the retinal image mapped onto sphere surfaces.

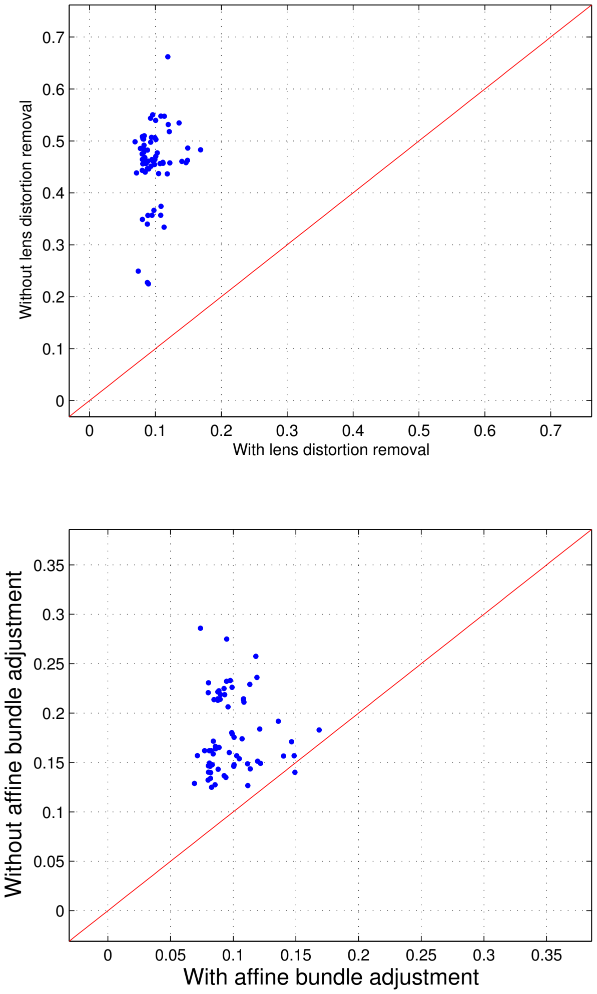


Figure 6.17: The errors between reconstructed 3-D points and the approximated spherical surface without/with radial distortion removal (a) and without/with affine bundle adjustment (with the radial distortion removed) (b).

CHAPTER 7

Constrained Optimization for 3-D Surface Reconstruction

7.1 Introduction

We have developed and proposed a framework for 3-D retinal surface reconstruction in the previous chapter [102]. In this chapter, we are dealing with the issue of constrained optimization for a SFM problem. As mentioned in chapter 6 that SFM is a process to recover objects' 3-D shapes, cameras' poses, and cameras' internal parameters. Our objective for this chapter is to solve the problem in an optimal way which means to estimate structure and camera parameters simultaneously by minimizing a physically meaningful cost function. Because geometric structures, e.g. geometric shapes of known objects, display certain regularities, adding surface models into a cost function would represent a geometrically meaningful definition.

All of the works dedicate to the subject of 3-D constrained optimization assume either lens distortion is removed before the reconstruction process or lens distortion is insignificant. In our case of retinal images, however, lens distortion is too prominent to be disregarded. As shown in Fig. 7.1, there is a series of optics involved in the retinal imaging process, which includes an actual fundus camera, an digital camera, and the human cornea. In chapter 6, we remove lens distortion prior to the reconstruction procedure. Nevertheless, we believe only lens distortion caused by a fundus camera has been removed while lens distortion caused by the human cornea still presents. Therefore, we propose a constrained optimization which includes lens distortion parameters in the process.

Another issue is that most of the works proposed in the subject of geometric

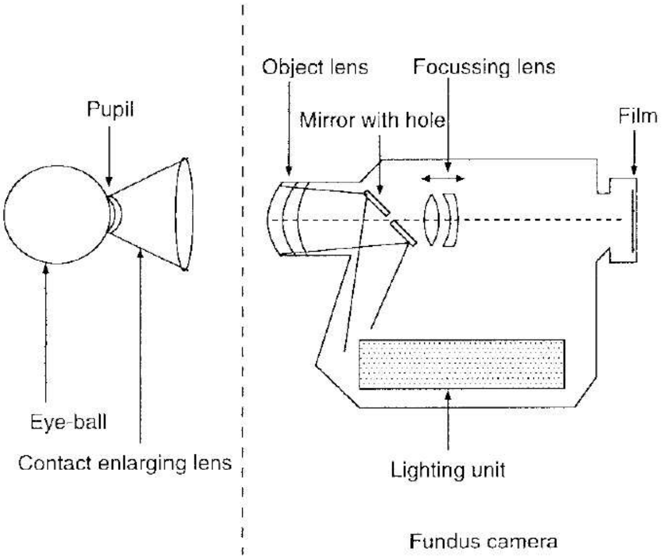


Figure 7.1: Retinal images are obtained from a fundus camera which composes of an actual camera and a digital camera attached to a fundus camera [3].

constrained optimization are done in a final step or in the Euclidean space. Only exceptions are the ones that assume planarity constraints [114, 115]. In this work, we deal with retinal images which are captured from the back of eyeballs. Because an eyeball can be exhibited by an approximated sphere, the spherical constraint could incorporated into an optimization process to improve the reconstruction results. Typically, geometric constraints are involved in the SFM process in a final processing stage. Szeliski et.al. [115] have suggested that prior geometric knowledge, which is incorporated early on in the reconstruction process, can improve the quality of the estimates. Since an optimization process in an affine space requires less computational time compared to that in a Euclidean space, we want to implement the constrained optimization in an affine space.

Three variations of bundle adjustment have been developed and tested on sets of retinal images and synthesized data. Specifically, the three implementations are affine bundle adjustment (ABA), constrained affine bundle adjustment (CABA), and constrained affine bundle adjustment with lens distortion updates (CABA-LDU). Affine bundle adjustment is discussed in chapter 6. This chapter will be dedicated to the CABA and CABA-LDU.

7.2 Related Works

Related works regarding retinal curvature estimation and affine SfM are reviewed in chapter 6. In this chapter, we will focus on the constrained optimization issue. If there is some prior knowledge about the 3-D geometry, adding the surface models or geometrical constraints into the SfM process would provide a more geometrically meaningful solution. Shan et.al. [116] proposed a model-based bundle adjustment algorithm with face modeling application. Their algorithm used a surface controlled by a small set of parameters by eliminating all the 3-D point position variables in a cost function. Fua [117] addressed the SFM problem in the context of head modeling. Based on the prior knowledge of the head's shape, they augmented the standard bundle adjustment with iterative reweighted least square and regularization. Gong et.al. [118] used sequential quadratic programming (SQP) [119] to recover 3-D quadratic surface parameters by using a quadratic surface as a constraint. Simulation results revealed that a constrained method produce more accurate results than that by traditional approaches. Wong et.al. [120] showed that 3-D reconstruction accuracy can be significantly improved by employing constraints inherent in the object's motion, namely the object is constrained to rotate around a fixed axis. Bartoli et.al. [114] merged the multi-coplanarity constraints with the traditional bundle adjustment approach. Simulation results showed that the accuracy of a constrained method is superior compared to that of traditional ones.

7.3 Constrained Optimization for 3-D Reconstruction

If point correspondences are error free and lens distortions can be completely removed, motion parameters and 3-D structure could be accurately obtained through standard SFM procedure. However, point correspondences are too sensitive to noise and lens distortion still presents. Moreover, the standard optimization procedure, i.e. bundle adjustment, does not carry any 3-D geometrically meaningful description. Hence, the results from the standard SFM would not be as accurate as one expects. In specific applications, prior knowledge can be involved to compensate the errors from point correspondences and lens distortions as well as provide a semantically meaningful cost function. In our case, we have prior knowledge regarding the shape of a retinal surface. We propose a constrained optimization in an affine space rather than the traditional Euclidean space since it is more computationally efficient. To improve performance in terms of robustness and accuracy, a constrained surface model, which carries geometrically meaningful definition, is incorporated into affine bundle adjustment. To increase accuracy and further remove the remaining lens distortion, we optimize motion parameters, 3-D structure, and lens distortions simultaneously. Therefore, we amend the affine bundle adjustment algorithm in two manners.

7.3.1 Constrained Affine Bundle Adjustment (CABA)

A quadratic surface constraint is included into a cost function to improve robustness and accuracy. The purpose of this supplement is to prevent severe deformations. A quadratic surface can be defined by the equation

$$\sum_{j=1}^{n} M_j^T Q M_j = 0, \tag{7.1}$$

where Q is a symmetric 4×4 matrix. M is a homogeneous 4-vector which represents a 3-D point on retinal surface. Incorporate the above equation into a standard affine bundle adjustment algorithm.

$$\min_{\breve{P}_i, \breve{M}_j, Q_e, \rho} \sum_{i=1}^{v} \sum_{j=1}^{n} \|\breve{m}(\breve{P}_i, \breve{M}_j) - m_{ij}\|^2 + \rho \sum_{j=1}^{n} \breve{M}_j^T Q_e \breve{M}_j, \tag{7.2}$$

where we have v views/images and n correspondences. Q_e is a symmetric 4×4 matrix representing an ellipsoid surface. $\breve{m}(\breve{P}_i, \breve{M}_j)$ is a projection of a point $\breve{M}_j$ from an image i. m_{ij} represents a retinal image point and ρ is a Lagrange multiplier.

Q_e is initialized as a spherical surface by using the point-based linear method introduced in Section 6.4.5. During the iterations, parameters in matrix Q_e are updated to represent an ellipsoid surface. The Lagrange multiplier ρ acts like a weighting parameter in the cost function. It can be chosen based upon the dynamic ranges of the two terms in Equation (7.2) where the first part is the residual error in all 2-D images and the second term the surface approximation error in the 3-D affine space. In our case, the two error terms have the similar dynamic range, and we initialize ρ to be 1 that is updated during the iterations.

7.3.2 Affine Bundle Adjustment with Lens Distortion Update

Lens distortions are incorporated into a constrained affine bundle adjustment to increase accuracy. Even though, lens distortions are removed in the first calibration as discussed in chapter 6, virtual lens effects from human cornea are still present. We propose to further remove those remaining lens distortions through an optimization procedure. There are several models describing different types of lens distortions. In this work, we only consider two lens-distortion models, radial distortion and tangential distortion since other distortions are either insignificant or can be included in radial and tangential distortions [105]. Radial distortion causes the actual image points to be displaced radially in the image plane [104]. Centers of curvature of lens surfaces are not always strictly collinear which cause tangential distortion [104] Radial distortion can be expressed as

$$\begin{aligned} \delta x_r &= \tilde{x}(k_{c1}r^2 + k_{c2}r^4 + k_{c3}r^3) \\ \delta y_r &= \tilde{y}(k_{c1}r^2 + k_{c2}r^4 + k_{c3}r^3), \end{aligned} \tag{7.3}$$

where $(\tilde{x}, \tilde{y})$ are image coordinates in metric unit, $r = \sqrt{\tilde{x}^2 + \tilde{y}^2}$, and k_{c1}, k_{c2}, k_{c3} are coefficients for radial distortion. The expression for tangential distortion is

$$\begin{aligned} \delta x_t &= 2k_{c4}\tilde{x}\tilde{y} + k_{c5}(r^2 + 2\tilde{x}^2) \\ \delta y_t &= k_{c4}(r^2 + 2\tilde{y}^2) + 2k_{c5}\tilde{x}\tilde{y}, \end{aligned} \tag{7.4}$$

where k_{c4}, and k_{c5} are coefficients for tangential distortion. Image pixel (x, y) and metric $(\tilde{x}, \tilde{y})$ coordinates are related by

$$\begin{aligned} x &= c_x + f_x\tilde{x} + s\tilde{y} \\ y &= c_y + f_y\tilde{y}, \end{aligned} \tag{7.5}$$

where (c_x, c_y) denotes the principal point, f_x and f_y represent the focal length, and s is a skew factor. With lens distortions included into a standard affine bundle adjustment.

$$\min_{\breve{P}_i, \breve{M}_j, k_c} \sum_{i=1}^{v} \sum_{j=1}^{n} \|\breve{m}(\breve{P}_i, \breve{M}_j, \delta_r, \delta_t) - m_{ij}\|^2, \tag{7.6}$$

where we have v views/images and n correspondences. $\breve{m}(\breve{P}_i, \breve{M}_j, \delta_r, \delta_t)$ is a projection of a point $\breve{M}_j$ in the ith image following by the radial $\delta_r \triangleq [\delta_{xr}, \delta_{yr}]$ and tangential distortions $\delta_t \triangleq [\delta_{xt}, \delta_{yt}]$ defined in Equations (7.3) and (7.4) respectively. m_{ij} represents a retinal image point.

7.3.3 Constrained Affine Bundle Adjustment with Lens Distortion Update (CABA-LDU)

With both 3-D geometric constraint and lens distortions are integrated into the equation, the cost function becomes

$$\begin{aligned} \min\nolimits_{\breve{P}_i, \breve{M}_j, Q_e, \rho, k_c} \quad & (\textstyle\sum_{i=1}^{v} \sum_{j=1}^{n} \|\breve{m}(\breve{P}_i, \breve{M}_j, \delta_r, \delta_t) - m_{ij}\|^2 \\ & \quad + \rho \textstyle\sum_{j=1}^{n} \breve{M}_j^T Q_e \breve{M}_j), \end{aligned} \tag{7.7}$$

where we have v views/images and n correspondences. $\breve{m}(\breve{P}_i, \breve{M}_j, \delta_r, \delta_t)$ is a projection of a point $\breve{M}_j$ in the ith image following by the radial $\delta_r \triangleq [\delta_{xr}, \delta_{yr}]$ and tangential distortions $\delta_t \triangleq [\delta_{xt}, \delta_{yt}]$ defined in Equations (7.3) and (7.4) respectively. m_{ij} represents a retinal image point. Q_e is a symmetric 4×4 matrix representing an ellipsoid surface. ρ is a Lagrange multiplier. Equation 7.7 shows that we associate two types of errors, both 2-D error (the first term) and 3-D error (the second term), into an optimization process. The cost function also incorporates both geometrically meaningful definition and lens distortion. The procedure optimizes all of the parameters, camera's parameters, 3-D points, physical shape of a retinal surface, and lens distortion, simultaneously.

7.4 Experimental Analysis

We tested three variations of bundle adjustment, ABA, CABA, and CABA-LDU, on both synthesized data and retinal images. As mentioned in chapter 6 that the synthetic data allow us to measure the algorithm's performance numerically because quantitative performance is impossible to obtained from real retinal images. We generate 3-D partial sphere point clouds with the spreading angle of 90^o in a world coordinate system. Then, four virtual cameras are positioned according to the ETDRS imaging setting. In the following experiments, we add various lens distortion coefficients and different noise variances to correspondence measurements as well as vary the number of synthesized points in order to compare different approaches under various circumstances. Four images are the minimum settings for Euclidean reconstruction using an affine camera. Four retinal images used are two from adjacent fields, fields 1 and 2, and the other two from their stereo pair. Two numerical criteria, discussed in chapter 6 section 6.5, are used to evaluate the effectiveness of the 3-D retinal surface reconstruction, i.e., the surface approximation error defined in and the spreading angle (i.e., the curvature).

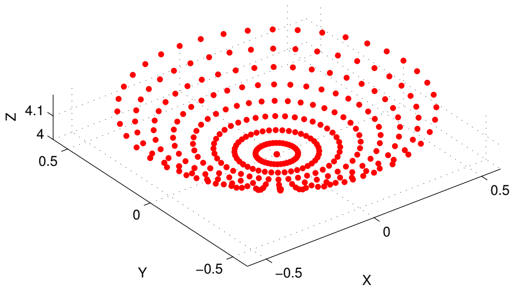


Figure 7.2: An original partial sphere with spreading angle = 90^o

7.4.1 Surface Approximation on Synthesized Data

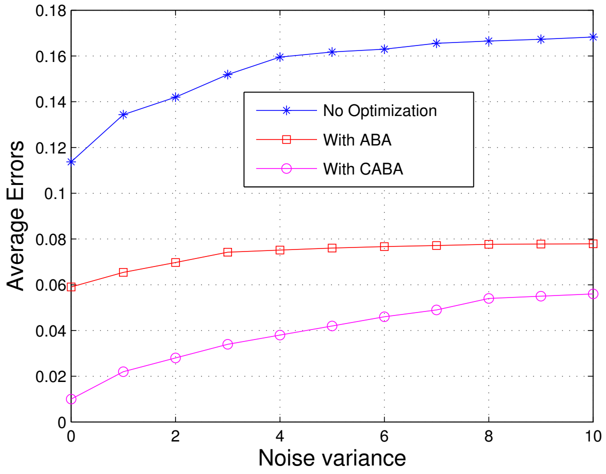


Figure 7.3: Lens distortion-free synthesized data with noise: the errors between reconstructed points and surface approximation versus noise variances.

Quantitative Analysis We, first, evaluate algorithms' performance on synthesized data based upon three conditions, synthesized data with noise shown in Fig. 7.8, synthesized data with lens distortion shown in Fig. 7.9, and synthesized data with both noise and lens distortion. Two numerical criteria defined in Section 6.5.1 are used to evaluate the effectiveness of the 3-D surface reconstruction, i.e., the surface

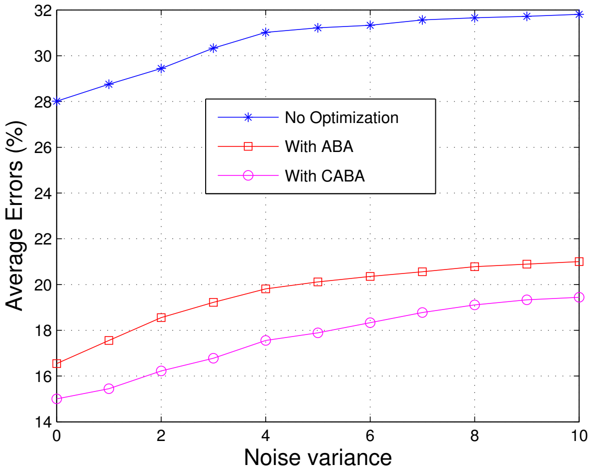


Figure 7.4: Lens distortion-free synthesized data with noise: the errors of spreading angles in percentage versus noise variances. The overall spreading angle of an original synthetic partial sphere is 90^o.

approximation error and the spreading angle.

In the case of synthesized data with noise, we tested on two optimizations, ABA and CABA. Zero-mean, isotropic, Gaussian noises with different variances are added to images. Fig. 7.3 shows the errors of surface approximation versus noise variances. Fig. 7.4 shows errors of spreading angle in percentage versus noise variances. At each noise level, the algorithm is performed ten times to obtain an average error. It is shown that constrained affine bundle adjustment improves the reconstruction performance and sustains good performance under strong noises.

In the second case of synthesized data with lens distortion, we tested on three optimizations, ABA, CABA, and CABA-LDU. Fig. 7.10 illustrates experiments on noise-free synthesized data with lens distortion coefficients set to $k_c = [0.03, -0.08, 0.02, 0.01, -0.02]$. Fig. 7.10(a) compares errors between the procedure without optimization versus the procedure with affine bundle adjustment. Fig. 7.10(b) relates errors between ABA and CABA. Fig. 7.10(c) associates errors between CABA and CABA-LDU. Fig. 7.10 demonstrates that an algorithm's performance can be im-

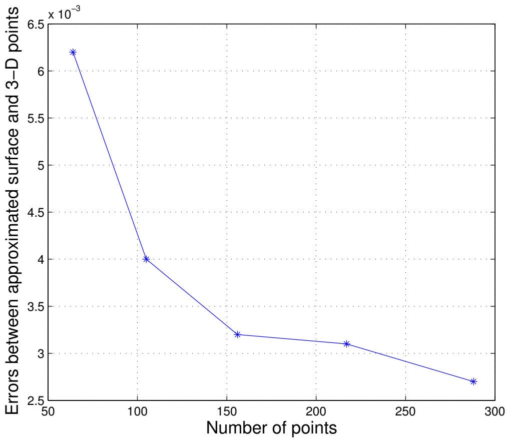


Figure 7.5: Noise-free synthesized data with lens distortion coefficients set to $k_c = [0.03, \ -0.08, \ 0.02, \ 0.01, \ -0.02]$: Tested on CABA-LDU optimization procedure. A plot between average errors between approximated surface and 3-D points versus the number of synthesized points.

proved step by step through appropriate optimization procedures. CABA-LDU yields the best performance in terms of accuracy. Regarding the lens distortion coefficients, the procedure can further remove the remaining lens distortion. However, the lens distortion coefficients are not accurate. This probably due to the fact that we run the optimization in an affine space.

To evaluate the performance of CABA-LDU in various conditions, we vary the number of synthesized points. Fig. 7.5 shows that number of point correspondences can improve the algorithm's performance in terms of accuracy. The higher the number of point correspondences, the better the accuracy.

In the last case of synthesized data with both noise and lens distortion, we only tested on a CABA-LDU procedure. Zero-mean, isotropic, Gaussian noises with different variances and lens distortion coefficients set to $k_c = [0.03, \ -0.08, \ 0.02, \ 0.01, \ -0.02]$ are added to images. Fig. 7.3 shows the errors of surface approximation versus noise variances. The plot demonstrates that the higher the noise variance, the worse

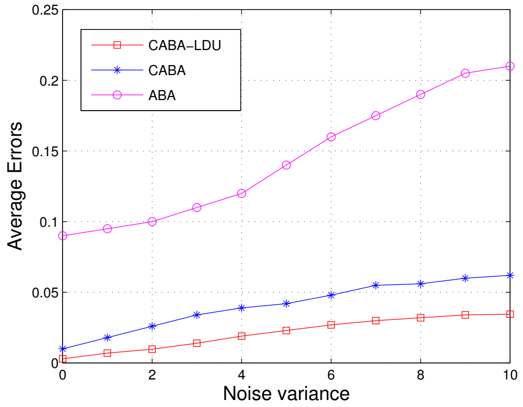


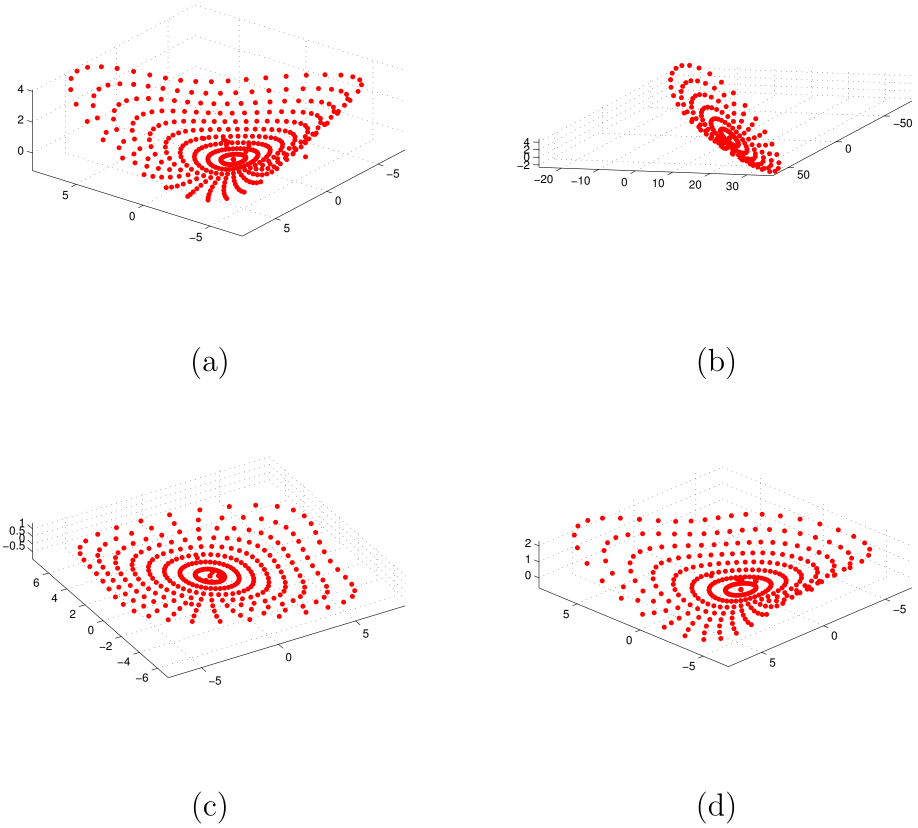
Figure 7.6: Synthesized data with both noise and lens distortion: the errors between reconstructed points and surface approximation versus noise variances.

the algorithm's performance in terms of accuracy.

Qualitative Analysis The original synthesize partial sphere is shown in Fig. 7.2 with spreading angle 90^o. The 3D reconstruction results on synthetic data are illustrated in Fig. 7.7 under the setting of lens distortion coefficients $k_c = [0.03, \ -0.08, \ 0.02, \ 0.01, \ -0.02]$. Fig. 7.7(a) shows the simulation result with no optimization. The shape possesses a quadratic-shape resemblance rather than a spherical surface. Fig. 7.7(b) shows the simulation result with ABA. The surface is less similar to a quadratic surface and gets closer to an ellipsoid shape. Fig. 7.7(c) and Fig. 7.7(d) demonstrate the simulation result with CABA and CABA-LDU. Compare to an original partial spherical surface shown in Fig. 7.2, a reconstructed surface in Fig. 7.7(d) holds the closest similar shape.

7.4.2 Surface Approximation on Retinal Images

With the same experiments on synthesized data, we tested on three optimizations, ABA, CABA, and CABA-LDU. Fig. 7.11(a) compares errors between the procedure without optimization versus the procedure with ABA. Fig. 7.11(b) relates errors between ABA and CABA. Fig. 7.11(c) associates errors between CABA and CABA-



(a)

(b)

(c)

(d)

Figure 7.7: Noise-free synthesized data with lens distortion coefficients $k_c = [0.03, \; -0.08, \; 0.02, \; 0.01, \; -0.02]$: (a) No optimization. (b) With ABA. (c) With CABA. (d) With CABA-LDU.

LDU. Fig. 7.11 demonstrates that the algorithm's performance can be improved step by step through appropriate optimization procedures. CABA-LDU yields the best performance in terms of accuracy. The experimental results of 3-D retinal surface reconstruction are shown in Fig. 7.12(a) and Fig. 7.12(b).

For two image sets, the estimated spreading angles in the overlapping part (the 3-D reconstruction area) are about $24^\circ \sim 26^\circ$ in both directions. This result is consistent with our assumption that the depth variation is relatively low in the overlapping part. The experimental result of one-field retinal curvature estimation is visualized in Fig. 7.12 where the horizontal and vertical curvatures are derived from the ratio between the size of the overlapping part (900×700) and the size of the original images (1728×1152). This 3-D retinal model could serve as a reference surface based on which the local depth recovery can be further estimated for the regions of interest.

7.4.3 More Discussions

There are two limitations in this work that need further investigation. First, since the optimization is in the affine space, the lens distortion parameters cannot be accurately estimated. Second, the spherical constraint is enforced indirectly as an ellipsoid one in the optimization process, and the surface approximation error added in CABA and CABA-LDU may not directly reflect the shortest distance between the 3-D points and the reconstructed surface in the Euclidian space. Nevertheless, this research is able to provide accurate and robust estimation of retinal curvature that can be further combined with other techniques for more detailed and accurate 3-D retinal reconstruction and visualization.

7.5 Conclusions

This chapter presents constrained optimization algorithms for more accurate and robust 3-D retinal surface reconstruction where we have considered both the known 3-D geometry of human retina and the virtual lens distortion introduced by human cornea. Specifically, we have defined a new optimization function for affine bundle adjustment that incorporates both the geometrically meaningful surface approximation error and the lens distortion removal. The proposed algorithm can effectively yield more accurate 3-D structures compared with previous affine bundle adjustment in Chapter 6.

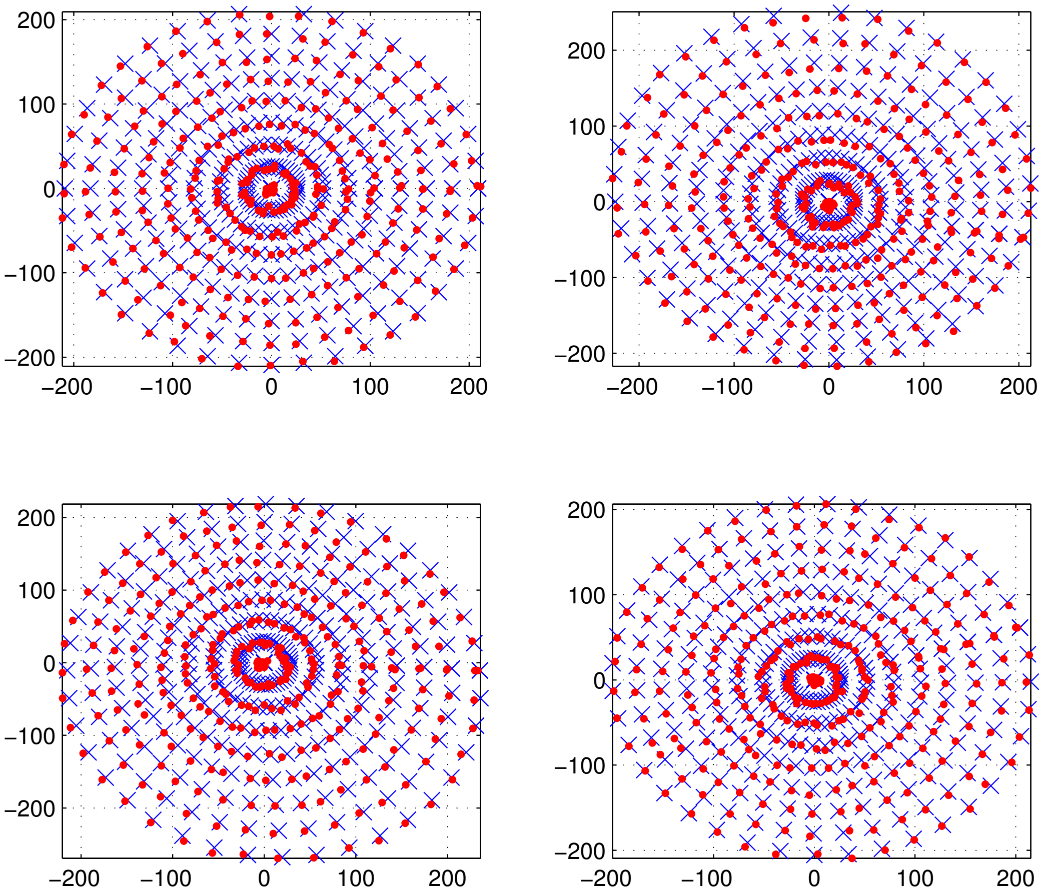


Figure 7.8: Four images generated by the four synthesized cameras: blue crosses (×) are original images and red dots (•) are images with noise (variance = 3).

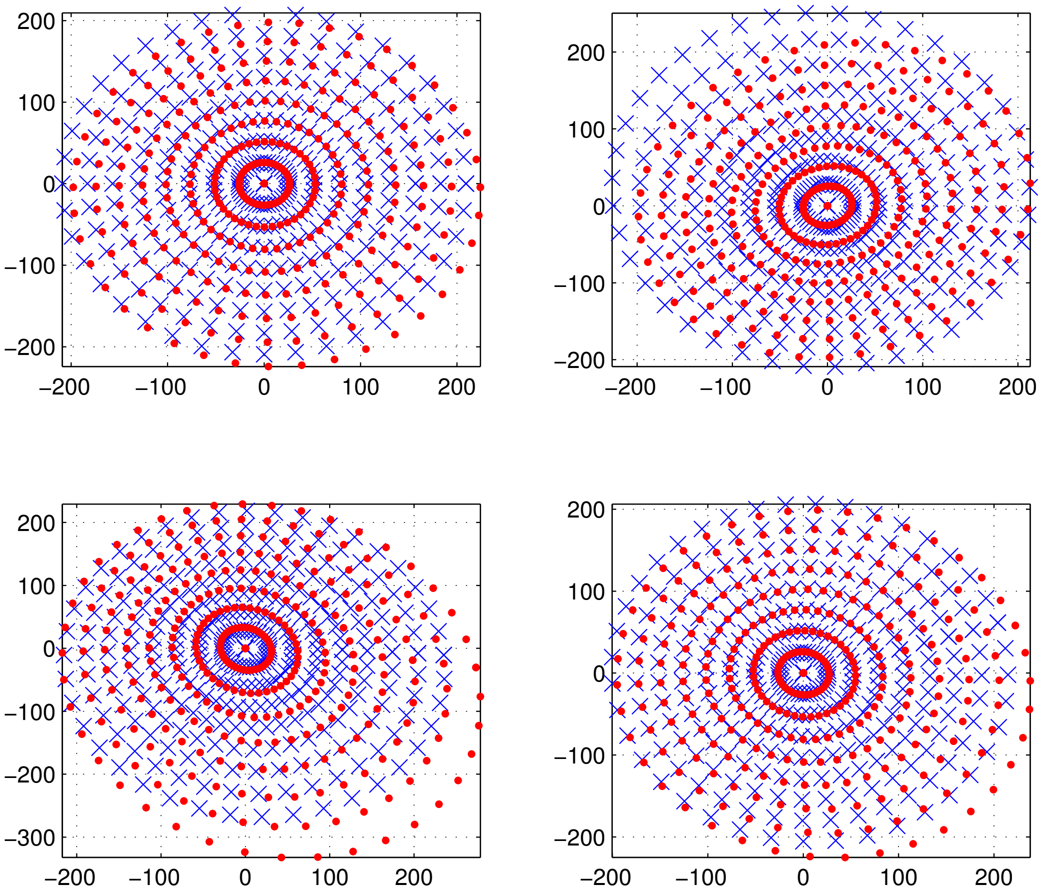
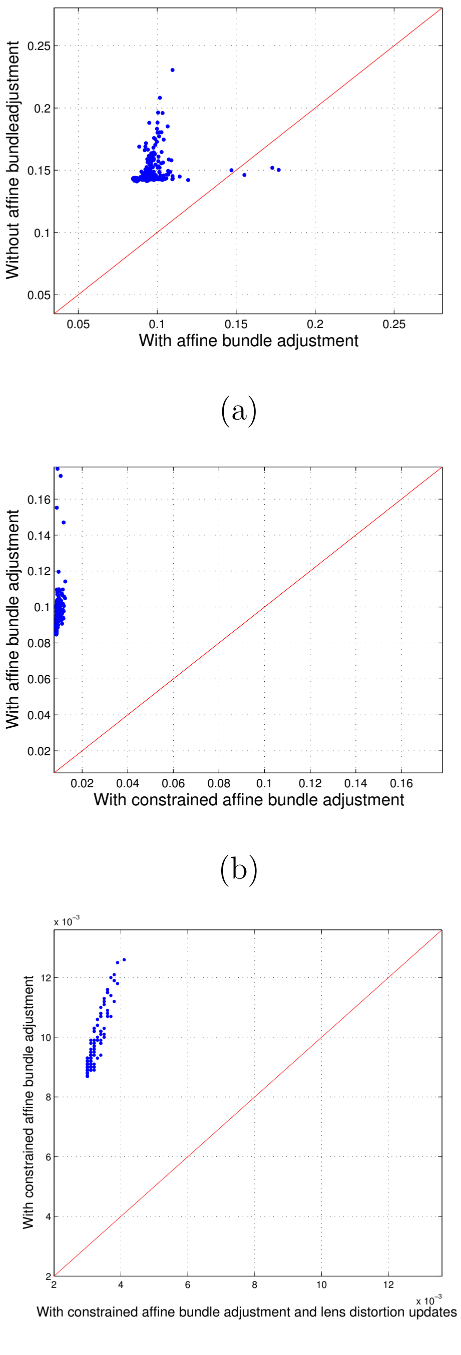


Figure 7.9: Four images generated by the four synthesized cameras: blue crosses (×) are original images and red dots (•) are images with lens distortions (coefficients $= [0.2, -0.3, -0.1, 0.1, 0.1]$).

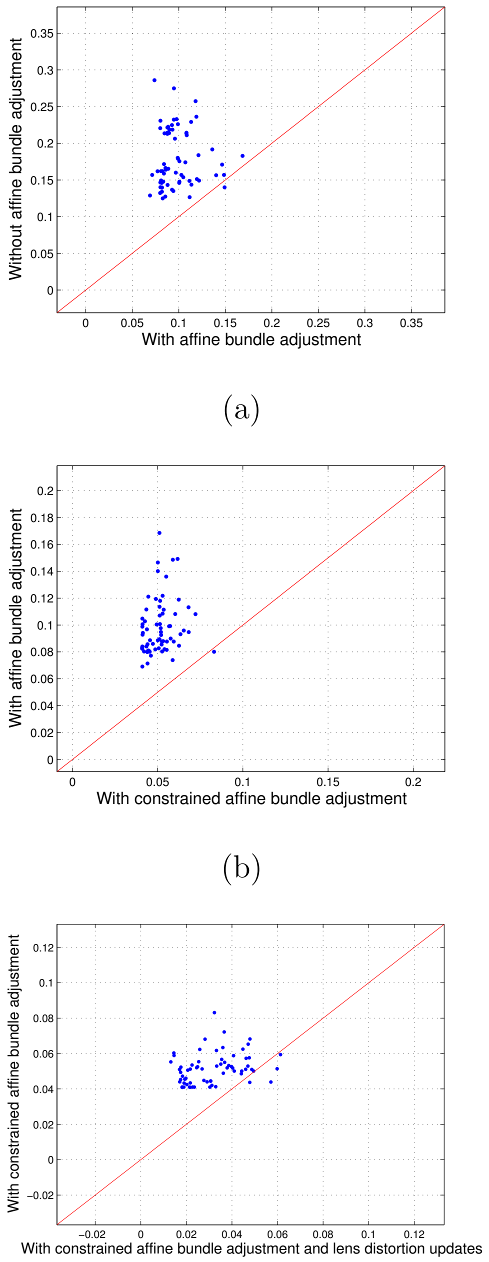


(a)

(b)

(c)

Figure 7.10: Noise-free synthesized data with lens distortion coefficients set to $k_c = [0.03,\ -0.08,\ 0.02,\ 0.01,\ -0.02]$: (a) No optimization versus ABA. (b) ABA versus CABA. (c) CABA versus CABA-LDU.

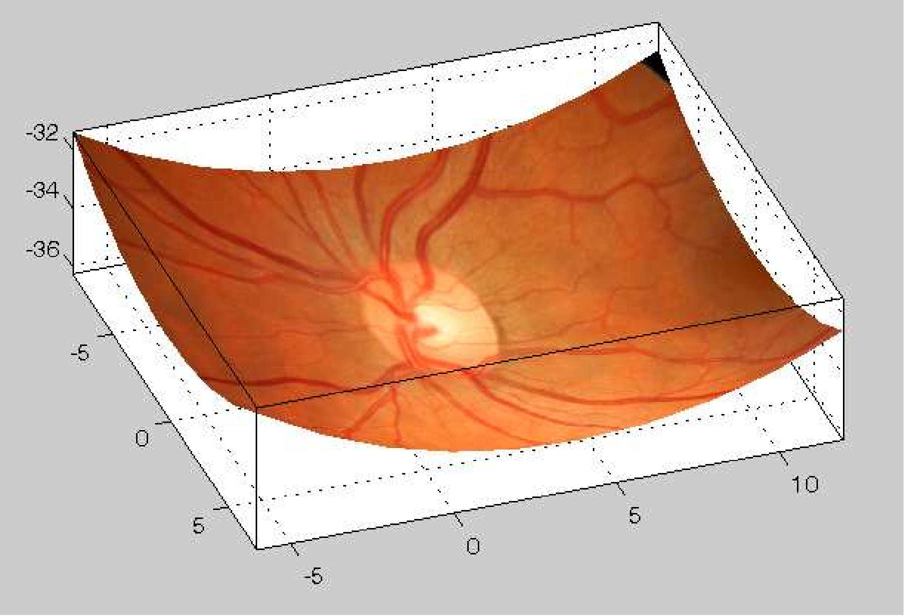


(a)

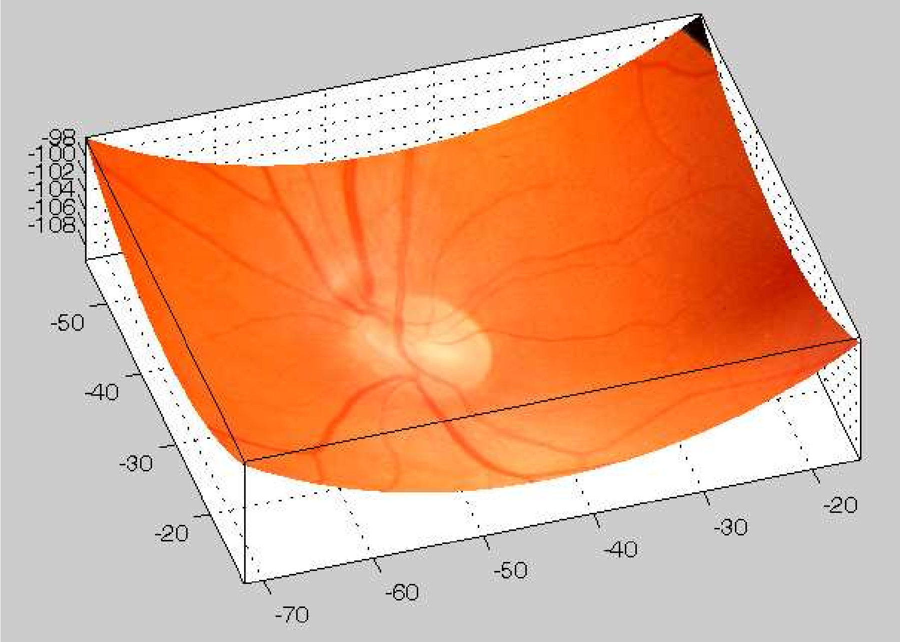
(b)

(c)

Figure 7.11: Retinal Images: (a) No optimization versus ABA. (b) ABA versus CABA. (c) CABA versus CABA-LDU.



(a)



(b)

Figure 7.12: The 3-D retinal reconstruction results with the retinal image mapped onto sphere surfaces.

CHAPTER 8

Conclusions and Future Works

In this dissertation, we have studied advanced retinal imaging research in the context of multi-view geometry that involves 2-D image registration (2-D/2-D) and 3-D structure reconstruction (2-D/3-D). With multiple retinal images, motion estimation is essential to infer structural information across images. As the prerequisite of this research, feature extraction, i.e., blood vessel segmentation, is also addressed by developing a modified local entropy thresholding algorithm. 2-D retinal image registration and 3-D retinal surface reconstruction algorithms are developed along with a feature extraction technique. Features serve as the input to motion estimation problems. 2-D retinal registration relates with the 2-D/2-D transformations. 3-D surface reconstruction requires a camera projection (a 3-D/2-D transformation) and the 3-D/3-D transformations.

The first problem of this work deals with feature extraction and correspondence selection. Inspired by the concepts of local entropy and relative entropy thresholding, we develop a new thresholding algorithm called modified local entropy thresholding algorithm where we develop a smoothed co-occurrence matrix to increase the entropy and to reduce the peak in the co-occurrence. Competitive experimental results on blood vessel segmentation are obtained by comparing with other state-of-the-art algorithms.

Next, a new 2-D retinal image registration framework is proposed. The proposed framework is able to overcome various limitations imposed by ETDRS image sets. Our unified hybrid framework is able (1) to select an appropriate transformation model, (2) to determine whether the sampling process is needed, and (3) to reject invalid registration results, so that we can maximize the success rate and minimize the registration error.

Then, we have studied 3-D retinal surface reconstruction method by using an affine camera model. An affine bundle adjustment algorithm based on a nonlinear optimization technique is established to refine an affine shape and affine cameras. A point-based sphere fitting method is introduced and the criteria for initial point selection are discussed. Simulation results on synthetic data show the robustness of our algorithm.

Last, a constrained optimization procedure is proposed to estimate structure and camera parameters simultaneously by minimizing a physically meaningful cost function. We introduce an optimization process which incorporates a geometrically meaningful measure and lens distortion into a cost function. We have proposed a new optimization algorithm called constrained affine bundle adjustment with lens distortion update. The procedure optimizes camera parameters, 3-D points, the physical shape of a retinal surface, and lens distortion, simultaneously.

The future research is discussed on the following issues.

- **Depth recovery of pathological areas** Currently we can obtain the global 3-D retinal surface. It is possible to utilize the current estimation as a baseline reference to estimate the local depth of pathological areas that is more valuable for disease diagnosis.

- **Constrained bundle adjustment in the Euclidean space** At present, we can improve the reconstruction result by moving all 3D points closer to a quadratic surface in an affine space. The lens distortion is estimated in the affine space, hence it may not be as accurate as one would expect. Although a constrained optimization in the Euclidean space is more computationally expensive, it could yields a better geometrically meaning result and better lens distortion approximation.

- **3-D retinal visualization** This research will require transformations of 3D/3D. 3-D surface models of several ETDRS views will be generated to fuse all of the partial retinal surfaces into one 3-D retinal model. This can be achieved by merging multiple parametric representations to a spherical surface that can be displayed and manipulated in a 3-D interactive visualization environment.

BIBLIOGRAPHY

[1] N. Ritter, R. Owens, J. Cooper, R. H. Eikelboom, and P. P. V. Saarloos, "Registration of stereo and temporal images of the retina," *IEEE Trans. Medical Imaging*, vol. 18, pp. 404–418, May 1999.

[2] G. K. Matsopoulos, N. A. Mouravliansky, K. K. Delibasis, and K. S. Nikita, "Automatic retinal image registration scheme using global optimization techniques," *IEEE Trans. Inform. Technol. Biomed.*, vol. 3, pp. 47–60, March 1999.

[3] K. Deguchi, D. Kawamata, K. Mizutani, H. Hontani, and K. Wakabayashi, "3d fundus shape reconstruction and display from stereo fundus images," *IEICE Trans. Inf. & Syst.*, vol. E83-D, pp. 1408–1414, July 2000.

[4] T. Choe, I. Cohen, and G. Medioni, "3-D shape reconstruction of retinal fundus," in *IEEE Conference on Computer Vision and Pattern Recognition CVPR*, vol. 2, pp. 2277–2284, June 2006.

[5] E. Corona, S. Mitra, M. Wilson, T. Krile, Y. H. Kwon, and P. Soliz, "Digital stereo image analyzer for generating automated 3-D measures of optic disc deformation in glaucoma," *IEEE Trans. on Medical Imaging*, vol. 21, pp. 1244–1253, October 2002.

[6] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Trans. Med. Imag.*, vol. 23, pp. 501–509, April 2004.

[7] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, “Classifying convex sets for vessel detection in retinal images,” in *Proc. of International Symposium on Biomedical Imaging*, pp. 269–272, July 2002.

[8] X. Jiang and D. Mojon, “Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal images,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, pp. 131–137, January 2005.

[9] A. Hoover, V. Kouznetsova, and M. Goldbaum, “Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response,” *IEEE Trans. Medical Imaging*, vol. 19, pp. 203–210, March 2000.

[10] US. National Institute of Health (NIH), National Eye Institute, “Diabetic eye disease..” http://www.nei.nih.gov/diabetes/.

[11] US. National Institute of Health (NIH), National Eye Institute, “Clinical studies database: Early treatment diabetic retinopathy study (ETDRS).” http://www.nei.nih.gov/neitrials/viewStudyWeb.aspx?id=53.

[12] US. National Institute of Health (NIH), National Eye Institute, “Diabetic retinopathy: What you should know..” http://www.nei.nih.gov/health/diabetic/retinopathy.asp.

[13] U. School of Medicine, University of Bermingham, “Diabetic retinopathy.” www.diabeticretinopathy.org.uk.

[14] A. D. Association, “Diabetes information.” http://www.diabetes.org.

[15] US. National Institute of Health (NIH), National Eye Institute, “Clinical trials: Early treatment of diabetic retinopathy study ETDRS.” http://www.clinicaltrials.gov/ct/show/NCT00000151.

[16] E. R. Group, “Fundus photographic risk factors for progression of diabetic retinopathy, ETDRS report number 12,” *Ophthalmology*, vol. 98, pp. 823–833, 1991.

[17] S. R. Fransen, T. C. Leonard-Martin, W. J. Feuer, and P. L. Hildebrand, “Clinical evaluation of patients with diabetic retinopathy: Accuracy of the inoveon diabetic retinopathy-3DT system,” *The American Academy of Ophthalmology*, vol. 109, pp. 595–601, March 2002.

[18] C. J. Rudnisky, B. J. Hinz, M. T. S. Tennant, A. R. de Leon, and M. D. J. Greve, “High-resolution stereoscopic digital fundus photography versus contact lens biomicroscopy for the detection of clinically significant macular edema,” *The American Academy of Ophthalmology*, vol. 109, pp. 267–274, February 2002.

[19] J. M. Brady, “Seeds of perception,” in *Proceedings of the 3rd Alvey Vision Conference, Cambridge University*, pp. 259–265, September 1987.

[20] J. Shi and C. Tomasi, “Good features to track,” in *IEEE Conference on Computer Vision and Pattern Recognition CVPR*, pp. 593–600, June 1994.

[21] Z. Zivkovic and F. V. der Heijden, “Better features to track by estimating the tracking convergence region,” in *IEEE Int'l Conference on Pattern Recognition*, vol. 2, pp. 635–638, August 2002.

[22] Y. S. Yao and R. Chellappa, “Tracking a dynamic set of feature points,” *IEEE Trans. Image Processing*, vol. 4, pp. 1382–1395, October 1995.

[23] V. Mahadevan, H. Narasimha-Iyer, B. Roysam, and H. L. Tanenbaum, “Robust model-based vasculature detection in noisy biomedical images,” *IEEE Trans. Information Technology in Biomedicine*, vol. 8, pp. 360–376, September 2004.

[24] L. Zhou, M. S. Rzeszotarski, L. Singerman, and J. M. Chokreff, “The detection and quantification of retinopathy using digital angiograms,” *IEEE Trans. Medical Imaging*, vol. 13, pp. 619–626, December 1994.

[25] C. Sinthanayothin, J. F. Boyce, H. L. Cook, and T. H. Williamson, “Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images,” *Br F Ophthalmol*, pp. 902–910, February 1999.

[26] A. Pinz, S. Bernogger, P. Datlinger, and A. Kruger, “Mapping the human retina,” *IEEE Trans. Medical Imaging*, vol. 17, pp. 606–619, August 1998.

[27] F. Zana and J. Klein, “Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation,” *IEEE Trans. Image Processing*, vol. 10, pp. 1010–1019, July 2001.

[28] C. L. Tsai, C. V. Stewart, B. Roysam, and H. L. Tanenbaum, “Model-based method for improving the accuracy and repeatability of estimating vascular bifurcations and crossovers from retinal fundus images,” *IEEE Trans. Information Technology in Biomedicine*, vol. 8, pp. 122–130, June 2004.

[29] Y. Bouaoune, M. K. Assogba, J. C. Nunes, and P. Bunel, “Spatio-temporal characterization of vessel segments applied to retinal angiographic images,” in *Pattern Recognition Letters*, vol. 24, 2003.

[30] T. Kondo, “Gradient orientation based feature detection: an application for extracting retinal blood vessels,” in *Proc. of Int’l Symposium on Inteeligent Multimedia, Video and Speech Processing*, pp. 194–197, October 2004.

[31] E. Grisan, A. Pesce, A. Giani, M. Foracchia, and A. Ruggeri, “A new tracking system for the robust extraction of retinal vessel structure,” in *IEEE Proc. of Int’l Conference on Engineering in Medicine and Biology*, vol. 1, pp. 1620–1623, 2004.

[32] C. Sinthanayothin, J. F. Boyce, H. L. Cook, T. H. Williamson, E. Mensah, S. Lal, and D. Usher, "Automated detection of diabetic retinopathy on digital fundus images," *Diabetic Med.*, no. 2, pp. 105–112, 2002.

[33] S. N. Kalitzin, J. J. Staal, B. M. T. H. Romeny, and M. A. Viergever, "A computational method for segmenting topological point sets and application to image analysis," *IEEE Trans. pattern Anal. Machine Intell.*, vol. 23, pp. 447–459, May 2001.

[34] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, "Detection of blood vessels in retinal images using two-dimensional matched filters," *IEEE Trans. Medical Imaging*, vol. 8, pp. 263–269, September 1989.

[35] A. Can, H. Shen, J. N. Turner, H. L. Tanenbaum, and B. Roysam, "Rapid automated tracing and feature extraction from retinal fundus images using direct exploratory algorithms," *IEEE Trans. Information Technology in Biomedicine*, vol. 3, pp. 125–138, June 1999.

[36] O. Chutatape, L. Zheng, and S. M. Krishnan, "Retinal bloodvessel detection and tracking by matched gaussian and kalman filters," in *International conference of the IEEE Engr. in Medicineand Biology Soc.*, vol. 20, pp. 3144–3149, 1998.

[37] D. Wu, M. Zhang, J. C. Liu, and W. Bauman, "On the adaptive detection of blood vessels in retinal images," *IEEE Trans. Biomedical Engineering*, vol. 53, pp. 341–343, February 2006.

[38] A. Can, C. V. Stewart, B. Roysam, and H. L. Tanenbaum, "A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, pp. 347–364, March 2002.

[39] A. Can, C. V. Stewart, B. Roysam, and H. L. Tanenbaum, “A feature-based technique for joint, linear estimation of higher-order image-to-mosaic transformations: Mosaicing the curved human retina,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, pp. 412–419, March 2002.

[40] F. Zana and J. Klein, “A multimodal registration algorithm of eye fundus images using vessels detection and hough transform,” *IEEE Trans. Biomedical Engineering*, vol. 18, pp. 419–428, May 1999.

[41] F. Zana and J. Klein, “A registration algorithm of eye fundus images using a bayesian hough transform,” in *IEEE International Conference on Image Processing and Its Applications*, vol. 2, pp. 479–483, July 1999.

[42] H. Shen, C. V. Stewart, B. Roysam, G. Lin, and H. L. Tanenbaum, “Frame-rate spatial referencing based on invariant indexing and alignment with application to online retinal image registration,” *IEEE Trans. pattern Anal. Machine Intell.*, vol. 25, pp. 379–384, March 2003.

[43] C. V. Stewart, C. L. Tsai, and B. Roysam, “The dual-bootstrap iterative closest point algorithm with application to retinal image registration,” *IEEE Trans. Med. Imag.*, vol. 22, pp. 1379–1394, November 2003.

[44] F. Laliberte, L. Gagnon, and Y. Sheng;, “Registration and fusion of retinal images - an evaluation study,” *IEEE Trans. Medical Imaging*, vol. 22, pp. 661–673, May 2003.

[45] C. L. Tsai, C. V. Stewart, B. Roysam, and H. L. Tanenbaum, “Covariance-driven retinal image registration initialized from small sets of landmark correspondences,” in *IEEE International Symposium on Biomedical Imaging, 2002*, pp. 333–336, July 2002.

[46] E. H. Zhang, Y. Zhang, and T. X. Zhang, “Automatic retinal image registration based on blood vessels feature point,” in *IEEE International Conference on Machine Learning and Cybernetics*, vol. 4, pp. 2010–2015, November 2002.

[47] W. E. Hart and M. H. Goldbaum, “Registering retinal images using automatically selected control point pairs,” in *IEEE Proc. of International conference on Image Processing*, vol. 3, pp. 576–580, November 1994.

[48] C. Heneghan, P. Maguire, N. Ryan, and P. de Chazal, “Retinal image registration using control points,” in *IEEE Proc. of International Symposium Biomedical Imaging*, pp. 349–352, July 2002.

[49] J. Park, J. M. Keller, P. D. Gader, and R. A. Schuchard, “Hough-based registration of retinal images,” in *IEEE Proc. of International Conference on Systems, Man, and Cybernetics*, vol. 5, pp. 4550 –4555, October 1998.

[50] P. Jasiobedzki, “Registration of retinal images using adaptive adjacency graphs,” in *IEEE Symposium on Computer-Based Medical Systems*, pp. 40–45, June 1993.

[51] A. M. Mendonca, A. Campilho, and J. M. R. Nunes, “A new similarity criterion for retinal image registration,” in *IEEE Proc. of International Conference on Image Processing*, vol. 3, pp. 696–700, September 1994.

[52] L. G. Brown, “A survey of image registration rechniques,” *ACM Computing Surveys*, vol. 24, pp. 325–376, December 1992.

[53] B. Zitova and J. Flusser, “Image registration methods: A survey,” *Image and Vision Computing*, vol. 21, pp. 977–1000, 2003.

[54] J. B. A. Maintz and M. A. Viegever, “A survey of medical image registration,” *Medical Image Analysis*, 1998. Oxford University Press.

[55] P. A. van den Elsen, E. D. Pol, and M. A. Viergever, “Medical image matching - a review with classification,” *IEEE Engineering in Medicine and Biology Mag.*, vol. 12, pp. 26–39, March 1993.

[56] R. Wan and M. Li, “An overview of medical image registration,” in *IEEE Proc. of fifth International Conference on Comp. Intell. and Multimedia Application*, pp. 385–390, September 2003.

[57] H. Lester and S. Arridge, “A survey of hierarchical nonlinear medical image registration,” *Pattern Recognit.*, vol. 32, no. 1, pp. 129–149, 1999.

[58] T. Chanwimaluang and G. Fan, “Hybrid retinal image registration,” *IEEE Trans. Information Technology in Biomedicine*, vol. 10, pp. 129–142, January 2006.

[59] T. Chanwimaluang and G. Fan, “Retinal image registration for NIH’s ETDRS,” in *International Symposium on Visual Computing, A Lecture Notes*, vol. 3804, pp. 51–59, December 2005.

[60] B. Fang and Y. Tang, “Elastic registration for retinal images based on reconstructed vascular trees,” *IEEE Trans. Biomedical Engineering*, no. 99, 2006.

[61] G. K. Matsopoulos, P. A. Asvestas, N. A. Mouravliansky, and K. K. Delibasis, “Multimodal registration of retinal images using self organizing maps,” *IEEE Trans. Medical Imaging*, vol. 23, pp. 1557–1563, December 2004.

[62] K. Deguchi, J. Noami, and H. Hontani, “3d fundus pattern reconstruction and display from multiple images,” in *IEEE Int’l conference on Pattern Recognition*, vol. 4, pp. 94–97, September 2000.

[63] S. Mitra, D. J. Lee, and T. F. Krile, “3-D representation from time-sequenced biomedical images using 2-D cepstrum,” in *IEEE Proc. of Conference on Visualization in Biomedical Computing*, pp. 401–408, May 1990.

[64] Z. Wang, A. C. Bovic, and L. Lu, “Why is image quality assessment so difficult?,” in *IEEE Proc. of International Conference on Acoustics, Speech, and Signal Processing, ICASSP*, vol. 4, pp. 3313–3316, May 2002.

[65] Z. Wang and A. C. Bovic, “A universal image quality index,” *IEEE Signal Processing Letters*, vol. 9, pp. 81–84, March 2002.

[66] H. R. Sheikh, A. C. Bovik, and G. de Veciana, “An information fidelity criterion for image quality assessment using natural scene statistics,” *IEEE Trans. on Image Processing*, vol. 14, pp. 2117–2128, December 2005.

[67] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, “Image quality assessment based on a degradation model,” *IEEE Trans. on Image Processing*, vol. 9, pp. 636–650, April 2000.

[68] S. C. Lee and Y. Wang, “Automatic retinal image quality assessment and enhancement,” in *Proc. SPIE Conf. on Image Processing*, pp. 1581–1590, February 1999.

[69] M. Lalonde, L. Gagnon, and M. C. Boucher, “Automatic visual quallity assessment in optical fundus images,” in *Proc. of Vision Interface*, pp. 259–264, June 2001.

[70] A. Awawdeh and G. Fan, “Pseudo cepstrum for assessing stereo quality of retinal images,” in *Proc. of the 37th Asilomar Conference on Signals, Systems, and Computers*, November 2003.

[71] S. Mann and R. W. Picard, "Video orbits of the projective group: A simple approach to featureless estimation of parameters," *IEEE Trans. Image Processing*, vol. 6, pp. 1281–1295, September 1997.

[72] S. Lertrattanapanich, "Image registration for video mosaic," Master's thesis, Pennsylvania State Univ., University Park, 1999.

[73] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision: Second Edition*. Cambridge University Press, 2003.

[74] J. J. Koenderink and A. J. V. Doorn, "Affine structure from motion," *Journal of Optical Society of America*, vol. 8, pp. 377–385, February 1991.

[75] L. Quan and R. Mohr, "Towards structure from motion for linear features through reference points," in *IEEE Workshop on Visual Motion*, pp. 249–254, October 1991.

[76] S. Demey, A. Zisserman, and P. Beardsley, "Affine and projective structure from motion," *Proc. British Machine Vision Conference (BMVC)*, pp. 49–58, 1992.

[77] L. S. Shapiro, *Affine Analysis of Image Sequences*. PhD thesis, Sharp Laboratories of Europe, Oxford, Oxford, UK, 1995.

[78] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: A factorization method," *Intl. Journal of Computer Vision*, vol. 9, pp. 137–154, November 1992.

[79] L. S. Shapiro, A. Zisserman, and M. Brady, "3d motion recovery via affine epipolar geometry," *Int'l Journal Computer Vision*, vol. 16, pp. 147–182, October 1995.

[80] I. Shimshoni, R. Basri, and E. Rivlin, “A geometric interpretation of weak-perspective motion,” *IEEE Trans. pattern Anal. Machine Intell.*, vol. 21, pp. 252–257, March 1999.

[81] T. Pun, “A new method for gray-level picture thresholding using the entropy of an histogram,” *Signal Processing*, vol. 2, pp. 223–237, 1980.

[82] T. Pun, “Entropic thresholding: A new approach,” *Comp. Graphics and Image Proc.*, vol. 16, pp. 210–239, 1981.

[83] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication.* The University of Illinois Press, 1949.

[84] J. N. Kapur, P. K. Sahoo, and A. K. C. Wong, “A new method for gray-level picture thresholding using the entropy of an histogram,” *Comp. Graphics Vision and Image Proc.*, vol. 29, pp. 273–285, 1985.

[85] N. R. Pal and S. K. Pal, “Entropic thresholding,” *Signal processing*, vol. 16, pp. 97–108, 1989.

[86] N. R. Pal and S. K. Pal, “Segmentation using contrast and homogeneity measures,” *Pattern Recognition Letters*, vol. 5, pp. 293–304, 1987.

[87] C. Chang, K. Chen, J. Wang, and M. Althouse, “A relative entropy-based approach to image thresholding,” *Pattern recognition*, vol. 27, no. 9, pp. 1275–1289, 1994.

[88] R. Fisher, S. Perkins, A. Walker, and E. Wolfart, “Connected components labeling.”

[89] A. Hoover, “The stare project.” http://www.ces.clemson.edu/∼ahoover/stare.

[90] N. Mouravliansky, G. K. Matsopoulos, K. Delibasis, and K. S. Nikita, "Automatic retinal registration using global optimization techniques," in *Proc. of International Conference on Eng. in Med. Biol. Society*, vol. 2, pp. 567–570, November 1998.

[91] D. Lloret, J. Serrat, A. M. Lopez, A. Soler, and J. J. Villaneuva, "Retinal image registration using creases as anatomical landmarks," in *IEEE Proc. of International Conference on Pattern Recognition*, vol. 3, pp. 203–206, September 2000.

[92] A. V. Cideciyan, "Registration of ocular fundus images: an algorithm using cross-correlation of triple invariant image descriptors," *IEEE Eng. in Med. Biol. Mag.*, vol. 14, pp. 52–58, January - February 1995.

[93] L. Ballerini, "Temporal matched filters for integration of ocular fundus images," in *IEEE Conference on Digital Signal Processing Proceedings*, vol. 2, pp. 1161–1164, July 1997.

[94] M. Skokan, A. Skoupy, and J. Jan, "Registration of multimodal images of retina," in *IEEE Conference Eng. Med. Biol.*, vol. 2, pp. 1094–1096, October 2002.

[95] T. Chanwimaluang and G. Fan, "An efficient blood vessel detection algorithm for retinal images using local entropy thresholding," in *IEEE Proc. of International Symposium on Circuits and Systems*, vol. 5, pp. V21–V24, May 2003.

[96] T. Chanwimaluang and G. Fan, "An efficient algorithm for extraction of anatomical structures in retinal images," in *IEEE Proc. of International conference on Image Processing*, vol. 1, pp. I1093–I1096, September 2003.

[97] J. P. W. Pluim, J. B. Antoine, and M. A. Viegever, "Mutual-information-based registration of medical images: A survey," *IEEE Trans. Medical Imaging*, vol. 22, pp. 986–1004, August 2003.

[98] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3d medical image alignment," *Pattern Recognit.*, vol. 32, no. 1, pp. 71–86, 1999.

[99] A. Collignon, *Multi-Modality Medical Image Registration by Maximization of Mutual Information.* PhD thesis, Catholic Univ., Leuven, Belgium, 1998.

[100] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality medical image registration by maximization of mutual information," *IEEE Trans. Medical Imaging*, vol. 16, pp. 187–198, April 1997.

[101] P. J. Besl and N. D. McKay, "A method for registration of 3-D shape," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 239–256, February 1992.

[102] T. Chanwimaluang and G. Fan, "Affine camera for 3-D retinal reconstruction," in *International Symposium on Visual Computing*, vol. 2, pp. 19–30, November 2006.

[103] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *IEEE Int'l Conference on Computer Vision*, vol. 1, pp. 666–673, September 1999.

[104] C. C. Slama, *Manual of Photogrammetry.* American Soiety of Photogrammetry, 4th ed., 1980.

[105] J. Heikkila and O. Silven, "A four-step camera calibration procedure with implicit image correction," in *IEEE Proc. Computer Vision and Pattern Recognition*, pp. 1106–1112, June 1997.

[106] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon, “Bundle adjustment: A modern synthesis,” *Vision Algorithms: Theory And Practice, Springer-Verlag*, 2000.

[107] P. Tresadern and I. Reid, “Uncalibrated and unsynchronized human motion capture: a stereo factorization approach,” in *IEEE Conference on Computer Vision and Pattern Recognition CVPR*, vol. 1, pp. 128–134, June 2004.

[108] D. Weinshall and C. Tomasi, “Linear and incremental acquisition of invariant shape models from image sequences,” in *IEEE Proc. of 4th Int'l Conference on Computer Vision*, vol. 17, pp. 512–517, 1993.

[109] D. Weinshall and C. Tomasi, “Linear and incremental acquisition of invariant shape models from image sequences,” *IEEE Trans. pattern Anal. Machine Intell. PAMI*, vol. 17, pp. 512–517, May 1995.

[110] C. J. Poleman and T. Kanade, “A paraperspective factorization method for shape and motion recovery,” in *Proc. of 3rd European Conference on Computer Vision*, pp. 97–108, May 1994.

[111] C. J. Poleman and T. Kanade, “A paraperspective factorization method for shape and motion recovery,” *IEEE Trans. Pattern and Machine Intelligent*, vol. 19, pp. 206–218, March 1997.

[112] L. Quan, “Self-calibration of an affine camera from multiple views,” *International Journal of Computer Vision*, vol. 19, no. 1, pp. 93–110, 1996.

[113] T. Kurata, J. Fujiki, and K. Sakaue, “Affine epipolar geometry via factorization method,” in *Proc. of 14th Int'l Conference on Pattern Recognition*, vol. 1, pp. 862–866, August 1998.

[114] A. Bartoli and P. Sturm, "Constrained structure and motion from multiple uncalibrated views of a piecewise planar scene," *Int'l J. of Computer Vision*, vol. 52, pp. 45–64, April 2003.

[115] R. Szeliski and P. H. S. Torr, "Geometrically constrained structure from motion: Points on planes," in *Workshop on 3D Structure from Mulitple Images of Large-scale Environment (SMILE)*, June 1998.

[116] Y. Shan, Z. Liu, and Z. Zhang, "Model-based bundle adjustment with application to face modeling," in *IEEE Int'l Conf. on Computer Vision ICCV*, vol. 2, pp. 644–651, July 2001.

[117] P. Fua, "Regularized bundle-adjustment to model heads from image sequences without calibration data," *Int'l J. of Computer Vision*, vol. 38, July 2000.

[118] R. Gong and G. Xu, "Quadratic surface reconstruction from multiple views using sqp," *Integrated Image and Graphics Technologies*, pp. 197–217, 2004.

[119] P. T. Boggs and J. W. Tolle, *Sequential Quadratic Programming.* Acta Numerica, 1995.

[120] K. H. Wong and Y. C. Ming, "3d model reconstruction by constrained bundle adjustment," in *IEEE Proc. of International Conference on Pattern Recognition, ICPR*, vol. 3, pp. 902–905, August 2004.

[121] I. Anonymous, "–," in –.

[122] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Trans. Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.

[123] G. Champleboux, S. Lavallee, R. Szeliski, and L. Brunie, "From accurate range imaging sensor calibration to accurate model-based 3D object localization," in

IEEE Conference on Computer Vision and Pattern Recognition CVPR, pp. 83–89, June 1992.

[124] S. Dua, N. Kandiraju, and H. W. Thompson, "Design and implementation of a unique blood-vessel detection algorithm towards early diagnosis of diabetic retinopathy," in *IEEE Int'l conference of Information Technology: Coding and Computing*, vol. 1, pp. 26–31, April 2005.

[125] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using hausdorff distance," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, pp. 850–863, September 1993.

[126] Inoveon Corporation, "Preserving sight is our vision." http://www.inoveon.com/index.html.

[127] C. Kerbas and F. K. H. Quek, "Vessel extraction techniques and algorithms: A survey," in *Proc. of International Conference on Image Processing*, vol. 2, pp. 837–840, October 2001.

[128] M. Lalonde, M. Beaulieu, and L. Gagnon, "Fast and robust optic disc detection using pyramidal decomposition and hausdorff-based template matching," *IEEE Trans. Medical Imaging*, vol. 20, pp. 1193–1200, November 2001.

[129] H. Li and O. Chutatape, "Automatic location of optic disk in retinal images," in *Proc. of International Conference on Image Processing*, vol. 2, pp. 837–840, October 2001.

[130] G. Lin, C. V.Stewart, B. Roysam, K. Fritzsche, G. Yang, and H. L. Tanenbaum, "Predictive scheduling algorithms for real-time feature extraction and spatial referencing: application to retinal image sequences," *IEEE Trans. Biomedical Imaging*, vol. 51, pp. 115–125, January 2004.

[131] A. Osareh, M. Mirmehdi, B. Thomas, and R. Markham, "Comparison of colour spaces for optic disc localisation in retinal images," in *IEEE 16th International Conference on Pattern Recognition*, vol. 1, pp. 743–746, 2002.

[132] H. S. Sawhney and R. Kumar, "True multi-image alignment and its application to mosaicing and lens distortion correction," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, pp. 235–243, March 1999.

[133] K. Sengupta, S. Wang, C. C. Ko, and P. Burman, "Automatic face modeling from monocular image sequences using modified non parametric regression and an affine camera model," in *IEEE Int'l Conf. on Automatic Face and Gesture Recognition*, pp. 524–529, March 2000.

[134] K. Sengupta and C. C. Ko, "Scanning face models with desktop cameras," *IEEE Trans. Industrial Electronics*, vol. 48, pp. 904–912, October 2001.

[135] N. H. Solouma, A. M. Youssef, Y. A. Badr, and Y. M. Kadah, "A new real-time retinal tracking systems for image-guided laser treatment," *IEEE Trans. Biomedical Engineering*, vol. 49, pp. 1059–1067, September 2002.

[136] N. H. Solouma, A. M. Youssef, Y. A. Badr, and Y. M. Kadah, "Robust computer-assisted laser treatment using real-time retinal tracking," in *Proc. of International Conference on Engineering in Medicine and Biology Society*, vol. 3, pp. 2499–2502, October 2001.

[137] W. Tan, Y. Wang, and S. Lee, "Retinal blood vessel detection using frequency analysis and local-mean interpolation filters," in *Proc. SPIE Medical Imaging: Image Processing*, vol. 4322, pp. 1373–1384, 2001.

[138] M. Turk and A. P. Pentland, "Face recognition using eigenfaces," *IEEE Conference on Computer Vision and Pattern Recognition CVPR*, pp. 586–591, June 1991.

[139] S. Ullman and R. Basri, “Recognition by linear combinations of models,” *IEEE Trans Pattern Recognition and Machine Intelligent PAMI*, vol. 13, no. 10, pp. 992–1006, 1991.

[140] P. Viola and W. M. W. III, “Alignment by maximization of mutual information,” in *Proc. IEEE International Conference on Computer Vision*, pp. 16–23, June 1995.

[141] D. J. Williams and M. Shah, “A fast algorithm for active contours,” in *Proc. IEEE International Conference on Computer Vision*, pp. 592–595, December 1990.

[142] Department of Ophthalmology and Visual Sciences, University of Wisconsin - Madison, “Fundus photograph reading center.” http://eyephoto.ophth.wisc.edu/Photographers.html.

VITA

Thitiporn Chanwimaluang

Candidate for the Degree of

Doctor of Philosophy

Dissertation: ADVANCED RETINAL IMAGING: FEATURE EXTRACTION, 2-D REGISTRATION, AND 3-D RECONSTRUCTION

Major Field: Electrical and Computer Engineering

Biographical:

Personal Data:

Education:

Thitiporn Chanwimaluang was born in Bangkok, Thailand, in 1975. She received her B.Eng.'s degree in Electrical Engineering from Chilalongkorn University, Thailand, and M.S.'s degree in Electrical Engineering from the Pennsylvania State University, State College, PA, in 1995 and 2001, respectively. In 1996, she was an engineer at the Siemens (Thailand). From 1997-1998, she was an engineer at the Thai Telephone and Telecommunication (TT&T). She receive her Ph.D. in Electrical Engineering in December 2006 from Oklahoma State University. Her research interests are signal and image processing. experience attained

Name: Thitiporn Chanwimaluang Date of Degree: December, 2006

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: ADVANCED RETINAL IMAGING: FEATURE EXTRACTION, 2-D REGISTRATION, AND 3-D RECONSTRUCTION

Pages in Study: 174 Candidate for the Degree of Doctor of Philosophy

Major Field: Electrical and Computer Engineering

In this dissertation, we have studied feature extraction and multiple view geometry in the context of retinal imaging. Specifically, this research involves three components, i.e., feature extraction, 2-D registration, and 3-D reconstruction. First, the problem of feature extraction is investigated. Features are significantly important in motion estimation techniques because they are the input to the algorithms. We have proposed a feature extraction algorithm for retinal images. Bifurcations/crossovers are used as features. A modified local entropy thresholding algorithm based on a new definition of co-occurrence matrix is proposed. Then, we consider 2-D retinal image registration which is the problem of the transformation of 2-D/2-D. Both linear and nonlinear models are incorporated to account for motions and distortions. A hybrid registration method has been introduced in order to take advantages of both feature-based and area-based methods have offered along with relevant decision-making criteria. Area-based binary mutual information is proposed or translation estimation. A feature-based hierarchical registration technique, which involves the affine and quadratic transformations, is developed. After that, a 3-D retinal surface reconstruction issue has been addressed. To generate a 3-D scene from 2-D images, a camera projection or transformations of 3-D/2-D techniques have been investigated. We choose an affine camera to characterize for 3-D retinal reconstruction. We introduce a constrained optimization procedure which incorporates a geometrically penalty function and lens distortion into the cost function. The procedure optimizes all of the parameters, camera's parameters, 3-D points, the physical shape of human retina, and lens distortion, simultaneously. Then, a point-based spherical fitting method is introduced. The proposed retinal imaging techniques will pave the path to a comprehensive visual 3-D retinal model for many medical applications.

ADVISOR'S APPROVAL: ______________________________