Automated 3-D optic disk reconstruction and feature extraction for glaucoma analysis

Xu, Juan

2006

Xu, J. (2006). Automated 3-D optic disk reconstruction and feature extraction for glaucoma analysis. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/3463>

<https://doi.org/10.32657/10356/3463>

Nanyang Technological University

Downloaded on 26 Apr 2023 14:25:49 SGT

Automated 3-D Optic Disk Reconstruction and Feature Extraction for Glaucoma Analysis

Xu Juan

School of Electrical & Electronic Engineering

A thesis submitted to Nanyang Technological University
in fulfilment of the requirement for the degree of
Doctor of Philosophy

2006

Acknowledgements

First and foremost, I would like to express my deep appreciation to my supervisors, Associate Professor Eric Sung and Dr. Opas Chutatape, for their patient guidance and support throughout the duration of this research. Without their encouragement and help, I would never finish this thesis. Credit goes to Dr. Shankar M. Krishnan in Biomedical Engineering Research Center (BMERC), for his support of my research.

I wish to thank Dr. Paul Chew Tec Kuan and Dr. Zheng Ce from the National University Hospital, and Dr. Luo Gang from Schepens Eye Research Institute, for their valuable helps in providing various image data, groundtruth, and interpretation of the medical background.

I also want to express my appreciations to School of Electrical and Electronic Engineering (EEE), Nanyang Technological University (NTU), Singapore, for offering me the study opportunities and scholarship, providing me with advanced research facilities and giving me chances to attend local and international conferences.

I am grateful to the people in our cluster. Thanks Mr. Zhou Jiayin, Mr. Li Peng, Ms. Fang Wen, Mr. Zhang Xiaohui for giving me many useful advices. I would like to thank all the friends in the BMERC who make a warm environment for research.

Finally, my appreciation goes to my parents, my sister and my dear husband, for their love, support and understanding. They keep encouraging me and give me the power to carry on my research.

Summary

Color fundus image processing is a challenging research topic in medical image processing. The visualization and parameter estimation of the optic disk, one of the main components on fundus image, is investigated in this research. Computer-aided processing for the measurement of optic disk boundary and its three dimensional (3-D) shape not only provides more clinical information which is difficult to be obtained manually, but also relieves the clinicians from time-consuming workloads.

As a result of this research, a glaucoma analyzing system is presented as an assisting tool for clinicians to quantitatively evaluate and monitor optic disk condition. A number of novel techniques are introduced in each module of this system, which includes 3-D optic disk reconstruction, optic disk boundary detection, cup boundary estimation, disk parameter measurement and analysis.

The 3-D reconstruction procedures include camera calibration, correspondence search, depth recovery, and eye-optics component estimation. Two camera calibration methods are compared and the better one is selected for the measurement of camera intrinsic parameters. Simultaneous stereo fundus camera is investigated and its focal length is measured for use in the 3-D reconstruction. A precise combined matching approach is employed to detect the matching points and then convert them into depths. The proposed algorithm fuses the advantages of both correlation method and feature-based method to prune uncertain matching and obtain more precise results. Different from the ordinary 3-D reconstruction of other objects, the optical effect of light media within the eyeball is considered in this study. A new method is proposed to calibrate and

integrate this effect into the reconstruction process to provide a real 3-D image of the optic disk.

A novel deformable model algorithm is used to precisely detect the optic disk boundary. The proposed algorithm modifies and extends the original snake (deforming-only technique) by including three modifications which are (i) radial deformation, (ii) self-adjusted clustering and (iii) smoothing update. The modifications not only directly solve the problem of blood vessel influence, but also show good performance for ill-defined edges, noises, and fuzzy shapes. Based on 100 randomly-selected testing images, the success rates are 94% for the proposed algorithm, as compared to 12% for the gradient vector flow (GVF) snake and 82% for the modified active shape model (ASM) method, which show the effectiveness of the proposed algorithm.

The previously obtained true-color 3-D optic disk image and precise disk boundary make it possible to automatically estimate cup boundary and related disk parameters. The deformable model technique is applied to estimate the cup boundary using combined energies of depth, gradient, smoothness, and shape. An optic disk coordinate system is established, based on which disk parameters are measured. The results are validated clinically to analyze and grade the stage of glaucoma.

In summary, an entire glaucoma analyzing system, consisting of various novel approaches on 3-D optic disk reconstruction and feature extraction, is presented in this research as an assisting tool for clinicians. The developed system achieves satisfactory results which can be used to facilitate the quantitative analysis, clinical diagnosis, and monitoring of glaucoma progression.

Table of Contents

Acknowledgements	i
Summary.....	ii
Table of Contents	iv
List of Figures.....	ix
List of Tables	xii
Chapter 1 Introduction.....	1
1.1 Motivation.....	3
1.2 Objectives	6
1.3 Major Contributions of the Thesis	7
1.4 Organization of the Thesis	9
Chapter 2 Background and Literature Review.....	11
2.1 Eye Structure.....	11
2.2 Related Eye Diseases	13
2.2.1 Glaucoma	13
2.2.1.1 Glaucoma Type.....	15
2.2.1.2 Glaucoma Evaluation.....	16
2.2.2 Diabetic Retinopathy	17
2.2.3 Cataract	18
2.3 Related Ophthalmic Instruments.....	18
2.3.1 Monocular Fundus Camera.....	19
2.3.2 Simultaneous Stereo Fundus Camera	22
2.3.3 Heidelberg Retina Tomograph (HRT)	23
2.3.4 Optical Coherence Tomography (OCT)	25
2.4 Review of 3-D Reconstruction.....	26
2.4.1 Camera Calibration	28

2.4.2 Correspondence Search and 3-D Reconstruction	31
2.5 Review of Boundary Detection.....	37
2.5.1 Deformable Models	38
2.5.1.1 Free-form Deformable Models (Active Contour Models).....	39
2.5.1.2 Parametrically Deformable Models	43
2.5.2 Optic Disk Localization and Approximate Boundary Estimation	44
2.5.3 Optic Disk Boundary Detection.....	46
2.6 Review of Other Related Projects.....	47
2.6.1 Blood Vessel Detection.....	48
2.6.2 Fundus Image Feature Extraction	49
2.6.3 Analysis of Disease.....	50
Chapter 3 Camera Calibration.....	51
3.1 Introduction of Geometry of 3-D Vision	52
3.1.1 Single Camera, Pinhole Camera Model.....	52
3.1.1.1 Four Coordinate Systems.....	52
3.1.1.2 The Relation of Four Coordinate Systems.....	54
3.1.2 Two Cameras, Stereopsis.....	56
3.1.3 Parallel Stereo Configuration.....	60
3.2 Camera Calibration	62
3.2.1 Camera Calibration Based on 3-D Object	63
3.2.1.1 Estimating Projection Matrix	64
3.2.1.2 Lens Distortion.....	66
3.2.1.3 Implementing Levenberg-Marquardt Method for Nonlinear Optimization	69
3.2.2 Camera Calibration Based on Planar Object	72
3.2.3 Results and Comparison	75
3.2.3.1 Results of 3-D-object-based Camera-calibration Method	75
3.2.3.2 Results of Planar-object-based Camera-calibration Method	79
3.2.3.3 Comparisons and Discussions.....	84

Table of Contents	vi
3.2.4 Conclusion	85
3.3 Study of Stereo Fundus Camera	86
Chapter 4 3-D Optic Disk Reconstruction.....	92
4.1 Introduction.....	92
4.2 Methodology of 3-D Reconstruction	95
4.2.1 Background of Stereo Pair System	95
4.2.2 Preprocessing and Initial Disparity Detection	97
4.2.3 Combined Matching.....	102
4.2.3.1 Normalized Cross Correlation Method.....	103
4.2.3.2 Feature-based Method.....	103
4.2.3.3 Sub-pixel Disparity and Constraint-based Modification	104
4.2.4 Auto-adjustments of Searching Range and Vertical Disparity	107
4.2.5 Dense Depth Recovery Based on Multiple Windows.....	109
4.3 Evaluation	113
4.4 Calibration of Optical Effect of the Eye	114
4.4.1 Modeling Eye Optical Effect	114
4.4.2 Estimation of Zr_0	116
4.5 Results and Discussions	119
4.6 Conclusion	130
Chapter 5 Optic Disk Boundary Detection.....	132
5.1 Introduction.....	132
5.2 Optic Disk Boundary Detection by Modified Snake	135
5.2.1 Initializing Optic Disk Center and Boundary	136
5.2.2 Contour Deformation Based on Combination of Local Energies	139
5.2.2.1 Energy Definition.....	139
5.2.2.2 Parameter Setting	141
5.2.2.3 Stop Criterion.....	142
5.2.3 Clustering.....	143

Table of Contents	vii
5.2.3.1 Feature Selection.....	145
5.2.3.2 Self-adjusted Clustering Process.....	147
5.2.4 Smoothing Update	148
5.2.4.1 Updating Process	151
5.2.4.2 Auto-adjusted Updating.....	153
5.3 Evaluation	154
5.4 Results and Discussions.....	155
5.5 Conclusion	164
Chapter 6 Disk Parameter Estimation.....	166
6.1 Introduction.....	166
6.2 Cup Boundary Estimation.....	170
6.2.1 Energy Definition and Contour Deformation	171
6.2.2 Parameter Setting.....	176
6.3 Extraction of Disk Parameters for Analysis.....	178
6.3.1 Optic Disk Coordinate System.....	178
6.3.2 Optic Disk Parameters	181
6.4 Results and Discussions.....	181
6.4.1 Cup boundary.....	181
6.4.2 C/D Vertical Ratio for Evaluation and Analysis.....	186
6.4.2.1 Evaluation	186
6.4.2.2 Analysis.....	188
6.4.3 Other Disk Parameters for Analysis	191
6.5 Conclusion	192
Chapter 7 Conclusion and Recommendations	194
7.1 Conclusion	194
7.2 Recommendations.....	197
Author's Publications	199

Table of Contents	viii
Bibliography	201
Appendix A Image Rectification.....	216
Appendix B K-means Algorithm	219
Appendix C Fundus Image Data	220

List of Figures

Figure 1.1: Features on fundus images	2
Figure 1.2: Stereo viewer	4
Figure 1.3: Glaucoma analyzing system.....	9
Figure 2.1: The structure of eyeball	12
Figure 2.2: Cross section of optic disk for normal eye and glaucoma eye	14
Figure 2.3: Glaucoma.....	14
Figure 2.4: Cross section of optic disk.....	16
Figure 2.5: Cup-to-disk ratio.....	17
Figure 2.6: Angle of view	19
Figure 2.7: Non-mydratic fundus camera	20
Figure 2.8: Inner structure of fundus camera.....	21
Figure 2.9: Fundus image	21
Figure 2.10: Inner structure of simultaneous stereo fundus camera	22
Figure 2.11: Inner structure of HRT	23
Figure 2.12: Confocal scanning laser system	24
Figure 2.13: OCT image of optic disk	26
Figure 3.1: The geometry of the camera model	53
Figure 3.2: The coordinates of a projected point	54
Figure 3.3: The relationship between image Euclidean coordinate system and image affine coordinate system.....	55
Figure 3.4: The epipolar geometry in stereopsis.....	57
Figure 3.5: Parallel stereo configuration.....	60
Figure 3.6: Geometry of parallel stereo configuration.....	61
Figure 3.7: The typical 3-D calibration object.....	63
Figure 3.8: The ideal and actual images of grid pattern	67
Figure 3.9: Images of 3-D calibration object	77
Figure 3.10: Images of planar calibration object	81

List of Figures	x
Figure 3.11: Corrected image	82
Figure 3.12: Stereo images of planar calibration object	88
Figure 3.13: Corrected stereo images	90
Figure 4.1: Parallel stereo image configuration	96
Figure 4.2: STD method for initial disparity detection.....	99
Figure 4.3: Feature vectors of left image and right image.....	100
Figure 4.4: Correspondence search.....	102
Figure 4.5: Sub-pixel disparity	106
Figure 4.6: Process of searching range adjustment.....	109
Figure 4.7: Stereo image pair with regular grid for correspondence search.....	110
Figure 4.8: The main procedure of 3-D reconstruction	113
Figure 4.9: Optical model of the eye and stereo fundus camera.....	115
Figure 4.10: Maximal cup depth.....	117
Figure 4.11: An example of color stereo fundus image pair.....	121
Figure 4.12: An example of combined matching method on sparse depth mesh	122
Figure 4.13: An example of multiple window process based on combined matching	123
Figure 4.14: Dense depths by two different operations computing on sparse depths.....	124
Figure 4.15: An example of 3-D optic disk reconstruction	125
Figure 4.16: One more example of 3-D optic disk reconstruction	126
Figure 4.17: Combination of original and recovered right images.....	127
Figure 4.18: An example of nonuniform illumination between the left and right images	130
Figure 5.1: Flowchart of optic disk boundary detection.....	135
Figure 5.2: An example of distance transformation.....	137
Figure 5.3: Deforming range and clustering sample.....	139
Figure 5.4: Normalized clustering samples	146
Figure 5.5: Two clusters	149
Figure 5.6: Clustered contour points.....	150
Figure 5.7: Smoothing weights based on different distributions	152
Figure 5.8: A successful and fair example of the three methods	158

List of Figures	xi
Figure 5.9: A successful example of all three methods	159
Figure 5.10: An example of elliptic optic disk	160
Figure 5.11: An example of ill-defined optic disk boundary	161
Figure 5.12: A failure example due to lesion disturbance on disk edge	162
Figure 5.13: Graph of the performance comparison among different methods	163
Figure 6.1: Cup boundary estimation based on pallor	168
Figure 6.2: Cup boundary estimation based on bending of small blood vessels	168
Figure 6.3: Flowchart of disk parameter estimation	169
Figure 6.4: One definition of cup boundary based on retinal surface	170
Figure 6.5: Another definition of cup boundary based on disk edge	170
Figure 6.6: Initial cup boundary	171
Figure 6.7: Alignment of shape model	174
Figure 6.8: Flowchart of cup boundary detection	176
Figure 6.9: Parameter setting of energy function	177
Figure 6.10: Definition of optic disk region	179
Figure 6.11: The optic disk coordinate system	179
Figure 6.12: Geometric model of optic disk with cross sectional profile and projected image	180
Figure 6.13: An example of optic disk image with obvious pallor	183
Figure 6.14: Another example of non-obvious pallor	185
Figure 6.15: An example of disk parameter estimation	190
Figure 6.16: Another example of disk parameter estimation	191
Figure A.1: Image rectification	216

List of Tables

Table 2.1: Comparison of local and global matching methods	33
Table 3.1: Comparison results of different camera model based on 3-D calibration object	78
Table 3.2: Camera calibration results based on 3-D calibration object with lens distortion considered	78
Table 3.3: Comparison results of different camera models based on planar calibration object.....	81
Table 3.4: Camera calibration results based on planar calibration object with lens distortion considered.....	82
Table 3.5: Camera calibration results at location of 190mm with lens distortion considered	83
Table 3.6: Comparison of two camera calibration methods	85
Table 3.7: Stereo camera calibration results of left viewpoint	88
Table 3.8: Stereo camera calibration results of right viewpoint	88
Table 4.1: Maximal cup depths from HRT and 3-D virtual image.....	118
Table 4.2: Evaluating correlations of 100 testing image pairs.....	129
Table 5.1: Comparison of MDCP among different methods.....	163
Table 5.2: Results of variable weight ratio of the proposed method	164
Table 6.1: C/D vertical ratio of two examples.....	186
Table 6.2: C/D vertical ratio.	189
Table 6.3: Statistical comparison of C/D vertical ratio.....	190
Table 6.4: Disk parameters for glaucoma analysis	190
Table 6.5: Another example of disk parameters in comparison with HRT results.....	191

Chapter 1

Introduction

Computer-aided medical image processing and analysis for clinical diagnosis is one of the research areas currently drawing intense interests from scientists. The availability of such a facility will help physicians to identify and measure important features in an image, compare sequential images, aggregate images similar in content and finally obtain automated diagnosis from images.

Ophthalmology is considered second only to radiology in the application of imaging in the medical field [1]. Color fundus image which is very common and quite useful to ophthalmologists is specifically studied in this research. A color fundus image can provide a large amount of information about pathological changes caused by some eye diseases such as glaucoma, cataract, which may lead to loss of vision, and it can also indicate early signs of some systemic diseases, such as diabetes or hypertension. It is important to detect harmful changes and abnormal 3-D shape of the retina as early as possible and monitor their progression during clinical therapy, to prevent major visual field loss and to extend the efficient period of usable vision.

The features in color fundus image can be simply classified into normal features and abnormal features. Normal features include optic disk with bright-white region inside called pallor, blood vessels, and fovea at the center of macular as shown in Figure 1.1(a). Abnormal features include exudates, hemorrhages, etc., which are caused by eye diseases as shown in Figure 1.1(b). Extraction of both normal and abnormal features is an important step in automated analysis which is normally based on two-dimensional (2-D) fundus images. On the other hand, three-dimensional (3-D) visualization of the retina is relatively more difficult to obtain but also very useful for ophthalmologists to understand and evaluate the condition of the retina. The 3-D presentation of the retina not only provides clinicians a clear and direct view of retinal structure, but also enables significant improvements in accuracy of retinal feature extraction and evaluation.

Figure 1.1: Features on fundus images

In this thesis, new approaches for 3-D optic disk reconstruction, cup/disk boundary detection and disk parameter measurement are presented, which could be applied in the mass screening and diagnosis of retinal diseases, especially for glaucoma.

1.1 Motivation

Color retinal photography is one of the first and fundamental steps in order to document, analyze and diagnose eye and systemic diseases in clinics. Optic disk, fovea and blood vessels are the basic components on retina. Any changes in the shape, color or depth of the optic disk is an indicator of various ophthalmic pathologies [2], especially for glaucoma, which is one of the most common causes of the blindness. Glaucoma affects approximately 70 million people all over the world. It accounts for 20% of blindness in Singapore [3]. Over 90% of cases of glaucoma are of a variety that may have no signs or symptoms because peripheral vision can be lost before a person's central vision is affected. Unfortunately, nearly 50% of people with glaucoma may go undetected until they have lost substantial vision. Many people are unaware they are losing vision until less than 20% of their visual field remains [4]. The damage of the retina and the loss of the vision due to this disease may not be reversible. However, most loss of vision from glaucoma is preventable if the disease is treated early enough. Hence, early detection of glaucoma is especially significant to make the timely treatment and to control the state of the illness.

Glaucoma is a disease caused by increased intraocular pressure (IOP) resulting either from a malformation or malfunction of the eye's drainage structures. Left untreated, an elevated IOP causes irreversible damage of the optic nerve and retinal fibers resulting in a progressive, permanent loss of vision [5]. There are three important factors in diagnosis of glaucoma: the intraocular pressure, the visual field and the condition of the optic nerve

head (optic disk). However, glaucoma can also occur with normal eye pressure in many cases; hence the diagnosis cannot only rely on IOP. Checking of optic disk is more important. The 3-D shape of the optic disk directly reflects the cupping progression of the optic disk. Moreover, the 3-D representation makes the measurement of disk parameters more accurate and easier. Those parameters are used to evaluate the condition of the optic disk for diagnosis. For example, cup-to-disk (C/D) ratio is one useful disk parameter, in which the definition of cup is based on the 3-D depth. According to Masunori *et al.* [6], the cup is defined at approximately 150 μ m down from the top edge of the disk boundary. Another definition locates cup at 50 μ m down from the retinal surface [7]. The 3-D structure of optic disk, the disk and cup boundaries, as well as the C/D ratio, etc., are the common parameters used for analysis by clinicians. Therefore, the fundus image processing for 3-D disk shape recovery and disk parameter estimation is an assisting tool for early glaucoma diagnosis as well as a follow-up tool to monitor the progression of the disease.

Figure 1.2: Stereo viewer

Relatively less work however has been done on the computer generated 3-D visualization of the retinal surface. One major reason is probably that the small retinal area of interest does not easily lend itself to an accurate and automated 3-D reconstruction.

Practically, the ophthalmologists simply view 2-D images and manually draw the cup and disk boundaries on the images based on their experience. Without a clear and correct impression of the 3-D visualization of the optic disk, the accuracy of resulting measurements and derived evaluations will be reduced. Generally, ophthalmoscope is widely used on the patient's eye for retinal observation to get the impression of 3-D optic disk shape. Some clinicians may also use a stereoscope to obtain 3-D visualization from a pair of stereo fundus images as shown in Figure 1.2. However, only blurred and small 3-D optic disk can be observed in both of these two methods. Moreover, the observation depends on the experience and skill of each clinician. An effective instrument is Optical Coherence Tomography (OCT), which is used to obtain cross-sectional images of the retina. It provides clinicians the impression of 3-D retinal shape at certain cross section. Nevertheless, the full 3-D visualization and true color (natural color) information cannot be obtained from OCT. The advanced alternative is to use the uncommon medical instruments such as Heidelberg Retina Tomograph (HRT), which can provide topographic image of optic disk based on laser scanning technique. The disk and cup boundaries are placed manually or semi-automatically by the clinicians using the HRT system. The limitation of this instrument is that the obtained 3-D image can be only assigned pseudo-colors based on depths, which make the measurements of cup and disk boundaries by the human expert less reliable. With natural color (true color), the optic disk and cup can be more easily segmented by the ophthalmologists. The 2-D color fundus image is still referred to by most clinicians in estimating the disk and cup boundaries from HRT's 3-D image. Moreover, the patient's cooperation is necessary in HRT and OCT imaging process. It takes 1.6 seconds to acquire the HRT image and 1.92 seconds to acquire the OCT image. If the patient cannot maintain eye fixation during this 1.6 or 1.92 seconds time period, which may happen for the old patients, it may be difficult to effectively

obtain the retinal image with good quality. On the contrary, turning on the flash light of the fundus camera takes only 2 milli-seconds.

Computer-generated color 3-D visualization of the optic disk is a valuable task still being investigated. The extended work of automated detection of the disk and cup boundaries could avoid a costly workload of redundant measurements and enable the clinician to see more patients in a given time. In addition, the computerized quantitative measurements of the disk parameters can also reduce the variation of manual measurements derived by different clinicians. The motivation of this thesis can be concluded as follows:

- The 3-D shape of the optic disk and its relative parameters are important information for clinicians to analyze eye diseases, especially glaucoma.
- The existing instruments cannot provide true color 3-D visualization of the optic disk together with the automated measurements of disk parameters.
- Various fundus cameras make it possible to recover the color 3-D representation of the optic disk and quantitatively evaluate it to help clinicians in analysis, diagnosis, and monitoring.

1.2 Objectives

The objective of this research work is to explore new approaches to help develop an automated glaucoma analyzing system as an aiding tool for clinical diagnosis, which is also useful in the mass screening. The system, including 3-D optic disk reconstruction, automated disk and cup boundary detection and disk parameter estimation, is based on

common color fundus images obtained by a stereo color fundus camera. The goal in each module of the system can be summarized as follows.

- One goal is to reconstruct true color 3-D optic disk. The challenge here is that the depth variation of the retinal area of interest is quite small (generally several hundred microns). Moreover, the unknown and complex inner structure of stereo fundus camera and the optical effect of biological layers within the eye are both the difficulties in the 3-D optic disk reconstruction.
- Automated detection of the optic disk boundary is another task. There are several sources of difficulties in disk boundary detection, such as occlusions of blood vessels, ill-defined edges, noises, fuzzy shapes, etc.
- Estimation of cup boundary and its related parameters is the extended task of 3-D optic disk reconstruction, which provides useful clinical parameters for analysis and diagnosis.
- Proposing the evaluation methods to verify the computerized results and comparing the results with those from other available approaches or other instruments are also the necessary assignments.

1.3 Major Contributions of the Thesis

Novel approaches are proposed in each module of this research work to achieve above-mentioned objectives and solve the difficulties pointed out. The major contributions are described as follows:

- An automated glaucoma analyzing system based on color stereo fundus images is presented in this research work as shown in Figure 1.3, which includes 3-D optic

disk reconstruction, optic disk boundary detection, cup boundary estimation, and disk parameter measurement and analysis.

- Two typical camera calibration techniques are implemented and applied on the monocular fundus camera for comparison and reported as the first time. The more suitable method is chosen in this research. Then the simultaneous stereo fundus camera is investigated as the first time to approximately estimate its focal length for use in the 3-D reconstruction.
- An automated, precise, 3-D optic disk reconstruction method from a pair of stereo images is developed. A combination of two constraint-based matching methods is introduced to detect the precise sub-pixel correspondences and prune the error matching. The optical effect within the eye is calibrated and integrated in the reconstruction process.
- A modified active contour model (modified snake) is proposed for fully automated detection of the optic disk boundary, which directly solves the problem of blood vessel occlusions on the disk boundary. The original snake technique is extended and modified in two aspects: clustering and updating after each radial deformation. The modifications make the proposed algorithm more accurate and effective to blood vessel occlusions, noises, weak edges and fuzzy contour shapes.
- The modified snake is extended to estimate the cup boundary, which is a relatively new task in fundus image processing. A disk coordinate system is established to measure the cup and disk parameters for a better description of the optic disk condition for glaucoma analysis and monitoring.

Figure 1.3: Glaucoma analyzing system

1.4 Organization of the Thesis

This thesis is organized according to the respective tasks of glaucoma analysis. The whole thesis consists of seven chapters. Introductory aspects of the research work, motivation and objectives are included in Chapter 1. The remainder of this thesis is structured as stated below:

Chapter 2 provides an overview of the related projects and methods. The medical background of eyeball structure, eye diseases and fundus cameras are introduced. Related projects such as optic disk identification, blood vessel detection and tracking, etc., are presented. Relevant methods on camera calibration, 3-D reconstruction, and boundary detection, etc., are surveyed in this chapter.

Chapter 3 addresses the camera calibration task. Two camera calibration methods based on different calibration object patterns are used on monocular fundus camera and the results are compared and analyzed. The advantage and disadvantage of these two methods are shown. Following this, the stereo fundus camera is later studied.

Chapter 4 performs the 3-D optic disk reconstruction. A simple and fast standard deviation (STD) method is introduced to measure the initial disparity. A sub-pixel combined matching method is proposed to reduce the error matching and detect more precise correspondences based on the initial disparity. Multiple windowing and interpolation are applied to recover the 3-D dense depths of the optic disk. The results are quantitatively evaluated by the proposed correlation method. The optical effect inside the eye is considered, calibrated and integrated in the 3-D reconstruction process to obtain true depths.

Chapter 5 investigates the optic disk boundary detection. A novel deformable model method is employed for fully automated detection of the optic disk boundary. The blood vessel, ill-defined edge and noise problems are effectively solved by modifications on the snake, such as, radial deformation, clustering, and smoothing update. The computerized results are compared with the reference boundaries manually identified under the supervision of an ophthalmologist to evaluate the accuracy of the proposed method.

Chapter 6 presents the cup boundary detection based on the 3-D image obtained from chapter 4 and disk boundary detected from chapter 5. The disk coordinate system is established to measure the disk parameters for analysis. This chapter concludes with the extracted results from the clinical validation.

Chapter 7 summarizes the main results delivered in this research and outlines the contributions and the future work in this field.

Chapter 2

Background and Literature Review

Many eye-relative diseases will lead to the evident fundus changes that can be observed from the color fundus image. Color fundus image processing, not only in two dimensions but also in three dimensions, can help ophthalmologists clearly observe, analyze and diagnose the diseases. It can also be used in mass screening. In this chapter, the medical background about the eye structure, eye diseases and ophthalmic instruments is first introduced. Then the related work, i.e., camera calibration, 3-D reconstruction and boundary detection will be surveyed. The strengths and weaknesses of current techniques to the fundus image processing will also be discussed in this chapter.

2.1 Eye Structure

The eye is considered to be one of the most complex and essential organs of the body. It allows us to see and perceive the environment around us. Figure 2.1 shows a cross section of human eye.

Figure 2.1: The structure of eyeball (from [8])

The eyeball is nearly a sphere, with the diameter of approximately 24mm for the adult. The main components on the anterior portion of the eye are cornea, iris, pupil, lens, etc. Pupil is the opening in the iris through which light enters the eye. The retina is a multi-layered sensory tissue on the posterior portion of the eye. It contains millions of photoreceptors that capture light rays and convert them into electrical impulses. These impulses travel along the optic nerve to the brain where they are turned into images [5]. The optic disk (or called optic nerve head, papilla) is the region on the retina where nerve fibers converge to become part of the optic nerves. Optic disk is also called the blind spot, because there is no photoreceptor existing here. The inner portion of the eye, including optic disk, fovea and retina is called ocular fundus. The function of the eye can be compared to the camera. Light enters the eye, passes through cornea, pupil, lens, aqueous humor, vitreous humor, and finally forms the image on the retina. Both the lens and the cornea are the transparent structures that play a role in bending and focusing light rays.

2.2 Related Eye Diseases

Many eye diseases may cause vision damage or even blindness, in which case the most common diseases are: glaucoma, diabetic retinopathy and cataract. Some of these will be introduced in this section. The pathological changes of glaucoma and diabetic retinopathy can be observed from fundus images.

2.2.1 Glaucoma

Glaucoma is a disease that damages the optic nerve and is one of the most common causes of blindness. It is a group of eye diseases in which the optic nerve at the back of the eye is slowly destroyed. In most people this damage is due to an increased pressure inside the eye, which is the result of blockage of the circulation of aqueous, or its drainage [9]. The eye constantly produces the aqueous which is the clear fluid filling the anterior chamber (the space between the iris and cornea). The fluid is pumped into the eye from the bloodstream carrying oxygen, sugars and other essential nutrients. Circulating around the structures inside the eye, the aqueous is then drained out of the anterior chamber, through a meshwork back to the blood to be renewed. The delicate balance between the production and drainage of aqueous determines the eye's intraocular pressure (IOP). If the circulation or drainage is blocked, the IOP will increase, which directly damages the optic nerve. The elevation of IOP cups (pushes out) the optic disk and changes its 3-D shape as shown in Figure 2.2, which can be observed from the fundus image under the assisting of some instruments. Most people's IOPs fall between 8 and 21. However, some eyes can tolerate higher pressures than others. That's why it may be normal for one person to have a higher pressure than another. For few of glaucoma patients, the damage is due to the causes of poor blood supply to the vital optic nerve fibers, a weakness in the structure of

the nerve, and/or a problem in the health of the nerve fibers themselves [9]. The damage of optic nerve by glaucoma leads to irreversible loss of vision, which makes early detection and treatment significantly important.

Figure 2.2: Cross section of optic disk for normal eye and glaucoma eye (from [8])

Figure 2.3: Glaucoma (from [8])

2.2.1.1 Glaucoma Type

There are two major classes of glaucoma: open-angle glaucoma and angle-closure glaucoma, as shown in Figure 2.3. Each kind of glaucoma can be subsequently classified into primary or secondary type.

Primary open-angle glaucoma

Primary open-angle glaucoma is the most common type of glaucoma, which accounts for 80% to 85% of all glaucoma cases. Its exact cause is unknown but is believed to be linked to an obstruction of the drainage system from the trabecular meshwork. Trabecular meshwork is a small sieve-like drain in the angle of the eye as shown in Figure 2.3, through which fluid must pass to return the blood stream. With this type of glaucoma, even though the anterior structures of the eye appear normal, fluid builds within the anterior chamber, causing the IOP to become elevated gradually [8]. People with this type of glaucoma often do not show any symptoms such as pain, high IOP, etc. However, the optic nerves are gradually damaged; and peripheral vision loss may keep several months or even years. Hence, glaucoma is known as the silent thief of vision.

Primary angle-closure glaucoma

The angle-closure glaucoma accounts for approximately 10% of all glaucoma cases. This type of glaucoma has obvious symptoms. It occurs when the iris moves front and makes the angle between the iris and the cornea narrow or even close. The drainage system is blocked because of this structural abnormality in the anterior chamber, which makes the IOP rapidly rise. This causes severe pain in the eye and face, as well as other symptoms, such as nausea, vomiting, colored halos around lights, and loss of vision [8]. An attack of angle-closure glaucoma is often severe. Immediate medical help should be

sought. If treatment is delayed there can be permanent visual damage in a very short time [9].

Secondary glaucoma

Secondary glaucoma for both open-angle and angle-closure types occurs as a result of another disease or problem within the eye such as: inflammation, trauma, previous surgery, diabetes, cataract, tumor, and certain medications. For this type, both the glaucoma and the underlying problem must be treated [5].

2.2.1.2 Glaucoma Evaluation

There are several components in glaucoma evaluation. In addition to measuring the intraocular pressure, the ophthalmologists will also evaluate the condition of the optic disk, test the peripheral vision, and examine the structure of anterior portion of the eye with a special lens (gonioscopy) before making a diagnosis.

Figure 2.4: Cross section of optic disk (adapted from [7])

The characteristic of glaucoma is cupping of optic disk. Cup is the depressed area in the center of the optic disk as shown in Figure 2.4. Rim is the region inside disk and outside cup. Many disk parameters can be used to evaluate the condition of optic disk such as cup-to-disk (C/D) ratio, rim-to-disk (R/D) area ratio, disk area, cup area, cup

volume, etc. However some of them are not easy to estimate without the aid of computer, thus the physicians generally evaluate the condition of optic disk only by noting the C/D vertical ratio. This is simply a comparison of the cup to the entire vertical diameter of the optic disk. The area of cupping, or depression, increases in glaucoma progress. Therefore, a patient with a higher ratio has more damage [5]. The C/D ratio is less than 0.3 for most of normal eyes. However, around 5% people with normal eyes have C/D ratios larger than 0.6. Hence, the difference of C/D ratio between the left eye and the right eye is another parameter for evaluation. Figure 2.5 gives the C/D vertical ratio in different stages of glaucoma. Currently, the R/D area ratio is preferred by clinicians because of its stability. However this parameter is difficult to be roughly estimated by merely observing the 2-D fundus image without the aid of computer.

Figure 2.5: Cup-to-disk ratio (from [5])

2.2.2 Diabetic Retinopathy

Patients with diabetes are more likely to develop eye problems. The effect of diabetes on the eye is called diabetic retinopathy, which occurs when diabetes damages the tiny blood vessels that provide oxygen to the retina. The earliest phase of the disease is known as background diabetic retinopathy. In this phase, the arteries in the retina become weakened and leak, forming small, dot-like hemorrhages. These leaking vessels often lead to swelling or edema in the retina and decreased vision. The next stage is known as

proliferative diabetic retinopathy. In this stage, circulation problems cause areas of the retina to become oxygen-deprived or ischemic. New, fragile, vessels develop as the circulatory system that attempts to maintain adequate oxygen levels within the retina. This is called neovascularization. Unfortunately, these delicate vessels hemorrhage easily. Blood may leak into the retina and vitreous, causing spots or floaters, along with decreased vision [5].

Diabetic retinopathy is a system disease causing of blindness. The abnormal features of this disease are exudates, hemorrhages, microaneurysms, etc., which can be observed on the color 2-D fundus image. Some researchers focus their work on automated extraction of these features to help clinicians in early disease detection and analysis.

2.2.3 Cataract

A cataract is a clouding of the normally clear lens of the eye. The lens is contained in a sealed bag or capsule. As old cells die they become trapped within the capsule. Over time, the cells accumulate causing the lens to cloud, making images look blurred or fuzzy. About three-quarters of cataracts are a natural result of aging [5]. The other causes of cataract are eye injuries, certain medications, and diseases such as diabetes and alcoholism, etc. Cataract is the leading cause of visual loss among adults 55 and older, and surgery is the only treatment to remove the cataract.

2.3 Related Ophthalmic Instruments

The optical instrument used to document the visual condition of the retina is called fundus camera. It is a specialized low-power microscope with an attached camera back [10]. Traditional fundus camera records the image on the film. Currently, with the large

requirement of electronic imaging, a number of fundus cameras are upgraded. Some of them can directly attach the digital camera back. Others use video camera back to record the analog image; it is then digitized by frame grabber and transferred to computer. Many different manufacturers make a wide variety of fundus cameras available.

2.3.1 Monocular Fundus Camera

The ordinary fundus camera is monocular fundus camera, from which one image is obtained in each exposure. Fundus cameras are described by the angle of view as shown in Figure 2.6. The angle of 30° is considered as the normal angle of view, which creates a film image 2.5 times larger than life. Wide-angle fundus camera captures image between 45° to 140° , and provides less image magnification. A camera with the angle of view of 20° or less is referred to as narrow-angle fundus camera.

Figure 2.6: Angle of view

Both mydriatic and non-mydriatic fundus cameras are available. Mydriatic fundus camera requires pupil dilation. This versatility allows both sequential stereo fundus photography and fluorescein angiography. Using this type of fundus camera, the pseudo stereo fundus images can be obtained by first taking one image and then slightly shifting

the camera left or right to take another image. Non-mydriatic fundus camera is generally used as a screening device because dilation is not required. It exploits the patient's natural dilation in a dark room. Only a single image is captured each time because it will take some time to let the patient's pupil naturally dilate again.

Figure 2.7: Non-mydriatic fundus camera (from [11])

The Canon non-mydriatic fundus camera CR6-45NM is studied in this thesis. The Canon EOS-D30 digital camera back is attached to the fundus camera to capture the high resolution digital image as shown in Figure 2.7. The inner structure of fundus camera is shown in Figure 2.8. The rays indicated by solid black lines are the imaging rays focusing on the film or imaging sensor. The rays indicated by gray lines and reflected by the mirror are the illumination rays, which are used to illuminate the object's eye. Two lamps are in the illumination unit; one is for observation, the other is for photography. The camera can automatically compensate eye refraction and delivers optimal retinal imaging. The image taken from this camera is given in Figure 2.9(a).

Figure 2.8: Inner structure of fundus camera

(a) Taking from monocular fundus camera

(b) Taking from stereo fundus camera

Figure 2.9: Fundus image

2.3.2 Simultaneous Stereo Fundus Camera

Stereo fundus camera uses only one exposure to place two images side by side on a single film simultaneously. The prototype of stereo fundus camera was designed by David Donaldson and Harold Edgerton (Massachusetts Institute of Technology) in 1964. Their fundus camera used two prisms, which split the light from a single front lens element and directed it to two different frames of film, as shown in Figure 2.10. The horizontal distance between two optical systems is the stereo base (or baseline). Variations of this design are used on current stereo fundus camera [10].

Simultaneous stereo fundus camera has obvious advantages. For example, patient cooperation is not needed while taking a pair of photographs compared with the sequential (pseudo) stereo fundus photography, and the stereo fundus camera provides the physician with stereo images of constant baseline. Its standardized baseline allows longitudinal studies of retinal topography as well as detailed electronic analysis of retinal topography [10]. Nidek 3-Dx stereo fundus camera is used in this research work, which provides color stereo fundus images as shown in Figure 2.9(b).

Figure 2.10: Inner structure of simultaneous stereo fundus camera (adapted from Saine's [10])

2.3.3 Heidelberg Retina Tomograph (HRT)

Figure 2.11: Inner structure of HRT (adapted from HRT tutorial [7])

The Heidelberg Retina Tomograph (HRT) is a confocal scanning laser system (cSLO) designed for acquisition and analysis of three-dimensional topography images of retina. A scanning laser system (SLO) operates by focusing a laser to one spot on the retina. As this laser beam is scanned across the retina, the reflected light goes the same way back through the optics, is separated from the incident laser beam, and recorded on a light detector as shown in Figure 2.11. In order to produce a two-dimensional (2-D) image, the illuminating laser beam is deflected periodically in two dimensions perpendicular to the optical axis using scanning mirrors. Therefore, the object eye is scanned point by point sequentially in two dimensions to create a laser image at a scan rate of 256×256 spots [7].

Figure 2.12: Confocal scanning laser system (adapted from HRT tutorial [7])

In the confocal scanning laser system of HRT, a small diaphragm is placed in front of the detector at a location which is optically conjugated to the focal plane, as shown in Figure 2.11, where the laser beam focusing on the focal plane should also focus on the diaphragm. The diaphragm is like a confocal pinhole. Only the reflected light which is focused to the pinhole can pass it and is detected, whereas the light reflected from layers of the three-dimensional object above or below the focal plane cannot focus to the pinhole. It is highly suppressed and only a small fraction of it can pass the pinhole and is detected. Hence, a 2-D laser image acquired at certain focal plane only carries the information of the object layer located at or near the focal plane [7]. When the focal plane is moved at

different depth locations, a series of optical section images (thirty-two 2-D laser images) are obtained, which form a layer-by-layer 3-D image of retina as shown in Figure 2.12.

The topography image is computed from layered 3-D image. For each spot (x, y), the series of 32 section image planes contains the distribution of the reflected light intensity along the optical axis, the z-axis. It is a symmetric distribution with a maximum at the location of the light reflecting surface. Therefore, the depth for each spot (x, y) is recovered by determining the position of the profile maximum. The recovered 3-D topography image from HRT is colorless. In order to visualize it more vividly, the pseudo colors are arranged on the image by translating each specific height into a specific color according to a color scale with dark colors representing prominent structures and light colors representing depressed structures.

The deficiencies of HRT are (1) HRT only provides 3-D tomography image with pseudo-color, which make the measurements of cup and disk boundaries less reliable. With natural color (true color), the optic disk and cup can be more easily segmented by the clinicians (2) It takes 1.6 seconds for HRT to acquire the 3-D image, and the patient's cooperation is important in imaging. However, maintaining eye fixation is difficult especially for older patients (3) HRT is more expensive for the patient. On the contrary, stereo fundus camera is much cheaper and so on the whole it is more suitable for mass screening.

2.3.4 Optical Coherence Tomography (OCT)

Optical Coherence Tomography (OCT) is a diagnostic imaging technique that uses near infrared light to produce high resolution cross-sectional images of the retina. OCT is similar to ultrasound B-scan imaging except that light rather than sound waves is used in

order to obtain a much higher longitudinal resolution of approximately $10\mu\text{m}$ in the retina. OCT generally has been shown to be clinically useful for imaging selected macular diseases including macular holes, macular edema, age-related macular degeneration, retinal inflammatory diseases, etc. In addition, OCT has the capability of measuring the retinal nerve fiber layer thickness in glaucoma and other diseases of the optic nerve [4]. One OCT image of normal optic disk is given in Figure 2.13.

The deficiencies of OCT are (1) OCT only provides cross sectional image, (2) It takes 1.92 seconds for OCT to acquire the image, the patient's cooperation is important in imaging. (3) OCT is more expensive for the patient, which is not quite suitable for mass screening.

Figure 2.13: OCT image of optic disk (from [12])

2.4 Review of 3-D Reconstruction

The objective of this research is development of a computer-based tool for the 3-D analysis of glaucoma symptoms. 3-D optic disk shape is first reconstructed from stereo image. The optic disk boundary is then detected on 2-D image. Finally, the cup boundary is estimated based on the information of both 3-D shape and the detected disk boundary; the disk parameters are measured for clinical analysis. The review of 3-D reconstruction and boundary detection are given in this section and next section respectively.

In the domain of 3-D scene reconstruction, the major techniques are passive stereovision, active stereovision, and shape from X, etc.

- Passive stereovision uses the geometric physical models of two or more cameras from different viewpoints in order to compute the physical depth. These models determine the position of a point in the 3-D scene using its projections in the images and the description of the cameras [13].
- Active stereovision is a technique of projecting the light patterns onto a scene while viewing from a different viewpoint by camera, which is similar to passive stereovision except that one camera is replaced by the light source (projector).
- Shape from X is a generic name for techniques that aim to extract 3-D shape from intensity images, where X is one of a number of options. This family of techniques includes methods such as shape from shading, shape from motion, shape from texture, shape from focus/defocus, shape from zoom, shape from contours, and etc [65][168]. Shape from shading (SFS) uses smooth variations in the brightness or shading of objects in an image to estimate the shape of depicted objects. To solve the SFS problem, it is important to study how the images are formed. Knowledge of photometric physical models and light source direction is required in order to calculate the depth [13]. Other shape from X techniques estimate the 3-D shape from motion in image sequence, image texture, images in focus change, images in zoom change, contours in one or more view directions, respectively.

Based on the structure of the fundus camera and physical characteristics of the eyeball, passive stereovision is selected in this research for retinal surface reconstruction, in which the basic theory is the epipolar geometry. The prime problems to be solved in 3-D reconstruction are camera calibration, correspondence search, and depth recovery.

2.4.1 Camera Calibration

Camera Calibration is the first and necessary step in 3-D reconstruction in order to extract the intrinsic and extrinsic parameters of the camera, such as focal length, pixel size and distortion, etc. It can also be described as the process of determining the relationship between what appears on the image plane and where it is located in the 3-D world. Much work has been done on camera calibration in the last two decades. However fundus camera calibration is very little reported. This thesis is probably among the only few reported in the area. The techniques can be roughly classified into two categories: self-calibration [14]-[28] and photogrammetric calibration [29]-[39].

Self-calibration

Self-calibration technique does not use any calibration object, but uses image information alone. It is first introduced in the field of computer vision by Faugeras and Maybank [14][15] for the study of motion. If images are taken by the same camera with fixed internal parameters, correspondences between three images (at least) are sufficient to recover both the internal and external parameters. Although the calibration objects are not necessary, a complicated task, i.e., the extraction of precise correspondences between each pair of images, is required. Moreover, large number of parameters needs to be estimated, resulting in a much complex mathematical problem. The lens distortion of the camera is usually not considered in self-calibration technique.

Self-calibration technique is not very suitable for the fundus camera calibration for the following reasons. One problem for this approach is that it is difficult to take three fundus images with fixed camera parameters. In fundus image capturing, unlike the general static scene, the object's eye cannot be kept exactly fixing in the series of camera shots. The fundus camera must be slightly adjusted for focusing in each capturing.

Besides, it is inconvenient and inefficient for both clinicians and patients to undergo many rounds of fundus image capturing. Although camera self-calibration technique does not require a calibration object with a known 3-D shape, it is not quite suitable for this research case, because various conditions of self-calibration cannot be well satisfied.

Photogrammetric calibration

Photogrammetric calibration is performed by observing a calibration object whose geometry in 3-D space is known with good precision. The intrinsic and extrinsic parameters of the camera are efficiently estimated from the information of 3-D object and its corresponding information on the 2-D image. According to the dimension of the calibration objects, the techniques are classified into 3-D, 2-D or even 1-D object-based calibration. Those techniques can also be subdivided into direct method and indirect method [168]. The direct method recovers directly the intrinsic and extrinsic camera parameters; the indirect method estimates the projection matrix first without solving explicitly camera parameters, which are then computed from projection matrix.

Sturm and Maybank [30] presented a camera calibration algorithm based on 3-D object (3-D object-based calibration). The calibration object was designed as a combination of three planes orthogonal to each other. Only one image was needed. However, the image of each plane may not totally contain the reliable information of camera intrinsic parameters because it is located at only a part of the whole image. The lens distortion was not considered in their method. Heikkila and Silven [31] described a four-step camera calibration approach based on 3-D calibration object, in which lens distortion was estimated.

Deguchi and Morishita [32] presented a linear camera calibration method using two parallel planes (2-D object-based calibration). The main idea of this technique is to

determine the spatial lines-of-sight corresponding to every image points, instead of directly identifying the intrinsic parameters. The two calibration planes must be exactly parallel. This setup may be realized by firstly setting the calibration plane at the first position and taking its first image, then, shifting it to the second position keeping it parallel to the plane at the first position, and taking its second image. Or, it may be effective to prepare one calibration object having two parallel plane surfaces which must be made of special transparent materials. However, it is an uneconomical method, and also, it is difficult to guarantee the shifting is exactly parallel. Because the planar object undergoing a parallel translation, the two-plane calibration object in this technique can be considered as a variation of 3-D object.

Zhang [33] introduced a technique to calibrate a camera based on planar calibration object (2-D object-based calibration). Three images of the plane object are needed in his method. Either the camera or the planar pattern can be freely moved in each image capturing in which the motion need not be known. The lens distortion was also considered by Zhang to obtain more precise results. Compared with the two-plane calibration method, this method is easy to use.

Currently, 1-D calibration object was employed in the camera calibration technique by Zhang [29] (1-D object-based calibration). In his investigation, the camera calibration was not possible with free-moving of 1-D object. It can be solved if one point of the object is fixed. Six images were required in this method. The lens distortion was not considered in this method.

Different models of lens distortion were introduced separately or in conjunction with basic intrinsic parameters of the camera in [34]-[39]. Generally, the authors optimized the distortion parameters using the property that the straight line of the object will be distorted

to a curve line in the image. The lens distortion model and the parameter optimization method were the main points to be investigated by them.

Photogrammetric calibration method is selected in this research work. The problem of these methods is that if the camera is adjusted after calibration, the focal length and other intrinsic parameters may be different from the ones measured in camera calibration operation. Fortunately, based on the structure of the fundus camera, the intrinsic parameters have minor variations in focusing; the major adjustment is shifting the position of the camera. Hence the detected intrinsic parameters can be used as the approximate values. Two photogrammetric calibration methods based on 3-D calibration object pattern and planar calibration object pattern will be presented in the following. The results from two methods will be compared and analyzed.

2.4.2 Correspondence Search and 3-D Reconstruction

3-D reconstruction specified in the optic disk region is more valuable than the recovery of the whole retinal surface. Little amount of information on 3-D retinal surface reconstruction is provided in [40]-[46]. Most of them used parallel stereo configuration in which the depth is calculated from correspondences by triangular computation and the correspondence search becomes the most important task. Hence, the literature on correspondence search and 3-D retinal surface reconstruction is reviewed together in the sequel.

Many researchers in this area focused their work on correspondence search. Brown *et al.* [47], Scharstein and Szelisky [48] reviewed the current approaches of stereo matching. They classified the stereo matching methods into local method [41][42][49] and global method [50]-[54]. The process of correspondence search can be seen as having two

distinct parts: the similarity measure and the matching search. The similarity can be measured based on correlation, absolute difference, rank transform, phase, optical flow, features, mutual information, etc. In local methods, the disparity computation at a given point depends only on similarity, where the individual correspondence is located at the point with largest similarity. Global methods estimate the disparity by minimizing a global cost function that combines both data and smoothness terms based on similarities, from which the available algorithms are based on dynamic programming, intrinsic curves, graph cuts, etc. Local methods are very efficient, but they are sensitive to locally ambiguous regions in images (e.g., occlusion regions or regions with uniform texture). Global methods are less sensitive to these problems but require much more computational expense. Moreover, since it is global optimization, the local errors may be propagated to corrupt other potentially good matching.

Some other methods are the variations of local method or global method, such as constraint-based method [40][45][55][56] and iterative method [57]–[61], which improve the matching accuracy. Constraint-based method is the variation of local method, which uses not only the similarities but also the constraints or the information from surrounding matching points to obtain correct correspondences. Iterative method is the variation of global method that does not explicitly state a global minimizing function, but iteratively optimizes the minimizing function. Hierarchical (coarse-to-fine) algorithm resembles such iterative method, where results from coarser levels are used to constrain a more local search at finer levels. The stereo matching methods are compared and summarized in Table 2.1.

	Local matching method	Global matching method
Basic steps	<pre> graph TD Similarity --> Local_matching[Local matching] Constraint --> Local_matching </pre>	<pre> graph TD Similarity --> Global_cost_function[Global cost function] Smoothness --> Global_cost_function Global_cost_function --> Optimization Optimization --> Global_matching[Global matching] </pre>
Approaches	Correlation Absolute difference Rank transform Phase Optical flow Features Mutual information	Dynamic programming Intrinsic curves Graph cuts Transformation model Iterative method
Advantages	Simple, fast and efficient	Reliable
Disadvantages	Sensitive to locally ambiguous regions	High computational expense Local errors may be propagated to corrupt other potentially good matching

Table 2.1: Comparison of local and global matching methods

Local method

The similarity definition is the essential work in local method. Corona *et al.* [41][42], introduced power cepstrum [43] and cross correlation techniques for similarity measure in correspondence search and surface reconstruction. They simplified the reconstruction model to a parallel stereo image pair, which assumed the two fundus images were located in the same plane only with a pure translation in scan-line direction. Therefore the

reconstruction procedure was simplified to determine the correspondences in the same horizontal line. The depth information was inversely proportional to the disparity between two corresponding points. The detected disparity in their method was only on a pixel level and also large window cubic B-spline operation was used for smoothing, which could make the final 3-D surface less accurate. Similar work can be found in [43][44].

Wang *et al.* [55] described a feature-based constraint searching approach for correspondence search. They used multiple constraints for coarse-to-fine matching point detection. In their method, the detection of latter matching points depended on the matching points that had been detected already. Hence the inaccurate matching points detected previously may affect the latter result and give out an unexpected final result. The accuracy of their approach was also on a pixel level which was suitable for large object but might not be suitable for small object such as retina.

Zhang *et al.* [56] measured the similarity by correlation and relaxation which was estimated from the surrounding matching points. They found an initial set of correspondences first and then used least median square technique to discard false correspondences and estimate the epipolar geometry. More matching points were eventually found by using the recovered epipolar constraint.

Deguchi *et al.* [40][45] used Zhang's method to estimate the dense correspondences to recover the whole 3-D retinal surface. They assumed the fundus formed part of sphere and created a sphere equation. Then the sphere parameters were optimized to recover the shape of the fundus. Although the result was successfully obtained, it might not be sufficiently useful for diagnosis because the recovered shape of the fundus was fixed to a sphere which would not allow some abnormal features to be accurately displayed.

Global method

In global method, the essential works are similarity definition, global function definition and optimization procedure. Ritter *et al.* [50] suggested an affine transformation model for image registration, which combined the use of mutual information as the similarity measure and simulated annealing as the search technique. Mutual information is a statistical measure which works on the basis of how well one image explains the other.

Stoyanov *et al.* [57] combined image calibration and rectification with constrained disparity registration for reliable depth recovery. The authors used multi-resolution registration approach based on piecewise bilinear maps (PBM), which was proposed by Veesser *et al.* [58]. In their method, the registration of two images consisted of finding a transformation, which was defined by PBM rather than certain affine transformation. The disparities obtained at low-resolution levels were propagated to higher levels and used as starting points for the optimization process.

In fundus image registration, some other researchers first detected the bifurcation points of blood vessels as the candidate points [52][59]-[63]. Then the minimization of global function was based only on these sparse vessel bifurcation points. Can *et al.* [59]-[61] provided a 12-parameter transformation model in image registration method for mosaic. They used weak perspective projection camera model and ignored the depth variation of the retinal surface when they set up the 12-parameter registration model. Thus the depth of the optic disk could not be measured precisely in their method. Moreover, sparse vessel bifurcation points were extracted as the candidate points for the following optimization of transformation parameters. The final matching result depended on the accuracy of bifurcation point extraction directly.

Zana and Klein [52] applied a matching technique using the surrounding vessel orientations of vascular bifurcations as the similarity measure. Then a Bayesian Hough transform was used to sort the transformations with their respective likelihoods and the best affine transformation was finally chosen for registration. Laliberte *et al.* [62] did the similar work based on vascular bifurcations. They tested three transformations (similarity, affine, and second-order polynomial) for comparison and found that the affine transformation gave the best results.

Matsopoulos *et al.* [63] used self-organizing maps to extract correspondences of vascular bifurcation points and then estimated the parameters of affine transformation to obtain dense matching points.

All the global methods of fundus image registration introduced earlier, model an affine transform or a polynomial transform between two images that is iteratively updated to minimize the global function. This is suitable for the reconstruction of full retinal surface, from which the fundus images are taken using a wide-angle camera as shown in Figure 2.9(a), because the wide-angle retinal surface can be considered as a quadratic surface. However, in the optic disk image reconstruction, the stereo fundus images are focused on the optic disk region as shown in Figure 2.9(b) which has a cup shape instead of a quadratic surface. The 3-D cup shape is variable in the progression of diseases. Setting the two images undergoing certain transform model may reduce the accuracy. Hence, local method is selected in this research work. A novel correspondence search approach is proposed in chapter 4 based on the combination of two local matching methods.

2.5 Review of Boundary Detection

Boundary detection and image segmentation methods may be roughly classified into several main groups as thresholding-based, edge-based, region-based, template matching, deformable model-based method, etc. [65][123][167]. Brief descriptions of these methods are given as follows.

- In thresholding-based method [65][123][170], global or local thresholding derived from image histogram is used to segment the object from the background. However, the noise of the image illumination and nonuniform brightness of the object make thresholds difficult to estimate.
- Edge-based method tries to detect the locations with rapid change of brightness or intensity from one region to the other region. In practice, an edge is seldom continuous because of noise, breaks in the edge from nonuniform illumination and other effects that introduce spurious intensity discontinuities. Thus edge detection algorithms typically are followed by edge linking procedures, such as Hough transform [65][123][167]. The limitation of this method is that edge detection schemes may suffer from incorrect detection of edges due to noise, artefact, and variability in the threshold selection in the edge image.
- Region-based method involves the division or separation of the image into regions of similar attribute by region splitting, region merging, and region growing techniques. Homogeneity is the precondition of this method [65][123].
- Template matching method is a fundamental means of object detection. Replica of an object of interest is compared to all the unknown objects in the image. If the template is sufficiently close, the unknown object is labelled as the template object.

This is known as the “knowledge-based method”, which may use the knowledge of the object or information from the training set [65].

- Deformable model-based method is an energy-minimizing operation. A deformable contour model grows to the object boundary by minimizing an energy function defined from contour shape, contour location, or priori known models, etc [64][65][171].

In this research, the optic disk boundary on the fundus image is estimated to provide the valuable information to clinicians. The difficult of this task is the occlusion of blood vessels and inhomogeneous property of the optic disk region. Thresholding-based method and region-based method are not quite suitable for disk boundary detection for the reason that their theories are based on homogenous property. Edge-based method and template matching method may only provide the approximate disk location and rough boundary because of their limitations; however they can be used as the initial results. The more flexible and attractive boundary detection approach is deformable models, which could overcome many of the difficulties in disk boundary detection, such as, the complexity and variability of the shape, etc.

2.5.1 Deformable Models

Deformable models offer a reasonable approach to solve the boundary detection and image segmentation problems, due to their stability, controllability, and flexibility. A deformable model can be characterized as a model, which under an implicit or explicit optimization criterion deforms the shape to match a known type of objects in an image. These approaches can be roughly classified in two categories: free-form deformable models and parametrically deformable models.

2.5.1.1 Free-form Deformable Models (Active Contour Models)

M. Kass [64] first introduced the active contour models (snakes) technique for boundary detection and image segmentation. Snake may be understood as a technique by matching a deformable model to an image by means of energy minimization [65]. The key point of this method is the design of suitable energy function and selection of optimization technique.

The snake is defined as a parametric curve as $\mathbf{s}(n) = [u(n), v(n)]$, where n is the index of the contour point, $1 \leq n \leq N$. For some applications, n is normalized in the range of $[0, 1]$. In snake growing, the contour is pushed to the location with the minimal global energy. The energy function to be minimized is written as

$$E_{snake} = \int_1^N [E_{int}(\mathbf{s}(n)) + E_{ext}(\mathbf{s}(n))]dn = \int_1^N [E_{int}(\mathbf{s}(n)) + E_{image}(\mathbf{s}(n)) + E_{con}(\mathbf{s}(n))]dn \quad (2.1)$$

which consists of internal energy $E_{int}(\mathbf{s}(n))$ due to bending and external energy $E_{ext}(\mathbf{s}(n))$ including image forces and constraint forces. $E_{int}(\mathbf{s}(n))$ represents the internal energy of the contour as

$$E_{int} = \frac{1}{2}[\alpha_1(n)|\mathbf{s}_n(n)|^2 + \alpha_2(n)|\mathbf{s}_{nn}(n)|^2] = \frac{1}{2}[\alpha_1(n)\left|\frac{d\mathbf{s}}{dn}\right|^2 + \alpha_2(n)\left|\frac{d^2\mathbf{s}}{dn^2}\right|^2] \quad (2.2)$$

where $\mathbf{s}_n(n) = d\mathbf{s}/dn$ and $\mathbf{s}_{nn}(n) = d^2\mathbf{s}/dn^2$ are the first and second derivatives of the contour, $\alpha_1(n)$ and $\alpha_2(n)$ specify the elasticity and stiffness of the snake [65]. $d\mathbf{s}/dn$ describes the elasticity of the contour and make it act like an elastic string, while $d^2\mathbf{s}/dn^2$ represents the bending of the contour and makes it act like a thin plate [1]. Setting the coefficients $\alpha_1(n)$ and $\alpha_2(n)$ is a task investigated by some researchers.

The image energy is derived from the image data, such as lines, edges and terminations to push the snake toward the desired object. The function is written as

$$E_{image} = w_{line} E_{line}(\mathbf{s}(n)) + w_{edge} E_{edge}(\mathbf{s}(n)) + w_{term} E_{term}(\mathbf{s}(n)) \quad (2.3)$$

Let the $I(\mathbf{s}(n))$ denote the image intensity at image location $\mathbf{s}(n) = [u(n), v(n)]$, the line-based functional may be very simple as

$$E_{line}(\mathbf{s}(n)) = I(\mathbf{s}(n)) \quad (2.4)$$

The sign of w_{line} specifies whether the snake is attracted to light or dark lines [66]. Traditionally, the magnitude of the gradient is set to be the edge-based functional as

$$E_{edge}^{(1)} = -|\nabla I(u, v)|^2 \quad \text{or} \quad E_{edge}^{(2)} = -|\nabla G_{\sigma}(u, v) * I(u, v)|^2 \quad (2.5)$$

where $G_{\sigma}(u, v)$ is 2-D zero-mean Gaussian function with standard deviation σ ; $*$ is the convolution operator; ∇ is the gradient operator defined as $(\frac{\partial}{\partial u}, \frac{\partial}{\partial v})$. E_{edge} attracts the snake to the contour to locations of strong edges. E_{term} attracts the snake towards to line terminations and corners (more details can be found in [65]). The snake behavior may be controlled by adjusting the weights w_{line} , w_{edge} and w_{term} . In most cases of boundary detection, only E_{edge} is used in image energy. The constraint energy in the third term of equation (2.1) is imposed either by a user or some other higher-level process which may force the snake toward particular features.

A snake that minimizes energy E_{snake} must satisfy the Euler-Lagrange equation [64][65] written as

$$\alpha_1(n) \mathbf{s}_{nn}(n) - \alpha_2(n) \mathbf{s}_{nnn}(n) - \nabla E_{ext} = 0 \quad (2.6)$$

This can be written as a force balance equation

$$F_{int} + F_{ext} = 0 \quad (2.7)$$

where $F_{int} = \alpha_1(n)\mathbf{s}_{nn}(n) - \alpha_2(n)\mathbf{s}_{nnnn}(n)$ and $F_{ext} = -\nabla E_{ext}$ are the internal force and external force respectively. The internal force prevents stretching and bending while the external force pulls the snake toward the desired image edges.

Traditional snake has some limitations. When the initial snake is located far away from the object boundary or there is concavity on the object boundary, the snake is difficult to converge to the actual object contour. The variations of snake, including distance snake [67], balloon snake [68], and gradient vector flow snake (GVF snake) [69], are introduced by some researchers to improve the behaviors of traditional snake.

- Distance snake uses an energy function which depends on the distance $dist(\mathbf{s})$ between snake point and its closest edge points on the edge map as $P(dist(\mathbf{s})) = \sum_n \|\mathbf{s}(n) - f(u, v)\|^2$. The external force is computed as the negative gradient of the energy function $P(dist(\mathbf{s}))$ as $F_{ext} = -\nabla P(dist(\mathbf{s}))$.
- In balloon snake, the external force is modified by normalizing it as $F_{ext} = -w_{balloon} \nabla E_{ext} / \|\nabla E_{ext}\|$. When a point of the curve is close to an edge point, it is attracted to the edge and stabilizes there. The normalization makes either smaller or larger gradient having the same influence on the curve.
- In GVF snake, gradient vector flow field is introduced as the external force denoted as $F_{ext} = G_{GVF}(u, v)$, to replace the negative gradient of the external energy.

GVF field is defined as $G_{GVF}(u, v) = [x(u, v), y(u, v)]$, which is a dense vector field from image by minimizing an energy functional

$$\varepsilon = \iint \mu(x_u^2 + x_v^2 + y_u^2 + y_v^2) + |\nabla f|^2 |G_{GVF} - \nabla f|^2 dudv \quad (2.8)$$

where $f(u, v)$ is the edge map and μ is the parameter used to adjust the tradeoff between the first term and the second term in (2.8). When $|\nabla f|$ is small, the energy is dominated by the first term. On the other hand, when $|\nabla f|$ is large the second term dominates the function and is minimized by setting $G_{GVF} = |\nabla f|$ [69].

Particular advantages of GVF snake over a traditional snake are its insensitivity to initialization and its ability to move into boundary concavities [69]. The distance snake and balloon snake can address the problem when the initial snake is located far away from the actual boundary. Dumitras and Venetsanopoulos [70] compared four snake models (traditional, distance, balloon and GVF snake models) with respect to their accuracy in describing the shape of an object and robustness to changes of model parameter values. Their experimental results showed that both distance snake and GVF snake obtained more accurate results; and GVF snake outperformed other snakes in terms of robustness.

However, both snake and GVF snake cannot work well on optic disk boundary detection because of blood vessels and inhomogeneous property of the optic disk region, although some researchers introduced the morphology operation to decrease the effect of the blood vessels [71][72]. The improvement of the snake technique in different applications is still under investigation.

2.5.1.2 Parametrically Deformable Models

Parametrically deformable models can be considered as deformable templates. It employs priori knowledge about the geometrical shape of the object. This priori shape information is available in the form of sketch, binary template or a parameter vector. The shape models can be learned from training set in order to obtain an accurate description of the shape. The deformation of the shape is presented by different values of the parameters and is restricted within the shape models to avoid taking any arbitrary shape [65][76][171].

The difference between free-form deformable models (snakes) and parametrically deformable models is that snakes are entirely free-form energy minimizing approach. There is no global structure of the object boundary. It is constrained only by local smoothness and continuity constraints, although the snake growing is global updating. Snakes do not use a priori knowledge of the shape in a direct manner. On the contrary, parametrically deformable models control the deformation using a set of parameters which are able to globally encode a specific shape. These models directly use a priori information about the geometrical shape of the object.

Parametrically deformable models are suitable to be used when more specific shape information is available and the detected object has relatively uniform shape with limited variation. However, in optic disk boundary detection, the pathological changes may arbitrarily deform the shape of optic disk and also make the blood vessels distorted. Hence, deformable templates may not be able to sufficiently encode various shapes of optic disk from different pathological changes, which might reduce the accuracy of the result.

2.5.2 Optic Disk Localization and Approximate Boundary Estimation

The optic disk boundary detection can be separated into two steps: optic disk localization and disk boundary detection. Correctly locating the optic disk is the precondition of the boundary detection that follows. Moreover, it can be used as the starting point for blood vessel tracking, or as a landmark for the identification and segmentation of other features such as fovea and lesions. Many previous researchers focused their work on locating the disk center. Most of them considered the optic disk as a circle or ellipse to obtain rough disk boundary based on the detected disk center.

Tamura and Okamoto [73] looked for the candidate region of optic disk by searching an area with highest gray-level. Edge detection and circular Hough transform were then used to estimate the size and position of the optic disk. Some researchers presented similar approach for optic disk localization [74][75]. The variations of their approaches were in different applications of edge detection and transform techniques. These approaches relied on conditions of disk shape. The optic disk, especially with pathological changes, is not always circular or elliptic, so these might reduce the accuracy. However it may be a simple method for rough localization or measurement of the potential regions of optic disk. Many other methods for locating optic disk started from this idea [1][76][77].

Sinthanayothin [78][79] located the position of optic disk by finding the region with the highest local variation in intensity. This method has been shown to work well if there are no or only few bright lesions. When the fundus image has a large number of bright lesions, this method often fails.

Li and Chutatape [76] firstly clustered the brightest pixels into several candidate regions of optic disk. Principle Component Analysis (PCA) was then applied to project the image to the 'disk space' specified by the eigenvectors which were estimated from the

training set. Then the center of optic disk was located by minimizing the distance between the original retinal image and its projection. This algorithm is a kind of template matching approach using priori knowledge. Osareh *et al.* [72] and Lowell *et al.* [80] also employed the template matching method to locate the optic disk.

Lalonde *et al.* [81] identified the potential regions of optic disk by means of a pyramidal decomposition on the green channel of the input RGB image. The final optic disk boundary was estimated by the Hausdorff-based matching on the edge map with the circular template of different sizes.

Trucco and Kamat [82] introduced a method of locating the optic disk, macula and main blood vessels together, but not singly locating the optic disk. They generated sets of plausible candidates for optic disk, macula and main vessels, and then searched the space of all possible triplets (optic disk, macula, vessels) to identify the one satisfying the priori anatomical constraints. This method efficiently reduced the uncertainty of brightness-only techniques by using the information of other components on the fundus image. Similar algorithms were utilizing geometrical relationship between the optic disk and blood vessels to identify the optic disk [83]-[85]. In these methods, the preliminary detection of the main blood vessels was required. However the blood vessel detection was a complex operation.

Although most of above methods can locate the optic disk with varying degrees of success, the accurate disk boundary is more useful for analyzing pathological change. The optic disk localization can be used to set the initial disk boundary for the following accurate disk boundary detection.

2.5.3 Optic Disk Boundary Detection

Li and Chutatape [2][86][87] produced a modified active shape model (ASM) to estimate the optic disk boundary based on point distribution model (PDM) from the training sets. The blood vessels were also used to identify the boundary. Their approach was a kind of parametrically deformable models, which could obtain relatively satisfactory global shape. The drawback of this method was that various optic disk shapes, sizes, as well as distortions of blood vessels with pathological changes were not easy to be represented by a number of shape models. Hence the accuracy might reduce. Moreover, placing of the landmark points for construction of the training set is clearly a labor intensive work [65].

Mendels *et al.* [71] introduced a method for the accurate boundary identification of the optic disk using free-form deformable model (snake) technique. In this approach, the blood vessels were first removed by gray-level morphology in image pre-processing step. Then, the disk boundary was initially located manually, and further refined through GVF snake algorithm. Osareh *et al.* [72] improved Mendels's approach in two aspects. Firstly, color morphology was employed on the original color image to remove the blood vessels. Secondly, the authors used a template matching to estimate the position of the optic disk to obtain the initial boundary automatically. The templates were estimated from huge training data. In their method, the morphology preprocessing reduced the effect of blood vessels. However, it could not totally remove the effect. Additionally, the morphology preprocessing might blur the optic disk boundary and change the location of the boundary, which made the disk boundary detection unreliable.

Lowell *et al.* [80] presented a different template to match and locate the optic disk. The disk boundary was then segmented by a deformable contour model, which used a

global elliptical model and a local deformable model with variable edge-strength dependent stiffness. The combination of global model and local deformable model could partially reduce the distractors along blood vessels and obtain smooth boundary. However, the authors indicated the performance to the images with variably pathological changes still needed to be further improved.

Walter *et al.* [88] detected the optic disk boundary by means of morphological filtering techniques and the watershed transformation. They identified the position of optic disk based on the intensity of image, before that a shade-correlation operator was applied on the input image in order to reduce the effect of bright lesions. The authors also used morphological filtering techniques to remove the blood vessels; and then estimated the contour of optic disk by the watershed transformation. However, the resulted boundary was distorted in the regions with outgoing vessels.

The methods as discussed earlier show that the occlusion of blood vessels is still one of the unsolved problems in boundary detection of optic disk. Li and Chutatape's approach could indirectly solve the vessel problem, but it might reduce the accuracy. Some of the above methods involved manual initialization. Hence, automated, robust and precise boundary detection of optic disk is one ongoing work in fundus image processing. A novel deformable model method for fully automated detection of the optic disk boundary is proposed in chapter 5, which can effectively eliminate the influence of blood vessels without reducing the accuracy.

2.6 Review of Other Related Projects

Besides the analysis of optic disk, there are some other relevant projects on analyzing and interpreting fundus images, such as localization of blood vessels and fovea, detection

of vessel width, extraction and classification of abnormal features (exudates, hemorrhages, drusens, microaneurysms, etc.). Some of these will be mentioned in the sequel.

2.6.1 Blood Vessel Detection

Information of blood vessels on retina can be used in grading disease severity or as a part of the process for automated diagnosis. As a result of systemic disease, the blood vessels may have measurable abnormalities in diameter, color and tortuosity. They can also be used as a landmark to locate the optic disk, fovea, and lesions. The techniques of blood vessel detection can be classified into two different strategies: scanning-based methods [75][78][89]-[91] and tracking-based methods [73][74][77][92]-[94].

Scanning-based methods search throughout the whole image for possible vessel pixels and then connect them together to set up the vascular network. Chaudhuri *et al.* [89] modeled the cross section of blood vessels by a Gaussian shape curve. The blood vessels were then detected using rotated matched filters and a window-based convolution. Various windows can be Sobel operator, morphological detector [95], gradient operator [75], and other matched filters [93]. Hoover *et al.* [90] introduced a thresholding method based on the matched filter response, which was built upon Chaudhuri's method. They used both local and global vessel features cooperatively to segment the vessel network. Multilayer neural network was employed in [78][91] to classify each pixel of the image into vessel or non-vessel cluster. The image was preprocessed in [78] by Canny edge operator and principal component analysis technique to enhance the blood vessels. The scanning-based methods usually provide complete segmentation of the blood vessels in the fundus image. However, in some situations the existence of random or structural noise may confuse the recognition of blood vessels and lead to erroneous results [165].

Tracking-based methods are based on the continuity properties of blood vessels. The blood vessels are segmented by repeating the convolving profile model at certain interval along the direction of the last-traced part of blood vessel. Profile model is the key point in these methods, which may base on second order derivative Gaussian matched filter [74][92][93], Canny edge operator, fuzzy model [77], or some other matched models [73][94]. Tracking-based method is not that susceptible to noise as the scanning method. One disadvantage of the tracking-based method is that it is insufficient to provide a complete vascular network in the case that the vessel fades away in some regions and appears again along the direction.

All of above-mentioned methods are based on local information to detect the blood vessels. Li and Chutatape [1][2] used point distribution model and modified active shape model to globally estimate the location of blood vessels. Prior knowledge of blood vessels was obtained from training set. After detection of blood vessels and optic disk, the fovea was often estimated according to the geometrical structure of optic disk and main blood vessels [1][2][78][82][87].

2.6.2 Fundus Image Feature Extraction

Detection, measurement and classification of lesions are also essential tasks for automated fundus image analysis and diagnosis. The bright lesions generally include hard exudates, cotton-wool spots and drusens, while hemorrhages and microaneurysms are the dark lesions.

Zhang and Chutatape [96] detected and classified exudates and cotton wool spots by local contrast enhancement and two-step improved fuzzy c-means approach. Liu *et al.* [74] applied dynamic thresholding for exudate detection. Sbeh *et al.* [97] introduced

mathematical morphology to segment drusens, which was based on the search for a new class of regional maxima components of the image. Sinthanayothin [78] introduced moat operator in order to create sharp edges of microaneurysms and hemorrhages, then recursive region growing was applied to segment microaneurysms and hemorrhages from the background. Other methods of lesion detection are discussed in [1][2][87][88][98].

2.6.3 Analysis of Disease

The investigated techniques on 3-D reconstruction and feature extraction in fundus image make the development of an automated or semi-automated analyzing system possible. Most of feature extraction systems could detect the diseases by identifying and classifying the lesions on 2-D fundus image [78][88][96][98]. The techniques have been introduced in section 2.6.2. Foveal coordinate system was generally set up in order to effectively measure the position of lesions [1][2][87]. The lesions close to the center of coordinate system, i.e., fovea, will affect the vision of patients more than the lesions in peripheral region. Analyzing the condition of optic disk was used to detect the retinal diseases. The cup-to-disk ratios in different regions on the fundus image were estimated based on 3-D optic disk image to evaluate the progression of glaucoma [7][41][42][99]. The appearance of the optic disk was proved to be an indicator of age-related macular degeneration (ARMD) [100]. Nowadays, fundus image analyzing systems presented are still not sufficiently automated. The reliability and effectiveness of various algorithms still remain the issues for improvement, and ongoing research work is being investigated by many research groups.

Chapter 3

Camera Calibration

In human visual system, a 3-D representation of the world can be reconstructed from two 2-D images projected into two eyes. A typical stereovision system, compared to human visual system, is made up of two cameras and a 3-D object to be recovered. The relationship between the two cameras is expressed as camera translation and camera rotation, which are the geometric information called extrinsic parameters. In addition, the relationship between camera Euclidean coordinate system and the image affine coordinate system (pixel coordinates on the image) is denoted by intrinsic parameters of the camera, which is an expression of the location of the optical center relative to the image plane. Camera calibration is a process to extract the intrinsic and extrinsic parameters [168]. After both the extrinsic and intrinsic parameters of the camera are determined, the 3-D object can be recovered. Here, camera calibration is defined to associate with intrinsic parameters only [65].

Epipolar geometry is the basic relationship to recover the 3-D information of the object from only a pair of images, in which the prime steps are camera calibration,

correspondence search, and depth recovery. The geometry of 3-D vision is first introduced in this chapter. The two camera calibration methods are then presented, analyzed and compared. Finally, the stereo fundus camera is investigated.

3.1 Introduction of Geometry of 3-D Vision

3.1.1 Single Camera, Pinhole Camera Model

There are a series of lenses in the fundus camera as shown in Figure 2.8. Combination of all the lenses can be modeled simply as a single lens. The camera is considered to function as the simplest and ideal pinhole camera plus distortion. The pinhole camera is modeled in such a way that light is allowed to pass through an infinitesimal hole, and then hit the image plane. The geometry of the camera model is described in Figure 3.1, where the middle plane is the image plane π to which the real world projects and the vertical dash line is the optical axis. The camera lens is located perpendicularly to the optical axis at the optical center $\mathbf{C}$ (also called the focal point). The principal point $\mathbf{m}_0$ is the intersection of the optical axis with the image plane π . The distance from the principal point $\mathbf{m}_0$ to optical center $\mathbf{C}$ is equal to the focal length f , which is one intrinsic parameter of the camera. In the real world, the optical center must be positioned in the middle of object and image plane. For convenience of analysis, the image plane and object are put on the same side of the optical center in the sequel.

3.1.1.1 Four Coordinate Systems

According to [65], four coordinate systems are defined in the camera model as follows.

- The *world Euclidean coordinate system* (subscript w) is denoted by $\mathbf{M}_w$, which is simplified to $\mathbf{M}$ in the following section.
- The *camera Euclidean coordinate system* (subscript c) denoted by $\mathbf{M}_c$ has the optical center $\mathbf{C} \equiv \mathbf{O}_c$ as its origin. The axis $\mathbf{Z}_c$ is coincident with the optical axis.
- The *image Euclidean coordinate system* (subscript i) denoted by $\mathbf{m}_i$ is lying in the image plane. The origin $\mathbf{O}_i$ is the intersection of the optical axis with the image plane π , coinciding with the principal point $\mathbf{m}_0$.
- The *image affine coordinate system* denoted by $\mathbf{m}$ is a pixel coordinate system with coordinate axes u and v . Its origin $\mathbf{O}_{uv}$ is at the top-left corner of the image. The axes u and v may not be orthogonal to each other.

Figure 3.1: The geometry of the camera model

3.1.1.2 The Relation of Four Coordinate Systems

A scene point $\mathbf{M}(X, Y, Z)$ is expressed in the world Euclidean coordinate system as a 3×1 vector. It can be expressed in the camera Euclidean coordinate system denoted by $\mathbf{M}_c$ through a translation vector $\mathbf{t}$ and a rotation matrix $\mathbf{R}$, written as

$$\mathbf{M}_c = [x_c \quad y_c \quad z_c]^T = \mathbf{R}(\mathbf{M} - \mathbf{t}) \quad (3.1)$$

The point $\mathbf{M}_c$ is projected to the image plane π at point $\mathbf{m}_c$ (expressed by camera Euclidean coordinate system). The coordinates of the projected point $\mathbf{m}_c$ are derived from the similar triangles as shown in Figure 3.2.

$$\mathbf{m}_c = [fx_c/z_c, fy_c/z_c, f]^T \quad (3.2)$$

The same point $\mathbf{m}_c$ is expressed in the image Euclidean coordinate system as $\mathbf{m}_i$

$$\mathbf{m}_i = [x_i, y_i, z_i]^T = [fx_c/z_c, fy_c/z_c, 0]^T \quad (3.3)$$

Figure 3.2: The coordinates of a projected point

Figure 3.3: The relationship between image Euclidean coordinate system and image affine coordinate system

The image affine coordinate system, with origin at the top-left corner of the image, represents a skew and rescaling of the image Euclidean coordinate system as shown in Figure 3.3. The image point (projected point) denoted by $\mathbf{m} = [u, v]^T$ can be expressed as

$$u = u_0 + k_u x_i - k_v y_i \cot \theta \quad \text{and} \quad v = v_0 + k_v y_i / \sin \theta \quad (3.4)$$

where $1/k_u$, $1/k_v$ are the pixel size along u axis and v axis respectively. The image point $\mathbf{m}$ can be represented in the 2-D image plane π in homogeneous coordinates as $\tilde{\mathbf{m}} = [u, v, 1]^T$ with a scale factor w , written as

$$\begin{aligned} w\tilde{\mathbf{m}} &= \begin{bmatrix} wu \\ wv \\ w \end{bmatrix} = \begin{bmatrix} fk_u & -fk_u \cot \theta & u_0 \\ 0 & fk_v / \sin \theta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} \\ &= \mathbf{KM}_c = \mathbf{KR}(\mathbf{M} - \mathbf{t}) \end{aligned} \quad (3.5)$$

where α and β depict scaling of f in pixels along the u and v axes respectively; γ measures in pixels the slant distance from the Y_i axis to the u axis. $\mathbf{K}$ is the intrinsic parameter matrix. All the intrinsic parameters are given in pixels. Rotation matrix $\mathbf{R}$ and

translation vector $\mathbf{t}$ are extrinsic parameters. After getting all intrinsic and extrinsic parameters, 3-D profile can be estimated.

Equation (3.5) can be rewritten using a single 3×4 matrix. A 3-D scene point is denoted by $\mathbf{M} = [X, Y, Z]^T$ and its 2-D image point is denoted by $\mathbf{m} = [u, v]^T$. Adding 1 as the last element, the two points are expressed as $\tilde{\mathbf{M}} = [X, Y, Z, 1]^T$ and $\tilde{\mathbf{m}} = [u, v, 1]^T$ respectively. By using the delimiter ‘|’ to denote that the matrix is composed of two sub-matrices, where the left partition matrix is 3×3 sub-matrix, and the right partition matrix is 3×1 vector, equation (3.5) is rewritten as

$$w\tilde{\mathbf{m}} = [\mathbf{KR} \mid -\mathbf{KRt}] \begin{bmatrix} \mathbf{M} \\ 1 \end{bmatrix} = [\mathbf{KR} \mid -\mathbf{KRt}] \tilde{\mathbf{M}} = \mathbf{H} \tilde{\mathbf{M}} \quad (3.6)$$

where $\mathbf{H}$ is a 3×4 matrix called projection matrix. It can be seen that the camera performs a linear projective transformation from the 3-D space to the 2-D image plane and the camera calibration is the operation to estimate the intrinsic parameter matrix $\mathbf{K}$.

3.1.2 Two Cameras, Stereopsis

Stereovision has enormous importance in computer vision. It has provoked a great deal of research into vision systems with two inputs that exploit the knowledge of their own relative geometry to derive depth information from the two views they receive [65].

The geometry of the stereovision system with two cameras is illustrated in Figure 3.4. $\mathbf{C}$ and $\mathbf{C}'$ are the optical centers of two cameras respectively, where the line $\mathbf{CC}'$ is called the *baseline*. Any scene point $\mathbf{M}$ and two optical centers $\mathbf{C}$, $\mathbf{C}'$ define an *epipolar plane* — the triangular shadow plane in Figure 3.4. The epipolar plane intersects two image planes in two corresponding *epipolar lines* $\mathbf{l}$ and $\mathbf{l}'$. The *epipoles* $\mathbf{e}$ and $\mathbf{e}'$ are the

intersections of the baseline with the respective image planes. To different scene points, all the epipolar lines must pass through the epipoles. Let $\mathbf{m}$ and $\mathbf{m}'$ be the projections of the scene point $\mathbf{M}$ in the left and right images respectively. The optical ray $\mathbf{Cm}$ (or $\mathbf{CM}$) represents all possible positions of the point $\mathbf{M}$ for the left image, and it is also projected into the epipolar line $\mathbf{l}'$ in the right image. The point $\mathbf{m}'$ in the right image that corresponds to the point $\mathbf{m}$ in the left image must thus lie on the epipolar line $\mathbf{l}'$ in the right image. This geometry provides a strong *epipolar constraint* that reduces the dimensionality of the search space for a correspondence between $\mathbf{m}$ in the left and $\mathbf{m}'$ in the right image from 2-D to 1-D. The scene point $\mathbf{M}$ can be determined by the intersection of two optical rays $\mathbf{Cm}$ and $\mathbf{C'm'}$.

Figure 3.4: The epipolar geometry in stereopsis (adapted from Sonka's [65])

Assuming the world Euclidean coordinate system coincides with the left camera Euclidean coordinate system, two equations derived from equation (3.6) are obtained

$$\begin{aligned} w\tilde{\mathbf{m}} &= [\mathbf{K} \mid \mathbf{0}]\tilde{\mathbf{M}} = \mathbf{KM} = \mathbf{KM}_c \\ w'\tilde{\mathbf{m}}' &= \mathbf{K}'\mathbf{R}[\mathbf{I} \mid -\mathbf{t}]\tilde{\mathbf{M}} = \mathbf{K}'\mathbf{M}_{c'} \end{aligned} \quad (3.7)$$

where $\mathbf{M}_c$ and $\mathbf{M}_{c'}$ are expressed in the left and right camera Euclidean coordinate system respectively. It is known that $\mathbf{M}_c$, $\mathbf{M}_{c'}$ and $\mathbf{t}$ are co-planar as shown in Figure 3.4. The outer product of two vectors $\mathbf{t} \times \mathbf{M}_{c'}$ produces a vector that is perpendicular to both $\mathbf{M}_{c'}$ and $\mathbf{t}$, which means it is perpendicular to the co-planarity. Therefore, the vector $\mathbf{M}_c$ is perpendicular to the vector $\mathbf{t} \times \mathbf{M}_{c'}$, the inner product of these two vectors is equal to zero. The equation to express the co-planarity is (more detail in [65])

$$(\mathbf{K}^{-1}\tilde{\mathbf{m}})^T (\mathbf{t} \times \mathbf{R}^{-1}(\mathbf{K}')^{-1}\tilde{\mathbf{m}}') = 0 \quad (3.8)$$

The translation vector $\mathbf{t} = [t_x, t_y, t_z]^T$ has already been introduced previously. A skew symmetric matrix $\mathbf{S}(\mathbf{t})$ can be created from it if $\mathbf{t} \neq \mathbf{0}$.

$$\mathbf{S}(\mathbf{t}) = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \quad (3.9)$$

The vector product can be replaced by the matrix multiplication for any matrix $\mathbf{A}$, as

$$\mathbf{t} \times \mathbf{A} = \mathbf{S}(\mathbf{t})\mathbf{A} \quad (3.10)$$

The equation (3.8) can be rewritten as

$$\tilde{\mathbf{m}}^T (\mathbf{K}^{-1})^T \mathbf{S}(\mathbf{t}) \mathbf{R}^{-1} (\mathbf{K}')^{-1} \tilde{\mathbf{m}}' = 0 \quad (3.11)$$

from which *fundamental matrix* $\mathbf{F}$ and *essential matrix* $\mathbf{E}$ in the epipolar geometry can be derived. Fundamental matrix $\mathbf{F}$ is defined as

$$\mathbf{F} = (\mathbf{K}^{-1})^T \mathbf{S}(\mathbf{t}) \mathbf{R}^{-1} (\mathbf{K}')^{-1} \quad (3.12)$$

Therefore the relationship between two images is obtained as

$$\tilde{\mathbf{m}}^T \mathbf{F} \tilde{\mathbf{m}}' = 0 \quad (3.13)$$

The intrinsic matrices $\mathbf{K}$ and $\mathbf{K}'$ give the relations as

$$\tilde{\mathbf{m}} = \mathbf{K}^{-1} \tilde{\mathbf{m}} \text{ and } \tilde{\mathbf{m}}' = \mathbf{K}'^{-1} \tilde{\mathbf{m}}' \quad (3.14)$$

Substituting equation (3.14) into (3.11) will give

$$\tilde{\mathbf{m}}^T \mathbf{S}(\mathbf{t}) \mathbf{R}^{-1} \tilde{\mathbf{m}}' = \tilde{\mathbf{m}}^T \mathbf{E} \tilde{\mathbf{m}}' = 0 \quad (3.15)$$

Equation (3.15) is a simplified version of equation (3.11). Essential matrix $\mathbf{E}$ is expressed as

$$\mathbf{E} = \mathbf{S}(\mathbf{t}) \mathbf{R}^{-1} \quad (3.16)$$

The Essential matrix contains all the geometric information about relative position from the first camera to the second camera. The extrinsic parameters (rotation matrix $\mathbf{R}$ and translation vector $\mathbf{t}$) can be calculated from the essential matrix $\mathbf{E}$. The intrinsic parameters $\mathbf{K}$ and $\mathbf{K}'$ can be estimated from camera calibration. After all the intrinsic and extrinsic parameters are determined, the 3-D scene point can be recovered. Recalling the equation (3.7) gives

$$w\tilde{\mathbf{m}} = [\mathbf{K} \mid \mathbf{0}] \tilde{\mathbf{M}} \text{ and } w'\tilde{\mathbf{m}}' = \mathbf{K}' \mathbf{R} [\mathbf{I} \mid -\mathbf{t}] \tilde{\mathbf{M}}$$

In these two equations, $\mathbf{K}$, $\mathbf{K}'$, $\mathbf{R}$, $\mathbf{t}$, $\tilde{\mathbf{m}}$ and $\tilde{\mathbf{m}}'$ are known. Each matrix equation can provide three scalar equations. The five unknowns are $\mathbf{M} = [X, Y, Z]^T$, w and w' . There are six equations totally for five unknowns; hence the 3-D scene point $\mathbf{M}$ can be determined.

3.1.3 Parallel Stereo Configuration

A special arrangement of the stereo camera rig, called parallel stereo configuration, is often used as shown in Figure 3.5, where the baseline is aligned to the horizontal scan-line of the image, the optical axes of two cameras are parallel, the epipoles move to infinity, and the corresponding epipolar lines in two image planes are located at the same horizontal scan-line.

Figure 3.5: Parallel stereo configuration (adapted from Sonka's [65])

For this configuration, the computation of depth is slightly simpler than general stereo configuration. Figure 3.6 illustrates the geometry of parallel stereo configuration with optical centers C_L and C_R , and baseline denoted as b . A 3-D object point $\mathbf{M}(X, Y, Z)$ is projected on the left and right image planes at points $\mathbf{M}_L(x_L, y_L)$ and $\mathbf{M}_R(x_R, y_R)$ respectively. For clearly viewing the parallel stereo configuration, the 3-D geometric image as shown in Figure 3.6(a) is projected onto the Y-Z plane as given in Figure 3.6(b).

(a) 3-D geometry of parallel stereo configuration

(b) Projection on $Y-Z$ plane

Figure 3.6: Geometry of parallel stereo configuration

Three equations can be obtained from the similar triangles as

$$\begin{cases} \frac{y_L}{f} = \frac{b/2 + Y}{Z} \\ -\frac{y_R}{f} = \frac{b/2 - Y}{Z} \\ \frac{x_L}{f} = \frac{x_R}{f} = \frac{X}{Z} \end{cases} \quad (3.17)$$

Then, point $\mathbf{M}(X, Y, Z)$ can be computed by

$$\begin{cases} X = \frac{bx_L}{(y_L - y_R)} = \frac{bx_R}{(y_L - y_R)} \\ Y = \frac{1}{2} \frac{b(y_L + y_R)}{(y_L - y_R)} \\ Z = \frac{bf}{(y_L - y_R)} \end{cases} \quad (3.18)$$

where $(y_L - y_R)$ is the *disparity* in the observation of $\mathbf{M}$. Therefore, in parallel stereo configuration, the prime steps of 3-D reconstruction are measurements of focal length, baseline, and corresponding points (disparities).

3.2 Camera Calibration

Camera calibration is the first step in the 3-D reconstruction, which is an operation to estimate matrix $\mathbf{K}$. Calibration of one camera together with the knowledge of the coordinates of one image point allows us to determine a ray in space uniquely, as shown by $\mathbf{Cm}$ in Figure 3.4. If two calibrated cameras observe the same scene point $\mathbf{M}$, its 3-D coordinates can be computed as the intersection of two such rays [65]. This is the basic principle of stereovision. As introduced in chapter 2, the photogrammetric calibration method is more suitable for the fundus camera. Two photogrammetric calibration methods

based on 3-D calibration object [101] and 2-D calibration object [33], are presented and compared in this chapter.

3.2.1 Camera Calibration Based on 3-D Object

A 3-D-object-based camera-calibration method was presented in [101]. Only four intrinsic parameters, α , β , u_0 and v_0 , were measured in this approach by both linear and non-linear methods, which will be briefly introduced in section 3.2.1.1. In this research, the parameter γ is also measured to improve the accuracy. The lens distortion is considered and added in the 3-D-object-based camera-calibration method. The Levenberg-Marquardt method is implemented to optimize the intrinsic parameters.

Figure 3.7: The typical 3-D calibration object

Recall the equations (3.6) $w\tilde{\mathbf{m}} = \mathbf{H}\tilde{\mathbf{M}}$, where $\mathbf{H} = \mathbf{KR}[\mathbf{I} | -\mathbf{t}]$. Given enough 3-D object points and their corresponding image points, $\mathbf{H}$ can be determined. Projection matrix $\mathbf{H}$ depends upon eleven parameters: five intrinsic parameters $\alpha, \beta, \gamma, u_0, v_0$, three independent degrees of freedom associated with the rotation matrix $\mathbf{R}$, and three translation parameters t_x, t_y, t_z . If more than eleven equations are available, the parameters of projection matrix can be determined, from which the intrinsic parameter matrix $\mathbf{K}$ can

be computed. The typical and simplest 3-D calibration object consists of two or three orthogonal planes as shown in Figure 3.7. The 3-D object points must not simply lie on one plane; otherwise the unique solution cannot be achieved.

3.2.1.1 Estimating Projection Matrix

The equation (3.6) $w\tilde{\mathbf{m}} = \mathbf{H}\tilde{\mathbf{M}}$ can be extended [101] as

$$\begin{bmatrix} wu \\ wv \\ w \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} & h_{14} \\ h_{21} & h_{22} & h_{23} & h_{24} \\ h_{31} & h_{32} & h_{33} & h_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (3.19)$$

From which

$$u = \frac{Xh_{11} + Yh_{12} + Zh_{13} + h_{14}}{Xh_{31} + Yh_{32} + Zh_{33} + h_{34}} \quad \text{and} \quad v = \frac{Xh_{21} + Yh_{22} + Zh_{23} + h_{24}}{Xh_{31} + Yh_{32} + Zh_{33} + h_{34}} \quad (3.20)$$

Rewriting this gives

$$\begin{aligned} Xh_{11} + Yh_{12} + Zh_{13} + h_{14} - uXh_{31} - uYh_{32} - uZh_{33} - uh_{34} &= 0 \\ Xh_{21} + Yh_{22} + Zh_{23} + h_{24} - vXh_{31} - vYh_{32} - vZh_{33} - vh_{34} &= 0 \end{aligned} \quad (3.21)$$

Then equation can be written as

$$\mathbf{L}\mathbf{h} = 0 \quad (3.22)$$

Given N 3-D object points as well as their 2-D image points denoted by $\mathbf{M}_n$ and $\mathbf{m}_n$ respectively where $1 \leq n \leq N$, and each point provides two equations, then $\mathbf{L}$ becomes a $2N \times 12$ matrix and $\mathbf{h}$ is a 12×1 vector. In general, at least 6 points are needed to estimate the projection matrix $\mathbf{H}$. The following matrix equation can be set up.

$$\begin{bmatrix}
X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -u_1 X_1 & -u_1 Y_1 & -u_1 Z_1 & -u_1 \\
0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -v_1 X_1 & -v_1 Y_1 & -v_1 Z_1 & -v_1 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
X_N & Y_N & Z_N & 1 & 0 & 0 & 0 & 0 & -u_N X_N & -u_N Y_N & -u_N Z_N & -u_N \\
0 & 0 & 0 & 0 & X_N & Y_N & Z_N & 1 & -v_N X_N & -v_N Y_N & -v_N Z_N & -v_N
\end{bmatrix}
\begin{bmatrix}
h_{11} \\
h_{12} \\
h_{13} \\
h_{14} \\
h_{21} \\
h_{22} \\
h_{23} \\
h_{24} \\
h_{31} \\
h_{32} \\
h_{33} \\
h_{34}
\end{bmatrix} = 0
\quad (3.23)$$

The solution is well known to be the eigenvector of $\mathbf{L}^T \mathbf{L}$ corresponding to the smallest eigenvalue [33][65][101]. The intrinsic parameters $\mathbf{K}$, rotation matrix $\mathbf{R}$ and translation vector $\mathbf{t}$ can be computed from $\mathbf{H}$. Let

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} & h_{14} \\ h_{21} & h_{22} & h_{23} & h_{24} \\ h_{31} & h_{32} & h_{33} & h_{34} \end{bmatrix} = \begin{bmatrix} \mathbf{h}_1^T & h_{14} \\ \mathbf{h}_2^T & h_{24} \\ \mathbf{h}_3^T & h_{34} \end{bmatrix} = \begin{bmatrix} \alpha \mathbf{c}_1 + \gamma \mathbf{c}_2 + u_0 \mathbf{c}_3 & (-\mathbf{K} \mathbf{R} \mathbf{t})_1 \\ \beta \mathbf{c}_2 + v_0 \mathbf{c}_3 & (-\mathbf{K} \mathbf{R} \mathbf{t})_2 \\ \mathbf{c}_3 & (-\mathbf{K} \mathbf{R} \mathbf{t})_3 \end{bmatrix} \quad (3.24)$$

In it

$$\mathbf{R} = \begin{bmatrix} \mathbf{c}_{11} & \mathbf{c}_{12} & \mathbf{c}_{13} \\ \mathbf{c}_{21} & \mathbf{c}_{22} & \mathbf{c}_{23} \\ \mathbf{c}_{31} & \mathbf{c}_{32} & \mathbf{c}_{33} \end{bmatrix} = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \mathbf{c}_3 \end{bmatrix} \quad (3.25)$$

where $\mathbf{c}_i$ is the i^{th} row of the rotation matrix $\mathbf{R}$, and $\mathbf{h}_i^T = (h_{i1}, h_{i2}, h_{i3})$, $i=1, 2, 3$. Using the knowledge that $\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3$ are orthonormal, the intrinsic parameters can be computed as

$$\begin{aligned}
u_0 &= \mathbf{h}_1^T \mathbf{h}_3 \\
v_0 &= \mathbf{h}_2^T \mathbf{h}_3 \\
\beta &= \sqrt{\mathbf{h}_2^T \mathbf{h}_2 - v_0^2} \\
\gamma &= (\mathbf{h}_1^T \mathbf{h}_2 - u_0 v_0) / \beta \\
\alpha &= \sqrt{\mathbf{h}_1^T \mathbf{h}_1 - u_0^2 - \gamma^2}
\end{aligned} \tag{3.26}$$

$\|\mathbf{h}_3\|^2 = \|\mathbf{c}_3\|^2 = 1$ is the constraint used to estimate the projection matrix $\mathbf{H}$. After all the intrinsic parameters are obtained, the extrinsic parameters $\mathbf{R}$ and $\mathbf{t}$ can be estimated easily. Rewrite $\mathbf{H} = [\mathbf{KR} \mid -\mathbf{KRt}] = [\mathbf{H}_{\mathbf{KR}} \mid \mathbf{H}_{\mathbf{KRt}}]$, then

$$\mathbf{R} = \mathbf{K}^{-1} \mathbf{H}_{\mathbf{KR}} \text{ and } \mathbf{t} = -\mathbf{R}^{-1} \mathbf{K}^{-1} \mathbf{H}_{\mathbf{KRt}} \tag{3.27}$$

The linear method can estimate the projection matrix $\mathbf{H}$ quickly, but the results are not precise enough. It is however possible to solve the equation (3.19) and obtain more accurate results from a nonlinear criterion by minimizing the sum square distances in the image plane between the image point $\mathbf{m}_n$ and the back-projected image point $\mathbf{m}_n(\mathbf{M}_n, \mathbf{H})$ associated with the object point $\mathbf{M}_n$ and estimated projection matrix $\mathbf{H}$. The function to be minimized is

$$\sum_{n=1}^N \|\mathbf{m}_n - \mathbf{m}_n(\mathbf{M}_n, \mathbf{H})\|^2 = \sum_{n=1}^N \left\| u_n - \frac{\mathbf{h}_1^T \mathbf{M}_n + h_{14}}{\mathbf{h}_3^T \mathbf{M}_n + h_{34}} \right\|^2 + \sum_{n=1}^N \left\| v_n - \frac{\mathbf{h}_2^T \mathbf{M}_n + h_{24}}{\mathbf{h}_3^T \mathbf{M}_n + h_{34}} \right\|^2 \tag{3.28}$$

3.2.1.2 Lens Distortion

Up to now, distortion-free camera model is discussed. The actual lens however is not ideal central projection as a pinhole camera model, because of several types of imperfections in design and assembly of lenses composing the camera optical system. For the accuracy requirement, lens distortion cannot be ignored. The lens distortions,

consisting of radial distortion, decentering distortion and thin prism distortion, will effectively map the straight lines into curves as shown in Figure 3.8(a), (b). Radial distortion is mainly caused by the flawed curvature curve of the lenses. Decentering distortion arises when the optical centers of the lenses are not strictly collinear. The slight tilt of some lenses or image sensor results in the prism distortion [35]. Based on the reports in [33][34][36], it is likely that the radial distortion especially the first term, dominates the whole distortion function.

Figure 3.8: The ideal and actual images of grid pattern

Let $\mathbf{m}(u, v)$ be the ideal (non-distorted) image point in the image affine coordinate system, and $\hat{\mathbf{m}}(\hat{u}, \hat{v})$ is the corresponding observed (distorted) image point. Similarly, let $\mathbf{M}_i(x_i, y_i)$ and $\hat{\mathbf{M}}_i(\hat{x}_i, \hat{y}_i)$ be the ideal and observed image points in the image Euclidean coordinate system (indicated by subscript i) respectively, the radial lens distortion can then be expressed as an infinite series as $\hat{r}_i = r_i + k_1 r_i^3 + k_2 r_i^5 + k_3 r_i^7 + \dots$, where $r_i = \sqrt{x_i^2 + y_i^2}$ and $\hat{r}_i = \sqrt{\hat{x}_i^2 + \hat{y}_i^2}$ [35][38]. Terms higher than the 5th-order one can be ignored as their contributions to the radial distortion are negligible [38]. From [33], the distortion function is written as

$$\begin{aligned}\hat{x}_i &= x_i + x_i(k_1 r_i^2 + k_2 r_i^4) \\ \hat{y}_i &= y_i + y_i(k_1 r_i^2 + k_2 r_i^4)\end{aligned}\quad (3.29)$$

The subscript i here indicates the image Euclidean coordinate system instead of the point number. $\mathbf{k} = [k_1, k_2]^T$ are the coefficients of the radial distortion. Since the lens distortion parameters are estimated after all the other five intrinsic parameters have been determined, the projected image point $\mathbf{M}_i = [x_i, y_i]^T$ can be computed as

$$[x_i, y_i]^T = [fx_c/z_c, fy_c/z_c]^T \quad (3.30)$$

where the focal length is estimated by $f = \alpha/k_u$ from (3.5). The center of the radial distortion is the same as the principal point. Substituting (3.29) into the modified (3.4):

$$\hat{u} = k_u \hat{x}_i - k_u \hat{y}_i \cot \theta + u_0 \quad \text{and} \quad \hat{v} = k_v \hat{y}_i / \sin \theta + v_0 \quad (3.31)$$

will give

$$\begin{aligned}\hat{u} &= u + (u - u_0)(k_1 r_i^2 + k_2 r_i^4) \\ \hat{v} &= v + (v - v_0)(k_1 r_i^2 + k_2 r_i^4)\end{aligned}\quad (3.32)$$

The radial distortion parameters can be computed by a simple linear method. The results will be later used as the initial guess for the nonlinear optimization method. Rewrite equation (3.32) as

$$\begin{bmatrix} (u - u_0)r_i^2 & (u - u_0)r_i^4 \\ (v - v_0)r_i^2 & (v - v_0)r_i^4 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} \hat{u} - u \\ \hat{v} - v \end{bmatrix} \quad (3.33)$$

The equations are written in a matrix form as $\mathbf{Bk} = \mathbf{b}$. Given N points, $\mathbf{B}$ is a $2N \times 2$ matrix and $\mathbf{b}$ is a $2N \times 1$ vector. The linear least square solution is given by

$$\mathbf{k} = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{b} \quad (3.34)$$

Once $\mathbf{k}$ is determined by linear method, the solution is to be used as the initial guess of the nonlinear method. The projection matrix will be optimized together with radial distortion parameters, by minimizing the following function

$$\sum_{n=1}^N \|\mathbf{m}_n - \hat{\mathbf{m}}(\mathbf{M}_n, \mathbf{H}, \mathbf{k})\|^2 \quad (3.35)$$

3.2.1.3 Implementing Levenberg-Marquardt Method for Nonlinear Optimization

Levenberg-Marquardt (L-M) method is a typical nonlinear downhill minimization algorithm, which has proved to be an effective and popular way to solve the nonlinear least square problem. It dynamically combines Gauss-Newton and gradient-descent iterations. In this experiment, the results of the linear method are used as the initial guess for the L-M algorithm.

The theory of Levenberg-Marquardt method

Assume (x_n, y_n) , $1 \leq n \leq N$, are N extracted data points. The model used to fit the data that depends on M adjustable parameters $\mathbf{a} = (a_1, a_2, \dots, a_M)^T$ is $y = y(x; \mathbf{a})$. Here, (x_n, y_n) is used to explain the principle of L-M method. It has different meaning from previous introduction. In the camera calibration application, 3-D scene point $\mathbf{M}_n(X_n, Y_n, Z_n)$ corresponds to x_n , and image point $\mathbf{m}_n(u_n, v_n)$ corresponds to y_n . 14 optimizing parameters, h_{ij} ($1 \leq i \leq 3, 1 \leq j \leq 4$) and k_1, k_2 , form the optimizing parameter vector $\mathbf{a} = (a_1, a_2, \dots, a_M)^T$, $M = 14$. The merit function χ^2 is given as

$$\chi^2(\mathbf{a}) = \sum_{n=1}^N \left[\frac{y_n - y_n(x_n; \mathbf{a})}{\sigma_n} \right]^2 \quad (3.36)$$

where σ_n ($1 \leq n \leq N$) are the estimated errors on the measurements for each y_n . Let

$$\alpha_{kl} = \frac{1}{2} \frac{\partial^2 \chi^2}{\partial a_k \partial a_l} = \sum_{n=1}^N \frac{1}{\sigma_n^2} \frac{\partial y_n(x_n; \mathbf{a})}{\partial a_k} \frac{\partial y_n(x_n; \mathbf{a})}{\partial a_l}, \quad 1 \leq k, l \leq M \quad (3.37)$$

$$\beta_k = -\frac{1}{2} \frac{\partial \chi^2}{\partial a_k} = \sum_{n=1}^N \frac{[y_n - y_n(x_n; \mathbf{a})]}{\sigma_n^2} \frac{\partial y_n(x_n; \mathbf{a})}{\partial a_k}, \quad 1 \leq k \leq M \quad (3.38)$$

In Levenberg-Marquardt method

$$\begin{aligned} \alpha'_{jj} &= \alpha_{jj}(1 + \lambda) \\ \alpha'_{jk} &= \alpha_{jk} \quad (j \neq k) \end{aligned} \quad (3.39)$$

where $j, k=1, 2, \dots, M$. A value λ is added to the diagonal elements of matrix α in L-M method, which allows the algorithm to smoothly switch between Gauss-Newton algorithm (small λ) and gradient-descent algorithm (large λ). λ is a constant equal for all j . The value of λ varies during the minimization procedure, which will be introduced later. Let $\mathbf{a}_{min}$ be the parameter to achieve the minimal χ^2 , and $\mathbf{a}_{curr}$ is the current optimized parameter, then $\delta \mathbf{a} = \mathbf{a}_{min} - \mathbf{a}_{curr}$ is the update value in a single iteration. We have

$$\sum_{l=1}^M \alpha'_{kl} \delta a_l = \beta_k \quad (3.40)$$

δa_l is the update value to the optimizing parameter a_l . Both α'_{kl} and β_k are required to calculate δa_l .

Implementing the Levenberg-Marquardt method

In this camera calibration application, since each image point $\mathbf{m}_n(u_n, v_n)$ is extracted independently using the same procedure, it can be assumed that the estimated error for each point is a constant as $\sigma_n = \sigma$. Consequently, the constant σ can be omitted in the

merit function because it will not affect the results of optimization. The minimizing merit function will therefore become

$$\chi^2 = \sum_{n=1}^N |u_n - \hat{u}_n(\mathbf{M}_n, \mathbf{H}, \mathbf{k})|^2 + \sum_{i=1}^N |v_n - \hat{v}_n(\mathbf{M}_n, \mathbf{H}, \mathbf{k})|^2 \quad (3.41)$$

The main steps for minimization are as follows:

- (1) Choose an initial guess of $\mathbf{a}$
- (2) Compute $\chi^2(\mathbf{a})$
- (3) Select a modest value of λ (such as $\lambda=0.001$)
- (4) Solve the linear equation (3.40) for $\delta\mathbf{a}$

The derivative function is determined as the following:

$$\frac{\partial y_n(x_n; \mathbf{a})}{\partial a_l} = \frac{y_n(x_n; a_l + \delta a_l) - y_n(x_n; a_l)}{\delta a_l}$$

where δa_l is set to be $\max(0.01 \delta a_l, 0.0001)$.

Build up the 14×14 matrix $\boldsymbol{\alpha}'$ and 14×1 vector $\boldsymbol{\beta}$ based on the 14 optimizing parameters.

$$\alpha_{kl} = \sum_{n=1}^N \left(\frac{\partial u_n(\mathbf{M}_n, \mathbf{a})}{\partial a_k} \times \frac{\partial u_n(\mathbf{M}_n, \mathbf{a})}{\partial a_l} \right) + \sum_{n=1}^N \left(\frac{\partial v_n(\mathbf{M}_n, \mathbf{a})}{\partial a_k} \times \frac{\partial v_n(\mathbf{M}_n, \mathbf{a})}{\partial a_l} \right)$$

$(1 \leq k \leq 14, 1 \leq l \leq 14)$

$$\beta_k = \sum_{n=1}^N (u_n - \hat{u}_n(\mathbf{M}_n; \mathbf{a})) \frac{\partial u_n(\mathbf{M}_n, \mathbf{a})}{\partial a_k} + \sum_{n=1}^N (v_n - \hat{v}_n(\mathbf{M}_n; \mathbf{a})) \frac{\partial v_n(\mathbf{M}_n, \mathbf{a})}{\partial a_k}$$

$(1 \leq k \leq 14)$

Compute α' and rewrite equation (3.40) as matrix product $\alpha' \delta \mathbf{a} = \beta$, therefore

$$\delta \mathbf{a} = \alpha'^{-1} \beta.$$

(5) Recompute χ^2

If $\chi^2(\mathbf{a} + \delta \mathbf{a}) \geq \chi^2(\mathbf{a})$, increase λ by a factor of 2 (or any substantial factor) and go back to (4)

If $\chi^2(\mathbf{a} + \delta \mathbf{a}) < \chi^2(\mathbf{a})$, decrease λ by a factor of 2, update the trial solution $\mathbf{a} + \delta \mathbf{a} \rightarrow \mathbf{a}$ and go back to (4)

(6) Repeat steps (4)(5), until satisfying the condition:

$\chi^2(\mathbf{a} + \delta \mathbf{a}) < \chi^2(\mathbf{a})$ and $|\chi^2(\mathbf{a} + \delta \mathbf{a}) - \chi^2(\mathbf{a})| < 0.1$ twice consecutively. The final optimized solution is $\mathbf{a}$.

3.2.2 Camera Calibration Based on Planar Object

This planar-object-based camera-calibration method was proposed in [33]. A brief introduction is given in this section. Assume the 2-D model plane is on $Z=0$ of the world Euclidean coordinate system. Let $\mathbf{r}_i$ denote the i^{th} column of the rotation matrix $\mathbf{R}$, equation (3.6) $w\tilde{\mathbf{m}} = \mathbf{H}\tilde{\mathbf{M}}$ can be written as

$$w \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{K} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & -\mathbf{Rt} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = \mathbf{K} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & -\mathbf{Rt} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \quad (3.42)$$

where $\tilde{\mathbf{M}}$ is changed into $\tilde{\mathbf{M}} = [X, Y, 1]^T$, since Z is always equal to 0. Therefore a model point $\tilde{\mathbf{M}}$ and its image $\tilde{\mathbf{m}}$ is related by a 3×3 projection matrix $\mathbf{H}$.

$$\mathbf{H} = [\mathbf{q}_1 \quad \mathbf{q}_2 \quad \mathbf{q}_3] = s\mathbf{K}[\mathbf{r}_1 \quad \mathbf{r}_2 \quad -\mathbf{Rt}] \quad (3.43)$$

where s is an arbitrary scalar; $\mathbf{q}_i$ is the i^{th} column of $\mathbf{H}$. Using the knowledge that $\mathbf{r}_1$ and $\mathbf{r}_2$ are orthonormal, two constraint equations are obtained:

$$\begin{aligned} \mathbf{q}_1^T \mathbf{K}^{-T} \mathbf{K}^{-1} \mathbf{q}_2 &= 0 \\ \mathbf{q}_1^T \mathbf{K}^{-T} \mathbf{K}^{-1} \mathbf{q}_1 &= \mathbf{q}_2^T \mathbf{K}^{-T} \mathbf{K}^{-1} \mathbf{q}_2 \end{aligned} \quad (3.44)$$

These are two basic constraints on the intrinsic parameters for a given projection matrix $\mathbf{H}$. Thus each projection matrix $\mathbf{H}$ can provide two equations. There are five unknown parameters in $\mathbf{K}$; hence at least three $\mathbf{H}$ are needed. Similarly, linear method is chosen to obtain the initial guess; the nonlinear method is then used to acquire the accurate result. Let

$$\begin{aligned} \mathbf{P} = \mathbf{K}^{-T} \mathbf{K}^{-1} &= \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{12} & p_{22} & p_{23} \\ p_{13} & p_{23} & p_{33} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{\alpha^2} & -\frac{\gamma}{\alpha^2 \beta} & \frac{v_0 \gamma - u_0 \beta}{\alpha^2 \beta} \\ -\frac{\gamma}{\alpha^2 \beta} & \frac{\gamma^2}{\alpha^2 \beta^2} + \frac{1}{\beta^2} & -\frac{\gamma(v_0 \gamma - u_0 \beta)}{\alpha^2 \beta^2} - \frac{v_0}{\beta^2} \\ \frac{v_0 \gamma - u_0 \beta}{\alpha^2 \beta} & -\frac{\gamma(v_0 \gamma - u_0 \beta)}{\alpha^2 \beta^2} - \frac{v_0}{\beta^2} & \frac{(v_0 \gamma - u_0 \beta)^2}{\alpha^2 \beta^2} + \frac{v_0^2}{\beta^2} + 1 \end{bmatrix} \end{aligned} \quad (3.45)$$

$\mathbf{P}$ is a symmetric matrix. Hence, a six dimensional vector $\mathbf{p}$ can be defined as

$$\mathbf{p} = [p_{11}, p_{12}, p_{22}, p_{13}, p_{23}, p_{33}]^T \quad (3.46)$$

Then

$$\mathbf{q}_i^T \mathbf{K}^{-T} \mathbf{K}^{-1} \mathbf{q}_j = \mathbf{q}_i^T \mathbf{P} \mathbf{q}_j = [q_{1i}, q_{2i}, q_{3i}] \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{12} & p_{22} & p_{23} \\ p_{13} & p_{23} & p_{33} \end{bmatrix} \begin{bmatrix} q_{1j} \\ q_{2j} \\ q_{3j} \end{bmatrix} \quad (3.47)$$

Extend equation (3.47), and let

$$\mathbf{v}_{ij} = [q_{1i}q_{1j}, \quad q_{1i}q_{2j} + q_{2i}q_{1j}, \quad q_{2i}q_{2j}, \\ q_{1i}q_{3j} + q_{3i}q_{1j}, \quad q_{2i}q_{3j} + q_{3i}q_{2j}, \quad q_{3i}q_{3j}]^T \quad (3.48)$$

The two constraint equations (3.44) can be rewritten as a linear matrix equation as

$$\begin{bmatrix} \mathbf{v}_{12}^T \\ (\mathbf{v}_{11} - \mathbf{v}_{22})^T \end{bmatrix} \mathbf{p} = \mathbf{V} \mathbf{p} = 0 \quad (3.49)$$

where $\mathbf{V}$ is 2×6 matrix. Given M images ($M \geq 3$), $\mathbf{V}$ is modified to a $2M \times 6$ matrix. The solution is the eigenvector of $\mathbf{V}^T \mathbf{V}$ corresponding to the minimal eigenvalue. The intrinsic parameters can be extracted from the matrix $\mathbf{P}$, which is estimated up to an arbitrary scale factor as $\mathbf{P} = s \mathbf{K}^{-T} \mathbf{K}^{-1}$. The unique solutions [33] are

$$\begin{aligned} v_0 &= (p_{12}p_{13} - p_{11}p_{23}) / (p_{11}p_{22} - p_{12}^2) \\ s &= p_{33} - [p_{13}^2 + v_0(p_{12}p_{13} - p_{11}p_{23})] / p_{11} \\ \alpha &= \sqrt{s / p_{11}} \\ \beta &= \sqrt{sp_{11} / (p_{11}p_{22} - p_{12}^2)} \\ \gamma &= -p_{12}\alpha^2\beta / s \\ u_0 &= \gamma v_0 / \beta - p_{13}\alpha^2 / s \end{aligned} \quad (3.50)$$

Similar to 3-D-object-based method, the radial distortion must be considered here. Given M images and N points for each image, the matrix equation of radial distortion is $\mathbf{B}\mathbf{k} = \mathbf{b}$, where $\mathbf{B}$ is a $2MN \times 2$ matrix. An initial guess of k_1 and k_2 can be obtained

from the linear solution. Nonlinear solution is optimized by L-M method to minimize the function

$$\sum_{i=1}^M \sum_{j=1}^N \left\| \mathbf{m}_{ij} - \hat{\mathbf{m}}(\mathbf{M}_j, \mathbf{H}_i, \mathbf{k}) \right\|^2 \quad (3.51)$$

3.2.3 Results and Comparison

The camera to be calibrated is Canon non-mydratic fundus camera CR6-45NM. The image resolution is 1140×2160 pixels. One pixel is equal to 10.5μm in both u and v directions on the imaging sensor. The calibration object was placed at approximately 100mm in front of the objective lens of the camera, which is close to the working distance of the fundus camera. In both of two approaches, i.e., 3-D-object-based camera-calibration method and planar-object-based camera-calibration method, linear method was first used to estimate the initial results of intrinsic parameters; the L-M nonlinear optimization method was then applied to obtain final results.

3.2.3.1 Results of 3-D-object-based Camera-calibration Method

The 3-D calibration object pattern applied in the experiments has two orthogonal checkboard planes with the dimension of 12.71mm×12.71mm for each checkboard. Ten images were captured from different angles of view as illustrated in Figure 3.9. The corners were semi-automatically extracted by modified Plessey corner detection algorithm [102] whose threshold was manually adjusted for different images. The average error distance (AED) between image points and their corresponding back-projected image points is used to evaluate the accuracy of the method, written as

$$AED = \sqrt{\frac{1}{N} \sum_{n=1}^N \|\mathbf{m}_n - \hat{\mathbf{m}}(\mathbf{M}_n, \mathbf{H}, \mathbf{k})\|^2} \quad (3.52)$$

The camera calibration results of Figure 3.9(1) are illustrated in Table 3.1 to compare the camera model including the lens distortion with the model without lens distortion. It can be observed, when lens distortion was considered, the back-projection error (AED) decreased from 8.14 pixels to 3.49 pixels. Hence, camera model with lens distortion was more suitable for fundus camera. The results of ten images are shown in Table 3.2, where the last two rows display the mean and standard deviation of the computed intrinsic parameters.

(1)

(2)

(3)

(4)

(5)

(6)

(7)

(8)

(9)

(10)

Figure 3.9: Images of 3-D calibration object. Corners are marked by 'x'.

	Intrinsic matrix K (pixel)	Distortion parameters k ($\times 10^{-3} \text{ mm}^{-2}$, $\times 10^{-3} \text{ mm}^{-4}$)	AED (pixel)
Camera model without lens distortion	$\begin{bmatrix} 1675.66 & 6.63 & 712.27 \\ 0 & 1700.37 & 1054.70 \\ 0 & 0 & 1 \end{bmatrix}$	–	8.14
Camera model with lens distortion	$\begin{bmatrix} 2007.90 & 12.27 & 720.54 \\ 0 & 2013.23 & 1092.97 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} -3.6 \\ 0.035 \end{bmatrix}$	3.49

Table 3.1: Comparison results of different camera model based on 3-D calibration object

Image no.	α (pixel)	β (pixel)	γ (pixel)	u_0 (pixel)	v_0 (pixel)	k_1 ($\times 10^{-3} \text{ mm}^{-2}$)	k_2 ($\times 10^{-5} \text{ mm}^{-4}$)	AED (pixel)
1	2007.90	2013.23	12.27	720.54	1092.97	-3.60	3.55	3.49
2	2015.10	2014.83	10.92	679.86	1066.51	-3.24	2.73	4.10
3	1991.20	1990.34	10.03	694.99	1086.35	-3.03	2.54	4.00
4	2034.19	2031.25	13.49	684.90	1094.83	-3.21	2.77	3.55
5	1994.68	1996.27	15.20	685.32	1175.86	-2.97	2.41	4.03
6	1972.33	1973.71	5.57	694.43	1184.90	-2.65	1.88	4.46
7	2031.33	2024.18	7.78	643.51	1166.75	-2.64	1.68	4.29
8	1972.80	1977.92	8.58	728.56	1105.86	-3.28	3.28	4.48
9	2027.15	2032.65	11.05	671.52	1131.99	-3.44	2.94	3.99
10	1990.74	1992.36	8.86	727.18	1122.28	-3.36	3.20	4.63
Mean	2003.74	2004.67	10.37	693.08	1122.83	-3.14	2.70	4.10
Std	22.93	21.46	2.84	26.67	41.02	0.32	0.59	0.37

Table 3.2: Camera calibration results based on 3-D calibration object with lens distortion considered

3.2.3.2 Results of Planar-object-based Camera-calibration Method

The planar calibration object is a smooth plane with checkboard pattern whose dimension is 12.14mm×12.14mm for each checkboard. Eight images of the planar calibration object were captured from different angles of view, and the corner points were extracted as shown in Figure 3.10. In these 8 images, 2 were taken from the straight front view (middle direction); 3 were taken from the left front view (left direction); and the other 3 were taken from the right front view (right direction). The evaluation criterion becomes

$$AED = \sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \|\mathbf{m}_{ij} - \hat{\mathbf{m}}(\mathbf{M}_j, \mathbf{H}_i, \mathbf{k})\|^2} \quad (3.53)$$

where M is the image number and N is the number of extracted corners on each image.

According to Zhang's [33], three images ($M=3$) could provide expected result. Moreover, the three-image combination undergoing pure translation is the degenerated case of this method, because the constraint equation (3.49) obtained from each image will not be independent to each other in this situation. Hence, in this experiment, to avoid the degenerated case, the three-image combination must follow the selection criterion that is one image taken from the left direction, one from the middle direction, and the other one from the right direction.

The camera calibration results of one three-image combination, Figure 3.10(1), (4), (7), are given in Table 3.3, to compare the effectiveness of two camera models. Similar to the results of 3-D-object-based camera-calibration method, when lens distortion was considered in the camera model, the back-projection error (AED) was ten-fold reduced from 10.67 pixels to 0.98 pixels. Ten different three-image combinations (groups) were

randomly selected from 8 images based on the selection criterion to estimate the intrinsic parameters. The results are given in Table 3.4.

(1) Left direction 1

(2) Left direction 2

(3) Left direction 3

(4) Middle direction 1

(5) Middle direction 2

(6) Right direction 1

(7) Right direction 2

(8) Right direction 3

Figure 3.10: Images of planar calibration object. Corners are marked by 'x'.

	Intrinsic matrix K (pixel)	Distortion parameters k ($\times 10^{-3} \text{ mm}^{-2}$, $\times 10^{-3} \text{ mm}^{-4}$)	AED (pixel)
Camera model without lens distortion	$\begin{bmatrix} 1907.74 & 4.32 & 715.96 \\ 0 & 1908.63 & 1082.43 \\ 0 & 0 & 1 \end{bmatrix}$	–	10.67
Camera model with lens distortion	$\begin{bmatrix} 2105.56 & 5.40 & 704.84 \\ 0 & 2107.44 & 1104.10 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} -1.73 \\ 0.0079 \end{bmatrix}$	0.98

Table 3.3: Comparison results of different camera models based on planar calibration object

Image group no.	α (pixel)	β (pixel)	γ (pixel)	u_0 (pixel)	v_0 (pixel)	k_1 ($\times 10^{-3}$ mm $^{-2}$)	k_2 ($\times 10^{-5}$ mm $^{-4}$)	AED (pixel)
1	2105.56	2107.44	5.40	704.84	1104.10	1.73	0.79	0.98
2	2118.50	2121.03	5.66	705.04	1100.79	1.77	0.84	0.93
3	2098.43	2099.53	5.03	704.36	1104.27	1.73	0.78	0.98
4	2108.74	2110.36	5.41	704.42	1100.83	1.77	0.84	0.94
5	2100.69	2102.13	5.46	706.05	1102.85	1.73	0.79	0.96
6	2114.24	2116.09	5.71	706.05	1100.55	1.77	0.83	0.92
7	2095.24	2095.80	5.10	705.94	1102.94	1.73	0.78	0.96
8	2105.74	2106.69	5.43	705.90	1100.43	1.77	0.83	0.93
9	2111.43	2113.08	5.04	704.56	1102.51	1.74	0.80	0.91
10	2116.48	2117.71	5.26	706.61	1099.76	1.77	0.85	0.85
Mean	2107.50	2108.99	5.35	705.38	1101.90	1.75	0.81	0.94
Std	7.81	8.22	0.23	0.81	1.62	0.020	0.027	0.038

Table 3.4: Camera calibration results based on planar calibration object with lens distortion considered

Figure 3.11: Corrected image. 'x' – corners, '•' – corrected corners.

From the estimated intrinsic parameters, the image distortion was corrected based on equation (3.32), as illustrated in Figure 3.11, where the curve lines were corrected to the straight lines. The cross symbols are the extracted real image points (curve line). The dot symbols are the corrected image points (the ideal non-distorted image points, straight line). It can be observed that the distortion is small at the center of the image, and becomes larger gradually at peripheral area of the image. The experimental results show the distortion is nearly zero in the center of the image and approximately 60 pixels in the peripheral area.

Image group no.	α (pixel)	β (pixel)	γ (pixel)	u_0 (pixel)	v_0 (pixel)	k_1 ($\times 10^{-3}$ mm ⁻²)	k_2 ($\times 10^{-5}$ mm ⁻⁴)	AED (pixel)
1	2195.57	2190.88	5.40	708.27	1087.72	-1.75	0.83	0.92
2	2189.83	2185.36	5.52	708.63	1089.22	-1.77	0.86	0.88
3	2168.94	2164.03	5.35	708.02	1087.17	-1.74	0.81	0.87
4	2181.44	2177.43	5.35	708.69	1088.08	-1.76	0.83	0.84
5	2204.55	2200.28	5.67	709.06	1088.94	-1.75	0.83	0.93
6	2192.30	2188.09	5.97	709.62	1089.93	-1.76	0.85	0.91
7	2176.44	2172.06	5.45	707.95	1088.17	-1.73	0.80	0.90
8	2186.55	2183.05	5.47	709.05	1089.27	-1.76	0.82	0.87
9	2180.76	2177.07	5.15	709.08	1089.76	-1.77	0.85	0.90
10	2189.77	2186.60	5.35	708.82	1090.13	-1.78	0.84	0.83
Mean	2186.61	2182.48	5.47	708.72	1088.84	-1.76	0.83	0.89
Std	10.17	10.26	0.22	0.52	1.0075	0.014	0.018	0.032

Table 3.5: Camera calibration results at location of 190mm with lens distortion considered

In the above-mentioned experiments, the calibration object was placed approximately 100mm in front of the objective lens of the camera. Next, the calibration object was moved to the location of approximately 190mm away from the objective lens of the camera. The camera was properly adjusted to make the image clear. Eight more images

were then captured, with 2 from the straight front view, 3 from the left front view, and the other 3 from the right front view. Similarly, the results of ten three-image combinations based on the selection criterion mentioned earlier are given in Table 3.5. It can be observed that standard deviations of all intrinsic parameters are close to the results measured from the images which were captured at the distance of 100mm from the objective lens. It implies that the location of the calibration object would not greatly affect the stability of this method.

3.2.3.3 Comparisons and Discussions

Two camera calibration methods were described based on 3-D calibration object and planar calibration object respectively. The advantage of the 3-D-object-based calibration method is that the calibration procedure is simple, in which only one image is needed. Both linear-method computation and optimization complexity are simpler than the method based on planar calibration object. The problem is that it is difficult to acquire the precise 3-D calibration object. Generally the angle between two orthogonal planes is not equal to 90° exactly. In addition, the three-dimensional coordinates of the object points are difficult to measure accurately. The imprecise calibration object will lead to the inaccurate results.

The planar-object-based calibration method is flexible, but a little complex. At least three images are needed in this method, whose poses and positions are not needed be known. There are two advantages of this method. Firstly, the planar calibration object is easy to be acquired in the laboratory, such as glass plane, compared with the precise 3-D calibration object which must be specially made. Hence the planar calibration object method is more economical. Secondly, in the observation of Table 3.2 and Table 3.4, both the average error distance and the standard deviation of every intrinsic parameter

estimated using the planar-object-based calibration method are less than those of the 3-D-object-based calibration method. Thus it shows that this method is more stable and accurate. The comparison of the advantages and disadvantages of these two methods is given in Table 3.6.

	Object requirement	Complexity	Image requirement	Accuracy	Stability
3-D-object-based calibration method	High precision	Low	One	Low	Low
Planar-object-based calibration method	Low precision	High	Three	High	High

Table 3.6: Comparison of two camera calibration methods

3.2.4 Conclusion

Two photogrammetric calibration methods based on 3-D calibration object and planar calibration object were presented, analyzed and compared. The Canon non-mydiatic fundus camera model CR6-45NM is the optical system under investigation. The linear algorithm was first used to obtain the initial guess, and then the nonlinear optimization was applied. Lens distortion was considered and finally optimized together with the other camera intrinsic parameters. L-M method was implemented to solve the nonlinear optimization problem of camera calibration.

The experimental results show the camera model with lens distortion is a more suitable model for the fundus camera to estimate the intrinsic parameters. The lens distortion is small in the center of the image and larger in the peripheral area of the image. The location of the calibration object does not greatly affect the stability of the calibration method.

From the comparison of the numerical results of two methods it can be summarized that the 3-D-object-based calibration method is a simple but unstable and less accurate method if used without a very precise 3-D calibration object. On the contrary, the planar-object-based calibration method is more stable and accurate although it requires a little more complexity. Based on these characteristics, the method based on the planar object is chosen in this research work to estimate the camera intrinsic parameters for further use in the following 3-D reconstruction.

3.3 Study of Stereo Fundus Camera

The camera to be studied is Nidek 3-Dx stereo fundus camera. It can photograph a fundus and gives two fundus images simultaneously on a single film. The illumination light is projected through the intraocular tissue. The light is reflected by the fundus and goes back to the camera through the objective lens. Then it is split and shifted into two beams by prism and projected on the film to create a pair of stereo fundus image. Compared with monocular fundus camera, stereo fundus camera has more complex optical configuration due to the addition of the prism.

The stereo fundus camera has a parallel stereo configuration as introduced in section 3.1.3. The focal length and baseline are the two intrinsic parameters to be estimated for 3-D reconstruction. The baseline is fixed at 3mm by the manufacturer [103]. The complex inner structure of the stereo fundus camera as shown in Figure 2.10 can be simplified as two separately optical systems (left optical system and right optical system). When capturing the image, the two optical systems are focused simultaneously by adjusting a single knob; moreover, in the two optical systems, the distances between the fundus object and the image are the same. Hence, it is assumed that the focal lengths from the left

viewpoint and the right viewpoint are the same. The average value of left and right focal lengths is set to be the final focal length. The planar-object-based camera-calibration method is applied on the stereo fundus camera to estimate the left and right focal lengths separately.

(1) Left direction 1

(2) Left direction 2

(3) Left direction 3

(4) Middle direction 1

(5) Middle direction 2

(6) Right direction 1

(7) Right direction 2

Figure 3.12: Stereo images of planar calibration object. Corners are marked by 'x'.

Image group no	α (pixel)	β (pixel)	γ (pixel)	u_0 (pixel)	v_0 (pixel)	k_1 ($\times 10^{-3}$ mm ⁻²)	k_2 ($\times 10^{-5}$ mm ⁻⁴)	AED (pixel)
1	2511.56	2498.73	3.82	632.75	898.84	0.66	-1.75	3.81
2	2501.61	2487.81	3.51	631.12	905.00	0.58	-1.59	3.93
3	2315.12	2302.23	7.13	650.91	834.17	0.22	-1.14	3.37
4	2565.96	2549.30	11.07	625.52	941.34	0.11	-0.30	3.46
5	2538.07	2525.31	7.54	644.44	910.28	0.51	-1.45	3.33
Mean	2486.46	2472.68	6.61	636.95	897.93	0.42	-1.25	3.58
Std	98.99	98.23	3.10	10.41	39.22	0.23	0.57	0.27

Table 3.7: Stereo camera calibration results of left viewpoint

Image group no	α (pixel)	β (pixel)	γ (pixel)	u_0 (pixel)	v_0 (pixel)	k_1 ($\times 10^{-3}$ mm ⁻²)	k_2 ($\times 10^{-5}$ mm ⁻⁴)	AED (pixel)
1	2597.57	2610.11	-0.45	730.93	1813.99	-0.85	-1.39	2.00
2	2343.43	2336.35	9.89	722.00	1718.31	0.090	-5.18	2.14
3	2472.86	2474.64	2.13	722.85	1808.54	-0.88	-1.42	2.07
4	2617.32	2612.32	2.39	715.79	1691.22	1.37	-9.67	2.22
5	2354.31	2344.59	11.49	721.31	1693.84	0.85	-8.24	2.19
Mean	2477.10	2475.60	5.09	722.57	1745.18	0.12	-5.18	2.12
Std	129.55	135.40	5.26	5.42	61.27	1.00	3.81	0.089

Table 3.8: Stereo camera calibration results of right viewpoint

Seven pairs of images were captured from different angles of view as illustrated in Figure 3.12 where 3 were taken from the straight front view (middle direction), 2 from the left front view (left direction), and the other 2 from the right front view (right direction). Five three-image combinations were randomly selected based on the selection criterion. The left optical system (left viewpoint) and right optical system (right viewpoint) were calibrated respectively using planar-object-based camera-calibration method. The results are presented in Table 3.7 and Table 3.8 where α and β are the focal length depicted in pixels along u and v axes respectively. The final focal length was measured to be 2481.78 pixels, which is the average value of α estimated from both left and right viewpoints.

It can be observed that the standard deviations of all the intrinsic parameters are much higher than the standard deviations measured from the monocular fundus camera. It could be because the stereo camera is made more sensitive to the variation of object-plane rotation due to the prism incorporated. Other distortion models, such as decentering distortion model or prism distortion model [35], were also tested; however all of them could not converge in the optimization of intrinsic parameters. The standard deviations estimated from the left viewpoints were lower than the ones estimated from the right viewpoints, which implied the left image is more stable than the right image. The reason is that the left image is the main image when capturing an image pair. The camera is first located and focused based on the left image observed from the eyepiece, and then its position is slightly adjusted to make both left image and right image clear. Hence, the right image is not as stable as the left image. It is also the reason why the measured focal lengths (α and β) of the left and right images in Table 3.7 and Table 3.8 are not exactly same.

Figure 3.13: Corrected stereo images. 'x' – corners, '•' – corrected corners.

After estimating the intrinsic parameters, the image distortion was corrected as shown in Figure 3.13, where 'x' symbols were the extracted corners and '•' symbols were corresponding non-distorted image points. It can be observed that the distortion is small at the centers and at the four sides of the left and right rectangle images, and it becomes larger at the corners. The right-bottom corner of left image and left-bottom corner of the right image have larger distortion than other corners, which may due to the splitting and shifting of the light by prism in imaging process. The distortion was measured to be approximately 20 pixels at the left-bottom corner of the right image, less than 2 pixels in the peripheral areas of vertical sides of two rectangle images, and almost zero at the centers of both images. The distortion is much smaller than that of monocular fundus camera, because the stereo fundus camera under study is a narrow-angle camera, whereas the monocular fundus camera is comparatively a wide-angle camera. The viewing angle

of Nidek 3-Dx stereo fundus camera is only 20° in horizontal direction and 25° in vertical direction, whereas the viewing angle of Canon CR6-45NM is 45° . It can be concluded that the narrow-angle fundus camera has small distortion whereas the wide-angle fundus camera has large distortion.

It can be observed that the real stereo fundus camera also has a slight vertical disparity, which makes the correspondences not locating on the same horizontal scan-line. This problem will be considered in Chapter 4 under the 3-D reconstruction.

Chapter 4

3-D Optic Disk Reconstruction

4.1 Introduction

Optic disk is the entrance region of optic nerves and blood vessels to the retina. Its shape, color, and 3-D dimension are often studied for clues of some eye diseases. Glaucoma in particular is the eye disease that damages the optic fiber, and thus changes the 3-D structure of the optic disk. Ophthalmologists usually quantify the cupping of optic disk to evaluate the progression of glaucoma. True color 3-D visualization of the optic disk may not be only useful for clinicians to evaluate the optic disk condition, but it could also make the measurement of disk parameters more accurate. The existing ophthalmic instruments used to measure the depth information, such as HRT and OCT, could not provide the true color 3-D visualization of the optic disk. The 2-D color fundus image is still referred to by most clinicians to manually or semi-automatically extract the optic disk boundary and its parameters. Hence, a computer-aided reconstruction of a true color 3-D

optic disk can be very useful, especially in the current situation where computer is a more accessible tool.

In this research, the true color 3-D optic disk is reconstructed based on a pair of stereo color fundus images to provide clinicians another choice for the analysis of optic disk. The prime problems to be solved in 3-D reconstruction are camera calibration, correspondence search, depth recovery, and eye-optics component estimation. Camera calibration has been introduced in Chapter 3. The main issues in this chapter concern sparse correspondence search and dense depth recovery. Compared with the common 3-D reconstruction of other objects, small depth variation and the optical effect of light media within the eye are the main challenges of 3-D optic disk reconstruction. Small depth variation makes it difficult to precisely detect matching points. In addition, the common stereo technique may not work well, because the stereo fundus images are obtained through both the camera lens and the media inside the eye. The media within the eye will affect the whole optics in imaging, which should be considered in reconstruction.

So far only few researchers focused their work on full 3-D optic disk reconstruction. Some methods of fundus image correspondence search make use of the sparse bifurcations of blood vessels to globally estimate a defined transformation model [52][59][62][63]. The objective of these projects was to fuse multiple fundus images to a single wide-angle image. Hence, the variation of depth was ignored when setting up the transformation model. In 3-D optic disk reconstruction, estimating depth information is the main objective; hence, the correspondence search of two images cannot be simplified as the optimization of a certain transformation. The reasonable approach is to rectify two images into parallel stereo configuration, and then detect the corresponding points by certain local matching method to compute the depth [41]-[44]. Only correspondence search was particularly emphasized in above-mentioned literatures; camera calibration

and the effect of eye-optics were not considered. A full retinal surface reconstruction approach was described in [40][45], where optical effect inside the eye was considered to optimize a spherical retina model. However, the spherical model of retinal surface could not sufficiently display some abnormal features, thus it might not be adequate for diagnosis. The 3-D optic disk reconstruction is still an unsolved problem which is being investigated.

A full 3-D optic disk reconstruction approach is proposed in this research. A fast, sub-pixel, combined matching method is introduced to precisely extract correspondences. As introduced in Chapter 2, local matching method is a suitable choice in this research. However, every individual local matching method has its own advantages and disadvantages, and it cannot guarantee to accurately obtain every correspondence in a pair of images, even if the constraints and hierarchical operation are used to help reduce the error. Hence a novel matching approach is proposed in this research work based on the combination of two constraint-based local matching methods, where the window-based matching method and feature-based matching method are fused to prune uncertain matching and get more precise result. The extracted correspondences are used to recover the depths. Optical media within the eyeball is considered and included by modeling the eye as a single lens. A new method is proposed to calibrate and integrate this effect into the reconstruction process to provide a real 3-D image of the optic disk.

The proposed 3-D reconstruction method is introduced in section 4.2. Section 4.3 describes how to evaluate the reconstructed 3-D image. The optical effect of light media within the eye is considered in section 4.4 to modify the recovered depths. The results and discussions are presented in section 4.5 and section 4.6 gives a conclusion of this chapter.

4.2 Methodology of 3-D Reconstruction

In Chapter 3, the stereo fundus camera was studied, where the lens distortion was measured to be very small in the center of the image and less than 2 pixels at the peripheral region of the image. Moreover, the optic disk is generally focused at the center of the image when capturing the stereo fundus images for checking of glaucoma. Hence, lens distortion is ignored in this chapter. Then, precisely measuring dense depths is one of the most challenging steps in 3-D reconstruction. Here a new auto-adjusted, sub-pixel matching method which combines advantages of these two methods, i.e., the window-based and feature-based methods, is proposed. A sub-pixel sparse matching method is used to obtain more accurate correspondences and also to compensate for the low resolution problem in the image. To make the process faster the sparse correspondences are employed and then followed by the piecewise cubic interpolation to obtain dense depths. Auto-adjustment of the searching range and multiple windows are introduced to drop the outliers, enhance the details and speed up the computation time.

4.2.1 Background of Stereo Pair System

In generally, there are rotation and translation between two cameras in stereo configuration. The corresponding points between two images must satisfy the epipolar constraint. For each given point, its corresponding point can be detected along the epipolar line on the other image. A special case of the stereo camera rig is the parallel stereo configuration, in which there is no rotation between two cameras and the translation is only along v axis as shown in Figure 4.1. The corresponding epipolar lines are aligned with horizontal coordinate axis (scan-line) and at the same height in the left and right images. Hence the corresponding point is taken place only along the v axis

instead of the whole image. Our stereo fundus camera has this special structure as introduced in Chapter 2, and accordingly the depth is inversely proportional to the disparity between the two corresponding points.

Figure 4.1: Parallel stereo image configuration

The object point $\mathbf{M}(x, y, z)$, expressed according to origin $\mathbf{O}$, is projected onto the left image at $\mathbf{m}_L(u_L, v_L)$ and the right image at $\mathbf{m}_R(u_R, v_R)$ as shown in Figure 4.1, where (u_L, v_L) and (u_R, v_R) indicate the row and column numbers in the left image and the right image respectively. $\mathbf{m}_L(u_L, v_L)$ and $\mathbf{m}_R(u_R, v_R)$ are called corresponding points or matching points. Assume the principal points on the left and right images are both at the image centers, according to equation (3.18), the depth for each corresponding pair is:

$$z = \frac{bf}{(v_L - v_R)} \quad (4.1)$$

where the term $(v_L - v_R)$ is disparity, b is the baseline which is the distance between two camera centers, f is the focal length of the two cameras. The units of $b, f, z, (u_L, v_L), (u_R, v_R)$, are all in pixels. Focal length has been measured to be 2481.78 pixels in Chapter 3, and the baseline is found to be 3mm constantly in stereo fundus camera. Therefore, the essential parts for reconstruction are correspondence search and dense depth recovery.

4.2.2 Preprocessing and Initial Disparity Detection

The fundus image has its own unique properties. Firstly, the depth of the optic disk is rather small, approximately a few hundred microns. Hence the disparity change in a pair of images is only in a few pixels. This is one difficulty in retinal surface reconstruction compared with a common object 3-D reconstruction. Sub-pixel measurement is therefore quite necessary to distinguish the disparity difference. Because of the small disparity change, the detection speed can be increased by searching the matching point in one previously set range instead of the whole row of the image. Secondly, there is no occlusion in the fundus and it is a relatively smooth surface. Thus a sharply changed disparity can be dropped as a wrong result in the processing. This property is used to automatically adjust the searching range, which will be introduced in section 4.2.4.

A pair of color fundus images is captured by a stereo fundus camera. In the preprocessing step, the color image is first converted to gray image, and then the image intensity is stretched to the range of $[0, 1]$ to enhance the contrast. The main components on fundus image are optic disk and blood vessels. The optic disk is the bright region which has quite different intensity value compared with blood vessels and background. The difference of optic disk location in the left and right images could be used to measure

the initial disparity. Assuming that the intensity values along each column in the image have certain distribution, obviously, the column contains more bright pixels, if it passes the optic disk region. Then each column will have different distribution. The feature of each column is extracted to set up a feature vector to estimate the disparity between the left and right images.

A string of standard deviation (STD) of each column creates a STD vector which represents an effective property of the image. The initial disparity between stereo pair can be measured by shifting and comparing two STD vectors to obtain the best matched location. For example, the column's STD of the left (right) image is computed and recorded in a vector STD_L (STD_R) as shown in Figure 4.2(a). While setting the STD_L as the reference vector, the vector STD_R is then shifted pixel by pixel labeled by m , and the average absolute difference of intersecting part of vector STD_L and vector STD_R , as shown in Figure 4.2(b), is computed. The shifting value m with minimal average absolute difference is the initially estimated disparity, denoted by t_0 . The disparity can also be measured by detecting the location with maximal correlation of STD_L and STD_R . To get the correct initial disparity, a priori knowledge of the image should be used to set the disparity searching range. In stereo fundus image pair, the disparity is generally a few dozen pixels, no larger than 40 pixels, hence the shifting range in disparity detection is set from zero to one-quarter width of the image ($0 \leq m \leq N/4$). The proposed simple and fast STD method is quite suitable for the initial rough disparity detection. It is suitable to the stereo pair with relatively small baseline and nonuniform texture. For the stereo pair with large baseline, the disparity may be quite large. Then the intersecting part of vector STD_L and vector STD_R is relatively small, which makes the disparity estimation less accurate. In the case of uniform texture image, the STD of each column is similar, which cannot be used to distinguish and measure the disparity.

(a) STD_L and STD_R . Solid line is STD of the left image; dash line indicates STD of the right image.

(b) STD method

Figure 4.2: STD method for initial disparity detection

(a) Left and right vectors of uniformities

(b) Left and right vectors of entropies

Figure 4.3: Feature vectors of left image and right image

Some other features could also be extracted from the columns to estimate the initial disparity, such as mean, entropy, uniformity, etc. The vector of means would only convey the brightness; it is sensitive to the illumination. Moreover, if the optic disk has the similar color to the background, which may happen for some eye diseases, vector of means could not provide effective disparity. Entropy, uniformity and STD do not change with the uniform variation of illumination. Since the detection of initial disparity is an approximate estimation, and the disparity of individual pixel will be refined in the following process, each of these three features could be used here to detect the initial disparity. However, compared with STD, entropy and uniformity are a little more sensitive to blood vessels and noises, as shown in Figure 4.3. It can be observed that the slight variation between adjacent columns can be detected by uniformity and entropy. Hence, vector of STDs is the first choice in this research to estimate the initial disparity.

In stereo fundus images, the corresponding points may not lie on the same horizontal scan-line, since there is a small vertical disparity (a few pixels) between the image pair. Based on the inner structure of the stereo fundus camera, the vertical disparity only comes from the global position shift along u axis because there is no rotation in stereo camera rig. Some researchers introduced relatively complex rectification method [104] for the general stereo rig, from which the rectification was realized by optimizing 2-D projective transforms or homographies and applying them to each image. This method can rectify both rotation and translation. To the parallel stereo rig, the pure global vertical translation can be fast rectified by using the proposed STD method. The STD method is applied in each row to detect the vertical disparity. The left image is set as reference image, and then the right image is shifted up or down based on detected vertical disparity for rectification and making the corresponding points located at the same horizontal scan-line. The matching point detection in the whole image is then reduced to one dimension detection

which is only along the same horizontal line. The vertical disparity can be auto-adjusted later based on the correspondences, which will be later introduced in section 4.2.4.

4.2.3 Combined Matching

Figure 4.4: Correspondence search

One point is selected from the left image as the reference point. Based on the background knowledge of stereo pair system, the corresponding point in the right image must lie on the same horizontal scan-line (epipolar line) as the reference point. A suitable searching range is set in the right image (the dash rectangle in Figure 4.4). Each point on the epipolar line in the right image inside the searching range called candidate point is compared with the reference point, as shown in Figure 4.4. The intensity value of an individual pixel does not give sufficient information; hence the surrounding pixels inside the windows (the solid squares in Figure 4.4) centered at reference point and candidate points are picked up for comparison. Generally, a 7×7 , 9×9 or 11×11 window may be used [65]. The window size also depends on the image size. The most similar point among the candidate points is the best matching point. Two operations (correlation method and feature-based method) are applied to detect the best matching point, and then the resulted two matching points are compared. If the difference of the two matching points from two

methods is less than one pixel, the average location is the final corresponding point to the reference point. Otherwise, there is no suitable corresponding point to the reference point. Then the next point is selected from the reference image; the operation is repeated until all the corresponding points are calculated.

4.2.3.1 Normalized Cross Correlation Method

In this method the corresponding point is computed by fixing a small window around a reference pixel in the left image, and measuring its correlations with the candidate pixels in the right image within the same window size. The candidate pixel with highest correlation is the corresponding point.

The normalized cross correlation between two points is

$$corr(\mathbf{m}_L, \mathbf{m}_R) = \frac{\sum_{i=-W}^W \sum_{j=-W}^W [g(u_L + i, v_L + j) - \bar{g}][h(u_R + i, v_R + j) - \bar{h}]}{\sqrt{\sum_{i=-W}^W \sum_{j=-W}^W [g(u_L + i, v_L + j) - \bar{g}]^2 \sum_{i=-W}^W \sum_{j=-W}^W [h(u_R + i, v_R + j) - \bar{h}]^2}} \quad (4.2)$$

Here, g and h are the intensity values of left and right images, $(2W+1)$ is the comparing window size, $\bar{g}$ and $\bar{h}$ are the average intensity values in the comparing windows, $\mathbf{m}_L(u_L, v_L)$ is the reference point in left image and $\mathbf{m}_R(u_R, v_R)$ is the candidate point in the right image. In the searching range, the point with maximal correlation is the best matching point.

4.2.3.2 Feature-based Method

Several features are extracted from the image pair. The average absolute differences of the features between the reference pixel in the left image and candidate pixels in the

right image are measured. The candidate pixel with minimal difference is the corresponding point. The conventional features are intensity, x gradient, y gradient, gradient magnitude, gradient orientation, curvature and so on.

In this research work, three features are extracted in each comparing window. They are intensity, x gradient magnitude and gradient orientation. Every feature is normalized to the range of $[0, 1]$. The average absolute difference of these features between two windows are computed as

$$FeatDiff = \frac{1}{3}|Diff_{int}| + \frac{1}{3}|Diff_{xgrd}| + \frac{1}{3}|Diff_{ori}| \quad (4.3)$$

In the searching range, the point that has the minimal feature difference is considered as the best matching point.

4.2.3.3 Sub-pixel Disparity and Constraint-based Modification

Sub-pixel matching point must be considered to obtain more accurate result. It also compensates for the low resolution problem of the image. The sub-pixel matching point is estimated by interpolation (but not approximation) from the neighboring pixels, since the interpolation gives more accurate results. What should be considered in this estimation is if the interpolation (curve fitting) is accurate, while whether the neighboring pixels are accurate to the true positions is not considered. The neighboring pixel may be accurate or noisy. Curve fitting is used for accurate estimation of the sub-pixel matching point. The curve must pass all the neighboring points. 5 neighbours are used for estimation; hence, fourth order polynomial fitting is chosen to compute the 5 parameters of the polynomial. Whether the estimated sub-pixel matching point is accurate or inaccurate to the true position will be evaluated by the following operation. If the 5 neighboring points are accurate determined, the estimated sub-pixel matching point is accurate. On the contrary,

if the 5 neighboring points are noisy, the result from the interpolation may be inaccurate. The matching point may be dropped as “bad matching point” from the constraint-based modification process, which is introduced in the follows. The algorithm ensures only accurate matching points remain, while all the other matching points (inaccurate points) are dropped. The number of neighbours used for interpolation can be changed into some other values, such as 7 points, etc.

Fourth order polynomial fitting is used to obtain the sub-pixel location. For example, based on equation (4.3) in feature-based method, the feature difference of each candidate point along epipolar line (v direction) of the right image in searching range is plotted in Figure 4.5, where $FeatDiff(v)$ is the average absolute difference of three features between the reference point in the left image and a candidate point in the right image. All the local minima in $FeatDiff(v)$ as shown in Figure 4.5 are detected, among which the pixels with the minimal $FeatDiff$ and the second minimal $FeatDiff$ are measured and recorded as first minimum and second minimum, as the follows:

$$\begin{aligned} v_{First\ Minimum} &= \arg \min_{v \in Local\ Minima} \{FeatDiff(v)\} \\ v_{Second\ Minimum} &= \arg \text{second min}_{v \in Local\ Minima} \{FeatDiff(v)\} \end{aligned} \quad (4.4)$$

Five pixels centered at each minimum are picked up as the star points shown in the second image of Figure 4.5 to optimize the parameters of the fourth order polynomial. Then the local minimum's location can be modified to sub-pixel level by the optimized polynomial. The continuous line and circle point (at pixel 23.2) in the second image of Figure 4.5 are the optimized polynomial and modified sub-pixel minimum respectively. Both the first and second sub-pixel local minima can be obtained. Similarly, the first and second local maxima are measured in correlation method.

Figure 4.5: Sub-pixel disparity

Two difference constraints are introduced here to drop the potentially uncertain matching point. The first constraint is operated inside the individual feature-based method or correlation method; and the second constraint is operated between two matching methods. The first difference constraint is based on the $FeatDiff$ values of the first and second minima in feature-based method (or in correlation method). The first minimum is considered the best matching point. However due to noise in the image pair, the point with the minimal $FeatDiff$ may not be the correct matching point. Further processing is

therefore needed to prune the incorrect points. The *FeatDiff* value difference between the first and second minima is computed as a criterion for selection, as shown in Figure 4.5. If the difference $FeatDiff(v_{Second\ Minimum}) - FeatDiff(v_{First\ Minimum})$ is larger than a threshold T_1 , the first minimum is the good matching point. Otherwise, the first minimum is arbitrarily considered as a false matching point which will not be used for the following computation. Similar constraint is also applied in the correlation method.

The second difference constraint is based on the locations of the two sub-pixel matching points. After applying the first difference constraint, some uncertain matching points are dropped. Hence, for each reference point in the left image, there are three cases of detected matching points in the right image, e.g., two best matching points from two methods are remains, one best matching point is the remain, and none of any best matching point is the remain. Then only those reference points, each with two sub-pixel best matching points from two methods are considered in the following computation. The location difference between the two best matching points from two methods is calculated. In the second difference criterion, if the difference is larger than $T_2 = 1$ pixel, this corresponding pair is dropped; otherwise the final matching point is set to be the average location of the two matching points. Finally a pair of sub-pixel correspondence is obtained.

4.2.4 Auto-adjustments of Searching Range and Vertical Disparity

From step 4.2.2 to 4.2.3, the sparse accurate matching points are obtained. However the computation is only based on the fixed searching range and initial vertical rectification. Both of these affect the matching result. Hence two adjustment methods are described here to auto-adjust the searching range and vertical disparity.

The disparity map $disp(u, v)$ is used to record correspondences based on the reference image (the left image). For a matching pair (u_L^*, v_L^*) and (u_R^*, v_R^*) , its disparity $(v_L^* - v_R^*)$ is computed and recorded at $disp(u_L^*, v_L^*)$. The searching range for the correspondences of the whole image can be convertibly written as disparity searching range $[dl, dr]$, where dl is the left disparity searching boundary and dr is the right disparity searching boundary. Setting a suitable searching range can reduce the computation time considerably. Moreover, relatively small searching range has lower probability to obtain incorrect matching point. The retina has a smooth surface without sharp changes; hence the disparity between image pairs only changes gradually. The searching range can be increased gradually from the initial range based on the previously detected disparities in each iteration (repeating step 4.2.2 to 4.2.3). If the detected disparity is very close to the disparity searching boundary, the boundary will be extended by two pixels. Let the initial disparity searching range be $[dl, dr]$. The sparse disparities as computed using the method introduced in section 4.2.2 to 4.2.3, are denoted as $disp(u_L, v_L)$ based on the reference image, where (u_L, v_L) is the reference point having the final matching point determined. The minimum and the maximum disparities are recorded as $disp_{\min} = \min\{disp(u_L, v_L)\}$ and $disp_{\max} = \max\{disp(u_L, v_L)\}$. If the $disp_{\min}$ or $disp_{\max}$ is close to the previous disparity searching boundary, the left and right searching boundaries are updated as the follows:

$$\begin{cases} dl(n+1) = dl(n) - 2 & \text{if } [disp_{\min}(n) - dl(n)] < 2 \text{ pixels} \\ dl(n+1) = dl(n) & \text{else} \end{cases}$$

$$\begin{cases} dr(n+1) = dr(n) + 2 & \text{if } [dr(n) - disp_{\max}(n)] < 2 \text{ pixels} \\ dr(n+1) = dr(n) & \text{else} \end{cases} \quad (4.5)$$

where n is the iteration times in searching range updating. The adjustment stops when both dl and dr are unchanged in the iteration. Then the final searching range is obtained. The process of searching range adjustment is depicted in a close loop block diagram as shown in Figure 4.6.

Figure 4.6: Process of searching range adjustment

The vertical disparity is adjusted by slightly shifting it up and down and then repeat section 4.2.2 and section 4.2.3 to obtain sparse correspondences. The total number of the available sparse correspondences is recorded for each vertical disparity shifting. The final vertical disparity is located at the place where the correspondence number is maximal.

4.2.5 Dense Depth Recovery Based on Multiple Windows

In order to reconstruct the 3-D retinal surface, the depth must be known for each pixel in the reference image. However, some correspondence pairs are dropped because they do not satisfy the constraints. Moreover, to reduce the computation time, not all the points in the left image are computed. A set of regular grid lines with separating space of Δx pixels in row direction and Δy pixels in column direction is established on the reference image as shown in Figure 4.7. Those crossing points on the grid lines are used

as reference points in the proposed method. Additionally, the edge points extracted by Canny edge detection on every row of grid lines are also considered as the reference points. Hence, after detecting the matching points, only sparse and precise correspondence pairs are obtained. Based on equation (4.1), the correspondence pairs are converted into sparse depths.

Figure 4.7: Stereo image pair with regular grid for correspondence search

To guarantee the correspondences are set up for most reference points, multiple windows are applied on disparity detection. A large window makes both correlation method and feature-based method more stable and robust to the noise. And also it can effectively detect the corresponding point in the feature-less area, because more surrounding pixels are put into computation. However the details of the disparity variation may not be detected by large window process. Hence the large window is used to obtain basic result and then small window is applied on the basic result to extract further details. Some matching points may not be established by a single small window. For the reference

point whose matching point cannot be established (generally in feature-less area), the window size is increased and computation is repeated until the available matching point is detected. The correspondence search with large window is used to complement the missing matching in the previous correspondence search with small window. Using various window sizes guarantees that matching for most of the reference points can be established. In this application, the window increase will stop when 80% correspondences are reached. Then the disparities on regular grid points and selected edge points are converted into sparse depths. One dimensional piecewise cubic interpolation is applied on every row and column of grid lines to obtain the continuous depths. This is later applied in the whole image to obtain the remaining rows and columns' depths. When each pixel's depth information is known, a median filter is used to reduce the noise and smooth the surface. The median filter can effectively reduce the noise and preserve the useful details of original 3-D image.

Another method to obtain dense depths is piecewise surface-fitting. The surface within a small region can be simplified as a quadratic surface. The sparse depths inside the region are used to optimize the parameters of quadratic surface. Then the dense depths are obtained from the combination of all optimized small quadratic surfaces. The result of the testing image shows that it is a computationally expensive work, whereas the accuracy of dense depths from 2-D quadratic surface-fitting is no better than that from 1-D interpolation. Hence the interpolation is selected in this research to get the final dense depths. The examples of 2-D quadratic surface-fitting and 1-D interpolation are given in section 4.5. The basic steps for sparse correspondence detection and dense depth recovery are summarized as follows:

- (1) The large window (31×31 pixels), large Δx and Δy are used for fast searching range adjustment and vertical disparity adjustment. Then the basic disparities are obtained based on large window.
- (2) The window is changed into small size (11×11 pixels). Δx and Δy are reduced and fixed in the following computation. Apply step 4.2.2 and 4.2.3 to obtain sparse precise matching points for regular grid points and additional edge points based on the basic disparities and the obtained searching range from step (1).
- (3) Increase window size and re-compute the disparities for the reference points which cannot establish available matching points based on the previous small window. The established matching points from the previous small window are kept.
- (4) Repeat steps (3) until 80% matching is set up for the reference points (regular grid points and edge points).
- (5) Convert sparse disparities into sparse depths, apply piecewise cubic interpolation and median filter to obtain final dense depths.

Once all of the above operations are done, the final 3-D retinal representation is obtained. Figure 4.8 gives out the basic steps of the proposed reconstruction method. The bracketed number in each box in the figure denotes the section identified in this chapter.

Figure 4.8: The main procedure of 3-D reconstruction

4.3 Evaluation

The accuracy of the reconstructed 3-D surface is evaluated by comparing the original right image with the recovered right image which is calculated from the left image and recovered 3-D depth information. Each pixel $\mathbf{m}_L(u_L, v_L)$ in the left image is back

projected to $\mathbf{m}'_R(u'_R, v'_R)$ in the right image based on the three dimensional depth of this pixel, written as:

$$\begin{cases} u'_R = u_L \\ v'_R = v_L - bf/z \end{cases} \quad (4.6)$$

The recovered right image denoted as R' is compared with the original right image denoted as R by computing their normalized cross correlation using equation (4.2) as the following:

$$correlation = corr(R, R') \quad (4.7)$$

The correlation of two images is normalized to the range of $[-1, 1]$. The higher the correlation between two images is, the better they match, which means the reconstruction is more accurate. If the correlation of two images is 1, it indicates that the two images are perfectly the same.

4.4 Calibration of Optical Effect of the Eye

4.4.1 Modeling Eye Optical Effect

So far, the effect of eye lens is not considered. Based on the report in literature [40], it is likely that the recovered 3-D image is the image of retina as appeared through the eye lens. If the complex eye structure, including cornea, eye lens, vitreous, and aqueous, etc., is lump-modeled as a single simple lens, the real retinal surface is mapped onto an image called virtual surface. The 3-D image recovered from stereo fundus images in section 4.2 is just this virtual surface as shown in Figure 4.9, where $\mathbf{O}_e$ and f_e are the optical center and focal length of the modeled single eye lens respectively. Let $\mathbf{M}(x, y, z)$ be the point on

recovered 3-D virtual surface according to origin $\mathbf{O}$, $\mathbf{M}(X, Y, Z)$ is the same point based on the origin $\mathbf{O}_e$, whose corresponding point on the real retinal surface is expressed by $\mathbf{Mr}(X_r, Y_r, Z_r)$. A calibration method is proposed in this section to extract the real depths of retina from the virtual surface.

Figure 4.9: Optical model of the eye and stereo fundus camera (adapted from Deguchi's [40])

Let Z_r and Z be the depth values of certain point on retinal surface and its corresponding point on virtual surface respectively, where Z_r and Z are both positive numbers for computational convenience, it gives

$$\frac{1}{Z_r} + \frac{1}{Z} = \frac{1}{f_e} \quad (4.8)$$

Hence, Z can be expressed as

$$Z = \frac{Z_r \cdot f_e}{Z_r - f_e} \quad (4.9)$$

Assume there is a reference plane located at Z_{r0} , which is perpendicular to the optical axis located at Z_{r0} , the corresponding image of this reference plane is located at Z_0 . The

absolute positions of Zr_0 and Z_0 need not to be known. The relative depth in virtual surface is

$$\Delta Z = Z - Z_0 = \frac{Zr \cdot f_e}{Zr - f_e} - \frac{Zr_0 \cdot f_e}{Zr_0 - f_e} \quad (4.10)$$

From the previous equation, the relationship between the relative depth of real retinal surface ΔZr and the relative depth of virtual image ΔZ can be set up, as

$$\Delta Zr = Zr - Zr_0 = \Delta Z \frac{2Zr_0 f_e - f_e^2 - Zr_0^2}{f_e^2 - \Delta Z f_e + \Delta Z Zr_0} \quad (4.11)$$

One example of focal length of the eye, f_e , is reported to be 17.05mm (58.64 diopters) [105]. If Zr_0 is given, the depth of real retinal surface can be measured from the recovered 3-D virtual surface.

4.4.2 Estimation of Zr_0

The maximal cup depth is used to calibrate Zr_0 , which is defined to be the relative depth from the periphery of retinal surface (reference plane) to the lowest region of optic disk. Circle of 2.5mm diameter is placed at the periphery of optic disk to locate the reference plane used to measure the maximal cup depth as shown in Figure 4.10. Maximal cup depth ΔZr as measured from HRT and belonging to a certain pixel will have the corresponding maximal cup depth ΔZ under the same definition in recovered 3-D virtual surface computed as a result from section 4.2. Given ΔZr , ΔZ and f_e , the value of Zr_0 can be computed; and then each pixel's real depth according to the reference plane can be recovered from equation (4.11).

(a) Periphery circle of optic disk

(b) Cross section of optic disk

Figure 4.10: Maximal cup depth

Assume focal length of the eye f_e is fixed; thus, every image has different Zr_0 . For the stereo image pair which has corresponding HRT image, Zr_0 can be directly measured by the above-mentioned method. In the case that corresponding HRT image is not available, the average Zr_0 is used to compute the approximate real 3-D image, which is measured from a group of stereo images and HRT images. Given N pair of ΔZ and ΔZr measured from N stereo image pairs and their corresponding HRT images, then Zr_0 can be estimated from

$$\arg \min_{Zr_0 \in [12mm, 24mm]} \left\{ \sum_n^N |\Delta Zr_n - \Delta Zr_n(\Delta Z_n, Zr_0)|^2 \right\} \quad (4.12)$$

where Zr_0 is restricted within the range of [12mm, 24mm] based on the anatomic structure of the eye. The cornea and real eye lens are in the anterior portion of the eyeball. Hence, the distance of the modeled single eye lens to the reference plane on the retinal surface should be larger than the radius of the eyeball (12mm) and smaller than the diameter of the eyeball (24mm). After Zr_0 is measured, the depths of real 3-D retinal surface can be approximately recovered from equation (4.11) based on the reference plane, as shown in Figure 4.10.

Eight pairs of images were selected for calibration. The selection was based on the criteria of low noise in HRT images and high quality in stereo fundus images. The maximal cup depths measured from HRT and their corresponding values from 3-D virtual surface are given in Table 4.1. ΔZ is negative here because it is expressed according to the optical center O_e . From equation (4.12), average Zr_0 was measured to be 20.14 mm.

No	ΔZr from HRT (mm)	ΔZ from virtual surface(mm)	No	ΔZr from HRT (mm)	ΔZ from virtual surface(mm)
1	0.55	-12.88	5	0.70	-16.49
2	0.46	-9.33	6	1.23	-19.52
3	0.76	-20.92	7	0.66	-17.44
4	1.10	-16.60	8	0.95	-26.18

Table 4.1: Maximal cup depths from HRT and 3-D virtual image

For every 3-D image (virtual surface) computed from section 4.2, substituting $Zr_0=20.14\text{mm}$ back to the equation (4.11), the 3-D image with true depths can be approximately obtained. It provides clinicians a direct 3-D view of optic disk; and also lets them easier to observe and monitor the variation of 3-D optic disk shape for diagnosis.

Moreover, cup boundary could be estimated based on reconstructed 3-D image. Since the 3-D image has approximate true depths, the cup could be defined based on relative depth. Details will be introduced in Chapter 6. If the image pair does have corresponding HRT image, the Z_{r_0} for this individual image pair could be measured, and then 3-D image with real depths could be obtained based on equation (4.11). Thus, the resulted 3-D image has the same maximal cup depth as that HRT image has. In this case, either relative depth or absolute depth can be used as a cup definition in the following estimation of cup boundary.

4.5 Results and Discussions

One hundred pairs of stereo images of optic disk were tested by the proposed algorithm, which were provided by National University Hospital. All the images were resized to 256×272 pixels. A pair of stereo RGB optic disk images is shown as an example in Figure 4.11. It was first converted into gray images. The initial disparity was measured as $t_0 = 32$ pixels by STD method. Based on the understanding of the stereo image pair, the initial searching range was set to be $[(t_0-15) \text{ pixels}, (t_0+15) \text{ pixels}]$. Sub-pixel correspondences were computed and then the disparities were converted to sparse depths.

For comparison, pure feature-based method, pure correlation method and combined matching method were all applied on the image pair to obtain sparse depths based on a fixed large window (21×21 pixels). The results are given out in Figure 4.12. It can be observed clearly that even if based on large window, some bad matching points were still detected by pure feature-based method (Figure 4.12(a)) or pure correlation method (Figure 4.12(b)); whereas the proposed combined matching method can effectively drop the bad matching points (Figure 4.12(c)).

The multiple-window process is illustrated in Figure 4.13. A small window (11×11 pixels) was selected first in combined matching method to obtain accurate sparse depths as shown in Figure 4.13(a). However, much matching was pruned by two difference constraints. Then the multiple windows were applied to make up for more matching which had been dropped in the previous small window operation. Figure 4.13(b) gives the final sparse depth result from multiple windows (11×11 pixels, 21×21 pixels and 31×31 pixels), in which more than 80% matching had been acquired.

The piecewise cubic interpolation and piecewise surface-fitting were both employed on final sparse depths to get the dense depths as shown in Figure 4.14(a), (b), from which the evaluating correlations (from equation (4.7)) are both 0.98. However, the computation time of piecewise surface-fitting is 523.0 seconds long which is much more than that of piecewise cubic interpolation, which is 8.9 seconds long. Hence the piecewise cubic interpolation is the first choice to be used for estimating the dense depth here. An 11×11 pixels median filter was applied on the dense depth image to remove noise and smooth the surface. The virtual depths were then modified to obtain the real depths based on equation (4.11); and the final recovered 3-D retinal surface is shown in Figure 4.15(a), whose correlation factor was evaluated as 0.98. The resulted 3-D image is comparable with the 3-D image measured from HRT as illustrated in Figure 4.15(b). Moreover, it provides true color information on the 3-D image. The entire procedure including sparse correspondence detection, interpolation, smoothing operation and depth modification was completed within the much improved 60.3 seconds computing time with the use of Matlab. For the same sample, the computational costs of the pure feature-based method and pure correlation method were 58.5 seconds and 56.9 seconds respectively, which were quite close to the computational cost of the proposed combined correspondence search method.

Figure 4.11: An example of color stereo fundus image pair

(a) Sparse depths from pure feature-based method

(b) Sparse depths from pure correlation method

(c) Final sparse depths from proposed combined matching method

Figure 4.12: An example of combined matching method on sparse depth mesh

(a) Sparse depths based on small window (11×11 pixels) (b) Sparse depths based on multiple windows (11×11 pixels, 21×21 pixels and 31×31 pixels)

Figure 4.13: An example of multiple window process based on combined matching

(a) The dense depths from piecewise cubic interpolation

(b) Dense depths from piecewise surface-fitting

Figure 4.14: Dense depths by two different operations computing on sparse depths, meshed with color

(a) Recovered 3-D optic disk by the proposed method

(b) 3-D optic disk from HRT

Figure 4.15: An example of 3-D optic disk reconstruction

(a) Color stereo fundus image pair

(b) Recovered 3-D optic disk by the proposed method

(c) 3-D optic disk from HRT

Figure 4.16: One more example of 3-D optic disk reconstruction

One more example is illustrated in Figure 4.16. The stereo images, as shown in Figure 4.16(a), were first converted into gray images. The virtual depths were later estimated by combined matching and then modified to obtain the real depths based on equation (4.11). The maximal cup depths were measured to be 26.18mm and 0.95mm respectively, from the virtual surface and its corresponding HRT image, from which Zr_0 was computed to be 19.80mm. The final recovered 3-D optic disk image is shown in Figure 4.16(b), whose correlation factor was evaluated as 0.96. The resulted 3-D image is comparable with the 3-D image measured from HRT as illustrated in Figure 4.16(c), where the sawtooth shape in the cupping area of HRT image is caused by noise. The 3-D shapes from HRT and proposed method are quite similar from observation. Quantitatively comparing the results of the proposed method with those of HRT will be introduced in Chapter 6, where C/D vertical ratio is used as a criterion of evaluation.

Figure 4.17: Combination of original and recovered right images
Dark squares are from the original right image; bright squares are from the recovered right image.

Table 4.2 gives the correlation factors of 100 stereo pairs. The average correlation was approximately 0.96. A visual example of indirectly evaluating the resulted 3-D image is given in Figure 4.17. The left image was back-projected onto the right image based on the recovered 3-D image as shown in Figure 4.17, where the dark squares and bright squares are from the original and projected right images respectively. The consistent combinations of two right images show the effect the recovered 3-D image. It can be observed that some image pairs in the table have obviously lower correlations (less than 0.90) than all the other image pairs. It is because of nonuniform illumination in the left and right images, which is one limitation of the proposed evaluation method. The image pair that has the lowest correlation factor (0.82) in Table 4.2, named 25L, is illustrated in Figure 4.18 as an example. It can be observed that the left region is dark and the right-top corner is bright in the left image; on the contrary, in the right image, the left-top corner is bright and right region is dark. This nonuniform illumination reduced the correlation value; it might give lower evaluation than the image really had; whereas if the illumination uniformly changed between the left and right image, the correlation would not be affected. This nonuniform illumination should be avoided by properly adjusting the camera when capturing the image pair. In the case that the illumination is obviously nonuniform in the left and right images, the image pair is unavailable which should be re-captured following the camera operation manual. The edge feature of image is robust to the illumination, which could be used as another evaluation method. The edges extracted from the left image were back-projected onto the original right image based on the resulted 3-D image, denoted by the solid lines, as shown in Figure 4.18. The consistent combination of edge map and right image implied the accuracy of the recovered 3-D image. However, this method is difficult to quantitatively evaluate the resulted 3-D image. How to automatically and quantitatively evaluate the results is still being investigated.

Image name	corr	Image name	corr	Image name	corr	Image name	corr
1L	0.97	26_3	0.99	55L	0.87	29R	0.92
2L	0.94	26L	0.99	1R	0.97	30R	0.99
3L	0.97	28L	0.97	2R	0.94	31R	0.99
4L	0.96	29L	0.91	3R	0.95	32R	0.96
5L	0.98	30L	0.99	4R	0.95	33R	0.97
6L	0.99	31L	0.98	5R	0.98	34R	0.98
7L	0.91	33L	0.97	6R	0.98	35R	0.97
8L	0.99	34L	0.98	8R	0.98	36R	0.96
9L	0.98	35L	0.98	9R	0.97	37R	0.97
10L	0.95	36L	0.93	10R	0.92	38R	0.88
11L	0.97	37L	0.92	12R	0.99	39R	0.98
12L	0.97	38L	0.92	14R	0.98	40R	0.98
14L	0.98	39L	0.96	15R	0.96	41R	0.91
15L	0.98	40L	0.98	16R	0.97	42R	0.97
16L	0.95	41L	0.89	17R	0.97	43R	0.98
17L	0.97	42L	0.98	18R	0.97	44R	0.98
18L	0.93	43L	0.98	19R	0.95	45R	0.94
19L	0.97	44L	0.98	20R	0.89	46R	0.97
20L	0.84	45L	0.96	21R	0.86	47R	0.98
22L	0.98	46L	0.97	22R	0.86	48R	0.98
23L	0.97	47L	0.98	23R	0.98	49R	0.98
24L	0.96	48L	0.98	24R	0.96	52R	0.93
25L	0.82	49L	0.98	25R	0.96	53R	0.98
26_1	0.99	52L	0.96	26R	0.97	55R	0.96
26_2	0.99	53L	0.99	28R	0.97	56R	0.97
—	—	—	—	—	—	average	0.96

Table 4.2: Evaluating correlations of 100 testing image pairs, ‘corr’ indicates correlation.

Figure 4.18: An example of nonuniform illumination between the left and right images

4.6 Conclusion

A full 3-D optic disk reconstruction approach has been proposed in this chapter, using a number of efficient steps including camera calibration, constraint-based combined matching, dense-depth recovery, and eye-optics inclusion. The focal length obtained in Chapter 3 is used to recover the virtual depth from matching pair, which is later approximately revised to the real depth by a modeled eye optical system.

A simple and fast STD method is proposed to estimate the initial disparity. The correspondences are measured by a precise and fast combined matching approach. The correlation method and feature-based method are fused together to prune uncertain matching points and to obtain more accurate disparities. This operation utilizes the advantages of both methods, i.e., the accuracy of feature-based method and noise stabilization in the correlation method. It is better than only using any single matching

method without post-processing. Two difference constraints are introduced to drop the outliers of the correspondences. Sub-pixel modification in this method helps compensate for the low resolution problem and also increases the accuracy of the result. Multiple windows guarantee most matching is established for the following interpolation to obtain dense depths.

The optical effect within the eye has been considered in this chapter. The complex eye structure is lump-modeled as a simple lens, and then calibrated by the selected images, using the maximal cup depth from HRT results. The modeled eye system makes it possible to approximately modify the recovered virtual depths to the real depths.

The experimental results show that the proposed method can reconstruct the retinal surface successfully. The test of result is done by comparing the original right image with the recovered right image, which shows a well matched pair. The average correlation between the original and recovered right images is 0.96. Compared with the result from HRT, the 3-D image reconstructed by the proposed method shows good consistency and compatibility which indicates that it could be used as an alternative mode of 3-D viewing of the optic nerve head. The recovered surface can conveniently provide clinicians a direct view of true color 3-D retinal surface from the basic stereo fundus camera. It can be used as an assisting tool for physicians to directly observe the cupping on the retina; and also it is a necessary and basic step for detection of useful optic disk parameters for clinical analysis.

Chapter 5

Optic Disk Boundary Detection

5.1 Introduction

In the computer-aided process of detection and screening of the disease using color fundus images, automated localization of the optic disk and estimation of its boundary are the first two essential steps before any further analysis. Optic disk boundary is an important feature on fundus image to indicate various ophthalmic pathologies; and also it acts as a landmark and reference to other structures. For instance, optic disk boundary is used as a reference contour to estimate other disk parameters, such as cup boundary, C/D ratio, etc., in diagnosis of glaucoma. The accurate detection of optic disk boundary can also be used to monitor the progression of eye diseases.

Practically, the shape of optic disk is assessed manually by ophthalmologists, which is a time consuming job. Many attempts have been made to identify the optic disk location and to detect the optic disk boundary automatically. However, it is not a simple task to segment the optic disk reliably and accurately, as there are problems of blood vessel

occlusions, ill-defined edges, noises, and fuzzy contour shapes due to pathological changes and variable photographic conditions. These problems to a large extent are very similar to other boundary detection and image segmentation problems in the medical imaging area that still require robust solution. The methods of optic disk boundary detection can be categorized into two groups due to different objectives: one is to identify the location of optic disk center and detect the approximate disk shape by means of estimating best-fit circle or ellipse; the other one is to extract the precise boundary of optic disk. Identifying the disk center and its rough boundary is generally used to provide a landmark for estimating the location of other features on retina or setting the starting point in blood vessel tracking. Precise disk boundary can provide more information in diagnosis which is often used to evaluate the condition of optic disk. In this research work, precise disk boundary is required to estimate cup edges for glaucoma analysis.

Many existing approaches can be used to locate the optic disk successfully. Recently, a principal component analysis (PCA) model-based approach was used in [2][76], and template matching was used in [72][80][81]. Hough transform was employed to obtain disk center and the outer circle of disk boundary in [73][75]. Another method is based on the knowledge that the optic disk locates on the cross region of main blood vessels. Although the main blood vessel detection was an complex operation, the main blood vessels were extracted, and the geometric relationship between the optic disk and blood vessels was utilized to identify the disk location in [82]-[84]. Once the location of optic disk is known, it can be used to initialize the disk boundary.

Currently, deformable models have attracted much interest in boundary detection and image segmentation, which can be roughly classified into free-form deformable models (snakes) and parametrically deformable models (ASM), as introduced in Chapter 2. The free-form deformation model could access the accurate positions of edges, but it is

sensitive to noises and vessel occlusions on the contour. In order to reduce such effects, some methods, such as increasing the stiffness coefficient of snake or removing the blood vessels from image, were proposed [71][72][80]. In [71][72], the blood vessels were removed by morphology; and then GVF snake algorithm was applied to detect the exact boundary of optic disk. This algorithm was sensitive to preprocessing. In [80], the disk boundary was segmented by a deformation approach with variable stiffness based on a global elliptic model and a local deformable model. A level set approach to model arbitrarily complex shapes was proposed in [106]. The advantage of this approach was its ability to evolve the model in the presence of sharp corners, cusps, shapes with pieces and holes, etc. However, many of these problems are different from those encountered in the boundary detection of the optic disk. Influence of blood vessels was one of the main difficulties in the detection of the exact optic disk boundary, which was still not well solved in above-mentioned methods. A model-based deformable model (parametrically deformable model) was employed in [2][86][87] to estimate the optic disk boundary based on point distribution model (PDM). This approach could indirectly handle blood vessel occlusion problem by restricting the contour deformation into certain models. The top-down strategy increased the stability of the contour, but might reduce the accuracy, because a number of shape models were difficult to represent every fuzzy shape due to pathological change.

In this chapter, a novel method for fully automated detection of the exact optic disk boundary is proposed. Through the proposed method the above-mentioned problems, especially the influence of blood vessels, are directly solved without affecting the accuracy. The disk boundary is initially estimated by distance transformation, and thereafter the snake technique is implemented with the energy function defined based on the knowledge of optic disk. With this proposed method the original snake which is a

deforming-only process is modified and extended in two aspects. Firstly, after each deformation, the contour points are classified into edge-point cluster or uncertain-point cluster by unsupervised learning, k-means algorithm. Secondly, the contour is updated by variable updating sample numbers. The updating is self-adjusted using both global and local information, so that the balance on contour stability and accuracy can be achieved. Under proper variations, the investigated method may be applied to many other applications with similar problems, such as occlusions on object boundary.

The proposed approach of optic disk boundary detection is introduced in section 5.2. An evaluation criterion is presented in section 5.3. Experimental results and discussions are given out in section 5.4. Section 5.5 concludes the whole chapter.

5.2 Optic Disk Boundary Detection by Modified Snake

Figure 5.1: Flowchart of optic disk boundary detection

The proposed method of optic disk boundary detection includes initialization of optic disk boundary, contour deformation, clustering and smoothing update, which is depicted in a closed loop block diagram as shown in Figure 5.1.

5.2.1 Initializing Optic Disk Center and Boundary

From the analysis of three color channels, the red channel of the RGB image is used here in the computation because the blood vessel has a relatively small effect and the optic disk is still sufficiently obvious in this channel. The candidate region of optic disk can be located with high probability by the largest bright region of the image $I(u, v)$, where u and v are the row and column numbers. The extraordinary case in which the pathological bright region is much larger and brighter than the optic disk is not considered in this chapter, because it is a relatively rare case in narrow-angle stereo optic disk images. The median filter with relative large window is first applied on the original image to remove the effect of the pathological bright regions. From the observation, the average radius of the optic disk is approximately 55 pixels in the stereo fundus image data. Hence, the average disk region is close to 13% of the whole image area. For every filtered image, all pixels are listed in descending order of gray-level; the top 13% pixels are selected as the candidate region of the optic disk. The centroid of the selected pixels is set to be the initial disk center denoted as $\mathbf{c}_0(u_0, v_0)$. Template matching could also be used to detect initial disk center. An error of the initial disk center within several dozen pixels could be tolerated because the initial center provides only the rough location of optic disk.

(a) Edge map for distance transformation. Solid line is the edges; '+' – initial disk center $\mathbf{c}_0$; '∇' – final disk center $\bar{\mathbf{c}}_0$; '×' – initial contour points based on $\bar{r}$.

(b) The radius accumulator for one disk center in distance space

Figure 5.2: An example of distance transformation

The final disk center and radius are detected by *distance transformation* denoted as Ψ , based on the edge map. In the Ψ operation, the distance of each edge point to the disk center is first computed and transformed to the distance space. A suitable bin is selected in distance space to accumulate the distance as shown in Figure 5.2(b). If the distance is in the range of certain bin, the accumulator of this bin will be added by 1. Canny edge detection with Gaussian smoothing is applied on the original image to obtain the edge map I_{edge} . Then the distance of each edge point $\mathbf{g}_i$ to $\mathbf{c}_0$ is computed and mapped to distance space. $\mathbf{c}_0(u_0, v_0)$ is also shifted in the range of $u_0 \pm 30$ pixels and $v_0 \pm 30$ pixels. The bin size is set to be 5 pixels here. Then the peak of the distance space is detected to be the final radius and center of the optic disk (denoted as $\bar{\mathbf{c}}_0$ and $\bar{r}$) as follows:

$$(\bar{\mathbf{c}}_0, \bar{r}) = \arg \max_{\mathbf{g}_i \in I_{edge}, \mathbf{c}_0 \in u_0 \pm 30 \text{ \& } v_0 \pm 30} \{\Psi(\|\mathbf{g}_i - \mathbf{c}_0\|)\} \quad (5.1)$$

Hence the initial disk boundary for the following contour deforming is set to be the circle based on the detected $\bar{\mathbf{c}}_0$ and $\bar{r}$, as shown in Figure 5.2(a), where solid lines are the edges, ‘+’ point is the initial disk center $\mathbf{c}_0$, ‘ ∇ ’ point indicates the final disk center $\bar{\mathbf{c}}_0$, ‘ $\times$ ’ points are the initial contour points based on $\bar{r}$. Although the shape of optic disk may not be ideally circular, to simplify the computation, the initial boundary is assumed to be circular shape, which can be pushed toward the final circular or non-circular contour in the deforming process. The distance transformation lets the initial boundary close to the real boundary as much as possible. The disk center is not fixed, but is later updated together with the disk boundary in the contour deforming operation.

5.2.2 Contour Deformation Based on Combination of Local Energies

In traditional snake, the contour is defined as a parametric curve $\mathbf{s}(n) = [u(n), v(n)]$, where n is the index of the contour point, $1 \leq n \leq N$. It globally deforms to the expected location by minimizing a global energy function, $E_{snake} = \int_1^N [E_{int}(\mathbf{s}(n)) + E_{ext}(\mathbf{s}(n))]dn$, as introduced in section 2.5.1. However, in global deformation, the distractors may affect the potential candidates for correct boundary points. Therefore, in the proposed method, the local radial deformation is used to drive the contour, based on local energy instead of the energy summation of all contour points. Hence, the deformation of each contour point depends only on itself, and is not affected by the incorrect points on the whole contour. This makes the process of searching for correct contour points insensitive to blood vessel influence.

5.2.2.1 Energy Definition

Figure 5.3: Deforming range and clustering sample

A dynamic ρ - θ polar coordinate system of the optic disk is set up for contour deformation as shown in Figure 5.2(a), for which the origin is adjustable. The disk center

$\bar{c}_0$ is set to be the initial origin. The initial contour points are selected at $\rho = \bar{r}$ for every 5° of θ . Hence there are $N = 72$ contour points. The contour growing is limited only in the radial direction (along $\pm\rho$). This constraint reduces the computation complexity and increases the accuracy because the candidate pixels are limited within a relatively small range. For each contour point n , the pixels inside the deforming range $[\rho_n - \rho w, \rho_n + \rho w]$ are considered to be the candidate pixels denoted as ni , as shown in Figure 5.3. In this application, ρw is set to be 10 pixels. The local energy for each candidate pixel ni consists of internal energy and external energy as $E_{ni} = E_{int}(ni) + E_{ext}(ni)$. The new contour point n is set to be the candidate pixel with the lowest local energy:

$$\mathbf{s}^t(n) = \underset{ni \in \text{deforming range}}{\operatorname{argmin}} \{E_{int}^{t-1}(ni) + E_{ext}^{t-1}(ni)\} \quad (5.2)$$

where t is the iteration number; n is the contour point; ni is the candidate pixel inside the deforming range.

Similar to the original snake, the local internal energy is the summation of the absolute value of the first and second derivatives, written as:

$$E_{int}(ni) = \alpha_1 \left| \frac{d\mathbf{s}(ni)}{dn} \right| + \alpha_2 \left| \frac{d^2\mathbf{s}(ni)}{dn^2} \right| \quad (5.3)$$

where α_1 and α_2 are constants specifying the elasticity and stiffness of the snake.

The local external energy is composed of three terms, the magnitude of gradient, the difference of gradient orientation and the difference of intensity value. The gradient magnitude is written as

$$|\nabla I(u, v)| = \sqrt{[\nabla_u I(u, v)]^2 + [\nabla_v I(u, v)]^2} \quad (5.4)$$

where $\nabla_u = \frac{\partial}{\partial u}$ and $\nabla_v = \frac{\partial}{\partial v}$ are the gradient operators in u and v directions. Magnitude of gradient can efficiently provide the boundary location. However the point located on the edge of blood vessel also has high gradient magnitude. Hence, the gradient orientation information is introduced into the external energy as

$$Ori(I(\rho, \theta)) = Ori(I(u, v)) + 90^\circ = \tan^{-1} \left[\frac{\nabla_v I(u, v)}{\nabla_u I(u, v)} \right] + 90^\circ \quad (5.5)$$

The 90° in equation (5.5) is used to adjust the difference between u - v image coordinate system and ρ - θ polar coordinate system. Considering the optic disk is close to a circular shape and the disk region is brighter than the other region, the gradient orientation of the disk boundary at point n denoted as $Ori(n)$ should be close to the angle θ_n of this point. Hence the difference $|Ori(ni) - \theta_n|$ is set to be a robust term of external energy. The intensity value is also included in the computation of the external energy to increase the accuracy. On the disk boundary, most contour points have similar intensity value. Let T_{median} be the median intensity of the contour points and $Int(ni)$ be the intensity value of individual candidate pixel. The absolute intensity difference $|Int(ni) - T_{median}|$ is another term of external energy. Therefore, the full local external energy can be written as

$$E_{ext}(ni) = -\beta_1 |\nabla I(ni)| + \beta_2 |Ori(ni) - \theta_n| + \beta_3 |Int(ni) - T_{median}| \quad (5.6)$$

where β_1 , β_2 and β_3 are the positive weights of three energy terms. The values of these three parameters are set based on the importance of each energy term, and they may be set to the same value by default.

5.2.2.2 Parameter Setting

The weight ratio, not the absolute value of each weight parameter, affects the deforming result; different weight parameter values with the same ratio between each term will provide the same deforming result. In contour deformation, every term in the internal and external energy functions is first normalized to the range of $[0, 1]$. This operation guarantees that each term contributes the same magnitude of energy. Then the weight parameter of each term could be set based on its level of importance in the whole energy function. α_1 and α_2 specify the weights of elasticity and stiffness terms of the snake. Based on the 100 testing data, the stiffness term shows less necessity in the energy function, since there is smoothing update process in each iteration. Hence its weight (α_2) is set to be very small, only 0.1 times that of α_1 . Considering that the gradient magnitude and gradient orientation are more important features, their weight parameter values can be set slightly higher than the others. Since the optic disk generally has an oval shape, not exactly circular one, the gradient orientation of a contour point may not be exactly the same as the point angle θ_n . As a result the gradient orientation will not be as robust as the gradient magnitude. Thus, the weight of gradient orientation is set to no more than the weight of gradient magnitude. For example, $\alpha_1 : \alpha_2 : \beta_1 : \beta_2 : \beta_3$ could be set as 1: 0.1: 1.2: 1: 1. The parameters are set manually only once for all the testing images. The proposed method is not sensitive to the reasonable variation of weight ratio between each term. The ratio of weights does not need to be set quite precisely. The examples of different weight setting will be given in section 5.4.

5.2.2.3 Stop Criterion

The origin of the ρ - θ polar coordinate system shifts to the centroid of the contour after each deformation. The average deforming distance in each iteration is used to set the

stop criterion. At iteration t , the average absolute distance (AAD) between the old contour and deformed contour is defined as

$$AAD(t) = \frac{1}{N} \sum_{n=1}^N \|\mathbf{s}^t(n) - \mathbf{s}^{t-1}(n)\| \quad (5.7)$$

The operation is repeated, until the AAD distance is less than one pixel for five consecutive iterations.

Compared with the original snake which only consists of contour deformation, in this proposed method, clustering and smoothing update are added after each deformation. These modifications make the algorithm more robust to the blood vessel occlusions, noises, and fuzzy contour shapes. The salient feature of this method is that the uncertain points (distractors) are self-extracted and updated whereas the correct edge points are not affected.

5.2.3 Clustering

After each contour deformation, the contour points go to the locations with minimal combined energies. Some points may not be located on the true edges of optic disk because of the effects of blood vessels, ill-defined edges, noises, etc. Deforming-only process could not give sufficiently good result; hence, more operations are required. A method to roughly separate uncertain points and edge points and then update these two groups using different operations is proposed here. Since the classification process is embedded into each iteration, a simple and fast technique should be selected. Precise classification is not necessary in this algorithm, because this is an iterative-update algorithm; the resulted contour does not only depend on one classification in an iteration.

There are many methods of classification, which can be roughly separated into two categories: supervised learning and unsupervised learning. In the supervised learning, the training data are accompanied by labels indicating the class of the samples, and then new sample is classified based on the training set. Unsupervised learning is to establish the existence of classes by natural association according to some similarity measures inside the data. Since the retinal color, optic disk shape, fundus tissue reflection and even photographic illumination, are quite different between each image, it is difficult to give a training set and estimate certain criterion to well classify the edge points and the uncertain points in different images. Hence, the unsupervised learning based on clustering is a more suitable technique which uses sample properties extracted from single image to separate them through self-learning.

There are a number of popular clustering methods such as k-means, fuzzy k-means, sequential k-means, self-organizing feature maps, hierarchical clustering, etc. In this work, the data space has only 72 samples, which is very small. Moreover, it has been already known that there are two classes. Hence, the simple method, i.e., k-means algorithm, is sufficient to approximately group the contour points into two clusters, i.e., edge-point cluster and uncertain-point cluster. Weak edges are enhanced by normalization. This operation is robust to the illumination changes, and to the different tissue reflection from different eyes, etc.

This unsupervised learning (k-means algorithm) is based on the assumption that more than 50% contour points are edge points. Then two features are automatically extracted from contour points for classification. Training data is not required. In the proposed method, a weight is introduced in the traditional k-means algorithm to control the number of edge points into the range of 65% to 83% of the total contour points, which is more than 50% as in the assumption. Because there is the criterion for the point number in

clustering, some edge points may be clustered into uncertain point group. This false clustering however will not affect the final update result, because it can be auto-adjusted by soomthing update operation.

5.2.3.1 Feature Selection

For each contour point (ρ_n, θ_n) , the intensity distribution of its neighboring pixels along radial line, in the range of $\rho_n \pm 10$ pixels, as shown in Figure 5.3, is used as a clustering sample. This sample is a 21×1 vector, denoted as $\mathbf{m}_n$, which must be normalized before clustering computation to enhance the weak edge. There are 72 contour points totally; hence a sequence of 72 sample vectors is obtained. If the contour point is located on the disk edge, its intensity distribution will be similar to a negative-step profile as shown in Figure 5.4(a), because the disk region is brighter than the background region. On the contrary, if the contour point is located on the blood vessel, disk region or background region, it will have different distribution as shown in Figure 5.4(b),(c).

(a) Sample of edge point

(b) Sample of blood vessel point

(c) Sample of background point

Figure 5.4: Normalized clustering samples

Two features are extracted from the sample sequence for classification. Assume that more than 50% of contour points are located on the true boundary after each radial deformation. This assumption is generally satisfied under a correct initialization. A 21×1 median vector $\mathbf{m}_0$ is computed from the 72 sample vectors as

$$\mathbf{m}_0(i) = \text{median}\{\mathbf{m}_1(i), \mathbf{m}_2(i) \cdots, \mathbf{m}_{72}(i)\}, i = 1, 2 \cdots 21 \quad (5.8)$$

which can be considered as an intensity distribution of the edge-point sample. The first feature is set to be the distance between a sample vector $\mathbf{m}_n$ and the median vector $\mathbf{m}_0$, written as $Dm_0(n) = \|\mathbf{m}_n - \mathbf{m}_0\|$, $n = 1, 2, \dots, 72$. Even if less than 50% of the contour points are located at the true boundary, this feature could still be used, because the deformation in the next iteration could attract more contour points to the true boundary. The second feature is set to be the standard deviation of each sample vector $\mathbf{m}_n$, denoted as $Pstd(n) = std(\mathbf{m}_n)$. Hence, a 21×1 sample vector with profile as shown in Figure 5.4 is now reduced to a 2×1 sample vector as $\mathbf{fm}_n = [Dm_0(n), Pstd(n)]^T$. If the sample is an edge point, it will have small Dm_0 and large $Pstd$, otherwise it will have large Dm_0 and small $Pstd$.

5.2.3.2 Self-adjusted Clustering Process

K-means algorithm is applied to separate the contour points into two clusters, the edge-point cluster S_1 and the uncertain-point cluster S_2 . This algorithm classifies the sample into the class which has the center nearest to the sample in the feature space. The features, Dm_0 and $Pstd$, are first normalized to the range of $[0, 1]$. Based on the prior knowledge of the extracted features, the edge samples are expected to be at the left-top corner while the uncertain samples (non-edge samples) are expected to be at the right-bottom corner, as shown in Fig 5.5. Hence, the initial cluster centers are set to be the left-top corner and the right-bottom corner, written as

$$\begin{aligned} \mathbf{cm}_1(t) &= [\min(Dm_0) \quad \max(Pstd)]^T \\ \mathbf{cm}_2(t) &= [\max(Dm_0) \quad \min(Pstd)]^T \end{aligned} \quad (5.9)$$

for the edge-point cluster and uncertain-point cluster respectively with t as the iteration number. At t^{th} iteration, the sample $\mathbf{fm}_n$ is distributed among two clusters $\{S_1(t), S_2(t)\}$ according to the following rule:

$$\begin{cases} \mathbf{fm}_n \in S_1(t) & \text{if } \|\mathbf{fm}_n - \mathbf{cm}_1\| \leq wm \|\mathbf{fm}_n - \mathbf{cm}_2\| \\ \mathbf{fm}_n \in S_2(t) & \text{if } \|\mathbf{fm}_n - \mathbf{cm}_1\| > wm \|\mathbf{fm}_n - \mathbf{cm}_2\| \end{cases} \quad (5.10)$$

The cluster centers are then updated by

$$\mathbf{cm}_j(t+1) = \frac{1}{N_j} \sum_{\mathbf{fm}_n \in S_j(t)} \mathbf{fm}_n, \quad j=1, 2 \quad (5.11)$$

where N_j is the number of the samples in cluster j . If $\mathbf{cm}_j(t+1) = \mathbf{cm}_j(t)$ for $j=1, 2$, the clustering iteration will stop. The final clusters are obtained as shown in Figure 5.5, where ‘×’ symbols are edge points and ‘o’ symbols are uncertain points. Because the following smoothing update is based on the information of S_1 (edge-point cluster), the sample number in S_1 should dominate, i.e., $|S_1| > |S_2|$. wm is the weight used to self-adjust the cluster number N_1 into the range of $[47, 60]$, which is 65% to 83% of the total number of contour points, $N=72$. If $N_1 < 47$, wm is increased; and if $N_1 > 60$, wm is decreased.

5.2.4 Smoothing Update

The uncertain-point cluster S_2 consists of points belonging to vessels or noises. The vessel points tend to scatter on the whole contour and generally have only several continuous points as shown by the circle points in Figure 5.6(a). The noise points are arbitrarily located anywhere on the contour. In some cases, the noise points occur almost continuously in certain large region, as shown by the circle points in Figure 5.6(b), which

may be the result of wrong disk center. Hence, in these cases, the disk center is first replaced by the centroid of edge points before smoothing update.

In each iteration, after contour deformation, clustering, and disk center modification, all the contour points, whether they are in S_1 or S_2 , are updated using the information of their neighbors in S_1 . The point information in S_2 however is not used for update. The points belonging to different clusters are updated by different criteria. This operation will maintain the edge points close to their original positions and update the uncertain points to the correct positions.

Figure 5.5: Two clusters

(a) Dispersedly uncertain points

(b) Continuously uncertain points

Figure 5.6: Clustered contour points. 'x' – edge points; 'o' – uncertain points.

5.2.4.1 Updating Process

Each point's radius ρ_n is updated by its neighboring edge points while keeping the angle θ_n fixed. The number of neighboring points denoted as $win(n)$ is varied for different contour points. An example is illustrated in Figure 5.6(a) to explain the smoothing update, where $win(n)$ is set to be 2, the '×' points are the edge points belonging to S_1 , and the 'o' points indicate the uncertain points belonging to S_2 . The radius ρ_n of point n is replaced by its first and second closest neighboring edge points ('×' points) in S_1 along both $\pm \Delta\theta_n$ sides, written as:

$$\rho_n^{smooth} = \frac{\sum_m w_m \rho_m}{\sum_m w_m}, \quad m = \begin{cases} nl_2, nl_1, n, nr_1, nr_2 & \text{if } n \in S_1 \\ nl_2, nl_1, nr_1, nr_2 & \text{if } n \in S_2 \end{cases}, \quad 1 \leq n \leq N \quad (5.12)$$

where w_m is the smoothing weight, which could be set to be a constant, such as 1, based on uniform distribution as shown in Figure 5.7(a). Gaussian distribution could be another choice to set the weights as shown in Figure 5.7(b). nl_2, nl_1, nr_1, nr_2 are the first and second closest '×' points to point n along $\pm \Delta\theta_n$ directions. The computed neighboring points must belong to S_1 . If the neighboring point is an uncertain point in S_2 , it is not used for smoothing computation. As shown in Figure 5.6(a), the 'o' points between n and nl_1 are not used to update the point n , because they are uncertain points, although they are closer to point n than point nl_1 . Hence, for each point n , $2 \times win(n) + 1$ points (if n is in edge-point cluster) or $2 \times win(n)$ points (if n is in uncertain-point cluster) are used for smoothing. The contour is smoothed in each iteration after deformation.

(a) Smoothing weights based on uniform distribution

(b) Smoothing weights based on Gaussian distribution

Figure 5.7: Smoothing weights based on different distributions

5.2.4.2 Auto-adjusted Updating

Small updating sample number of neighbors, such as $win(\bullet) = 2$, is able to solve the problem of blood vessel occlusion as shown in Figure 5.6(a) and approximately maintain the original positions of edge points. Since the updating here relies on local information only, it may however not provide reliable boundary points, if the contour has continuously uncertain points as shown in Figure 5.6(b). Hence, $win(\bullet)$ should be increased in the continuous noise region, where the updating is based on semi-global information. If $win(\bullet)$ increases to the half of the whole contour, the updating will become an operation based on global information. Global information therefore helps restrict the contour within certain model; local information provides the detailed variation of the contour. Therefore, variable updating sample number for different contour points is introduced in this algorithm to balance the stability and accuracy of the contour.

The *combination force* of uncertain points within the neighbors is used to set the updating sample number $win(n)$. For a contour point n , the orientation of the contour point could be expressed as $Norm_n = (\cos(\theta_n), \sin(\theta_n))$. The combination force of its nearby uncertain points (along $\pm \Delta\theta_n$ directions) in the range of $[n-5, n+5]$ is computed as

$$\chi_{win}(n) = \sum_i w_i Norm_i, \quad i \in [n-5, n+5], \quad w_i = \begin{cases} 0 & \text{if } n \in S_1 \\ 1 & \text{if } n \in S_2 \end{cases} \quad (5.13)$$

The range of $[n-5, n+5]$ is equal to 50° , which is a fixed setting for all the images. It is a suitable semi-global window for smoothing update based on the knowledge. Large $\|\chi_{win}(n)\|$ denotes the continuation of uncertain noise region; small $\|\chi_{win}(n)\|$ indicates blood vessel region or small noise region, as shown in Figure 5.6(b). Hence, the updating sample number $win(n)$ is set based on $\|\chi_{win}(n)\|$, written as

$$win(n) = \begin{cases} 2 & \text{if } \|\mathbf{x}_{win}(n)\| \leq 5 \\ 2 + (\|\mathbf{x}_{win}(n)\| - 5) & \text{if } \|\mathbf{x}_{win}(n)\| > 5 \end{cases} \quad (5.14)$$

Variable updating sample numbers combine both the global information and local information, and make the smoothing update auto-adjusting process between global and local information. Hence, for the large uncertain region of disk edge, the shape of the contour is restricted in the model by using global information; on the contrary, for the certain region of disk edge, the contour is maintained close to its original position.

5.3 Evaluation

The groundtruth of the boundary is manually labeled under the supervision of an ophthalmologist. The average distance from the detected boundary point to the groundtruth is measured for evaluation. The groundtruth denoted by A , consists of the individual pixel a_i , $1 \leq i \leq M$, where M is the amount of the pixel on the groundtruth boundary. $\hat{\mathbf{s}}(n) = [u(n), v(n)]$ is the computed final contour, $1 \leq n \leq N$. For each contour point n , the distance to the closest point (DCP) of groundtruth is defined as

$$DCP(\hat{\mathbf{s}}(n), A) = \min \|\hat{\mathbf{s}}(n) - a_i\|, \quad 1 \leq i \leq M \quad (5.15)$$

The accuracy of the detected boundary is evaluated by the mean of DCP (MDCP) as follows:

$$MDCP(\hat{\mathbf{s}}, A) = \frac{1}{N} \sum_{n=1}^N DCP(\hat{\mathbf{s}}(n), A) \quad (5.16)$$

The smaller MDCP is, the closer the detected boundary is to the groundtruth.

5.4 Results and Discussions

All color stereo fundus image data were provided by the hospital and first resized to 256×272 pixels. One hundred fundus images were randomly selected and tested. The weight ratio, $\alpha_1 : \alpha_2 : \beta_1 : \beta_2 : \beta_3$, was set to be 1:0.1:1.2:1:1, as discussed in section 5.2.2.2. Among them, 94 images were successfully processed, 5 images required further improvement and 1 image failed.

Two examples given the successful or fair results are illustrated in Figure 5.8 and Figure 5.9. In Figure 5.8, the measured MDCPs are respectively 1.1 pixels, 3.1 pixels and 1.5 pixels for the proposed method, GVF snake and the modified ASM, where the proposed method and the modified ASM method provide the successful results and the GVF snake give the fair result. In Figure 5.9, all the three methods give the successful results. Their measured MDCPs are respectively 1.8 pixels, 2.8 pixels and 1.4 pixels for the proposed method, GVF snake and the modified ASM.

An example of typical elliptic optic disk is illustrated in Figure 5.10. The initial disk contour was estimated by distance transformation in the edge map and then deformed to the final boundary as shown in Figure 5.10(a),(b), where ‘×’ indicates the initial contour point, the bright (green) line is the final disk boundary. Groundtruth was manually placed as dark (blue) line in Figure 5.10 to evaluate the accuracy of the detected boundary. The MDCP was measured to be 1.9 pixels in this example. For comparison, GVF snake [72] and modified ASM algorithm [2][86][87] were applied on the images with the same starting disk center. In the modified ASM algorithm, PCA based method for disk center localization was also used, but it did not improve the accuracy of final disk boundary compared with estimating disk center by distance transformation. In GVF snake, the images were preprocessed by morphological operation; and the parameters in the energy

functions were carefully set to make a balance between the smoothness and the accuracy on the resulted boundary. The resulted boundaries for GVF snake and modified ASM are given in Figure 5.10(c), (d), in which the MDCPs are 6.7 pixels and 3.5 pixels respectively. It can be obviously seen that the proposed method provides better result.

One more example given in Figure 5.11 is a relatively small optic disk with ill-defined boundary and noises from the surrounding tissue which make the initial contour far away from the real boundary. The contour deformation, clustering and smoothing update in the proposed method let the boundary grow to the correct location. The measured MDCPs are respectively 2.4 pixels, 23.6 pixels and 24.2 pixels for the proposed method, GVF snake and the modified ASM.

The failure case due to false estimating the disk center under the disturbance of large bright lesion is shown in Figure 5.12. In this example, PCA based approached [76] was applied in modified ASM method to identify disk center; however, it also gave wrong position. Three different methods were applied on this image, and all of them failed.

The MDCP results of 100 images are plotted in Figure 5.13 to clearly compare the performances of different methods. In the graph, the red line denotes the results of the proposed method; the green line gives the results of GVF snake; the blue line indicates the results of modified ASM method. It can be observed that the red line is almost lower than the green line and blue line. Table 5.1 shows the statistical comparison among these three methods, where $MDCP \leq 3$ pixels indicates success, $3 < MDCP \leq 5$ pixels denotes a fair result requiring improvement, and $MDCP > 5$ pixels means a failure. The comparison shows that the proposed method achieves more successful number of results; and also obtains more accurate boundaries in the successful cases than the other two methods.

The weight parameters of energy function were changed into different values of ratio to show the stability of the proposed method, which were applied on the same 100 testing images. Based on the knowledge that the weights of gradient magnitude (β_1) and gradient orientation (β_2) should be set to slightly higher than the other weights, α_1, α_2 and β_3 were fixed to be 1, 0.1 and 1; β_1 and β_2 were varied in the range of $[1, 2]$. The results of five different settings are given in Table 5.2. The mean and standard deviation of the success rate were measured to be 93% and 0.8% respectively, and mean MDCP of the successful cases were all 1.3 pixels for these five settings, which showed that the proposed method was not sensitive to the reasonable variation of weight ratio.

Figure 5.8: A successful and fair example of the three methods, where the proposed method and modified ASM got successful results, GVF snake got fair result. Dark (blue) contour ‘—’ is the groundtruth; bright (green) line ‘—’ is resulted contour; ‘×’ indicates initial contour points.

Figure 5.9: A successful example of all three methods. Dark (blue) contour ‘—’ is the groundtruth; bright (green) line ‘—’ is resulted contour; ‘×’ indicates initial contour points.

Figure 5.10: An example of elliptic optic disk. Dark (blue) contour ‘—’ is the groundtruth; bright (green) line ‘—’ is resulted contour; ‘×’ indicates initial contour points.

Figure 5.11: An example of ill-defined optic disk boundary. Dark (blue) contour ‘—’ is the groundtruth; bright (green) line ‘—’ is resulted contour; ‘×’ indicates initial contour points.

Figure 5.12: A failure example due to lesion disturbance on disk edge. Dark (blue) contour ‘—’ is the groundtruth; bright (green) line ‘—’ is resulted contour; ‘×’ indicates initial contour points; ‘Δ’ is the disk center from PCA.

Figure 5.13: Graph of the performance comparison among different methods

Result \ Method	Proposed method	GVF snake	Modified ASM
Success (MDCP < 3 pixels)	94	12	82
Fair (MDCP = 3-5 pixels)	5	33	9
Fail (MDCP ≥ 5 pixels)	1	55	9
Success rate	94%	12%	82%
Mean MDCP of success (pixels)	1.3	3.6	1.8

Table 5.1: Comparison of mean distance to closest point (MDCP) among different methods

No	Weight ratio of energy function $\alpha_1 : \alpha_2 : \beta_1 : \beta_2 : \beta_3$	Success rate	Mean MDCP of success (pixel)
1	1 : 0.1 : 1.2 : 1.0 : 1	94%	1.3
2	1 : 0.1 : 1.2 : 1.2 : 1	93%	1.3
3	1 : 0.1 : 1.5 : 1.2 : 1	94%	1.3
4	1 : 0.1 : 1.0 : 1.0 : 1	92%	1.3
5	1 : 0.1 : 1.6 : 1.3 : 1	94%	1.3
Mean	—	93%	1.3
STD	—	0.8%	0

Table 5.2: Results of variable weight ratio of the proposed method. STD denotes standard deviation.

5.5 Conclusion

A novel and accurate algorithm of optic disk boundary detection is proposed in this chapter. The initial boundary is estimated by distance transformation. The final boundary is later segmented by modified snake. The original snake is improved and extended in two aspects: clustering and updating the contour points by variable updating sample numbers. Clustering process effectively self-extracts the uncertain contour points from the correct edge points, which directly solves the blood vessel problem. The variable updating sample numbers combine both global and local information of the correct edge points to update the contour points after each radial deformation. This operation updates the uncertain points to the correct locations and keeps the edge points close to their original locations. The modifications make the proposed approach effective to blood vessel occlusions, ill-defined edges, noises, and fuzzy disk shapes, while maintaining the accuracy.

In the 100 testing images, up to 94% disk boundaries can be measured successfully. The average MDCP of success images is 1.3 pixels. The comparison results show that the proposed method is more accurate than GVF snake and modified ASM algorithm, from which the success rates are 12% and 82% respectively. The proposed algorithm is not limited only to optic disk boundary detection, but it is expected to be widely applicable to many other similar applications in boundary detection and image segmentation, especially in the medical imaging area.

Chapter 6

Disk Parameter Estimation

6.1 Introduction

With glaucoma, the eye's drainage system is clogged or blocked, which results in the elevation of eye's intraocular pressure (IOP). High pressure in the eye damages the optic nerve by killing the nerve fibers, which causes loss of vision. The obvious feature of glaucoma on retina is cupping of the optic disk. Millions of nerve fibers run from the retina to the optic nerve meeting at the optic disk. As the nerve fibers are damaged and die, the optic disk begins to hollow out, which is known as cupping. As the nerve fibers keep dying, the cup becomes deeper and bigger.

The 3-D shape of optic disk is generally observed by clinicians under the aid of some instruments to evaluate the optic disk condition; moreover, the quantitative measurements of optic disk parameters, such as cup contour, cup-to-disk (C/D) ratio, rim-to-disk (R/D) area ratio, etc., are also important for evaluation of optic disk. Automated extraction of

these disk parameters could be an assisting tool for physicians to analyze, diagnose and monitor the diseases.

Cup boundary estimation is the first step for extraction of other disk parameters. Practically, the ophthalmologists manually draw the cup boundary and estimate its related parameters. This is a time-consuming work; moreover, the cup boundary measurement requires good experience. The specialized training is necessary for the clinicians to correctly estimate the cup boundary. The cup edge is defined at certain depth down from the disk boundary or retinal surface plane based on 3-D depth information. However, lacking of direct visualization of 3-D optic disk, the ophthalmologists will lose the most important feature for cup edge estimation.

Since 3-D images are not generally available, some definitions are provided to estimate the cup boundary on 2-D images. Previously, the clinicians used the pallor to estimate the cup boundary, while pallor is defined as the area of maximum color contrast inside the disk area. However, in many cases, there is no obvious pallor in disk area as illustrated in Figure 6.1(b). Even if the pallor is clear, its edge is just close to, but not the exact cup edge. Moreover, the pallor edge may be occluded by the main blood vessels in some regions, as shown in Figure 6.1(a). Hence, pallor is not an effective feature for cup boundary estimation. Currently, the bending of small blood vessels at the cup edge is used as a clue to measure the cup boundary as shown in Figure 6.2(a). Nevertheless, this method can only provide several points of cup boundary in the area where there are small blood vessels; for the area without small blood vessels, the cup boundary is not easy to be estimated as show in Figure 6.2(b). A true color 3-D optic disk image is important for clinicians to obtain a more reliable and accurate result of cup contour.

Figure 6.1: Cup boundary estimation based on pallor

Figure 6.2: Cup boundary estimation based on bending of small blood vessels

Cup is the depressed area inside the optic disk, hence the 3-D depth is the primary feature of the cup boundary, for which the detection is a relatively new task in fundus image processing. So far, very few researchers focused their work on cup boundary detection due to the fact that the 3-D image is not easily available. Results of disk and cup boundaries were shown in [41][42], however, the authors did not clearly describe how to obtain these boundaries. In HRT instrument, the disk contour is manually placed by clinicians, and then the cup is extracted using the disk contour and depth information. In this research work, the 3-D optic disk image obtained from Chapter 4 and the accurate disk boundary obtained from Chapter 5 are used to estimate the cup boundary and its related parameters automatically. True color 3-D image not only provides depth information, but also maintains original information of 2-D image. Figure 6.3 gives the basic steps of disk parameter estimation.

Figure 6.3: Flowchart of disk parameter estimation

A deformable model technique will be introduced in section 6.2 to estimate the cup boundary based on the combination of 3-D depth information and 2-D feature information. In section 6.3, an optic disk coordinate system is set up and the clinically useful optic disk parameters are described. This information is useful for glaucoma analysis and monitoring. The results of cup contour and disk parameters are presented in section 6.4. The summary of this chapter is given in section 6.5.

6.2 Cup Boundary Estimation

Compared with the disk boundary detection on the 2-D fundus image, the variable shapes and colors of cup make it more difficult to be automatically segmented. Moreover, in many images, the cup contour could not be easily measured even manually without some experience. Hence, automated detection of cup boundary is a challenge in fundus image processing.

Figure 6.4: One definition of cup boundary based on retinal surface

Figure 6.5: Another definition of cup boundary based on disk edge

3-D depth information is a relatively reliable feature to estimate the cup contour. There are several different cup definitions based on depths. In [7], cup is defined to be 50 μ m down from a reference plane which is estimated from the periphery of retinal surface, as shown in Figure 6.4. Another definition locates cup at 150 μ m down from the optic disk edge [107][108] or 150 μ m down from the highest point of optic disk [6], as

shown in Figure 6.5. Other approaches, such as searching for 50% drop from the optic disk edge to the deepest point as the cup edge, or searching for 120 μ m down from the optic disk edge as the cup edge, were introduced in [109]-[111]. All of these definitions only depend on depth information, because the 3-D image obtained in most of these literatures is colorless or only with pseudo-color. In this research, true color 3-D image has been obtained in Chapter 4; hence not only the depths but also the other features on 2-D images, such as gradient, are fused together to estimate the cup boundary by means of deformable model.

6.2.1 Energy Definition and Contour Deformation

Figure 6.6: Initial cup boundary

The modified deformable model technique described in Chapter 5 is applied to extract the cup boundary by using a different energy definition. Let $I(u,v)$ be the intensity value of the image, a cup contour is defined as a parametric curve $\mathbf{s}(n) = [u(n), v(n)]$, $1 \leq n \leq N$, where n is the index of the contour point. A dynamic ρ - θ polar coordinate system is set up for a deformable cup contour which can be expressed as $\mathbf{s}(n) = [\rho_n, \theta_n]$, $1 \leq n \leq N$. The contour deforms along radial direction to the cup boundary location by minimizing an energy function. The points inside the deforming range of $[\rho_n - 10 \text{ pixels}, \rho_n + 10 \text{ pixels}]$ are the candidate pixels denoted as ni . The origin is initialized at the center of the detected disk boundary. The initial cup points are set to the locations where the disk points are shrunk by 10 pixels toward the origin as illustrated in Figure 6.6, and written as

$$\rho_n^{\text{initial cup}} = \rho_n^{\text{disk}} - 10 \text{ pixels}, \quad 1 \leq n \leq N \quad (6.1)$$

The energy function that fuses together the information of depth, gradient, smoothness, and shape, is written as

$$E_{\text{cup}}(ni) = \lambda_1 E_{\text{depth}}(ni) + \lambda_2 E_{\text{gradient}}(ni) + \lambda_3 E_{\text{smooth}}(ni) + \lambda_4 E_{\text{shape}}(ni) \quad (6.2)$$

where depth and gradient terms give local information, whereas the shape term provides the global information of cup shape so as to make the contour deformation more reliable. λ_1 to λ_4 are the weights for each energy term.

Let d_n be the depth of each optic disk point and $d(ni)$ be the depth of each candidate pixel. Here cup boundary is defined at the location of $150\mu\text{m}$ down from each point of the disk boundary, denoted as $\Delta_n = (d_n - 150\mu\text{m})$. Then the depth term of the energy is

$$E_{\text{depth}}(ni) = |d(ni) - \Delta_n| \quad (6.3)$$

Gradient is useful information, if there is pallor in optic disk. Gradient term consists of both gradient magnitude $\nabla I(ni)$ and gradient orientation $Ori(ni)$ as

$$E_{gradient}(ni) = -|\nabla I(ni)| + |Ori(ni) - \theta_n| \quad (6.4)$$

The definitions of $\nabla I(ni)$ and $Ori(ni)$ are the same as the ones introduced in Chapter 5.

The internal energy of the contour is set to be the smoothness term of the energy, where the coefficient of the second derivative is set to be 0.5 time to the first derivative.

$$E_{smooth}(ni) = \left| \frac{d\mathbf{s}(ni)}{dn} \right| + 0.5 \times \left| \frac{d^2\mathbf{s}(ni)}{dn^2} \right| \quad (6.5)$$

The free-form deformation may give uncertain shape if the cup features are not obvious. Therefore shape model is introduced in the energy function to make the deformation close to certain predefined shape. Let $\mathbf{s}^{shape}$ be the shape model of the cup, which is expressed in ρ - θ coordinate system as $\mathbf{s}^{shape}(n) = [\rho_n^{shape}, \theta_n^{shape}]$, $1 \leq n \leq N$. The shape model could be estimated from training set, or set to be an ellipse (or circle). Based on the observation that the shape of the optic disk is generally similar to the shape of the cup, the disk shape is used as the shape model for cup contour deformation. Alignment of shape model is required before the radial deformation in each iteration. An example of shape model alignment is illustrated in Figure 6.7 where ‘×’ symbols are the cup points computed from previous iteration, and ‘.’ symbols are the points of shape model. The center of the shape model is first moved to the cup center, and then the average radius of the shape model is aligned to be the average radius of the cup contour points, as

$$\bar{\rho}^{shape} = \bar{\rho}, \text{ where } \bar{\rho}^{shape} = \frac{1}{N} \sum_{n=1}^N \rho_n^{shape} \text{ and } \bar{\rho} = \frac{1}{N} \sum_{n=1}^N \rho_n \quad (6.6)$$

(a) Original shape

(b) Aligned shape

Figure 6.7: Alignment of shape model

The shape term of the energy is set to be the radial distance between the candidate pixel and the model pixel.

$$E_{shape}(ni) = |\rho_{ni} - \rho_n^{shape}| \quad (6.7)$$

where the candidate points close to the shape model have low energy, and the points far away from the shape model have high energy.

The full energy function is written as

$$\begin{aligned} E_{cup}(ni) = & \lambda_1 |d(ni) - \Delta_n| + \lambda_2 \{-|\nabla I(ni)| + |Ori(ni) - \theta_n|\} \\ & + \lambda_3 \left\{ \left| \frac{ds(ni)}{dn} \right| + 0.5 \times \left| \frac{d^2s(ni)}{dn^2} \right| \right\} + \lambda_4 |\rho_{ni} - \rho_n^{shape}| \end{aligned} \quad (6.8)$$

In each iteration, the shape model is first aligned based on the cup contour obtained from previous iteration, and then the cup contour deforms to the location where the combined energies of depth, gradient, smoothness, and shape are smallest, finally the contour points are classified and updated. Different from Chapter 5, only the intensity value of each contour point is used here as the clustering feature, because the cup edge may not have a negative-step profile. The contour point with lower intensity value is considered to be an uncertain point based on the knowledge that the blood vessel is always darker than the surrounding area. The separated contour points are then updated by the same criteria as described in Chapter 5. The main steps of cup boundary detection are depicted in a closed loop block diagram as shown in Figure 6.8. The deformation will stop, when the average contour variation is less than 1 pixel for five consecutive iterations.

Figure 6.8: Flowchart of cup boundary detection

6.2.2 Parameter Setting

All deformable model techniques have the weight setting problem. The contour deforms to the location with minimal energy in terms of the weight ratio in energy function, but not the absolute value of each weight parameter. The depth term (λ_1) is based on 3-D depth information, which is the more important feature for cup boundary measurement. The gradient term (λ_2) contributes the gradient information which is useful for the image with high contrast pallor. Since in many cases the pallor is not visible, the weight of the gradient term is set to be 0.5 time to the weight of the depth term. The smooth term (λ_3) adjusts the elasticity and stiffness of contour for smoothness, which could avoid the sawtooth on the contour. The shape term (λ_4) introduces the global shape model to control the free-form deformation into certain model. Every term is first normalized to the range of $[0, 1]$, then the weights $\lambda_1 - \lambda_4$ are set based on the importance

of the energy term. The results from the testing examples show that, if the weights of the smooth and shape terms are set to be the same value of the weight of the depth term at beginning, the contour cannot move to the location with proper depth value and thus a wrong contour is the result. It may be due to the control of the smoothness and shape being too tight to let the contour deform to the proper depth value. Hence the weights of the smooth and shape terms are set to be less than the depth weight at beginning. When the contour deforms to the location with the proper depth value, the weights of the smooth and shape terms are increased automatically to control the shape and smoothness of the contour. The weights are variable with iterations as shown in Figure 6.9, and written as

Figure 6.9: Parameter setting of energy function

$$\begin{cases} \lambda_2 = 0.5\lambda_1 \\ \lambda_3 = \frac{\lambda_1}{1 + e^{-0.8(t-5)}} \\ \lambda_4 = \frac{\lambda_1}{1 + e^{-0.3(t-5)}} \end{cases} \quad (6.9)$$

where t is the iteration number. Depth energy term dominates the whole energy function at the beginning of iterations, then the shape term and smoothness term gradually increase to the same value of the depth term in the following iterations. The gradient term is always set to be a half of the depth term, because the pallor is not obvious in many cases.

6.3 Extraction of Disk Parameters for Analysis

6.3.1 Optic Disk Coordinate System

The fundus areas generally used in analysis are illustrated in Figure 6.10 to determine the positions of the fundus image, where the fundus images are separated into nasal side and temporal side based on anatomic structure. For glaucoma analysis, the disk coordinate system is established with six sectors as illustrated in Figure 6.11, whose origin is set to be the optic disk center. The angle of temporal sector and nasal sector is 90° ; and the angle of the other four sectors is 45° . The optic nerve fibers in superior and inferior regions of temporal side are more frequently damaged than the other sectors [112]-[114]. Hence, inferior temporal and superior temporal sectors are easier to show the pathological changes than other sectors. Moreover, the local damages in six sectors can be used to monitor the diseases, and also used to classify the disease progression into the early glaucomatous stage, moderate and moderately advanced stages.

Figure 6.10: Definition of optic disk region

Figure 6.11: The optic disk coordinate system

(a) Disk parameters

(b) C/D vertical ratio

Figure 6.12: Geometric model of optic disk with cross sectional profile and projected image

6.3.2 Optic Disk Parameters

Many parameters can be used to evaluate the optic disk condition, such as disk area, cup area, rim area, cup and rim volumes, maximal cup depth, cup-to-disk (C/D) vertical ratio, cup-to-disk (C/D) area ratio, rim-to-disk (R/D) area ratio, etc. After the disk and cup boundaries are obtained, those disk parameters can be defined and measured as illustrated in Figure 6.12. Since the size and depth of the normal optic disk are varied with different persons, it is difficult to evaluate the optic disk by the absolute values of area or depth, whereas the relative values such as the ratios are more reasonable parameters to evaluate the optic disk. Among them, C/D vertical ratio is generally estimated by ophthalmologists to quantitatively evaluate the optic disk condition, because this parameter is relatively easy to obtain by observing 2-D image. With the aid of computerized analysis, R/D area ratio is found to be a stable parameter for detecting a glaucomatous abnormality. Hence, in this research, C/D vertical ratio is measured and compared with the result estimated from an experienced clinician. R/D area ratio in each sector of the disk coordinate system is also computed as the referenced parameter for further analysis. These results are shown in section 6.4.

6.4 Results and Discussions

6.4.1 Cup boundary

Thirty pairs of stereo optic disk images were tested. In all the images, the green channel is obvious for definition of pallor edge, which is not very visible in the red channel and blue channel. Hence, the green channel is used for cup boundary estimation.

The 3-D image was first computed using the approach described in Chapter 4, and then the maximal cup depth was calibrated to obtain the real depths. Then the cup boundary was estimated by the proposed deformable model method from the combination of depth, gradient, smoothness, and shape energies.

One example of obvious pallor is illustrated in Figure 6.13. The recovered 3-D true color image with the cup and disk boundaries measured by the proposed methods is shown in Figure 6.13(a). The corresponding 2-D image is given in Figure 6.13(b), where the outer contour is disk boundary and the inner contour is the cup boundary. It can be seen the resulted cup contour passes the bending of small blood vessel as shown in Figure 6.13(b). The cup contour was manually placed by a senior ophthalmologist, denoted by dark line as shown in Figure 6.13(c). Mean distance to closest point (MDCP) defined in Chapter 5 was used to compare how close the two contours are. It was measured to be 3.4 pixels in this example. It can be observed that the boundary placed by the ophthalmologist and the boundary estimated by the proposed method are close at temporal side. However the two boundaries at the nasal side have slightly different, because there is no obvious feature of cup on nasal side, and also the main blood vessels occlude on that region. Hence, it is difficult to decide the exact location of cup edge on the nasal side in this example.

(a) Cup and disk boundaries in 3-D image

(b) Cup and disk boundaries on 2-D image

(c) Comparison of two cup boundaries

Figure 6.13: An example of optic disk image with obvious pallor

Another example of non-obvious pallor is shown in Figure 6.14. The bending of small vessels is obvious at temporal side in this example. Compared with the contour manually placed by the ophthalmologist as shown by the dark line in Figure 6.14(c), the MDCP was measured to be 6.8 pixels. The corresponding HRT result is given in Figure 6.14(d), where the red region is cup, the blue and green regions are rim. The cup is estimated only based on depth information in HRT.

(a) Cup and disk boundaries in 3-D image

(b) Cup and disk boundaries on 2-D image

(c) Comparison of two cup boundaries

(d) HRT

Figure 6.14: Another example of non-obvious pallor

6.4.2 C/D Vertical Ratio for Evaluation and Analysis

6.4.2.1 Evaluation

For many cases, the whole cup contour may not be easily measured even manually. Hence, C/D vertical ratio is not only a disk parameter for analysis, but also can be used to evaluate the results. Table 6.1 gives comparison results of the C/D vertical ratios of the two examples illustrated in section 6.4.1 in which the C/D vertical ratios measured from the proposed method and HRT are both close to the results estimated by the ophthalmologist.

Example name	C/D vertical ratio		
	Proposed method	HRT	Ophthalmologist
Example 1 (Figure 6.13)	0.70	–	0.64
Example 2 (Figure 6.14)	0.72	0.78	0.76

Table 6.1: C/D vertical ratio of two examples

In the 30 testing images mentioned in section 6.4.1, 24 of them have corresponding HRT results. The proposed method of cup boundary detection was employed on these 24 images and the C/D vertical ratios were computed as given in Table 6.2. Different cup definitions were used to measure the cup boundary, the results were compared. The first definition denoted as D1 is that cup edge is located at 150 μ m down from the disk edge as described in section 6.2. From the introduction in section 4.4, it is known that generally the approximately true depths but not the exactly true depths are obtained from a pair of stereo images, if its corresponding HRT image is not available. Since the obtained depths are the approximate values, the cup definition based on the relative depth may be a more suitable definition. Moreover, one limitation of locating the cup boundary based on a

fixed drop value (150 μ m) is that the depth of the optic disk varies with different people which may cause unreliable result. Hence another cup definition donated as D2 is proposed here based on the comments from the ophthalmologists, in which the cup edge is defined at the location of 1/3 drop from the disk edge to the deepest point inside the optic disk. Then Δ_n (the cup edge location) in the energy function (6.3) is changed into $\Delta_n = [d_n - (d_n - d_{min})/3]$, where d_{min} denotes the deepest depth and d_n is the depth of each disk point. For comparison, C/D vertical ratios measured from HRT and estimated by the ophthalmologist were both given in Table 6.2. The C/D vertical ratio estimated by the ophthalmologist is set to be the groundtruth. Then the correlations of the results with groundtruth are 0.62, 0.72 and 0.67 respectively for the proposed method with D1, the proposed method with D2, and HRT.

Automated detection of the cup boundary on the stereo image is first studied in this thesis. Generally, the ophthalmologist labels the cup boundary based on the combined information of depth change, blood vessel tortuosity, pallor edge, etc. How to define the cup boundary in quantitative criterion is still investigated by many clinicians, since the pathological changes are quite different for different patients. D1 and D2 are two definitions generally used. If the cupping is 450 μ m, D1(150 μ m drop from disk edge) and D2 (1/3 drop from disk edge) will give the same result. If the cupping is deeper or shallower than the 450 μ m, the results of D1 and D2 disagree. The deeper (or shallower) the cupping is, the greater the disagreeing is. The cup is defined at 50 μ m down from the retinal surface in HRT system. Similar to the previous discussion, if the depth of retinal surface plus 50 μ m equals 1/3 of the cupping, the results of HRT and D2 are same, otherwise they are different.

Let $Diff$ express the absolute difference of C/D vertical ratios estimated by the ophthalmologist and other three methods. The statistical comparison among the results from the proposed methods with two definitions and HRT is given in Table 6.3, where $Diff \leq 0.1$ indicates success, $0.1 < Diff \leq 0.2$ denotes a fair result requiring improvement, and $Diff > 0.2$ means a failure. The proposed method with D1 achieves results similar to those of the proposed method with D2, both have success rate of 75%, which is slightly higher than the success rate of 67% for HRT. It implies the proposed method of 3-D reconstruction and disk parameter estimation is comparable with HRT, and locating cup boundary based on the relative depth could be a feasible cup definition.

6.4.2.2 Analysis

C/D ratio is a parameter widely used in glaucoma analysis. In normal people, C/D vertical ratio of most people is less than 0.3; 15% is larger than 0.4; 10.2% is larger than 0.5; and 5.8% is larger than 0.6. This ratio has some relationship with IOP, but not entirely relates with IOP. If the C/D vertical ratio is larger than 0.6, or if the difference of C/D vertical ratios between the left eye and right eye is larger than 0.2, it is considered to be pathological [105]. The analysis results are given in Table 6.2, based on the C/D vertical ratios estimated from the proposed method with the first cup definition.

Image no. (Image name)	C/D vertical ratio				Analysis Result
	Proposed method (D1)	Proposed method (D2)	HRT	Ophthalm- ologist	
1 (1L)	0.75	0.59	0.58	0.54	A
2 (1R)	0.63	0.49	0.66	0.72	A
3 (5L)	0.73	0.65	0.63	0.66	A
4 (5R)	0.75	0.56	0.61	0.67	A
5 (8L)	0.62	0.64	0.45	0.52	A
6 (8R)	0.71	0.66	0.46	0.69	A
7 (12L)	0.72	0.82	0.94	0.85	A
8 (12R)	0.67	0.70	0.21	0.76	A
9 (14L)	0.82	0.68	0.82	0.78	A
10 (14R)	0.80	0.73	0.79	0.77	A
11 (15L)	0.64	0.59	0.41	0.48	A
12 (15R)	0.56	0.47	0.37	0.52	N
13 (16L)	0.51	0.45	0.27	0.44	N
14 (16R)	0.56	0.43	0.46	0.52	N
15 (22L)	0.77	0.70	0.47	0.65	A
16 (22R)	0.71	0.61	0.69	0.70	A
17 (24L)	0.89	0.78	0.88	0.82	A
18 (24R)	0.84	0.70	0.86	0.65	A
19 (25L)	0.64	0.61	0.76	0.72	A
20 (25R)	0.72	0.71	0.78	0.76	A
21 (28L)	0.67	0.58	0.60	0.68	A
22 (28R)	0.77	0.66	0.58	0.62	A
23 (30L)	0.69	0.62	0.58	0.72	A
24 (30R)	0.69	0.67	0.56	0.68	A
correlation	0.62	0.72	0.67	—	—

Table 6.2: C/D vertical ratio

D1– 1st cup definition; D2–2nd cup definition; N–Normality; A–Abnormality.

Result \ Method	Proposed method (D1)	Proposed method (D2)	HRT
Success ($Diff \leq 0.1$)	18	18	16
Fair ($0.1 \leq Diff \leq 0.2$)	5	5	5
Fail ($Diff \geq 0.2$)	1	1	3
Success rate	75%	75%	67%

Table 6.3: Statistical comparison of C/D vertical ratio

Figure 6.15: An example of disk parameter estimation

Parameter \ Sector	Global	T	ST	SN	N	IN	IT
C/D area ratio	0.50	0.40	0.49	0.60	0.58	0.52	0.39
R/D area ratio	0.50	0.60	0.51	0.40	0.42	0.48	0.61
C/D vertical ratio	0.70	–	–	–	–	–	–

Table 6.4: Disk parameters for glaucoma analysis

Figure 6.16: Another example of disk parameter estimation

Parameter \ Sector	Global	T	ST	SN	N	IN	IT
C/D area ratio	0.42	0.36	0.57	0.57	0.32	0.33	0.51
R/D area ratio	0.58	0.64	0.43	0.43	0.68	0.67	0.49
C/D vertical ratio	0.78	—	—	—	—	—	—
C/D area ratio (HRT)	0.53	0.56	0.68	0.65	0.37	0.38	0.64
R/D area ratio (HRT)	0.47	0.44	0.32	0.35	0.63	0.62	0.36
C/D vertical ratio (HRT)	0.78	—	—	—	—	—	—

Table 6.5: Another example of disk parameters in comparison with HRT results

6.4.3 Other Disk Parameters for Analysis

More disk parameters could be extracted, if the cup and disk boundaries are available. R/D area ratio is considered as one stable parameter for glaucoma analysis, which can be

measured in six different disk sectors. The optic fibers in the superior area and inferior area are generally first damaged. Hence, these areas are more sensitive to pathological changes. In this section, the first cup definition is used for discussion. An example is shown in Figure 6.15 which is the same as Figure 6.13 with coordinate system indicated. The disk boundary was extracted by the method proposed in Chapter 5 following by the cup boundary estimated on the 3-D optic disk image. The C/D area ratios and R/D area ratios in different sectors were then measured with the results as given in Table 6.4, where global ratio is the whole area ratio for C/D or R/D. One more example is illustrated in Figure 6.16. The resulted ratios were compared with the results computed from HRT as given in Table 6.5. It can be observed in this particular case that the R/D ratios in ST, SN and IT regions are obviously smaller than the other three regions. This result gives useful information to clinicians, since IT and ST sectors are easier to show the pathological changes than other sectors as introduced in section 6.3.1.

6.5 Conclusion

A deformable model approach is introduced in this chapter to estimate the cup boundary, which is used for further measurement of disk parameters for glaucoma analysis. The cup boundary is computed based on the results obtained previously, i.e., 3-D optic disk image obtained from Chapter 4 and optic disk boundary extracted from Chapter 5. The energy function in deformable model technique combines the information of depth, gradient, smoothness, and shape. C/D vertical ratios are measured and compared with the results from HRT and ophthalmologist for evaluation. Based on 24 images, and the C/D ratio evaluation made by a senior ophthalmologist as the reference, the proposed method with D1 and proposed method with D2 have both achieved 75% success rate where the defined $Diff \leq 0.1$ is observed for success. This is similar to the results from HRT,

implying that it can offer a comparable analytical result. The disk coordinate system is incorporated to help separate the optic disk region into six sectors so that other disk parameters, i.e., R/D and C/D area ratios can be measured straightforwardly for glaucoma analysis. The proposed system for disk parameter estimation could therefore be an effective alternative for use in clinical analysis and diagnosis.

Chapter 7

Conclusion and Recommendations

7.1 Conclusion

Early detection of eye diseases is essential to make the timely treatment and to avoid irreversible loss of vision. Moreover, monitoring the disease progression during the clinical therapy can help the physicians to better understand and control the state of the illness. Automated processing and analyzing fundus images not only reduce the time consuming workloads, but also provide more clinical information which is difficult to be obtained or estimated manually.

In this research, a glaucoma analyzing system based on color stereo fundus image has been presented as a computer-aided tool, for fulfilling the increasing requirements on eye disease analysis and monitoring tool in clinical practice. The work consists of 3-D optic disk reconstruction, optic disk boundary detection, cup boundary estimation, disk parameter measurement and analysis. The newly developed techniques in each module of the system are summarized as follows:

- Fundus camera calibration was first studied in this research work. Two popular camera calibration methods based on 3-D calibration object and planar calibration object respectively were implemented and first applied on the monocular fundus camera (Canon CR6-45NM) for comparison. The later one was proved to be more stable and accurate. Simultaneous stereo fundus camera (Nidek 3-Dx) was first investigated. The camera calibration method based on planar object was employed to estimate the focal length for use in the following 3-D optic disk reconstruction.
- A fast, sub-pixel, combined matching method has been proposed to precisely detect the correspondences and then convert them into depths. This operation utilizes the advantages of both methods, i.e., the accuracy of feature-based method and noise stabilization in the correlation method. A new method has been proposed to calibrate and integrate the optical effect within the eyeball into the reconstruction process to provide a real 3-D image of the optic disk. This is the difference of the 3-D optic disk reconstruction from the common object reconstruction. An evaluation method based on correlation has been introduced to indirectly evaluate the accuracy of the 3-D image. The experimental results show that the proposed method gives accurate matching from which the average value of correlation factor has been achieved up to 0.96. Compared with the result from Heidelberg Retina Tomograph (HRT), the resulted 3-D image from the proposed method also shows good consistency which indicates that it can be an alternative for use in clinical analysis and diagnosis.
- A novel and accurate algorithm based on deformable model technique has been proposed to detect the optic disk boundary. It extends the original snake in two aspects, i.e., clustering and updating, to directly solve the blood vessel problem. Firstly, the contour points are self-separated into edge-point cluster and uncertain-

point cluster by k-means algorithm after each contour deformation. Then, the contour points in two clusters are updated by different criteria using both the global and local information. The modifications make the proposed approach more effective to blood vessel occlusions, ill-defined edges, noises, and fuzzy disk shapes, while maintaining the accuracy. The comparison results show that the proposed method is more accurate than GVF-snake and modified ASM algorithm.

- A deformable model approach with combined energies has been employed to estimate the cup contour based on the available information of 3-D optic disk image and disk boundary. A disk coordinate system has been established, from which the disk parameters were extracted with the validation of clinical evaluation. The comparison results show that the proposed system gives similar results as those from HRT. The results of disk parameter measurements can potentially be used for analyzing, monitoring and grading of glaucoma.

In the present work, an in-depth study of two main problems has been carried out for the development of a glaucoma analyzing system. This includes the development of a novel, suitable and accurate 3-D reconstruction method for optic disk evaluation, and automated extraction of optic disk parameters with clinical validations. The most significant contributions in this research work are summarised as follows:

- (1) In this study, an entire system for assisting in glaucoma analysis is presented. This system successfully recovers true-color 3-D optic disk image, detects the optic disk boundary, estimates cup contour, and measures disk parameters.
- (2) A promising combined matching approach is proposed to precisely detect the matching points in 3-D reconstruction. The optical effect of biological layers

within the eyeball is considered, calibrated and integrated into the reconstruction process to provide a real 3-D image of the optic disk.

- (3) A novel approach based on deformable model technique is introduced to detect the optic disk boundary. The algorithm extends original snake technique into two aspects, clustering and smoothing update, to efficiently solve the problems of blood vessel occlusions, noises, ill-defined edges, etc.
- (4) The cup contour and disk parameters are estimated with clinical validation. The whole system is proved to be successful and its performance is comparable with that of HRT.

7.2 Recommendations

Fundus image processing is a challenging research topic with great application potential. A glaucoma analyzing system has been studied and very promising results have been achieved. However, there is still much room for improvement, such as automatic analysis and diagnosis. In particular, recommendations for further research are listed below.

- It is already known that the 3-D shape of the optic disk gives signs of abnormality. The disk parameters, such as C/D ratio indirectly provide the cupping stage of the optic disk. How to directly evaluate 3-D optic disk shape to grade the damage of optic nerve should be further studied. And also, clues of pathologies in 2-D shape of optic disk contour and color of retina should be identified to give more useful information of diagnosis. So far, our study is directed on automated glaucoma analysis. Fusing the optic disk parameters with other parameters from other

medical examination such as IOP and vision field, etc., a robust and intelligent diagnostic aiding system could be set up.

- Computerized diagnostic aiding system for monitoring and analyzing the progression of diseases can also help clinicians to better understand the diseases, and then guide them to adjust the therapies. A follow-up system could be set up based on the current study. The extracted clinical parameters can be compared with the previous records for analysis.
- As described in Chapter 4, the presented evaluation methods for 3-D optic disk reconstruction have certain limitations. The proposed correlation evaluation method is sensitive to the nonuniform illumination, whereas the method based on edge map is not easy to give quantitative evaluation. How to automatically and quantitatively evaluate the recovered 3-D optic disk image should be further investigated.
- Detection of the main components on the fundus image, such as optic disk, blood vessels and fovea, has been investigated by many researchers. Accurate detection of lesions on fundus image such as hemorrhages, microaneurysms, cotton-wood spots, drusens could be another topic for further study. Hemorrhages and microaneurysms are dark spots, whereas cotton-wood spots and drusens are similar bright spots. How to detect these abnormal features and classify them is a challenge work. The lesion detection results could be fused together with other information of the eye to establish an auto-analyzing system used in multiple-disease mass screening.

Author's Publications

1. Juan Xu, Opas Chutatape, "Auto-adjusted 3-D optic disk viewing from low-resolution stereo fundus image", International Journal of Computers in Biology and Medicine, in press.
2. Juan Xu, Opas Chutatape, Ce Zheng, Paul Chew Tec Kuan, "3-D optic disk visualization from stereo images via dual registration and ocular media optical correction", British Journal of Ophthalmology, in press.
3. Juan Xu, Opas Chutatape, Paul Chew Tek Kuan, "Automated optic disk boundary detection by modified active contour model", IEEE Trans. on Biomedical Engineering, in review.
4. Juan Xu, Opas Chutatape, Eric Sung, Ce Zheng, Paul Chew Tec Kuan, "Optic disk feature extraction via modified deformable model technique for glaucoma analysis", Pattern Recognition, in review.
5. Juan Xu, Ce Zheng, Opas Chutatape, Paul Chew Tec Kuan, "3-D rebuild technique", Visual Perception of the 3D Space and Applications in Robotics, Singapore 2005 Workshop, November 24th – 25th, 2005.
6. Juan Xu, Opas Chutatape, "3-D optic disk reconstruction via combined registration and inclusion of eye optical effects", 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Sep 1st- 4th, 2005.
7. Juan Xu, Opas Chutatape, "Automated detection of optic disk boundary by a new deformable model technique", 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Sep 1st- 4th, 2005.

8. Juan Xu, Opas Chutatape, "Automatic 3-D optic disk image reconstruction from low-resolution fundus image for glaucoma analysis", 1st International Bioengineering Conference, Sep 8th-10th, 2004.
9. Juan Xu, Opas Chutatape, "3-D optic disk reconstruction from low-resolution stereo images based on combined registrations and sub-pixel matching", 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Sep 1st-5th, 2004.
10. S. Dandapat, Juan. Xu, Opas Chutatape, S. M. Krishnan, "Wavelet transform domain data embedding in a medical image", 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Sep 1st-5th, 2004.
11. Juan Xu, Opas Chutatape, "Comparative study of two calibration methods on fundus camera", 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Sep 17th-21st, 2003.
12. Juan Xu, Opas Chutatape, "Quadratic matching of retinal surface", 1st SERI-ARVO Meeting on Research in Vision and Ophthalmology, Feb 6th-9th, 2003.

Bibliography

- [1] H. Li, "A model-based approach for automated feature extraction in color fundus images", a Ph.D thesis, in school of EEE, Nanyang Technological University, Singapore, 2003.
- [2] H. Li, O. Chutatape, "Automated feature extraction in color retinal images by a model based approach", *IEEE Trans. on Biomedical Engineering*, vol. 51, issue 2, pp. 246 – 254, February 2004.
- [3] Singapore National Eye Center, http://www.shec.com.sg/clinical_services/glaucoma.asp, 2005
- [4] New England Eye Center, <http://www.neec.com>, 2005
- [5] <http://www.stlukeseye.com/Conditions/Default.asp>, 2005
- [6] K. Masunori, N. Manabu, N. Tsuguo, "A method for analyzing a stereoscopic image of a fundus, and an apparatus for executing that method", patent no. EP0846439, June 1998.
- [7] "Quantitative three-dimensional imaging of the posterior segment with the Heidelberg Retina Tomograph", HRT tutorial, <http://www.heidelbergengineering.com/hrtf/hrt.html>, 2005
- [8] http://www.merckfrosst.ca/e/health/glaucoma/the_eye/anatomy/home.html, 2005.
- [9] <http://www.glaucoma.org.au/whatis.htm>, "Glaucoma", 2005
- [10] R. J. Saine, M. E. Tyler, "Ophthalmic photography: retinal photography, angiography, and electronic imaging", Second Edition, Butterworth-Heinemann, 2002.
- [11] <http://www.lpopproducts.com/products/LPO12.asp>, 2005
- [12] http://www.bhj.org/journal/2002_4403_jul/md_346.htm, 2005
- [13] H. Lange, "Advances in the cooperation of shape from shading and stereo vision", in *Proc. the 2nd International Conference on 3-D Digital Imaging and Modeling*, pp. 46 – 58, October 1999.
- [14] O. D. Faugeras, Q. T. Luong, S. J. Maybank, "Camera self-calibration: theory and experiments", in *Proc. the 2nd European Conference on Computer Vision*, pp. 321 – 334, May 1992.
- [15] S. Maybank and O. D. Faugeras. "A theory of self-calibration of a moving camera," *Int. Journal of Computer Vision*, vol. 8(2), pp.123 – 151, August 1992.
- [16] Z. Zhang, Q. T. Luong, O. Faugeras, "Motion of an uncalibrated stereo rig: self-calibration and metric reconstruction", *IEEE Trans. on Robotics and Automation*, vol. 12, no. 1, pp. 103 – 113, February 1996.

- [17] Q. T. Luong, O. D. Faugeras, “Self-calibration of a stereo rig from unknown camera motions and point correspondences”, Technical report RR-2014, INRIA, 1993.
- [18] Q. T. Luong, O. D. Faugeras, “Self-calibration of a camera using multiple images”, in *Proc. the 11th IAPR International Conference on Pattern Recognition, Conference A: Computer Vision and Applications*, vol. 1, pp. 9 – 12, September 1992.
- [19] Q. T. Luong, O. D. Faugeras. “Self-calibration of a moving camera from point correspondences and fundamental matrices”, *International Journal of Computer Vision*, vol. 22, issue. 3, pp. 261 – 289, March/April 1997.
- [20] R. I. Hartley, “Kruppa’s equations derived from the fundamental matrix”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 2, pp. 133 – 135, February 1997.
- [21] G. Xu, J. Terai, H. Y. Shum, “A linear algorithm for camera self-calibration, motion and structure recovery for multi-planar scenes from two perspective images”, in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 474 – 479, 2000.
- [22] M. I. A. Lourakis, R. Deriche, “Camera self-calibration using the Kruppa equations and the SVD of the fundamental matrix: the case of varying intrinsic parameters”. Technical report RR-3911, INRIA, 2000.
- [23] S Ying, B. G. Wooi, “Camera self-calibration from video sequences with changing focal length”, in *Proc. IEEE International Conference on Image Processing*, vol. 2 , pp. 176 – 180, October 1998.
- [24] P. R. S. Mendonca, R. Cipolla, “A simple technique for self-calibration”, in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 1999 – 505, June 1999.
- [25] M. Agrawal, “On automatic determination of varying focal lengths using semidefinite programming”, in *Proc. IEEE International Conference on Image Processing*, pp. 3379 – 3382, October 2004.
- [26] F. Shen, H. Wang, “A new stratified self-calibration algorithm under small camera rotations”, in *Proc. Image and Vision Computing New Zealand (IVCNZ2002)*, November 2002.
- [27] A. Habed, B. S. Boufama, “Camera self-calibration from two views”, in *Proc. IEEE International Conference on Systems, Man and Cybernetics*, vol. 4, pp. 1 – 5, October 2002.
- [28] E. E. Hemayed, “A survey of camera self-calibration”, in *Proc. IEEE Conference on Advanced Video and Signal Based Surveillance*, pp. 351 – 357, July 2003.

-
- [29] Z. Zhang, "Camera calibration with one-dimensional objects", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26, issue. 7, pp. 892 – 899, July 2004.
  - [30] P. F. Sturm, S. J. Maybank, "On plane-based camera calibration: a general algorithm, singularities", in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 432 – 437, June 1999.
  - [31] J. Heikkila, O. Silven, "A four-step camera calibration procedure with implicit image correction", in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1106 – 1112, June 1997.
  - [32] K. Deguchi, I. Morishita, "A unified linear camera calibration method using two parallel calibration planes", *Trans. Society of Instrumentation and Control Engineers*, vol. 29, no. 9, pp. 1023 – 1032, 1993.
  - [33] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations", in *Proc. the 7th IEEE International Conference on Computer Vision*, vol. 1, pp. 666 – 673, September 1999.
  - [34] G. Q. Wei, S. D. Ma, "Implicit and explicit camera calibration: theory and experiments", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 16, no. 5, pp. 469 – 480, May 1994.
  - [35] J. Weng, P. Cohen, M. Herniou, "Camera calibration with distortion models and accuracy evaluation", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 14, issue. 10, pp. 965 – 980, October 1992.
  - [36] R. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses", *IEEE Journal of Robotics and Automation*, vol. 3, issue. 4, pp. 323 – 344, August 1987.
  - [37] M. T. El-Melegy, A. A. Farag, "Nonmetric lens distortion calibration: closed-form solutions, robust estimation and model selection", in *Proc. the 9th IEEE International Conference on Computer Vision*, vol. 1, pp. 554 – 559, October 2003.
  - [38] R. Swaminathan, S. K. Nayar, "Non-metric calibration of wide-angle lenses and polycameras", in *Proc. IEEE Computer Society Conference on. Computer Vision and Pattern Recognition*, vol. 2, pp. 413 – 419, June 1999.
  - [39] M. T. Ahmed, A. A. Farag, "Differential methods for nonmetric calibration of camera lens distortion", in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. II-477 – II-482, December 2001.
-

-
- [40] K. Deguchi, D. Kawamata, K. Mizutani, H. Hontani, K. wakabayashi, "3D fundus shape reconstruction an display from stereo fundus images", *IEICE Trans. INF. & SYST.*, vol. E83-D, no. 7, July 2000.
  - [41] E. Corona, S. Mitra, M. Wilson, T. Krile, Y.H. Kwon, 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, no.10, pp. 1244 – 1253, October 2002.
  - [42] E. Corona, S. Mitra, M. Wilson, P. Soliz, "Digital stereo optic disc image analyzer for monitoring progression of glaucoma", in *Proc. SPIE on Medical Imaging*, vol. 4684, pp. 82 – 93, 2002.
  - [43] M. Ramirez, S. Mitra, A. Kher, J. Morales, "3-D digital surface recovery of the optic nerve head from stereo fundus images", in *Proc. the 5th Annual IEEE Symposium on Computer-Based Medical Systems*, pp. 284 – 291, June 1992.
  - [44] J. M. Ramirez, S. Mitra, J. Norales, "Visualization of the three-dimensional topography of the optic nerve head through a passive stereo vision model", *Journal of Electronic Imaging*, vol. 8, no. 1, pp. 92 – 97, January 1999.
  - [45] K. Deguchi, J. Noami, H. Hontani, "3D fundus pattern reconstruction and display from multiple fundus images", in *Proc. the 15th International Conference on Pattern Recognition*, vol. 4, pp. 94 – 97, September 2000.
  - [46] M. Okutomi, O. Yoshizaki, G. Tomita, "Color stereo matching and its application to 3-D measurement of optic nerve head", in *Proc. the 11th IAPR International Conference on Pattern Recognition, Conference A: Computer Vision and Applications*, vol. 1, pp. 509 – 513, September 1992.
  - [47] M. Z. Brown, D. Burschka, G. D. Hager, "Advances in computational stereo", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, issue. 8, pp. 993 – 1008, August 2003.
  - [48] D. Scharstein, R. Szelisky, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms", *International Journal of Computer Vision*, vol. 47, no. 1, pp. 7 – 42, April 2002.
  - [49] E. Corona, S. Mitra, M. Wilson, "A fast algorithm for registration of individual frames and information recovery in fluorescein angiography video image analysis", in *Proc. the 5th IEEE Southwest Symposium on Image Analysis and Interpretation*, pp. 265 – 268, April 2002.
-

- [50] N. Ritter, R. Owens, J. Cooper, R. H. Eikelboom, P. P. Van Saarloos, "Registration of stereo and temporal images of the retina", *IEEE Trans. on Medical Imaging*, vol. 18, issue. 5, pp. 404 – 418, May 1999.
- [51] L. Kubecka, J. Jan, "Registration of bimodal retinal images-improving modifications", in *Proc. the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1, pp. 1695 – 1698, September 2004.
- [52] F. Zana, J. C. Klein, "A multimodal registration algorithm of eye fundus images using vessels detection and Hough transform", *IEEE Trans. on Medical Imaging*, vol. 18, issue. 5, pp. 419 – 428, May 1999.
- [53] C. Tomasi, R. Manduchi, "Stereo matching as a nearest-neighbor problem", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, issue. 3, pp. 333 – 340, March 1998.
- [54] S. Roy, I. J. Cox, "A maximum-flow formulation of the N-camera stereo correspondence problem", in *Proc. the 6th International Conference on Computer Vision*, pp. 492 – 499, January 1998.
- [55] J. Wang, R. Hsien, H. Chiu, "A progressive constraint search approach for disparity matching in stereo vision", in *Proc. Natl. Sci. Counc. ROC(A)*, vol. 23, no. 6, pp.789-798, 1999.
- [56] Z. Zhang, R. Deriche, O. Faguaras, Q. T. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry", Research Report RR-2273, Sophia-Antipolis: INRIA Sophia-Antipolis, 1994.
- [57] D. Stoyanov, A Darzy, G. Z. Yang, "Dense 3D depth recovery for soft tissue deformation during robotically assisted laparoscopic surgery", in *Proc. the 7th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI2004)*, pp. 41 – 48, 2004.
- [58] S. Veesser, M. Dunn, G. Z. Yang, "Multiresolution image registration for two-dimensional gel electrophoresis", *Proteomics*, vol. 1, pp. 856 – 870, 2001.
- [59] A. Can, C. V. Stewart, B. Roysam, H. L. Tanenbaum, "A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 3, pp. 347 – 364, March 2002.
- [60] A. Can, C. V. Stewart, B. Roysam. "Robust hierarchical algorithm for constructing a mosaic from images of the curved human retina", in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 286-292, June 1999.

-
- [61] A. Can, C. V. Stewart, B. Roysam, H. L. Tanenbaum, "A feature-based technique for joint, linear estimation of high-order image-to-mosaic transformations: application to mosaicing the curved human retina", in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 585 –591, June 2000.
  - [62] F. Laliberte, L. Gagnon, Y. Sheng, "Registration and fusion of retinal images-an evaluation study", *IEEE Trans. on Medical Imaging*, vol. 22, issue. 5, pp. 661 – 673, May 2003.
  - [63] G. K. Matsopoulos, P. A. Asvestas, N. A. Mouravliansky, K. K. Delibasis, "Multimodal registration of retinal images using self organizing maps", *IEEE Trans. on Medical Imaging*, vol. 23, issue. 12, pp. 1557 – 1563, December 2004.
  - [64] M. Kass, A. Witkin, D. Terzopoulos, "Snakes: active contour models", *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321-331, 1987.
  - [65] M. Sonka, V. Hlavac, R. Boyle, "Image Processing, Analysis, and Machine Vision", Second Edition, Pacific Grove, CA: PWS Pub., 1999.
  - [66] <http://www.icaen.uiowa.edu/~dip/LECTURE/Understanding2.html>, "Image understanding: Active contour models - snakes", Lecture Notes, 2005.
  - [67] L. D. Cohen, I. Cohen, "Finite-element methods for active contour models and balloons for 2-D and 3-D images", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 15, issue. 11, pp. 1131 – 1147, November 1993.
  - [68] L. D. Cohen, "On active contour models and balloons", *CVGIP: Image Understanding*, vol. 53, no. 2, pp. 211 – 218, March 1991.
  - [69] C. Xu, J. L. Prince, "Snakes, shapes, and gradient vector flow", *IEEE Trans. on Image Processing*, vol. 7, issue. 3, pp. 359 – 369, March 1998.
  - [70] A. Dumitras, A. N. Venetsanopoulos, "A comparative study of snake models with application to object shape description in bi-level and gray-level images", in *Proc. IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, June 2001.
  - [71] F. Mendels, C. Heneghan, J. P. Thiran, "Identification of the optic disc boundary in retinal images using active contours", in *Proc. Irish Machine Vision and Image Processing (IMVIP) Conference*, pp. 103 – 115, 1999.

-
- [72] A. Osareh, M. Mirmehd, B. Thomas, R. Markham, "Comparison of colour spaces for optic disc localisation in retinal images", in *Proc. the 16th International Conference on Pattern Recognition*, vol. 1, pp. 743 – 746, August 2002.
  - [73] S. Tamura, Y. Okamoto, "Zero-crossing interval correction in tracing eye-fundus blood vessels," *Pattern Recognition*, vol. 21, no. 3, pp. 227 – 233, 1988.
  - [74] L. Zheng, O Chutatape, S. M. Krishnan, "Automatic image analysis of fundus photograph", in *Proc. the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2, pp. 524 – 525, October 1997.
  - [75] A. Pinz, S. Bernogger, P. Datlinger, and A. Kruger, "Mapping the human retina," *IEEE Trans. on Medical Imaging*, vol. 17, issue 4, pp. 606 – 619, August 1998.
  - [76] H. Li, O. Chutatape, "Automatic location of optic disk in retinal images", in *Proc. International Conference on Image Processing*, vol. 2, pp. 837 – 840, October 2001.
  - [77] Y. A. Tolias, S. M. Panas, "A fuzzy vessel tracking algorithm for retinal images based on fuzzy clustering", *IEEE Trans on Medical Imaging*, vol. 17, issue. 2, pp. 263 – 273, April 1998.
  - [78] C. Sinthanayothin, "Image analysis for automatic diagnosis of diabetic retinopathy", a Ph.D thesis, in King's College London, University of London, UK, 1999.
  - [79] C. Sinthanayothin, J. A. Boyce, H. L. Cook, and T. H. Williamson, "Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images", *British Journal of Ophthalmology*, vol. 83, pp. 902–910, 1999.
  - [80] J. Lowell, A. Hunter, D. Steel, A. Basu, R. Ryder, E. Fletcher, L. Kennedy, "Optic nerve head segmentation", *IEEE Trans. on Medical Imaging*, vol. 23, issue. 2, pp. 256 – 264, February 2004.
  - [81] M. Lalonde, M. Beaulieu, L. Gagnon, "Fast and robust optic disc detection using pyramidal decomposition and Hausdorff-based template matching", *IEEE Trans. on Medical Imaging*, vol. 20, issue. 11, pp. 1193 – 1200, November 2001.
  - [82] E. Trucco, P. J. Kamat, "Locating the optic disk in retinal images via plausible detection and constraint satisfaction", in *Proc. International Conference on Image Processing*, pp. 155 – 158, October 2004.
  - [83] M. Foracchia, E. Grisan, A. Ruggeri, "Detection of optic disc in retinal images by means of a geometrical model of vessel structure", *IEEE Trans. on Medical Imaging*, vol. 23, issue. 10, pp. 1189 – 1195, October 2004.
-

-
- [84] A. Hoover, M. Goldbaum, "Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels", *IEEE Trans. on Medical Imaging*, vol. 22, issue. 8, pp. 951 – 958, August 2003.
  - [85] K. L. Hwee, O. Chutatape, "Blood vessel tracking technique for optic nerve localisation for field 1-3 color fundus images", in *Proc. the 5th Pacific Rim Conference on Multimedia, the 2003 Joint Conference of the 4th International Conference on Information, Communications and Signal Processing*, vol. 3, pp. 1437 – 1441, December 2003.
  - [86] H. Li, O. Chutatape, "Boundary detection of optic disk by a modified ASM method", *Pattern Recognition*, vol. 36, no. 9, pp. 2093 – 2104, 2003.
  - [87] H. Li, O. Chutatape, "A model-based approach for automated feature extraction in fundus images", in *Proc. the 9th IEEE International Conference on Computer Vision*, vol. 1, pp. 394 – 399, October 2003.
  - [88] T. Walter, J. C. Klein, P. Massin, A. Erginay, "A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina", *IEEE Trans. on Medical Imaging*, vol. 21, issue. 10, pp. 1236 – 1243, October 2002.
  - [89] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, M. Goldbaum, "Detection of blood vessels in retinal images using two-dimensional matched filters", *IEEE Trans. on Medical Imaging*, vol. 8, issue. 3, pp. 263 – 269, September 1989.
  - [90] A. D. Hoover, V. Kouznetsova, M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response", *IEEE Transactions on Medical Imaging*, vol. 19, issue. 3, pp. 203 – 210, March 2000.
  - [91] R. Nekovei, Y. Sun, "Back-propagation network and its configuration for blood vessel detection in angiograms", *IEEE Trans. on Neural Networks*, vol. 6, issue. 1, pp. 64 – 72, January 1995.
  - [92] O. Chutatape, Z. Liu, S. M. Krishnan, "Retinal blood vessel detection and tracking by matched Gaussian and Kalman Filters", in *Proc. the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 6, pp. 3144 – 3149, November 1998.
  - [93] L. Gang, O. Chutatape, S. M. Krishnan, "Detection and measurement of retinal vessels in fundus images using amplitude modified second-order Gaussian filter", *IEEE Trans. on Biomedical Engineering*, vol. 49, issue. 2, pp. 168 – 172, February 2002.
-

-
- [94] A. Can, H. Shen, J. N. Turner, H. L. Tanenbaum, B. Roysam, "Rapid automated tracing and feature extraction from retinal fundus images using direct exploratory algorithms", *IEEE Trans. on Information Technology in Biomedicine*, vol. 3, no. 2, pp 125 – 138, June 1999.
  - [95] D. Ye, L. Zheng, "Fundus image processing and feature classification based on mathematical morphology method", in *Proc. the 17th Annual Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2, pp. 1015 – 1016, September 1995.
  - [96] X. Zhang, O. Chutatape, "Detection and classification of bright lesions in color fundus images", in *Proc. IEEE International Conference on Image Processing*, vol. 1, pp. 139 – 142, October 2004.
  - [97] Z. B. Sbeh, L. D. Cohen, G. Mimoun, G. Coscas, "A new approach of geodesic reconstruction for drusen segmentation in eye fundus images", *IEEE Trans on Medical Imaging*, vol. 20, issue. 12, pp. 1321 – 1333, December 2001.
  - [98] M. J. Cree, J. A. Olson, K. C. McHardy, J. V. Forrester, P. F. Sharp, "Automated microaneurysm detection", in *Proc. IEEE International Conference on Image Processing*, vol. 3, pp. 699 – 702, September 1996.
  - [99] Y. Mo, D. Xiao, "Recognizing the glaucoma from ocular fundus image by image analysts", in *Proc. the 12th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 178 – 179, November 1990.
  - [100] S. K. Law, Y. H. Sohn, D. Hoffman, K. Small, A. L. Coleman, J. Caprioli, "Optic disk appearance in advanced age-related macular degeneration", *American Journal of Ophthalmology*, pp. 138(1):135 – 6, July 2004.
  - [101] http://www.dai.ed.ac.uk/CVonline/LOCAL_COPIES/OWENS/LECT9/, Lecture Notes, 2005.
  - [102] C. Harris, M. Stephens, "A Combined Corner and Edge Detector", in *Proc. the 4th Alvey Vision Conference*, pp. 147 – 151, 1988.
  - [103] <http://www.revoptom.com/archive/issue/ro08f6t4.htm>, 2005.
  - [104] C. Loop, Z. Zhang, "Computing rectifying homographies for stereo vision", in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 125 – 131, June 1999.
  - [105] Q. Chen, W. Wang, X. Ji, "Ophthalmology", Second Edition, Shanghai Medical University Press, 1993.
-

-
- [106] R. Malladi, J. A. Sethian, B. C. Vemuri, "Shape modeling with front propagation: a level set approach", *IEEE Trans on Pattern Analysis and Machine Intelligence*, vol. 17, issue. 2, pp. 158 – 175, February 1995.
  - [107] G. J. Jaffe, J. Caprioli, "Optical coherence tomography to detect and manage retinal disease and glaucoma", *American Journal of Ophthalmology*, pp. 137:156 – 169, January 2004.
  - [108] J. Caprioli J, U. Klingbeil, M. Sears, B. Pope, "Reproducibility of optic disc measurements with computerized analysis of stereoscopic video images", *Arch Ophthalmol.*, pp. 104(7):1035 – 9, July 1986.
  - [109] U. Klingbeil, "Fundus geometry measured with the analyzing stereo video ophthalmoscope", book chapter, internal circulation documents.
  - [110] R. Varma, G. L. Spaeth, "The PAR IS-2000: a new system for retinal digital image analysis", *Ophthalmic Surg.*, pp. 19(3):183 – 92, March 1988.
  - [111] P. Hrynchak, N. Hutchings, D. Jones, T. Simpson, "A comparison of cup-to-disc ratio measurement in normal subjects using optical coherence tomography image analysis of the optic nerve head and stereo fundus biomicroscopy", *Ophthalmic and Physiological Optics*, vol. 24, issue 6, pp. 543 – 550, November 2004.
  - [112] K. H. Mok, V. W. Lee, K. F So, "Retinal nerve fiber loss pattern in high-tension glaucoma by optical coherence tomography", *Journal of Glaucoma*, pp. 12(3):255 – 9, June 2003.
  - [113] T. A. El Beltagi, C. Bowd, C. Boden, P. Amini, P. A. Sample, L. M. Zangwill, R. N. Weinreb, "Retinal nerve fiber layer thickness measured with optical coherence tomography is related to visual function in glaucomatous eyes", *Ophthalmology*, pp. 110(11):2185 – 91, November 2003.
  - [114] K. H. Mok, V. W. Lee, K. F. So, "Retinal nerve fiber loss in high- and normal-tension glaucoma by optical coherence tomography", *Optom Vis Sci.*, pp. 81(5):369 – 72, May 2004.
  - [115] N. Ayache, C. Hansen, "Rectification of images for binocular and trinocular stereovision", in *Proc. IEEE 9th International Conference on Pattern Recognition*, vol. 1, pp. 11 – 16, November 1988.
  - [116] <http://fconyx.ncifcrf.gov/~lukeb/kmeans.html>, "K-means clustering", 2005.
  - [117] "Computer vision and machine intelligence", lecture notes, internal circulation documents, 2004.
  - [118] T. S. Huang, O. D. Gauger, "Some properties of the E matrix in two-view motion estimation", *IEEE trans. on Pattern Analysis and Machine Intelligence*, vol. 11, no. 12, pp. 1310 – 1312, December 1989.
-

- [119] J. Xu, O. Chutatape, "3-D optic disk reconstruction from low-resolution stereo images based on combined registrations and sub-pixel matching ", in *Proc. the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1, pp. 1825 – 1828, September 2004.
- [120] J. Xu, O. Chutatape, "Comparative Study of Two Calibration Methods on Fundus Camera", in *Proc. the 25th Annual international Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1, pp. 576 – 579, September, 2003.
- [121] E. H. Zhang, Y. Zhang, T. X. Zhang, "Automatic retinal image registration based on blood vessels feature point", in *Proc International Conference on Machine Learning and Cybernetics*, vol. 4, pp. 2010 – 2015, November 2002.
- [122] K. Muhlmann, D. Maier, R. Hesser, R. Manner, "Calculating dense disparity maps from color stereo images, an efficient implementation", in *Proc. IEEE Workshop on Stereo and Multi-Baseline Vision, (SMBV 2001)*, pp. 30 – 36, December 2001.
- [123] R. C. Gonzalez, R. E. Woods, "Digital image processing", Second Edition, Prentice Hall, 2002.
- [124] C. Davatzikos, X. Tao, D. Shen, "Hierarchical active shape models, using the wavelet transform", *IEEE Trans on Medical Imaging*, vol. 22, issue. 3, pp. 414 – 423, March 2003.
- [125] V. Chalana, Y .Kim, "A methodology for evaluation of boundary detection algorithms on medical images", *IEEE Trans on Medical Imaging*, vol. 16, issue. 5, pp. 642 – 652, October 1997.
- [126] A. Manivannan, P. Hossain, P. Vieira, J. Hipwell, N. Kirkpatrick, C. Lim, P. Sharp, J. Forrester, "Digital imaging with a scanning laser ophthalmoscope", in *Proc IEE Colloquium on Biomedical Applications of Photonics (Digest No. 1997/124)*, pp. 8/1 – 8/6, April 1997.
- [127] R. I. Hartley, C. Silpa-Anan, "Reconstruction from two views using approximate calibration", in *Proc. the 5th Asian Conference on Computer Vision*, pp. 338 – 343, January 2002.
- [128] R. Deriche, Z. Zhang, Q. T. Luong, O. D. Faugeras, "Robust recovery of the epipolar geometry for an uncalibrated stereo rig", in *Proc. the 3rd European Conference on Computer Vision*, vol. 1, pp. 567 – 576, May 1994.
- [129] K. Yogesana, C. J. Barry, L. Jitskaia, R. H. Eikelboom, W. H. Morgan, P. H. House, P. P. van Saarloos, "Software for 3-D visualization/analysis of optic-disc images", *IEEE Engineering in Medicine and biology*, vol. 18, issue 1, pp. 43 – 49, January/February 1999.

-
- [130] T. Kanade, M. Okutomi, "A stereo matching algorithm with an adaptive window: theory and experiment", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 16, no. 9, pp. 920 – 932, September 1994.
  - [131] M. Okutomi, T. Kanade, "A multiple-baseline stereo", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 15, no. 4, pp. 353 – 363, April 1993.
  - [132] H. Nguyen, A. Roychoudhry, A. Shannon, "Classification of diabetic retinopathy lesions from stereoscopic fundus images", in *Proc. the 19th International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1, pp. 426 – 428, November 1997.
  - [133] J. Jeon, K. Kim, C. Kim, Y. Ho, "A robust stereo-matching algorithm using multiple-baseline cameras", in *Proc. IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*, vol. 1, pp. 263 – 266, August 2001.
  - [134] A. P. Whichello, Y. Hong, "Document image mosaicing", in *Proc. the 14th International Conference on Pattern Recognition*, vol. 2, pp. 1081 – 1083, August 1998.
  - [135] H. Hattori, A. Maki, "Stereo matching with direct surface orientation recovery", in *Proc. British Machine Vision Conference*, 1998.
  - [136] http://www.vision.caltech.edu/bouguetj/calib_doc/htmls/links.html, 2005.
  - [137] S. J. Eglen, A. V. Ooyen, "Modelling retinal mosaic development with dendritic outgrowth and lateral cell movement", in *Proc. the 9th International Conference on Artificial Neural Networks*, September 1999.
  - [138] H. Y. Shum, R. Szeliski, "Construction and refinement of panoramic mosaic with global and local alignment", in *Proc. the 6th International Conference on Computer Vision*, pp. 953 – 956, January 1998.
  - [139] L. Di Stefano, M. Marchionni, S. Mattoccia, and G. Neri, "Dense stereo based on the uniqueness constraint", in *Proc. the 16th International Conference on Pattern Recognition*, vol. 3, pp. 657 – 661, August 2002.
  - [140] G. Li, Y. He, "A hierarchical combined feature- and area- based stereo matching algorithm", in *Proc. IEEE International Symposium on Circuits and Systems (ISCAS 2002)*, vol. 2, pp. 277 – 280, May 2002.
  - [141] M. Shimizu, M. Okutomi, "Precise sub-pixel estimation on area-based matching", in *Proc. the 8th IEEE International Conference on Computer Vision*, vol. 1, pp. 90 – 97, July 2001.
-

-
- [142] O. D. Faugeras, S. Maybank, "Motion from point matches: multiplicity of solutions", in *Proc. Workshop on Visual Motion*, pp. 248 – 255, March 1989.
  - [143] P. Sturm, "A case against Kruppa's equations for camera self-calibration", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1199 – 1204, October 2000.
  - [144] R. I. Hartley, "An algorithm for self calibration from several views", in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 908 – 912, June 1994.
  - [145] R. I. Hartley, "In defence of the 8-point Algorithm", in *Proc. the 5th International Conference on Computer Vision*, pp. 1064 – 1070, June 1995.
  - [146] R. I. Hartley, R. Gupta, T. Chang, "Stereo from uncalibrated cameras", in *Proc. IEEE Computer Society Conference on Computer vision and Pattern Recognition*, pp. 761 – 764, June 1992.
  - [147] <http://www.cs.umd.edu/~adeli/Calib.html>, "A method for camera calibration", 2005.
  - [148] S. Bougnoux. "From projective to Euclidean space under any practical situation, a criticism of self-calibration", in *Proc. the 6th International Conference on Computer Vision*, pp. 790 – 796, January 1998.
  - [149] A. Heyden, K. Astrom, "Euclidean reconstruction from constant intrinsic parameters", in *Proc. the 13th International Conference on Pattern Recognition*, vol. 1, pp. 339 – 343, August 1996.
  - [150] J. Gluckman, S. K. Nayar, "Planar catadioptric stereo: geometry and calibration", in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 1999 – 28, June 1999.
  - [151] H. Yu, F. Wu, Z. Ruan, B. Zhang, "The study of Kruppa equations", in *Proc. the 5th International Conference on Signal Processing Proceedings (WCCC-ICSP 2000)*, vol. 2, pp. 966 – 970, August 2000.
  - [152] F. Devernay, O. D. faugeras, "Automatic calibration and removal of distortion from scenes of structured environments", in *Proc. the SPIE Conference on Investigative and Trial Image*, vol. 2567, 1995.
  - [153] O. Chutatape, Y. Liu, "Determination of fundamental matrix by Levenberg-Marquardt method", in *Proc. IEEE 1st Pacific-Rim Conference on Multimedia (IEEE-PCM2000)*, pp 302 – 305, December 2000.
  - [154] Q. T. Luong, R. Deriche, O. D. Faugeras, T. Papadopoulo, "On determining the fundamental matrix: analysis of different methods and experimental results", Technical Report RR-1894, INRIA 1993.
-

-
- [155] <http://www.tnt.de/vispar/e-pv2.htm>, “Visual papilla recognition and glaucoma diagnosis support”, 2005.
  - [156] K. G. Gundersen, A. Heijl, B. Bengtsson, “Comparability of three-dimensional optic disc imaging with different techniques. A study with confocal scanning laser tomography and raster tomography”, *Acta Ophthalmol Scand*, pp. 78(1):9 – 13, February 2000.
  - [157] G. Lin, C. V. Stewart, B. Roysam, K. Fritzsche, G. Yang; H. L. Tanenbaum, “Predictive scheduling algorithms for real-time feature extraction and spatial referencing: application to retinal image sequences”, *IEEE Transactions on Biomedical Engineering*, vol. 51, issue. 1, pp. 115 – 125, January 2004.
  - [158] H. Shen, B. Roysam, C. V. Stewart, J. N. Turner, H. L. Tanenbaum, “Optimal scheduling of tracing computations for real-time vascular landmark extraction from retinal fundus images”, *IEEE Trans. on Information Technology in Biomedicine*, vol. 5, issue. 1, pp. 77 – 91, March 2001.
  - [159] <http://luna.msr.uni-bremen.de/glaucad/workplan.html>, 2005.
  - [160] N. J. Sheen, J. E. Morgan, J. L. Poulsen, R. V. North, “Digital stereoscopic analysis of the optic disc: evaluation of a teaching program”, *Ophthalmology*, pp. 111(10):1873 – 9, October 2004.
  - [161] J. P. W. Pluim, J. B. A. Maintz, M. A. Viergever, “Mutual-information-based registration of medical images: a survey”, *IEEE Trans on Medical Imaging*, vol. 22, issue. 8, pp. 986 – 1004, August 2003.
  - [162] T. Chanwimaluang, G. Fan, “An efficient algorithm for extraction of anatomical structures in retinal images”, in *Proc. International Conference on Image Processing*, vol. 1, pp. I– 1093–6, September 2003.
  - [163] J. Lowell, A. Hunter, D. Steel, A. Basu, R. Ryder, R. L. Kennedy, “Measurement of retinal vessel widths from fundus images based on 2-D modeling”, *IEEE Trans on Medical Imaging*, vol. 23, issue. 10, pp. 1196 – 1204, October 2004.
  - [164] J. T. Fulton, “Standardized human eye”, <http://www.4colorvision.com/standeye.htm>, 2005.
  - [165] Y. Liu, “3-D reconstruction of ocular fundus images”, a MEng thesis, in school of EEE, Nanyang Technological University, Singapore, 2002.
  - [166] <http://academia.hixie.ch/bath/eye/home.html>, 2005
  - [167] D. A. Forsyth, J. Ponce, “Computer Vision, A Modern Approach”, Prentice Hall, 2003.
  - [168] E. Trucco, A. Verri, “Introductory Techniques for 3-D Computer Vision”, Prentice Hall, 1998.
-

-
- [169] R. Hartley, A. Zisserman, “Multiple View Geometry in Computer Vision”, Cambridge University Press, June 2000.
- [170] P. K. Sahoo, S. Soltani, A. K. C. Wong, Y. C. Chen, “Survey of thresholding techniques”, Computer Vision , Graphics, and Image Processing, vol. 41, no. 2, pp. 233 – 260, 1988.
- [171] T. F. Cootes, C. J. Taylor, “Active shape models – ‘smart snakes’ ”, in Proc. the British Machine Vision Conference, Leeds, UK, pp. 266 – 275, Springer Verlag, London, 1992.

Appendix A

Image Rectification

Image rectification is an important component of stereovision computation. It is a process to transform the common stereo configuration into parallel stereo configuration to facilitate the correspondence detection. This section gives a brief introduction of image rectification citing from [65].

Figure A.1: Image rectification

Image rectification recalculates pixel coordinates using a linear transformation in projective space. This is illustrated in Figure A.1, where $\mathbf{C}$, $\mathbf{C}'$ are optical centers. Image planes with dashed borders show input before rectification, and image planes with solid borders and parallel horizontal epipolar lines (dotted lines in rectified images) are the desired result. Points in the left and right images are bilinearly related through the fundamental matrix $\mathbf{F}$, $\tilde{\mathbf{m}}^T \mathbf{F} \tilde{\mathbf{m}} = 0$. We seek the two 3×3 transformation matrices $\mathbf{A}$, $\mathbf{B}$ that rectify coordinates (denoted with $\tilde{\cdot}$) of points in left and right images, respectively, $\tilde{\mathbf{m}} = \mathbf{A} \tilde{\mathbf{m}}$, $\tilde{\mathbf{m}}' = \mathbf{B} \tilde{\mathbf{m}}'$. The fundamental matrix of the rectified images $\tilde{\mathbf{F}} = \mathbf{A}^{-T} \mathbf{F} \mathbf{B}^{-1}$ should correspond to epipoles that moved along horizontal axes to $-\infty$ or ∞ for the left and right images respectively.

To complete this task we need to set values in the transformation matrices $\mathbf{A}$ and $\mathbf{B}$ – a solution is given in [115] which we summarize here. Since the transformation is constrained, the number of unknowns is reduced; image coordinate transformations using matrices $\mathbf{A}$, $\mathbf{B}$ should not change the position of optical centers, but should align two distinct image planes into one common image plane that is parallel to the line joining $\mathbf{C}$ and $\mathbf{C}'$, and perpendicular to both newly recalculated optical axes. Moreover, it is desired that the corresponding epipolar lines should have the same vertical coordinate, as this simplifies calculations.

Image coordinates after rectification are

$$\tilde{\mathbf{m}} = \begin{bmatrix} U \\ V \\ W \end{bmatrix} = \mathbf{A} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}, \text{ and } \tilde{\mathbf{m}}' = \begin{bmatrix} U' \\ V' \\ W' \end{bmatrix} = \mathbf{B} \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} \quad (\text{A.1})$$

Recall the projection matrix from equation (3.19)

$$\begin{bmatrix} wu \\ wv \\ w \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} & h_{14} \\ h_{21} & h_{22} & h_{23} & h_{24} \\ h_{31} & h_{32} & h_{33} & h_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (\text{A.2})$$

We have two such matrices $\mathbf{H}$, $\mathbf{H}'$ for left and right images before rectification. The 3×3 submatrix on the left side of $\mathbf{H}$ is composed of three rows (1×3 vectors) that we denote as $\mathbf{h}_1$, $\mathbf{h}_2$ and $\mathbf{h}_3$; similarly for the right image, the matrix $\mathbf{H}'$ gives three vectors $\mathbf{h}'_1$, $\mathbf{h}'_2$ and $\mathbf{h}'_3$. $\mathbf{C}$ and $\mathbf{C}'$ are coordinates of the optical centers. The transformation matrices that perform rectification are then

$$\mathbf{A} = \begin{bmatrix} ((\mathbf{C} \times \mathbf{C}') \times \mathbf{C})^T \\ (\mathbf{C} \times \mathbf{C}')^T \\ ((\mathbf{C} - \mathbf{C}') \times (\mathbf{C} \times \mathbf{C}'))^T \end{bmatrix} \begin{bmatrix} \mathbf{h}_2 \times \mathbf{h}_3, & \mathbf{h}_3 \times \mathbf{h}_1, & \mathbf{h}_1 \times \mathbf{h}_2 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} ((\mathbf{C} \times \mathbf{C}') \times \mathbf{C}')^T \\ (\mathbf{C} \times \mathbf{C}')^T \\ ((\mathbf{C} - \mathbf{C}') \times (\mathbf{C} \times \mathbf{C}'))^T \end{bmatrix} \begin{bmatrix} \mathbf{h}'_2 \times \mathbf{h}'_3, & \mathbf{h}'_3 \times \mathbf{h}'_1, & \mathbf{h}'_1 \times \mathbf{h}'_2 \end{bmatrix} \quad (\text{A.3})$$

This procedure is computationally inexpensive. Only two 3×3 transformation matrices need be stored. Notice that the rectification is a linear transformation in projective space that preserves straight lines. If an image consists of linear segments, then it is sufficient to rectify end points of these segments. The procedure can be easily generalized to three and more images [115]. After image rectification, the corresponding epipolar lines are aligned to the same horizontal scan-lines. Detection of correspondences is then simplified to search only along horizontal scan-line.

Appendix B

K-means Algorithm

K-means algorithm is one of the clustering methods to divide the sample set of processed patterns into subsets (clusters) based on the mutual similarity of subset elements. The number of clusters, k , needs to be determined. The procedure of the k-means algorithm is presented as follows citing from [65][116][117].

1. Define the number of clusters, k , and arbitrarily choose k samples as initial cluster centroids, i.e. $\mathbf{z}_1(t), \mathbf{z}_2(t), \dots, \mathbf{z}_k(t)$, where t is the iteration number.
2. Assign each sample $\mathbf{m}$ to the group that has the closest centroid. At the t^{th} iterative step, distribute the sample $\mathbf{m}$ among k clusters $\{S_i\}$ according to the following rule:

$$\mathbf{m} \in S_j(t) \quad \text{if } \|\mathbf{m} - \mathbf{z}_j(t)\| < \|\mathbf{m} - \mathbf{z}_i(t)\| \text{ for } i=1, 2, \dots, k, i \neq j.$$

3. When all samples have been assigned, recalculate the positions of the k centroids.

$$\mathbf{z}_j(t+1) = \frac{1}{N_j} \sum_{\mathbf{m} \in S_j(t)} \mathbf{m}, \quad j=1, 2, \dots, k.$$

where N_j is the number of the samples in cluster S_j .

4. Repeat steps 2 and 3 until all the centroids no longer move.

Appendix C

Fundus Image Data

The tested stereo images provided by National University Hospital are demonstrated in the following.