Image Mosaicing of Neonatal Retinal Images

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

Master of Science (by Research)
in
Computer Science and Engineering

by

Akhilesh Bontala
200601019
akhilesh.bsr@research.iiit.ac.in

Center for Visual Information Technology
International Institute of Information Technology
Hyderabad - 500 032, INDIA
December 2013

International Institute of Information Technology
Hyderabad, India

CERTIFICATE

It is certified that the work contained in this thesis, titled "Image Mosaicing of Neonatal Retinal Images" by Akhilesh Bontala, has been carried out under my supervision and is not submitted elsewhere for a degree.

Date	Adviser: Dr. Jayanthi Sivaswamy

Abstract

Image mosaicing is a data fusion technique used for increasing the field of view of an image. Deriving the mosaiced image entails integrating information from multiple images. Image mosaicing permits overcoming the limitations of a camera lens and help create a wide field of view image of a 3D scene and hence has a wide range of applications in various domains including medical imaging. This thesis concerns the task of mosaicing specific to neonatal retinal images for aiding the doctors in the diagnosis of Retinopathy of prematurity (ROP). ROP is a vascular disease that affects low birth-weight, premature, infants. The prognosis of ROP relies on information on the presence of abnormal vessel growth and fibrosis in periphery. Diagnosis is based on a series of images obtained from a camera (such as RetCam), to capture the complete retina. Typically, as many as 20 to 30 images are captured and examined for diagnosis. In this thesis, we present a solution for mosaicing the RetCam images so that a comprehensive and complete view of the entire retina can be obtained in a single image for ROP diagnosis. The task is challenging given that the quality of the images obtained is variable. Furthermore, the presence of large spatial shift across consecutive frames makes them virtually unordered. We propose a novel, hierarchical system for efficiently mosaicing an unordered set of RetCam images. It is a two-stage approach in which the input images are first partitioned into subsets and images in each sub-set are spatially aligned and combined to create intermediate results. Given n images, the number of registrations required to generate a mosaic by conventional approaches to mosaicing is $O(n^2)$ whereas it is $O(n)$ for the proposed system. These images are then again spatially aligned and combined to create a final mosaic. An alignment technique for low quality retinal images and a blending method for combining images based on vessel quality is also designed as part of this framework. Individual components of the system are evaluated and compared with other approaches. The overall system was also evaluated on a locally-sourced dataset consisting of neonatal retinal images of 10 infants with ROP. Quantitative results show that there is a substantial increase in the field of view and the vessel extent is also improved in the generated mosaics. The generated mosaics have been validated by the experts to provide sufficient information for the diagnosis of ROP.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Image Mosaicing

In optics, field of view (FOV) is defined as the part of scene that is visible at an instance from a given position. It is measured as the angular extent of the scene in degrees. Humans have almost 180° FOV in the horizontal direction and depending on the placement of the eye, it varies from one animal to other. Birds, in general, have much larger FOV compared to animals (Eg: Pigeons has nearly 360° FOV) as their eyes are positioned on the opposite sides of the head. But when it comes to cameras, they have very limited FOV. With a fixed image sensor, the field of view of a camera depends upon the focal length of its lens. A standard lens has a FOV ranging from 40°– 60° and in case of a wide-angle lens it can be from 60°– 80°. There are special type of lenses which provide around 180° FOV, but these lenses introduce barrel distortion (Ex: Fisheye lens). Also these type of lenses are rare and are used for very specific applications. Therefore, in order to obtain an image of a broad area/wide scene, mosaicing is used.

Image mosaicing is a technique used to combine multiple overlapping images of a 3D scene and create a single high resolution image. Using mosaicing, we can construct mosaics of scenes which are generally very large to be captured using a single image. It can be argued that the scene can be covered in the FOV of the camera by moving away from the scene. This is not always possible as in many cases new objects occlude the scene and even if the scene is captured, the finer details are compromised. For example, a full mountain landscape can be captured from farther distance at the cost of losing the finer details of rocks and bushes. Therefore, the amount of visible scene is traded with the level of details.

In order to solve this problem, various views of the scene are obtained such a way that each view captures a sub-region of the scene preserving its details. These are then aligned to a single co-ordinate system and combined to create a large field of view mosaic of the scene without compromising the image resolution (See Fig:1.1). These images are captured from various view points in such a way that the required area of interest is covered and also that have fair amount of overlap between them. Currently, most of the digital cameras in the market can perform similar task. By panning the camera in a particular direction, the panoramic view of the scene can be obtained. But for general mosaicing, the camera motion is not constrained as in many cases a much more complex motion is required to obtain the complete scene.

Figure 1.1 A high resolution mosaic of the complete Sydney skyline (ImageSource:[1])

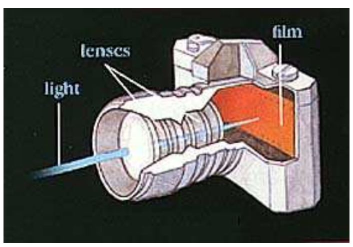
Mosaicing has been in use as early as from the start of 20^{th} century [29]. The images captured from hill-tops and airplanes were manually mosaiced to create large photo-maps. After the improvement in the computer technology, the need to develop automatic computerized technique increased. In the recent years, image mosaicing has been an active area of research in computer vision and computer graphics. Image mosaicing is now used in remote sensing [10], medical [58] and industrial applications [28]. Apart from the primary goal (to improve field of view and resolution), the construction of mosaics have attracted a wide range of applications like video compression [32], virtual reality application [39] etc.

1.2 Background of Retinopathy Of Prematurity

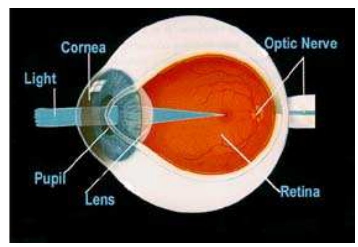
In this section, we introduce the medical domain of retina and in particular discuss about a disease Retinopathy Of Prematurity (ROP) which is of interest for this thesis. A mosaicing algorithm for retinal images is proposed in this work which is useful for ROP diagnosis.

1.2.1 About Retina

An eye is an organ that detects light and converts it into electrical impulses and sends them to brain. The overall anatomy of eye can be compared to a camera as they work in a similar way. The light rays enter through the cornea, bend through the pupil and are focused by the lens on to the back of the eye. The thin layer of membrane at the back of the eye is the retina. It contains the light receptors which convert the light rays into electrical impulses and sends them to the brain through optic nerve. Analogous to a camera, the cornea, pupil and retina act as lens cover, aperture and film respectively (see Fig: 1.2).



(a) Camera

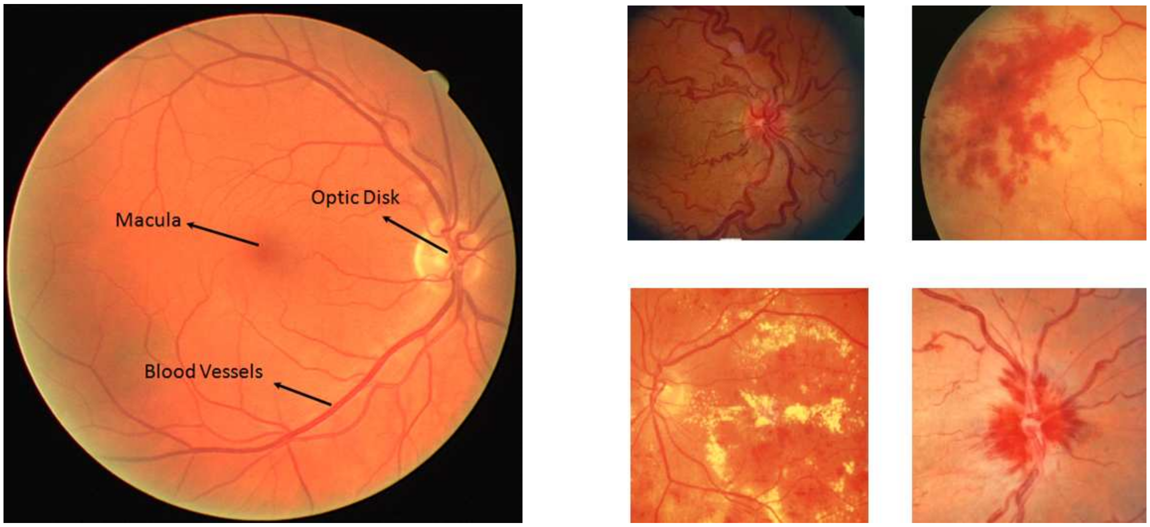


(b) Human Eye

Figure 1.2 Comparison between a Camera and a Human Eye (Image Source:[2])

Retina is a spherical structure which can be compared to an inner surface of a sphere. The major structures in retina are the optic disk, macula and blood vessels (see Fig 1.3). Optic disk is a region through which the blood vessels enter and exit (arteries enter with oxygenated blood and veins exist with deoxygenated blood). Also the optic nerve, which transmits the visual information to the brain starts here. Macula is the central region of retina and consists of fovea. Fovea, a small dip, is the true focus point which is responsible for the high resolution central vision. There are many diseases and disorders which affect retina eg: Age-related Macular Degeneration (AMD), Diabetic Retinopathy (DR), Retinopathy of Prematurity (ROP), Glaucoma etc. Various types of abnormalities in different regions of the retina are seen in these diseases (Fig 1.3).

With the advancement in the digital technology, fundus cameras came into existence. Using these cameras, the digital image of the retina can be captured and stored for further analysis. In general, the pupil is dilated through drugs so that the retina is sufficiently illuminated to obtain a good quality image. Due to the spherical nature, the complete retina cannot be captured in a single instance. Therefore, in order to obtain the periphery, multiple images from different viewpoints are taken. For many diseases, the central part of the retina is the region of interest where the abnormalities occur. So a view containing OD and macula is sufficient for diagnosis.



(a) Central view of a retina with major structures (b) Abnormalities in different regions

Figure 1.3 Structure of a retina and various types of abnormalities

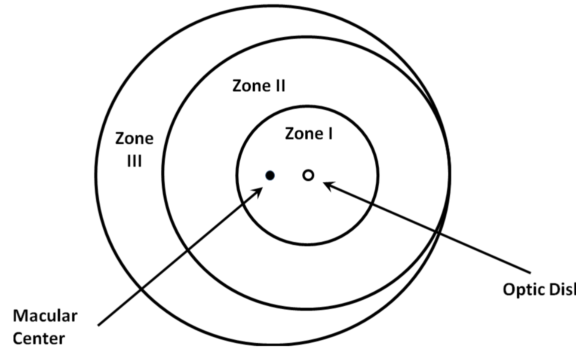
1.2.2 Retinopathy of Prematurity

Retinopathy of prematurity (ROP) is an eye disease that affects low birth-weight premature babies. In a normal mature baby, blood vessels are fully developed at the time of birth spreading across the complete retina. However, in a premature infant, they are not yet fully developed, which leads to the growth of abnormal vessels in the regions where the development was disrupted. These vessels may cause bleeding in the eye or form a scar tissue and result in retinal detachment. In severe cases, ROP may lead to complete blindness [21]. Due to the advancement in neonatal intensive care units, there is a huge increase in the survival of very low birth weight and extremely premature babies whose chances of survival were very little in past. As these infants have high risk of developing ROP, its detection is of interest.

ROP is described using a number of parameters [27] like stage, zone, extent etc which are used for estimating the prognosis and carrying out the treatment. It is classified into different stages based on findings observed at the junction between the vascularized and avascular retina (Fig 1.4a). The circumferential extent of ROP is based on the clock hours (1-12) and the location of the disease is described using zones centered on optic disk (Fig 1.4b). Any stage of disease present in zone I is considered critical and infant needs to be monitored continuously. As we move away from zone I, the disease is less dangerous.

Another important parameter for the classification is the presence of Plus disease. Plus disease describes the abnormal levels of vascular tortuosity and dilation in the retinal vessels which reflects the increase in blood flow. For ROP diagnosis, a physical examination of the complete retina is done using a binocular

(a) An elevated ridge found in stage 3 ROP



(b) Different zones of retina

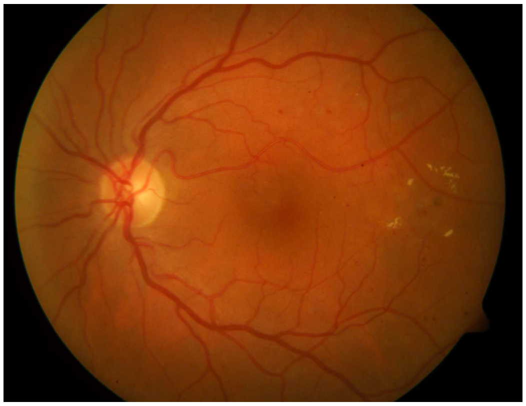
Figure 1.4 Classification of ROP

indirect ophthalmoscope. During the examination, it is determined how far the retinal blood vessels have grown (the zone) and whether the vessels are growing flat along the wall of the eye (stage) or not. The posterior pole (central region around OD and maccula) is examined for Plus disease and the periphery of the retina for finding the stage of ROP. Early diagnosis is very important to make the treatment successful as the disease can advance very quickly. As this disease is usually seen in both the eyes this is a potentially blinding condition. Therefore, if detected and treated early, can result in prevention of the blindness in these children [26]

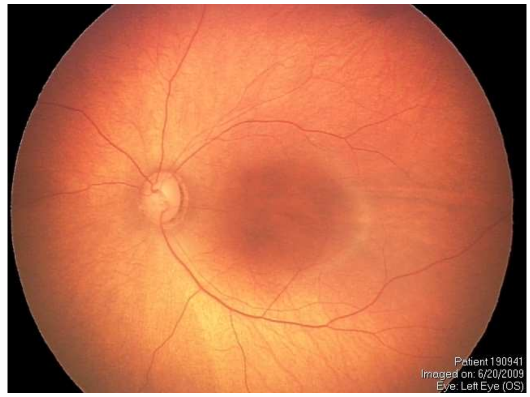
For imaging a neonatal retina, a special type of camera called RetCam is used. RetCam is a wide-angle fundus camera designed primarily for imaging the retina in babies. It is an easy to use, portable system with rapid image capture. Also since the periphery of the retina is of interest in this case, all the views of the retina are needed for diagnosis of ROP. The images obtained from a RetCam are very different from a fundus image of an adult retina because in premature babies the retina is not full developed (Fig 1.5). The retinal pigment epithelium is very thin, due to which even the choroidal vessels are visible. Also sub-optimal pupil size in infants results in uneven illumination and makes it difficult for the camera to focus. Even the pixel resolution of images is very low compared to normal fundus images as the camera is designed to capture the extreme periphery of the retina which is not possible using a standard fundus camera. Apart from the optics, even the acquisition is challenging as the babies constantly move which results in motion-induced blur in many images.

1.2.3 Existing Research

In ROP, the parameters needed for diagnosis are stage of the disease, zone of the disease and presence of Plus disease. A lot of research has been done in the area of retinal vessel analysis not just for ROP but for general retinal image analysis. Therefore, many systems targeting the Plus disease have been designed as it depends on the tortuosity and dilation of the vessels. Also it is very difficult for experts to



(a) An adult retina



(b) A neonatal retina

Figure 1.5 Adult vs Neonatal retina

judge the degree of vascular change and the decision is often an educated guess, so this introduces large inter-observer variability. Therefore, in order to maintain consistency and eliminate subjectivity in the assessment of Plus disease, a computer aided tool is necessary. A detailed review of these tools is given in [4]. These tools try to accurately obtain the blood vessels and derive a toruosity measure such that it correlates with the experts decision. These measures are then used for automatically detecting ROP. Here, we briefly discuss a few important systems.

In 2002, Heneghan [31] first proposed a technique for characterizing the changes in the blood vessels in ROP. Segmentation is done using morphological filtering followed by hysteresis thresholding. The vessel width at each point is calculated by finding the minimum distance between one edge of vessel image to the other. Significant difference in the vessel width and tortuosity is found between infants with and without ROP. Using this method in ROP screening, sensitivity of 82% and specificity of 75% was achieved by testing it on 23 subjects. Jomier et al. [33] proposed a computerized assessment of Plus disease in ROP. Best quality images taken from an indirect ophthalmoscope are used. The system requires user input to identify major vessels and optic disk margin. It then automatically traces each blood vessel using multi-scale ridge extraction and then calculates the vessel width and tortuosity. The system was able to achieve sensitivity of 80% and specificity of 92% when tested on 20 images. However, both these systems were tested on a small dataset.

A new system called ROPtool was designed by Wallace et al. [66] addressing some of the limitations of [33]. The metric for tortuosity measurement is modified by generating a smooth curve from several points and then calculating the ratio of the length of the vessel to the length of the new smoothed curve. Very high sensitivity of 97% and specificity of 94% was achieved on 185 images of which 37 were abnormal and 148 were normal. However, images chosen were of highest quality and sharp focus. Gelman et al. [25] proposed a new tortuosity measure for diagnosis of Plus disease. A semi-automatic segmentation algorithm called RISA is used and the average diameter is calculated by dividing the area of the pixels by the length of the segment. Tortuosity is defined as the ratio of the sum of angles of deviation along the skeleton to the length of the vessel. The results showed good sensitivity and specificity values and high area under ROC curve. However, even in this method, only high quality images have been selected which might limit the practical validity of the algorithm. A recently published method called Computer Aided Image Analysis of the Retina (CAIAR) was proposed by Wilsone et al [68]. The vessels are segmented using filtered detection measurement based upon maximum likelihood estimation of vessel parameters. 14 tortuosity measurements and 2 width measurements are used for analysis. Satisfactory results were obtained and the correlation of tortuosity with the experts marking was very high. However, its performance in a clinical level ROP detection from a typical image is yet to be tested.

To summarize, most of the tools like ROPtool, RISA and CAIAR are successful in detecting Plus disease and achieved high specificity and sensitivity. However, the main draw of these approaches is that they require a high quality image of the posterior pole of the retina.

1.3 Thesis focus

This thesis is focused on designing a image mosaicing system for low quality retinal images. We propose a novel approach for aligning various views captured and combining them to create a single mosaic. The resultant mosaic contains the complete retina including the posterior pole and the periphery.

1.4 Motivation

In the field of medicine, computer-aided tools have been in use from a long time. They are designed to assist doctors and improve the efficiency and performance of the diagnostic mechanism. The idea is to reduce the amount of effort and eliminate subjectivity in diagnosis. These systems can be at different levels of the process. They can be used for preprocessing the data, detecting and segmenting the regions of interest, quantitative analysis etc and assist in the decision making. Also, to make use of the past data, machine learning and artificial intelligence are used to design these systems. The holy grail would be to have a fully automatic system which can replace the doctor. But currently, most of the systems solve a part of the task automatically or semi-automatically (with some manual intervention) and the final judgment is generally left to the doctor. In the domain of retina, there exist algorithms for localizing optic disk, macula, segmentation of blood vessels, lesions etc.

For ROP diagnosis, a video stream or a set of snapshots is captured by the operator. The view point of the camera is frequently adjusted to obtain various views till the complete retinal surface has been covered. As mentioned earlier, the quality of neonatal retinal images is poor due to various factors. In

Fig:1.6, we can observe the dark regions and poor contrast due to uneven illumination. Varying degree of blur (both focal blur and motion blur) can be seen in these images (see Fig:1.7). Therefore, the operator captures a lot of images from the same view till the required information from the region is obtained. Of the captured set of images, many have either no information or redundant information. To filter out these images for expert diagnosis, we need to manually select good quality images. Even if a good quality set is obtained, doctors manually need to observe these images and virtually reconstruct and imagine the retinal surface for diagnosing the disease. Also, the zone of ROP is of interest for diagnosis and atmost the first zone is completely visible in the OD centric images.

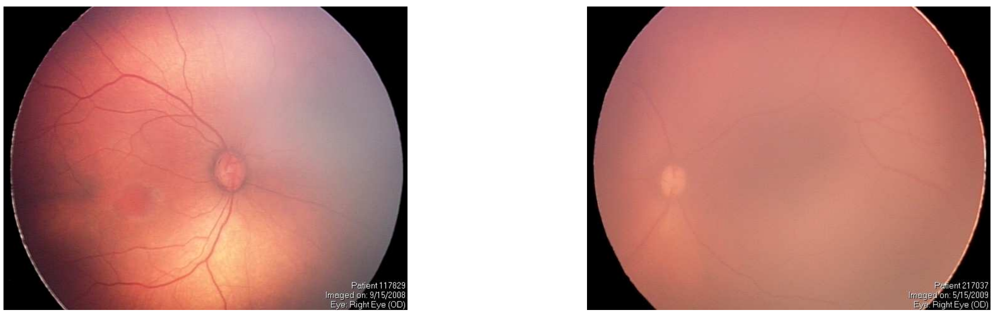


Figure 1.6 Uneven illumination and low contrast of RetCam images

In this work, we try to make the process of diagnosing easier for doctors. Given a large set of images captured by the operator, a mosaic of the complete retina is created which has the best available information for each region. Using such a mosaic, the doctors can inspect the complete retina in a single shot eliminating the need of manual inspection of the large set. All the zones of the retina can be observed in a single image. Also, tools designed for quantitative analysis of tortuosity and dilation (described in 1.2.3) need a single high quality image limiting their practical use. With the mosaic as input, these tools can be more widely used. Also by increasing the extent of vessels to the periphery, the performance of the Plus disease detection increases, making these tools more accurate. The mosaic can also be used to provide feedback to the operator to verify whether all the regions have been captured with reasonable quality or not.

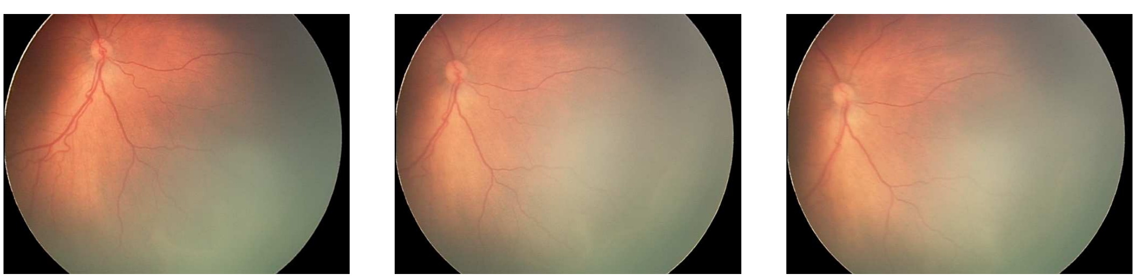


Figure 1.7 Images with variable amount of blur

1.5 Contributions

The key contributions of the thesis are:

Major:

- Mosaicing is the main objective of the thesis. A hierarchical framework for mosaicing of large set of retinal images is proposed.
- An image registration technique for low quality retinal images using an existing descriptor is presented
- Since the technique is hierarchical, the location of the optic disk is desirable for grouping the images. Hence a fast and accurate, optic disk detection algorithm is presented.

Minor:

- Adapted an image stitching method for blending retinal images based on the vessel quality

1.6 Organization

Chapter 2 describes the various steps involved in an image mosaicing framework, existing works in the area of medical imaging and motivation to the proposed method. In Chapter 3, we describe in detail our proposed approach for mosaicing. A color-distance based optic disk detection algorithm is presented in chapter 4. In chapter 5, we evaluate the proposed system and show various qualitative and quantitative results. We conclude the thesis in chapter 6.

Chapter 2

Image Mosacing: Background and Related Work

2.1 Introduction

Data fusion is a broad area of research which deals with integration of multiple data into an accurate and useful representation. In case of images, the process of combining multiple overlapping images of a scene to produce a single superior quality image is called Frame Fusion [15]. Different categories of methods come under the class of frame fusion eg: image mosaicing, superresolution (SR), high dynamic range imaging (HDRI) etc. The purpose of fusing the information varies from one method to other; image mosaicing is employed to extend the field of view, SR to increase the spatial resolution and HDRI to increase the dynamic range of the images. In this work, the problem of image mosaicing is investigated.

Image mosaicing is a process of integrating the information from multiple images to create a mosaic which has extended field of view compared to any single image. As discussed earlier, image mosaicing is used in various domains and has wide range of application. Therefore various techniques have been proposed to accomplish the task in different scenarios. In any mosaicing approach, the two main steps are image alignment and image stitching. In the first stage, the correspondence relationship among the images is established so that the images are spatially aligned. Using the alignment estimates, images are combined in a seamless manner to create a composite image. In the next two sections, we discuss these two problems in detail and the previous work done to solve these problems.

2.2 Image Alignment

A lot of research has been carried out in image alignment/registration as it is widely used in computer vision, medical imaging and remote sensing, not just for mosaicing but in various other frame fusion tasks. Images of same scene can be obtained in multiple ways, they may be captured from different viewpoints, at different times, using different sensors, etc. To accomplish image mosaicing, we assume that the images have been captured using the same sensor, at the same time (with minimum delay) but from different viewpoints. We will first discuss the problem of aligning two images and then proceed to multi-image alignment.

2.2.1 Aligning a Pair of Images

Given a pair of images, the goal of registration is to spatially align them. It is the process of transforming the images into a single co-ordinate system. It is achieved by accurately estimating the underlying transformation between the images and bringing them a into common frame of reference. Therefore, modeling the transformation is the primary step. Assuming a planar surface, it can be a simple Euclidean transform where only rotation and translation are allowed or a complete projective transformation. The number of parameters to be estimated depend upon the degrees of freedom (DOF) allowed in the model. Table 2.1 summarizes all planar transformations. Once a suitable model for transformation is chosen, we need to accurately estimate its parameters. The alignment methods are generally classified into two types [71]: direct method and feature based method. We now describe the basic steps involved in these two categories.

Transformation	Operations	DOF
Euclidean	Translation (T), Rotation (R)	3
Similarity	T, R + Scaling (Sc)	4
Affine	T, R, Sc + Sheer (Sh)	6
Projective	T, R, Sc, Sh + Projection	8

Table 2.1 Planar Transformation

2.2.1.1 Direct Method

Direct methods or area based methods consider the complete image information in order to estimate the parameters. A similarity metric is defined to compare two images and the parameters which maximize the similarity between the fixed and transformed image are considered as an accurate estimate. Some basic similarity measures used are Sum of Squared Difference (SSD) [11], Normalized Cross Correlation (NCC) [51] and Mutual Information (MI) [63].

- SSD: The simplest way to compare the images would be to find the sum of squared differences of their intensity and minimize it. To make it robust, some variations like SAD (sum of absolute difference), WSSD (weighted sum of squared difference), etc. have been proposed

- NCC: Instead of taking difference, the product of intensity can be performed to compare the aligned images. The cross-correlation of the two images is used as similarity metric and its maximum is searched. As the product of higher intensities cause bias, NCC is used. Several improvement on the original NCC are also developed to handle its limitation

- MI: Mutual information is a measure derived from information theory in which the mutual dependence of two images is quantified. It is defined in terms of the marginal entropies of individual images and their joint entropy. In general, it is described as *the measure of amount of information contained in one image about the other image*. Therefore, maximization of the mutual information leads to alignment.

After deciding upon the similarity measure to be used, finding the maximum is a multidimensional optimization problem (number of dimensions correspond to the number of parameters). Although an exact solution can be obtained using a brute force search, it is computationally expensive and not feasible at high dimensions. To speed up the process, a hierarchical approach is often used [8] where a multi-scale image pyramid is constructed. The search is first performed at a coarser level and the result is used to localize the search in the finer levels. In order to obtain an accurate solution, iterative methods are used which refine the estimate after each step and converge to an optimum solution. Advanced optimization techniques like gradient descent, Gauss-Newton minimization, Powell's multidimensional set, etc. are used to efficiently locate the maximum.

2.2.1.2 Feature based Method

In feature based methods, rather than matching the intensities of the images, distinctive features from each image are extracted and matched to obtain the corresponding points. Based on these correspondences, the parameters of the transformation model are estimated. The pipeline of feature based alignment is as follows:

1. Feature Extraction:

(a) Feature Detection: A feature is an interesting piece of information in an image which can be based on color, texture, gradient, etc. In feature detection, a subset of points in the image are selected as interest points. What constitutes to be an interest point completely depends upon the problem. The desired characteristic of any feature is its repeatability as the same points of the scene have to be detected in different images independent of the changes in viewpoint, lighting etc. Some of the well known feature are the edges [14], corner [30], blobs [45], etc. Based on the properties of the features, image processing operations are designed to detect these points.

(b) Feature Description: Once these points are obtained, the patch/region around the point is described using a feature vector. In order to achieve accurate matches, the representation should be robust to noise, invariant to transformations and also distinctive (very diverse descriptors for different points). They generally try to capture the shape, texture or colour information. The popular descriptors used in the literature are Scale-invariant feature transform (SIFT) [42], Speeded Up Robust Features (SURF) [6], Histogram of Oriented Gradients (HOG) [20], etc.

2. Feature Matching

After feature extraction, each image is represented by a set of feature vectors. The simplest way to find the matches is to compare all the features in one image to all other features. To make matching more efficient, different kinds of indexing schemes have been developed which try to find the nearest neighbor in a high dimensional space. [47] proposed a method which uses 1D binary search to efficiently select a list of candidates that lie within a hypercube of the query point. Data structures like k-d tree [54], metric tree [49], parametric-sensitive hashing [56] have been proposed for efficient search.

3. Correspondence Problem

Even by using complex feature extractors which are very robust and distinctive, the matches obtained are not accurate and consist of large number of incorrect or false matches. In this step, the task is to refine the matches and produce a set of points in each image such that they correspond to the same region of the scene. As the transformation model has been chosen (from Table: 2.1), this can be seen as model fitting problem with data containing large number of outliers (data points which do not follow the model). Robust statistical methods like LMedS (Least Median of Squares) [53] and Random Sample Consensus (RANSAC) [22] are used which fit a function to a set of data points without being affected by the presence of outliers.

4. Transformation Estimation

After obtaining the final corresponding points, a guided matching strategy is generally applied to refine the spatial location of the matches. As an estimate of the transformation is known, a small neighborhood around the location is considered and using similarity metrics (described in direct method) the location which results in highest similarity is taken as the exact location. An accurate estimate of the transformation parameters is obtained using these final matches.

2.2.1.3 Comparison: Direct vs Feature based Method

Direct methods make use of the complete information available and every pixel in the image is used to obtain the solution. These methods are preferred when images have no prominent details and colour/intensity (rather than shape or structure) is the only distinctive information. The main drawback is that they have very limited range of convergence. Also these methods fail when the amount of overlap between images is low. Direct methods can be used only when translation and small rotation is allowed between the images.

On the other hand, feature based methods are generally applied when structural information is more prominent than the intensity information. They handle complex transformation and robustly match images that differ in scale, orientation and also projection. The key point in feature based methods is to design robust and distinctive features which are well distributed over the images. This is a very challenging task which also changes from application to application. Hybrid approaches [17] have also been proposed that combine both direct methods and feature based methods and make use of the advantages of both the categories.

2.2.2 Multi-Image Alignment

The goal of image alignment is to transform all the images into a single co-ordinate system. In case of two images, the methods for estimating the transformation between them is discussed earlier. For aligning a set of images, the pair-wise registration can be applied sequentially multiple times. A mosaic is constructed by combining new images as soon as they are available. A new image is aligned with the previous frame (F2F) or with the updated mosaic (F2M). Such incremental methods are computationally efficient and can be used in a real-time scenario. The drawback is that the solution is only locally optimal. Also sequential combination leads to accumulation of error as the number of images increase, which results in visual artifacts.

Multi-Image alignment can also be achieved by extending the pair-wise registration techniques to simultaneous align the images. In simultaneous registration methods, the best transformation among several images is computed by simultaneously minimizing the misregistration error between all pair of images. Both direct and feature based method for pair-wise registration techniques have been extended to achieve global registration. In [16], SIFT features are extracted from all the images and matched using k-d tree to find approximate nearest neighbors. For every image, the images which have high number of feature matches are considered as potential matches. Finally a probabilistic model is used to verify the matches based on the inliers and outliers of RANSAC algorithms. A multi-image alignment based on a direct method is proposed in [55] where parameters of all the images are jointly optimized using Levenberg-Marquardt technique. Another class of methods based on graph algorithms have been proposed in [34, 44] where nodes in the graph represent images and the edge represents the registration error. By using algorithms like shortest path and minimum spanning tree, the overall accumulation error is minimized. The drawback of simultaneous registration is that they are computationally expensive and require all the images to be known in advance.

2.3 Image Stitching

Once all the images are spatially aligned to a global co-ordinate system, the images need to be stitched together to combine a single image. Adding all the images should ideally result in the mosaic as all the overlapping regions represent the same scene and should have same intensity. However, in practice this is not the case as the change in the illumination and exposure differences causes visible seam. Therefore for creating a seamless mosaic, it is to be decided which pixels to use and how to blend them.

The simplest method used for combining the images is a technique called feathering [59]. A weighted averaging is used to add the pixels at the overlapping regions. For each image, the pixels at the center of the image are assigned the maximum value and the weight is gradually decreased near the edges. These weights can be calculated using distance transform. Feathering can handle exposure differences but blurring is present in the combined image. The size of the neighborhood window at the overlap is an important factor as it depends on the size of the features present at the boundary. Coarse structures should blend slowly between images, requiring a larger window whereas a smaller window is required for finer details. An optimal window size is required for a particular feature and increasing the size introduces ghosting artifacts whereas decreasing the size makes the seam visible.

In order to address the above problem, a multi-band blending technique based on Laplacian pyramids has been proposed in [13]. Laplacian pyramids are constructed for each of the image and the pyramids are combined using blending mask obtained from overlaying the registered images. Reconstructing the image using the resultant pyramid gives a seamless image. Using multiple levels of pyramids, information of different frequency is combined over different window sizes. A new technique is presented in [50] where instead of using a 2^{nd} derivative (Laplacian), the images are combined using the 1^{st} derivative (gradient) information. The key idea is to ensure that the gradient of the boundary pixel in one image is made equal to the gradient of the second image making the transition smooth. The modified gradient is reintegrated by iteratively solving the Poisson partial differential equation. An important step in using these blending functions is to find the optimal seam in the overlapping regions such that the images agree. Algorithms like graph cuts and dynamic programming based methods are used to obtain the optimal seam with minimal overlap error. Also, based on the domain, criteria for choosing the right pixels from each image changes from problem to problem.

2.4 Image Mosaicing in Medical Domain

Image mosaicing has attracted a wide range of applications in the medical domain. Due to various factors, the data obtained from many imaging techniques suffer from small field of view, which makes it difficult to obtain a broader view. By applying image mosaicing in such cases, experts have access to information at a macro scale while retaining the micro level details. Therefore, a lot of research is being carried out for mosaicing medical images. Here, we summarize the literature for few types of imaging.

2.4.1 Ultrasound Images

Ultrasound is an imaging technique that uses sound waves (2–15 MHz) and their echoes to visualize the internal structure of human body. A sound pulse is transmitted using a probe and based on the time and intensity of the echoes, the distance, size and shape of the objects inside are calculated. To obtain a 3D ultrasound image of a large organ, multiple sweeps of the probe are needed. Gee et.al [24] proposed a new registration algorithm to combine these sweeps to create a wide field of view ultrasound image. Instead of matching the overlapping regions, the comparison is done only on a single dividing plane which approximately bisects the overlapping region. Therefore, the 3D-3D registration problem is converted into 2D-2D one. A similarity measure based on mutual information is used and a multiresolution search is performed to obtain the matching slices. The major pitfall of this method is that only a part of the available information is taken into account resulting in misregistration. In order to address this problem, Wachinger et.al [65] used simultaneous registration in which information from all the images is considered at the same time to align the images. A multivariate similarity measure consisting of sum of squared differences (SSD), normalized cross-correlation (NCC), mutual information (MI) and correlation-ratio (CR) is used. Though this strategy is computationally very expensive, the resultant mosaics are accurate.

A mosaicing method based on multimodal registration where the information from CT image is used to improve the registration of ultrasound images is proposed in [36]. A simultaneous optimization similar to [65] is used for finding the transformation. In order to reduce the computational costs, the complete algorithm is developed on GPU processors resulting in a 400X speedup compared to a single thread CPU version. The use of GPU in ultrasound mosaicing is further investigated in [12]. A real time system for mosaicing and visualization of 3D ultrasound sequence is developed. Both mosaicing and volume rendering are implemented on GPU to achieve a real-time solution. Instead of an intensity based registration, Ni et.al [48] proposed a feature based registration for ultrasound volumes suggesting that similarity measures used [24, 65], in general may not perform well due to low SNR and the presence of speckle noise. A modified 3D SIFT is used for finding the feature points. These 3D feature points are matched for stitching multiple volumes into a mosaic. It is also shown that the features matched are robust to noise and change in illumination.

2.4.2 Endoscopy

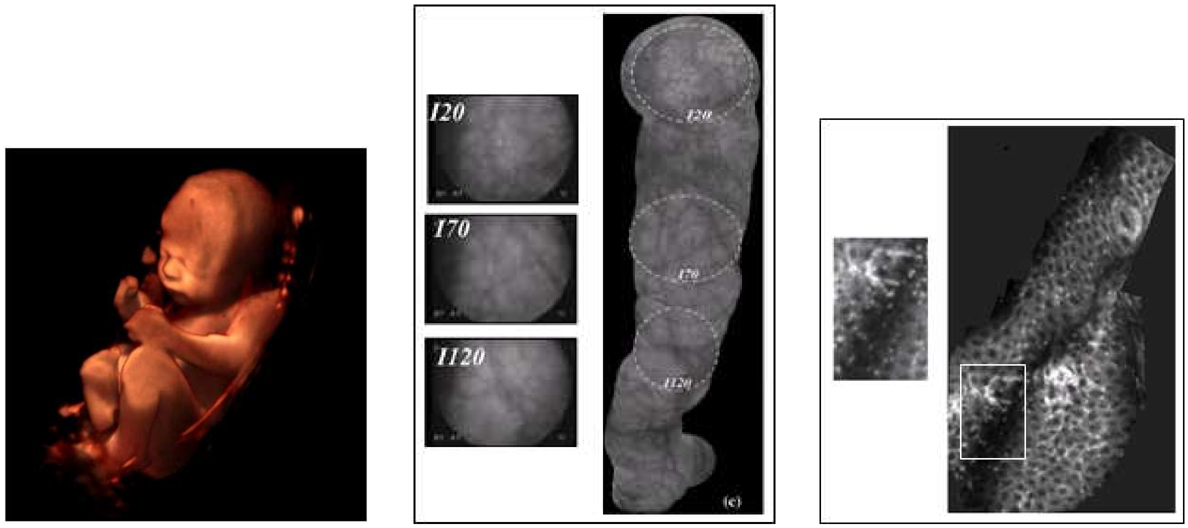
In endoscopic imaging, a camera is attached to a flexible tube and inserted into the organ to examine the interior cavity. It is performed on various organs like intestines, stomach, ear, etc. Stitching of endoscopic images is very useful in generating wide-field panoramas of internal anatomy. A mosaicing algorithm for bladder endoscopic images was proposed in [46]. The registration is based on maximization of mutual information. To make the optimization process more accurate and robust, mutual information is expressed analytically using a Parzen window. As endoscope motion path consists of loops, the alignment error is distributed along the loop in order to eliminate the artifact caused due to accumulation of error. In [7] a new method for creating and visualizing the mosaics of the bladder is presented. Images are matched using SIFT features and then homography is estimated using a modified RANSAC. In order to handle the geometric distortions caused due to planar projection, the bladder surface is modeled as a hemicube in which each of the faces contain the local panorama. Extending the previous approach, a new method with focus on global matching is proposed in [58]. A frame selection strategy is incorporated to reduce the amount of data. Instead of an exhaustive search for global matching, a three stage sparse matching strategy is employed. After global alignment, the surface is reconstructed using a thin plate spline and the final 3D mosaic is then obtained by texture mapping the image data onto the surface.

Real-time algorithms have also been proposed for mosaicing endoscopic videos [35, 9]. In [35], the motion field between consecutive frames is calculated using an optical flow algorithm. The affine parameters are then iteratively estimated based on the local motion field. In order to improve the accuracy of real-time mosaicing, a feature based algorithm is proposed in [9]. Feature tracking is carried out using Kanade-Lucas-Tomasi algorithm and the transformation is estimated by refining the matches using RANSAC. Using this method, significant improvement in accuracy and computational time over [35] is achieved.

2.4.3 Microscopy

In medicine, fibered confocal microscopes (FCM) are used for in vivo and in situ imaging of tissues which enables doctors to see inside the organism without actually damaging it. As the input frames are noisy, mosaicing helps us to recover the true information. Vercauteren et.al [62] proposed an algorithm to create a mosaic of the video sequence of FCM. A simple similarity based registration is applied to obtain rigid transformation between consecutive frames based on which the global transformation is then estimated. The global transformation is then refined by adding new pairwise registration results (non consecutive frames). This approach eliminates the need of exhaustive search for global optimum. The algorithm is extended in [61], in which distortions caused by the motion of the probe are modeled and then compensated to obtain sharper mosaics. Loewke [41] proposed a new approach combining two different methods. A global registration algorithm is presented handling the problem of accumulation of registration error and a local registration algorithm to accommodate non-rigid deformations. These two methods are integrated into a single framework which is suitable for solving general deformable mosaicing problem.

A real-time mosaicing algorithm was given in [40], where the registration is carried out by optical flow and then fine tuned by minimizing SSD using a gradient descent approach. The registration error is high in this approach as a trade-off for the real-time nature. A graph based approach addressing the problem of accumulation of registration error was proposed in [69]. Graph algorithms are used to obtain the anchor frame and the optimal path from each image to the anchor such that the overall registration error is minimized. Several approaches like minimum cost spanning tree, shortest path spanning tree, etc. are experimented and compared based on the quality of obtained mosaics and the computation time. Mosaicing algorithms have been designed in various other scenarios in medical imaging like X-ray [67], MRI [64] etc.



(a) (b) (c)

Figure 2.1 Mosaics generated by various algorithms (a) 3D Ultrasound of a baby[65], (b) Sample endoscopic images and the mosaicing result of a bladder[46], (c) Mosaicing of microscopic images of a cancer tissue[62]

2.5 Proposed Approach and Motivation

The need for mosaicing of neonatal retinal images has been discussed in 1.4. The sequence of images obtained from the RetCam are of poor quality with uneven illumination and variable contrast. Using these images, our aim is to automatically create a good quality mosaic of the complete neonatal retina. The images have to be spatially aligned to a global co-ordinate system and then information of the regions present in multiple views need to be combined. Pair-wise registration of RetCam is a challenging task as the spatial resolution and the quality of the images is very low compared to adult retinal images. Variable contrast and the blur makes it very difficult for accurate registration. Therefore, we pose it as a dense correspondence problem where a large number of features are obtained from both the images and matched. Since the pair-wise registration in this case is computationally heavy, finding a global alignment using simultaneous registration or graph based optimization is practically not possible. Therefore, a hierarchical approach is proposed which require fewer number of registrations. Finally, a blending schema is proposed for combining the information present in multiple images.

2.6 Summary

In this chapter, the technical challenges in a mosaicing system are introduced. The general steps involved in image mosaicing namely image alignment and image stitching with various methods for each step are discussed in detail. A brief literature of the mosaicing techniques used in medical mosaic is provided and the motivation for the proposed method is given. We present our hierarchical approach in the following chapter.

Chapter 3

A Hierarchical Approach for Image Mosaicing

3.1 Introduction

As discussed in Chapter 2, image mosaicing is an essential tool for many medical applications. For ROP diagnosis, a set of neonatal retinal images are captured using RetCam, which is a wide field fundus camera. Currently doctors need to manually inspect these images and based on the extent of vessel growth and abnormalities, decide upon the stage and zone of the disease respectively. This is a tedious process when the number of images is large and the experts need to reconstruct and visualize the complete surface from these images. Aiding the doctors calls for a solution for combining these images such a way that all the required information for the ROP diagnosis is present in a single image. Combining these images, requires their alignment to a single spatial co-ordinate system and then stitching. A simple method to align multiple images would be to sequentially register the images to the previous image or the initial image. Since the ordering of the images is unknown, the amount of overlap between successive frames may not be sufficient for registration. Therefore for accurate alignment, all the images need to be registered to each other and the set of mappings which achieve the global optima need to be chosen.

In this chapter, we investigate if mosaicing of neonatal retinal images can be achieved without exhaustive registration of all pairs of images. Due to the inherent limitation in the acquisition, degradation in quality is observed in some regions of an image (Discussed in Section 1.4). We assume that the required information (vessels and peripheral abnormalities) is available among the set of images. After aligning these images, the information is combined and the mosaic is created such a way that it contains the best available information for all the regions of the retina.

3.2 Proposed Method

The proposed mosaicing technique is a hierarchical approach for spatial alignment of multiple views of retina. Based on the background knowledge about the structure of a retina, a heuristic algorithm is designed for fast mosaicing. The overview of the algorithm with a sample set of images can be seen in Fig.3.1. The input images are first partitioned into subsets according to the region captured. An anchor image is chosen in each set which is of best quality using an image quality measure and all the images in the subset are registered to its anchor image. Images which have been accurately registered are combined using a blending technique to create intermediate mosaics. These intermediate mosaics are then registered and blended to obtain the final mosaic. Each of these are discussed in detailed in the following sections.

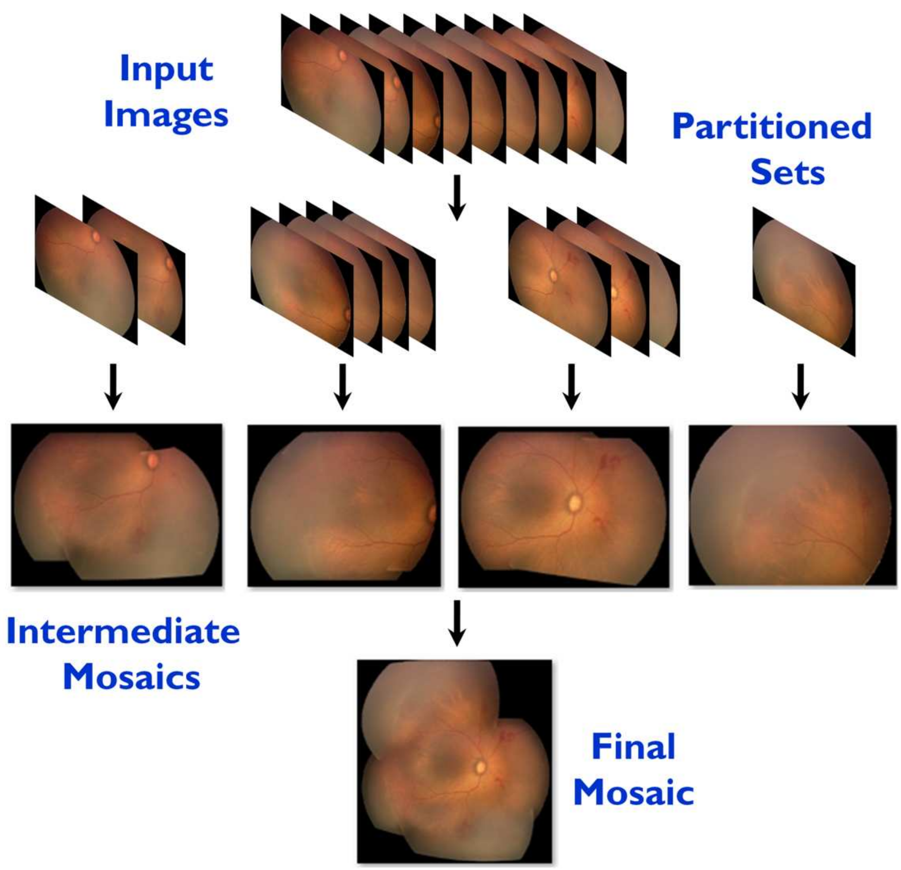


Figure 3.1 Proposed Hierarchical mosaicing approach

3.3 Partitioning

Given a set of retinal images, the aim of this module is to partition them into subsets. The criterion for partitioning is that all the images in a subset should represent nearly the same region of retina i.e. the view points from which they have been captured are close to each other. This criterion helps to organize the images and categorize them as knowledge about the motion of the probe is unavailable. Also, the operator revisits the same area multiple times till it is properly captured. Therefore, by this step we group together all the images which have maximum amount of overlapping information into one set.

This is necessary for accurate alignment of the images as the information content in each of these images is of low quality and estimation of transformation in such cases require a high number of matches. Maximizing the overlap among the images ensures this requirement. In many traditional applications, a continuous video stream is provided with very minimal spatial shift among consecutive frames. Due to high overlap, an accurate registration can be performed between consecutive frames. But in our scenario, since imaging is of premature infant, the stream is very sparsely sampled in time and the jump or the spatial shift among consecutive frames is very high. Therefore, the order of the images is generally random and hence partitioning serves to establish spatial relationship between these unordered set of images.

Formally, a set of images is denoted as $S = \{I_n; n = 1, 2..N\}$. This needs to be partitioned into $S_l; l = 1, 2.., L \leq N$; where S_l corresponds to a distinct region of the retina. This is achieved by using the structural information in the images, namely, the optic disk (OD) and blood vessels. OD is a bright elliptical region from which the blood vessels emerge radially outwards. Existing methods for detecting OD make use of the shape of the OD and the vessel information. We propose a new method in chapter 4 which can also be used for locating optic disk in the RetCam images. Once the OD is detected, images can be categorized as those containing OD and those that do not have OD. The latter are of peripheral region. Based on the relative position of OD from the image center, 5 partitions are derived ($L = 5$)

- S_1 is the set of OD-centric images (C)
- S_2 is the set with OD positioned in the left side (L)
- S_3 is the set with OD positioned in the right side (R)
- S_4 is the set with OD positioned in the top side (T)
- S_5 is the set with OD positioned in the bottom side (B)

For images with OD detected, its (x, y) coordinates are used to decide the sets. The x-coordinate is used to decide between center, left and right and y-coordinate is used to decide between center, top and bottom. Figure 3.2 describes the partition criteria based on the OD location in an image. Based on the region the detected OD falls in, the image is included in the corresponding subsets.

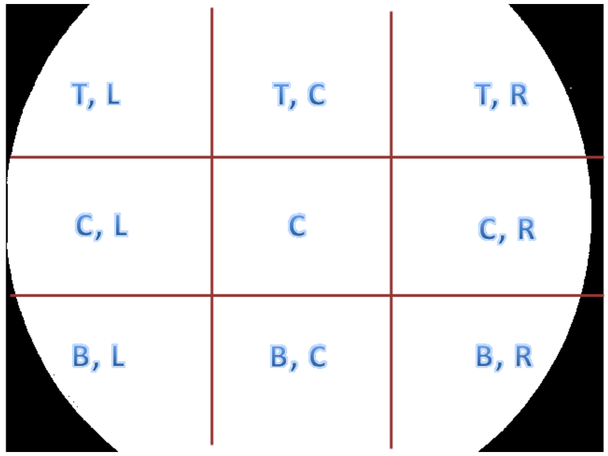


Figure 3.2 Membership based on OD

There are images of the periphery of the retina which do not have optic disk in their view. These images are classified based on the blood vessel information and domain knowledge about the anatomical structure of the retina. The retinal vessel network has an unique pattern as they radiate in a parabolic function from the OD in two main directions. Using the direction of the vessel and the location of the vessel, the image is placed in the appropriate subset.

The partitioning is a key step in our mosaicing framework. With partitioning, the need for exhaustive registration of all pair of images is eliminated. Using the knowledge about the scene of interest i.e the retina, we divide the images in such a way that registration needed to be performed only within the set. Since images of a set represent same region, the intra-set pairs have high probability that they are accurately registered. Also, it is highly unlikely that inter-set pairs accurately register. Therefore, for all images pairs $I \in S_x$ and $J \in S_y$, pair-wise registration is performed only if $x = y$.

3.4 Image quality Assessment

Once the input set is partitioned, images in each set S_l need to be registered. Therefore, as a frame of reference for each set, an anchor image is chosen based on image quality.

The definition of quality in images is task dependent. Something which can be termed as poor quality in one scenario may have the required information and termed as good quality in another scenario. Therefore, any quality assessment system defines what is the information of interest. In medicine, quality assessment systems are mainly used to judge whether an image obtained is good enough for expert diagnosis for a particular disease. In our system, we select the best quality image from a set which acts as an anchor image and the rest of images can be registered to the anchor image. This ensures accurate registration.

Since the primary structure of interest in ROP is the vessel network, an assessment based on the definition of the vessels is considered appropriate. The specific features of interest are the sharpness and the contrast of the vessels. Therefore, we convert the colour retinal images into gray scale images in which each pixel signifies the strength of the vessel at that location. The green channel of the color images is taken for processing as it contains the best information. Considering a 2D intensity image as a surface map (intensity corresponds to height at that location), blood vessels can be visualized as 'ridges' (see Fig3.3). Therefore, a multi-scale ridge strength measure is used for obtaining the vessel map. It is defined as follows.

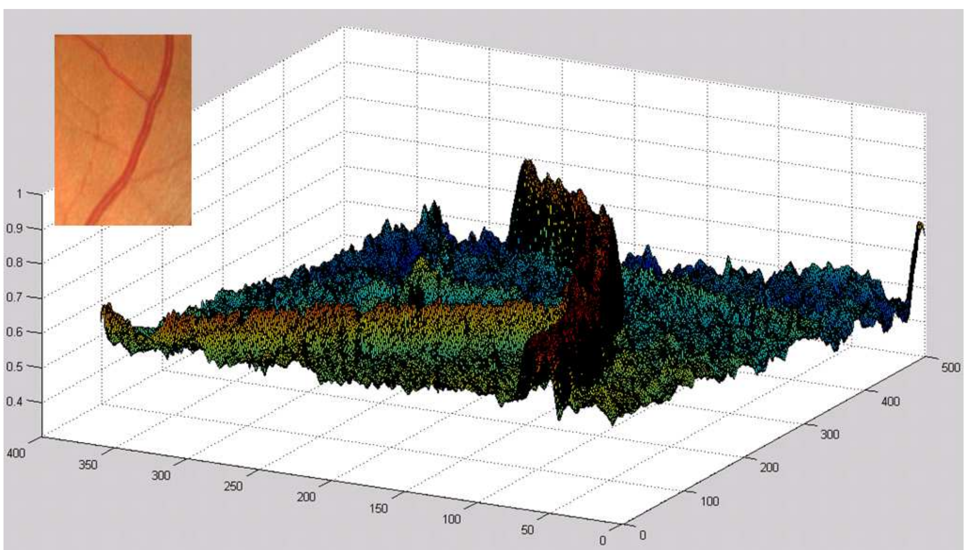


Figure 3.3 A 3D visualization of a retinal image as a height map. Inset: Corresponding retinal sub-image

Ridge Strength: Let $I(x, y)$ denote a 2D image. The ridge strength is defined based on the eigen values of Hessian matrix. The Hessian matrix H describes the second order partial derivatives of the image

$$H = \begin{bmatrix} L_{xx} & L_{xy} \\ L_{yx} & L_{yy} \end{bmatrix}$$

where L_{xx}, L_{yy} are the second derivative of $I(x, y)$ with respect to x and y and L_{xy}, L_{yx} are the mixed partial derivative.

The eigen vectors and eigen values of a Hessian matrix describe the nature of the geometry of the

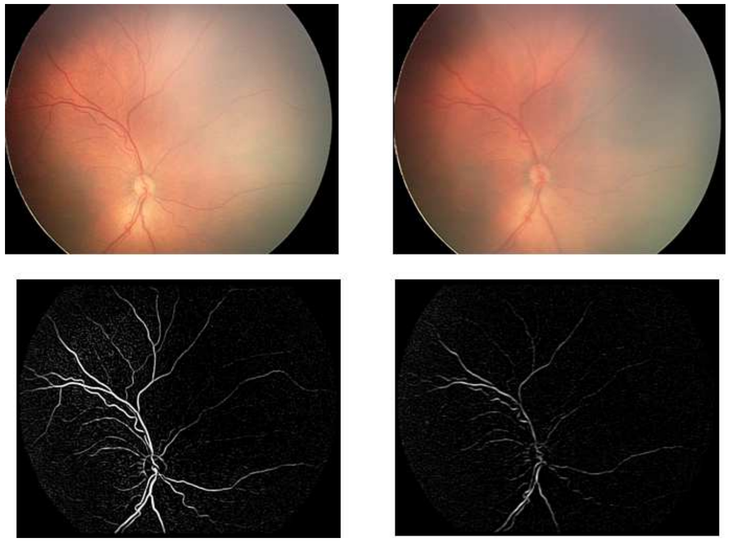


Figure 3.4 Two views of the same region of retina with the corresponding vesselmaps

curve at that point. The eigen vectors of H; u,v are called 'principal directions' and the eigen values L_{uu},L_{vv} are called 'principal curvatures'. The first eigen vector u corresponds to direction of greatest curvature (L_{uu}) and second eigen vector v corresponds to the direction of least curvature (L_{uu}). We use the multi-scale ridge strength measure proposed by Lindeberg [38]. It is calculated at various scales and the maximum across the scales is chosen as the final ridge strength. It is defined as

$$R_S(x, y, s) = (L_{uu}^2 - L_{vv}^2)^2$$
$$R(x, y) = argmax_s R_S(x, y, s)$$

where R_s is the ridge strength at scale s and R is the multi-scale ridge strength.

Using the ridge strength measure, we can compare the vessel quality of any two images using R. Figure 3.4 shows two colour images and their corresponding vessel maps. Visually it can be observed that the first image is of better quality with sharp and high contrast vessels. The same correlation is reflected in the vessel maps as the vessels in first image have higher ridge strength that second image. Therefore, a simple measure like the *mean* ridge strength of the complete image can be used for quality comparison. The quality measure Q of an image of size N is

$$Q = \sum_x \sum_y \frac{R(x, y)}{N}$$

In our problem, quality assessment is used to select the anchor image. Using this measure, an anchor image is selected in each of the sets which has best vessel information so that all the other images in the set are accurately registered to it. Also, by thresholding the measure, images of low quality are discarded from the set. Note that, such a measure can be used only because all these images belong to a single set and represent similar regions of the retina. Hence they should contain nearly the same vessels. Two arbitrary retinal views cannot be compared using ridge strengths as they correspond to different vessels.

3.5 Pair-wise Registration

Once the input set is partitioned and an anchor image is chosen in each partition, the task is to register all the images in each set to the corresponding anchor image. In this section, a pair-wise registration algorithm for retinal images is described which accurately registers even low quality RetCam images and is robust.

The goal of pair-wise image registration is to accurately estimate the underlying transformation between the images. As previously discussed (see Sec 2.2.1), this can be accomplished by direct methods or feature-based methods. In direct methods or area-based methods, the complete information of the images is used and a similarity measure is derived for comparing the images. This similarity is maximized using an optimization algorithm and the transformation parameters are obtained. Whereas in feature-based method, salient features are extracted and matched. Solving the correspondence problem using these matches, the parameters are obtained. In general, feature based methods have been used for retinal image registration in literature, as they are robust to changes in illumination and can handle partial overlap. Also in our problem, the changes in the non-vascular regions across the image will degrade the performance of direct methods. The pipeline of our registration algorithm is shown below.

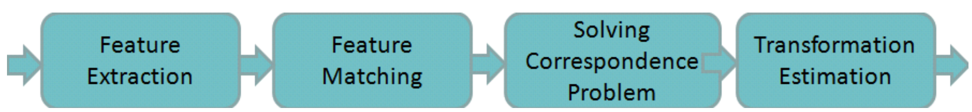


Figure 3.5 Keys steps in the registration algorithm

3.5.1 Feature Extraction

The key stages of feature extraction are the detection of interest points and then the description of regions around them. Feature Detection is first discussed. In many retinal image registration methods, branching points or vessel cross-over points are detected and used for registration. However, due to the low quality of information in RetCam images, these points are inadequate and result in fewer matches. Estimating the transformation using these matches leads to incorrect solution. Also, amount of overlap among the images may not be high even though partitioning is employed. Therefore, to overcome these issues, we solve registration by posing it as a dense correspondence problem. Large number of interest points which contain useful information are detected from the images.

A well known blob detection method called determinant of Hessian (DOH) is used for finding the interest points. DOH is computed at multiple scales and the scale-space maximum is calculated. Thresholding the DOH gives us the interest points. Since a dense set is needed, a low threshold is chosen. The obtained points are refined using curvature oriented histograms (COH) proposed in [52]. Using the entropy of the COH at a point, its saliency is determined. A subset of points is then chosen based on their saliency. Figure 3.6 shows a sample image with the interest points marked. It can be observed that many points are located near the vessel as they are of high curvature which get detected by DOH.

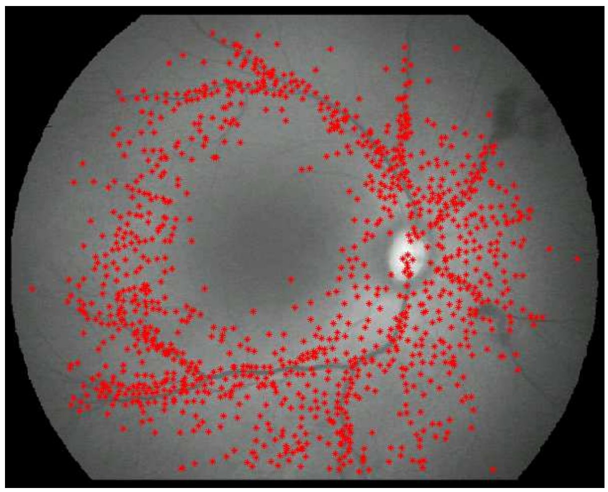


Figure 3.6 The feature points obtained using Determinant of Hessian

Once the interest points are obtained, the region around the points need to be described. Instead of computing the descriptor on the original images it is computed on the vessel map (described in Quality Assessment). As shown previously, images captured from RetCam of even the same region have different contrast with non-uniform illumination. Therefore, to eliminate the effect of these changes on the registration, the ridge strength is used to boost the vessels and generate a vessel map. Suppressing the background and using the vessel map makes the descriptor invariant to these changes.

Previously, many features based on DOH have been proposed. SURF [6], a popular feature also uses DOH for interest points and the sum of haar wavelets to describe the region. In our method, we use a Radon based descriptor previously proposed [5] for matching multi-modal retinal images which is shown to be accurate and robust for retinal image registration. Radon transform is widely used to describe the shape of the objects. It is a transformation based on projection. In this transform, 2D data is represented as a set of 1D projections. Each projection is obtained by performing a line integral at different offsets from the origin. By changing the angle of projection, we obtain the complete set. Formally it can be written as

$$R(r,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(x,y)\delta(xcos\theta + ysin\theta - r)\mathrm{d}x\mathrm{d}y \tag{3.1}$$

where r is offset and θ is the angle of projection and the image in (x, y) coordinate system is transformed into (r, θ) space. The descriptor is made rotationally invariant by using the histogram of principal direction in the region which is found from the eigen vectors of the Hessian. The peak of the histogram is calculated and the patch is rotated in that direction. Computing the Radon transform on the rotated patched and appending all the projections, results in the final descriptor. For our problem, large number of projections and offsets are chosen as more information is needed due to low quality of images.

3.5.2 Feature Matching

After obtaining feature descriptors for both the images, the next step is to match these descriptors and obtain the correspondence. A bilateral matching scheme [18] is used to find reliable matches. For all the features in both sets, the best match in the other set is computed based on Euclidean distance. In bilateral matching, only matches which are two way (i.e. vector v1 is the best match for v2 and v2 is the best match for v1) are considered. These matches are considered as the initial corresponding points.

3.5.3 Solving the Correspondence Problem

The matches obtained in the previous step have to be refined to produce an accurate set of points in each image which correspond to the same region of the scene. Using the knowledge of the transformation model, the initial matches can be refined. RANSAC [22] is a popular estimator used in most of the feature based approaches. It is used for model fitting and estimating the parameters of the model from the given data. It is mainly used when the data consists of both inliers (which follow the model) as well as a significant number of outliers (which do not follow the model) as a simple least square method would not produce a best estimate in such a case.

RANSAC is an iterative probabilistic algorithm where in each iteration n random points are selected as hypothetical inliers (Here n is the minimum number of corresponding points needed for estimation of the parameters). The model parameters are estimated using these hypothetical inliers and the rest of the data is tested against the fitted model. The estimated model is considered to be good if many points follow the model(with minimal error). This procedure is repeated a fixed number of times and the model which is followed by maximum number of points is chosen as the final model.

A family of algorithms based on RANSAC have been developed to make it fast, robust and accurate (MAPSAC, PROSAC, MLESAC etc). A detailed comparison and evaluation of these algorithms is described in [19]. For our problem a variant of RANSAC, called MSAC [60] is used for finding the model parameters. In the original algorithm, a model is evaluated based only on the number of points satisfying the model. In MSAC (M-estimator SAC), the loss function for evaluating a model is modified. The overall error for all the points satisfying the model is chosen as the loss function and the model with the least error is chosen as the best estimate.

As retina is a spherical structure, a quadratic transformation model is chosen for registration whose degree of freedom (DOF) is 12. Since directly fitting the data with a model with such high DOF results in inaccurate estimate, we first assume an affine transformation model whose DOF is 6. Therefore, only three corresponding points are needed for finding the parameters. In the MSAC algorithm, in each iteration, 3 matches are randomly chosen as inliers. Once the affine transformation is estimated, the matches are refined by discarding the correspondences which do not satisfy this transformation. The quadratic transformation parameters are then estimated using the refined matches.

3.5.4 Transformation Estimation

Once an accurate set of matches are obtained, they are spatially localized. A normalized cross-correlation is used as a similarity measure. A guided matching strategy is applied around a small window and the location which results in the highest similarity is chosen. Using these accurate matches, the quadratic transformation is estimated. A simple bicubic interpolation is used for applying the transformation. Registration results of the two views can be seen in fig 3.7. This type of checker board pattern visualization is a common practice in the literature [37] where consecutive patches belong to different images. This helps in visual inspection of the registered images by tracing the objects across the patches.

3.6 Image blending

In the previous section, a pair-wise registration algorithm has been presented for spatially aligning two images. Using this algorithm, images in each set are registered to the anchor image. In this section, a blending mechanism is proposed by adapting a method in [13] to combine the registered images in each set. The images in each partitioned set have large amount of overlap. Although they represent similar regions, the change in appearance among these images is very high. The vessel quality is different in these images and also the best available information for each region is shared among different images. Therefore, even after accurate registration, combining these images is a challenging task. We use a two step approach for combining the images where blending is employed in each step. In the first, blending is used to bring all the images to a common level with respect to the vessel quality and then, it is used to seamlessly combine the modified images.

The main idea behind the blending technique proposed in [13] is to combine the high frequencies over a small spatial extent and low frequencies over a large spatial extent. Assume that two images A and B need to be blended with R, a binary blending mask. The blending mask dictates which regions from each of the images have to be included in the composite image. Laplacian pyramids LA and LB are

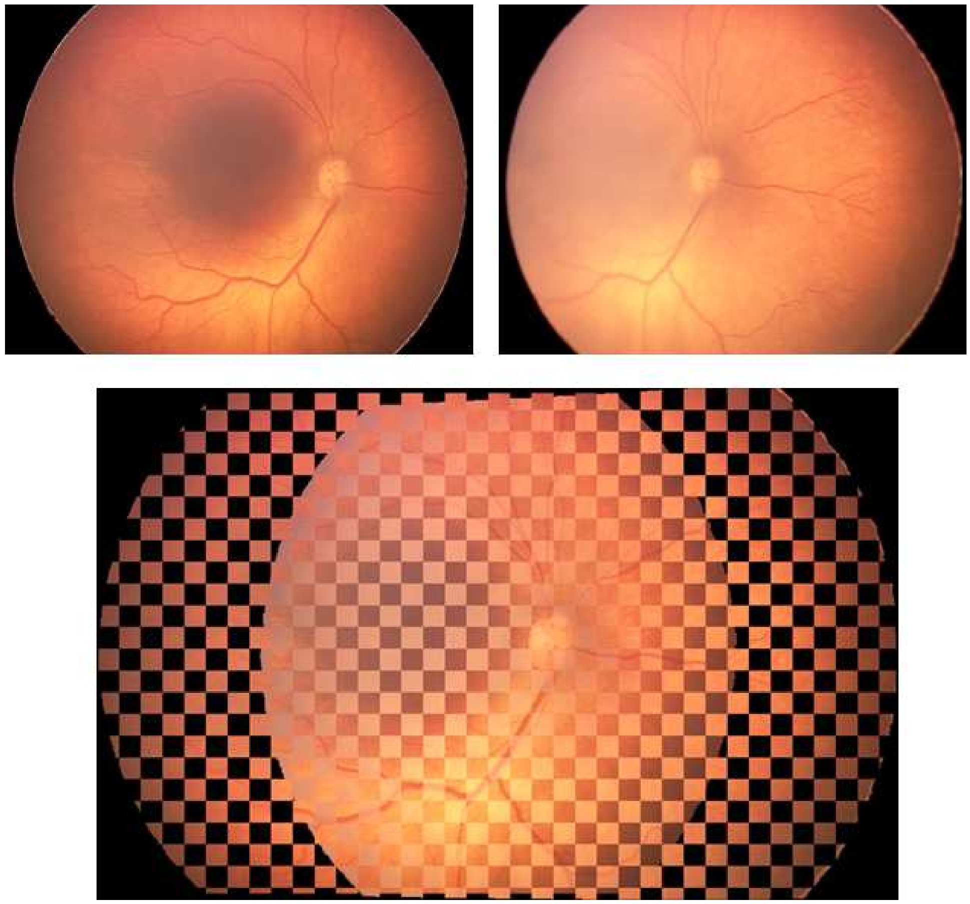


Figure 3.7 Registration result

constructed from images A and B and a Gaussian pyramid GR is constructed based on R.

Pyramids are multi-scale representation of images. In a Gaussian pyramid, a series of images are generated from the original image using successive low-pass filtering and down sampling. Similar to Gaussian, a Laplacian pyramid is constructed by applying successively Laplacian operator whose approximation can be obtained by taking the difference of Gaussians. Therefore, the difference of images at the successive levels of the Gaussian pyramid results in a Laplacian pyramid. The pyramids LA, LB are then combined using GR to create a new Laplacian pyramid LS.

$$LS[i,j] = GR[i,j] * LA[i,j] + (1 - GR[i,j]) * LB[i,j] \tag{3.2}$$

For using such a blending technique in our problem, the important step is to create the blending mask. In our two step approach, different blending masks are employed in each step to achieve different goals. Initially, blending two aligned images is discussed which can be extended to multiple images. As stated previously, in the first step we aim to equalise the vessel quality in all the images. Consider a pair of images A and B, where some regions of the scene are of good quality in image A whereas rest of the regions are clearly seen in B. Information from A is needed to be added to B and vice-verse. For this, a patch level comparison of the images is needed to find which image has more the information.

For find the blending masks for such a task, the quality assessment measure (mean Ridge strength) proposed in 3.4 is used. Previously it was used as a global image level measure, whereas in this case it is used for local comparison at patch level. This is possible since the images are accurately registered and the patches represent the exact same region. Figure 3.8 shows two patches of a same vessel from different view and their ridge strengths. It can be seen that the ridge strength can be used compare these patches.

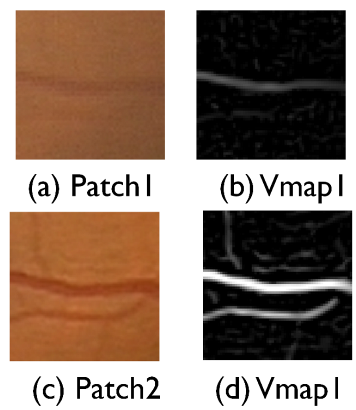


Figure 3.8 Vessel maps of two registered patches

Although a mean ridge strength is sufficient for comparison, it is possible that the patches do not contain vessels. These patches need to be excluded. As the ridge strength maps for these patches contain noise values, the comparison is invalid. For the patches containing the vessels, although the strength of the ridges is different in the images, the overall shape or structure is exactly same. Using this observation, we compare the projection of the patches to decide whether they represent two vessel regions or not. Using such a local quality assessment, corresponding patches in two image are compared and a binary blending mask is created.

Extending this to multiple images, for blending a set of images A,B,C, information from A and B is added to C, from B and C, to A etc. After this process, the vessel information in all the images is of nearly identical quality. In the second step, blending is again used to combine these modified images in a seamless composite image. For this purpose, a simple blending mask is constructed by overlaying the registered images with averaging at the overlapping region. Finally, all the images in each set are combined using this approach to create an intermediate mosaic.

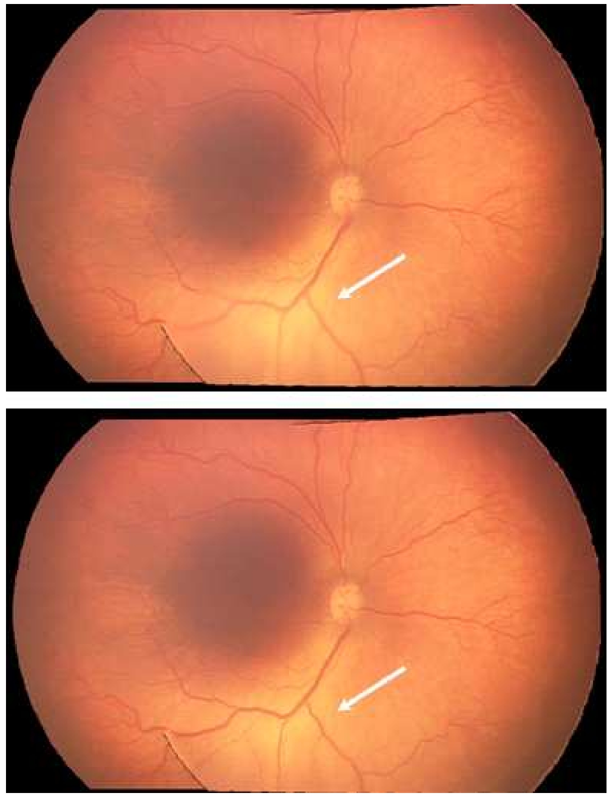


Figure 3.9 Blending original images and modified images (after adding the the information)

3.7 Image mosaicing

Recalling the complete flow of the algorithm, the input image sequence is partitioned in to sets based on the region they represent in order to obtain high amount of overlap among them. Using a global image quality metric, the image with the highest quality is chosen as an anchor image from each of the set. The rest of images in the set are then compared with the anchor image and based on a threshold low quality images are removed from the set. The remaining images in the set are registered to the anchor image. Once the pair-wise registration is performed, images which are not accurately registered are discarded for subsequent processing. Based on reprojection error and the number of accurate matches, these image pairs are detected. Eliminating these images helps in avoiding blurring or ghosting artifacts in the mosaics. The images which are accurately registered are then combined using a blending mechanism to create intermediate mosaics.

At this stage, the problem of registering arbitrary number of unordered images is converted into registration of these intermediate mosaics which have more spatial information and also have better vessel quality than a single image. In general, the images of the central region of the retina (with the optic disk at the center), tend to be of good quality, therefore the intermediate mosaic created from S_1 is chosen as the anchor image and the rest of the mosaics are registered using the same pair-wise registration algorithm. Finally these images are blended to create a complete good quality mosaic of the retina. The algorithm is summarized below

Algorithm for image mosaicing:

1. Partition the given set $S : [S_k]$ based on OD location
2. Based on image quality, choose an anchor I_{Ak} and remove low quality images for each S_k
3. Register the images $I_j \in S_k$ to the anchor I_{Ak}
4. Discard the images which do not register accurately
5. Blend the registered images in S_k to create an intermediate mosaic
6. Register and blend the intermediate mosaics to create a complete mosaic using steps similar to 3 and 5

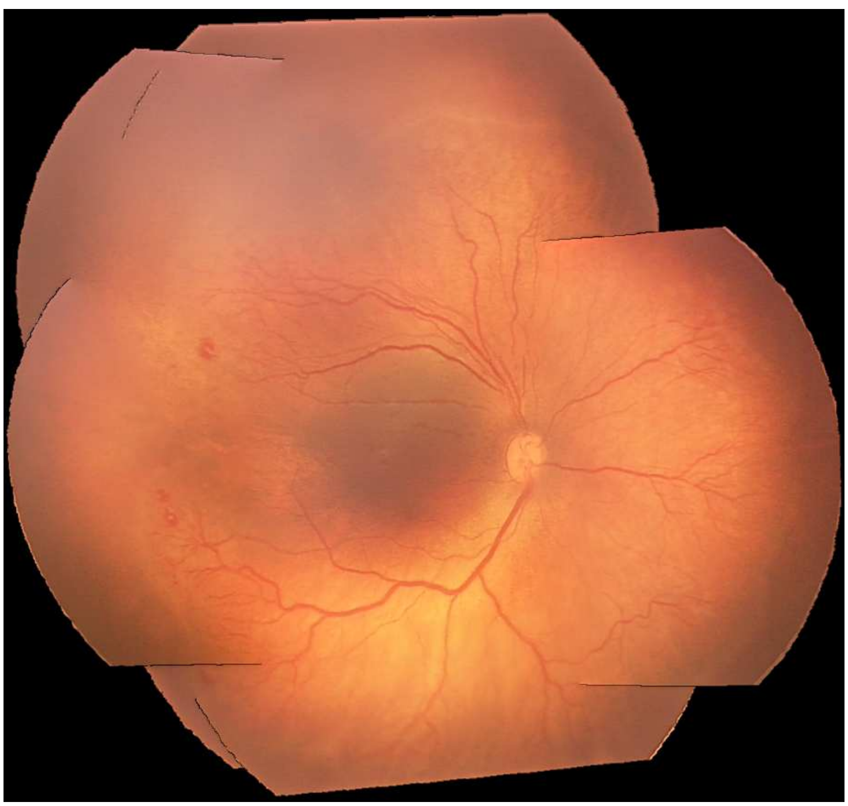


Figure 3.10 A generated mosaic

3.8 Summary

In this chapter, we have proposed a framework for mosaicing of retinal images. For global alignment, the structural information is used to design a hierarchical approach instead of an exhaustive method. Only the image pairs which are highly probable to register among themselves, are considered for creation of the mosaic. Also a smart blending mechanism is employed based on the local vessel quality such that the best quality information is retained in the mosaic. The mosaic created using such an approach has wide field of view containing even the periphery of the retina.

Chapter 4

Optic Disk Detection

4.1 Introduction

Optic disk is the location from which optic nerve exits retina. The visual information from retina is transmitted to brain using the optic nerve. It is also an entry point for the blood vessels which supply blood to retina. The central retinal artery and vein emerge from it and cover the entire retina by branching further. In fundus images, optic disk can be seen as a elliptical shaped bright yellowish region but in general, significant variations in size, color and shape can be observed.

Optic disk detection and segmentation is an important step for many retinal image analysis systems. It serves as a major landmark for detecting various other anatomical structures in retina. To segment blood vessels, vessel tracking methods require initial seed points and the vessels points within or near OD are used. As the distance between OD and macula is constant, OD position is used for estimating the location of macula. Masking the OD also leads to decrease in false positives in detection of bright legions. Segmentation of OD is required for automatic diagnosis for diseases which affect the OD region. Diseases like glaucoma are characterized by the change in shape of OD and quantitative analysis of its diameter is helpful for diagnosis.

Automatic localization of OD is a challenging problem due to the variations in its appearance and various methods have been proposed for detection of OD. Model based methods analyze the structure of the blood vessels and localize the OD as the point where the vessels converge. These methods achieve high accuracy but are computationally very expensive. In [23], geometric model-based method is presented which achieves 97.5% but with a computation time of 2 minutes. A directional matched filter based approach in [3] reports an accuracy of 98.8% with average time of 3.5 minutes. In order to reduce the computation time, a fast technique for OD detection is proposed in [43] which reduces the dimensionality of the search space by converting a 2D localization problem to two 1D problems. Accuracy of 92.6% is obtained with average time reduced to 0.46 seconds.

In our mosaicing framework, the set of RetCam images are partitioned into subsets using optic disk location. As previously discussed, the quality of retinal images obtained using RetCam is degraded which makes the detection of optic disk a difficult task. [23, 3] proposed sophisticated methods with high accuracy, but are computationally inefficient. Therefore, an accurate and efficient method is proposed for detection which makes use of the color difference at the boundary of OD. The method is tested on multiple public datasets and also RetCam images.

4.2 Color Distance based detection

The fundus images obtained by RetCam are blurred with variable contrast. However, even in a low quality image, OD can be treated as a homogeneous region of uniform color which has color discontinuity at the boundary of the region. Irrespective of the color or shape of OD this property holds true. Based on this observation, a measure is derived which captures this information and is maximum at OD location. Unlike most of the approaches proposed in literature which use either red or green channels, all the color channels are used in this technique which makes it robust and invariant to the color of OD.

As assumed earlier, OD consists of a region with uniform color, therefore the color distance (defined as the Euclidean distance in the RGB space) among the points in OD would be very low (or ideally 0). Also since there is a color change at the boundary of OD, the color distance between points inside the OD and the boundary points would be very high. Since these conditions hold true only at OD, a measure is derived by combining the two conditions. To optimize the measure, the color distance is both cases is computed only from the brightest point inside OD rather than all the points. Both the conditions hold true even in this case. The measure is formally described below.

First, an annulus-shaped template is designed whose radii r_i & r_o are calculated based on the resolution and DOF of the camera such that the OD boundary lies in the annulus i.e. the radius of optic disk lies between r_i & r_o. Let R_i and R_o be the inner and outer regions of the annulus (Fig: 4.1). At every point p in the image, the template is centered at p and the local maximum is determined within R_i. This is computed in RGB space and denoted as C_p. Next, the color variation relative to C_p, within the template is computed as follows:

$$O(p) = \sum_{r \in R_o} \|I(r) - C_p\| \tag{4.1}$$

$$I(p) = \sum_{r \in R_i} \|I(r) - C_p\| \tag{4.2}$$

where r is the position vector and $\|\|$ is the Euclidean distance in RGB space. For any p inside the OD, $O(p)$ would be very high and $I(p)$ would be very low. Therefore the OD measure $H(p)$ is defined as

$$H(p) = O(p) - I(p) \tag{4.3}$$

Figure 4.1 Sample OD region and OD template

This function would be maximum when p is at the OD center. Also, it would be high for all the points inside optic disk and choosing any of these points is sufficient. Therefore, this is made computationally efficient by calculating $H(p)$ at regular intervals rather than each and every point. One of the points inside OD which has high measure is detected in the sub-sampled version. Fig 4.2 shows a sample image captured by RetCam and its corresponding measure H. It can be observed that even in case of blurred and low contrast image, the measure is able to capture the required information. Also in such cases, model based approaches which rely on vessel information cannot be used.

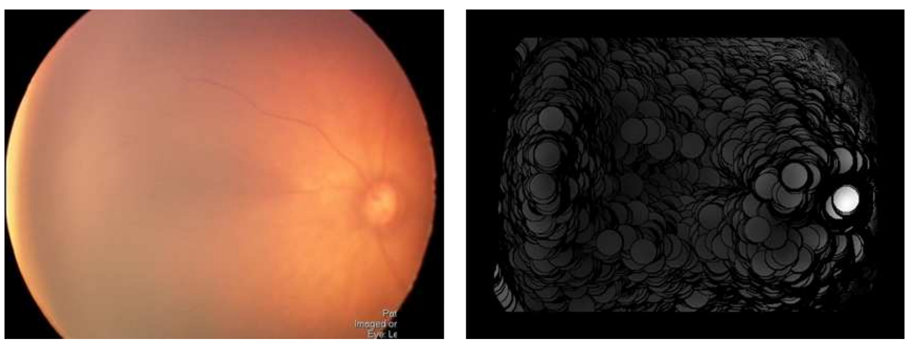


Figure 4.2 Retinal image and its OD measure

4.3 Evaluation and Results

For evaluating the performance and computation time of the algorithm three publicly available datasets are used. The focus of this thesis is on images captured using RetCam, therefore two others datasets ROP1 and ROP2 are created for testing the system on low quality images. The images are divided into two sets separating the challenging cases for analyzing the performance in detail. Table 4.1 list the information of all the databases used. The proposed method was implemented in MATLAB and the results show below are obtained by running on a PC with 2.4GHz Intel Core 2 Duo and 2 GB RAM. As an accurate center of the OD is not required, the detected location is considered correct if it falls anywhere on the OD. Based on the resolution of the datasets, r_i & r_o are obtained and for DIARETDB

Database	No.of Images	Resolution
DRIVE	40	565 X 584
DIARETDB0	130	1500 X 1152
DIARETDB1	89	1500 X 1152
ROP1 (easy)	75	640 X 480
ROP2 (difficult)	50	640 X 480

Table 4.1 Datasets for OD detection

they are chosen to be 80 & 120 and 40 & 60 for the remaining datasets.

Table 4.2 summarizes the results on the three public datasets and two ROP datasets. A success rate of 96.6% is achieved with OD being detected in 371 images out of 384 images. With average computation time for high resolution images (1500 X 1152) as nearly 1sec. Fig 4.3 and 4.4 shows examples of images with detected OD on DIARETDB and ROP images respectively. The last row in both figures show examples of failure cases. The algorithm fails to detect OD in cases when either the contrast of OD is low or when bright lesions are predominant. Preprocessing the image for image enhancement may improve the performance in the cases when contrast of OD is low.

Dataset	Drive	Diaretdb0	Diaretdb1	ROP1	ROP2
Number of Images	40	130	89	75	50
OD detected	40	125	87	74	45
Success Rate (%)	100	96.2	97.8	98.7	90
Computation time (secs)	0.30	1.05	1.05	0.32	0.32

Table 4.2 OD detection performance

4.4 Summary

Localization of OD is an important step in any retinal image analysis system. This chapter presented a detection method which can be used for localizing OD even in blurred and low contrast images. The algorithm was shown to detect with good accuracy on multiple public datasets and on RECTCAM images. This method is used in our proposed mosaicing framework in the partitioning stage where the input set is partitioned using the OD location.

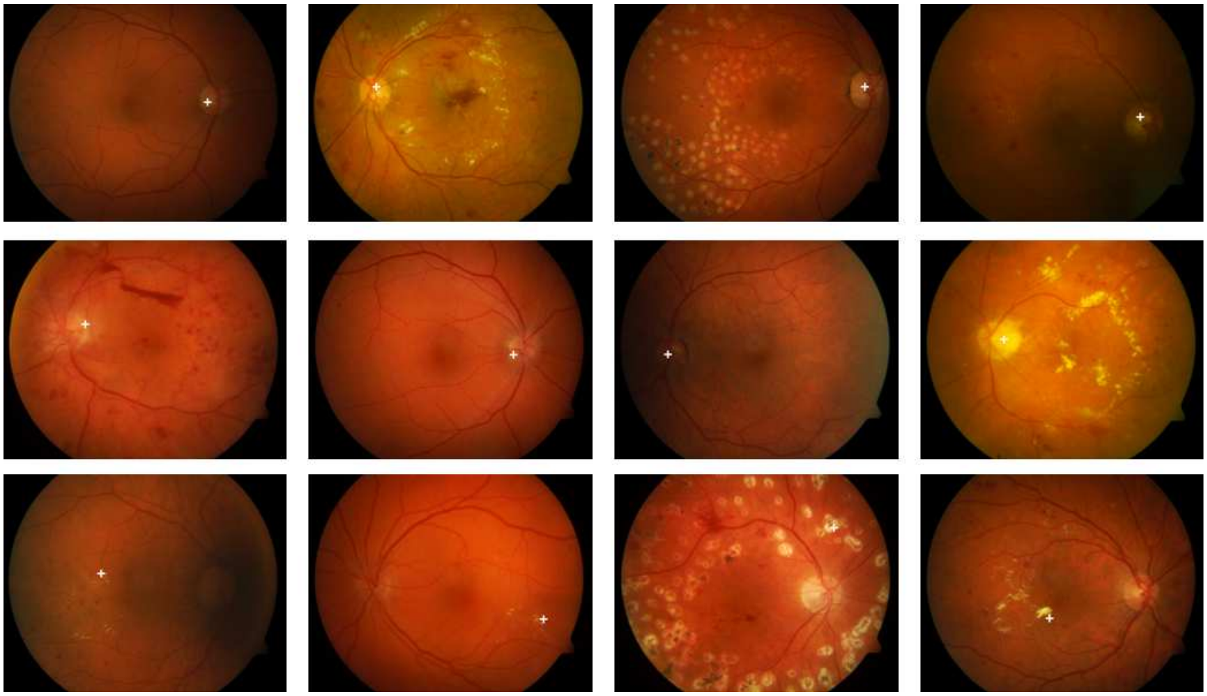


Figure 4.3 Sample results on Diaretdb1,2. Last row shows failure cases

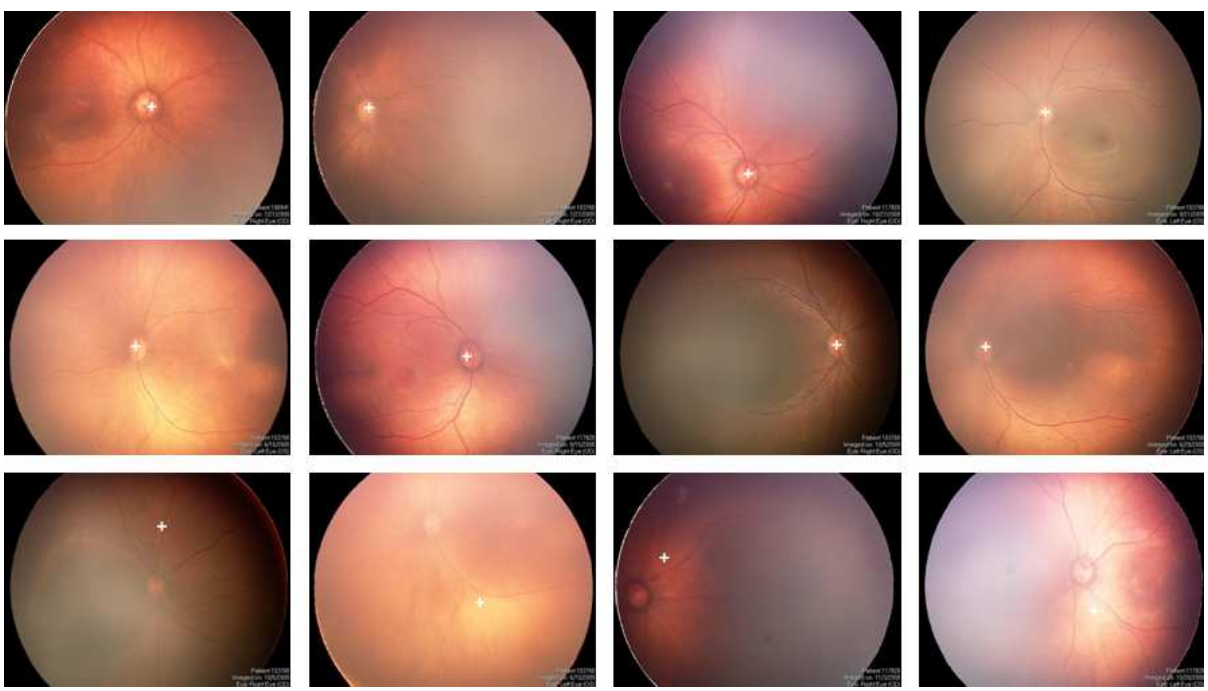


Figure 4.4 Sample results on ROP1,2. Last row shows failure cases

Chapter 5

Experimental Evaluation

5.1 Dataset

The proposed system is evaluated on a custom-built dataset consisting of neonatal retinal images of babies with ROP. These images captured using RetCam, were collected for clinical purpose in a local hospital. Since each retina is captured using multiple views, a set of images are present for each eye. The dataset consists of 125 images of 10 premature babies with the number of images in each set varying from 10 to 19. The images were captured by trained operators and all the subjects had been diagnosed by experts as belonging to different stages (1 to 4) of ROP. The images in the dataset are of compressed JPEG format with pixel resolution of 640x480. Although the images are captured after dilating the eyes, non uniform illumination is observed due to sub optimal pupil size. This results in a central shadow region in many images. A sample set of images from the dataset is shown in Fig 5.1.

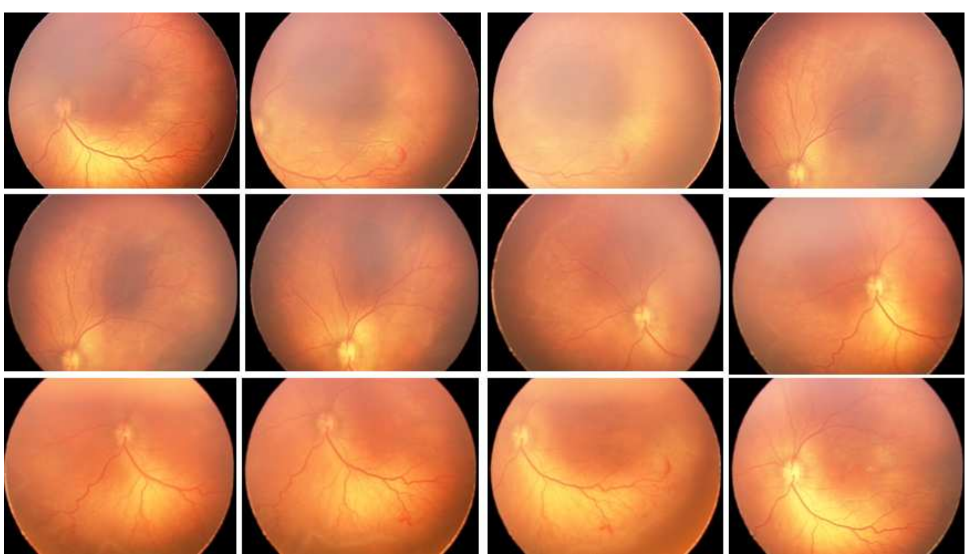


Figure 5.1 Images of a neonatal retina captured from different viewpoints

5.2 Evaluation of the components

Before presenting the results of the complete system, the individual components of the system are evaluated and results are shown for the three major modules; pair-wise registration, global alignment and image blending. Once these components are tested, both quantitative and qualitative results on the generated mosaics are presented.

5.2.1 Pair-wise registration

The pair-wise registration algorithm was tested on two sets of ROP images - REG1 and REG2. REG1 has 40 pairs of images with more than 40% of overlap and with significant vessel information and REG2 has 25 challenging pairs of images with low overlap. The method is tested on these 65 pairs of RetCam images and compared against GDBICP[70], a standard retinal image registration algorithm. The code for GDBICP was made available by the authors. It should be noted that GDBICP was not developed for RetCam type low quality images. In our methods, in order to accurately register the images, the features in the images are very densely taken by choosing a low threshold for DOH. Feature points as high as 3000-4000 are selected and the Radon descriptor is computed around a 40x40 patch. For GDBICP, the parameters for the executable have been chosen as per the guidelines given. The proposed method takes 50–60secs while GDBICP takes 20–30secs (even with -complete option which is significantly slower). The difference in the times is due to exhaustive feature extraction and matching.

As the quality of the images is low the performance of the methods are compared using visual inspection. Each result is classified into one of the three groups; accurate, satisfactory and failure. Table 5.1 shows the performance of proposed method and GDBICP on both the datasets. It can be seen that the proposed method performs well on both the sets. Even for challenging pairs, our method is able to register most of the pairs up to adequate alignment. The performance of GDBICP depends on the quality of images. This can be seen in the high number of failures. Sample cases from the datasets are show in Fig 5.2, 5.3 in which GDBICP completely failed to estimate the transformation whereas our approach accurately registered these views.

Method	Dataset	Accurate	Adequate	Failure
Our method	REG1	22	14	4
GDBICP	REG1	16	13	11
Our method	REG2	11	10	4
GDBICP	REG2	6	9	10

Table 5.1 Pair-wise Registration Performance

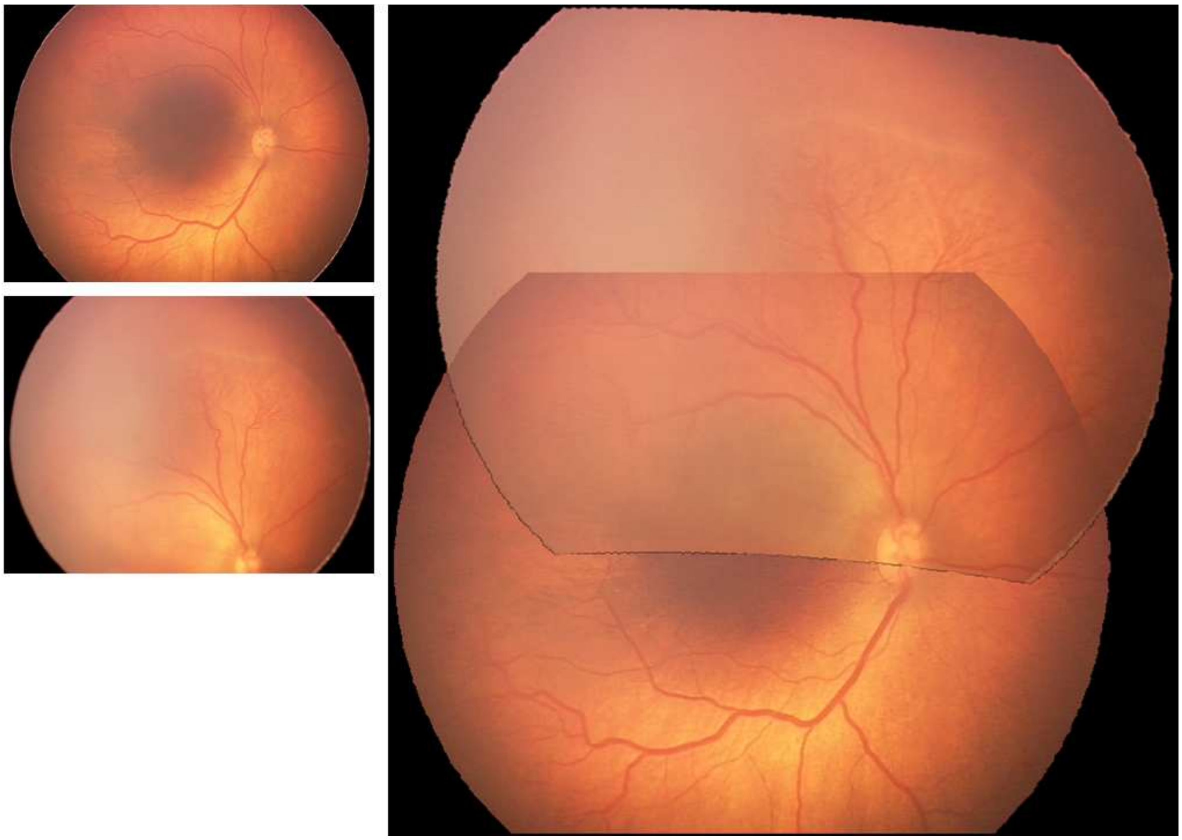


Figure 5.2 Registration result using the proposed method for images which failed to register using GDBICP

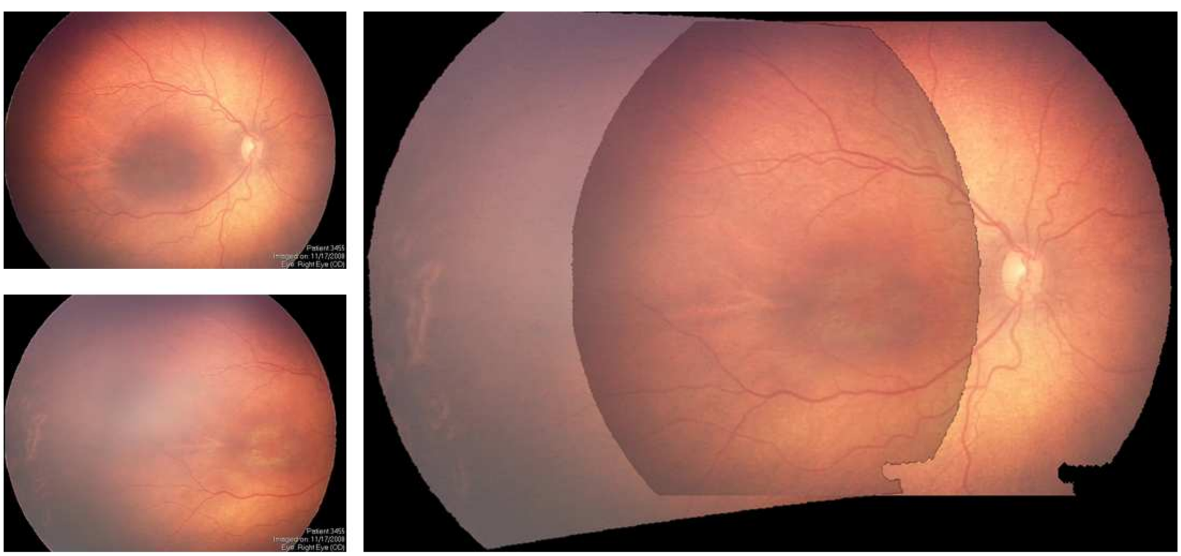


Figure 5.3 Registration result using the proposed method for images which failed to register using GDBICP

5.2.2 Global alignment

Globally aligning all the RetCam images using a sequential approach is not possible as the motion of the camera is discontinuous and the alignment between the neighbors cannot be guaranteed. Therefore, we compare the method with simultaneous global methods which are based on graph algorithms, to obtain optimum global transformation. Fig 5.4 shows a connectivity graph on a sample set of 12 images. Using our pair-wise registration algorithm (any registration algorithm can be used), the graph is constructed where the pairs which cannot be registered are discarded. This is done by thresholding the reprojection error and the number of final matches obtained. The partitions obtained by our mosaic approach are shown using the ellipses where node 1 is chosen as an anchor image for set red and green and node 4 for the third set. It can be seen that the obtained partitions are well connected among themselves and without explicitly registering all pairs of images the inherent connectivity of the structures are obtained. As the images which have good quality vessel information are chosen as anchors, high connectivity to these nodes is reflected in the graph.

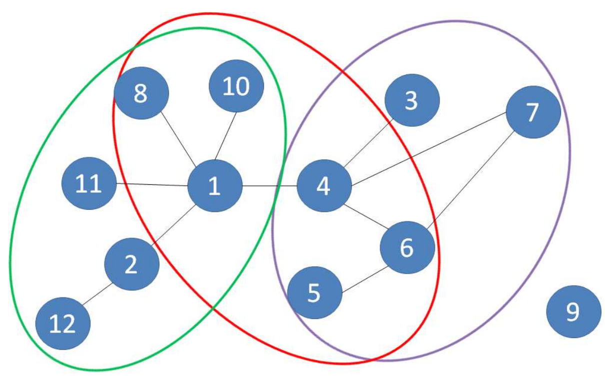


Figure 5.4 Connectivity graph of different views of retina

Similar correlation is observed even for other sets when compared to the graph based methods. The obtained partitions are generally found to have higher number of intra-set edges denoting that they can be accurately registered whereas fewer number of inter-set edges denoting that they are unlikely to register. Therefore, based on partitioning we select a subset of pairs which are highly probable to register and perform registration on only those pair. For a graph based approach, all possible pairs of images have to be registered and hence requires $O(n^2)$ registrations for n images where as in our approach only $O(n)$ registration are sufficient. The number of mosaic-mosaic registration between the intermediate mosaics is constant (number of partitions which is chosen to be 5). Fig 5.5 shows a plot of number of pair-wise registrations required against number of images in the set for both graph based methods and proposed approach.

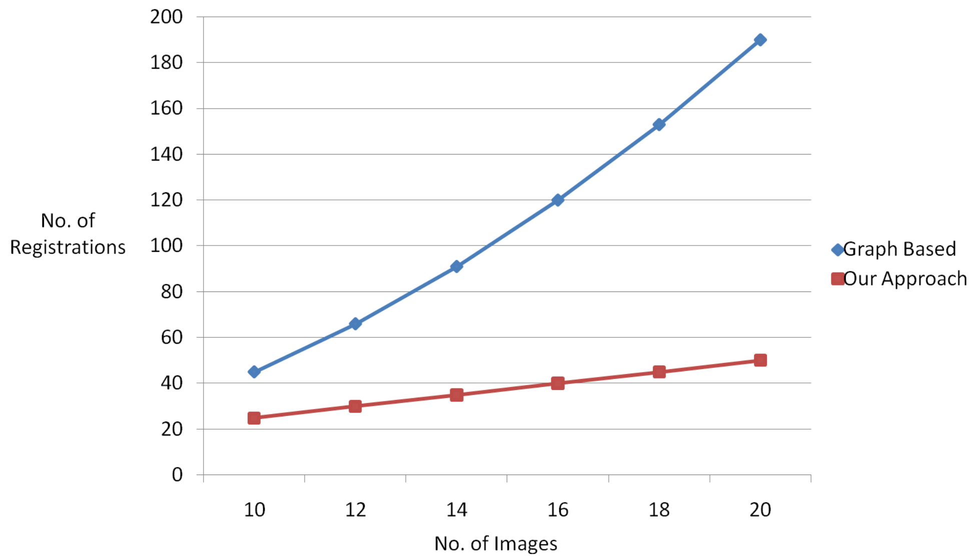


Figure 5.5 Comparison of number of pair-wise registrations

5.3 Image Blending

Image Blending is employed in order to combine the registered images and produce seamless mosaics. Blending proposed in our method is compared with simple averaging and center weighted alpha blending also known as feathering. In simple averaging, the images are combined by choosing the mean intensity value in the overlapping regions. In feathering, the blending masks are constructed based on the distance of a pixel to its nearest boundary. The weights are calculated using a distance transform and combined as shown below.

$$I_{blend} = \alpha * I_1 + (1 - \alpha) * I_2 \tag{5.1}$$

$$\alpha = \frac{dis_1}{dis_1 + dis_2} \tag{5.2}$$

where I_1, I_2 are the two registered images and dis_1, dis_2 their distance transforms. The blending results of these methods is shown in Fig 5.6. The input images are shown in the first column (a) and the output of simple averaging, center weighted alpha blending and proposed method are shown in (b), (c) and (d) respectively. It can be observed that center weighted alpha blending performs much better than simple averaging but subtle transitions are still visible whereas our proposed method combines the images without any visible seam even in case of change in illumination and contrast across the images.

The aim of the proposed blending method is to combine the images seamlessly and also preserve best vessel information. Fig 5.7 shows the blending results using original multi-band blending and the modified approach. In our method, the vessel quality information is used to create the blending mask such that the best available information is present. The improvement in the vessel information over the original method can be observed in the zoomed section.

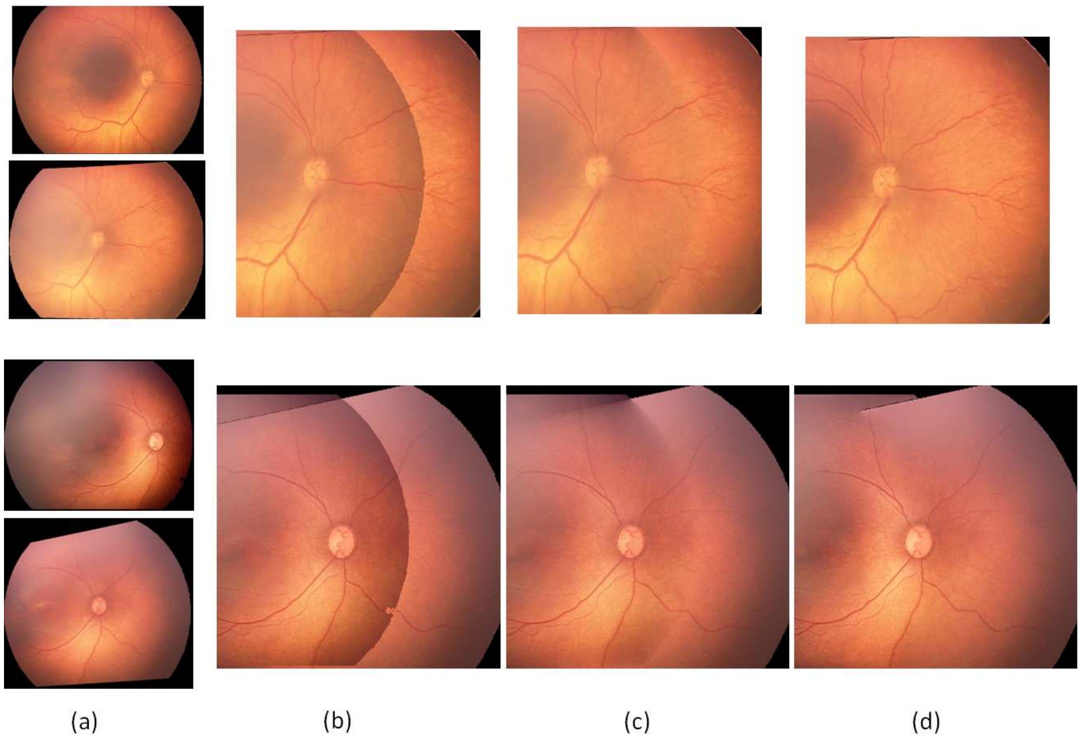
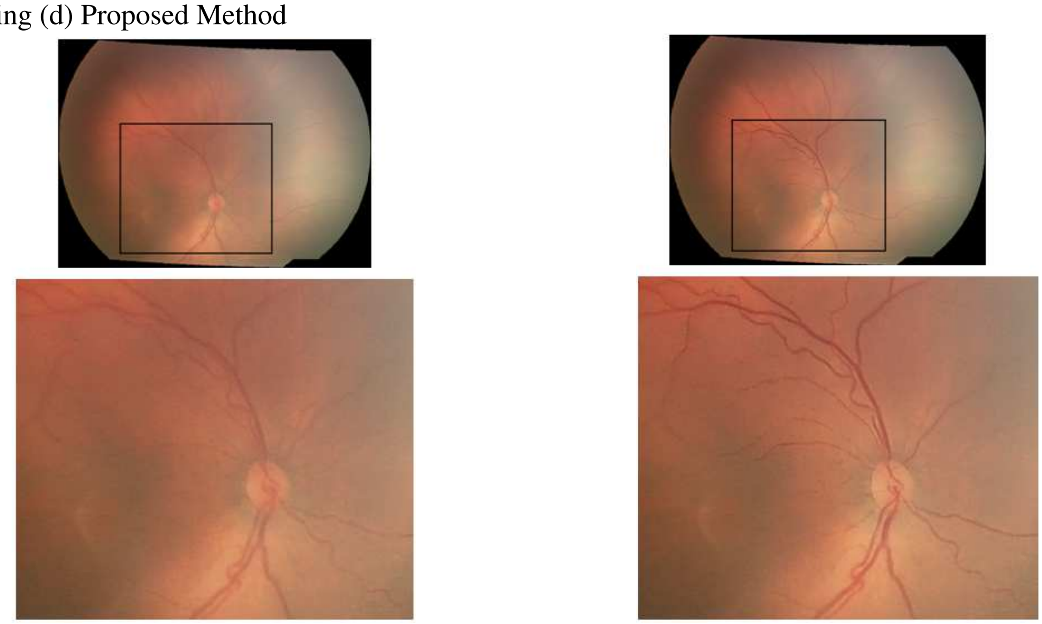


Figure 5.6 Blending results using different techniques (a) Input Image (b) Simple averaging (c) Alpha Blending (d) Proposed Method



(a) Using original multi-band method

(b) Using proposed modification

Figure 5.7 Improvement in blending using vessel quality information. Top row: Sample mosaics. Bottom row: Zoomed view of the subimage within the black box

5.4 Quantitative Results

The generated mosaics were evaluated based on FOV. The main motivation behind the thesis is to increase the field of the view so that the extent of the visible scene is increased. The abnormalities present at the periphery of the retina can be seen in the mosaic. Therefore, we measure the distance from the OD to periphery which can be seen in the mosaic. The increase in the visible area is reported for four different directions: nasal, temporal, superior and inferior (In medical domain, the directions are denoted with reference to anatomy where nasal/temporal denotes towards the nose/temple and superior/inferior denotes towards the head/feet). Also as per expert guidelines, the increase is reported in units of OD diameter.

Fig 5.8 shows a mosaic generated with the base image marked. From the set of views, best quality image which captures the central region is chosen as the base image. The % increase in the each of the directions in terms of OD diameter is reported. For instance, there are 3 white circles in the nasal direction in the base image. This increases to a further 3 after mosaicing heading to a 100% increase in FOV. Similarly, in the temporal direction, 7 units in the base image increases to 13. For the complete dataset, the average increase in the visible area across the four directions is found to be 95%.

	Nasal	Temporal	Superior	Inferior
Base Image	3	7	4	3.5
Mosaic	3	13	8	6
% Increase	100	85	100	70

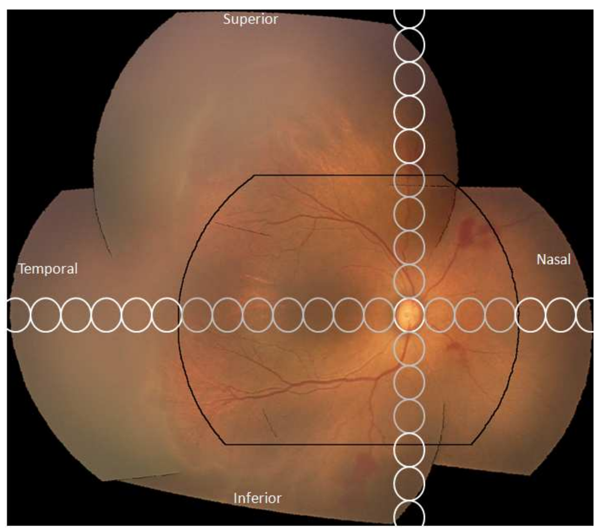


Figure 5.8 Increase in visible area after mosaicing 7 images. Base image is shown with black border

We also measure the increase in the visible scene in terms of pixel resolution. Original images of retina have a resolution of 640×480 in which the scene occupies nearly 0.265 megapixels. This is compared with the number of pixels of the retina in the generated mosaics to obtain the increase in the field of view. Fig 5.9 shows the % increase in number of pixels representing the retina for each case in the dataset. Apart from the increase in the visible area, the increase in the vessel extent in the mosaic is also of interest. An existing vessel centerline extraction algorithm [57] is used to compared the vessels in the base image to the generated mosaics. The centerlines were extracted in both the cases using the same set of parameters. Fig 5.10 shows % increase in the vessels detected for mosaics generated for all the cases. It can be clearly seen from both the graphs that there is significant increase in the visible region of the retina and also more vessel information is obtained.

The generated mosaics have been validated by the experts to be sufficient for diagnosis of ROP. Therefore, instead of examining all the views of the retina, the experts can now diagnose by examining the complete organ in a single image. On a average, each retina is captured using 15 images. If the time spend on each image is considered to be t secs then the time taken to examine a mosaic is nearly 2t secs as the number of pixels are nearly doubled. Therefore, there is a decrease of 7 folds in the time taken for diagnosis. This is a huge reduction and can be of great use in case of screening and telemedicine.

In many scenarios, using information technology, better health care is provided to rural communities. The data collected at various remote locations is either stored and forwarded or remotely monitored by an expert at a distant location for diagnosis. Instead of storing and transmitting a set of images, a single mosaic can be sent to the expert. The size of each individual image is 400KB and that of the mosaic is nearly 750KB. Therefore, apart from the reducing in the time taken for diagnosis, there is a 8 fold decrease in the amount of data to be transmitted which is a very useful due to the limitation of bandwidth in developing countries like India. Also for storing and archiving, the amount of storage can be reduced by using mosaics.

5.5 Qualitative Results

Some of the generated mosaics are shows for visual inspection in Fig: 5.11, 5.12. It can observed that the complete periphery is visible in the mosaics. Fibrosis and while demarcation lines can be observed and their location with respect to the central retina can also be seen giving the complete view of the organ.

5.6 Summary

In this chapter, we have evaluated the proposed mosaicing approach. The major three modules: Pairwise registration, global alignment and image blending have been individually tested and compared with other methods. Quantitative results have been presented for the mosaics showing the increase in the FOV. Also, qualitative results of mosaics showing the complete retina surface including the periphery have been shown.

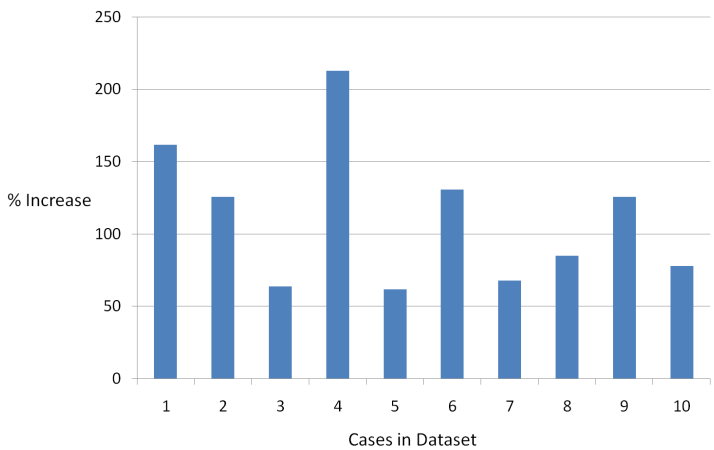


Figure 5.9 Graph representing the % increase in the number of pixels for each case in the dataset

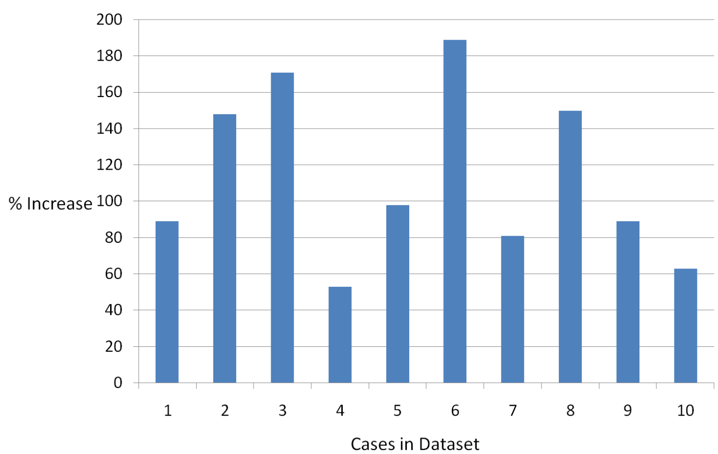


Figure 5.10 Graph representing the % increase in the vessels detected for each case in the dataset

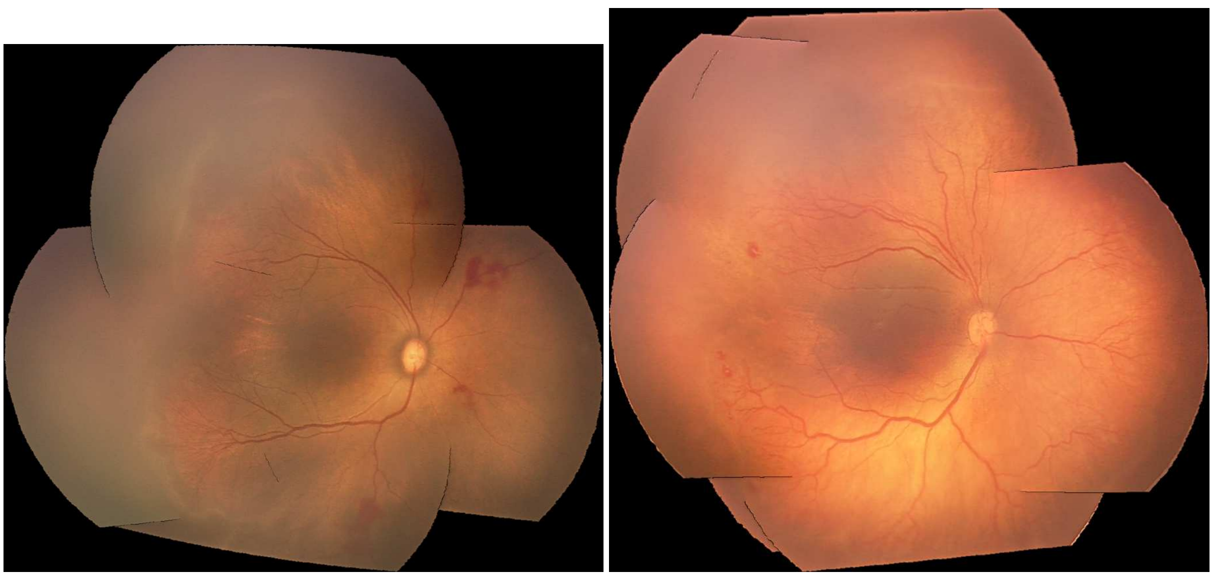


Figure 5.11 Generated Mosaics

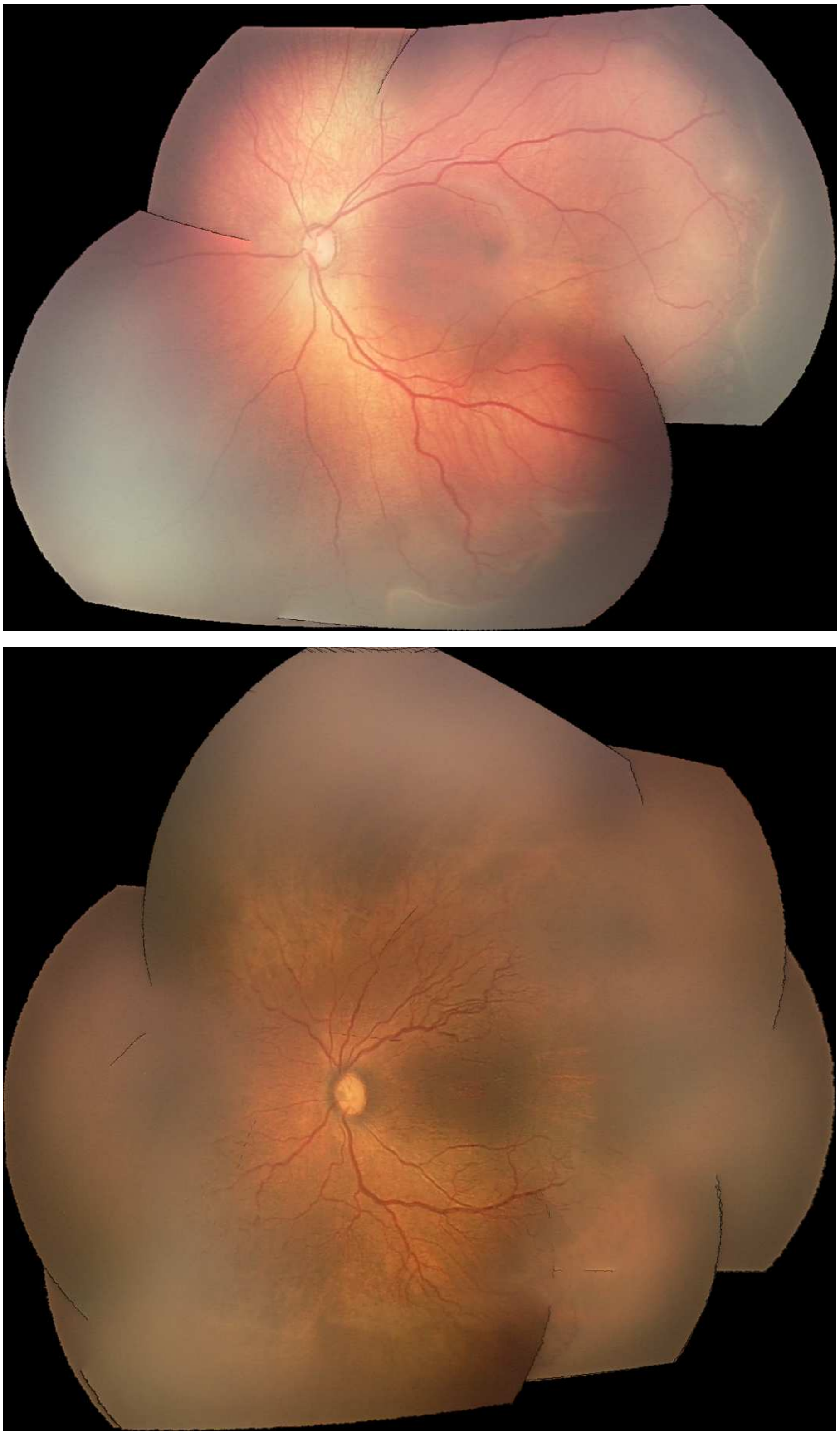


Figure 5.12 Generated Mosaics

Chapter 6

Conclusions

ROP is an eye disease that affects low birth-weight premature babies. In order to diagnose ROP, the experts observe the central region for abnormal vessels and the periphery for tissue fibrosis. For imaging the complete retinal in premature babies, a wide-angle fundus camera called RetCam is used. Multiple images of retina are obtained from various view points to capture the complete surface. Currently, doctors need to manually inspect the image sequence for diagnosing ROP and virtually reconstruct the retinal surface. This can be a tedious process when the number of images is large. Hence, there is a need for developing an automated system for aiding doctors in ROP diagnosis.

In this thesis, we have presented an image mosaicing system for neonatal retinal images which generates a single high quality image of the complete retina. The created mosaic can be used for diagnosis of ROP instead of observing the large set of images. A hierarchical framework for mosaicing large set of retinal images is introduced. This is efficiently solved by partitioning the input set and performing registration within the set rather than registering all possible pairs of images. A fast and efficient OD detection method is proposed which makes use of the color difference at the boundary of OD. This is used for partitioning the input set such that each subset represents nearly the same region of retina. The images in each of the set are spatially aligned using a pair-wise registration method. We proposed an algorithm for accurate registration of low quality retinal images using an existing feature descriptor based on Radon transform. A vessel quality metric based on ridge strength is presented which is used for global and local level quality comparison. Using the global measure, the anchor image is chosen in each set to which rest of the images are registered. For blending the registered images, the local measure is used to compare two corresponding vessel patches and choose the best information for fusion. We evaluated individual modules and show that our method performs better against other approaches. Quantitative and qualitative results are shown for the generated mosaics showing the increase in the field of view and visible vessel area. The generated mosaics have been validated by the expert to be sufficient for diagnosis of ROP.

6.1 Future Work

Possible extensions to the work presented in this thesis are

- Along with mosaicing, super-resolution techniques can be used to fuse the information across the view. This would help in generating better quality images with increase in pixel resolution
- The proposed system is currently implemented in Matlab. By porting it into C++ along with parallelization would help in building a real-time system.
- For better data fusion, quality metric for other vessel structures like Optic Disk, lesion, background tissue etc can be investigated
- Detailed clinical study of the proposed system would help in understanding the impact in general usage and also required improvements.

Related Publications

- Akhilesh Bontala, Jayanthi Sivaswamy and Rajeev R Pappuru, “Image Mosaicing of Low Quality Neonatal Retinal Images”, IEEE International Symposium on Biomedical Imaging (ISBI), Barcelona, Spain, 2012.

Bibliography

[1] http://en.wikipedia.org/wiki/File:Sydney_Harbour_Bridge_night.jpg.

[2] http://www.pasadenaeye.com/faq/faq15/faq15_text.html.

[3] AA-H Abdel-Razik Youssif, Atef Zaki Ghalwash, and AAS Abdel-Rahman Ghoneim. Optic disc detection from normalized digital fundus images by means of a vessels' direction matched filter. *IEEE Transactions on Medical Imaging*, 27(1):11–18, 2008.

[4] T. Aslam, B. Fleck, N. Patton, M. Trucco, and H. Azegrouz. Digital image analysis of plus disease in retinopathy of prematurity. *Acta ophthalmologica*, 87(4):368–377, 2009.

[5] Yogesh Babu Bathina, MV Medathati, and Jayanthi Sivaswamy. Robust matching of multi-modal retinal images using radon transform based local descriptor. In *ACM International Health Informatics*, pages 765–770, 2010.

[6] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. *European Conference on Computer Vision*, pages 404–417, 2006.

[7] A. Behrens, T. Stehle, S. Gross, and T. Aach. Local and global panoramic imaging for fluorescence bladder endoscopy. pages 6990–6993, 2009.

[8] J. Bergen, P. Anandan, K. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. pages 237–252, 1992.

[9] T. Bergen, S. Ruthotto, C. Munzenmayer, S. Rupp, D. Paulus, and C. Winter. Feature-based real-time endoscopic mosaicking. pages 695–700, 2009.

[10] F. Bignalet-Cazalet, S. Baillarin, and D. Greslou. Automatic and generic mosaicing of multisensor images: an application to pleiades hr. 8407:84070M, 2012.

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

[12] L. Brattain and R. Howe. Real-time 4d ultrasound mosaicing and visualization. *Medical Image Computing and Computer-Assisted Intervention*, pages 105–112, 2011.

[13] Peter J Burt and Edward H Adelson. A multiresolution spline with application to image mosaics. *ACM Transactions on Graphics (TOG)*, 2(4):217–236, 1983.

[14] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):679–698, 1986.

[15] D. Capel. Image mosaicing and super-resoution. 2001.

[16] T. Chanwimaluang, G. Fan, and S.R. Fransen. Hybrid retinal image registration. *IEEE International Conference on Computer Vision*, 2(5):1218–1225, 2003.

[17] T. Chanwimaluang, G. Fan, and S.R. Fransen. Hybrid retinal image registration. *IEEE Transactions on Information Technology in Biomedicine*, 10(1):129–142, 2006.

[18] Jian Chen, R Theodore Smith, Jie Tian, and Andrew F Laine. A novel registration method for retinal images based on local features. In *IEEE Engineering in Medicine and Biology Society*, pages 2242–2245, 2008.

[19] S. Choi, T. Kim, and W. Yu. Performance evaluation of ransac family. 2009.

[20] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. 1:886–893, 2005.

[21] W.M. Fierson et al. Screening examination of premature infants for retinopathy of prematurity. *Pediatrics*, 100(2):273–274, 1997.

[22] M.A. Fischler and R.C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[23] Marco Foracchia, Enrico Grisan, and Alfredo Ruggeri. Detection of optic disc in retinal images by means of a geometrical model of vessel structure. *IEEE Transactions on Medical Imaging*, 23(10):1189–1195, 2004.

[24] A.H. Gee, G.M. Treece, R.W. Prager, C.J.C. Cash, and L. Berman. Rapid registration for wide field of view freehand three-dimensional ultrasound. *IEEE Transactions on Medical Imaging*, 22(11):1344–1357, 2003.

[25] R. Gelman, M.E. Martinez-Perez, D.K. Vanderveen, A. Moskowitz, and A.B. Fulton. Diagnosis of plus disease in retinopathy of prematurity using retinal image multiscale analysis. *Investigative ophthalmology & visual science*, 46(12):4734–4738, 2005.

[26] C. Gilbert et al. Retinopathy of prematurity: a global perspective of the epidemics, population of babies at risk and implications for control. *Early human development*, 84(2):77–82, 2008.

[27] G.A. Gole et al. The international classification of retinopathy of prematurity revisited. *Arch Ophthalmol*, 123(7):991–999, 2005.

[28] C. Guestrin, F. Cozman, and M. Godoy Simoes. Industrial applications of image mosaicing and stabilization. 2:174–183, 1998.

[29] Sevket Gumustekin. An introduction to image mosaincing. 1999.

[30] C. Harris and M. Stephens. A combined corner and edge detector. 15:50, 1988.

[31] C. Heneghan, J. Flynn, M. OKeefe, and M. Cahill. Characterization of changes in blood vessel width and tortuosity in retinopathy of prematurity using image analysis. *Medical image analysis*, 6(4):407–429, 2002.

[32] M. Irani, P. Anandan, J. Bergen, R. Kumar, and S. Hsu. Efficient representations of video sequences and their applications. *Signal Processing: Image Communication*, 8(4):327–351, 1996.

[33] J. Jomier, D. Wallace, and S. Aylward. Quantification of retinopathy of prematurity via vessel segmentation. *Medical Image Computing and Computer-Assisted Intervention*, pages 620–626, 2003.

[34] E-Y Kang, Isaac Cohen, and Gerard Medioni. A graph-based global registration for 2d mosaics. 1:257–260, 2000.

[35] W. Konen, B. Breiderhoff, and M. Scholz. Real-time image mosaic for endoscopic video sequences. *Bildverarbeitung für die Medizin 2007*, pages 298–302, 2007.

[36] O. Kutter, W. Wein, and N. Navab. Multi-modal registration based ultrasound mosaicing. *Medical Image Computing and Computer-Assisted Intervention*, pages 763–770, 2009.

[37] France Laliberté, Langis Gagnon, and Yunlong Sheng. Registration and fusion of retinal images-an evaluation study. *Medical Imaging, IEEE Transactions on*, 22(5):661–673, 2003.

[38] Tony Lindeberg. Edge detection and ridge detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):117–156, 1998.

[39] D. Liu. Novel seamless image mosaicing algorithm for virtual environment. *International Journal of Computer Mathematics*, 84(2):219–229, 2007.

[40] K.E. Loewke, D. Camarillo, W. Piyawattanametha, D. Breeden, and K. Salisbury. Real-time image mosaicing with a hand-held dual-axes confocal microscope. 6851, 2008.

[41] K.E. Loewke et al. In vivo micro-image mosaicing. *IEEE Transactions on Biomedical Engineering*, 58(1):159–171, 2011.

[42] D.G. Lowe. Object recognition from local scale-invariant features. 2:1150–1157, 1999.

[43] Ahmed E Mahfouz and Ahmed S Fahmy. Fast localization of the optic disc using projection of image features. *IEEE Transactions on Image Processing*, 19(12):3285–3289, 2010.

[44] Roberto Marzotto, Andrea Fusiello, and Vittorio Murino. High resolution video mosaicing with global alignment. 1:I–692, 2004.

[45] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, 2004.

[46] R. Miranda-Luna, C. Daul, W.C.P.M. Blondel, Y. Hernandez-Mier, D. Wolf, and F. Guillemin. Mosaicing of bladder endoscopic image sequences: Distortion calibration and registration algorithm. *IEEE Transactions on Biomedical Engineering*, 55(2):541–553, 2008.

[47] S.A. Nene and S.K. Nayar. A simple algorithm for nearest neighbor search in high dimensions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(9):989–1003, 1997.

[48] D. Ni et al. Volumetric ultrasound panorama based on 3d sift. *Medical Image Computing and Computer-Assisted Intervention*, pages 52–60, 2008.

[49] D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. 2:2161–2168, 2006.

[50] Patrick Pérez, Michel Gangnet, and Andrew Blake. Poisson image editing. 22(3):313–318, 2003.

[51] WK Pratt. Digital image processing - 3rd edition. 2001.

[52] Keerthi Ram, Yogesh Babu, and Jayanthi Sivaswamy. Curvature orientation histograms for detection and matching of vascular landmarks in retinal images. In *SPIE Medical Imaging*, 2009.

[53] Peter J Rousseeuw. Least median of squares regression. *Journal of the American statistical association*, 79(388):871–880, 1984.

[54] H Samet. The design and analysis of spatial data structures. 1989.

[55] Harpreet S. Sawhney and Rakesh Kumar. True multi-image alignment and its application to mosaicing and lens distortion correction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(3):235–243, 1999.

[56] G. Shakhnarovich, P. Viola, and T. Darrell. Fast pose estimation with parameter-sensitive hashing. pages 750–757, 2003.

[57] Michal Sofka and Charles V. Stewart. Retinal vessel extraction using multiscale matched filters, confidence and edge measures. *IEEE Transactions on Medical Imaging*, 25(12):1531–1546, 2006.

[58] T.D. Soper, M.P. Porter, and E.J. Seibel. Surface mosaics of the bladder reconstructed from endoscopic video for automated surveillance. *IEEE Transactions on Biomedical Engineering*, 59(6):1670–1680, 2012.

[59] Richard Szeliski and Heung-Yeung Shum. Creating full view panoramic image mosaics and environment maps. pages 251–258, 1997.

[60] Philip HS Torr and Andrew Zisserman. Mlesac: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78(1):138–156, 2000.

[61] T. Vercauteren, A. Perchant, G. Malandain, X. Pennec, and N. Ayache. Robust mosaicing with correction of motion distortions and tissue deformations for in vivo fibered microscopy. *Medical Image Analysis*, 10(5):673–692, 2006.

[62] T. Vercauteren, A. Perchant, X. Pennec, and N. Ayache. Mosaicing of confocal microscopic in vivo soft tissue video sequences. *Medical Image Computing and Computer-Assisted Intervention*, pages 753–760, 2005.

[63] P. Viola and W.M. Wells III. Alignment by maximization of mutual information. *International journal of computer vision*, 24(2):137–154, 1997.

[64] C. Wachinger, B. Glocker, J. Zeltner, N. Paragios, N. Komodakis, M. Hansen, and N. Navab. Deformable mosaicing for whole-body mri. *Medical Image Computing and Computer-Assisted Intervention*, pages 113–121, 2008.

[65] C. Wachinger, W. Wein, and N. Navab. Three-dimensional ultrasound mosaicing. *Medical Image Computing and Computer-Assisted Intervention*, pages 327–335, 2007.

[66] D.K. Wallace, Z. Zhao, and S.F. Freedman. A pilot study using roptool to quantify plus disease in retinopathy of prematurity. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, 11(4):381–387, 2007.

[67] L. Wang, J. Traub, S. Weidert, S.M. Heining, E. Euler, and N. Navab. Parallax-free intra-operative x-ray image stitching. *Medical Image Analysis*, 14(5):674–686, 2010.

[68] C.M. Wilson et al. Computerized analysis of retinal vessel width and tortuosity in premature infants. *Investigative ophthalmology & visual science*, 49(8):3577–3585, 2008.

[69] YB Xie, P. Yang, and Y. Gong. On the graph-based panorama construction for 2d large-scale microscope images. int. *Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37:711–8, 2008.

[70] Gehua Yang, Charles V Stewart, Michal Sofka, and Chia-Ling Tsai. Registration of challenging image pairs: Initialization, estimation, and decision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(11):1973–1989, 2007.

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