Robust Registration of Retinal Images

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

Master of Science (by Research)
in
Computer Science

by

Yogesh Babu Bathina
200750024
Yogesh_b@research.iiit.ac.in

International Institute of Information Technology
Hyderabad, INDIA
April 2013

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
Hyderabad, India

CERTIFICATE

It is certified that the work contained in this thesis, titled “Robust Registration of Retinal Images” by Mr. Yogesh Babu Bathina, has been carried out under my supervision and is not submitted elsewhere for a degree.

Date

Prof.Jayanthi Sivaswamy,
Professor,
IIIT, Hyderabad

Copyright © Yogesh Babu Bathina, 2013
All Rights Reserved

This is for you, Mom, Dad and Sowmya. Thank you for believing in my dreams.

To CVIT

Acknowledgements

It is with immense gratitude that I acknowledge my supervisor Dr. Jayanthi Sivaswamy who has taught me that the best kind of knowledge to have is that which is learnt for its own sake. I am greatly indebted to her for giving me an opportunity to pursue research. Without her steadfast support, motivational ideologies, insightful criticisms and patient encouragement this thesis would not have been possible.

I would like to thank Keerthi Ram who's knowledge and valuable suggestions provided the foundation for this thesis. My deepest gratitude to Gopal Joshi for his never ending support. I specially thank Mayank Chawla who has been my technical and philosophical counterpart. I thank our lab assistant Satya for his help. Finally, I thank the almighty, my parents, my relatives and all those from CVIT and others who had at some point or the other helped me with their invaluable suggestions and feedback, and my research center, Center for Visual Information Technology (CVIT) for funding my MS by research in IIIT Hyderabad.

April 26th, 2013

Abstract

Ever so often a need arises in clinical scenarios, for integrating information from multiple images or modalities for the purposes of diagnosis and pathology tracking. Registration, the most fundamental step in such an integration, is the task of spatially aligning a pair of images of the same scene acquired from different sources, viewpoints and time. This thesis concerns the task of registration specific to three most popular retinal imaging modalities namely Color Fundus Imaging (CFI), Red-Free Imaging (RFI) and Fluoroscein Fundus Angiography (FFA). CFI is obtained under white light which enables the experts to examine the overall condition of the retina in full color. In RFI, the illuminating light is filtered to remove red color which improves the contrast between vessel and other structures. FFA is a set of time sequence images acquired under infrared light after a fluorescent dye is injected intravenously into the blood stream. This provides high contrast vessel information revealing blood flow dynamics, leaks and blockages.

Retina is a part of the central nervous system (CNS) which is composed of many different types of tissues. Given this distinctive feature, a wide variety of diseases affecting different body systems uniquely affect the retina. These Systemic diseases include Diabetes, Hypertension, Atherosclerosis , Sickle cell disease, Multiple sclerosis to name a few. Recent advancements reveal a close association of retinal vascular signs to cerebrovascular, cardiovascular and metabolic outcomes. Simply put, the health of blood vessels in the eye often indicates the condition of the blood vessels (arteries and veins) throughout the body.

Registration of multimodal retinal images aids in the diagnosis of various kinds of retinal diseases like Glaucoma, Diabetic Retinopathy, Age Related Macular degeneration etc. Single modality images acquired over a period of time are used for pathology tracking. Registration is also used for constructing a mosaic image of the entire retina from several narrow field images, which aids comprehensive retinal examination. Another key application area for registration is surgery, both in the planning stage and during surgery for which only optical range information is available. Fusion of these modalities also helps increase the anatomical range of visual inspection, early detection of potentially serious pathologies and assess the relationship between blood flow and the diseases occurring on the surface of the retina.

The task of registering retinal images is challenging given the wide range of pathologies captured via different modalities in different ways, geometric and photometric variation, illumination artifacts, noise and other degradations. Many successful methods have been proposed in the past for the registering retinal images. A review of these methods show good performance over healthy retinal images. However, the scope of handling a wide range of pathologies is limited for most of the approaches. Further, these methods fail to register poor quality images, especially in the multimodal case. In this work, we propose a feature based retinal image registration algorithm capable of handling such challenging image

pairs.

At the core of this algorithm is a novel landmark detector and descriptor scheme. A set of landmarks are detected on the topographic surface of retina using Curvature dispersion measure. The descriptor is based on local projections using radon transform which characterizes local structures in an abstract sense rendering it less sensitive to pathologies and noise. Drawing essence from the recent developments in robust estimation methods, a modified MSAC(M-estimators Sample and Consensus) is proposed for false correspondence pruning. On the whole, the minor contributions at each stage of feature based registration scheme presented here are of significance. We evaluate our method against two recent schemes on three different datasets which includes both monomodal and multimodal images. The results show that our method gives better accuracy for poor quality and pathology affected images while performing on par with the existing methods on normal images.

Contents

List of Figures

List of Tables

Chapter 1

Image Registration

1.1 Introduction

Unlike machines, humans perform vision related tasks like recognizing similar objects, patterns, locations, etc with utmost ease. Image registration is the primary step to facilitate machines to perform some of these tasks. The goal of registration is to spatially align two or more images of the same scene/object acquired from different sources, view points and time. Registration estimates the deformation between images and transforms one image into the coordinates of the other to align them.

Fig 1.1 shows an example of the 2-D registration where two Landsat images of the same area are taken at different times using different sensors. In abstract terms, here registration is achieved by identifying similar/corresponding regions (labeled in Fig 1.1) in both the images and overlaying one on top of the other to align them.

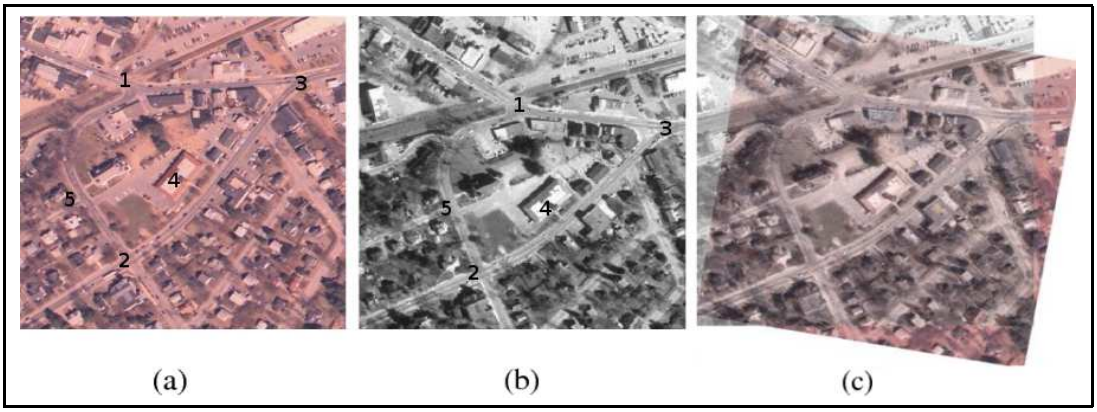


Figure 1.1: (a) & (b) showing two aerial photos of the same scene and their corresponding points (c) registered images

Registration is required in remote sensing (multi-spectral classification, environmental monitoring, change detection, image mosaicing, weather forecasting,integrating information into geographic infor-

mation systems (GIS)), in medicine (diagnosis, surgery planning and during surgery, mosaicing, tracking, fusion etc), in cartography (map updating) and in computer vision (tracking,stereo-vision, object recognition, target localization, super-resolution images etc).

The last two decades have seen a remarkable development in imaging technology. Given the diverse nature of the images captured and their use in day to day life, many researchers have tried to solve the most primary, yet a challenging problem in image processing and computer vision - *Image registration*. It has grown from being perceived as a minor precursor to a sub-discipline of its own [1].

Though significant progress has been made in this subfield, due to the wide variety of images, modalities and numerous degradations, it is impossible to develop a generic registration algorithm to cater to all the applications. Many details have to be taken into account to develop a registration technique for a specific application, like geometric and radiometric deformations, noise, image characteristics, required accuracy and stability, etc. This thesis concentrates on retinal image registration. The domain specific background is introduced in the next section.

1.2 Background

The retina is a light sensitive tissue lining the inner surface of the eye Fig 1.2(a). The image of the visual world is created on the retina, much similar to a basic camera. The retina is a complex interconnected multi-layered network of neurons with the surface lined with special photosensitive cells called the photoreceptors. The basic anatomy of the retina includes the optic disk, macula, fovea and vasculature Fig 1.2(b). The optic disk is an oval shaped bright disk where all the vessels appear to converge. The optic nerve is made up of thousands of nerve fibers which pass electrical signals to the brain for further processing. It is also called the blind spot due to the lack of photoreceptors cells in the area. Next to the optic disk is the macula with fovea at its center. Fovea has high concentration of cone cells (a type of photoreceptor) which is responsible for our sharp vision, but is less sensitive to the light.

There are many inherited and acquired diseases or disorders that may affect the retina, like macular degeneration, hypertensive retinopathy, diabetic retinopathy, etc (Fig 1.3). As the retina is a part of the central nervous system, it is an excellent indicator of systemic pathologies. The first digital fundus camera was developed in 1990, to diagnose these diseases related to the retina. The digital fundus camera (retina camera) is a specialized low powered microscope attached to a camera to capture images of the retina. The retinal imaging equipment today magnifies up to 2.5x with a resolution of 3000x3000 and angular resolution ranging between 10° to 50°. However, imaging the retina is analogous to peeping inside a closed room through a key hole, which means that only a part of the retina can be imaged at a time. As a part of National Institute of Health (NIH -U.S.A), Early Treatment Diabetic Retinopathy Study (ETDRS) committee has set standards for the retinal imaging. The ETDRS imaging protocol specifies seven fields of view for the retina. The specifications include the minimum region of overlap,

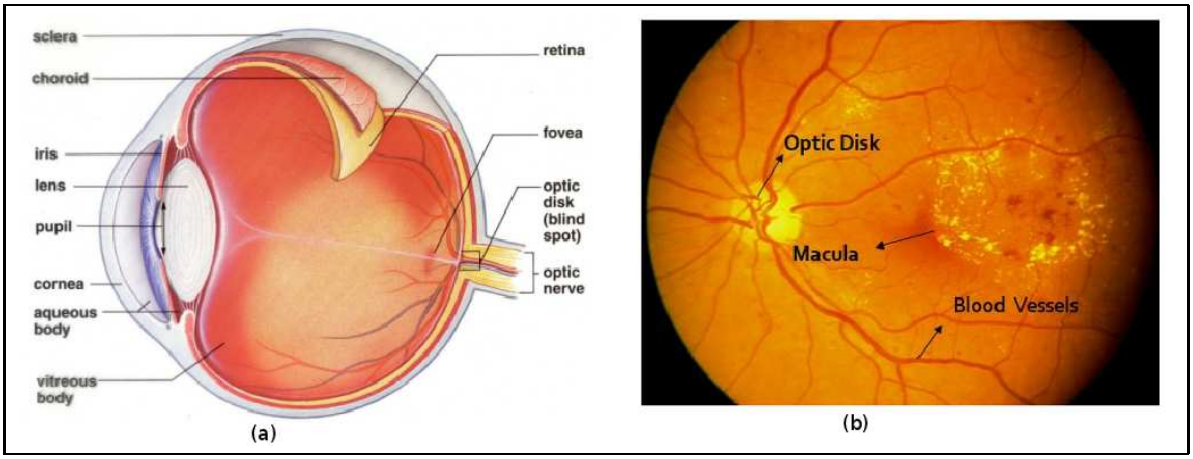


Figure 1.2: (a)Anatomy of the Human Eye (b) Retinal Image

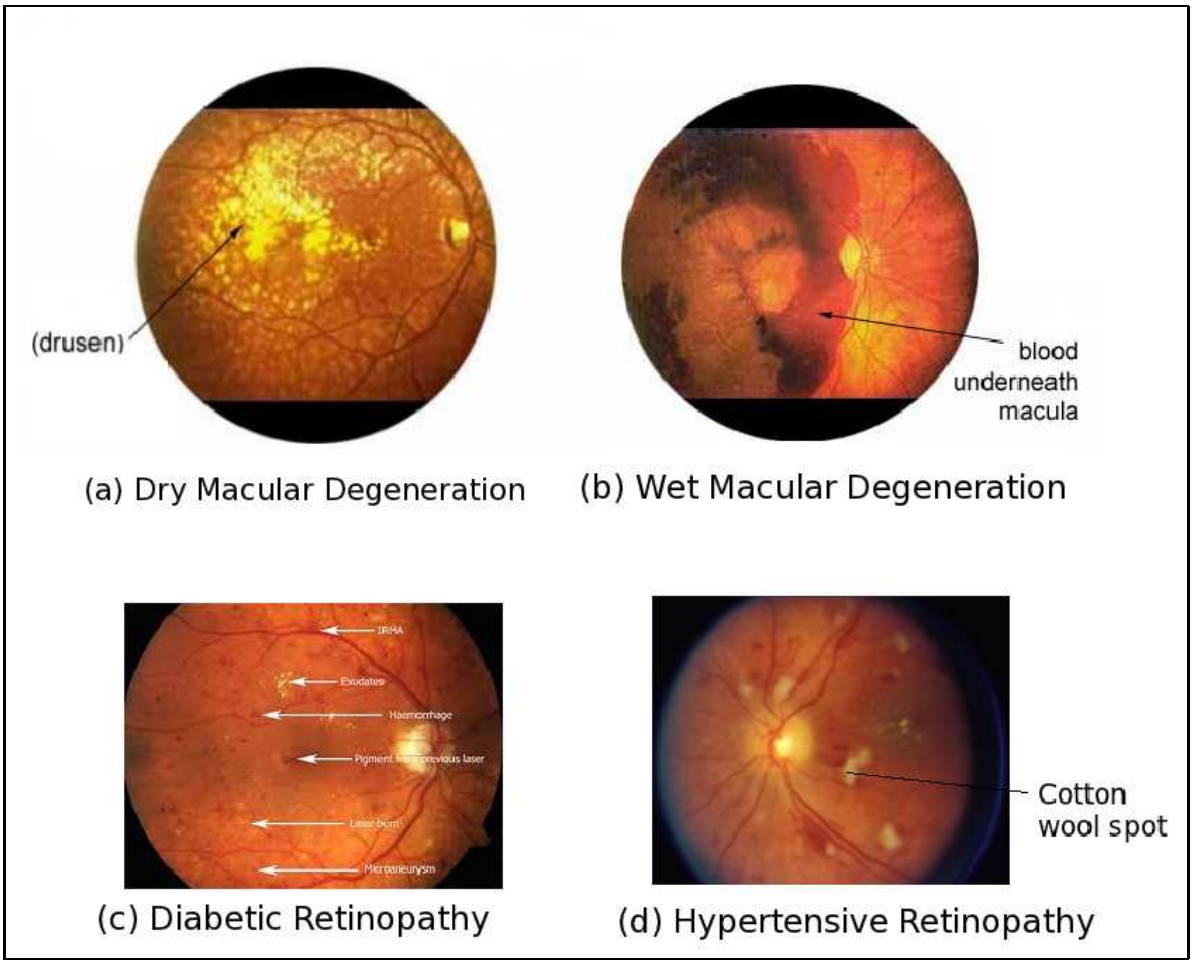


Figure 1.3: Common Retinal Disorders

minimum number of views, resolution, quality etc. However in practice these rules are seldom followed. In practical scenarios, images vary in resolution, overlap, illumination changes, etc.

In medical imaging, various types of equipment, probes or methods used to acquire images of the body are referred to as modality (Example: Computed Tomography, X-ray, Ultrasound, etc). The most popular fundus photography modalities are Color Fundus Image (CFI), Red-Free Image (RFI) and Fluorescein Fundus Angiogram (FFA), Fig 1.4 . CFI is obtained under white light which enables the experts to examine the overall condition of the retina in full color. In RFI, the illuminating light is filtered to remove red color which improves the contrast between vessel and other structures. FFA is a set of time sequence images acquired under ultraviolet light after a fluorescent dye is injected intravenously into the blood stream. This provides high contrast vessel information revealing blood flow dynamics, leaks and blockages. Despite the contrast difference between CFI and RFI, they reveal similar optical information about retina. Aligning images acquired from a single modality is called monomodal registration and we refer to term multimodal registration when images from more than one modality are to be aligned.

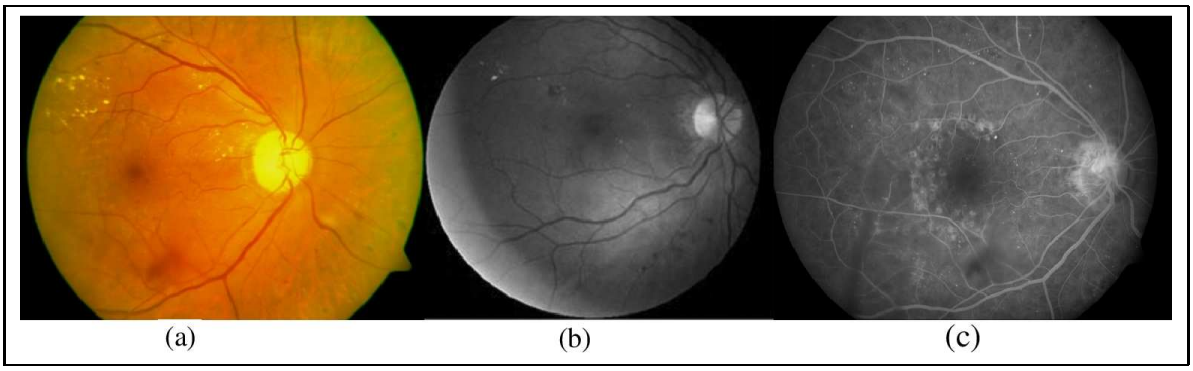


Figure 1.4: Image showing the following modalities of images:(a) Color Fundus Image (b) Red Free Image and (c) Fluorescein Fundus Angiogram

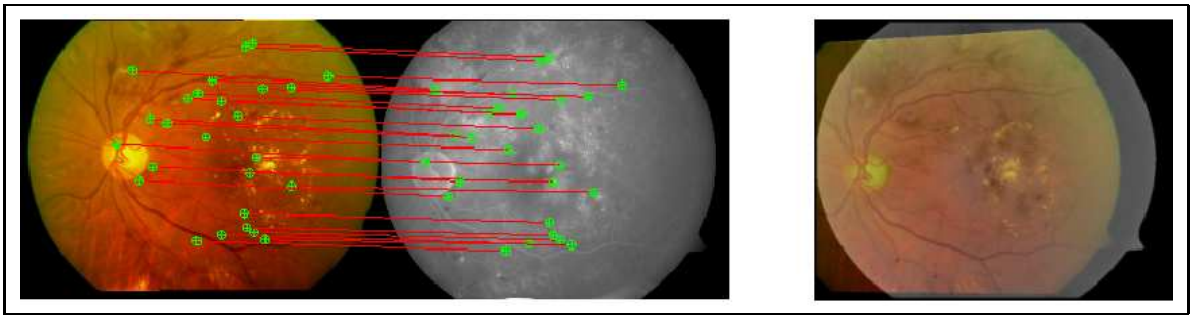
1.3 Problem Statement

Given two multimodal images $f(X)$ and $g(X')$ where $X = (x_1, x_2...x_n)$, n =2 and $X' = (x'_1, x'_2...x'_n)$for 2-D images, the goal of registration is to estimate a transformation function T that establishes pixel to pixel mapping between these images.

$$T\colon f(X) \rightarrow g(X') \Leftrightarrow T(f(X)) = g(X') \tag{1.1}$$

The transformation function T defines the deformation between the images, Fig 1.5. This deformation may include rotation, shift, scale, etc, which are controlled by a set of parameters. The number of parameters is called “Degrees of freedom”.

Another way of defining the problem would be through the coordinate system. The images captured by the acquisition device are stored in different coordinates frames due to the lack of any reference coordinate system. Registration can be defined as the process of transforming one of the images into the



(a) (b) (c)

Figure 1.5: Registration of a sample multimodal image pair (a) $f(X)$ (b) $g(X')$ and (c) $T(f(X)) = g(X')$

coordinate system of the other. The image taken as the reference is called the fixed image and the image to be transformed is known as moving image.

1.4 Motivation & Challenges

In this thesis, we propose a generic framework for monomodal and multimodal registration. Monomodal images taken over a period of time are used for pathology tracking. Different views of the retina obtained from the same modality are combined into a single image to form a mosaic to inspect overall health of the retina. Multimodal registration is the primary step in fusing complementary information contained in different imaging modalities for diagnostic purposes and also to track pathologies over a period of time. Information obtained through registration of two imaging modalities aids in the diagnosis of various kinds of retinal diseases such as glaucoma, diabetic retinopathy and age related macular degeneration. A key application area for registration is surgery, both in the planning stage and during surgery at which time only optical range information is available. Fusion of these modalities also helps increase the anatomical range of visual inspection and provides a means for early detection of potentially serious pathologies and reveals the relationship between the blood flow and the diseases occurring on the surface of the retina.

The challenges involved in retinal image registration can be summarized as follows: (1) The retina is a curved surface and its projection onto the imaging plane induces radial distortion. To model this, a second order polynomial or a quadratic transformation model is usually employed [2]. (2) The images may be obtained from two uncalibrated cameras. (3) The imaging setup parameters might vary as per the requirement of the clinical expert. The CFI image is usually low in resolution (as compared to FFA) which is enough to examine the overall health of the retina for a preliminary diagnosis. (4) The images are obtained from the same retina using different sensors captures complementary information due to which the pathologies(MA, Exudates, etc) may not be perceived in all of them, resulting in false matches, (Fig 1.6(a)-(b)). (5) FFA has better contrast and resolution than CFI as it is obtained under infrared light invasively but the time of acquisition in FFA plays a major role. The image may not be taken before the dye enters and leaves the retina, which are perceived as dull parallel edges in FFA

images. We refer to this as ill acquisition timed FFA. Fig 1.6(d). (6) The captured field of view is variable and is specified in degrees of visual angle between 30°-50°. This affects the magnification level of structures visible in the images to be registered as well as the degree of overlap between them. (7) Variability in contrast across multimodal images is a common problem in practical scenarios. (8) The common artifacts in the CFI is the non-uniform illumination and green channel noise which is due to the natural light entering the eye and the internal reflection at the surface of the retina, thus the quality of the image is compromised, Fig 1.6(c)-(e). (9) The overlap between the multimodal images is critical for the accurate registration, the strength of registration method depends on its ability to align images with minimum percentage of overlap, Fig 1.6(g)-(h). (10) In longitudinal studies, images can have a wide temporal separation. Hence, new pathologies can appear and disappear over time.

A variety of methods have been reported in the literature which address these challenges to varying degrees of success [1-30]. Existing approaches for retinal image registration perform poorly in practical scenarios due to variabilities mentioned above. In this thesis we especially lay emphasis on low resolution poor quality multimodal retinal images in the presence of diverse pathologies Fig 1.6(i)-(t).

1.5 Overview

The goal of registration is to estimate the deformation between the images while taking the domain specific information into consideration. A closer look at the problem statement intuitively reveals two methods of solving it. The first method operates directly on image intensity values, continuously transforming the entire image so as to align it with the other. The image is considered to be registered when desirable alignment is obtained for the respective transformation. These methods are called area based methods. The second method relies on a few salient points which are most prominent in both the images. The goal here is to detect the corresponding pairs of points/regions across the images from which the deformation is estimated. These are known as feature based methods.

Feature based methods have gained popularity over the area based methods as they are more robust to illumination changes, partial overlap between the images, occlusion, changes in background, and viewpoint [3]. Despite these advantages, area based methods are still preferred over feature based methods in the medical domain due to two main factors:(1) their ability to handle local deformations, which are especially the case with human organs and (2) dealing with information from different imaging modalities/sources.

Retina is the only part of the central nervous system which can be imaged directly. Retinal imaging is unique when compared to its other radiology(medical) counter parts. Retina is a curved surface and only a part of it can be imaged at an instance. The most important difference is that the transformation between any two images of the same retina do not have any significant local deformations. So, a feature based strategy can be used if the issue of multi-modal information is addressed.

In computer vision, feature based matching methods like SIFT, SURF, GLOH etc, which belong to the class of invariant detectors/descriptors, have received significant attention over the last decade [3].

However, these methods are limited to natural images which are obtained by directly recording the incident light. But medical imaging, in a restricted sense, is seen as the solution of mathematical inverse problem. This means that cause (the properties of living tissue) is inferred from effect (the observed signal) [1]. This inference is biased by modality specific noise, artifacts and other degradations. In monomodal case, often artifacts like green channel noise are present in only a single image which leads to large variations between both the images. In multimodal datasets, the information captured by a specific modality reveals only certain characteristics of the object. For example, CT scan reveals hard and dense structures better while MRI shows details in the soft tissues. Hence, general feature based methods perform poorly on multimodal medical data due to the complementary information present in them. In this work, we propose a novel feature based registration method capable of handling these monomodal and multimodal variations across images.

Feature based methods typically follow a three step approach - detection of significant landmarks across images, establishing correspondence using features extracted around landmarks and the estimation of the transformation function using correspondences. Each step in registration has its typical problems and the contribution of this thesis lies in these individual modules. For each module, the contribution level is graded as minor or major based on the novelty and innovation.

1. *Vessel Enhancement*: A normalized vesselness measure on the lines of scale space theory proposed by [4] has been put forth for vessel enhancement in retinal images. The role of this step is to bring a given image pair into a single representation thus bypassing the multimodal changes and rendering it invariant to illumination, contrast changes and other noise factors. -***Minor***

2. *Landmark Detection*: We propose a novel measure of curvature dispersion on the topographic surface of the image to detect landmarks.-***Major***

3. *Radon Descriptor*: A Radon based descriptor is introduced for scale invariant robust matching in retinal images. This projection based local shape descriptor captures abstract higher level information thus rendering the descriptor less sensitive to lesions and noise. This helps establishing accurate correspondence in poor quality cases. -***Major***

4. *False match rejection and initial transformation estimation* A Variant of MSAC (M-estimator sample and consensus) from robust statistics is proposed to reject false matches and estimate the initial transformation across images. -***Major***

5. *Transformation model selection* A novel transformation model selection scheme is introduced which exploits the information on the spatial distribution of the matches. -***Minor***

1.6 Organization of the Thesis

This thesis is organized as follows: Chapter 2, gives a detailed survey of general registration schemes. Chapter 3 gives an elaborate report on our initial work on monomodal retinal image matching. In Chapter 4, the drawbacks of this work are addressed and a robust framework for multimodal and monomodal

retinal image registration scheme is proposed. We conclude the thesis in the last section of Chapter 4 and show additional retinal image registration results in Appendix I.

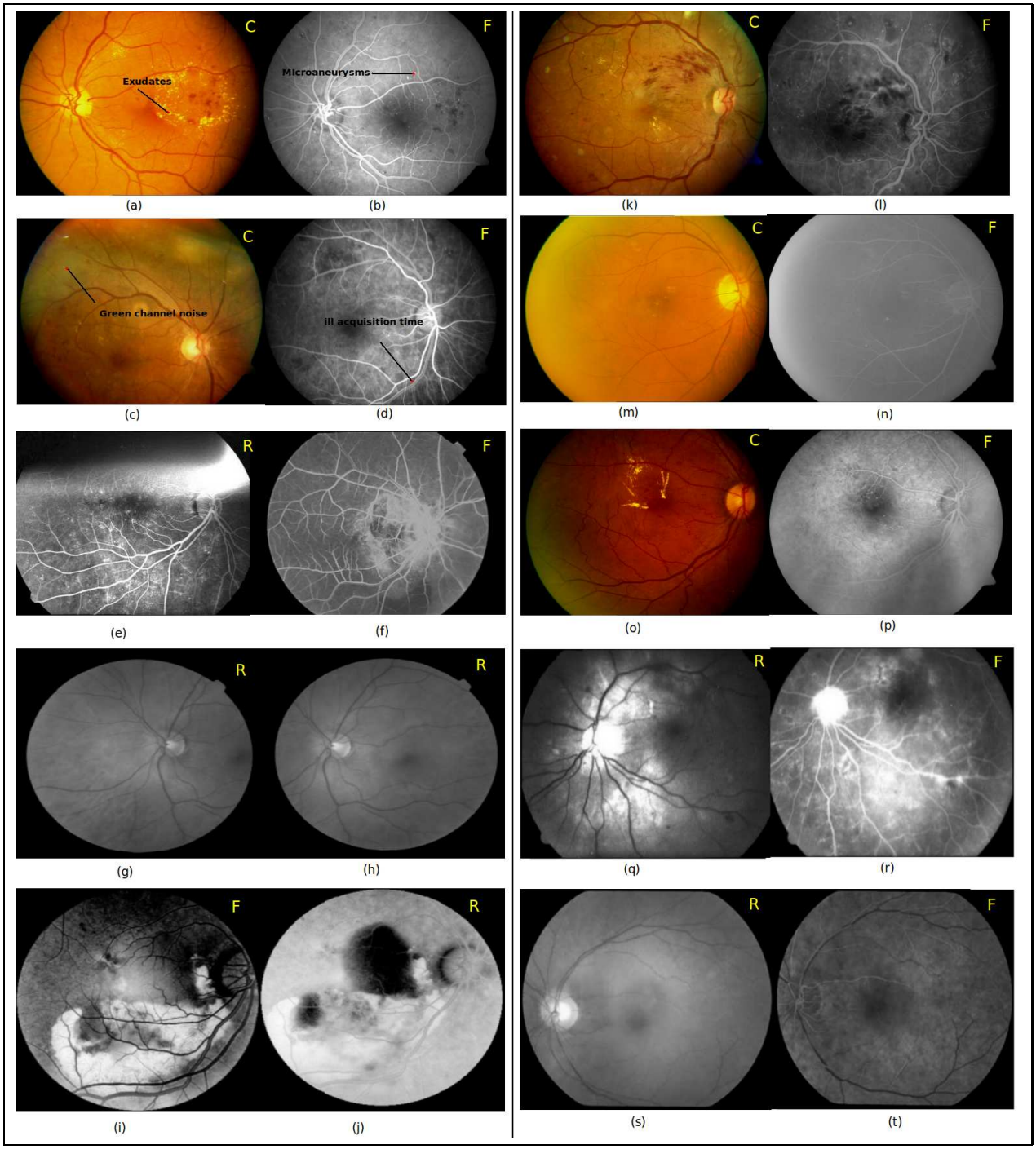


Figure 1.6: (a)-(b) Image pair with complementary information,(c) shows green channel noise in CFI image, (d) ill timed FFA capture showing dull edges, (e) Image showing shine through artifact, (f) pathology affected case, (g)-(h) low overlap case, (i)-(j), (k)-(l), (m)-(n), (o)-(p), (q)-(r), (s)-(t) show some challenging pairs. The yellow labels C, F, R on the images represent CFI, FFA and RFI images respectively.

Chapter 2

Background

Overview: This chapter gives a review of existing registration schemes while classifying them based on application, method and strategy. The background introduced in this chapter provides insights into various techniques, their advantages and shortcomings. The major goal here is to present generic registration schemes which help justify and motivate the design of our proposed method for multimodal retinal image registration.

The first classification is based on the types of applications, which narrows down the rest of the registration survey to methods designed for applications similar to the ones at hand. Next classification criteria is the nature of the method: Area based Vs Feature based methods. Feature based methods are reviewed in details where as the area based methods are restricted to classical schemes, keeping the scope of this thesis in view. Finally, the classification of the transformation models is reviewed here which paves way to model the deformations specific to the retinal images in the later chapters.

2.1 Classification of Image registration methods

All large systems which are used in evaluating images use registration or a similar operation as an intermediate step. As mentioned previously, to devise a generic registration method for all the applications is impossible due to variations in modalities, deformations and other degradations. Over the years, broad range of techniques have been proposed for various applications in image processing and computer vision. Several successful attempts have been made in the past [5] [6] to classify these techniques and wrap flexible frameworks around the problem for assisting in the selection of the most suitable technique for a specific problem.

There are many ways to categorize existing registration methods, detailed explanation can be found at [7], [8] and [3]. [3] gives the recent classification of registration methodologies based on application, feature detection, matching, optimization methods and transformation estimation. Borrowing ideas from [3], we classify the registration methods under three main categories

- Type of applications
- Area Vs Feature based methods

- Global Vs Local mapping models

The next few section have been adopted from [1].

2.1.1 Types of applications

According to [3], the applications of image registration can be categorized into four classes. (i) *Different view point*: The images acquired belong to the same scene but taken from different viewpoints. here registration is used for mosaicing, shape recovery, super resolution, depth estimation, stereo-vision etc. (ii) *Different times*: The images are taken at different times instances are used for disease tracking, change detection, motion tracking etc. (iii) *Different sensors*: Often in the medical domain there is a need to combine or fuse the information obtained from different sensors so that the complementary data can be integrated into a single image. For example, CT (Computed Tomography) reveals anatomical structures likes bones, hard tissues etc and PET (Positron Emission Tomography) reveals the functional information. Fusing these two modalities aid in surgery planning and intervention. (iv) *Scene to Model*: Images of a scene are registered to a model, which is a computer representation of the scene. These are quite popular in developing atlas and specimen classification.

2.1.2 Area Vs Feature based methods

The goal of the registration is to estimate the deformation between the images, a closer look at the problem statement intuitively reveals two methods of solving it. The first method continuously transforms the entire moving image so as to align it with the fixed image. The image is considered to be registered when desirable alignment is obtained for the respective transformation parameters. These methods are called area based methods. The second method relies on a few salient points which are most prominent in both the images. The goal here is to detect the corresponding pairs of points/regions across the images from which the deformation is estimated. These are known as feature based methods. They shall be dealt in detail in the next sections.

Area based methods:

The area based methods operate directly on the image intensity values of specific region of interest(ROI-window) or the entire image without deriving any structural information. The idea behind area based method is quite simple. Given two images $f(X)$ and $g(X')$, fixed and moving image respectively. The moving image is transformed(shifted, rotated and scaled which are called transformation parameters) to create a perfect alignment with the fixed image. The alignment or correspondence is validated for its goodness with the help of a similarity criteria. The images are considered to be registered when the similarity criteria reaches its maximum value for a given set of parameters. This methods typically use an optimization framework since exhaustively searching in this space explodes with the increase in the transformation parameters.

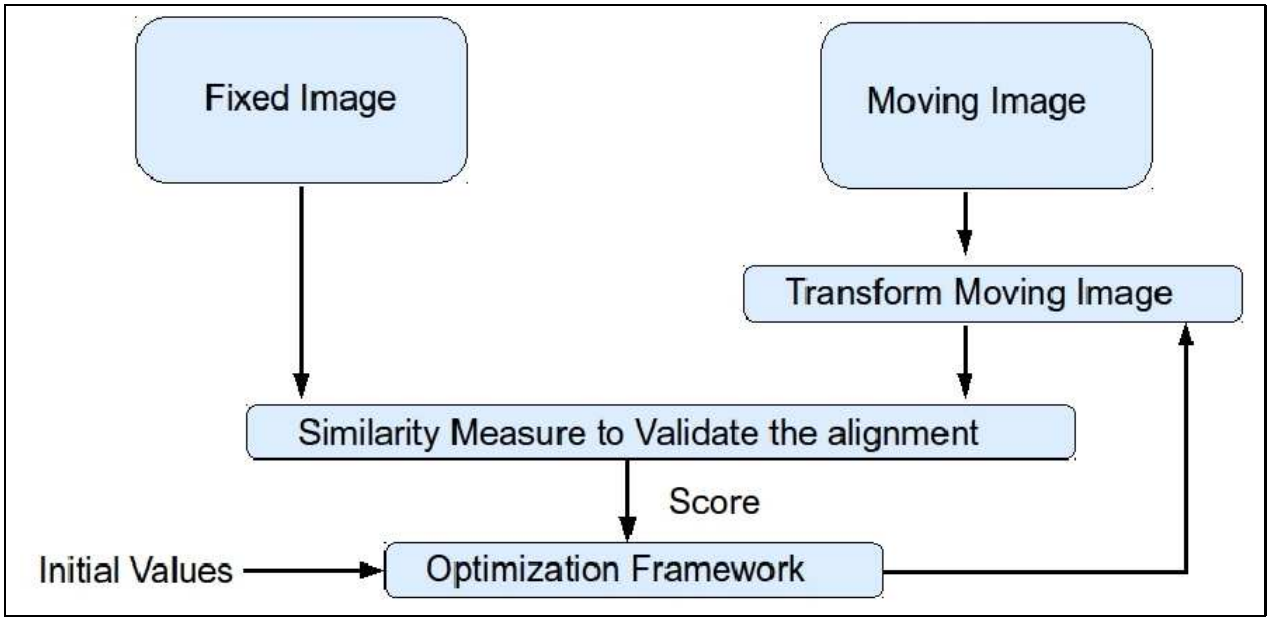


Figure 2.1: Area based method for image registration

Essentially area based methods have three components

1. similarity measure
2. optimization method
3. transformation model.

(i) *Similarity Measure*: As the name implies similarity measures determine how closely two signals/ images are related. The higher the score of the measure the better is the alignment between the images or specific window pairs. So the objective in this formulation is to find the transformation function which maximizes the similarity measure. One of classic similarity measures used for registration is Normalized Cross-Correlation.

$$CC(i,j) = \frac{\sum(f(X) - E(f(X)))\ (g(X'_{(i,j)}) - E(g(X'_{(i,j)})))}{\sqrt{\sum((f(X) - E(f(X))^2\ \ \sum((g(X'_{(i,j)}) - E(g(X'_{(i,j)}))^2}} \tag{2.1}$$

where (i, j) are the spatial coordinates and $E(.)$ is the expectation of the image.

Several other measures have been proposed like covariance criterion, correlation coefficient, cosine angle criterion,etc ([9]). The major drawback of this method lies in its inability to handle large scale and rotation factors. Also when the region of interest is smooth, the discriminability decreases and leads to false matches. Even with these limitations they have been used extensively used in the medical domain, especially dealing with multi-sensor data.

Mutual Information similarity measure has emerged from the information theory and was first proposed by [10] in 1992. It gives the statical dependency of the intensity values especially when the structural information is not prominent.Mutual information is given by

$$MI(f(X), g(X')) = H(f(X)) + H(g(X')) - H(f(X), g(X')) \tag{2.2}$$

Where $H(.)$ and $H(.,.)$ are the marginal and joint entropies respectively, given by

$$H(x) = -E_x(log(P(X))) \tag{2.3}$$

and $P(X)$ is the probability distribution of X.

Mutual Information is recent e technique for registration of multi-sensor images especially in the medical domain. Anatomical images like CT, MRI and function images like PET, SPECT differ in structure as well as information content. Mutual Information exploits the intensity distribution between the images to align the complementary data. The other similarity measure which have recently been proposed like cross-entropy [11] , Entropy [12] have proven to be more effective than MI.

Mutual information only exploits the statistical dependency between the two distributions but does not embed any neighborhood information. Recently, [13] proposed mutual information of regions which takes even the neighborhood information into account. Further modifications and improvements have been made in this area, which is beyond the scope of this thesis.

(ii)*Optimization*: To find the maxima of a similarity measure(or minima of a dissimilarity) for given set of transformation parameters exhaustively over the entire transformation space explodes with the increase in the degrees of freedom. Brute force may be used when the transformation includes only translation, but in case of higher transformation models an optimization framework is required to localize the maxima. Various optimization methods have been used for the registration of multi-modal images, like Gauss-Newton minimization, Gradient Descent, Levenberg-Marquardt, Powell optimization etc. A detailed literature review of these optimization schemes and its application can be found at [5], [6], [3]. Often in this framework, a regularization term(penalty term) is included with the objective function. The regularization term interconnects the transformation and the data to be transformed. There are referred as the energy minimization methods in the literature.

(iii)Transformation models with be dealt in detail in the later sections.

Feature based method

Feature based methods are used when local structural information has better distinctive signatures as compared to the intensity distributions. This method is posed as a correspondence problem i.e given a region/point in fixed image, find the homologous region/point in the moving image. By establishing the relationship between corresponding points across both the images, the deformation is estimated.
Feature based methods typically have four steps:

1. Feature detection
2. Feature matching
3. Transformation(deformation) estimation
4. Image resampling and transformation

(i)***Feature detection***: Feature detection aims at computing higher abstractions of local intensities of an image. Feature based methods are driven by a set of easily detectable salient points/regions in both the images. These features may be points,lines, curves and regions which are well spread across the images. This method demands a reasonable amount of overlap between the images even in the presence of object occlusion or any other noise factors. The features must be well localized and should be invariant to local degradations.

Point Features: As the name implies these are the salient points in an image which may represent corners, high variance points, centroids of regions or local curvature discontinuities. Point features range from simple corners, edge intersections to sophisticated transformation(geometric) invariant detectors. These are generally known as interest point detectors [14] [3]. Point based features shall be dealt in detail in the following chapters.

Line Features: line features represent discontinuities in image intensity. These discontinuities are referred to as edges (lines, curves, region contours,etc.). There are two main methods to detect edges in a image- Search based and Zero-crossing based. Search based methods usually compute the edge strength(usually first order derivative) and search for a local maxima direction to detect the edges. Zeros crossing based methods compute the second order derivatives directly to find these discontinuities. Recently edge detection via phase congruence [15], which is a frequency domain approach, has proven to be more robust to noise and other degradations.A survey of a number of different edge detection methods can be found in [16].

Region Features: Regions are closed-boundary areas with specific properties in intensity distribution, texture, color etc. Salient region detection can also be associated to process of image segmentation. The goal is to classify the image into two (in some cases more) classes- region and non-region, based on the properties mentioned above. These regions which are detected are usually represented by their center of gravity or by any other higher order moments [17]. The region detected must be stable in the presence of deformations like skewing, scaling, rotation,etc, and invariant to contrast and other random noise variations. These region based detectors are generally referred to as blob detectors. In early literature, LOG,DOG, determinant of hessian have been used extensively [18]. Recently, [6] proposed a method based on this scheme to simultaneously segment and register remote sensing images. Regions detected based on homogeneity of intensities was proposed by [19] called maximally stable extremal regions(MSER), the regions detected are invariant to a wide range of deformations.

The features detected in both the images shall be referred to as control points.

(ii)***Feature matching*** At the feature detection stage a set of control points have been extracted from both the images which are to be matched to establish pairwise correspondence across images. Feature matching step is the most crucial part of image registration framework. Features that appear similar may not be matched accurately due to false detections at the detections stage, unpredictable imaging conditions and different sensors. A trade off is usually made between the discriminability of different features and their ability to handles variations in noise and other artifacts. Also the matching strategy has to handle features that do not have any correspondence pair.

Various methods like spatial relations, descriptors, relaxation methods wavelets and pyramidal approaches exist in literature for feature matching. In this thesis we focus on the first two methods only.

Spatial relation based method: These set of methods exploit the spatial relations between the control points to establish correspondence. Goshtasby [20] proposed a method based on graph matching to register remote sensing images. This method does not incorporate any local neighborhood information, which makes it apt for applications where the local information is ambiguous or corrupted. It is particularly useful for applications like mosaicing [21]. [22] described a clustering based technique to establish correspondence. For all the each control point (CP) pair to its the transformation parameters are represented as a point in transformation space. The control points with accurate matches are clustered together where as the other combinations are spread out in this space. The centroid of this cluster represents the best transformation parameters which yield accurate matches. In this method, both matching and transformation estimation are coupled together. A method based on minimization of edge distances was presented by Borrow [23] called the chamfer matching. This method was further improved by [24] by using distance transform in root mean square minimization framework. Iterative close point algorithm introduced by Besl and Kay [25] is well known for its success in 3D registration. This algorithm refines matches based on spatially closest feature around a single accurate correspondence established.

*Invariant descriptors:*Descriptors establish correspondence using the information extracted from the close neighborhood of the detected features. It is a compact representation which describes the neighborhood of a feature at a higher(abstract) level. For every feature that is detected, a descriptor is extracted which represents unique structural characteristics of the region. Given a set of CPs and their respective descriptors for both the images, all the points in the fixed image are matched with ones in the moving image to compute a matching score. Every point in fixed image that has a highest matching score to its moving image counterpart, are tagged as corresponding pairs.

With a goal of establishing a generalized criteria for descriptors,they are to fulfill several conditions. Invariance is an important property which enables matching even when the image is deformed. This assumes that the descriptor for a known correspondence will not significantly vary under geometric or photometric distortions. Uniqueness defines the discriminative property of the descriptor to characterize a feature. The descriptor must be similar to its corresponding counterpart and differentiate false matches with a large margin. Repeatability defines the ability to yield similar matches under varying imaging condition, time instances and noise.

The most simple descriptor is the image intensity. Around each feature, a window of known size is extracted and correspondence is established by finding window pair across images which have maximum similarity. Correlation Coefficient [26], Cross-Correlation and Mutual information [27] are well established similarity metrics for ROI matching based on image intensities.

To represent local shape, specific descriptors are used which embed the structural information in them. These shape descriptors include chain code representation, polygonal approximations, shape numbers and Fourier descriptors [28]. For close boundary regions, moment based descriptors have been successfully used for registration. Hu [17] introduced invariant moments which describes the region contours by projecting the local binary patterns into higher order dimensions. Flusser and Suk [29] gave affine invariant moments which was illustrated over landsat images. Holm [30] integrated moments with geometric properties like perimeter and area. Flusser [31] introduced blur invariant moments which were further improved by incorporating rotational invariance called combined blur- rotational invariant descriptors.

Another class of detectors and descriptors which have gained popularity over the recent years are invariant interest point/region detectors and descriptors like SIFT(scale invariant feature transform) [32], SURF(speed up robust features) [33], etc which will be discussed in detail in chapter 4.

(iii) ***Transformation Estimation:*** After the correspondence is established the next step is to estimate the transformation/mapping between the images. The transformation function must align the moving image with the fixed image irrespective of the occlusion and overlap. Transformation must handle errors and distortions induced by the imaging equipment, while attaining the required alignment error. As mentioned above, given the variety of images available, a single transformation model cannot be applied to all the registration problems. For example, given two images which are deformed by just translation, then assuming a model with rotation,translation and scale would be an overkill. Assuming the right model for the problem is called the model selection problem.

Appropriate selection of the transformation model assists in selecting/ developing the techniques for registration. So, the next section is dedicated to the classification of transformation models, their variants and geometric properties.

(iv) ***Image resampling and transformation:*** Once the transformation function has been estimated, the next step is to transform the moving image into the the coordinates of the fixed image by resampling. The transformation may be realized as forward or backward mapping. Forward mapping methods are often complicated to implement and also produce holes in the image, which is why backward mapping techniques are a better choice. To transform the moving image into the fixed images coordinates interpolation schemes are used to fill holes which occur due to the discrete nature of the grid. Interpolation is a convolution operation via the interpolation kernel. The ideal kernel is a 2-D sinc function which is difficult to implement due to the infinite extent of the function. Thus approximations are used to reduce the computational cost. Popular interpolation functions are bilinear, bicubic, quadric splines, cubic splines and gaussians. These interpolation schemes are usually subjected to a trade off between accuracy and computational cost, which is why Bilinear and Bicubic are often used. A detail survey and comparison of resampling are investigated in [34].

2.1.3 Global Vs Local mapping

In mathematics,transformations models are studied under the name geometric transforms. They deal with the properties of objects under various deformations. These Transformation models can be classified as : Global mapping and Local Mapping methods. A more popular way of classification is rigid Vs non-rigid, but for the purpose of developing background for this thesis, we stick to the initial classification. Global mapping is used when the deformation between images can be expressed a single function which describes the mapping for the entire image. These method are used only in the absence of local deformations between the images. On the other hand, Local mapping methods are specifically designed to handle the local changes by giving preference to individual control points and regions. These methods cannot be defined succinctly and therefore represented by a large matrix, whose values represent the displacement of each pixel from moving image to fixed image.

Global mapping methods:

Often in practical applications, the deformation between images may be simple, like rotation, translation and scaling, which means that the shape is preserved between the images. The transform which preserves the shape is called *Similarity transform*. A similarity transform is an isometry composed with anisotropic or uniform scaling, given by

$$x' = s(xcos(\theta) - ysin(\theta)) + t_x \tag{2.4}$$

$$y' = s(xsin(\theta) + ycos(\theta)) + t_y \tag{2.5}$$

the homogeneous coordinate representation of similarity transform is given by,

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} s\ cos(\theta) & -s\ sin(\theta) & t_x \\ s\ sin(\theta) & s\ cos(\theta) & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{2.6}$$

compactly represented by

$$X' = s\ \mathbf{R}\ X + \mathbf{t} \tag{2.7}$$

This transform model has four degrees of freedom, where $s,\ \theta,\ t_x,\ t_y$ are the scale, rotation and translation parameters respectively. Since **R** is an orthogonal matrix, this transform preserves the angles between the lines and ratio of lengths. It requires two corresponding pairs to estimate a similarity transform. Figure 2.2 depicts an example of similarity transform.

Affine is another linear transform with six degrees of freedom, given by

$$x' = a_0 + a_1x + a_2y \tag{2.8}$$

$$y' = b_0 + b_1x + b_2y \tag{2.9}$$

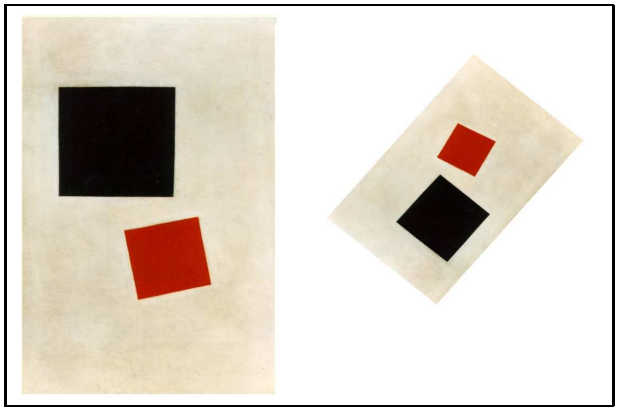


Figure 2.2: Example of similarity transform (a) Original image (b) Transformed image

matrix representation is given by

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{2.10}$$

compactly represeted by

$$X' = \mathbf{A}\, X + \mathbf{t} \tag{2.11}$$

The additional degrees of freedom come from the shear parameters incorporated in **A** matrix, it produces distortion in one axis direction proportional to the other.

$$shear_x = \mathbf{A}_{sh\ x} = \begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix} \quad shear_{sh\ y} = \mathbf{A}_y = \begin{pmatrix} 1 & 0 \\ \beta & 1 \end{pmatrix} \tag{2.12}$$

Another distortion which affine can handle is the change in aspect ratio

$$scale = \mathbf{A}_{sc} = \begin{pmatrix} s_x & 0 \\ 0 & s_y \end{pmatrix} \tag{2.13}$$

Aspect ratio is the relative scale between s_x and s_y, by scaling them independently their ratio is altered.This model requires three corresponding pairs to estimate the transformation parameters. It maps a parallelogram into square. It does not preserve angles due to non-isotropic scaling, but parallel lines remain parallel and the ratio of lengths between lines is preserved. Figure 2.3 shows an example of affine transform.It is used typically when distance between camera to scene is large as compared to the object.

The next in hierarchy of geometric transforms is a non linear transform called Projective. It given by,

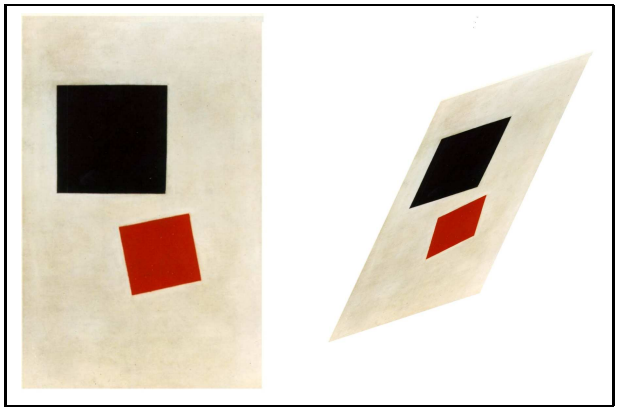


Figure 2.3: Example of Affine transform (a) Original image (b) Transformed image

$$x' = \frac{a_0 + a_1 x + a_2 y}{1 + c_1 x + c_2 y} \tag{2.14}$$

$$y' = \frac{b_0 + b_1 x + b_2 y}{1 + c_1 x + c_2 y} \tag{2.15}$$

Compact representation is given by,

$$X' = \begin{pmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{v}^T & 1 \end{pmatrix} X \tag{2.16}$$

where $\mathbf{v} = (v_1, v_2)$. The projective transform has 8 parameters and require four non collinear corresponding pairs in order to estimate the parameters. This does not preserve parallelism but the straightness of lines and planarity of surfaces are intact.It also preserves the cross ratio(ratio of ratio of lengths). This model is apt to capture a flat scene whose optical axis is not perpendicular to the camera. Example of a projective transform is shown in Figure 2.4.

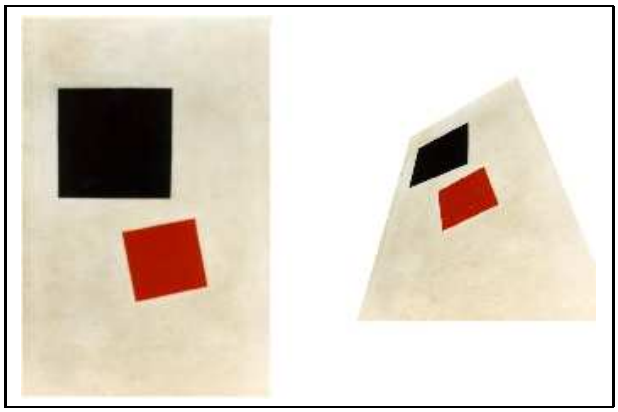


Figure 2.4: Example of Projective transform (a) Original image (b) Transformed image

Polynomial transforms are another class of transformation models which map curved lines into straight lines in the image. It will be dealt in greater detail in chapter 6.

Local mapping functions : These functions are used when the global transforms cannot handle local deformations. The weighted least squares and weighted mean methods have been proposed to register the images locally by weighting the CPs based on image data. Piece-wise linear and Piece wise cubic are two other methods which employ triangulations to map each triangle individually which accounts for the local deformations.

Under the banner of deformable registration, many methods have been proposed in literature which are specifically designed handle local deformations. Gaussian weighted,Thin-plate, Multiquadric and B spline methods belong to a popular class of global mapping methods called radial basis which have the ability to handle deformations locally. To handle more complex deformations Elastic and Fluid registration are often used in the field of medical image analysis. A few more methods of deformable registration are Diffusion based registration, Level sets based registration and optical flow based registration.

Chapter 3

Initial Work on Monomodal Retinal Image Matching

Overview: *Our initial work on retinal image registration was focused on registration of monomodal images with limited deformation and photometric variations. In this work, we illustrate a fast method for obtaining landmarks/interest points based on changes in a topographic descriptor of a retinal image. Building on the curvature primal sketch introduced by Asada and Brady [35] for describing interest points on planar curves, we extend the notion to grayscale images. We view an image as a topographic surface and propose to identify interest points on this surface using curvature as a descriptor. This is illustrated by modeling retinal vessels as trenches and identifying landmarks as points where the trench behavior changes, such as it splits or bends sharply. Based on this model, we present a method which uses the surface curvature to characterize landmark points on retinal vessels as points of high dispersion in the curvature orientation histogram computed around the points. This approach yields junction/crossover points of retinal vessels and provides a means to derive additional information about the type of junction. A scheme is developed for using such information and determining the correspondence between sets of landmarks from two images related by a rigid transformation. We conclude this chapter by reporting the limitations of this initial work and motivate the need for more robust schemes to fit practical scenarios.*

3.1 Introduction

Retinal images provide visual information to clinical experts on pathological changes, and early signs of systemic diseases like diabetes and hypertension. Changes in the retinal image over time are essential to observe and track in the diagnostic process. In automated analysis, a set of known image primitives (features) like points, lines and curves are used for finding the change/transformation. For instance in point-based registration, key points are extracted from the two images and the transformation is estimated using only the coordinates of the matched key points. Such key features are called landmarks. Landmarks are anatomically significant, visually salient, distinct features in the image that are identifiable and comparable across images. Apart from registration, landmarks are useful in several tasks including localization of disorders, surgery planning, constructing mosaics and synthesizing panoramic

views.

3.2 Background on Landmark Detection

Several methods use the bifurcation points (junctions) and crossover points of vessels as landmarks [36–38]. This is because they are meaningful landmarks, at a semantically higher level than points, lines and curves on the image. Morphological processing using revolving structural elements of T-shape has been used to locate vessel bifurcation points, after reducing the vessels to 1-pixel wide paths [37]. Changes in image gradient information have also been used to select landmark points [36]. Here, an edge-direction dispersion measure is computed in a window W around every edge pixel and its local maximas are declared as landmark points. A fixed window W will result in inconsistencies in localization of landmarks as well as their density along the tree, since the branches in the vessel tree of a retinal image are of varying thickness. Consequently, after additional processing like smoothing of histograms and pose clustering, only a subset of the landmarks are used to derive a set of corresponding landmarks from a pair of images. Motivated by the need to extract the full vessel tree accurately, a recent approach uses a multiscale approach and matched filters to trace the medial axis of thick and very thin vessels [38]. Landmark points on this vessel tree are detected at intersections of multiple traces. However, the vessel extraction step is computationally quite complex.

In point-set based registration, the need is only for detection of a set of landmark points with maximal information content. Hence, an approach that does not require accurate vessel tree extraction is of interest. In this work, we present such a method, which still guarantees the points to be located on vessel branching/crossover points and encapsulates additional information which is directly useful in matching landmarks across image pairs.

3.3 Method

Interest points have been described for planar curves, by mapping discontinuities in the curve to the local tangent-orientation space and obtaining descriptions from the new space [35]. The underlying hypothesis is that discontinuities have higher information content and are hence of interest. We propose a grayscale analogy and view a given image as a topographical surface S and examine how it changes to locate interesting points, specifically by analyzing the way S *bends* at any point. Vessel branching points and crossovers are subsets of such points of significant bend in S.

Let us consider detecting landmarks from a given color fundus image I_c. We restrict our analysis to the green plane I_g of the image which has maximum contrast. We embed the grayscale image I_g in 3-D where it can be viewed as a surface S_I, Fig.[3.1]. The intrinsic surface curvature is a suitable descriptor of S_I, assuming I_g to be twice differentiable. The curvature descriptor consists of four quantities based on the Hessian matrix of I_g [39]. These are two Eigen values a_1 and a_2, (with $a_2 \leq a_1$) and the corresponding vectors v_1 and v_2. Rules for identifying specific topographic features such as ridge, trench (inverted ridge), pit, saddle, plateau regions, are based on different combinations of the descriptor quantities [39, 40]. Vessels, being darker than the background, appear as trenches and have

been segmented using the curvature descriptor [41]. The Eigenvectors v_2 at neighboring points in a trench are oriented in the same direction, provided the trench does not bend or branch at any of these points. This observation is used to detect the vessel junction and crossover points: Consider a some neighborhood N_p around a point p in I_g. Let h_p be the curvature orientation histogram (COH) of the directions of v_2 of points in N_p. We define a dispersion measure $E(p)$ based on the *entropy* in h_p to determine the saliency of p and declare the local maxima of $E(p)$ as landmarks.

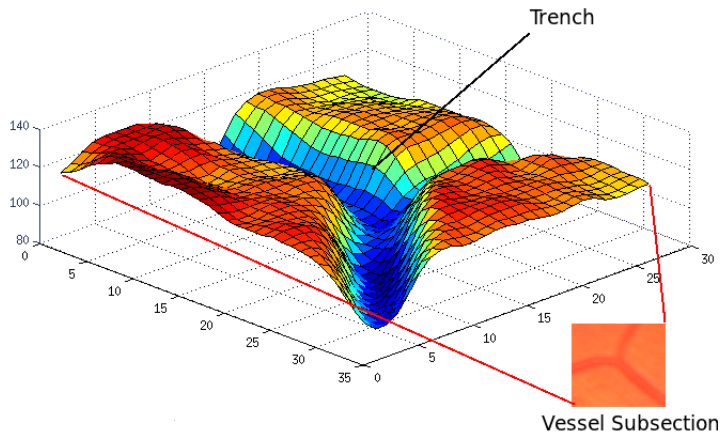


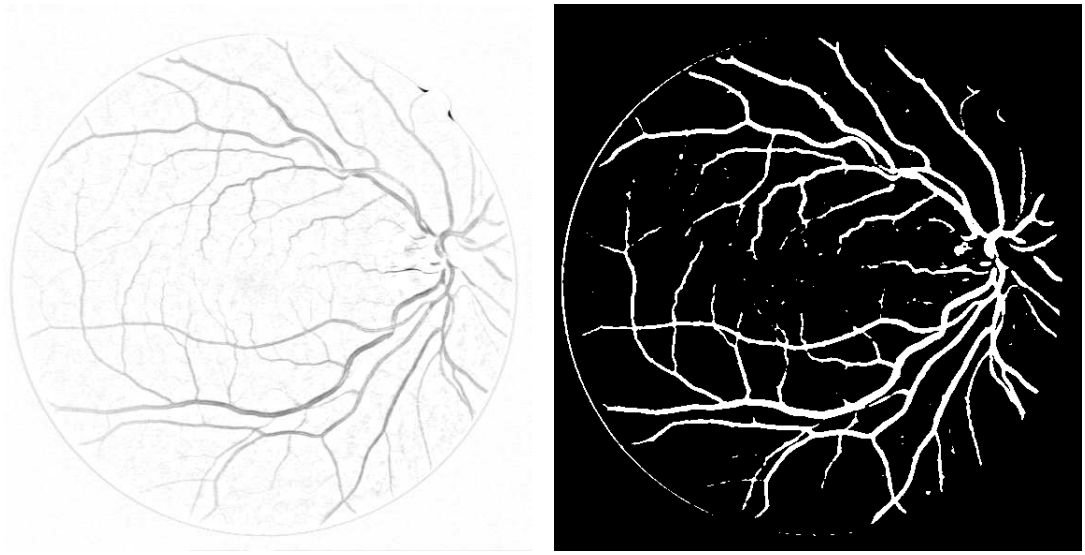
Figure 3.1: Image showing topographic surface of the subsection

3.3.1 Pre-processing

In a flat region, the curvature magnitudes are negligible, and $v_2 \approx \mathbf{0}$. Flat regions do not yield points of high information content, and hence we restrict the computation of H at only image discontinuities, specifically vessel pixels. Since extraction of all junction points is not necessary, a simple method suffices to extract vessel pixels. A background estimation [42] by median filtering I_g is performed and suppressed from I_g by subtraction, to obtain a shade corrected image I_{sc}. Blood vessels are extracted based on the fact that they have negative values in I_{sc}. They are then shifted to positive values by adding $|\min(I_{sc})|$ to get pre-processed image I_{pp} Fig.[3.2(a)]. The background has uniform high value in I_{pp}, while vessels and dark regions occupy lower intensity levels. Since, majority($\approx 90\%$) of the pixels are in the background region a binary vessel map is obtained via simple thresholding . Next, a morphological thinning followed by closing is applied to retain only the pixels on the medial axis of vessels. This vessel map is denoted as I_v Fig.[3.2(b)].

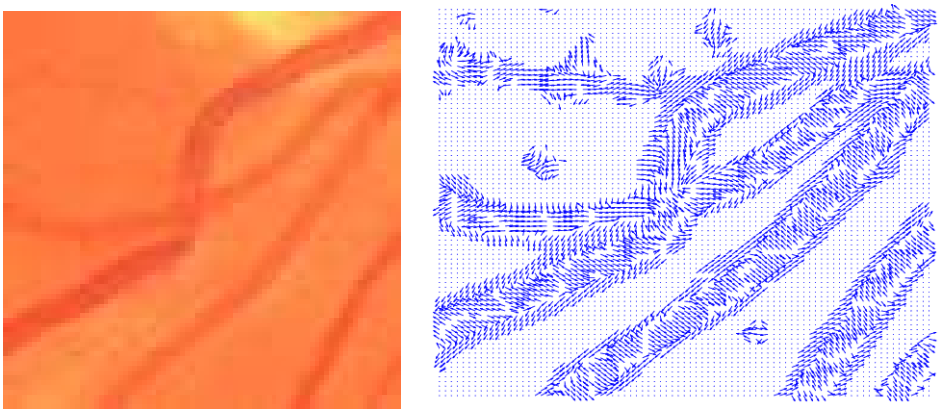
3.3.2 Dispersion measure

For every vessel pixel p in I_v, we compute orientation vectors v_2 from I_g Fig.[3.3] and then the COH h_p, by considering a neighborhood N_p on I_v. The dispersion measure E, shown in Fig.[3.4], which is



(a) Background subtracted image (b) Vessel map after thresholding

Figure 3.2: Preprocessing



(a) Image Subsection (b) Direction of v_2 on and around vessel pixels

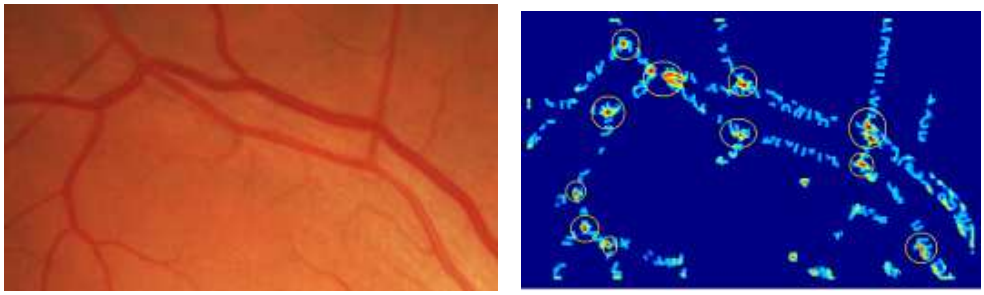
Figure 3.3: Quiver plot of Image subsection

the entropy of this COH is found as

$$E(p) = \sum_{i=1}^{n} h_p(i) \log \frac{1}{h_p(i)}. \tag{3.1}$$

Thus, corresponding to the vessel map I_v, we now have a dispersion map E. Note that at any point p the value of $E(p)$ will be high when the orientations of v_2 within N_p do not align, signifying the

presence of a junction at p, while $E(p)$ will be low in the absence of a junction as the orientations of v_2 directions are roughly along the progression of the vessel. We obtain landmarks $P = \{\hat{p}\}$ from E by applying non-maximum suppression,See Fig.[4.9].



(a) Image subsection (b) Dispersion map

Figure 3.4: Dipsersion Map showing high entropy regions in red color. It can be inferred visually that high entropy regions correspond to vessel crossover or junction points.

3.3.3 Landmark Detection Results

Images from the STARE [43] and DRIVE [44] dataset have been used to test our approach. The Hessian matrix was computed at the vessel pixels on I_g using Gaussian smoothing with a 9×9 mask ($\sigma = 1$), and 3×3 Prewitt mask. The Eigen vector v_2 was normalized and a 5×5 neighborhood (N_p) was considered for obtaining a 12-bin orientation histogram h_p (angular resolution of 15^o). The entropy map E is computed and non-maximum suppression of E is performed, with radius 12 and threshold $th = \max(E)/2$, to obtain the landmarks. The performance of the proposed method in different local context is shown in Fig.[3.6].

Fig.[3.7] shows results of our method and of two other methods: one which uses a dispersion measure on gradient vectors [36] and another which uses vessel tracing [38]. Results show that the proposed method provides landmarks that are sparser and maximally informative (as they always coincide with junctions) in comparison to the gradient-based method where they occur anywhere along the vessel. Both the proposed method and the vessel tracing-based method yield landmarks only if the vessels are detected. Branch points on some minor vessels and a T-shaped branch are not detected in Fig.[3.7(a)], because the incident vessels were not detected. Likewise in our approach, the misses depend on quality of I_v. This step can be refined, for instance by improving contrast of vessels, independent of the detection step, to increase the number of detected junctions if desired.

3.3.4 Correspondence computation scheme

The COHs that are the basis of the proposed landmark point detection, are a rich source of information. Sample subimages related by rotation and the corresponding COH pairs are shown in Fig. 3.8. We

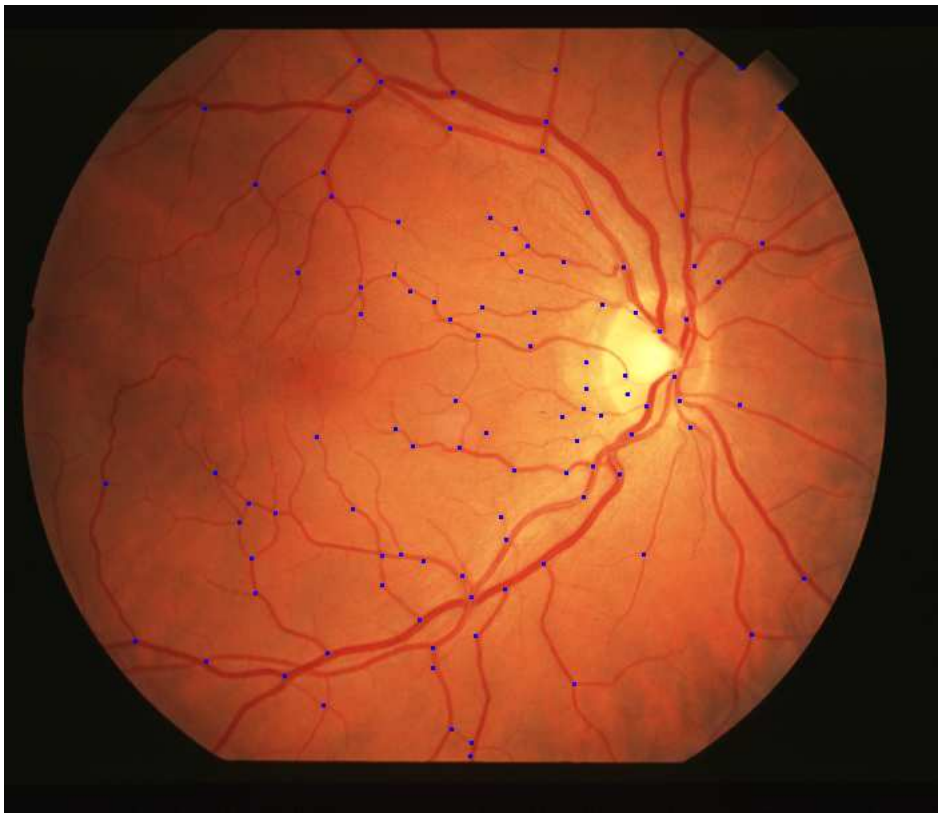


Figure 3.5: Landmark detection on stare images

now show a way to to use them to find the correspondence between two sets of landmark points. Let $P_1 = \{\hat{p}_1\}$ and $P_2 = \{\hat{p}_2\}$ be the set of landmark points computed from two images I_1 and I_2 related by an affine transformation. Let $\Phi_1 = \{h_{1i}, 1 \leq i \leq m\}$ and $\Phi_2 = \{h_{2j}, 1 \leq j \leq n\}$, $m \leq n$ be the respective COHs.

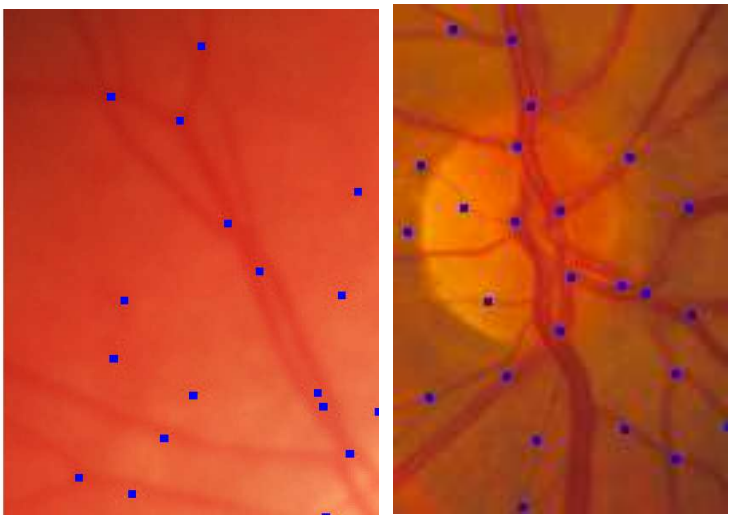
We compute a $m \times n$ distance matrix D between the histograms in Φ_1 and Φ_2, where D_{ij} is the distance between h_{1i} and h_{2j} using an appropriate histogram distance measure. Smaller values of D_{ij} signify a good match between P_{1i} and P_{2j}. An affine transformation affects the COHs in a predictable manner. They are translation invariant while rotations induce cyclic shifts. Thus, correspondence can be determined by matching cyclically shifted histograms.

For each $1 \leq i \leq m$, we find $\hat{j}$ for which $D_{i\hat{j}}$ is minimum. We then check if $D_{\alpha\hat{j}}$ is minimum at $\alpha = i$ for all $1 \leq \alpha \leq m$, to add the tuple (P_{1i}, P_{2j}) to the initial correspondence set C_0.

Different correspondence sets C_k are obtained by applying circular shifts to COHs in Φ_2 and recomputing the new distance matrix D_k, k denoting the shift size. A cost function λ_k is computed for each C_k as follows.

$$\lambda_k = \frac{\sum_{(i,j)\in C_k} D_{ij}}{|C_k|} \tag{3.2}$$

The best correspondence set is found as C_b where $b = \arg\min \lambda_k$.



(a) landmark detection on blur sub image

(b) landmark detection around optic disk

Figure 3.6: Landmark detection in different local contexts

k	0	1	2	3	4	5	6	7	8	9	10	11
λ_k	0.685	**0.242**	0.535	0.647	0.786	0.627	0.716	0.668	0.518	0.614	0.751	0.629
$\lvert C^k \rvert$	8	10	5	5	5	4	4	5	4	5	5	3
$\lvert w_k \rvert$	5	10	2	1	0	0	1	1	0	0	0	1

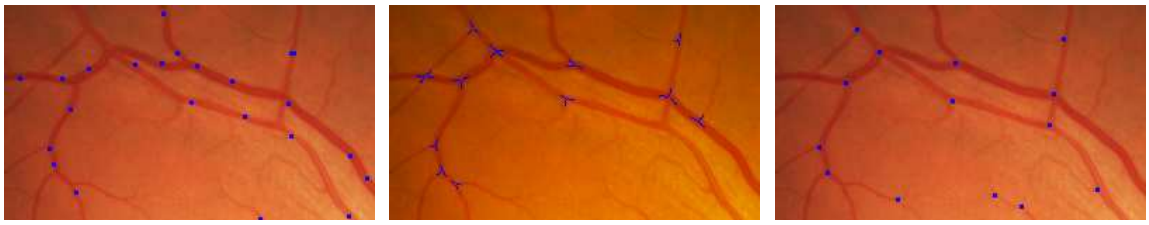
Table 3.1: λ_k values for our 12-bin histograms. $|w_k|$ denotes the number of matches visually found to be correct in C^k

A sample result of testing the above scheme using a rotation of 16^o is shown in Fig.[3.9]. The earth mover's distance [45] was used to compute D_k and ground distances were provided considering the cyclic nature of the orientation space. Table 3.1 shows the values of λ_k for 12 shifts of COHs computed over a 15×15 window. The match illustrated in Fig.[3.9] is obtained for $k = 1$ which indicates that the images are related by a rotation between 15 and 30 degrees.

From the above results, we see that the COH has potential to be used directly in correspondence computation. We have not considered scale variation, which requires parameterizing the COHs on scale (or neighborhood size).

3.4 Conclusion & Limitations

A simple approach to detect retinal landmark points on vasculature has been proposed based on entropy of the COH computed in the neighborhood of a point. Our method provides a set of sparse, yet maximally informative set of landmarks (junctions). The attractive feature of the method is that the COHs



(a) Gradient vector method (b) Vessel tracing method (c) Proposed method

Figure 3.7: Comparison of different approaches to landmark detection

implicitly capture the vessel branching information at a landmark point including the angles between them. In retinal images, this information remains invariant to rigid transformation. This is very useful in establishing correspondence between sets of landmark points obtained from images related by rigid transformations.

Though the proposed method is fast, its applications are limited. The assumption of rigid transformation severely limits the scale and view change handling capabilities of the method. The other drawback is the amount of information overlap between the images. For images with less than 40% information overlap, establishing one to one correspondence between a small set of landmarks across images may not be enough to estimate the transformation function. Also, there is no mechanism in the pipeline to prune false correspondence which may have been established between similar looking region across images. This method cannot be adapted to multimodal retinal image matching as the COH is sensitive to local information and cannot establish one to one correspondence when the information across images is complementary. We address these limitations in the next chapter and propose a method capable of registering both monomodal and multimodal retinal images.

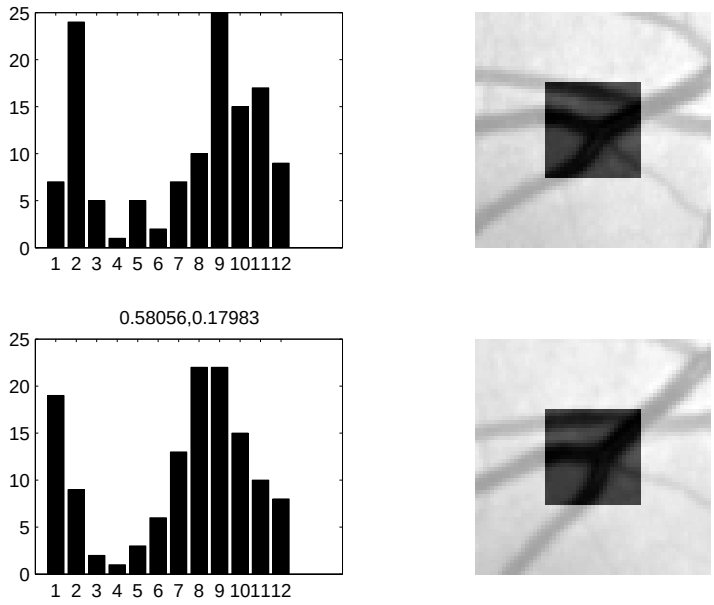
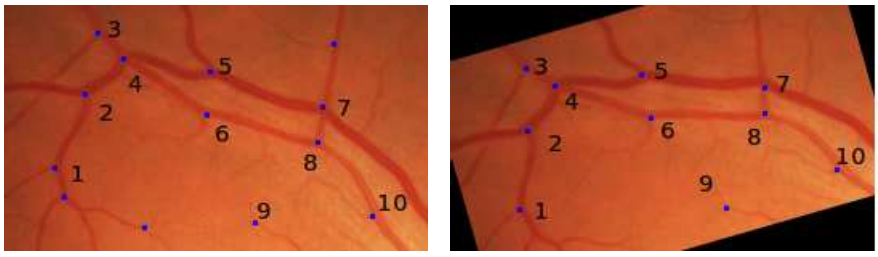


Figure 3.8: COH at a junction before and after transformation



(a) Detections on a sub-image (b) Detections after rotation by 16^o

Figure 3.9: Corresponding landmarks between an image and its transformed version

Chapter 4

A Unified Registration Framework for Monomodal and Multimodal Retinal Images

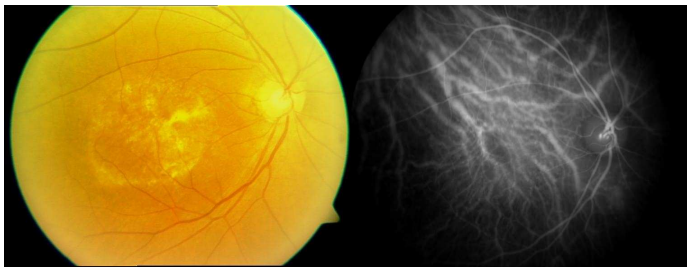
Overview: Previously proposed methods, including our own, are severely limited in its ability to register images in the presence of diverse pathologies, poor quality, low information overlap and other degradations. To address these drawbacks, we propose a robust retinal image registration algorithm capable of handling challenging monomodal and multimodal image pairs. At the core of this method is the novel feature detector and descriptor scheme. The detector is based on the extracting high curvature points on the surface of the retina using Curvature Dispersion measure. The descriptor is based on local projections using radon transform which characterizes local structures in an abstract sense. Thus, rendering it less sensitive to pathologies and noise. Drawing essence from the recent developments in robust estimation methods, a modified MSAC(M-estimators Sample and Consensus) is proposed. On the whole, the minor contributions at each stage of feature based registration scheme presented here is of significance. We evaluate our method against two recent schemes on three different datasets which includes both monomodal and multimodal images . The results show that our method is able to perform well for poor quality and pathology affected images while performing on par with the existing methods on normal images.

4.1 Introduction

Registration of multimodal retinal images aids in the diagnosis of various kinds of retinal diseases. Single modality images acquired over a period of time are used for pathology tracking. Registration is also the primary step in constructing a mosaic image of the entire retina from several narrow field images, which aids comprehensive retinal examination. Another key application area for registration is surgery, both in the planning stage and during surgery for which only optical range information is available. Fusion of these modalities also helps increase the anatomical range of visual inspection, early detection of potentially serious pathologies [46] and assess the relationship between blood flow and the diseases occurring on the surface of the retina [47].

The presence of pathologies alter the appearance of the retinal images captured via different modalities in different ways. For instance, drusen which occur in age related macular degeneration appears as yellowish blobs in CFI but are not visible in FFA. Such modality-specific impact of pathologies is a challenge for registration. Another source of difficulty is the varying quality of the images to be registered, the variability caused both by imaging conditions and patient-dependent variations. Sample multimodal image pairs are shown in Fig4.1 and Fig4.2 illustrate these variations.

Many successful methods have been proposed in the past for the registering retinal images. A review of these methods shows exceptional accuracy in terms of alignment error. However, the scope of handling a wide range of pathologies is limited for most of the approaches [48]. Further, these methods fail to register poor quality images [49]. Recently, [49] proposed a novel approach to handle poor quality cases. Though this approach successfully handles such poor quality images, the performance in terms of accuracy is restricted. Our interest lies in finding a solution for registration that is robust to pathologies and image quality changes without sacrificing accuracy. In this paper, we propose a new method that addresses these requirements. In the next section the literature specific to retinal image registration is presented.

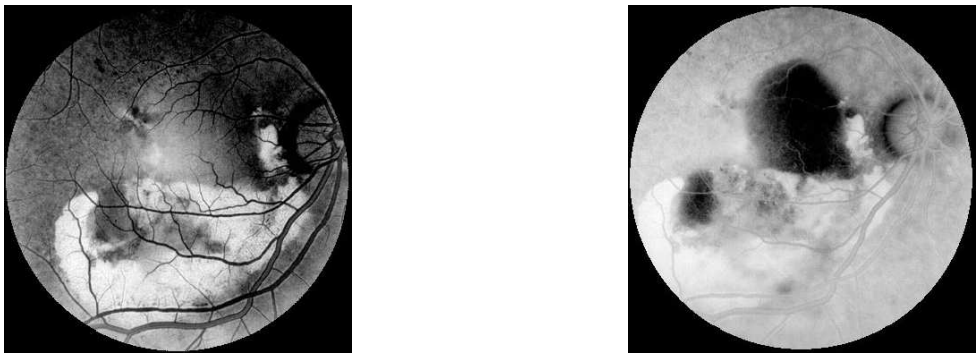


(a) (b)

Figure 4.1: Sample multimodal image pair from Dataset-I (a) Color Fundus Image and (b) Fluoroscein Fundus Angiogram

4.2 Related work

Existing approaches for retinal image registration can be classified into two broad categories: area based methods [1-6] and feature based methods[7-30]. The survey presented here is specific to retinal images. It is not he same as literature mentioned in chapter 2. Area based methods operate directly on the intensity values at a global level and choose a suitable similarity measure to drive the registration. These methods typically use an optimization framework with the objective of maximizing the similarity measure, while estimating the transformation between the images. Feature based methods, on the other hand, typically follow a three step approach - detection of significant landmarks across



(a) (b)

Figure 4.2: Sample image pair from Dataset-III (a) Redfree Image and the corresponding (b) Angiogram

images, establishing correspondence using features extracted around landmarks and the estimation of the transformation function using correspondences. Since the retina can be modeled well using global transformation models, deformable transformation models are considered unnecessary.

An early example of an area based approach is that of Metsopoulos et al [50]. They adopt a scheme based on Measure of Match similarity criterion driven by Genetic algorithms and Simulated Annealing optimization. In [51], Mutual Information (MI) based similarity measure was used with simulated annealing for aligning stereo and temporal images under rigid transformation. In general, feature based approaches have gained popularity as they are more robust to occlusion, illumination changes, partial overlap between the images as well as changes in background and viewpoint. Also, the texture-less, non-vascular regions and a non-uniform contrast across modalities degrades the performance of area based methods. The search space for area based methods increases exponentially with an increase in degrees of freedom in the transformation function, which is also undesirable.

Feature based approaches aim at establishing accurate correspondences by extracting local descriptions across images. They do not rely on the complementary information present in multimodal images and can handle variations in single modality images. Methods based on this approach may be subdivided into three classes depending on the type of feature used: (1) Intensity based (2) Vessel based and (3) Region based features. The first class of methods rely on local intensity measures like Sum of squared differences (SSD), Cross-Correlation etc. Peli et al [52] proposed a sequential similarity detection scheme using template matching over vessel junctions and bifurcations. This method is less sensitive to local contrast changes than the traditional schemes. Nagin et al [53] used edge enhancement and correlation to extract vessels followed by maximization of cross correlation. Markow et al [54] proposed a similar method, where the blood vessel templates are extracted using cross correlation and edge detection following which a correspondence is established using maximization of the cross correlation framework. Jagoe et al [55] proposed dimensionality reduction of the vessel junction feature set followed by triangulation of points to establish matching. Pinz et al [56] also performed vessel extraction, followed by affine matching of symbolic representation, using rigid transformation. These methods are

not robust to image quality degradation and pathology presence as they rely on intensity distributions to establish matches. Furthermore, they usually extract statistical information around a large fixed sized window of each landmark which leads to low similarity if pathologies are present within a window in one image. The fixed size of the window also poses a challenge when the images to be registered differ in magnification.

The second class of feature based methods rely directly on the vasculature information. A Hough transform based scheme is used in [37] to detect vessels following which a Bayesian estimation is done to find the best fit parameters of a rigid transformation. In [2], the similarity weighted matrix for all possible correspondences is computed based on the orientations of vascular centerlines and the similarity measure is converted to a prior probability. The transformation is estimated in a hierarchical manner from lower order to higher order models. Vessel junctions have also been used as landmarks in a expectation maximization framework to establish correspondences [46]. A dual-bootstrap iterative closest point (dual-bootstrap ICP) algorithm was introduced in [47]. Here, starting with one or more initial, low-order estimates of transformation (that are accurate in small image regions called bootstrap regions) an iterative refinement is done by expanding the bootstrap region. A testing phase in each iteration assesses the need for a higher order transformation model. A hybrid approach using both vessel features as well as intensity features to establish correspondence has been proposed in [57] via a hierarchical registration model. In [58] radial distortion is corrected prior to the registration and vessel junctions are used to create a composite image of the retina. A scheme based on phase correlation was proposed in [59]. Recently, a new monomodal registration framework based on graph matching has been proposed [60]. This is based on the observation that the vessel structures posses unique local signatures. STRUC-SAC estimator method is used to find matches across images.

All the above methods rely on segmented vasculature information which may be unreliable in the case of poor quality and severely pathology-affected images. These methods also tend to fail in low overlap cases due to the lack of sufficient number of landmark matches (except in [60]). There is a third class of methods which do not rely on vasculature and instead employ popular local descriptors such as Scale Invariant Feature Transform (SIFT) [32] and Speed Up Robust Features (SURF) [33]. They are designed for monomodal images and posses several desirable properties such as scale invariance, illumination invariance and noise invariance. Since they rely on gradient information they are not appropriate for multimodal medical images. An attempt to address this problem is Gradient Mirror based SIFT (GM - SIFT) [61] which is able to handle non-linear changes in intensity found in generic multimodal images. SIFT features have been used in [6] to establish correspondence followed by bundle adjustment to create a 3D metric reconstruction of the retinal surface from multiple views. In [62], Salient regions are extracted and their descriptions are used for the matching of monomodal retinal images.

A variant of [47], called the General dual bootstrap ICP was proposed in [63] for generic images, where the faces and corners extracted at multiple scales are used as landmarks. The strength of this method lies in the fact that it requires only a single accurate initial match to drive the entire registration, however in many cases it fails to do so due to the presence of a wide variety of pathologies. Chen et al [49] have derived a descriptor specifically for registration of poor quality retinal images which is called a Partially Intensity Invariant Feature Descriptor (PIIFD). Here, starting with Harris corners as landmarks, the orientation information of the local neighborhood is extracted to construct the descriptor.

PIIFD is partially invariant to affine, viewpoint and intensity changes and is reported to perform better than the SIFT. [49]and [63] do not rely on vessel information and hence perform well in low overlap cases. However, since they rely on gradient information they are not robust to pathologies.

We propose a scheme for registration of both multimodal and monomodal retinal images. The goal is to handle poor quality and pathology affected retinal images while not compromising on performance in terms accuracy, even in the case of low overlap images. The proposed scheme follows the feature descriptor based registration framework: landmark detection followed by a 2-step matching and estimation of transformation. The contributions are in the individual modules of this framework. The proposed method has been evaluated against two methods which also use a feature descriptor based approach, namely, [63] and [49]. While the former gives excellent performance in terms of accuracy, it fails on poor quality and pathology affected images. The latter is robust to quality degradation but trades off the performance. We next present the proposed method and results of its evaluation in detail.

4.3 Method

The proposed registration scheme has the following steps:

- Vessel enhancement based on a *vesselness* measure.
- Landmark detection using Curvature Dispersion Measure (CDM).
- Radon based descriptor computation for each landmark.
- Initial matching using a bilateral matching scheme.
- False match rejection and initial transformation estimation using variant of MSAC.
- Refinement and accurate localization using Normalized Cross Correlation.
- Transformation model selection and final transformation estimation using M-estimators.
- Image resampling using bicubic interpolation.

A schematic diagram of the proposed pipeline of processing for registration is given in Fig [4.3]. Before we look at the each of these modules in detail, we highlight the contributions made: (1) A normalized vesselness measure, which on the lines of scale space theory proposed by [64], has been put forth for vessel enhancement. The role of this step is to bring a given image pair into a single representation thus rendering it invariant to illumination, contrast changes and other noise factors. (2) Extraction of landmarks is based on a novel measure of curvature dispersion. (3) A Radon based descriptor is introduced for scale invariant robust matching in retinal images. This projection based local shape descriptor captures abstract higher level information thus rendering the descriptor less sensitive to lesions and noise. This helps establishing accurate correspondence even when the lesions are in the proximity of the landmarks. (4) A Variant of MSAC (M-estimator sample and consensus) is designed to reject false

matches and estimate the initial transformation to a lower order type (Affine). The resulting accurate matches are refined and localized using normalized cross correlation. (6) A novel transformation model selection scheme is introduced which exploits the information on the spatial distribution of the matches.

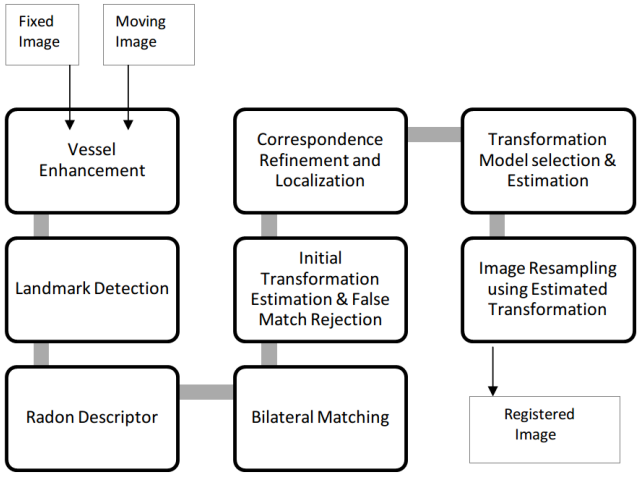


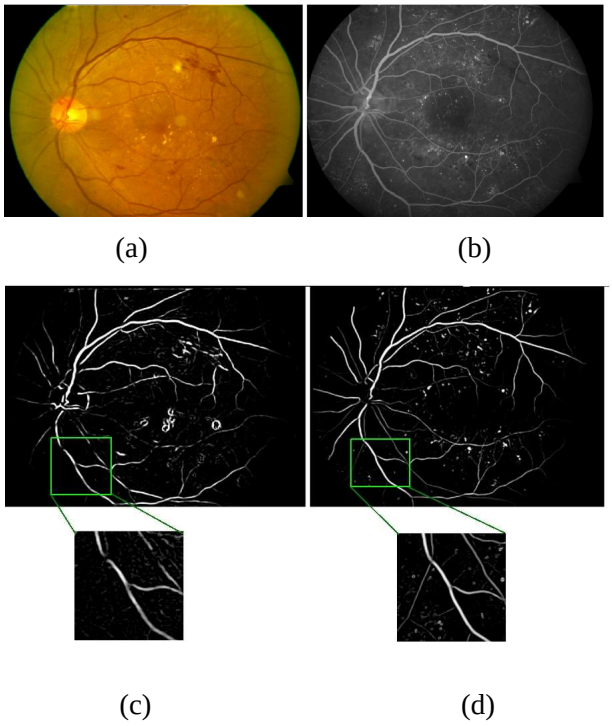
Figure 4.3: Block diagram of the proposed method

Through out this work only the green channel of the color image (CFI) is considered as it provides the maximum contrast between the vessels and the background.

4.3.1 Vessel Enhancement

Irrespective of the modality, blood vessels are the most appropriate representatives of a retinal image in the context of registration. They are well spread and anatomically significant structures which posses unique local characteristics that rarely change over time. However, extracting reliable vessel features based on its original representation (intensity) is a challenging task due to non-linear intensity difference between modalities of interest (CFI and FFA), degradations like non-uniform illumination, poor contrast and green channel noise. To address these issues, we bring a given image pair (both multimodal or monomodal images) into a single representation by enhancing vessel like structures and suppressing the background. On the lines of scale space theory, Lindeberg [64] proposed a generalized ridge strength measure which was successfully used to enhance blood vessels by modeling the vasculature as ridges. In this work, we further the generalized measure to obtain a better representation of the vasculature. We use the term *vesselness* measure in accordance with the literature [40], to refer to ridge strength in the present context.

We enhance the image as follows: Given an image *f(*$\mathbf{x}$*)*, where $\mathbf{x} = \{x, y\}$, find its scale space representation and compute vesselness measure at every point $\mathbf{x}$; maximize this measure across scales.
The image *f* and its coordinates $\mathbf{x} = \{x, y\}$, when embedded in 3D, is viewed as a topographical surface in which the vessels appear as trenches/ridges. These can be characterized based on eigen analysis of the Hessian matrix H - surface curvature descriptor[31]. The four quantities of curvature descriptors are eigen values λ_1,λ_2 (principle curvatures) where($\lambda_1 > \lambda_2$) and their respective eigen vectors υ_1, υ_2.



(a) (b)

(c) (d)

Figure 4.4: (a) Color Fundus Image (b) Fluoroscein Angiogram and (c)-(d) their respective vesselness measures. The sub-figures show a zoomed view.

In [64], the vesselness is computed as a measure of ridge strength based on the principle curvatures λ_1 and λ_2, which is given by

$$R_N(\mathbf{x},\sigma) = (\lambda_1^2(\sigma) - \lambda_2^2(\sigma))^2 \tag{4.1}$$

where R_N is referred to as the *square of the γ - normalized square of principle curvature difference*.

This measure gives strong response to elongated structures i.e the vessels, while minimizing the affect of blob like structures [64]. This important property ensures that the affect of small lesions in the retinal image are minimized. Due to the square term on principle curvatures, the dynamic range of the response is quite high. For extracting meaningful features the range of the response is constrained by computing $\sqrt[4]{R_N(\mathbf{x},\sigma)}$. This operation remaps the responses into a narrow range, but affects the structural discriminability of the extracted features.

In order to overcome this drawback, we compute the modified vesselness measure as

$$R(\mathbf{x},\sigma) = \frac{|\lambda_1^2(\sigma) - \lambda_2^2(\sigma)|}{\sqrt{\lambda_1^2(\sigma) + \lambda_2^2(\sigma)}} \tag{4.2}$$

Unlike the original formulation the entire numerator is not squared but only its absolute value is used. This ensures that the responses are not mapped into a broader range. The denominator in this expression

acts as a normalization factor, thus mapping the response into a narrow range compensating for the individual squared terms. Our experiments reveal 30% increase in the number of accurate initial matches using the modified vesselness measure. This means that the remapping of responses into a narrow range maintains good structural discriminability for the purpose of feature extraction.

The maximal response over multiple scales of the proposed vesselness measure R enhances the vessel like structures on the topography.

$$R(\mathbf{x}) \quad = \quad \text{argmax}_{\sigma}(R(\sigma)) \tag{4.3}$$

The orientation information Θ_R is also computed as the principle curvature minima direction(υ_2) corresponding to vesselness measure R on multiple scales.The vesselness representation (R , Θ_R) thus obtained is used for detecting landmarks as well as deriving local descriptions.

Through out the literature, many vesselness measures have been proposed as an intermediate step to achieve segmentation [38], [40]. These measures capture vessel structures with high discriminability which is apt for segmentation. For multimodal registration, high discriminability in vasculature yields less number of matches across images due to the complementary information across images. In this framework, we use the vesselness only to obtain an abstract structural representation of the vasculature, i.e. to boost long tubular structures and suppress the background. The goal here is to achieve an abstract vasculature representation where both the modalities are represented as closely as possible, See Fig.[4.4].

4.3.2 Landmark Detection

Landmarks are anatomically significant, visually salient, distinct features in an image that are identifiable and comparable across images. Traditional feature based registration schemes use vessel junctions and cross-over points as landmarks. However, since the amount of overlap between the images is not known apriori, using just vessel junctions or cross-over points may not yield enough common landmarks for registering an image pair. Also, due to the complementary information and contrast variations in multimodal images, these traditional landmarks(vessel junction and cross-over points) are seldom sufficient. A dense set of landmarks is needed to address this issue which is both meaningful(present on vasculature) and available in plenty.

As mentioned previously, in the context of retinal image registration, vessel structures are the most important representatives of the retinal image. Though the traditional landmarks reasonably describe how the vasculature is distributed on the surface of the retina, they are inadequate for the task of registering images with large variability and low overlap. *To characterize the vasculature better, we seek to find a dense set of landmarks which indicate subtle changes in the vessel profile, primarily how the vessel bends or changes orientation.* These landmarks are both meaningful and available in plenty. Incidentally, the vessel junctions and cross-over points would be a subset of such landmarks, See Fig.[4.9]. On the topographic surface of the retinal image, introduced in the previous chapter, these subtle changes map directly to discrepancies in the local curvature orientations [65]. The landmarks are detected by identifying points with high discrepancies using a dispersion measure over the local curvature orientation histogram, we refer to this as *Curvature Dispersion Measure*.

We want the landmarks to characterize only the vasculature, and so computing curvature dispersion measure over the entire image is unnecessary. For this we introduce a candidate selection(CS) stage before the extraction of landmarks. This stage identifies suitable points on the vasculature reducing the overhead of computing landmarks through out the image and ensures that the following important attributes are incorporated.

- *They are widely spread or distributed across the retinal image*. Wide spread ensures that there are enough landmarks for matching even if the overlap between the images is small. Also, this attribute is crucial during the initial transformation estimation stage as they form supports while fitting an affine transformation model across images, See section [4.3.5].

- *They have maximum presence on visually and anatomically significant structures such as vasculature and minimum presence on pathologies*. This ensures that the same types of landmarks are largely visible in different modalities and hence aid in establishing correspondences.

From the set of points extracted using the CS stage, the landmarks are detected by computing the curvature dispersion measure. A schematic representation is shown in fig.[4.5].

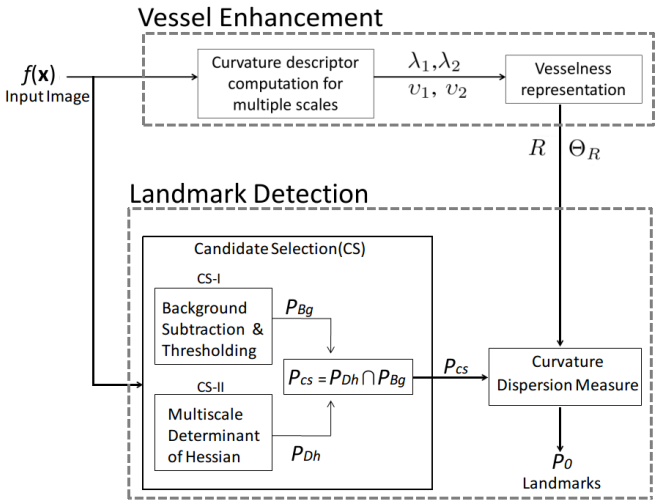


Figure 4.5: A schematic showing different stages in Vessel Enhancement and Landmark Detection.

The input to the landmark detection stage is the original image $f(\mathbf{x})$, the vessel enhanced image R and its curvature orientation map Θ_R. $f(\mathbf{x})$ is used for the candidate selection stage while R and Θ_R are used to compute the curvature dispersion measure. Next, we describe the candidate selection stage which has two modules, CS-I and CS-II.

In the first stage of candidate selection CS-I, a rough structural map P_{Bg} is to be estimated, which can be interpreted as a coarse level vessel segmentation. The goal here is to obtain a structural representation of the vasculature while suppressing blob like pathologies- Micro Aneurysms. First, the background is

estimated using a median filter with a large window size(typically 31x31). The median filter gives a coarse representation of the background and minimizes the presence of blob like structures in the image. We then subtract the background from the original image to retain the vasculature. The vessel structures have negative values in this background subtracted image f_{bg}. They are then shifted to positive values by subtracting from $max(f_{bg})$ to get pre-processed image f_s. The background has uniform low values in f_s while the vessels occupy the higher intensity levels. Since, majority (90%) of the pixels are in the background (lower range), a rough structural vessel map is obtained by extracting the top 10% of the f_s via simply thresholding. Next, morphological operations are used to reject isolated structures smaller than a fixed radius. The median filtering used in this method minimizes the influence of lesions by suppressing them as shown in fig.[4.6].

$$f_{bg}(\mathbf{x}) = f(\mathbf{x}) - medianfilter(f(\mathbf{x})). \tag{4.4}$$

$$f_s = \mid \max(f_{bg}(\mathbf{x})) - f_{bg}(\mathbf{x}) \mid \tag{4.5}$$

$$P_{Bg}(\mathbf{x}) = f_s > t_1 \tag{4.6}$$

t_1 is a threshold selected as 10% of $\max(f_s)$.

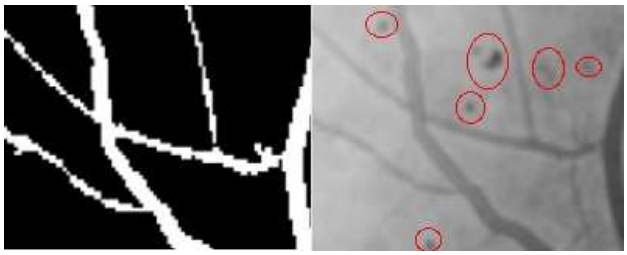


Figure 4.6: (a) shows the structural map P_{Bg} against (b) its corresponding patch from the green channel of CFI. The lesions in this image are highlighted in red.

In CS-II stage, a candidate set of high curvature points are selected using a blob detector. High curvature points on the topographic surface directly relates to discontinuities in the image which are salient. These high curvature points are a superset of the landmarks detected through the curvature dispersion measure (described next). This means that by restricting the computation of curvature dispersion measure only to high curvature points, we can further reduce the computational burden.

The Determinant of Hessian(DOH) is a well known blob detector [64], that can extract high curvature points from an image. The DOH of f is computed at multiple scales and non-maximal suppression is used to extract the desired set P_{Dh},see fig.[4.7].

$$\det H_f(\mathbf{x}, \sigma) = \sigma^2(L_{xx}L_{yy} - L_{xy}^2) = \sigma^2(\lambda_1\lambda_2) \tag{4.7}$$

where L is the gaussian convolved image and L_{aa} represents the second derivative computed in the direction of a.

$$P_{Dh}(\mathbf{x}) = \mathrm{argmaxlocal}_{(\mathbf{x},\sigma)}(\det H_f(\mathbf{x}, \sigma)). \tag{4.8}$$

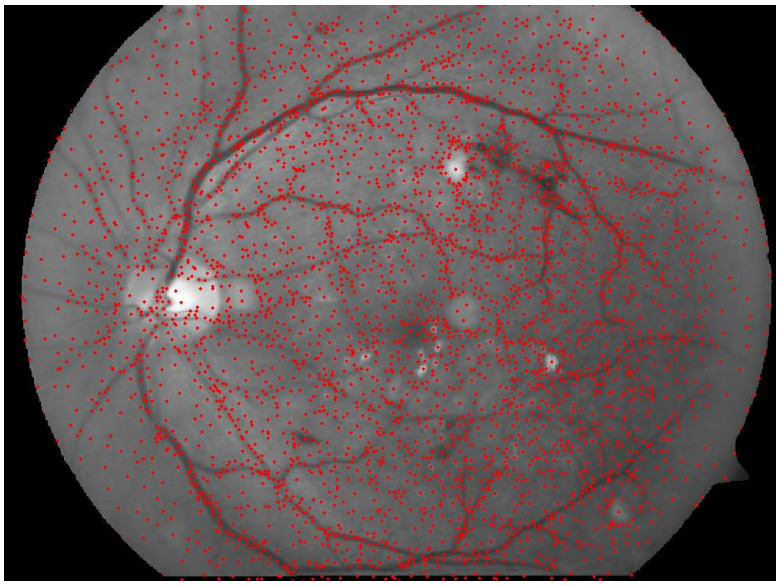


Figure 4.7: Image showing the high curvature points P_{Dh} detected by the determinant of hessian operator.

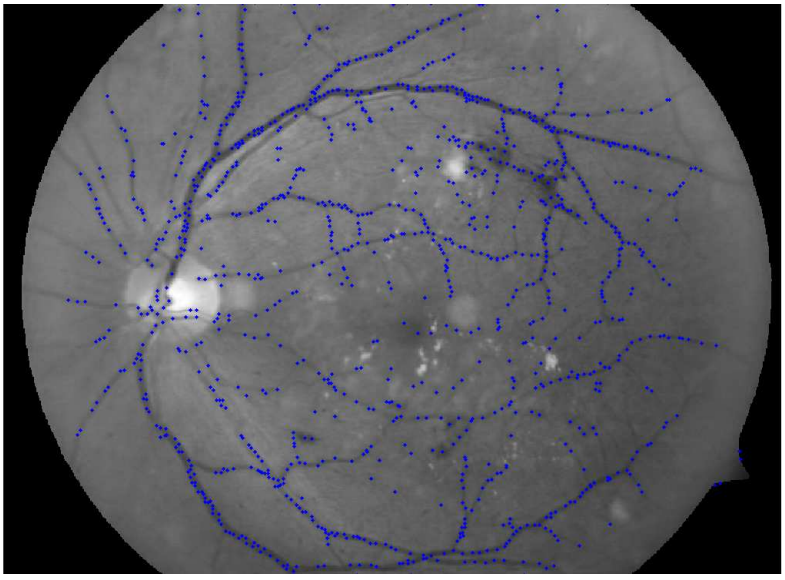


Figure 4.8: Image showing the selected candidates P_{cs} obtained though the CS stage. The candidates are maximally present on the vasculature.

A final candidate set is obtained as P_{cs}=$P_{Dh} \cap P_{Bg}$, see fig.[4.8]. P_{Dh} holds the rough structural binary map and P_{Bg} is a set of high curvature points. The candidate set P_{cs} , which are the common set of points extracted from CS-I and CS-II, ensures that the candidates are wide spread and maximally present on the vasculature while minimizing their presence on background and pathologies. Next, a set of landmarks are extracted from P_{cs} following a method we presented in [65].

When considering f as a topographic surface, the direction of the principle minima Θ_R of the vessels are oriented parallel to each other. The discrepancy in the local orientation in the neighborhood of Θ_R indicates the subtle change in vessel profile, i.e its direction and width. Based on this observation, a local dispersion measure is defined and used to obtain landmarks from a curvature orientation histogram (COH).

The method proposed here is different from the one proposed in the previous chapter, primarily in how the landmarks are defined. Here, we seek a dense set of landmarks on the vasculature and not just junctions and cross-over points. In the previous chapter, the curvature orientation histogram is constructed by simply binning the curvature orientations. But here, we modify the histogram construction by weighting each bin with the local vesselness measure r_k associated with each orientation. This modification helps characterize subtle changes in the vessel profile and not just vessel junction / cross-over points.

If $\theta \subset \Theta_R$ on the neighborhood $N_p$of every candidate point $p \in P_{cs}$, the local COH h_p is constructed as

$$h_p(\theta_k) = \left(\frac{n_{\theta_k}}{M}\right) r_k \tag{4.9}$$

Where θ_k refers to θ binned into k bins, n_{θ_k} is the no. of pixels with respect to the orientation θ_k, M the total no. of pixels in the neighborhood N_p and r_k is sum of the local vesselness measure corresponding to θ_k .

The dispersion measure map E is the entropy of the COH given by

$$E(p) = \sum_{k=1}^{n} h_p(\theta_k) \log \frac{1}{h_p(\theta_k)} \tag{4.10}$$

The value of $E(p)$ will be high when the vesselness R is significant and the orientations of Θ_R within N_p do not align, indicating the change in vessel profile which are our desired landmarks. We derive the desired set of landmarks P_0 from $E(p)$ through non-maximal suppression.

The above method yields well localized landmarks characterized by the discrepancies in the local curvature orientations. The candidate selection step plays a critical role in both localization of landmarks as well as suppressing candidates from pathology affected and homogeneous regions. The CS-II stage helps in localization while the CS-I stage helps minimize the effect of pathology affected areas and other homogeneous regions on landmark detection. Overall the landmarks obtained though this process are maximimally present on the vasculature that correspond to the changes in the local vessel orientation, which in turn implicitly indicates the presence of locally unique vessel profile. This observation is consistent with the results shown in Fig.[4.9]. In comparison, the Harris corner detector used in [49] is sensitive to pathology affected and non vascular areas [4.10]. Hence, a majority of these points may not have consistent matches across modalities.

4.3.3 Radon based Descriptor

Computation of a descriptor in general can be seen as an attempt to represent a signal or image data in a compact format while retaining relevant information. In addition to compact representation, it is

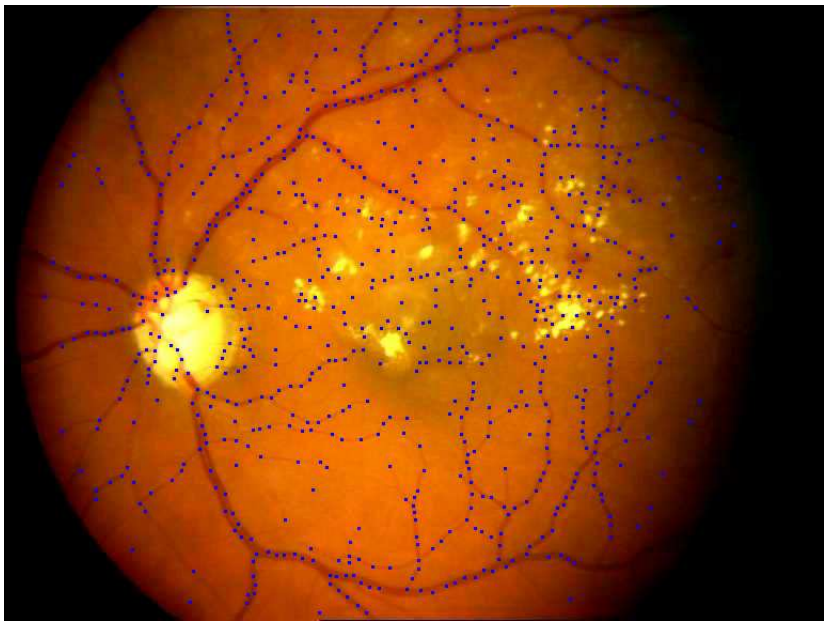


Figure 4.9: Detected landmarks, using the Curvature Dispersion measure.

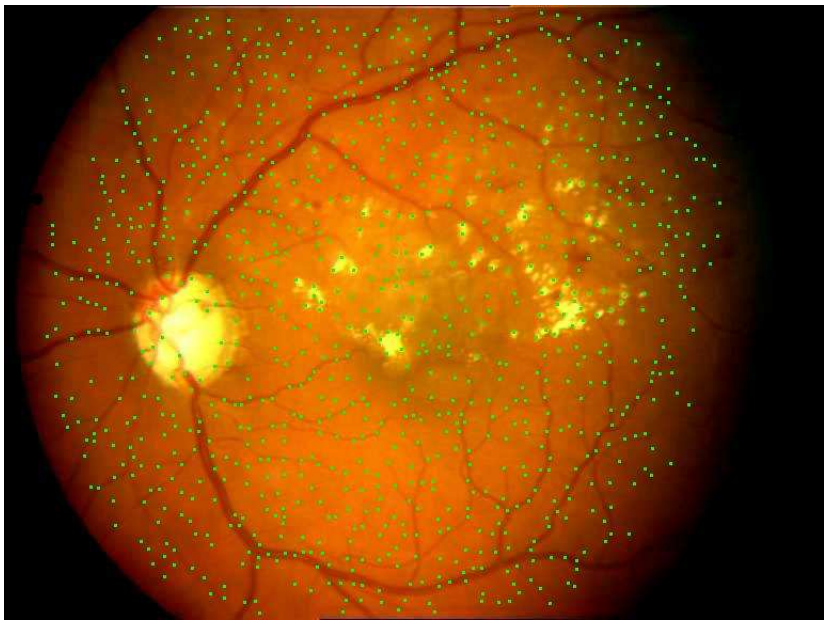


Figure 4.10: Results of Harris corner detector on CFI.

also important for a descriptor to be robust to geometric and photometric changes that occur across images. In the case of retinal images, we seek to find a compact representation which is less sensitive to pathologies while still holding enough discriminability to establish correspondence across multimodal pairs. Even if such a representation is obtained, it is useless until the notion of scale is addressed. By scale we refer to the size of the window used for each landmark to extract the descriptor. We provide a

novel solution to handle scale changes across multimodal retinal images. Next, we motivate and propose a new robust descriptor for retinal image matching.

Gradient based descriptors for retinal landmarks have been popular [61] [49]. However, they are adversely affected by the presence of lesions in the neighborhood of a landmark [48]. Also the window size is influenced by the scale at which the landmarks are detected and in multimodal retinal images two similar points/regions may not correspond to the same scale. Hence, we propose a projection based method termed the Radon descriptor (RD) to address these issues. The radon descriptor represents information around each landmark point in a compact form while being uninfluenced by pathologies. Instead of relying on scale to determine the windows size, for each landmark, we compute multiple descriptors for varying window sizes. This approach of having multiple descriptions for a single landmark point is computationally feasible because the descriptor is constructed using simple projections.

Radon transform , a well known shape descriptor [66], can be used to capture the local structures around each landmark point. The Radon transform of a function $f(x, y)$ denoted by $g(s, \theta)$ is given by

$$g(s,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta\left(xcos\theta + ysin\theta - s\right)dxdy, \tag{4.11}$$

with $-\infty < s < \infty$, $0 \leq \theta < \pi$. where s is the distance from the origin and θ is the angle of projection. Rotating a function by an angle θ_0 results in a shift in the projection (Radon) domain by θ_0.

$$\tilde{R}f_\rho(r, \phi + \theta_0) = g(s, \theta + \theta_0) \tag{4.12}$$

where $\tilde{R}$ is the transform operator and $f_\rho(r, \theta)$, is the representation of the function f in polar co-ordinate frame. Since we wish to derive a descriptor that is rotation invariant, the axis of the RT is aligned to the dominant local direction at a landmark point prior to the descriptor derivation. We next explain this in detail.

The inputs to this stage are the landmarks P_0 and the vesselness representation R and Θ_R. For each landmark point, a patch of size WxW is extracted on the vessel enhanced image R. The principle minima directionsΘ_R within the patch are binned into 18 equally spaced bins. The bin corresponding to $max(\Theta_R)$ is taken as the dominant orientation. The patch is then rotated to align to this dominant direction. Starting with the dominant orientation, the Radon transform (projection) is computed for uniform angular intervals. The individual projection profiles are normalized and appended as the feature descriptor for the given landmark. The resulting Radon descriptor RD_w (w refers to the window size) at location p_i is given as

$$RD_w(p_i) = \{g_{\theta_1}\ g_{\theta_2}\ g_{\theta_3}\ .....g_{\theta_n}\} \tag{4.13}$$

Here, n represents the angular resolution and its choice is a critical factor. If n is high, the descriptor can be too fine grained and hence lose out on robustness. Thus it may not allow for matching across regions between multimodal images. On the other hand, a low n can lead to the descriptor not capturing the local structure well, losing discriminative power. The balance lies in the selection of n, for all the experiments presented in this work we use $n = 12$.

The next important step is to address the issue of window size. For each landmark window size refers to the size of the local neighborhood around which the descriptor is constructed. In general computer

vision techniques like SIFT [32] and SURF [33] the scale is determined at the landmark detection stage. The characteristic scale σ_c of every landmark is determined through the principles of scale space theory proposed by [64] and the descriptor is constructed from a windows size of 6xσ_c around each landmark.

This method of scale selection cannot be employed for multimodal images. The characteristic scale of the two corresponding landmarks between images may not be the same due to the change in modality. To address this issue, for each landmark we compute multiple descriptors for varying window sizes and perform matching as if they belong to different landmarks. This means that we handle local scale differences at the descriptor level instead of the detector level. The descriptors are computed for multiple window sizes as follows.

$$RD(p_i) = \{RD_{w1}, RD_{w2}, RD_{w3}.....RD_{wm}\} \tag{4.14}$$

where $w1 > w2 > w3 > wm.$ for each landmark p_i, m descriptors of varying window sizes are extracted. The descriptors thus obtained are all is resized to 720 using bilinear interpolation.

4.3.4 Matching based on Descriptors

After the computation of descriptors, the next important stage for feature based registration is to obtain correspondence between the two images using appropriate matching strategy. In our approach, Bilateral matching technique [49] is used to ensure one to one correspondence. For two sets of landmarks P_0 and Q_0 with descriptors D_p and D_q, respectively, a set of correspondence C_{M1} is obtained by finding the best matches by minimizing the Euclidean distance between the descriptors in the fixed image D_p to the moving image D_q. Similarly, a second set C_{M2} is obtained by starting from the moving image to fixed image. The final correspondence is given by $C_1 = C_{M1} \cap C_{M2}$. If matching is only one way, one to one correspondence cannot be guaranteed.

The nearest neighbor in the descriptor space is obtained based on Euclidean distance (in $[0, \infty)$, with 0 closest) between points. This distance can be converted into a similarity measure (in $[0, 1]$, with 1 closest) by a monotonic decreasing function. In this paper, we choose the following similarity measure known as Euclidean-normalized similarity [67] given by

$$Similarity(\mathbf{D}_p, \mathbf{D}_q) = e^{-\|\mathbf{D}_p - \mathbf{D}_q\|_2^2} \tag{4.15}$$

Given the overhead of computing multiple descriptors for each landmark, the distance can be computed in a single operation as

$$\|\mathbf{D}_p - \mathbf{D}_q\| = \sqrt{\|\mathbf{D}_p\|^2 + \|\mathbf{D}_q\|^2 - 2\mathbf{D}_p\mathbf{D}_q} \tag{4.16}$$

The result of bilateral matching on a sample image pair is shown in Fig. [4.11].

We restrict the number of initial correspondences C_1 to N based on the similarity measure, where $N = 300$.

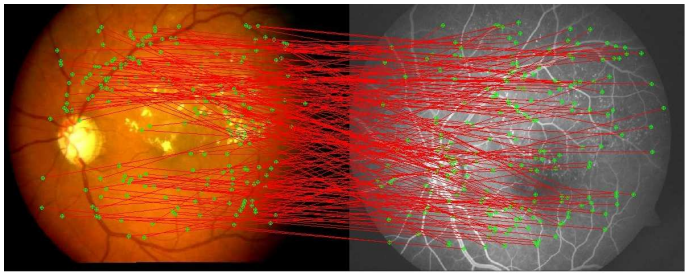


Figure 4.11: Result of Bilateral Matching showing Correspondence set C_1

4.3.5 Initial transformation estimation and outlier rejection using modified MSAC

After the initial matching step, we now have C_1 set of correspondences which are much more than the Degrees of freedom(DOF) in the transformation model. The transformation models, described in the later sections, like second order polynomial transform have DOF-12 , bilinear Transform have DOF-8. With the increase in the DOF, the predictability outside the region of fitting decreases [46]. To overcome this behavior we restrict the initial transformation model to a lower order one, i.e Affine. After outlier rejection and further localization of landmarks, the higher order transformations are estimated as the final transformation. This assumption is valid as the perspective effects in the images are minor , i.e the retina being imaged is roughly parallel to the image plane of the camera [2].

Initial matches established using any feature based descriptor method do not guarantee 100% accurate correspondence as matching is performed around a small neighborhoods around landmarks, the false matches are called outliers. In computer vision, RANSAC(Random Sample and Consensus) is a popular parameter estimation technique if the data is corrupted by outliers. RANSAC is a Hypothesis and Verify scheme, which generates the hypothesis from a small sample set and validates its consistency across the entire set using a cost function. The hypothesis which minimizes the overall cost function of the entire set yields the optimal parameters. Though, the method is general, the modifications we propose are of significance.

Hypothesis Generation

In RANSAC, from the initial set C_1 with N number of correspondences, a Minimal Sample Set(MSS) is randomly selected and the model parameters are estimated only from the MSS. The MSS is defined by the minimum number of correspondences required to estimate the parameters. For affine case the cardinality of MSS is 3. The selection of MSS are random, which means, all the correspondences are given equal importance. In the current scenario, the matching scores computed from bilateral matching is a reasonable measure to grade the correspondence according to their quality. So instead of randomly selecting points from the entire correspondence set, we restrict the hypothesis generation to a few correspondences which have highest matching scores, but use the entire set to evaluate the cost function. By this we assume that the correspondences with high matching scores are less likely to be contaminated. So this can be a called semi randomized algorithm. The number of top quality correspondences T_n are

selected as

$$T_N = H(N) = \begin{cases} N & \text{if } N \leq \vartheta \\ \vartheta + e^{(0.01N)} & \text{if } N < \vartheta \end{cases} \tag{4.17}$$

where ϑ is constant, which is set to 40. The function $H(\cdot)$ exhibits a linear behavior until ϑ and transforms into a monotonically increasing function beyond it. If C_{T_N} represents the top quality match set with minimum outlier percentage, three corresponding pairs are selected at random from this set to create M_0. Given the fact that the perspective effect in retinal images is minor, additional constrains can be imposed on such selection even before the hypothesis is generated. From any three randomly selected corresponding pairs, a triangle can be constructed in both moving and fixed image. If the points roughly correspond to each other then the two triangle are similar. This can be easily computed using the Angle-Angle Similar triangle criteria with a minimum overhead. All the points which obey this criteria are pooled into M_1 which is our final minimal sample set.

Verification

Verification is the process of estimating the goodness of fit of the generated hypothesis to the entire sample set. The elements consistent with the generated hypothesis are called the Consensus Set(CS). The goodness of fit is expressed in terms of cost function which is to be minimized/maximized over the entire dataset. The original RANSAC framework uses the cardinality of the consensus set as the cost function to be maximized. [68] provided an alternative cost function on the lines of M-estimators,and dubbed this method as MSAC(M-estimator Sample and Consensus). In this work, we adopt this cost function as the criteria to be minimized. The cost function CF is given by

$$CF = \textstyle\sum_i \rho(e_i^2) \quad \text{where } i = 1...n \tag{4.18}$$

Where e_i is the error on individual points and $\rho()$ is

$$\rho(e^2) = \begin{cases} e^2 & \text{if } e^2 \leq \delta \\ \delta & \text{otherwise} \end{cases} \tag{4.19}$$

Assuming that the elements are affected by gaussian noise of σ_n, δ can be computed as

$$\delta = \sigma_n \sqrt{F^{-1}_{\chi^2_n}(Pr(inliers))} \tag{4.20}$$

where $F^{-1}_{\chi^2_n}$ is the inverse cumulative distribution associated with chi-squared distribution of the random variable $\sum_i^n (\frac{e_i}{\sigma_n})^2$ and $Pr(inliers)$ is the probability of inliers. See [69] for implementation details.

The hypothesis and verification steps are iterated consequently until the stopping criterion is achieved. The stopping criterion is given by

$$\hat{S_{iter}} = \left[\frac{log\varepsilon}{log(1-q)} \right] \tag{4.21}$$

where $(1-q)$ is the probability of picking an MSS from M_1 with at least a single outlier and ε is the threshold.

The new correspondence set is C_2 and the affine parameters estimated as A. The moving image and its Vessel enhanced image are transformed into a new coordinate system based on the estimated parameter A. This transformation process is necessary for the accurate localization of landmarks in the next step. As a consequence of this operation, the scale difference between moving and fixed image is removed.

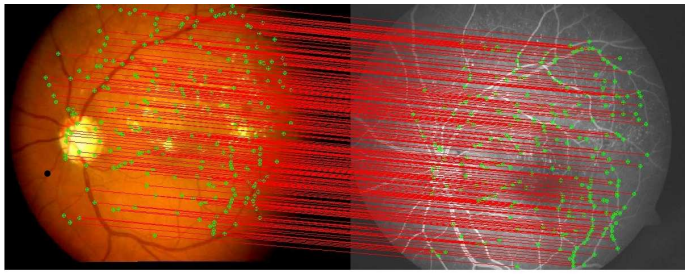


Figure 4.12: True correspondence set C_2 after outlier rejection.

4.3.6 Refinement & localization of correspondences

Given the accurate correspondence set C_2 and the estimated initial(affine) transformation A, the refinement step extracts all the possible correspondences from the original landmark set P_0 and Q_0 which fit the estimated transformation parameters. These include landmarks which failed to find matches across images at the initial matching stage.

$$q_i = \mathrm{argmin}_{\mathrm{q}\in\mathrm{C_2(p_i)}} \|q - Ap_i\| \tag{4.22}$$

$$C_3 = \|q_i - pi\| \leq dist_1 \tag{4.23}$$

The set of new correspondences are given by C_3 in $P_3 = \{p_i\} \in P_0$ and $Q_3 = \{q_i\} \in Q_0$. $dist_1$ is a distance threshold, here $dist_1 = 4$.

The estimation of higher order transformations are sensitive to localization errors. To minimize the error in the final transformation estimation step, it is necessary to localize the correspondence set C_2. We use normalized cross correlation (NCC) in the local neighborhood of each landmark in C_2 to achieve localization. NCC is computed over the vessel enhanced image instead of the intensity, which is given by.

$$NCC(u,v) = \frac{1}{n-1} \sum_{x,y} \frac{(p(x,y) - \overline{p})(q(x-u, y-v) - \overline{q})}{\sigma_p \sigma_q} \tag{4.24}$$

$$q_{\hat{i}} = \mathrm{argmax}_{\mathrm{u,v}}\, NCC(u,v) \tag{4.25}$$

where p & q are the corresponding landmarks. $\overline{p}$ & $\overline{q}$ are the mean and u, v represent a local neighborhood around each landmark. n is the number of pixels in u and v.

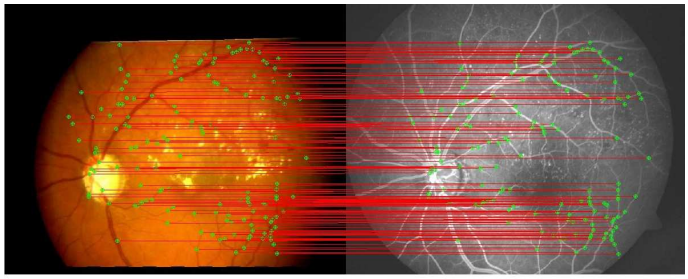


Figure 4.13: Result after refinement and localization.

4.3.7 Model selection and transformation estimation

[2] has shown that a second order transformation is the most appropriate model to register retinal images given their curved nature.

$$\begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} & \theta_{14} & \theta_{15} & \theta_{16} \\ \theta_{21} & \theta_{22} & \theta_{23} & \theta_{24} & \theta_{25} & \theta_{26} \end{bmatrix} \begin{bmatrix} x^2 \\ xy \\ y^2 \\ x \\ y \\ 1 \end{bmatrix} \tag{4.26}$$

The initial criteria for the model selection are the no. of correspondences in set C_3, i.e $|C_3|$. If the number of correspondences is less than the parameters to be estimated a lower order transformation model is selected. The lower order transformations are affine and bilinear transforms. The bilinear transform is given by

$$\begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} & \theta_{14} \\ \theta_{21} & \theta_{22} & \theta_{23} & \theta_{24} \end{bmatrix} \begin{bmatrix} xy \\ x \\ y \\ 1 \end{bmatrix} \tag{4.27}$$

which is

$$\mathbf{X'} = \phi\mathbf{X} \tag{4.28}$$

C_3 holds the best set of localized correspondences that is used to estimate the final transformation. The spatial spread of these correspondences is a vital factor in estimating the final transformation. Even if the no. of correspondences are greater than the number of parameters to be estimated, in few cases(especially if the information overlap between the images is low) dense set of correspondences are established in small regions. The lack of well spread correspondences leads to registration error. This is called correspondence cluttering. In such cases, it would be appropriate to fall back to lower order transformations to minimize the overall registration error.

To select the appropriate transformation model, we propose a spatial dispersion measure which estimates the spread of these established correspondences. If $X \in C_3(P_3)$ and $\bar{X}$ is the centroid of the

correspondences. The euclidean distances between the $\bar{X}$ and X are binned into 18 equally spaced histograms. The entropy score of this histogram gives the spatial dispersion of the correspondences. Based on the spatial dispersion of correspondences and the number of correspondences the transformation model is selected.

Given the model, we now estimate the transformation. The transformation estimation using least squares is sensitive to errors in the localization of correspondences. This issue is addressed by using robust estimation techniques. One of the most popular class of these estimators are M-estimators. The general form for parameter estimation using M-estimators is given by

$$\hat{\phi} = \text{argmin}_\theta \left(\sum_{i=1}^{n} \rho(\|\mathbf{X}' - \phi\mathbf{X}\|\backslash\sigma) \right) \tag{4.29}$$

where ρ is the Bi-weighted loss function [2] and σ is the scale estimate. Since no closed form solution exists for this loss function, it is implemented using Iteratively re-weighted Least squares as given in [2]. After the estimation of the final transformation, the moving image is resampled into the coordinates of the fixed image using bicubic interpolation.

4.4 Discussion & Results

We evaluate the performance of our method on three different datasets. Dataset I consists of 126 multimodal image pairs(CFI and FFA), acquired from same number of patients. Dataset II has 20 monomodal image pairs(CFI images) and Dataset III consists of 18 challenging image pairs(both monomodal and multimodal which includes CFI, FFA and Redfree images) collected from various internet sources. These datasets have a wide variety of pathologies and the images exhibit various acquisition artifacts like non-uniform illumination, motion blur etc. The resolution varies from 256x256 to 1204x1200 and angular resolution between 30°-50°. The lowest overlap case in the dataset is 30% and the highest rotation angle is 25°. The images have been obtained from Zeiss fundus camera.

We evaluate the proposed method against two other methods: GDBICP [70] and PIIFD [49]. GDBICP was chosen for comparison as out it outperforms most of the existing algorithms and has become a standard for retinal image registration. And PIIFD was selected as it is the only method which is specifically tailored to handle poor quality images. We evaluate our method against these two schemes on three different datasets which includes both monomodal and multimodal images . The results show that our method is able to perform well for poor quality and pathology affected images while performing on par with the existing methods on normal images.

4.4.1 Implementation Details and Parameter Settings

The proposed method has been implemented in Matlab 7.11.0 on AMD 64x processor. GDBICP algorithm is available as an executable file written in C++ at [70]. The experiments are validated in "-complete" mode, which enables it to register difficult pairs. Another optional parameter is the transformation model, all the given models have been tried exhaustively from higher to lower order models only if registration failed. Matlab code for PIIFD has been obtained from the authors and the parameters

have been set as per the author's guidelines. For the purpose of fair evaluation, the number of interest points detected by both the methods are made approximately equal(600-800 per image). In [49] it has been shown that PIIFD performs better than SIFT and so it has not been included in our validation process.

For each image pair in Dataset-I, the proposed method takes about 70-80 Sec where as the PIIFD takes 40-45 Sec. The increase in time is due to the landmark detection stage and a more elaborate initial transformation estimation scheme incorporated in our framework. It has been observed that the Harris corners detected in PIIFD are well spread across the images while the point features detected by our method are confined to vessel structures. However the corners are poorly localized across images if the scale difference between the images is $>$1.3. In next stage, vessels are enhanced using multiscale hessian computation over 5 scales $\sigma = \{1, 1.5, 2, 2.5, 3\}$. The window size for the scale invariant computation of the radon descriptor is $w = \{41, 49, 51, 71, 85\}$ and the relation between the window sizes is $w_{i+1} = 1.2w_i$. The value 1.2 was experimentally determined based on the the discriminability of the descriptor with increase in the scale.

4.4.2 Evaluation

Appropriate evaluation methodology is critical for unbiased and accurate qualification of the performance of various registration schemes. Unfortunately for fundus images no benchmark dataset or ground truth exist. This has prompted researchers to develop alternative comparison schemes, one such popular scheme is Centerline Measurement Error(CME). The vessel centerlines are traced for the pair of unregistered images and the centerline is densely sampled. The sampled points are then transformed from the moving image to the fixed image based on the estimated transform. The euclidean distance between the closest traced point in the fixed image and the transformed points in the registered image is taken as the sample error. The median of all the sample errors is the final Centerline measurement error(CME). We use a similar strategy to validate the performance but report on the lines of mean centerline measurement error(M-CME), here the vessels are traced on the registered pair instead of the original images. This modification considers the final resampling step in section[4.3.6] also for validation purpose. The Vessel tracer is available as an executable at [70]. The CME criteria is used to evaluate the performance of Dataset I&II. Since detecting the vessel centerline of the poor quality images in dataset III is a challenging task, the performance is validated through visual inspection.

Based on the standard validation process of retinal registration schemes [49], we examine our method on the lines of rotation invariance, scale handling capabilities and overlap criteria. In all the experiments presented below, the registration is considered to have failed if the CME error is above 0.96. For all practical purposes, the failed cases are mapped to CME error of 1.

Rotational Invariance test

The goal of this test is to examine the rotational invariance capabilities of the proposed descriptor. The M-CME error is calculated between moving image and fixed image by rotating the moving image from 0-180 at 10°, between 10 multimodal pairs. The results show M-CME error is almost constant($\pm.18$) for all the rotated angles of the moving image, demonstrating the rotational invariance of the proposed

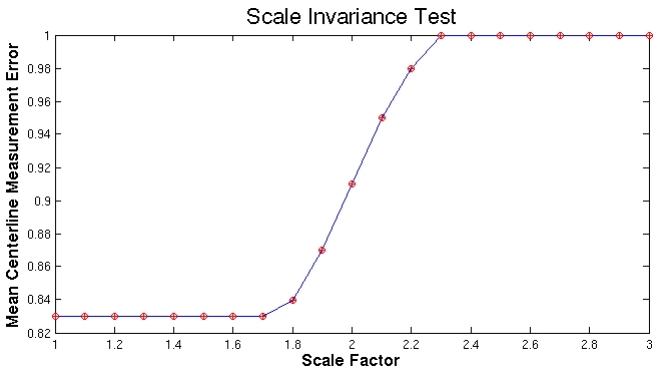


Figure 4.14: Scale test: Mean centerline measurement error relative to the scale factor.

descriptor. The slight variation in M-CME error is due to the random nature of the initial transformation estimation step.

Scale Invariance test

To demonstrate the scale invariance of the descriptor, M-CME is computed on the registered pair while the scale of the moving image is increased from 1 to 2.5 in steps of 0.1. The results show that our method is scale invariant up to scale factor of 1.6. In clinical practices the scale differences are below 1.5 [22], so the invariance factor is adequate.

Overlap Test

Overlap is the percentage of common area between the registered pair. Five pairs with varying percentage of overlap area has been used to validate this test. In all three datasets, the minimum percentage of overlap is 35%. Images of varying degrees of overlap are created from 10-20 percent using GIMP [71]. The results show that the M-CME error quickly falls after 30%. PIIFD has been shown to perform well till 20%, it has been observed that the number of corners detected are well spread across the image which helps cope with the low percentage of overlap if the angular resolution of the image remains the same.

4.4.3 General Discussion

In our previous publication [48], after the computation of radon descriptor, a FFT (Fast Fourier Transform) step was included in the descriptor computation which was used for better discriminability and as a dimentionality reduction step. The FFT step is excluded in the present study as more number of initial matches are obtained per image without it. Though the present descriptor has less discriminating capabilities without the FFT step, after the initial matching step, the robust parameter estimation technique can be relied upon to reject the false matches. The key would be to balance the discriminability of the descriptor between corresponding and non corresponding pairs by the selecting appropriate number of

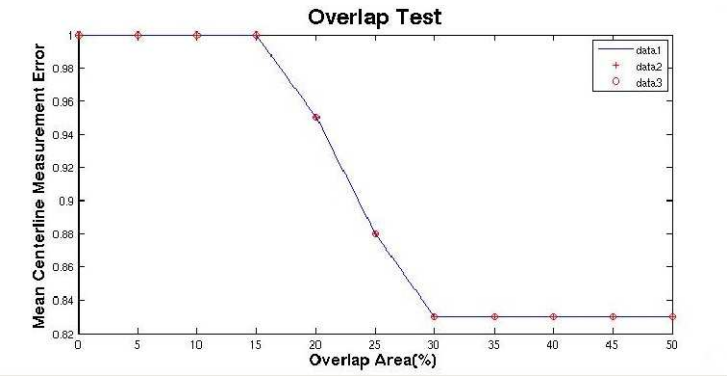


Figure 4.15: Overlap test: Mean centerline measurement error relative to the percentage of overlap.

angular resolutions. The length of the PIIFD descriptor is 128 and radon transform yields a descriptor of length 720.

On the lines of PCA-SIFT [72] a dimensionality reduction step for the proposed descriptor was studied. About 500 corresponding image patches were manually picked for PCA based dimensionality, the method is similar to eigen faces [73]. As a result, the number of accurate correspondences decreased by 30%. This tends to increase registration error in poor quality, pathology effected and low overlap cases. In the matching step, a two way matching technique has been adopted which avoids the consequence of hard thresholding based on similarity, thus identifying unique matches even with low similarity, especially in the presence of lesions.

One of the drawbacks of our method is that it cannot handle contrast reversal changes in images. In general retinal images, the vessels are either dark on bright background or vice-versa. In very rare cases, see Fig.[4.19], the vessels are both bright and dark at the same time and the registration fails. But in practical scenarios this is of little interest.

4.4.4 Results

Based on the M-CME criteria described in the above section, overall performance on two datasets is given below.

Table 4.1: Mean Centerline Measurement Error

Dataset	GDBICP	PIIFD#1	our Method
Dataset-I(126 pairs)	0.845	0.956	0.853 ± 1.8
Dataset-II(20 pairs)	0.823	0.906	0.84 ± 1.8

The success of registration can be categorized into two categories: accurate and acceptable. Pairs registered with 0.90 M-CME error are categorized as accurate and acceptable ones are between 0.9-0.96. Errors above 0.96 are considered as failed cases.

Performance on dataset-III has been evaluated based on visual inspection. Three volunteers have ranked each registered pair for all three methods. It has been carefully analyzed by viewing the reg-

Table 4.2: Overall Pairs Registered

Method	Dataset	Pairs Registered	Category	# of pairs
GDBICP	Dataset-I	78	Accurate	78
			Acceptable	-
PIIFD	Dataset-I	94	Accurate	27
			Acceptable	67
Our method	Dataset-I	112	Accurate	84
			Acceptable	28
GDBICP	Dataset-II	18	Accurate	18
			Acceptable	-
PIIFD	Dataset-II	20	Accurate	6
			Acceptable	14
Our method	Dataset-II	20	Accurate	11
			Acceptable	9

istered mosaic in checked-board view. The continuity of vessels in this view is the prime cue for this inspection.

Table 4.3: Dataset-III Evaluation

Method	Failure	Success
GDBICP	14	4
PIIFD	6	12
Our method	4	14

Out of the 13 images registered between PIIFD and our method, PIIFD performed better in 3 cases where as our method performed well in 10 cases. The results on three different datasets are given in Fig.[4.16], [4.17],[4.18] and the failed cases are shown in Fig.[4.19].

From the results presented in this section, it can be inferred that GDBICP well, but comparatively for only a subset of the Dataset which contain normal images. The method fails if a single initial point correspondence cannot be established. This is a common scenario in poor quality images. PIIFD performs better for poor quality and pathology affected images but the accuracy is compromised due to the nature of final transformation estimation step in the framework. Our method shows better registration capabilities in terms of both accuracy and registering poor quality and pathology affected images.

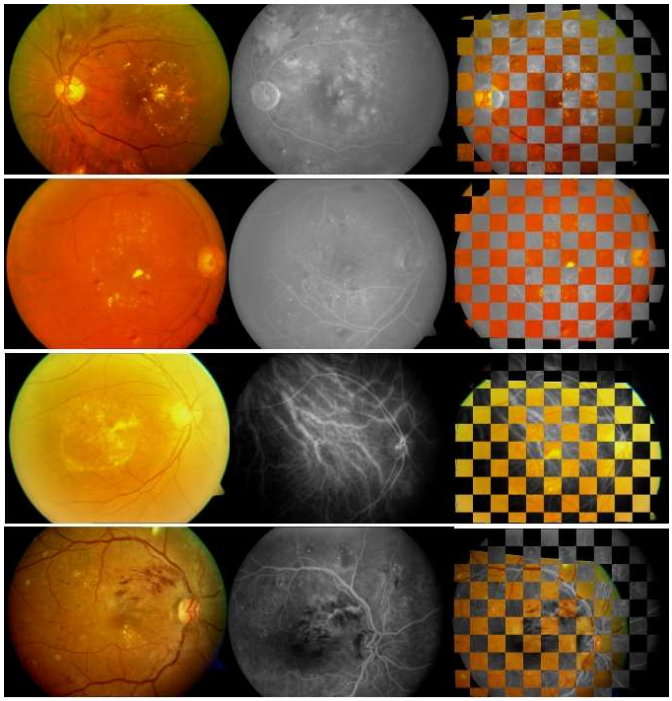


Figure 4.16: Registration of multimodal images from dataset I using the proposed method.

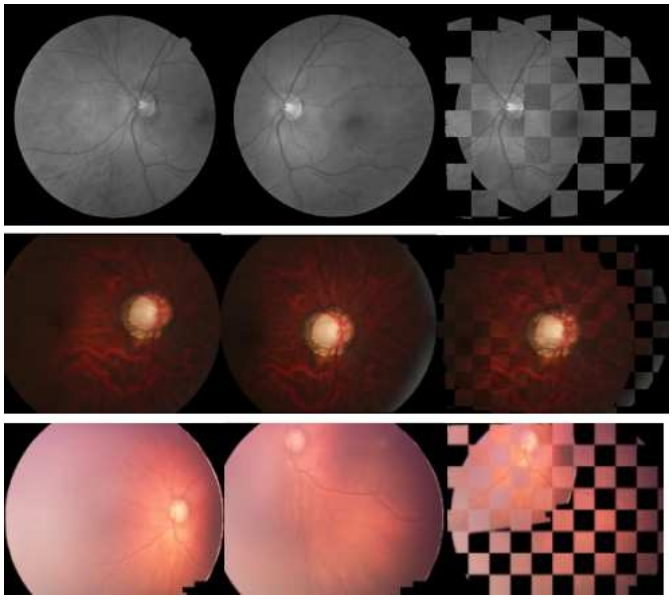


Figure 4.17: Registration of monomodal images from dataset II using the proposed method

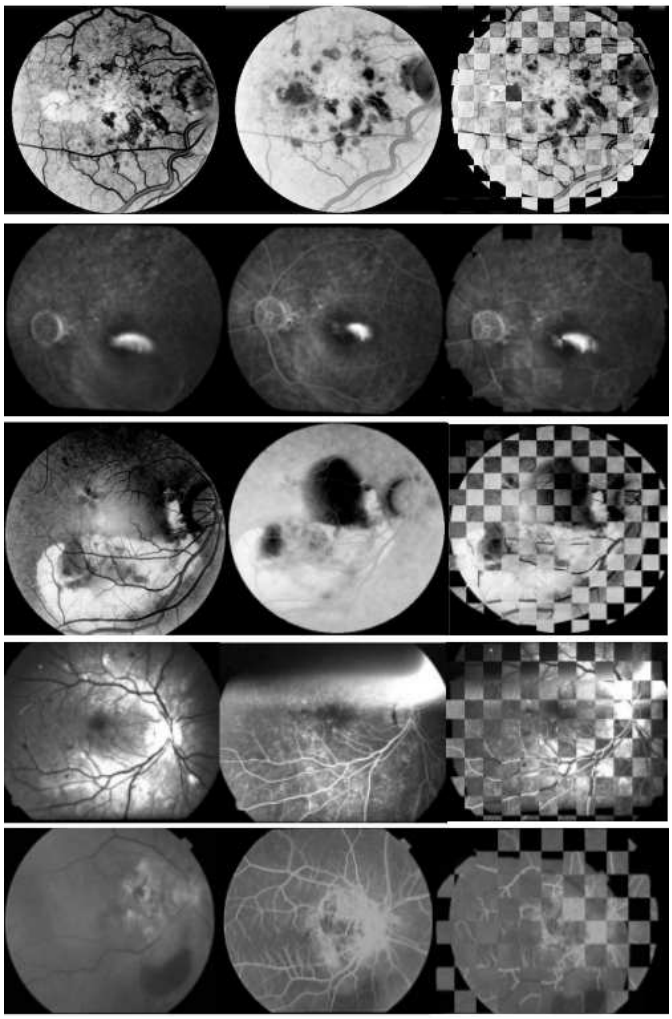


Figure 4.18: Registration of challenging image pairs from dataset III

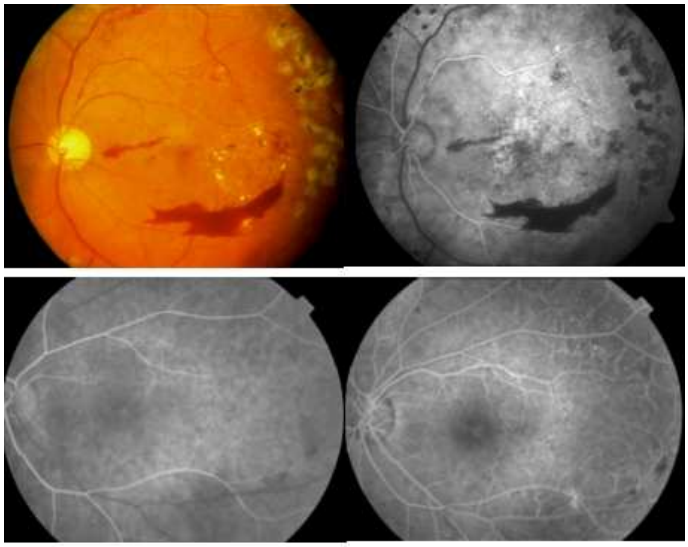


Figure 4.19: Failed cases due to the contrast reversal in multimodal image pairs

Chapter 5

Conclusion and Futurework

5.1 Conclusion

The human retina can be affected by a variety of pathologies like Glaucoma, Diabetic Retinopathy, Age Related Macular degeneration etc. In clinical scenarios, the presence of these diseases and poor quality of the images makes the task of feature based registration challenging. To address these issues a novel feature based registration scheme has been introduced. A set of salient landmarks are detected based on the curvature changes on the topographic surface of the retina using Curvature Dispersion Measure. A local descriptor based on Radon transform has been proposed for robust matching across retinal images. The proposed method uses curvature(Hessian) based enhancement to boost the vessel structures and Radon transform based representation to make it invariant to geometric changes. The attractive feature of this descriptor is its robustness to the presence of lesions while still retaining the required structural information. A modified MSAC(M-estimators Sample and Consensus) has been proposed for pruning false correspondence and estimating the initial transformation function. On the whole, the minor contributions at each stage of feature based registration scheme presented here is of significance. We evaluated our method against two recent schemes on three different datasets which includes both monomodal and multimodal images. The results show that our method is able to perform well for poor quality and pathology affected images while performing on par with the existing methods on normal images.

5.2 Futurework

Possible extension to this work includes (i) Optimization of the proposed method in terms of computational efficiency. One such promising direction is to use multi-resolution optimization procedures. (ii) Multiscale filtering is the most computationally expensive part of the algorithm. Integral Image based representation may be pursued for approximating the filter responses. (iii) An extension to 3D medical data would be an interesting direction to explore. (iv) Feature based registration algorithms are in general easily parallelized for real-time intervention systems. A GPU based implementation would yield rapid speeds for such systems. (iv) The dimension of the radon descriptor is 720 long, dimensionality reduction methods beyond PCA may yield a much more compact description. (V) The proposed method may be generalized to a wide varieties of modalities if the transformation model selection is automated. (vi) Extension of registration algorithms to Superresolution, Mosaicing, Fusion, Vessel Segmentation, automatic pathology detection are still an active area of research.

Chapter 6

Appendix I

In this section we show additional retinal image registration results for the purpose of visual inspection. We show the results using the standard checkerboard pattern and the primary cues for visually evaluating the result is to trace the vessel structures through the checkerboard. In this view the FFA images are inverted for the purpose of easy interpretation.

6.1 Registration of Multimodal Retinal Images- DataSet-I

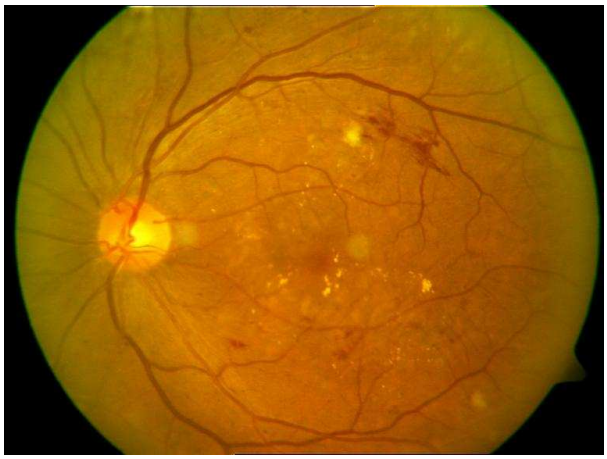


Figure 6.1: Color Fundus Image (CFI)

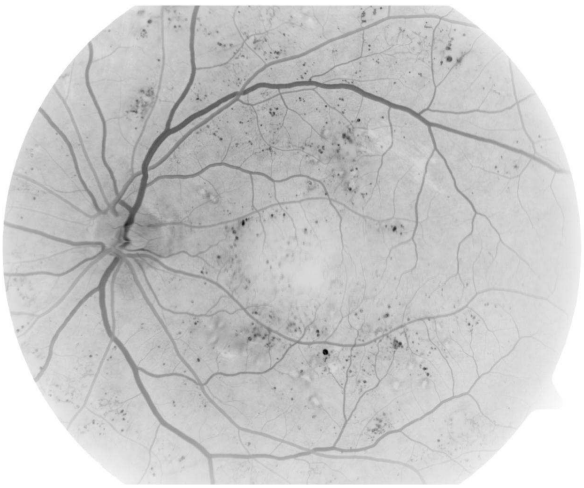


Figure 6.2: Fluroscein Fundus Angiogram FFA)

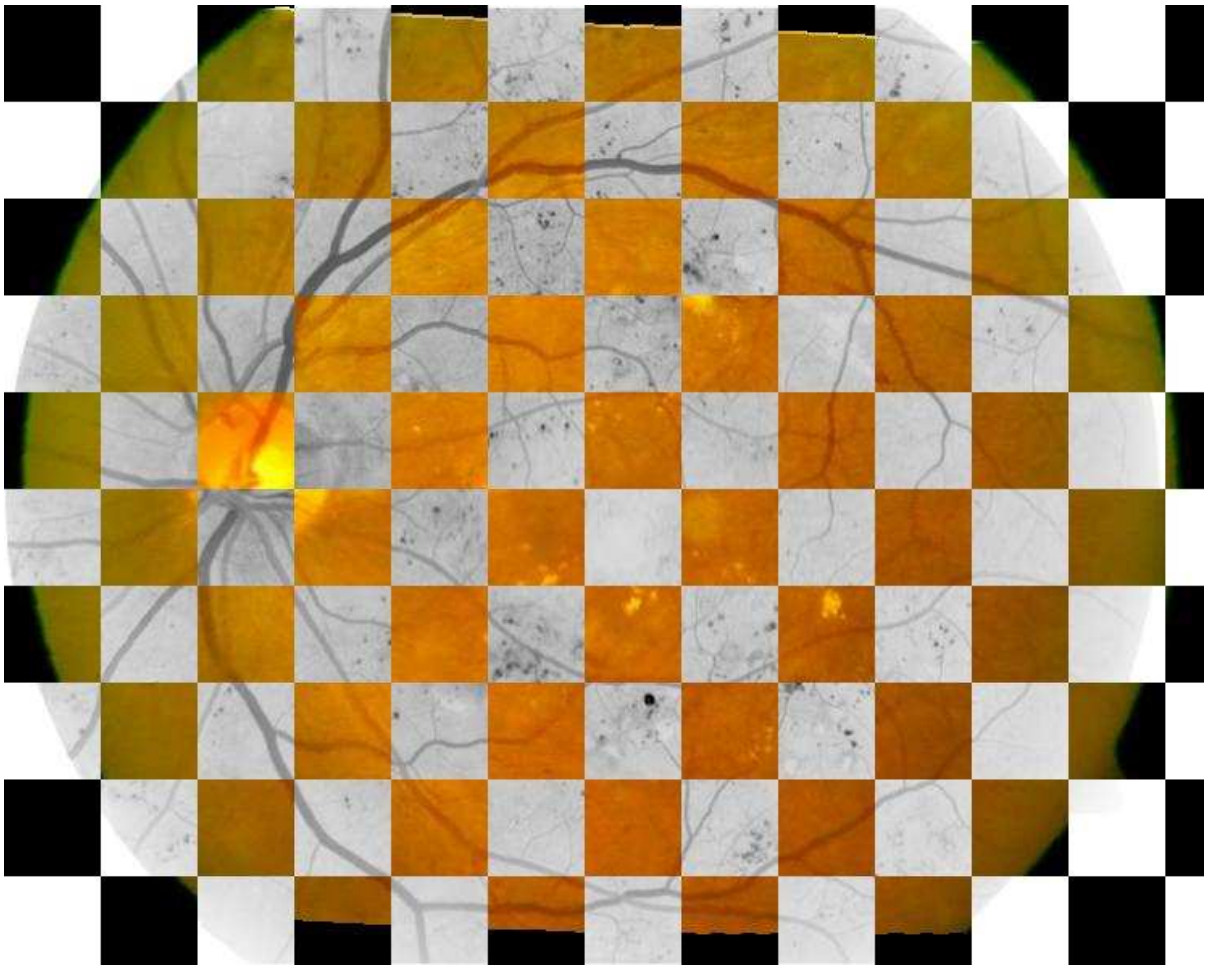


Figure 6.3: Image showing registration of CFI/FFA.

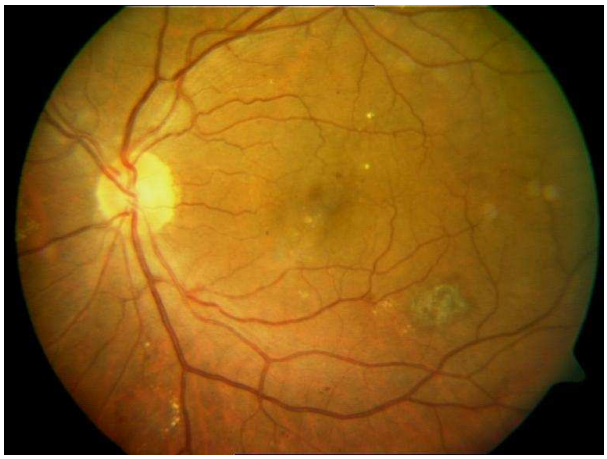


Figure 6.4: Color Fundus Image (CFI)

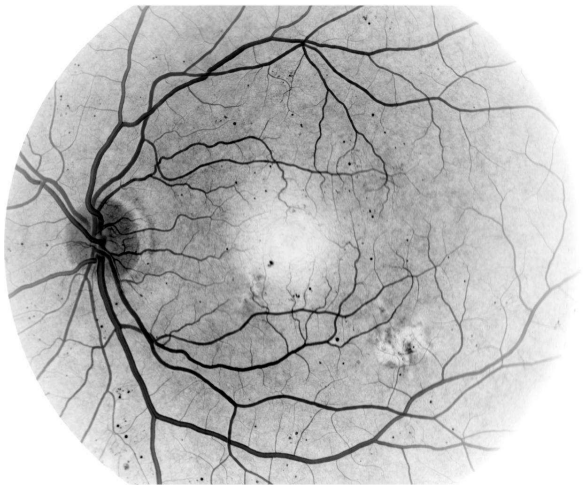


Figure 6.5: Fluroscein Fundus Angiogram FFA)

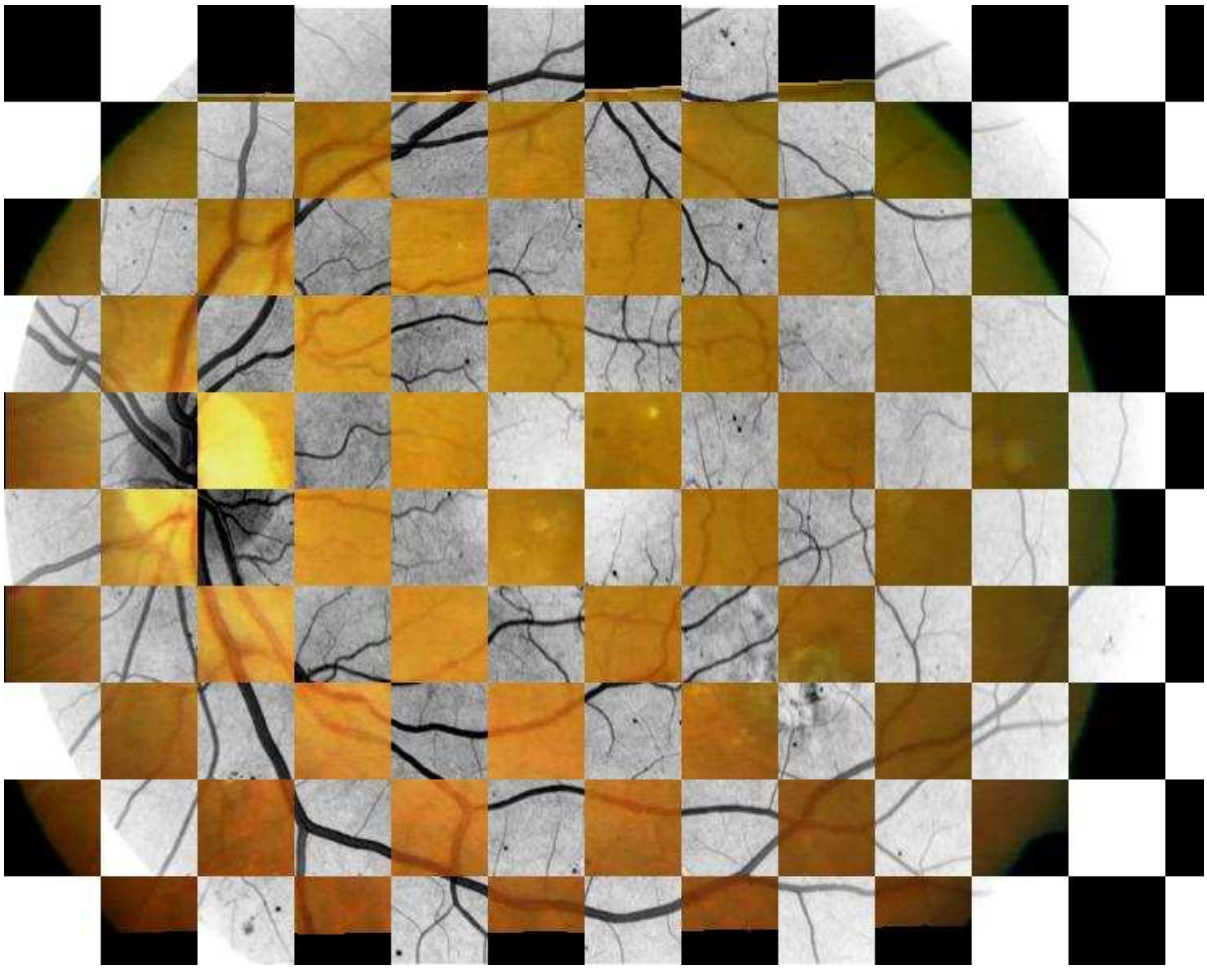


Figure 6.6: Image showing registration of CFI/FFA.

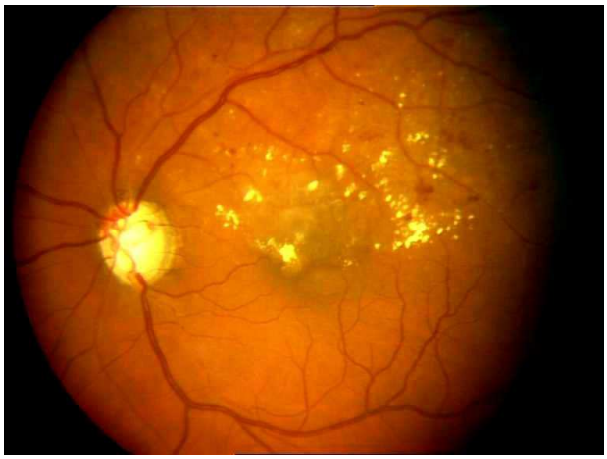


Figure 6.7: Color Fundus Image (CFI)

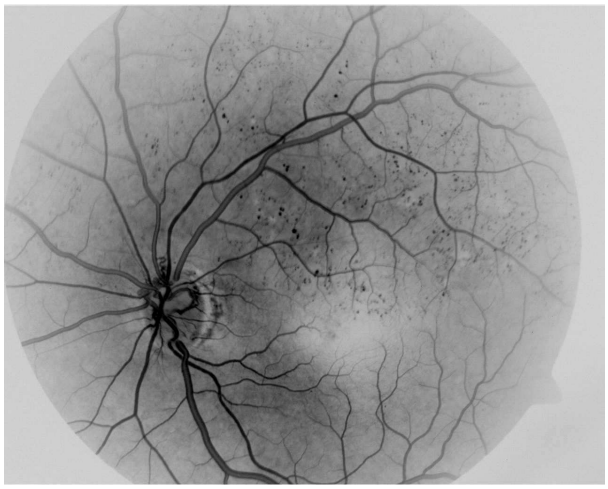


Figure 6.8: Fluroscein Fundus Angiogram FFA)

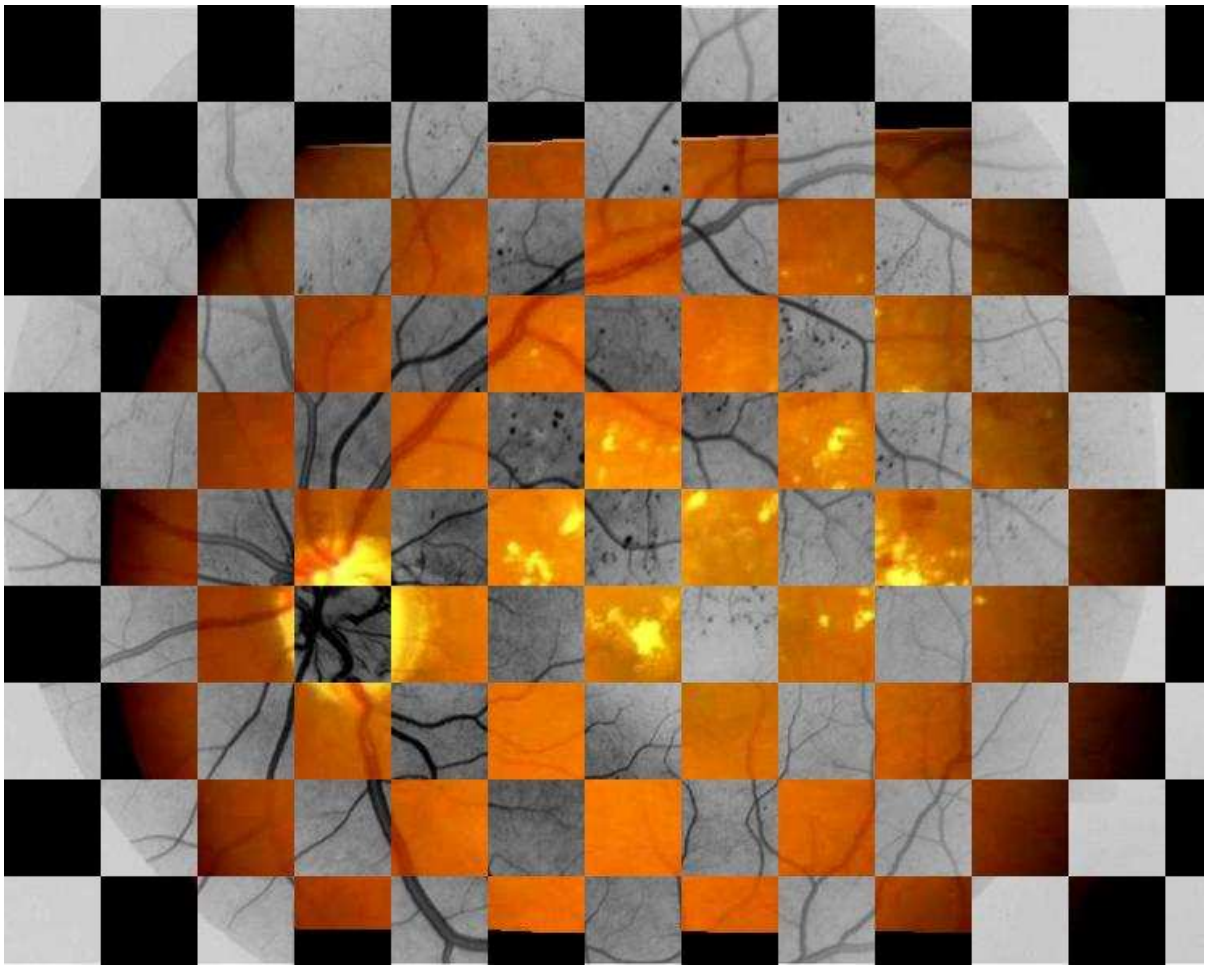


Figure 6.9: Image showing registration of CFI/FFA.

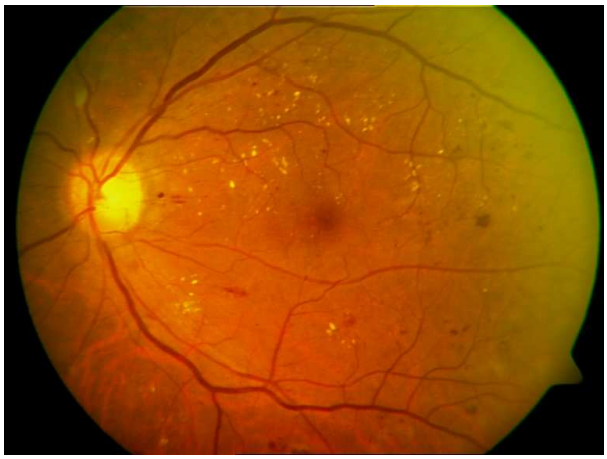


Figure 6.10: Color Fundus Image (CFI)

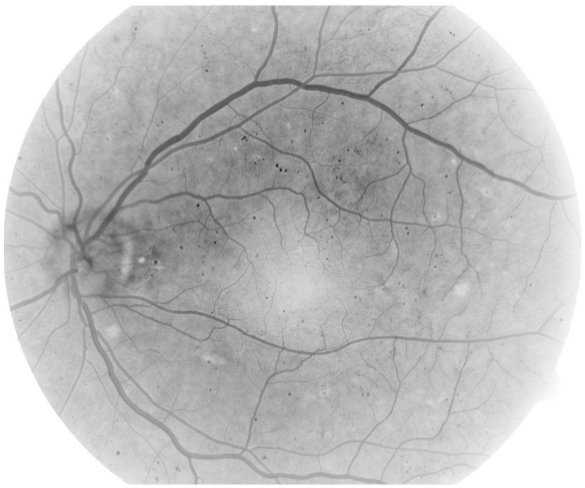


Figure 6.11: Fluroscein Fundus Angiogram FFA)

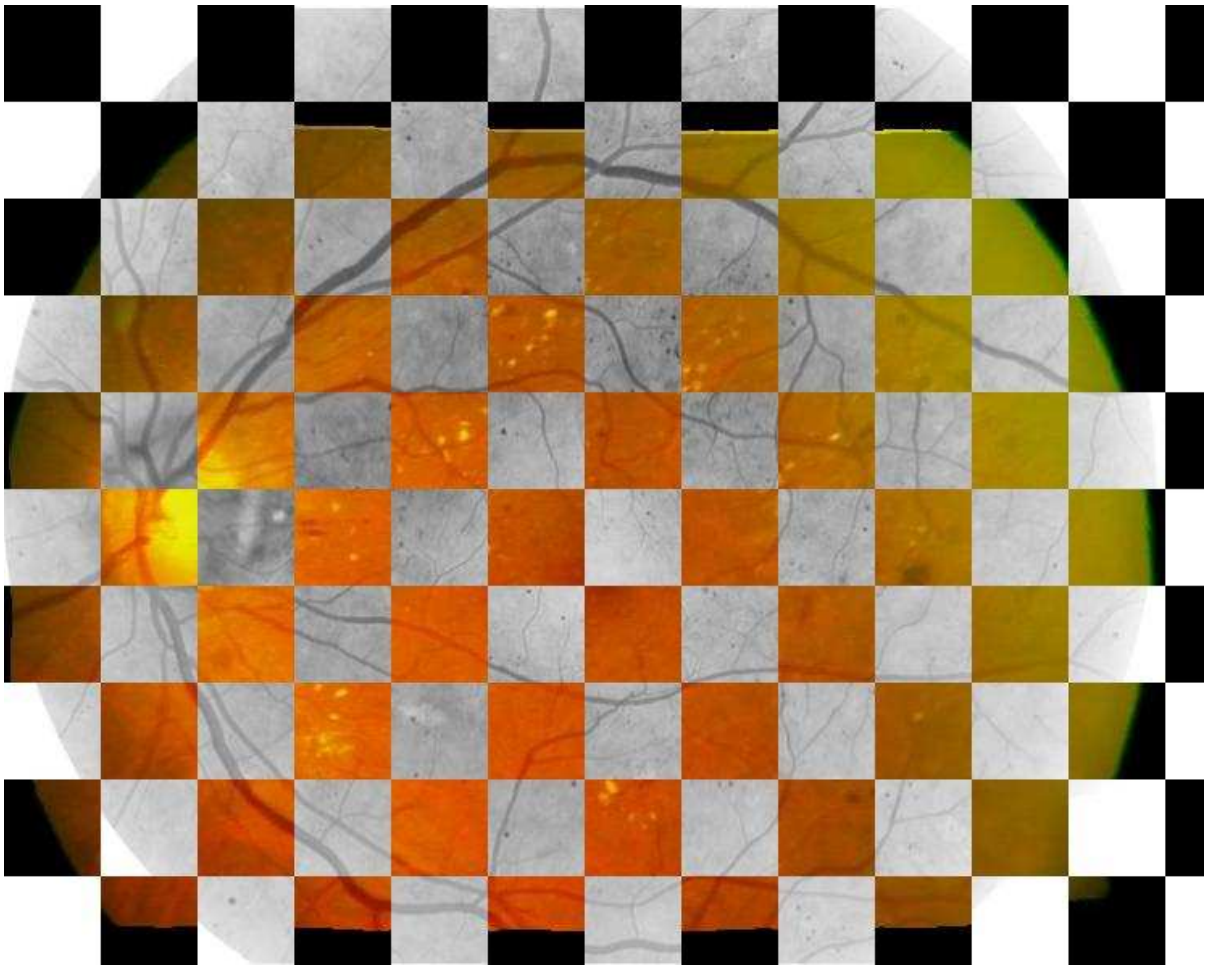


Figure 6.12: Image showing registration of CFI/FFA.

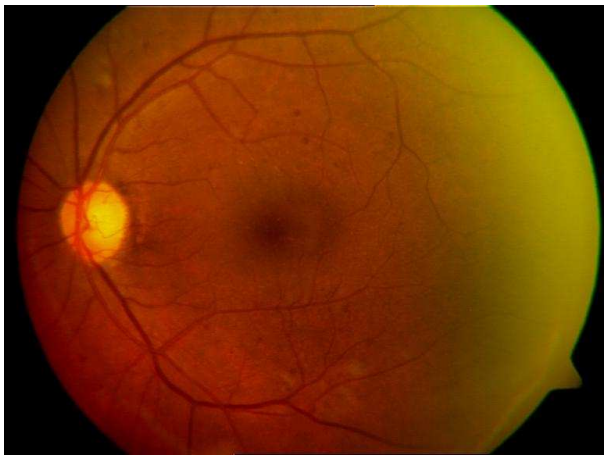


Figure 6.13: Color Fundus Image (CFI)

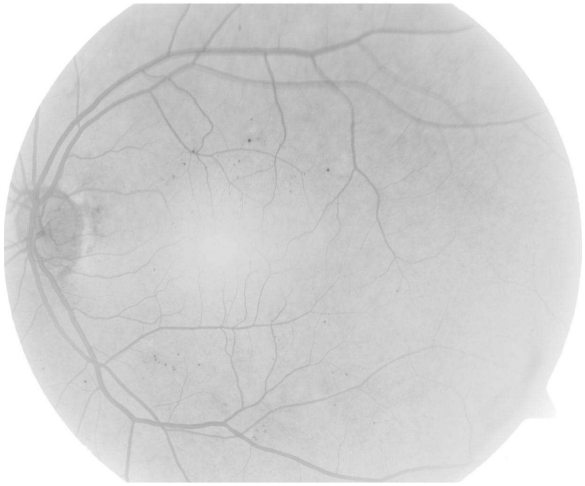


Figure 6.14: Fluroscein Fundus Angiogram FFA)

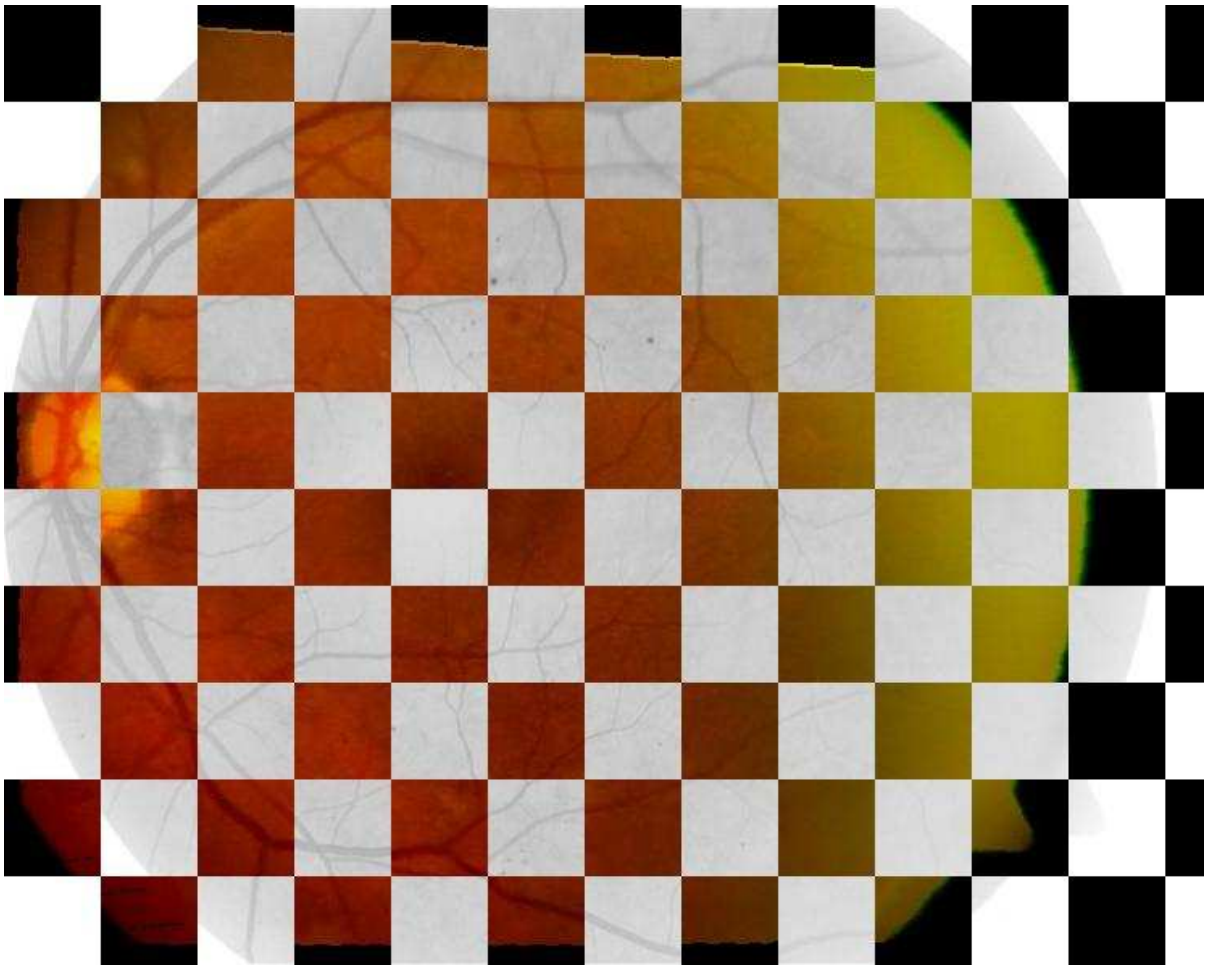


Figure 6.15: Image showing registration of CFI/FFA.

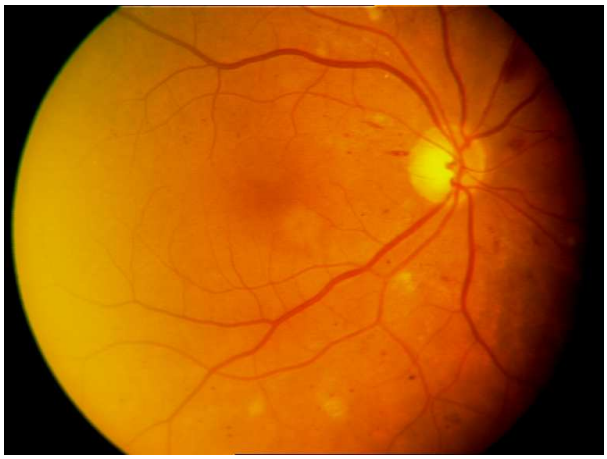


Figure 6.16: Color Fundus Image (CFI)

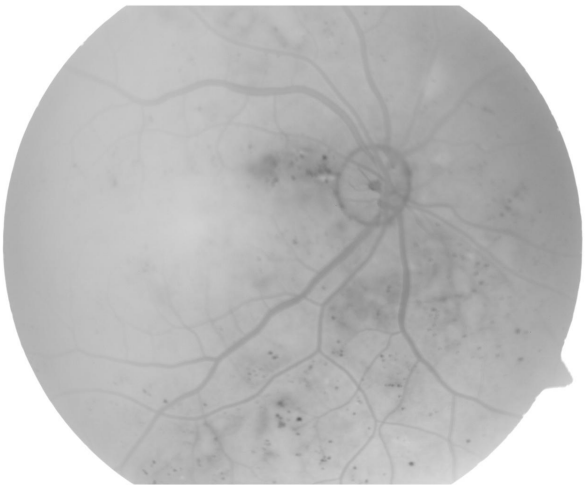


Figure 6.17: Fluroscein Fundus Angiogram FFA)

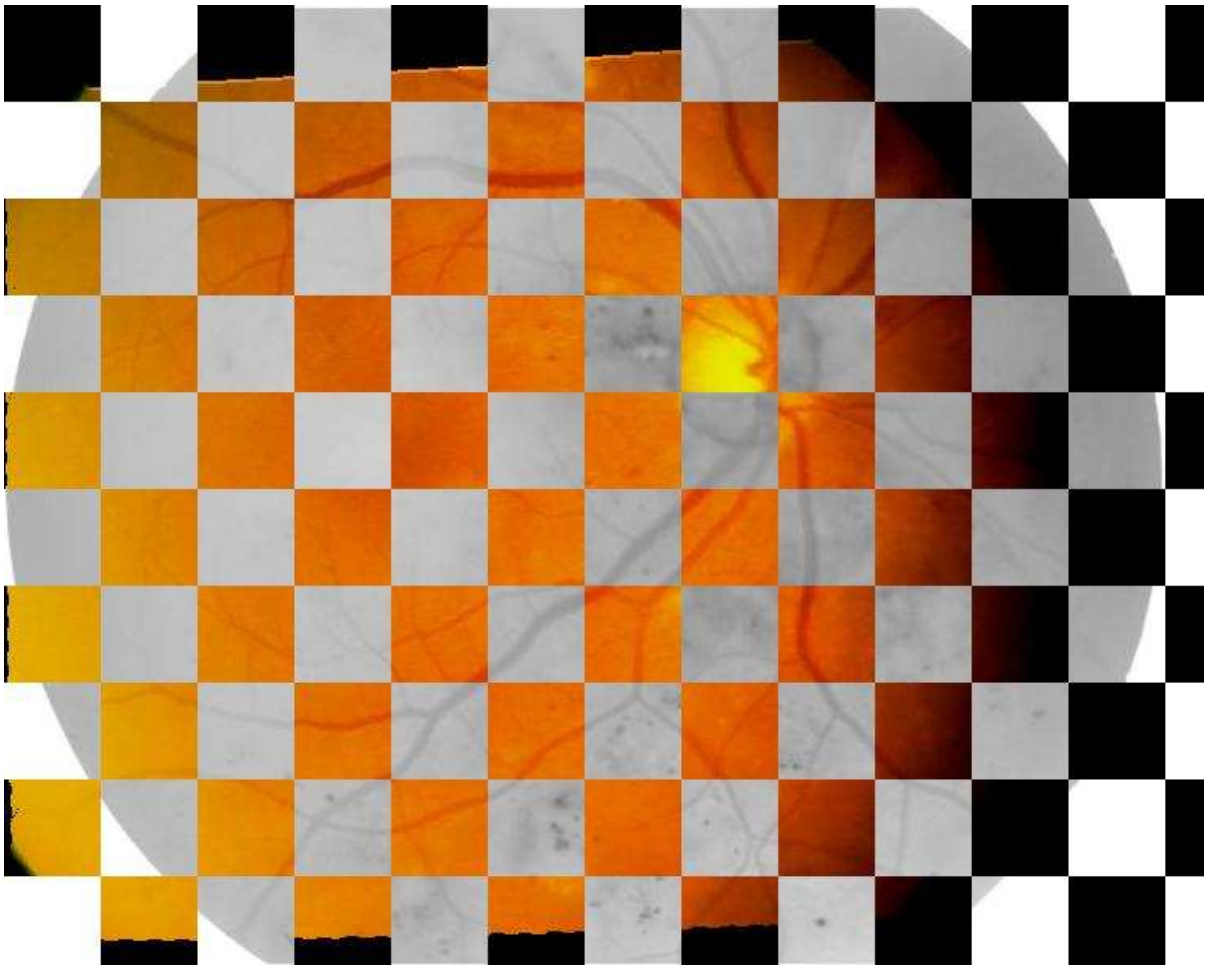


Figure 6.18: Image showing registration of CFI/FFA.

Figure 6.19: Color Fundus Image (CFI)

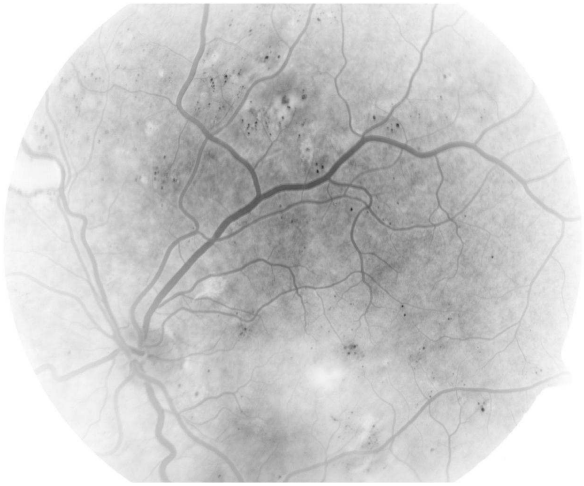


Figure 6.20: Fluroscein Fundus Angiogram FFA)

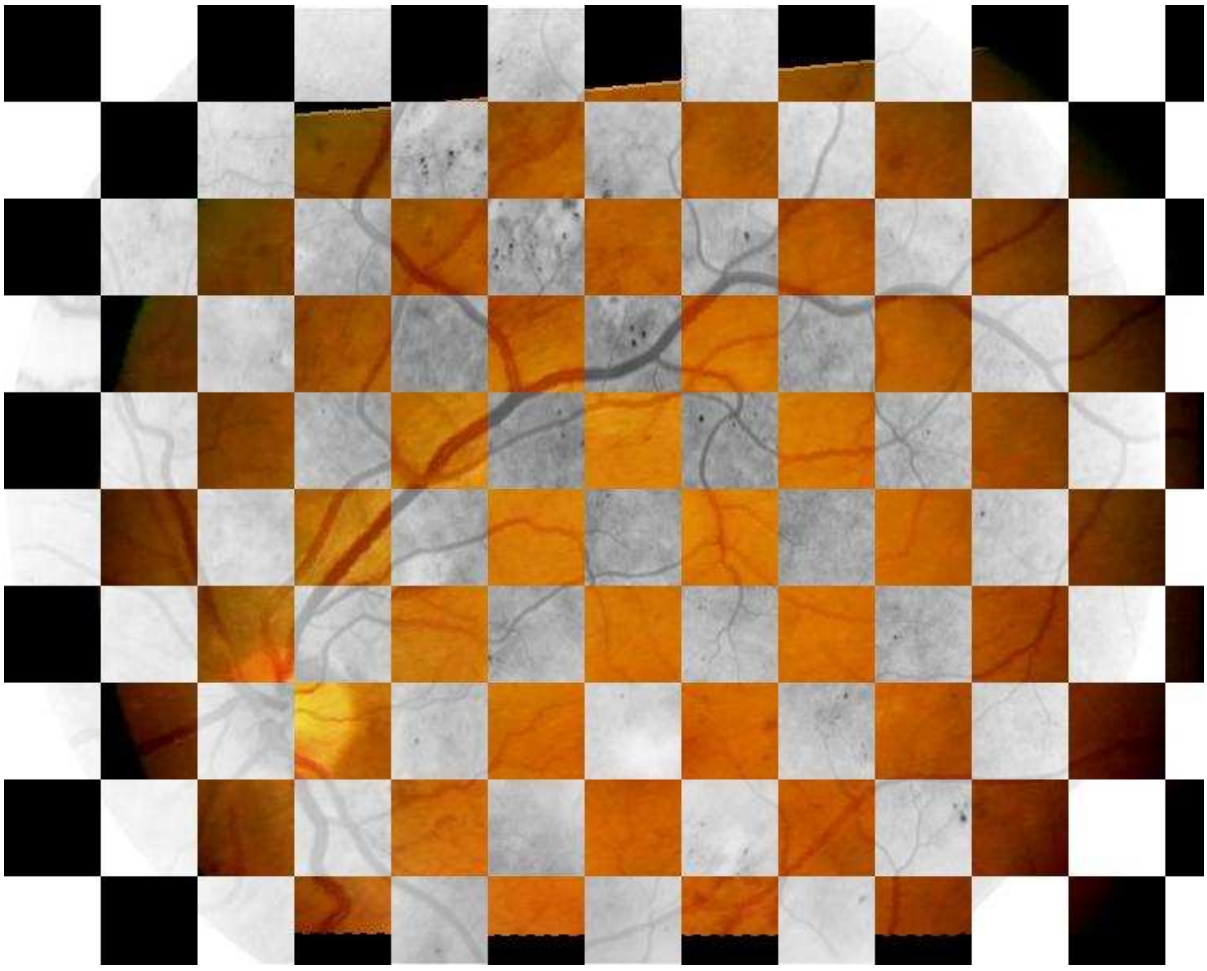


Figure 6.21: Image showing registration of CFI/FFA.

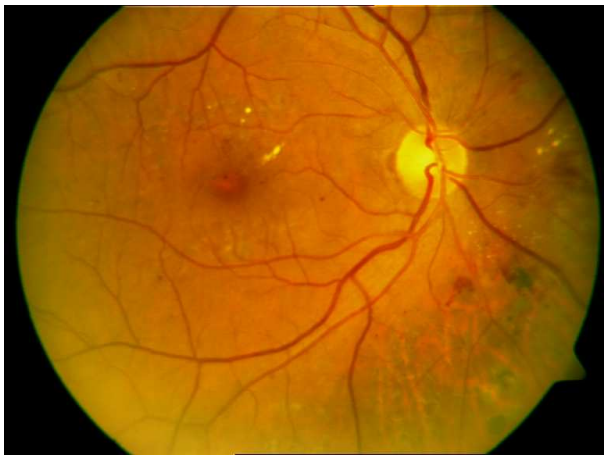


Figure 6.22: Color Fundus Image (CFI)

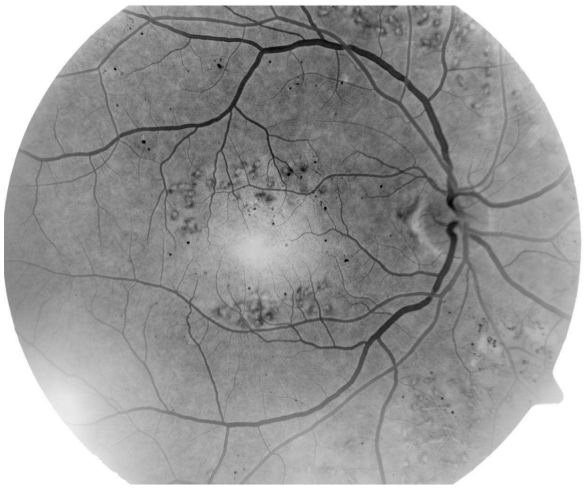


Figure 6.23: Fluroscein Fundus Angiogram FFA)

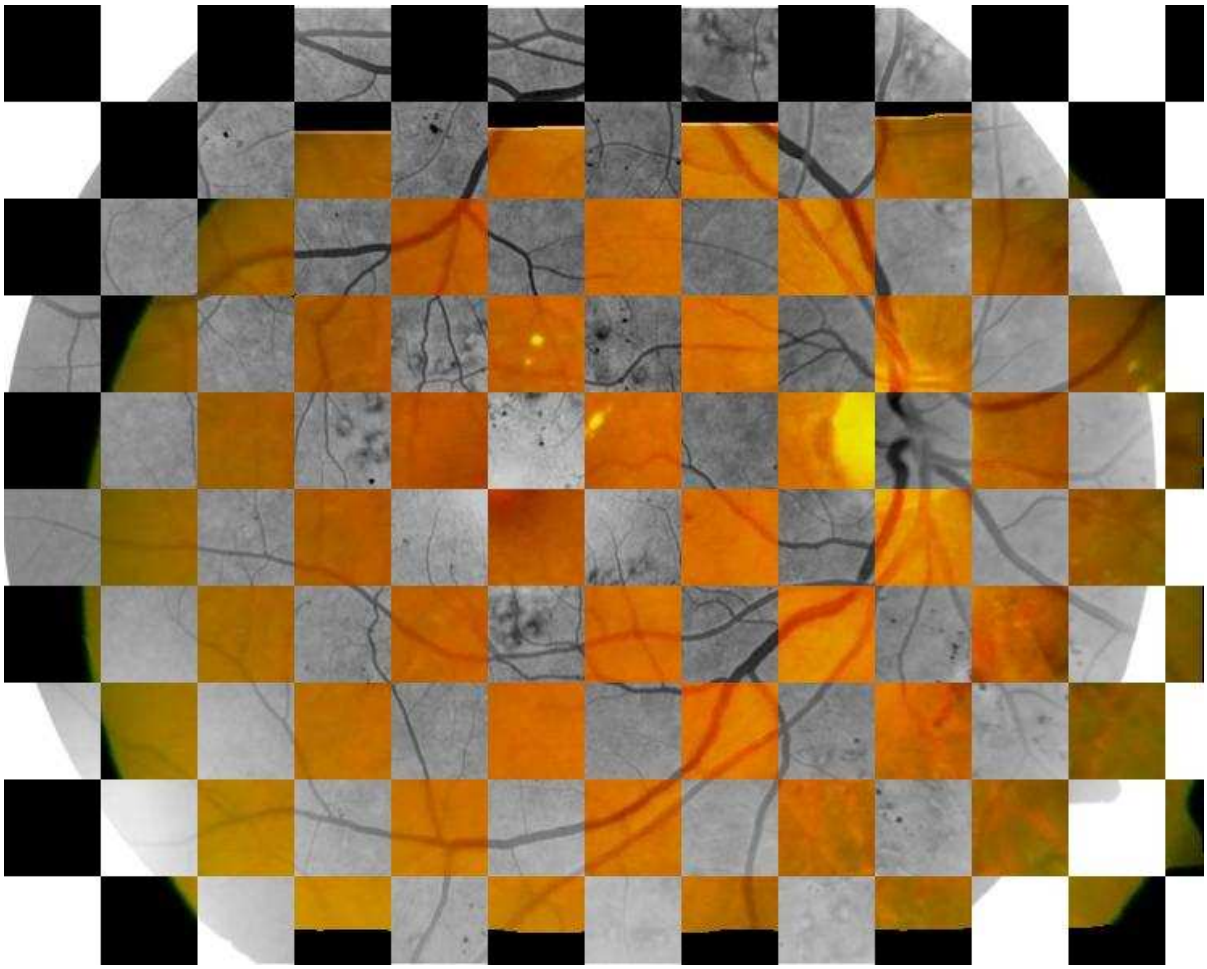


Figure 6.24: Image showing registration of CFI/FFA.

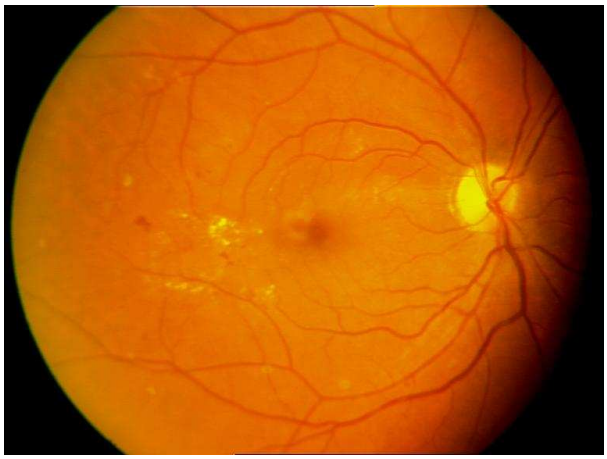


Figure 6.25: Color Fundus Image (CFI)

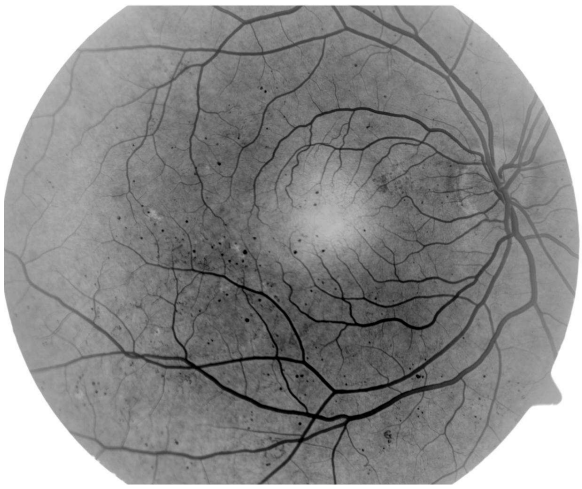


Figure 6.26: Fluroscein Fundus Angiogram FFA)

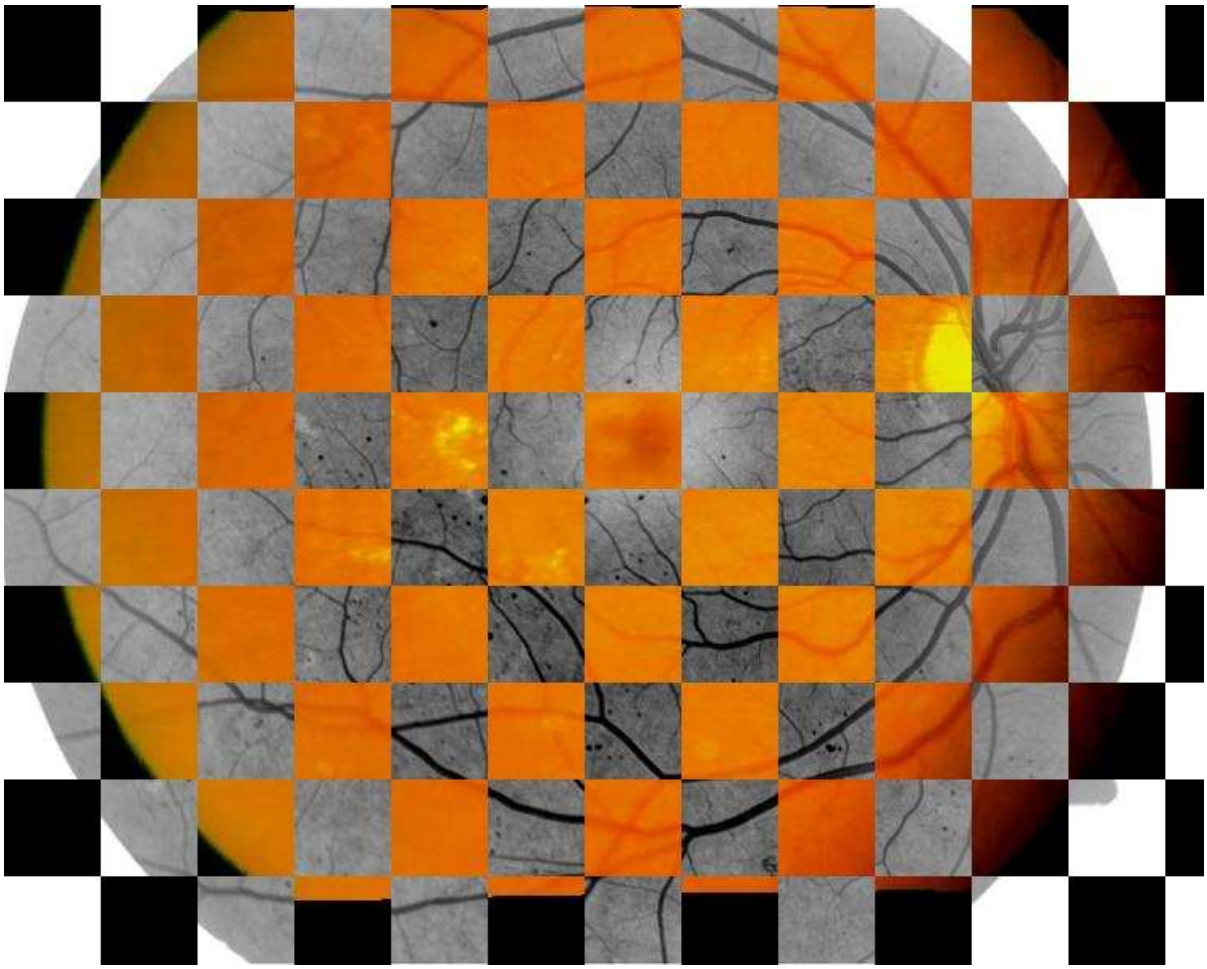


Figure 6.27: Image showing registration of CFI/FFA.

6.2 Registration of Retinal Images- DataSet-III

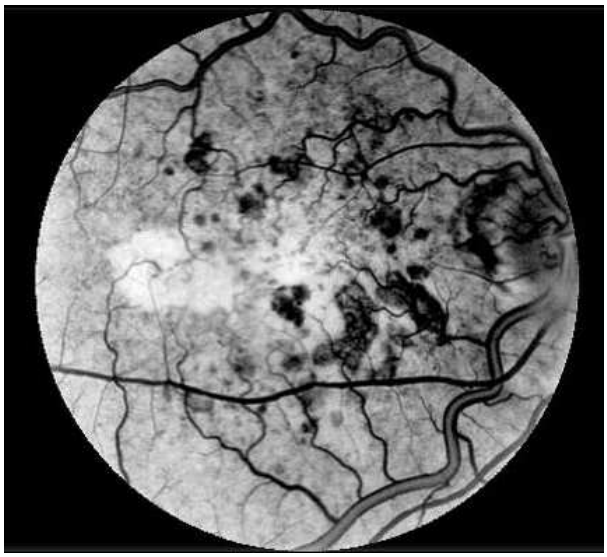


Figure 6.28: Color Fundus Image (CFI)

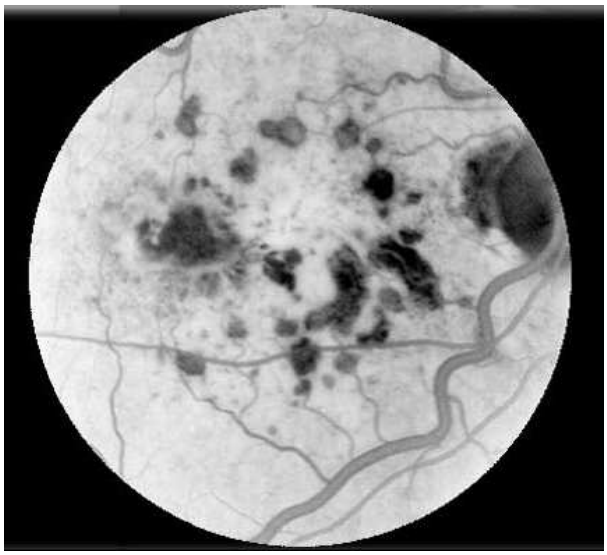


Figure 6.29: Fluroscein Fundus Angiogram FFA)

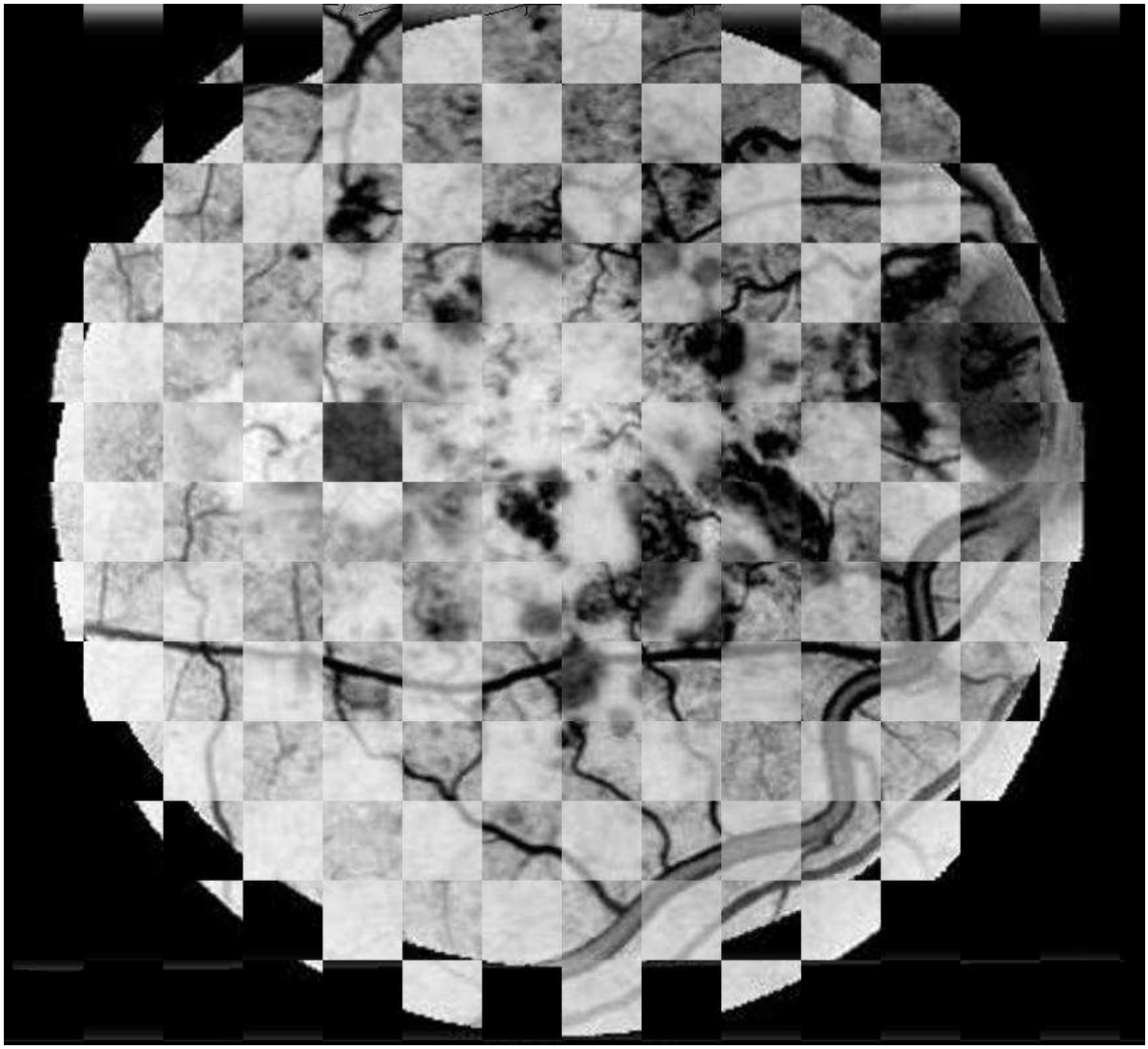


Figure 6.30: Image showing registration of CFI/FFA.

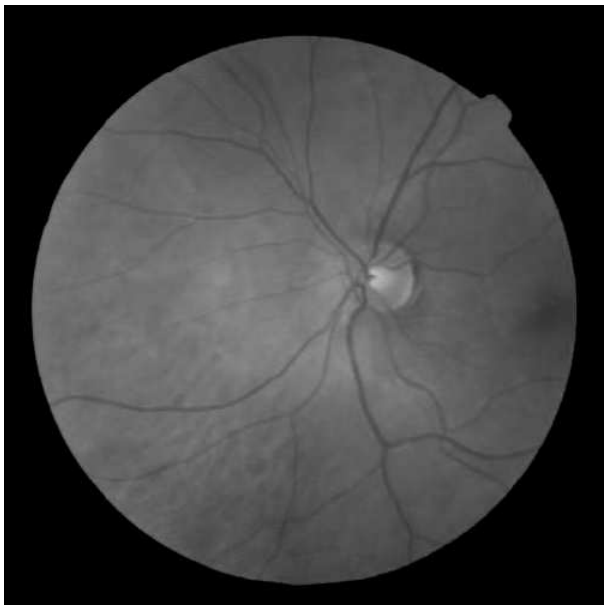


Figure 6.31: Color Fundus Image (CFI)

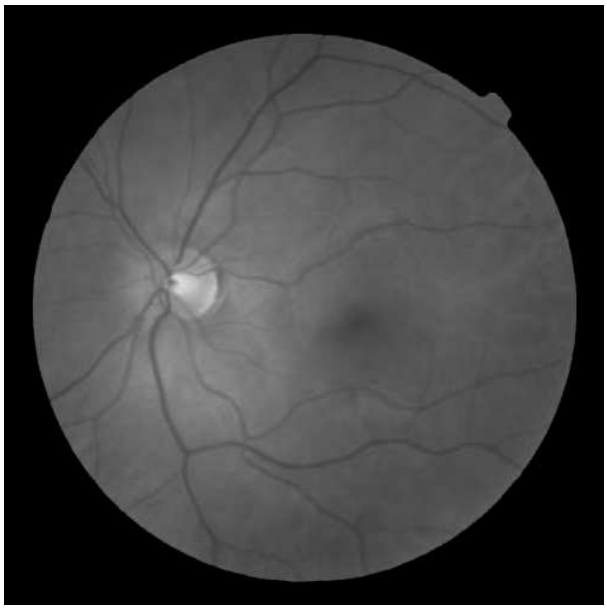


Figure 6.32: Fluroscein Fundus Angiogram FFA)

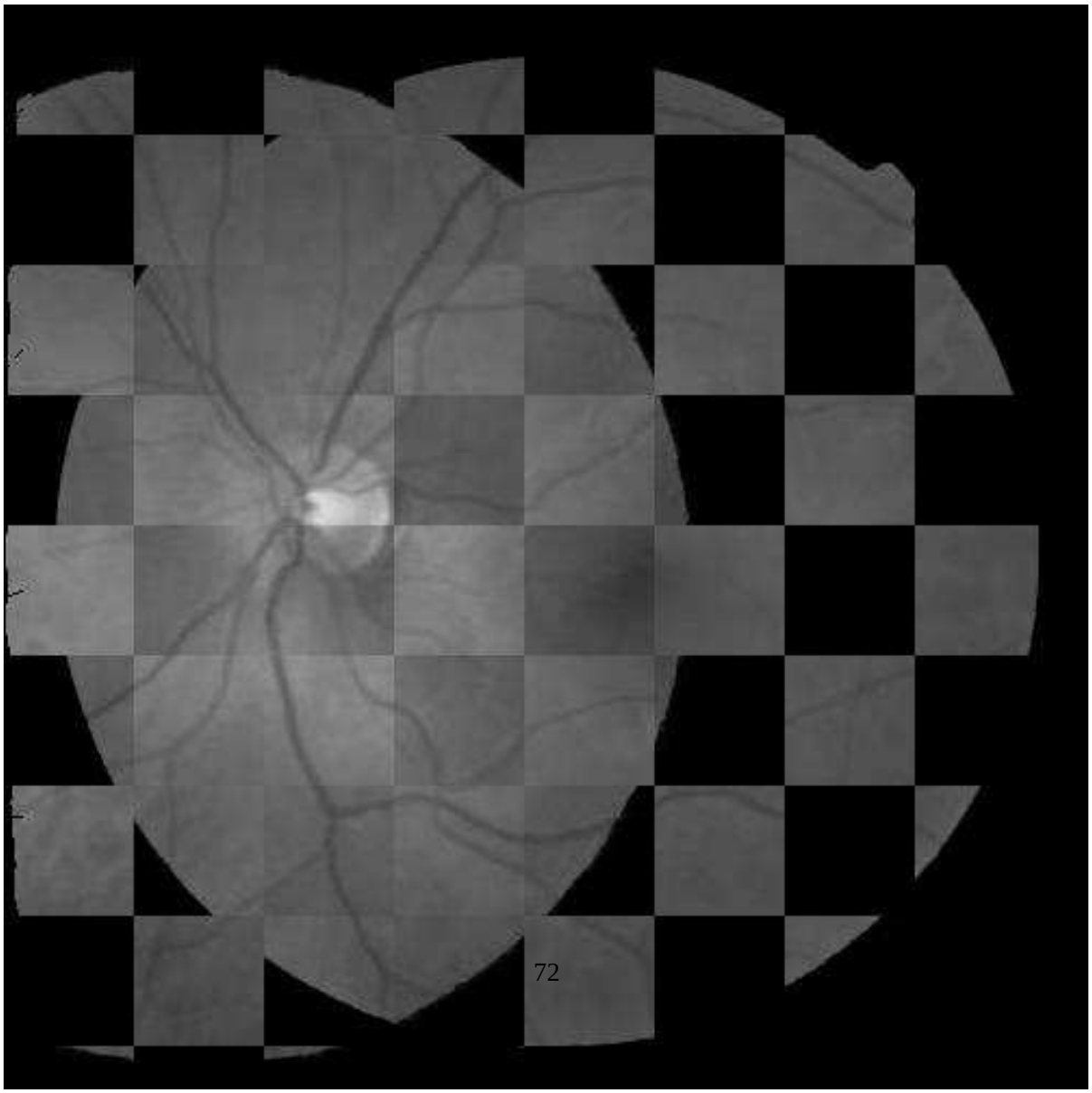


Figure 6.33: Image showing registration of CFI/FFA.

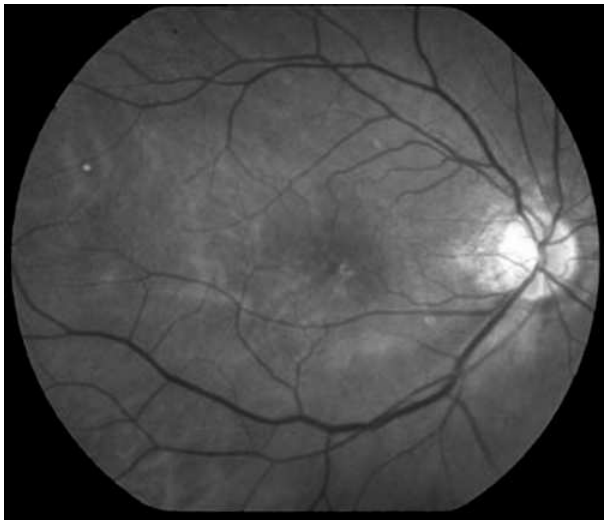


Figure 6.34: Color Fundus Image (CFI)

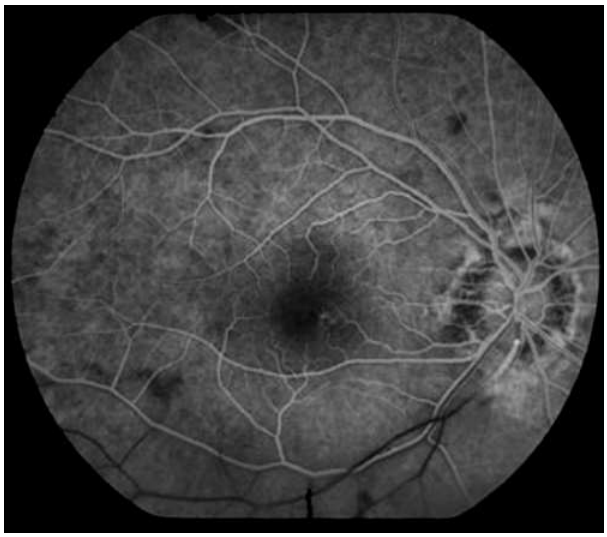


Figure 6.35: Fluroscein Fundus Angiogram FFA)

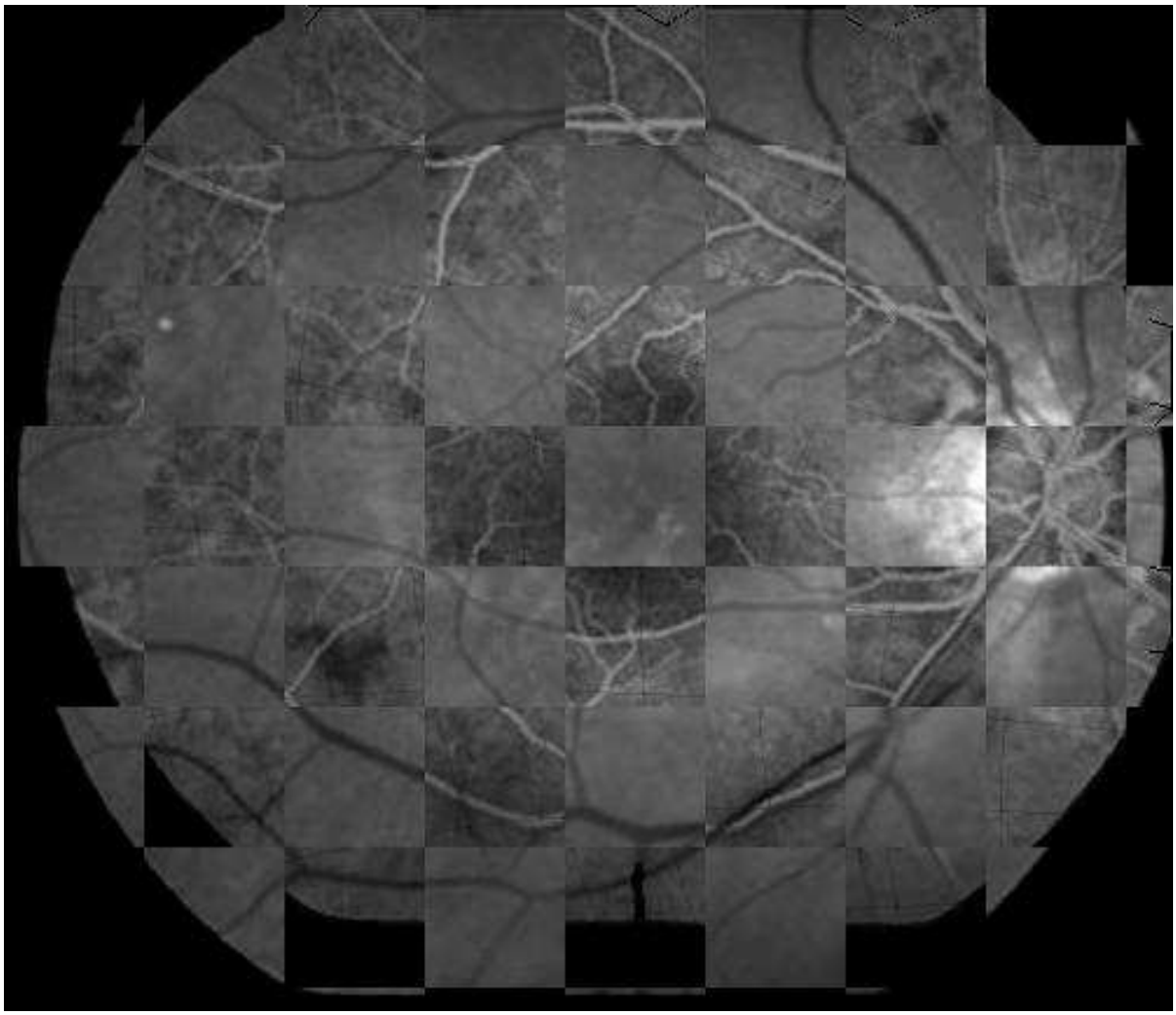


Figure 6.36: Image showing registration of CFI/FFA.

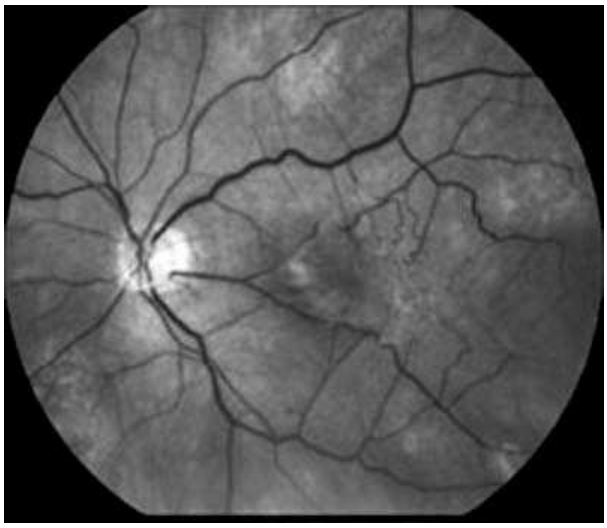


Figure 6.37: Color Fundus Image (CFI)

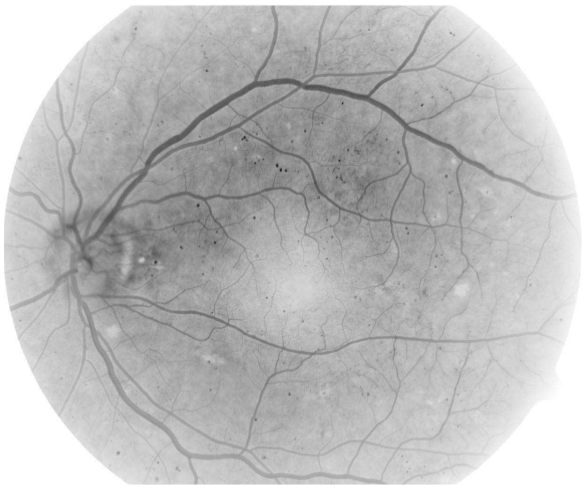


Figure 6.38: Fluroscein Fundus Angiogram FFA)

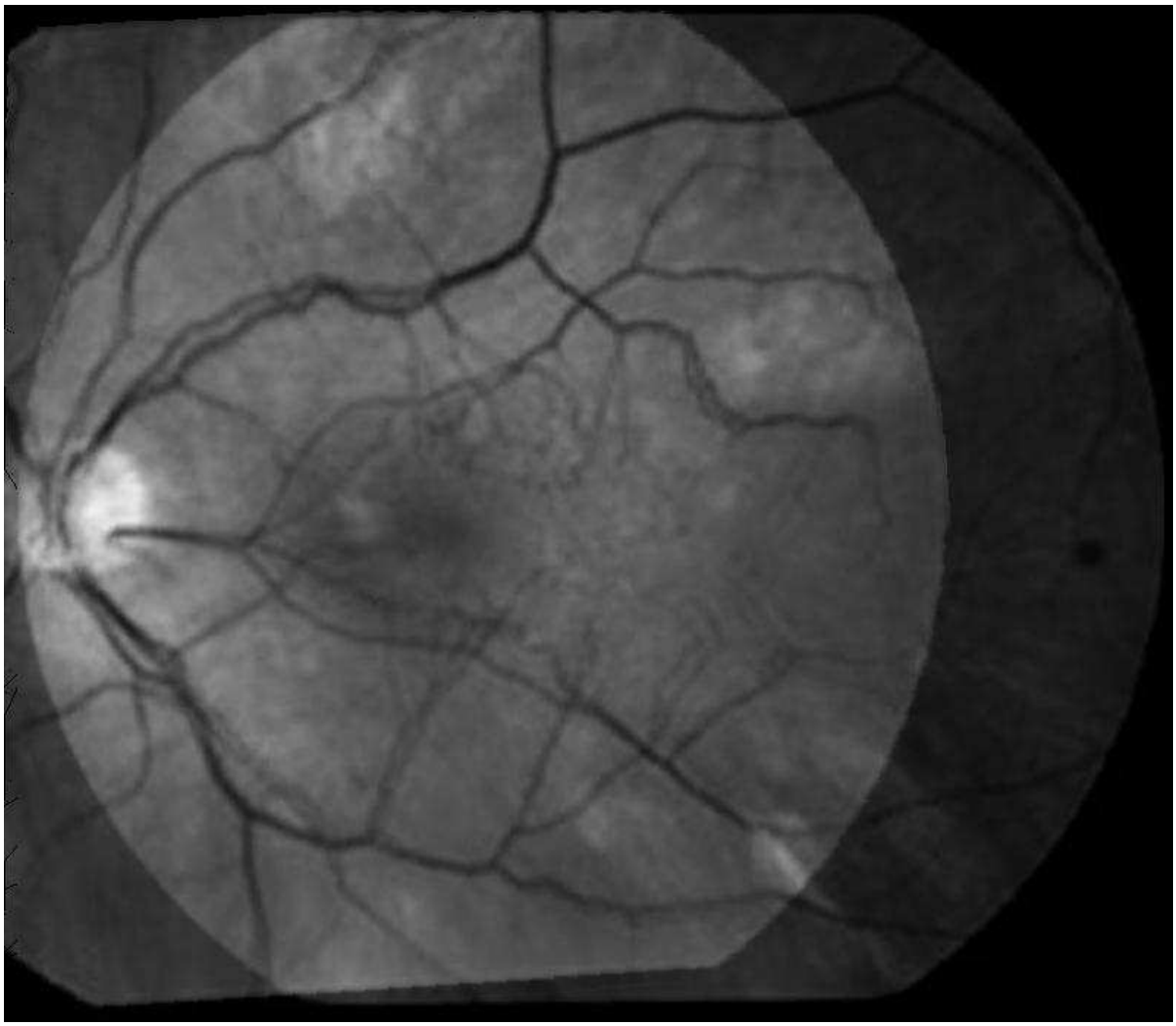


Figure 6.39: Image showing registration of CFI/FFA.

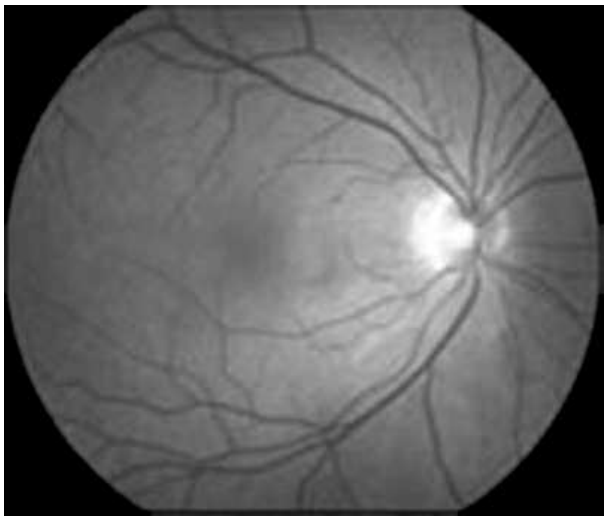


Figure 6.40: Color Fundus Image (CFI)

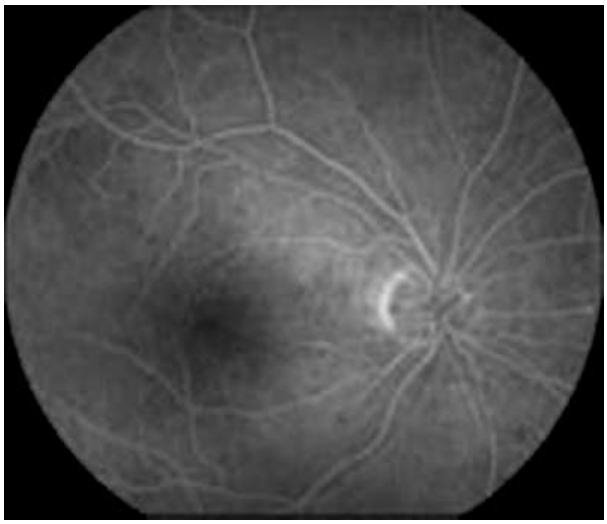


Figure 6.41: Fluroscein Fundus Angiogram FFA)

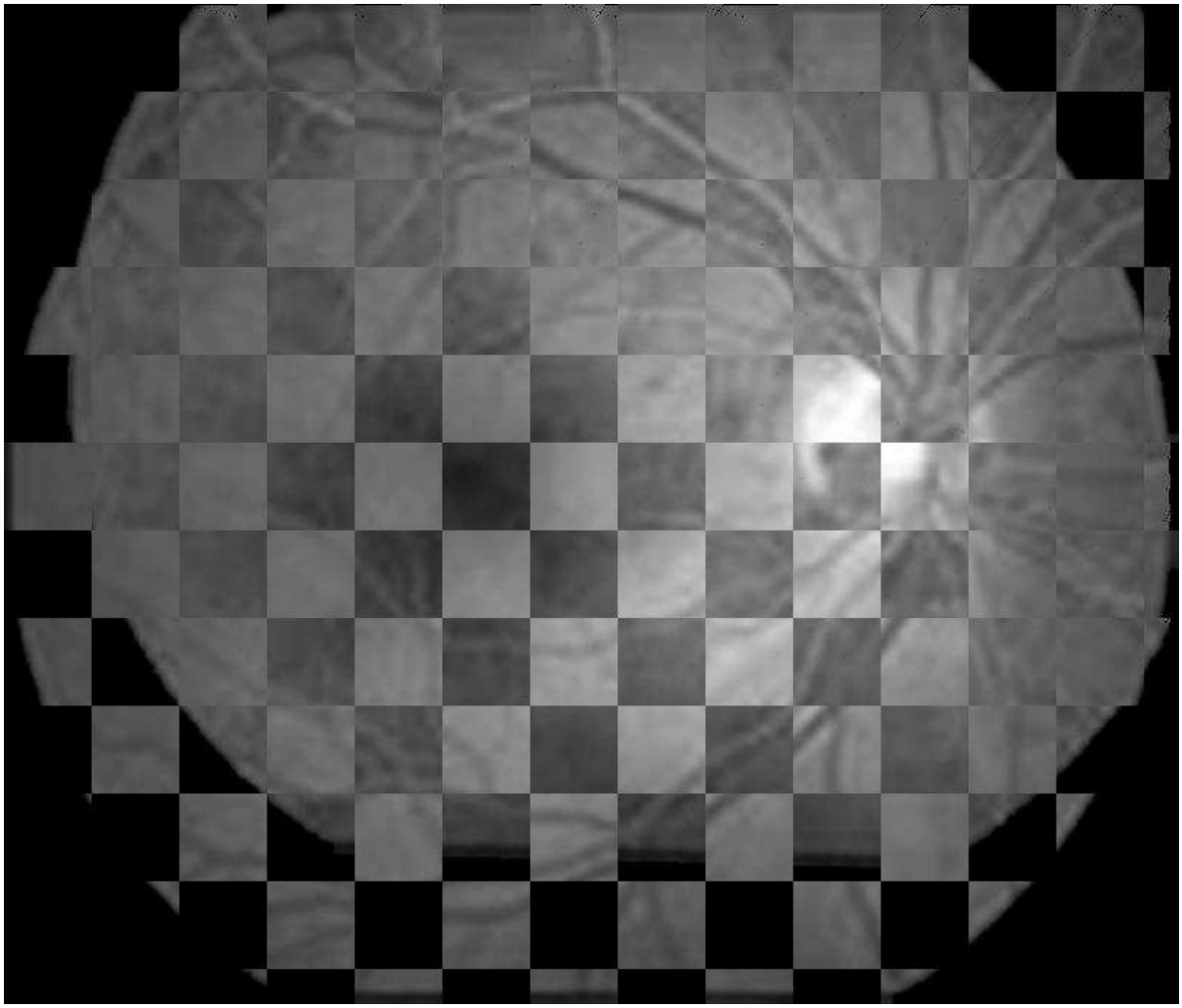


Figure 6.42: Image showing registration of CFI/FFA.

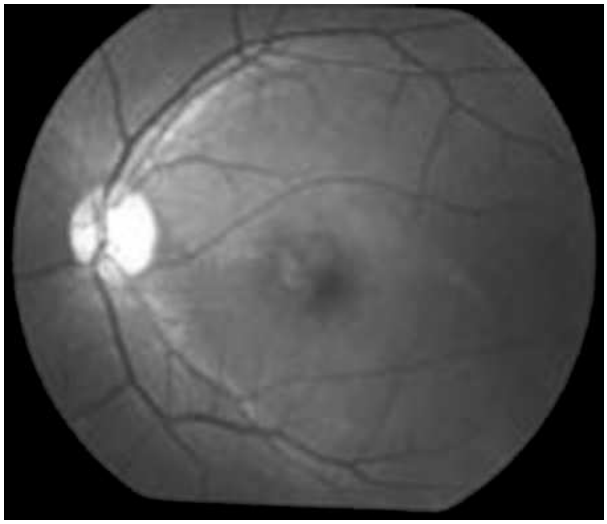


Figure 6.43: Color Fundus Image (CFI)

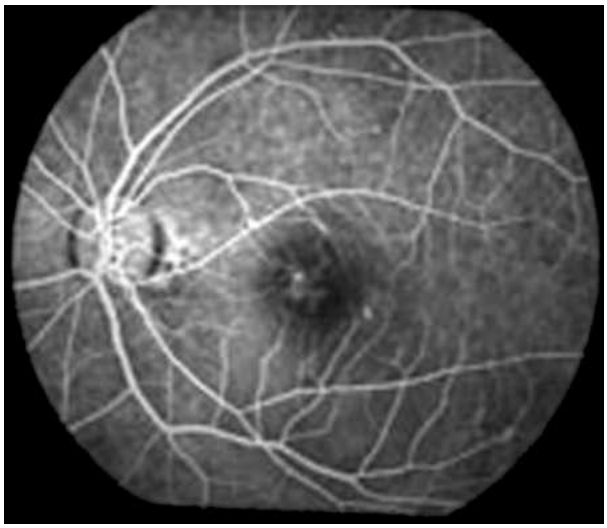


Figure 6.44: Fluroscein Fundus Angiogram FFA)

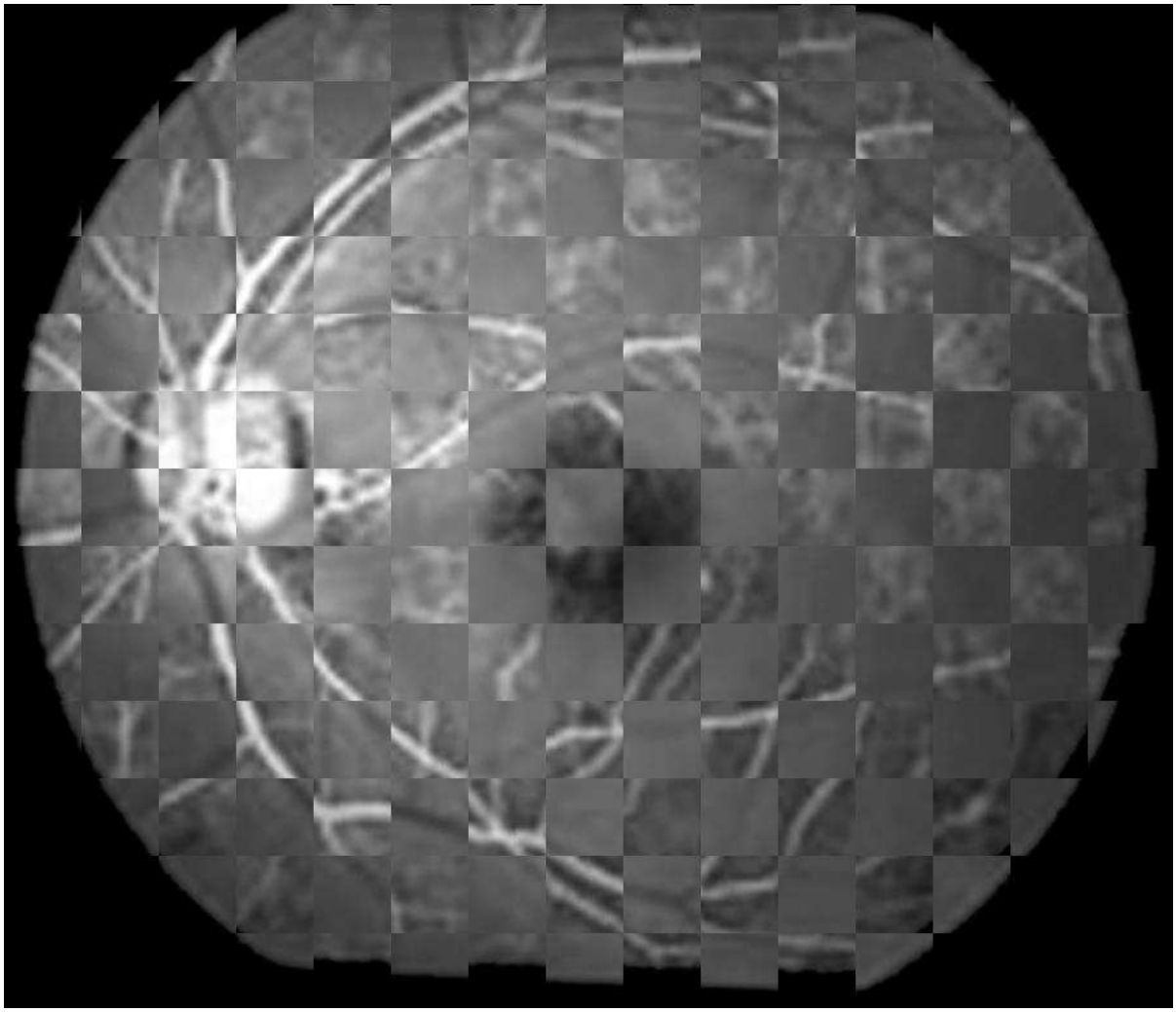


Figure 6.45: Image showing registration of CFI/FFA.

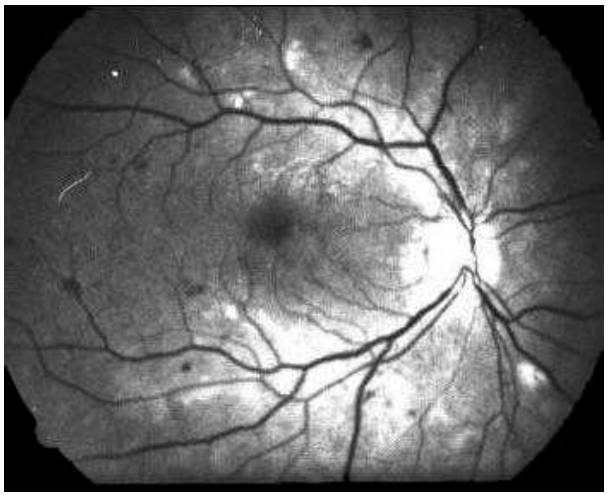


Figure 6.46: Color Fundus Image (CFI)

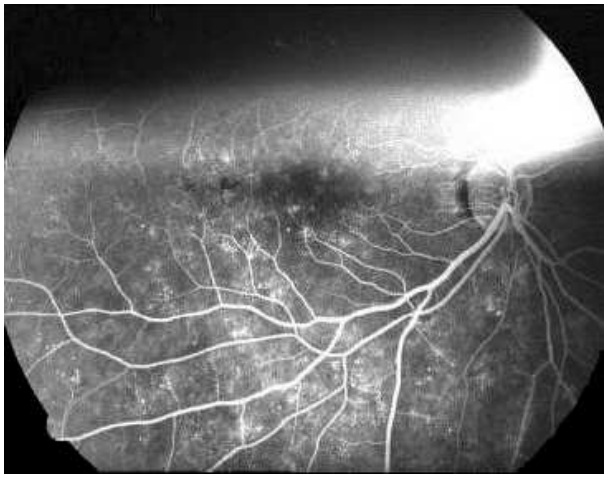


Figure 6.47: Fluroscein Fundus Angiogram FFA)

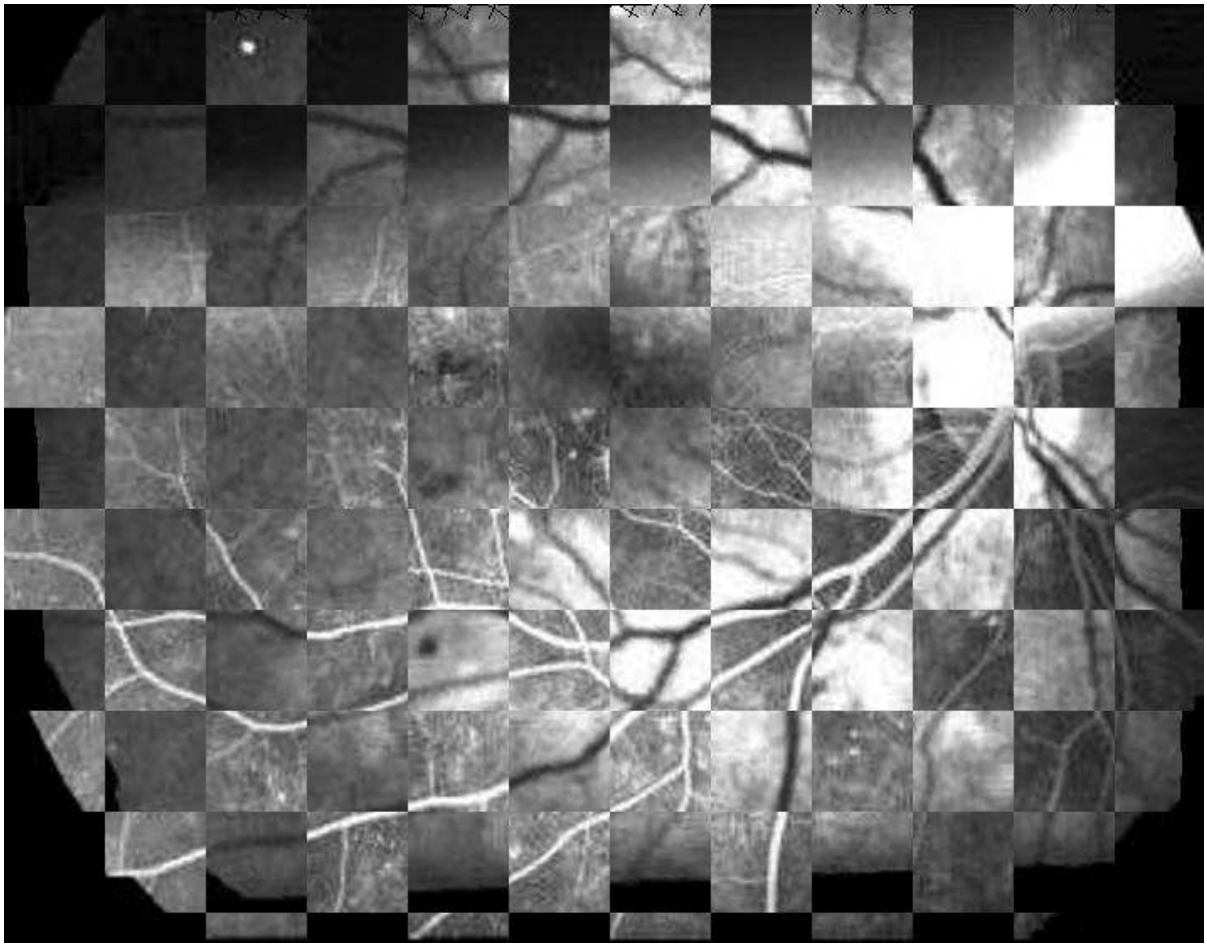


Figure 6.48: Image showing registration of CFI/FFA.

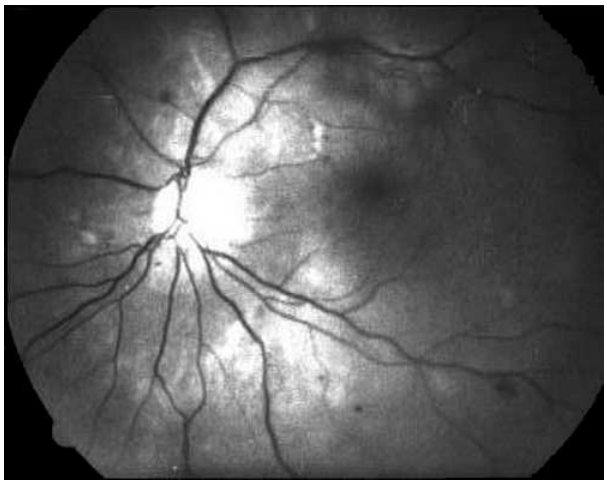


Figure 6.49: Color Fundus Image (CFI)

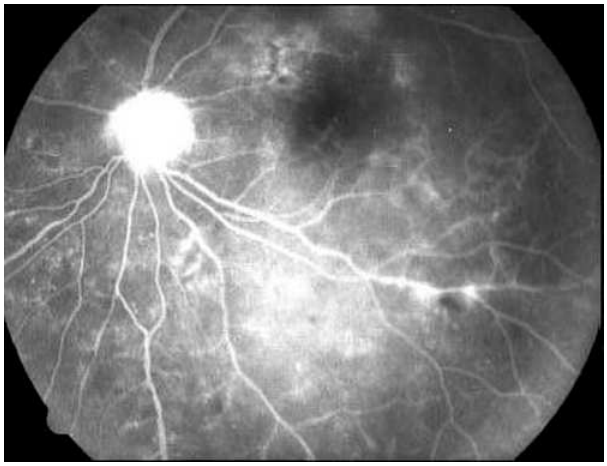


Figure 6.50: Fluroscein Fundus Angiogram FFA)

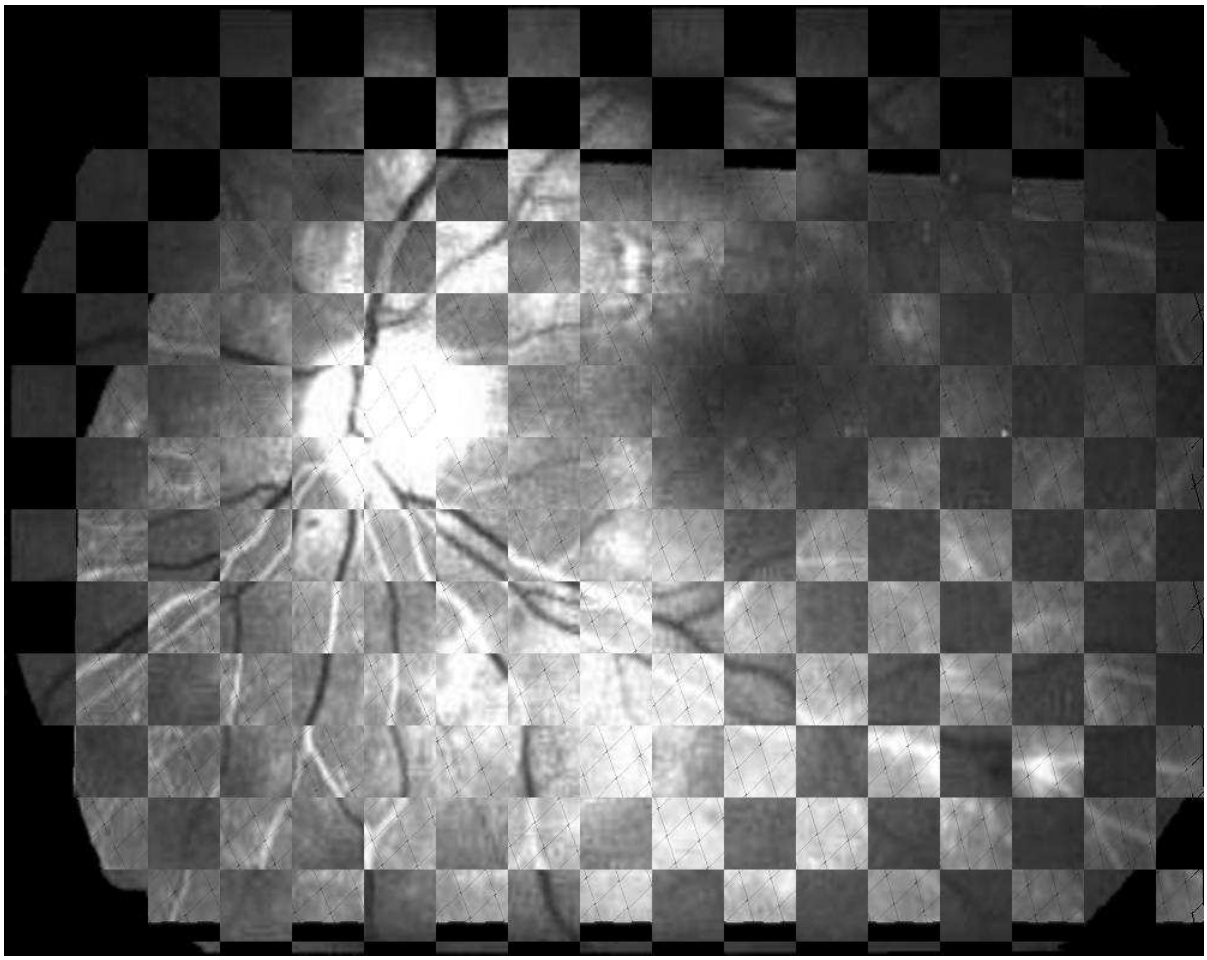


Figure 6.51: Image showing registration of CFI/FFA.

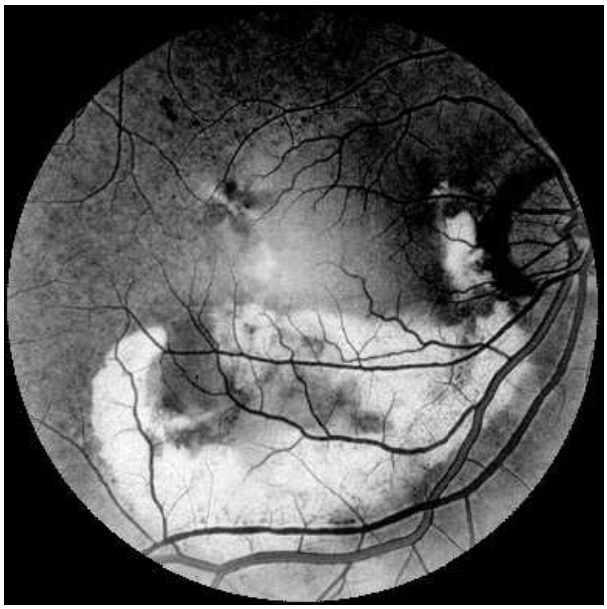


Figure 6.52: Color Fundus Image (CFI)

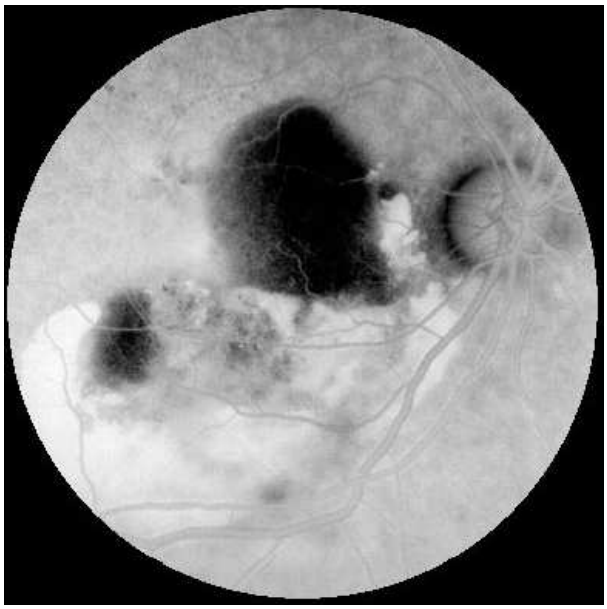


Figure 6.53: Fluroscein Fundus Angiogram FFA)

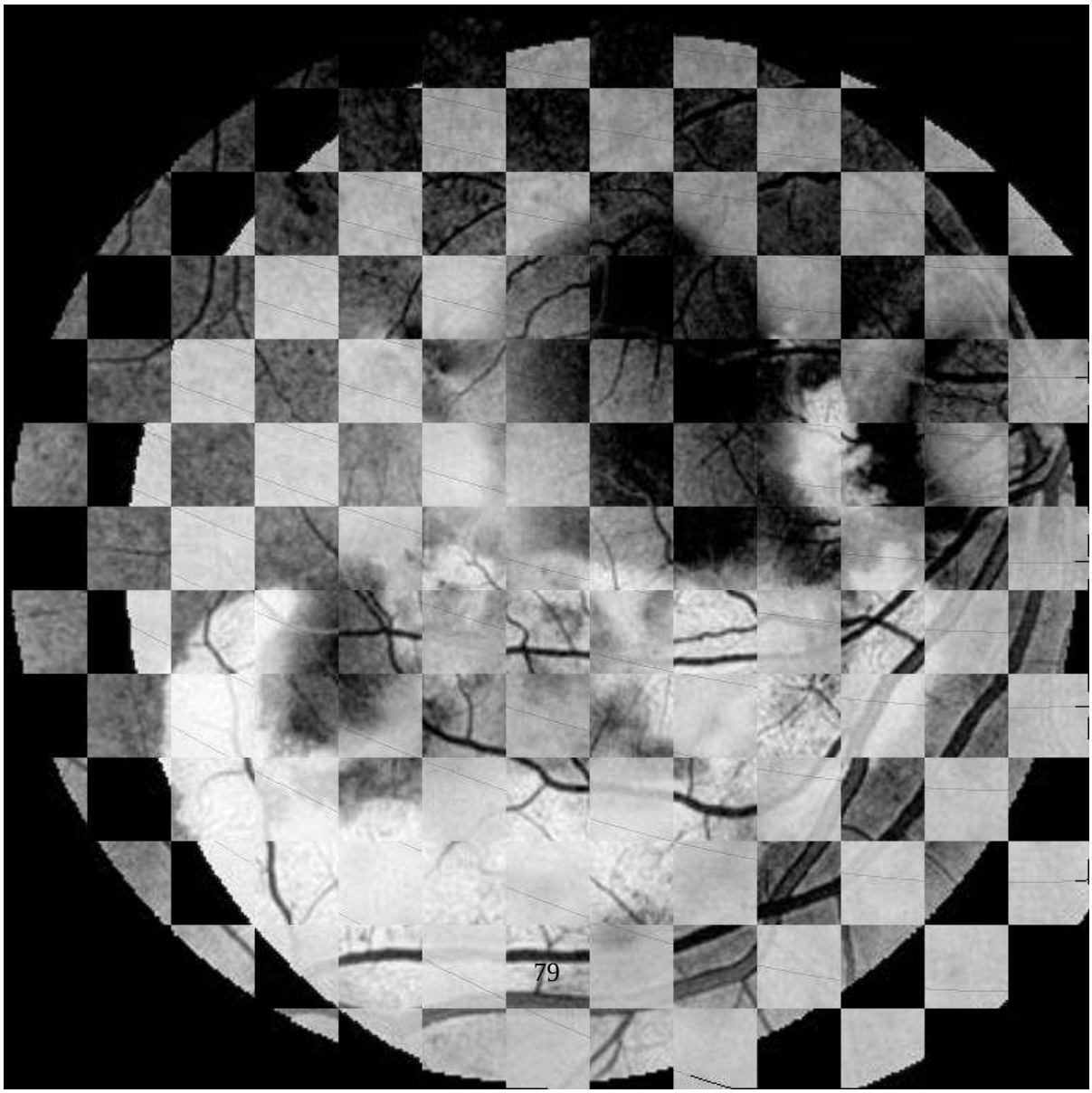


Figure 6.54: Image showing registration of CFI/FFA.

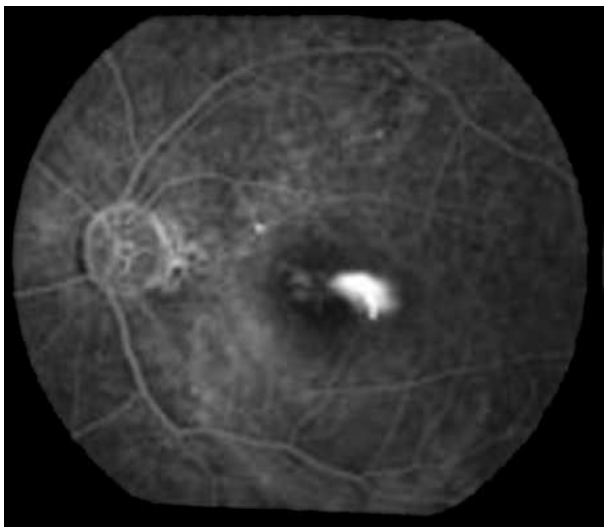


Figure 6.55: Color Fundus Image (CFI)

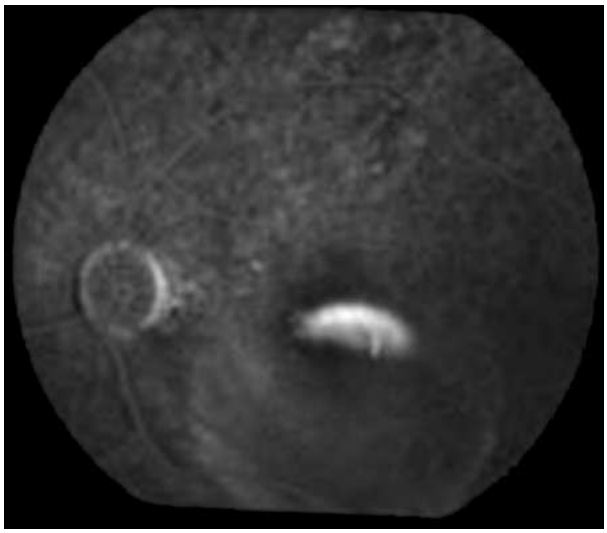


Figure 6.56: Fluroscein Fundus Angiogram FFA)

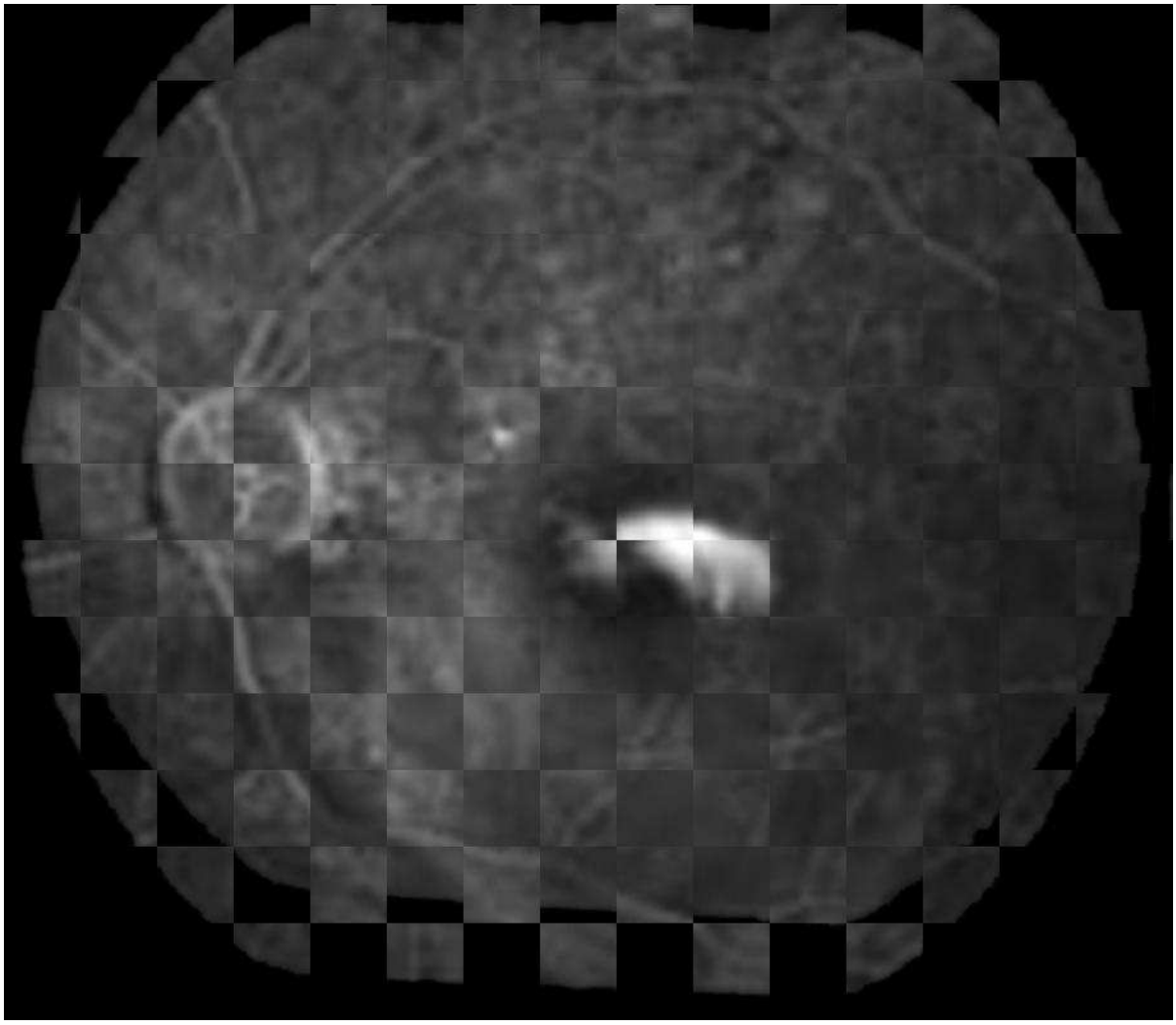


Figure 6.57: Image showing registration of CFI/FFA.

Bibliography

[1] Jiri Jan. *Medical Image Processing, Reconstruction and Restoration Concepts and Methods*. CRC Press, 2006.

[2] Ali Can, Charles V. Stewart, Badrinath Roysam, and Howard L. Tanenbaum. A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina. *PAMI, IEEE Transactions on*, 24(3):347–364, Mar 2002.

[3] Barbara Zitov and Jan Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21:977–1000, 2003.

[4] Tony Lindeberg. *Scale-Space*. 2007.

[5] David J . Hawkes Joseph V . Hajnal and Derek L . G . Hill. *Medical Image Registration*. CRC Press, 2001.

[6] Yuping Lin and Gerard Medioni. Retinal image registration from 2d to 3d. *CVPR*, 2008.

[7] J.B.Antoine Maintz and Max A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1 – 36, 1998.

[8] Lisa Gottesfeld Brown. A survey of image registration techniques. *ACM Comput. Surv.*, 24(4):325–376, 1992.

[9] Jiri Jan. *Medical Image Processing, Reconstruction and Restoration: Concepts and Methods (Signal Processing and Communications)*. CRC, nov 2005.

[10] Frederik Maes, Andr Collignon, Dirk V, Guy Marchal, and Paul Suetens. Multimodality image registration by maximization of mutual information. *IEEE transactions on Medical Imaging*, 16:187–198, 1997.

[11] Yang-Ming Zhu. Volume image registration by cross-entropy optimization. *IEEE transactions on Medical Imaging*, 21:174–180, 2002.

[12] C. Studholme, D. L. G. Hill, and D. J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32(1):71–86, jan 1999.

[13] Daniel B. Russakoff, Carlo Tomasi, Torsten Rohlfing, Calvin R. Maurer, and Jr. Image similarity using mutual information of regions. In *8th European Conference on Computer Vision (ECCV*, pages 596–607. Springer, 2004.

[14] T. Tuytelaars and K. Mikolajczyk. Local invariant feature detectors - a survey. *Foundations and Trends in Computer Graphics and Vision*, 2008.

[15] Peter Kovesi. Phase congruency detects corners and edges. In *in The Australian Pattern Recognition Society Conference: DICTA 2003*, pages 309–318, 2003.

[16] Djemel Ziou and Salvatore Tabbone. Edge detection techniques - an overview. *International Journal of Pattern Recognition and Image Analysis*, 8:537–559, 1998.

[17] Ming-Kuei Hu. Visual pattern recognition by moment invariants. *Information Theory, IRE Transactions on*, 8(2):179–187, feb 1962.

[18] Wikipedia. Blob detection, 2006.

[19] 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. ¡ce:title¿British Machine Vision Computing 2002¡/ce:title¿.

[20] G.C. Stockman A. Goshtasby. Point pattern matching using convex hull edges. *IEEE Transactions on Systems, Man and Cybernetics*, pages 631–637, 1985.

[21] Philippe C. Cattin, Herbert Bay, Luc Van Gool, and Gabor Szkely. Retina mosaicing using local features. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 4191 of *LNCS*, pages 185–192, October 2006.

[22] G.C. Stockman A. Goshtasby. A region-based approach to digital image registration with subpixel accuracy. *IEEE Transactions on Geoscience and Remote Sensing*, pages 390–399, 1986.

[23] R.C. Bolles H.C. Wolf. H.G. Barrow, J.M. Tenenbaum. Parametric correspondence and chamfer matching: Two new techniques for image matching. *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, pages 659–663, 1977.

[24] G. Borgefors. Hierarchical chamfer matching: a parametric edge matching algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 849–865, 1988.

[25] N.D. McKay P.J. Besl. A method for registration of 3d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 239–254, 1992.

[26] R. Chellapa Q. Zheng. A computational vision approach to image registration. *IEEE Transactions on Image Processing*, pages 311–325, 1993.

[27] Paul Viola and William M. Wells, III. Alignment by maximization of mutual information. *Int. J. Comput. Vision*, 24(2):137–154, sep 1997.

[28] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2nd edition, 2001.

[29] Jan Flusser, Barbara Zitova, and Tomas Suk. *Moments and Moment Invariants in Pattern Recognition*. Wiley Publishing, 2009.

[30] M. Helm. Towards automatic rectification of satellite images using feature based matching. *Geoscience and Remote Sensing Symposium, 1991. IGARSS '91. Remote Sensing: Global Monitoring for Earth Management., International*, pages 2439–2442, 1991.

[31] Jan Flusser and Barbara Zitova. Combined invariants to linear filtering and rotation. *Intl. J. Pattern Recognition Art. Intell*, 13:1123–1136, 1999.

[32] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004.

[33] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Comput. Vis. Image Underst.*, 110(3):346–359, 2008.

[34] Thomas M. Lehmann, Claudia Gnner, and Klaus Spitzer. Survey: Interpolation methods in medical image processing. *IEEE Transactions on Medical Imaging*, 18:1049–1075, 1999.

[35] H Asada and M Brady. The curvature primal sketch. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(1):2–14, jan 1986.

[36] D. E. Becker, J. N. Turner, H. Tanenbaum, and B. Roysam. Real-time image processing algorithms for an automated retinal laser surgery system. In *Proceedings of the 1995 International Conference on Image Processing (Vol. 1)-Volume 1 - Volume 1*, ICIP '95, pages 426–, Washington, DC, USA, 1995. IEEE Computer Society.

[37] F Zana and J C Klein. A multimodal registration algorithm of eye fundus images using vessels detection and hough transform. *IEEE Trans Med Imaging*, 18(5):419–28, 1999.

[38] 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, dec 2006.

[39] Robert M. Haralick, Layne T. Watson, and Thomas J. Laffey. The topographic primal sketch. *The International Journal of Robotics Research*, 2(1):50–72, mar 1983.

[40] Alejandro F. Frangi, Ro F. Frangi, Wiro J. Niessen, Koen L. Vincken, and Max A. Viergever. Multiscale vessel enhancement filtering. pages 130–137. Springer-Verlag, 1998.

[41] Saurabh Garg, Jayanthi Sivaswamy, and Siva Ch. Unsupervised curvature based retinal vessel segmentation, 2005.

[42] IA US) Van Ginneken Bram (Utrecht NL) Niemeijer Meindert (Utrecht NL) Abrmoff, Michael D. (University Heights. Automatic detection of red lesions in digital color fundus photographs, January 2009.

[43] Stare project website. Stare dataset.

[44] Drive dataset- digital retinal images for vessel extraction.

[45] Yossi Rubner, Carlo Tomasi, and Leonidas J. Guibas. A metric for distributions with applications to image databases. In *Proceedings of the Sixth International Conference on Computer Vision*, ICCV '98, pages 59–, Washington, DC, USA, 1998. IEEE Computer Society.

[46] Neil Ryan, Conor Heneghan, and Philip de Chazal. Registration of digital retinal images using landmark correspondence by expectation maximization. *Image and Vision Computing*, 22(11):883 – 898, 2004.

[47] Charles V. Stewart, Chia ling Tsai, and Badrinath Roysam. The dual-bootstrap iterative closest point algorithm with application to retinal image registration. *IEEE Trans. Med. Img*, 22, 2003.

[48] Yogesh Babu Bathina, M. V. Kartheek Medathati, and Jayanthi Sivaswamy. Robust matching of multi-modal retinal images using radon transform based local descriptor. In *Proceedings of the 1st ACM International Health Informatics Symposium*, IHI '10, pages 765–770, New York, NY, USA, 2010. ACM.

[49] J. Chen, J. Tian, N. Lee, J. Zheng, R. Smith, and A. Laine. A partial intensity invariant feature descriptor for multimodal retinal image registration. *IEEE Trans Biomed Eng*, 2010.

[50] G. K. Matsopoulos, N. A. Mouravliansky, K. K. Delibasis, and K. S. Nikita. Automatic retinal image registration scheme using global optimization techniques. *IEEE Trans. Inform. Technol*, 1999.

[51] Nicola Ritter, Robyn Owens, James Cooper, Robert H. Eikelboom, and Paul P. Van Saarloos. Registration of stereo and temporal images of the retina. *IEEE Transactions on Medical Imaging*, 18:404–418, 1999.

[52] Eli Peli, Reed A. Augliere, George, and T. Timberlake. Feature-based registration of retinal images. *IEEE Trans. Med. Imaging*, pages 272–278, 1987.

[53] G. Reynolds P. Nagin, B. Schwartz. Measurement of uorescein angiograms of the optic disc and retina using computerized image analysis. *Opthalmology*, 1982.

[54] M.S. Markow, III Rylander, H.G., and A.J. Welch. Real-time algorithm for retinal tracking. *Biomedical Engineering, IEEE Transactions on*, 40(12):1269 –1281, 1993.

[55] Christopher Blauth Peter L. C. Smithy Kenneth M. Taylor Roger Jagoe, John Arnold and Richard Wootton. Retinal vessel circulation patterns visualized from a sequence of computer-aligned angiograms. *Investigative Opthalmology and Visual Science*, 1993.

[56] Axel Pinz, Stefan Berngger, Peter Datlinger, and Andreas Kruger. Mapping the human retina. *IEEE Transactions on Medical Imaging*, 17:606–619.

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

[58] M. D. Abramoff S. Lee and J. M. Reinhardt. Feature-based pairwise retinal image registration by radial distortion correction. *Proc. SPIE Conf. Medical Imaging*, 6512, 2007.

[59] Michal Base Radim Kolar, Viktor Sikula. Retinal image registration using phase correlation. *Biosignal*, 2010.

[60] Jian Zheng Xing Zhang Xiaoqian Dai Kexin Deng, Jie Tian and Min Xu. Retinal fundus image registration via vascular structure graph matching. *International Journal of Biomedical Imaging*, 2010.

[61] Avi Kelman, Michal Sofka, and Charles V. Stewart. Keypoint descriptors for matching across multiple image modalities and non-linear intensity variations. In *CVPR,2007. IEEE Conference on*, pages 1–7, June 2007.

[62] Jie Tian Jian Zheng, Kexin Deng Yakang Dai, and Jian Chen. Retinal image registration based on salient feature regions. *IEEE EMBS*, 2009.

[63] Gehua Yang, Charles V. Stewart, Michal Sofka, and Chia-Ling Tsai. Registration of challenging image pairs: Initialization, estimation, and decision, 2007.

[64] Tony Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, Norwell, MA, USA, 1994.

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

[66] Salvatore Tabbone, Oriol Ramos Terrades, and Sabine Barrat. Histogram of Radon Transform. A useful descriptor for shape retrieval. In *ICPR*, Tampa United States, 2008.

[67] Euclidean normalised similarity measure. http://www.lans.ece.utexas.edu/ strehl/diss/node53.html.

[68] P. H. S. Torr and A. Zisserman. Mlesac: a new robust estimator with application to estimating image geometry. *Comput. Vis. Image Underst.*, 78:138–156, April 2000.

[69] M. Zuliani. Ransac toolbox for matlab, Nov. 2008.

[70] Vessel tracer and GDBICP executable. http://www.cs.rpi.edu/ sofka/downloads.html.

[71] Gimp - the gnu image manipulation program.

[72] Yan Ke and R. Sukthankar. Pca-sift: a more distinctive representation for local image descriptors. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–506 – II–513 Vol.2, june-2 july 2004.

[73] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. *J. Opt. Soc. Am. A*, 4(3):519–524, Mar 1987.