Integrated Computational System for Portable Retinal Imaging

by

Jason Boggess

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2012

Author .....
Department of Electrical Engineering and Computer Science
May 23, 2012

Certified by .....
Ramesh Raskar
Associate Professor
Thesis Supervisor

Accepted by .....
Professor Leslie A. Kolodziejski
Chair, Department Committee on Graduate Students

Integrated Computational System for Portable Retinal Imaging

by

Jason Boggess

Submitted to the Department of Electrical Engineering and Computer Science
on May 23, 2012, in partial fulfillment of the
requirements for the degree of
Master of Science in Computer Science

Abstract

This thesis introduces a system to improve image quality obtained from a low-light CMOS camera—specifically designed to image the surface of the retina. The retinal tissue, as well as having various diseases of its own, is unique being the only internal tissue in the human body that can be imaged non-invasively. This allows for diagnosis of diseases that are not limited to eye conditions, such as diabetes and hypertension. Current portable solutions for retinal imaging such as the Panoptic and indirect ophthalmoscopes require expensive, complex optics, and must be operated by a trained professional due to the challenging task of aligning the pupillary axis with the optical axis of the device. Our team has developed a simple hardware/software solution for inexpensive and portable retinal imaging that consists of an LED light source, CMOS camera, and simple LCD display built into a pair of sunglasses. This thesis presents a multistage solution that registers the retinal tissue on the sensor; identifies its shape and size; performs local integration with phase correlation; and performs global integration with panoramic mosaicing. Through this process we can increase signal to noise ratio, increase image contrast, create super-resolution images, and obtain a large field of view image of the retina. We also lay the groundwork for possible 3D reconstruction to increase the amount of information present. The main contributions of this thesis are in computational methods for improving overall image quality, while other team members focused on the illumination source, optics, or other hardware components.

Thesis Supervisor: Ramesh Raskar
Title: Associate Professor

Acknowledgments

I would first and foremost like to thank my parents for all their love and support throughout the years. I never would have made it this far in life without their continued help. My extended family from Iowa, Texas, and the surrounding areas have always been there for me with laughter, love, and comfort whenever I needed it. I really appreciate all the support and dedication to my success throughout the years.

I really admire my peers in the Camera Culture Group of the Media Lab. Doug has always amazed me with his 3D skills, and Gordon is a charismatic leader and understanding individual. Otkrist has been a great friend and companion and always there to grab ice cream when you need him. Nikil has the craziest laugh, and while I may not have known Aydin or Kshitij for long I know they will be successful in all their future endeavors.

I could not have even begun this work without our team members. Everett Lawson, Research Assistant, originally came up with the project years ago and never dreamed it could turn into what it is today. Other members Siddharth Khullar (Visiting PhD student from RIT), Alex Owal (Post-Doctoral fellow), Gordon Wetzstein (Post-Doctoral fellow), and Ramesh Raskar (Advisor) have been instrumental in making this project a success.

Technical Contributions. This research is made possible through the MIT Media Lab, and specifically the Camera Culture Group whose facilities and darkrooms were used extensively to capture data from the eyes. Ramesh Raskar was a key player in finding funding opportunities for the team.

Funding Acknowledgments. Funding was provided by the Media Lab Corporate sponsors and by the National Science Foundation for Jason Boggess's graduate fellowship. Prize money was also obtained through the CIMIT Primary Healthcare Prize and the Ideas Global Challenge for this research. Google provided us with a fundus camera which was instrumental in doing our research.

Thank you to all the family members who helped proofread this thesis and who will be attending my graduation ceremony.

Contents

List of Figures

Chapter 1

Introduction

1.1 Research Team

The team of researchers were instrumental in making this research project a success. Everett Lawson was a key player who motivated the team, thought about illumination challenges, subjected himself to many experiments and focused primarily on building hardware, buying cameras, lighting to develop prototypes, and wiring LEDs. Di Wu also helped significantly in her background research, designing computer algorithms for multimodal capture, and preparing hardware for data capture. Jennifer Wang, an undergrad we worked with for several weeks, aided the team doing validation on retina motion. Ramesh Raskar, our advisor initiated the project and contributed heavily to the direction of the project.

Siddharth Khullar, Alex Olwal, and Gordon Wetzstein were great members of the group helping us write and publish papers, and we could not have done it without the clinicians in the Boston area such as Caio Regatieri and Paolo Antonio Silva. Our clinicians were instrumental in our understanding of retinal diseases, their detection, and the validation of our hardware.

Ramesh Raskar joined the Media Lab from Mitsubishi Electric Research Laboratories in 2008 as head of the Media Lab's Camera Culture research group. His research interests span the fields of computational light transport, computational photography, inverse problems in imaging and human-computer interaction. He was a

head researcher on NETRA, refractive error assessment on a cellphone, and CATRA, cataract assessment on a cellphone, and continues to lead research in eyecare development in retinal image at the MIT Media Lab.

Vitor Pamplona was also a key member of the team, designing NETRA and CATRA, and laying the foundation for the retinal imaging work.

Some of the text in this thesis, especially in the related work section, was used in a recent paper submission and underwent editing by other members of the team.

1.2 Low-Light Imaging

1.2.1 Thesis Statement

I present a computational system designed to take images captured in very low lighting conditions, extract the most meaningful data from these images, and assemble them back together in a meaningful way. By using some recent developments in image translation identification, I am able to provide a comprehensive, robust, computational system that recursively integrates images to increase the signal-to-noise ratio in the data. My multistep process is able to take a database of images and autonomously reconstruct them into a panorama without user input.

1.2.2 Challenges in Computational Processing

While the enhancement of retinal images may seem straightforward, photographs captured from our system are unique when compared to natural images. As the eye moves, only a narrow window of the retina is seen through the pupil on the image. This requires registration and identification. Moreover lens distortions such as vingetting and normal aberrations cause the edges of the data area to become dark and distorted. Traditional computational photography systems rely on feature matching to detect motion between video frames, for the creation of fused image stacks, super-resolution, and panoramic stitching. However, under our low signal-to-noise ratio feature matching is not able to accurately detect features, and thus

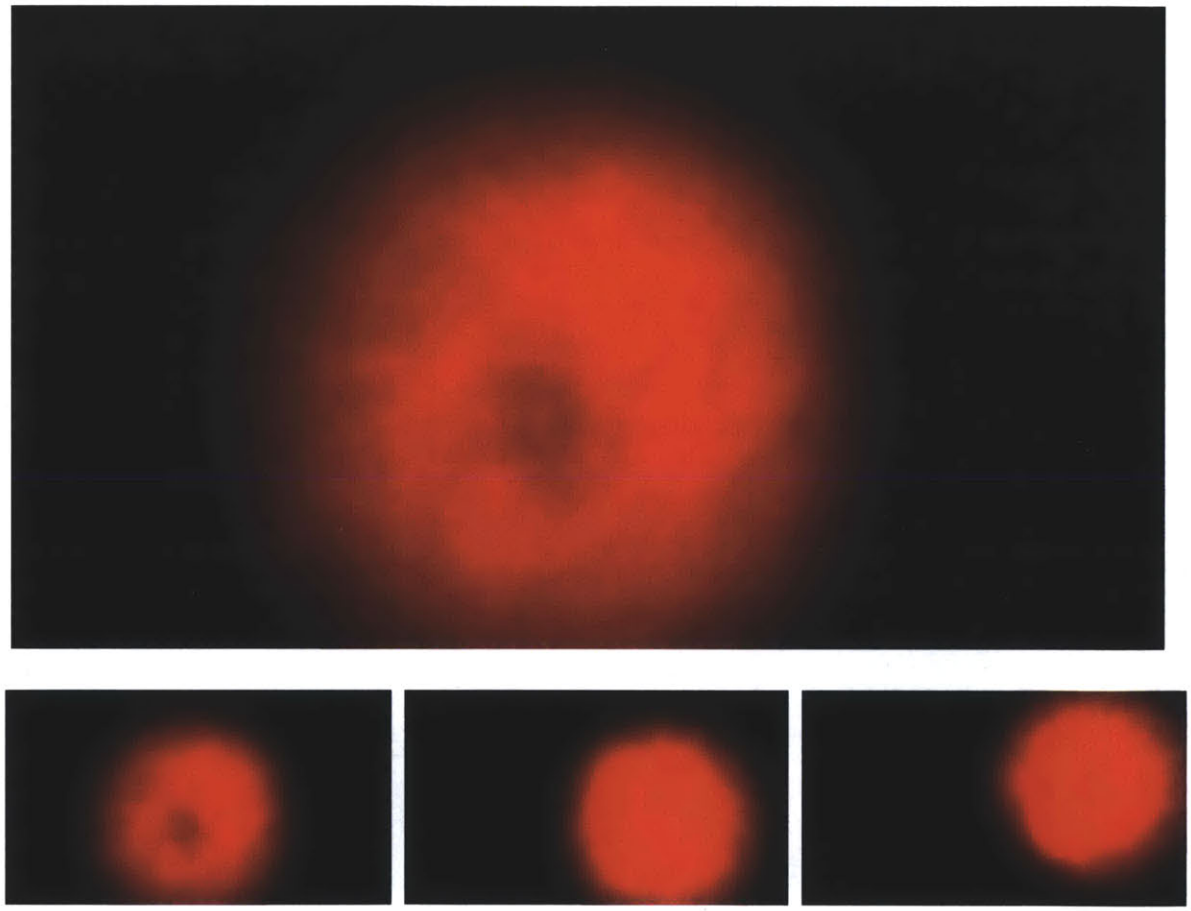


Figure 1-1: Raw data as obtained from our camera system.

reconstruction is impossible with this technique (See Figure 1-1).

1.2.3 Contributions

Specifically for this thesis I will concentrate on the following issues.

- Registration and recognition of the retina in a captured image, including the position and size of the retinal tissue on the sensor.
- Correction for lens falloff and distortion, due to vignetting and other optical interference.
- Time multiplexing and integration of frames to reduce noise and lens distortions.
- Production of super-resolution images from their low-resolution counterparts.
- Panoramic stitching to produce a large field of view from several smaller field of view images.
- Integration on mobile platforms to capture data portably and efficiently.
- Framework for three-dimensional representations of the retina using multiple lighting techniques.

1.3 Background

1.3.1 The Human Eye

The human eye is one of the most dynamic and important features of the human body. Giving us our strongest sense, the eye gives meaning to our environment and helps us develop spacial reasoning. Without our eyes, we would have never discovered outer space, or climbed Mount Everest, or seen the Grand Canyon. There are many tasks that require eyes to perform such as driving a car, that would be nearly impossible to do without sight.

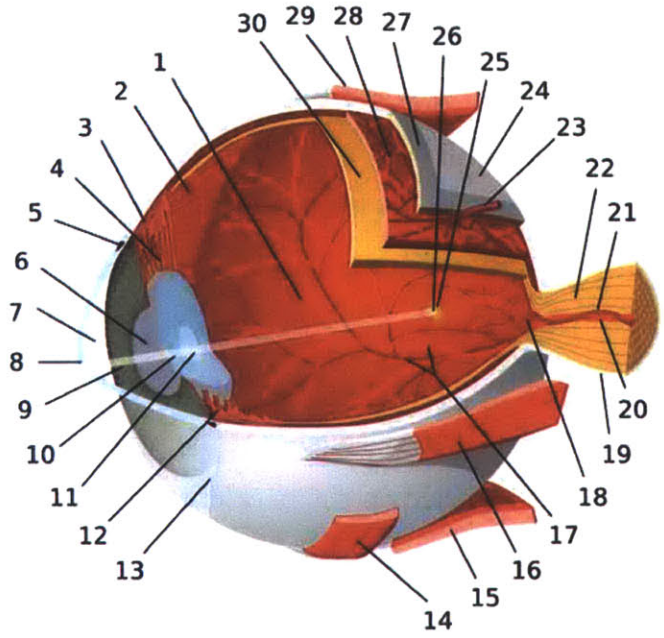


Figure 1-2: The Human Eye. Notice that the light must pass through the pupil into the vitreous humor and finally diverge into the retina. 1. posterior compartment - 2. ora serrata - 3. ciliary muscle - 4. ciliary zonules - 5. canal of Schlemm - 6. pupil - 7. anterior chamber - 8. cornea - 9. iris - 10. lens cortex - 11. lens nucleus - 12. ciliary process - 13. conjuntiva - 14. inferior oblique muscule - 15. inferior rectus muscule - 16. medial rectus muscle - 17. retinal arteries and veins - 18. optic disc - 19. dura mater - 20. central retinal artery - 21. central retinal vein - 22. optical nerve - 23. vorticose vein - 24. bulbar sheat - 25. macula - 26. fovea - 27. sclera - 28. choroid - 29. superior rectus muscule - 30. retina

The human eye is a one inch diameter sphere contained in a tough, white shell called the sclera. Light from the environment strikes the cornea at the front-most part of the eye, transfers through the pupil, enters the internal portion of the eye called the vitreous cavity, and is then focused through the lens of the eye onto the retina, a photosensitive tissue [29].

Like the aperture of a camera, the iris opens and closes to adjust the amount of light that enters the eye through the pupil. This function is aided by the use of ganglian cells which respond to the levels of incident light on the eye. The reaction time of the ganglian cells is critical in keeping the retinal tissue from absorbing too much light, a condition known as phototoxicity [29].

The center of the eye is filled with a dense, clear, viscous fluid called the vitreous humor which also contributes to the reflective nature of the light entering the human eye. The average focal length of the eye is between 17 and 24 mm, and most scientific models describe the eye as a perfect sphere. If the eye is too short or too long the lens has trouble focusing the image on the back of the retina, causing what is known as refractive error. Refractive error can be easily corrected with the aid of eyeglasses.

The retina is a complex tissue lining the back of the eye. The retina transfers the incoming light through photoreceptors to the optic nerve which then carries the signal to the lower rear of the brain for processing. Blood vessels and nerves appear on the exterior surface of the retina and vitreous humor. The vessels concentrate to a single bundle at place called the optic disc which is the brightest part of the retina appearing as a white circle in most retinal images.

The center of focus of the human eye is called the fovea [19]. The fovea is the darkest spot appearing almost as a dent on the surface of the retina. Surrounding the fovea is an area called the macula which spans between the optic disc and two large arteries on either side of the fovea. Most retinal diseases manifest themselves in the macula or in the area around the optic disc.

The retina makes use of photoreceptor cells to transmit incoming light into a signal that can be processed in the brain. There are two types of these photoreceptor cells called rods and cones corresponding to the shape of the photoreceptor. Rods

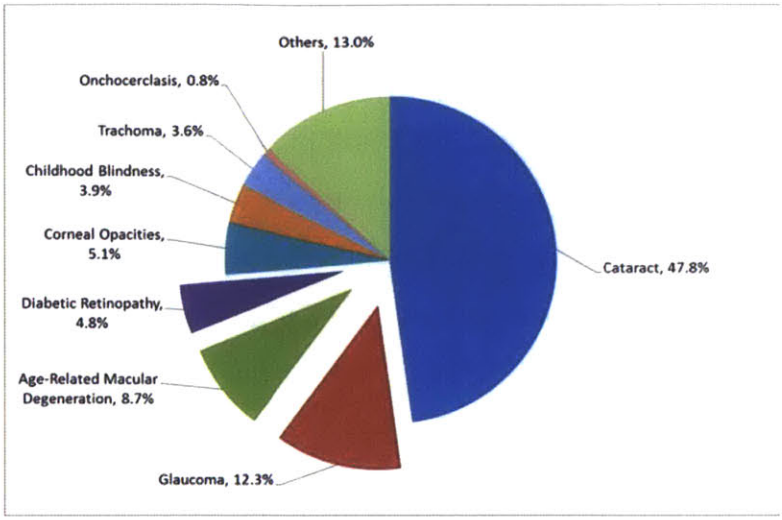


Figure 1-3: Causes of blindness in the world as of the year 2002. Notice that AMD, Glaucoma, and Diabetic Retinopathy can all be diagnosed through retinal images

are long and skinny and are most receptive to dim light. The signal transmitted from rods is interpreted in the brain as black-and-white vision. While there are many rods in the center region of vision and in the macula, there are more rods than cones in the peripheral areas of vision.

There are three different types of cones that respond to three different wavelengths of light. Mostly concentrated in the center of the macula, cones transmit red, green, and blue light information to the brain to be processed. Beyond the retina is the choroid layer which is a muscle group responsible for positioning and moving the eye. In some of our captured images, the choroid can be seen through thin retinal tissue.

1.3.2 Need for Retinal Imaging

In 2010, the World Health Organization(WHO) estimated that 40 million of the 6.7 billion people the world were completely blind and another 264 million of them had low vision [30]. Out of these nearly 300 million people, those who are 50 years old or older have an 82% chance of being blind. Most visual impairment comes from the lack of eyeglasses, something which the Camera Culture Group of the MIT Media Lab addressed in 2009 with their cellphone attachment NETRA [31]. The next major cause of visual impairment according to the WHO is untreated cataracts

which accounts for 33% of the visual impairment cases. The Camera Culture Group and the Media Lab also had a device to measure cataracts on a cell phone, called CATRA, which presents users with a picture of the cataract map of their lens [33]. The primary cause of all blindness globally is cataracts (See Figure 1-3).

Cataracts and uncorrected refractive error are not the only causes of vision loss. The next three major causes of blindness are glaucoma, age-related macular degeneration, and diabetic retinopathy [27]. According to the University of Nebraska Medical Center, Department of Ophthalmology and Visual Sciences, every 5 seconds someone in the world loses the ability to see forever. However, 80% of blindness-causing diseases are preventable and/or can be cured to near 100% vision restoration [30, 27]. The number of blind people worldwide is expected to double by the year 2020 [30].

Blindness is not distributed equally throughout the world with more than 90% of the world's visually impaired living in developing countries [30]. Southeast Asia (20%) and Western Pacific (26%) have the highest incidence of avoidable blindness, followed by Africa (17%), Eastern Mediterranean (10%), and America (10%).

Vision 2020 is a group that is dedicated to ending preventable blindness by the year 2020. They estimate that the limitations to eye care largely revolve around the lack of a cost-effective programs for screening and detection of diseases such as diabetic retinopathy [28]. While the prevalence of cataracts can be assessed using portable light field displays [31, 33], other diseases such as glaucoma, age-related macular degeneration, and diabetic retinopathy, require fundus images, or images of the central retina, to detect. Many imaging solutions (discussed in the next section) exist to adequately image the retina, but they are too large, expensive, and bulky to meet the needs of the developing world.

Dr. Caio Regatieri and friend of the MIT Media Lab is a leading researcher and ophthalmologist in the Boston area with specific expertise in retinal diseases. As a Brazil native, he is actively associated with eye screening programs and organizations in his home country. In a recent interview, he mentioned there are very few fundus cameras (one of which he owns) for the entire country of Brazil— 200 million people.

Age-related macular degeneration (AMD) is the leading cause of blindness in de-

veloping countries, and in industrialized countries causes half of all blindness. It results in a progressive loss of central vision and is detected by visualizing the retina. As advancements and access to health services lead to increases in average life expectancies around the world, we will observe a global rise in the elderly population, making AMD a greater problem for both the developed and developing world. A major WHO strategy for preventing blindness due to AMD is to increase diagnosis and referral for treatment.

Diabetic Retinopathy is the leading cause of blindness in American adults and disproportionately affects ethnic minorities. The WHO notes that "treatment of established retinopathy can reduce the risk of visual loss by more than 90%". As per Center for Disease Control (CDC) - "50% of 25.8 million diabetics are not getting their eyes examined or are diagnosed too late for treatment to be effective in the U.S." Many do not have access to screening with the U.S. alone facing a shortage of 6,000 ophthalmologists by 2020.

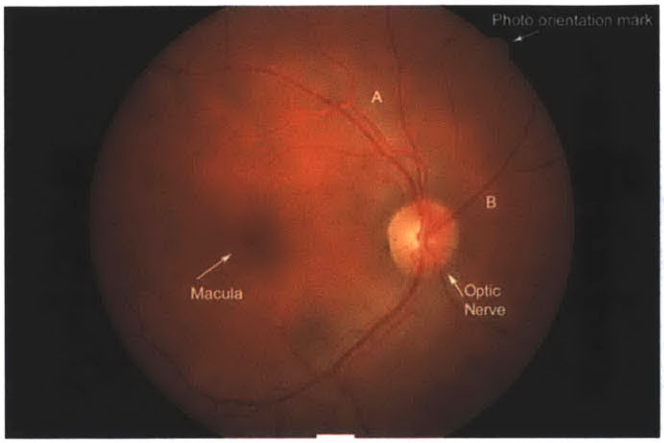
Other diseases affect the retinal surface that are not solely eye conditions. Hypertension (high blood pressure) and diabetes can sometimes appear in the eye months and possibly years before they show up in blood tests. Recently, it has been revealed that retinal imaging may even detect brain tumors and high cholesterol [15]. As retinal imaging technology becomes more available and common, even more diseases may be found to manifest themselves in retinal tissue (See Figure 1-4).

1.4 Scope and Limitations

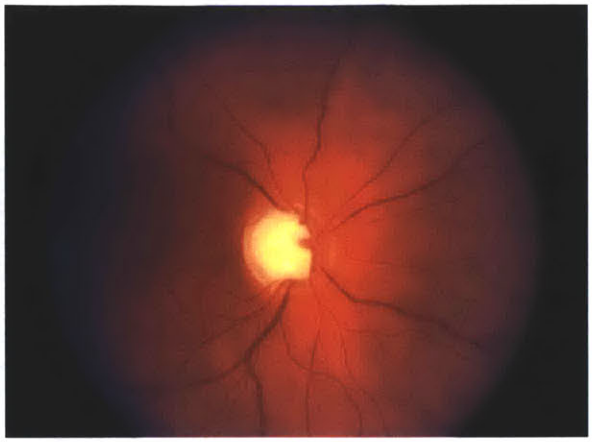
While we aim to provide retinal screenings and health care to people in the world in many different places, this work is just the beginning of a long line of retinal imaging publications. Our team goals are as follows:

- Develop a computational system to significantly reduce the need for fundus cameras.

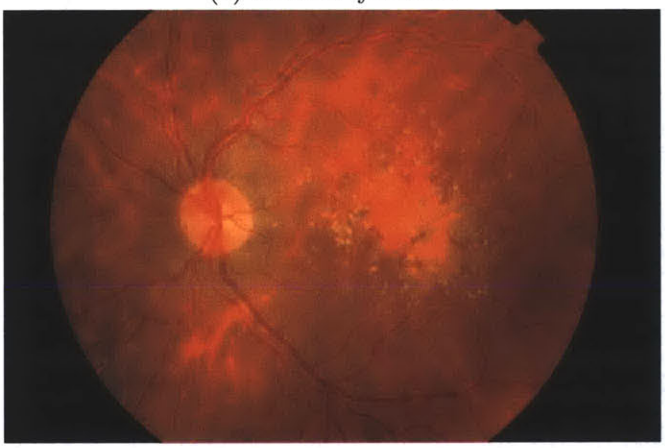
- Integrate our system with existing hardware technologies (such as mobile de-



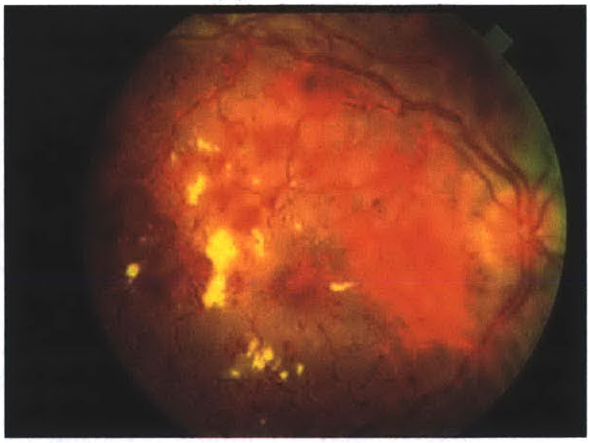
(a) A Healthy Retina



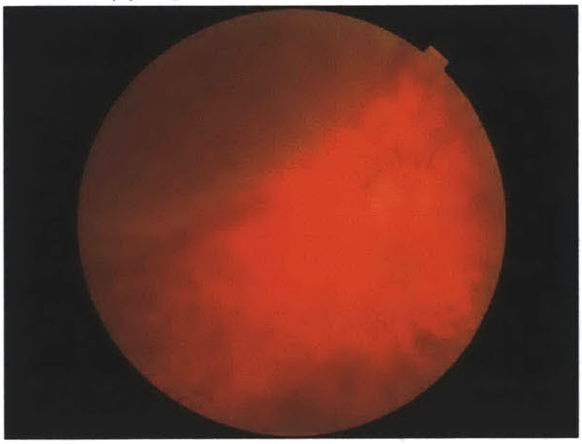
(b) Glaucoma



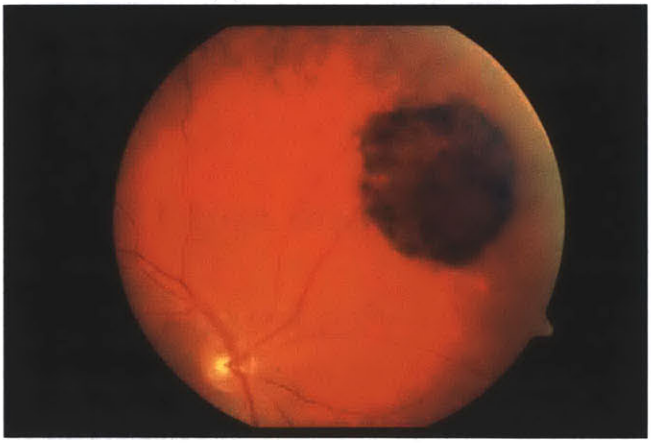
(c) Age-Related Macular Degeneration



(d) Diabetic Retinopathy



(e) Retinal Detachment



(f) Retinal Tumor

Figure 1-4: Comparison of the various Retinal Diseases. These images were captured using a standard fundus camera

vices) to increase portability across different venues of clinical scenarios.

- Enable individuals to operate the system without the need for a trained professional.

- Integrate imaging with systematic disease detection/prevention capabilities through computational algorithms and system integration (with hospital databases). Such an imaging solution will facilitate early detection of diabetes, hypertension, and specific brain abnormalities that manifest themselves first in the retina.

Limitations

- The computational system presented in this thesis while robust may not incorporate all data captured or result in an optimum use of the available data. The online algorithm makes use of a clustering heuristic to speed up processing and may miss frames interspersed between noisy data.

- Ideal processing will be real-time, but this system currently does not operate at the rate of capture. Future optimization and hardware improvements could make this a possibility.

- Certain restrictions to the position and scale are exercised in order to increase precision with low feature images. Specifically, the phase correlation of the image processing assumes images have been translated, but not rotated or scaled.

Image quality, resolution, noise and fixed depth of field are limitations of the CMOS sensor used in our system, which currently allows us to resolve structures down to $45\mu m$. There is a trade-off between the field of view and the camera's proximity to the eye. Fundus cameras achieve high magnification and large field of view in a single photograph without proximity constraints [11].

Chapter 2

Retinal Imaging

Some of the text and figures in this section were developed as part of a team effort for a paper submission.

2.1 Challenges in Retinal Imaging

	Current Methods	Our Method
Lighting	Direct, Xenon Bulb / Light Tubes	Indirect, Light Emitting Diode
Alignment	External Hardware, Expert Clinician	Interaction via Binocular Coupling
Large FOV	Complex Optics	Image Mosaicing
Focus Blur	Complex Optics	Small Aperture Camera
Motion Blur	Strobe Lighting / Short Exposure	Lucky Imaging

Table 2.1: Computer graphics can exploit modern technologies to simplify devices, e.g. for retinal imaging.

While retinal imaging has emerged as one of the most powerful and rapidly developing ways to detect both eye-specific diseases and other bodily ailments, the barrier to retinal imaging is still high. The main challenges in obtaining a good retinal image are as follows.

Lighting. The retinal tissue is designed to absorb light and is enclosed within the sclera tissue, making it difficult to image. Traditional devices use bright xenon bulb flashes through the pupil to sufficiently illuminate the retina. In order to take an adequate photograph of the internal eye, we must illuminate the retina effectively

and efficiently. With reduction in cost of ultrabright LEDs and portable light sources, we hope our indirect illumination approach will inspire other illumination techniques for existing devices.

Alignment. In order to accurately capture an image of the retina, light from the eye must travel linearly through the pupil and down the optical axis of an imaging device before it can be observed by the clinician. Aligning two lines in three-space so they are coincident is a very difficult task, even for a trained professional. Commercial fundus photography requires non-trivial alignment, presence of an operator, and a controlled environment, such as a clinic or hospital. Coupled illumination and stimulus control, as discussed later in this work, can overcome alignment challenges.

Field of View. While most current retinal imaging systems can function without the aid of mydriatic (dilation) eye-drops, there still exists a trade-off between field of view, and period of discomfort for the subject. The pupil is a narrow opening into the eye roughly about 4 mm in diameter, and CMOS camera sensors from digital SLRs are in the range of 20 to 30 mm in diameter. Therefore, only a small portion of the camera sensor will contain images of retinal tissue, unless complex relay optics are added to magnify the image of the retina before it hits the camera sensor. The handheld ophthalmoscope can image about a 10-15° field of view. The tabletop fundus cameras and scanning laser ophthalmoscopes can achieve about a 30° field of view in one photograph. However, both of these techniques rely on complex optics. Our system uses mosaicing to post-process the images and achieve a larger field of view. We ensure smooth, wide coverage using binocular coupling.

Focus Blur. The lenses in the human eye and camera must be focused in conjunction for sharp images of the retina. Traditional setups use complex alignment techniques, while our approach makes use of a small aperture and large depth of field.

Motion Blur. Long exposure times in retinal imaging yield photographs with motion blur, and short exposure times yield high signal-to-noise ratio. Traditional methods use short exposures and bright lights, but our methods compensate for motion blur by using the "Lucky" imaging approach, taking many images and selecting

only the best [23].

2.2 Current Imaging Solutions

There are three main types retinal imaging solutions doctors utilize in making retinal diagnostics: 1) Handheld Ophthalmoscopes (direct and indirect), 2) Fundus cameras, and 3) Laser rasterizing systems such as Scanning Laser Ophthalmoscopes (SLO) and Optical Coherence Technology (OCT).

2.2.1 Handheld Ophthalmoscopy

The smallest and most inexpensive group of ophthalmoscopes were invented by Charles Babbage in 1847 [39, 21]. These devices rely solely on optics to obtain an adequate view of the retina, similar to a telescope for the eye. These devices do not allow for image capture or postprocessing. The image produced by a direct ophthalmoscope is virtual and direct, but the examination must take place as close to a person's eyes as possible to increase the field of view.

The indirect ophthalmoscope [18] requires a condensing lens which is normally in the range of about 50 diopters. Attending physicians will wear a headmounted optical set up and hold the condensing lens at arm's length in front of the patient's face. Since the image focused from the condensing lens is projected out into real space, the entirety of the retina can be seen by both eyes simultaneously. This allows for stereopsis, or the ability to see 3-D structures on the retina. This feature is not present in direct ophthalmoscopy, which only can result in a 2-D image.

The most common, handheld device used by doctors today in developing regions of the world is an indirect ophthalmoscope because it offers the clearest picture to the attending physician. While both direct and indirect ophthalmoscopes operate on refractive principles, indirect ophthalmoscopy can only see up to the Ora serrata (one third of the way around the eye) while direct ophthalmoscopy can see slightly beyond the equator (halfway around the eye).

During the operation of handheld ophthalmoscopes, light is shone in through the

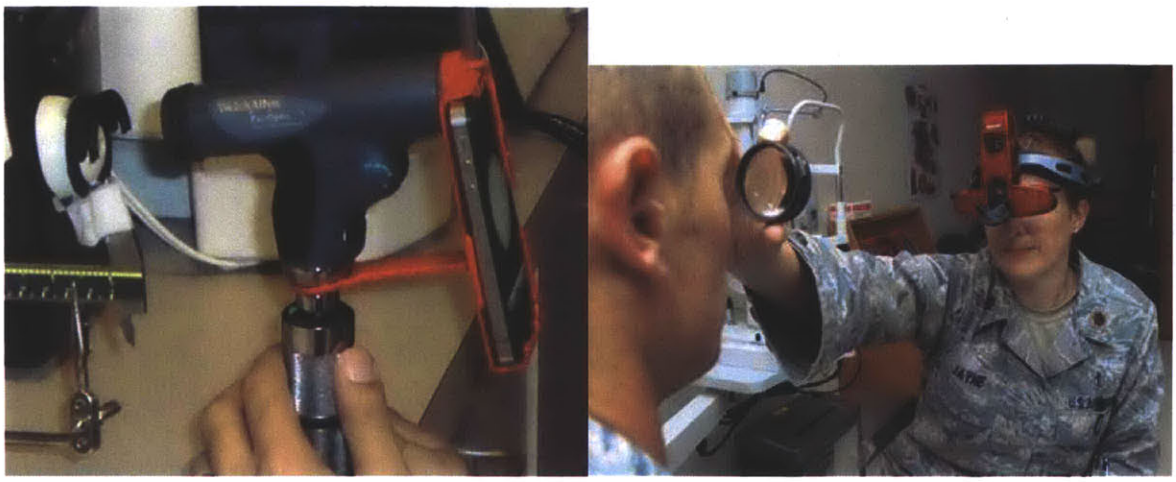
pupil to the rear of the eye where it is reflected back towards the ophthalmoscope. However, bright light shone in the human eye causes the iris to constrict around the pupil limiting the field of view seen by the doctor. This can be mitigated with mydriatic, dilating, drops that cause the iris to relax and the pupil to dilate. Dilation allows for a larger field of view, but reduces vision in the patient for several hours following the examination.

For most of these devices, there is no direct way to capture an image and therefore doctors are required to be present for the examination. Recently there have been developments of many attachments that can be affixed to the end of a direct ophthalmoloscope to record such an image. The iExaminer [1] is an iPhone mount that allows the iPhone camera to take a picture through the direct ophthalmoloscope. While you are able to record many pictures on your iPhone using the iExaminer, only a small field of view is captured without the aid of mydriatic drops. Such pictures are almost useless in diagnosing diseases. (Figure 2-2)

The cost of these devices is large despite their small size. A direct ophthalmoscope itself costs around $500. In indirect ophthalmoscopy, the condensing lens alone costs around $500. A binocular helmet used in conjunction with the indirect ophthalmoscopy can cost $1000-$3000. The iExaminer attachment for the iPhone costs $150 in addition to the direct ophthalmoscope for a total cost of over $600 (see Figure 2-1).

2.2.2 Fundus Imagery

A fundus camera is based on the principle of monocular indirect ophthalmoscopy, but allows the retinal image to be captured on a conventional SLR camera [20] or smaller CMOS camera [14], as opposed being seen by a doctor directly. The biggest advantage over handheld ophthalmoscopes is being able to image the eye without mydriatic drops and still obtain a large field of view image of the retina. This is achieved with a combination of rigid alignment procedures, relay optics, and quick flash illumination that does not cause the eye to constrict until after the photo has been taken. Most fundus cameras have robust alignment mounts with chin and forehead rests ensuring the patient stays in place during the the procedure which



(a) (b)

Figure 2-1: (a) A direct ophthalmoscope with the iPhone 4 attached to a Panoptic with a red plastic called the iExaminer. (b) A worn example of an indirect ophthalmoscope (Courtesy of the Pacific Airforce, http://www.pacaf.af.mil/)

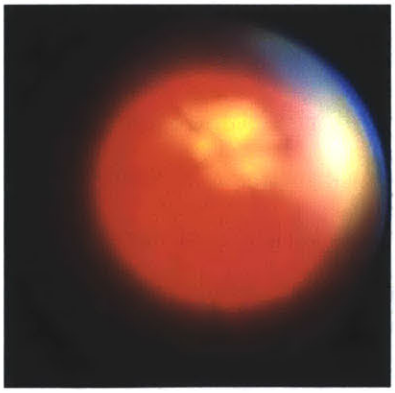


Figure 2-2: An image captured through the Panoptic, a direct ophthalmoscope

Figure 2-3: A fundus camera photographs the retina using complex optics and a standard SLR camera. These devices are easier to align than handheld devices and do not require mydriatic drops to operate.

reduces the amount of motion blur in the image.

Fundus cameras have another advantage in alignment whereas by shining infrared light into the person's eye, alignment can be done optically before the picture is taken. Since the invisible infrared light does not cause the pupil to constrict, this allows for a wider field of view without the need for mydriatic drops.

Light shone into the eye during fundus imagery passes through two dissimilar paths: one into the pupil illuminating the retina, another out of the pupil from the retina [2]. This minimizes scattering and corneal reflections. Illumination through the sclera has also been addressed in [35], but has never been used in any real-world application.

While robust and efficient, fundus cameras are still large and costly. Tabletop fundus cameras cost around tens of thousands of dollars and require separate cameras that must be attached for image capture. Furthermore, since these devices have large alignment mechanisms such as forehead rests, chin rests, and alignment knobs, they

can weigh 50 pounds and be larger than a small refrigerator. In developing regions of the world where access to health care is limited, access to one of these machines is difficult (See Figure 2-3 for a fundus camera).

2.2.3 Scanning Laser Ophthalmoscopy and Optical Coherence Tomography

Scanning laser ophthalmoscopes (SLOs) operate on a completely different principle than fundus imagery or standard ophthalmology. Instead of relying on conventional imaging and optics, scanning laser ophthalmoscopy uses confocal laser scanning microscopy to image the retina in a rasterized fashion. This provides sharper contrast images that allow for better diagnosis of glaucoma, macular degeneration and other disorders [40].

Invented in the late 1980s, SLO technology is the newest form of ophthalmology. A collimated laser beam is bounced off a beam splitter and sent into the eye. As the light returns from the retina it interferes with the original beam in a meaningful way to produce a high contrast image of the retina.

Scanning laser ophthalmoscopy also requires dilation drops, to keep the eye from focusing during the procedure. Accommodation due to eye focus would drastically change the magnification and shape of the data returned as the laser rasterizes through the eye.

SLOs can cost more than a quarter of $1 million (See Figure 2-4).

2.3 Our Solution

Our solution attempts to simplify the five challenges in retinal imaging by reducing the complex optics used in a conventional photograph and replacing them with a computational backend (See Table 2.1). Adequate lighting is achieved, not through the pupillary axis of the eye, but through the external tissue next to the eye via cool to the touch, low power light emitting diodes. We call this technique indirect

(a) A standard Scanning Laser Ophthalmoscope.

(b) Using the SLO in our research lab.

Figure 2-4: Scanning Laser Ophthalmoscopes (SLO).

Figure 2-5: We have developed a simple solution that trades complex optics for computation in order to take low quality images and extract the most data from them.

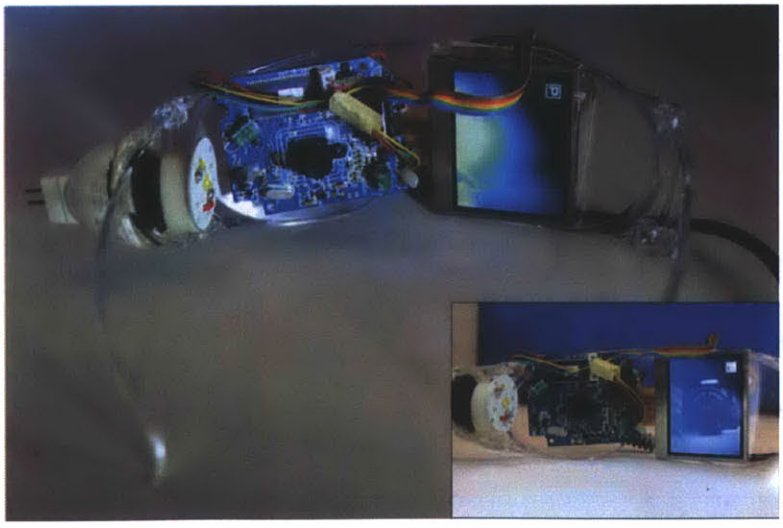


Figure 2-6: Our glasses consist of an LED light source and inexpensive CMOS camera

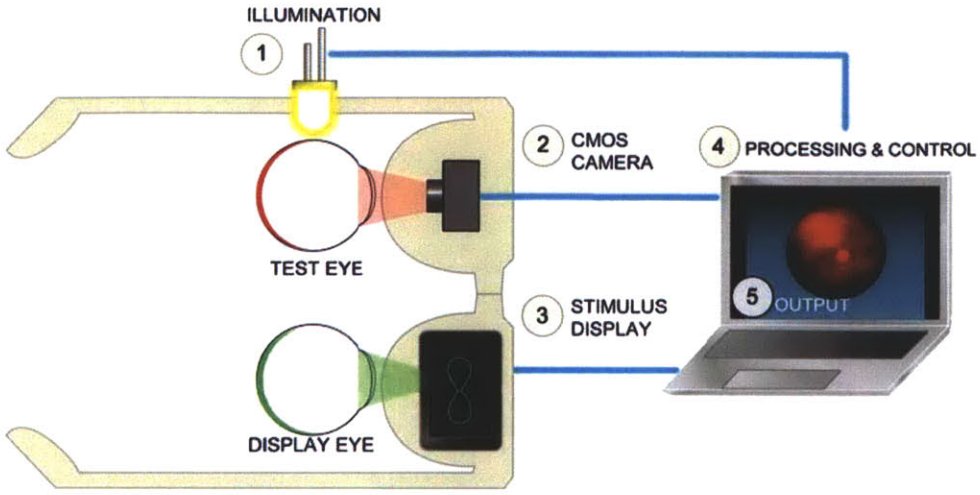


Figure 2-7: An LCD display allows the user to control one eye with the other to obtain a large field of view of the retina. Images captured can be digitally processed to greatly enhance the quality.

	Our Device	Handheld Ophthalmoscopes	Table-top Fundus Cameras	SLO and OCT
Cost	<$50	$500+	$5,000+	$100,000+
Self-capture	Yes	No	No	No
Ease of Alignment	High	Low	Moderate	High
Corneal Reflections	None	Yes	Yes	Moderate
Field of View	High	Low	High	Moderate
2-D resolution	Moderate	Low	High	High
Image Quality	Moderate	Moderate	High	High
Optical Simplicity	High	Moderate	Low	Low

Figure 2-8: A comparison of various retinal imaging devices.

diffuse illumination, for its transmissive properties through the surrounding tissue. We make use of small CMOS cameras with an aperture size less than the pupil size which increases the captured field of view and removes issues associated with focus blur.

Motion blur alignment issues are mitigated through computation and the lucky imaging approach [23]. The paradigm of traditional fundus photography is to capture one amazing image, while our philosophy is to capture many images and take the best subset of them, by selecting high quality images through computational procedures. Even though motion blur may be present in some of our images, chances are there is at least a subset that are crisp and sharp, which we can then integrate together computationally.

The entire form factor of our prototypes can fit into a standard eye-worn device.

Recent developments in technologies make our hardware setup possible. Super bright, low-power LED light sources have only become become available in the recent history. The proliferation of cameras into our cellphones have not only made the cameras small, but also high resolution and constantly evolving. The high prevalence of mobile devices into our society allows us to compute things cheaply, and effectively, on-demand anywhere in the world.

We have developed a pair of safety goggles that contain an LCD screen on one side and a camera on the other, pointed inward towards the eyes. The user puts on the device, turns on the light, and instantly sees a picture of his or her retina on the LCD screen. This allows for self-alignment without the need for a trained clinician. A stimulus is presented on the LCD screen for the eye which is not being tested to follow. By using gaze lock we can use one human organ to control the test eye, in a process known as binocular coupling. As the free eye moves around following the stimulus control, the test eye follows in parallel space while we are capturing data from the retina. Images can be subsequently mosaiced together into a large panorama.

We contribute the following to the realm of retinal imaging:

- **Optical Simplification.** We observe that small-aperture imaging systems, possible today with micro cameras, allow for reduction in optical constraints

and simplify alignment procedures.

- **Indirect Diffuse Illumination.** Realization of an unexpected avenue to provide illumination to the retina via emerging solid state light sources such as LEDs with a unique form factor that can be placed near the skin.

- **Self-imaging via Binocular coupling.** Novel user feedback mechanism to dramatically simplify alignment and avoid the need for wide field of view imaging where coupled gaze allows control of one eye by providing stimulus to the other eye while recording images with a narrow field of view camera. As far as we know this is the first near eye system to use coupled gaze control.

- **Image Composition.** Instead of highly engineered optics to overcome challenges such as alignment, blur, field of view, and timing, we use a computational photography approach to fuse images together from a varying-quality image sequence.

2.4 Related Image Processing Work

2.4.1 Retinal Imaging and Light Transport

The eye is a key component for translating the external world to comprehensive representations. In recent years, computer graphics researchers have shown an increased interest in exploring the human visual system. Computational models describing cone distribution [10], anatomy [38], pupil light reflex [32], and light transport [22], have been formulated to improve the performance of various graphics and display applications. Modeling real-world objects and their corresponding reflective, and refractive properties has especially been of great interest. Rendering the anterior segment, especially the iris has been widely explored for a great variety of applications such as modeling realistic video game characters and surrounding objects [8, 24, 25, 36, 41].

2.4.2 Low-light Imaging and Integration

Various methods have been used to reduce noise in images, some using a single image, and others using a composite of many images. The Wiener Filter [42] tries to reduce noise by using an iterative method that first applies a filter and then measures the error from the resulting image. The Wiener Filter was the state of the art filter until [37] which used a nonlinear invariant for noise removal.

Phase correlation has been used in video to stabilize capture [13] due to its speed and accuracy. The work done by [16] shows that phase correlation can be used at a subpixel level which makes super-resolution possible.

super-resolution has been explored in [34] which describes the process by which a higher resolution image can be created with multiple images of lessor resolution. The method created by MIT Prof. Freeman [17] gives an example-based approach to super-resolution by first creating a database of real-images and then matching an upscaled image to natural formations in images. However, this method is not appropriate for medical applications, as many images captured could contain diseases that we would not want superimposed into high-resolution images. Techniques as described in [34]are used in this thesis to create super-resolution images.

The work done by Microsoft researchers in [44] conquers both de-noising and de-blurring in imaging conditions that contain low light (e.g. the retina).

2.4.3 Retinal Mosaicing

There has been much previous work done in panoramic stitching and blending, sometimes called mosaicing in multiple dimensions. [3, 4] was one of the first comprehensive panoramic stitching algorithms available. A program called Hugin [9] builds on the panoramic tools to provide exposure correction and other advanced features.

The paper [7] describes a robust method of retinal mosaicing. For multi-modal image registration, we use a more robust method which is widely used in retinal imaging related research. Those methods described by [43] can deal with small portion of overlapping, and varying illumination images, a robust method for registration.

Chapter 3

Algorithms and Methodology

3.1 Processing Overview

3.1.1 Overview of Limitations of Captured Data

To capture images of the retina, we developed a hardware setup consisting of an LED light source (or fluorescent light tube) and a CMOS webcam. This allows us to capture many images of the retina sequentially in video format and utilize data across the temporal domain. However, each individual frame captured is not as high of quality as you would find in a typical fundus camera image due to the simplified optics and low lighting. As our design exchanges the complex optics and lighting for computation, the simplified images returned from the camera require processing in order to show in a single image as much data as on a typical fundus camera.

Image Ailments

- **Fractional Sensor Saturation**. Light leaving the eye through the pupil is concentrated in a circular cone which strikes the camera sensor in a particular manner, producing a red sphere on a black background. However, occasionally light leakage around the corner of the eye will cause a background to be non-black. Specifically, the areas around the sides of the retina are often a blurry white, pink, or red. In order to extract the most data that we can from the

sensor, we must identify the part of the sensor that contains the most data.

- **Vignetting**. Retinal images incident on the sensor appear more like spheres than flat planes. The lens curvature of the eye creates a cosine falloff, or vignetting, near the edges of the data. Therefore, not only must the region with the most data be determined, but a circular fit must be performed on the data to correct for vignetting. Color channels must also be balanced.

- **Low Signal-to-Noise Ratio**. Due to the lowlight nature of the eye, captured images suffer from high gain and, as a result, a low signal-to-noise ratio. This makes it difficult for doctors and nurses to identify diseases that exist at low levels of contrast. For instance, the presence of diabetic retinopathy can sometimes show up as yellow dots on the surface of the eye. Without a high contrast image, we cannot accurately detect the small lesions.

- **Partial Field of View**. Each frame captured by our retinal imaging camera contains about a $20 - 30^\circ$ field of view. However, due to the vignetting and other lens distortions, only the central $10 - 15^\circ$ of the field of view contains useful data. We would like to capture a full 180° field of view of the retina to see all the way from fovea to the equator on either side of the eye interior.

(See Figure 1-1 and Figure A-1 for examples of raw data)

3.1.2 Processing Outline

To address the aforementioned problems, we developed a four-step solution to derive the most data from retinal images.

1. **Image Registration & Correction.** Captured images are analyzed for retinal tissue. Once the circular region of the retina is identified, anti-vignetting is applied to the region to reduce lens effects and create a "flat" image. A selection is made from the central region of the retina circle as a region of interest. Only frames which contain a full view of the retina and are in sharp focus are used. Other frames are discarded.

2. **Frame Clustering.** Once images have been vignetting corrected and a region of interest has been identified, corrected images are clustered into groups of similar images to be integrated locally. An online clustering algorithm is used in conjunction with phase correlation to detect similar images.

3. **Local Integration.** Noise reduction and super-resolution creation are done by integrating similar images together locally on top of one another. Phase correlation is used to align images that may have shifted during capture. This is much faster and more accurate on our data set than traditional feature matching.

4. **Panoramic Stitching.** Integrated frames have reduced noise, higher resolution, and increased contrast. These processed frames can now be stitched together into a large field of view mosaic using standard panorama techniques.

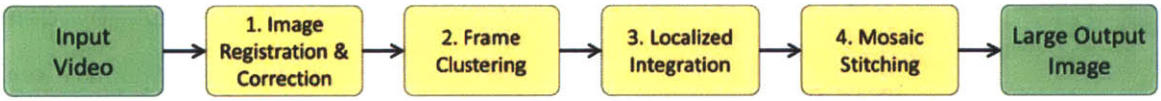


Figure 3-1: Flowchart of the image processing system.

3.2 Image Registration & Correction

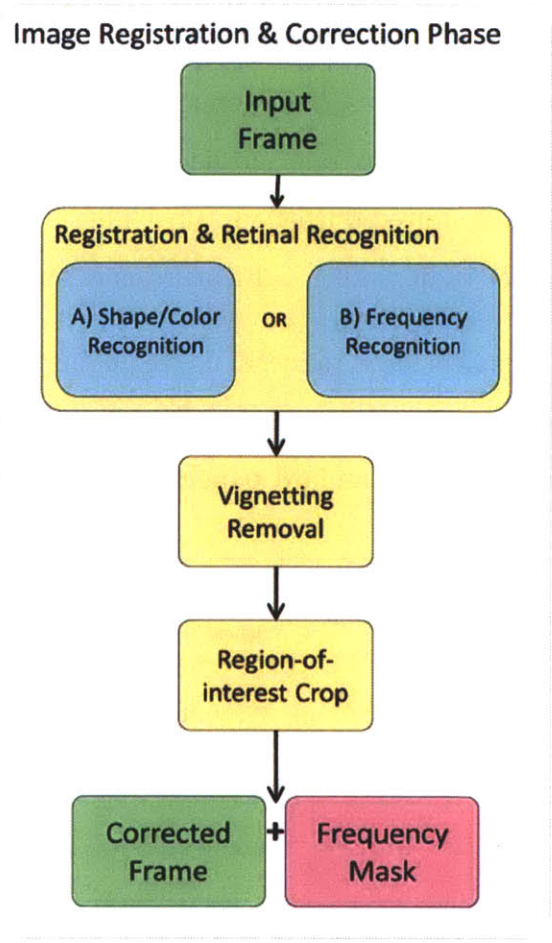


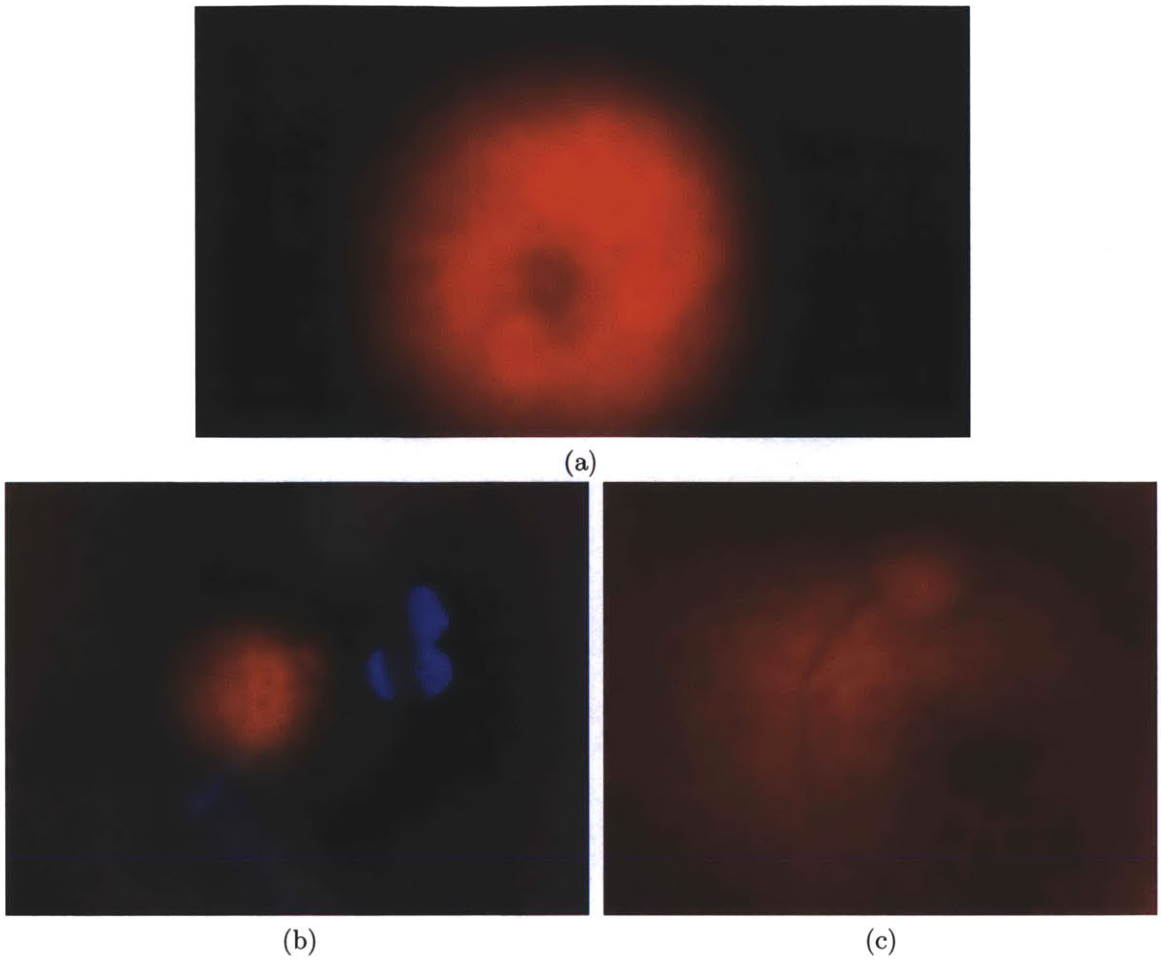
Figure 3-2: Image Registration & Correction

When images are captured from our setup, even though the retina (as through the pupil) is the primary object of focus, sometimes parts of the iris and sclera are apparent in the resulting image (See raw data in Figure 1-1 and Figure A-1). Generally since these objects are close to the camera, they appear largely out of focus, and often times very dark/black. This allows us to easily identify the part of the image sensor that contains the retina, by doing a simple thresholding operation identifying the region of the circle that is the brightest.

However, at times, light from our LED light source leaks around the sclera and the iris causing regions around our area of interest on the sensor to glow. Thus, a straightforward algorithm to identify the region of interest on the image sensor presents a challenge (See Figure 3-3 for examples of input data). We present a robust algorithm to work through a large set of captured images.

3.2.1 Frequency Recognition

We try to identify the region of the image that we are interested in by performing a mid-band filter on the image, zeroing the highest and lowest frequencies. The highest frequencies in the image represent the noise, of which there is a lot due to the low



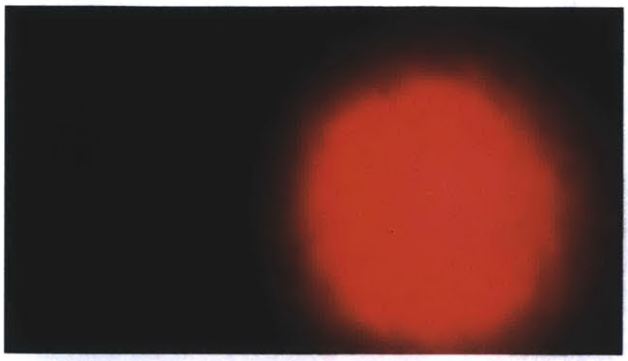
(a)

(b) (c)

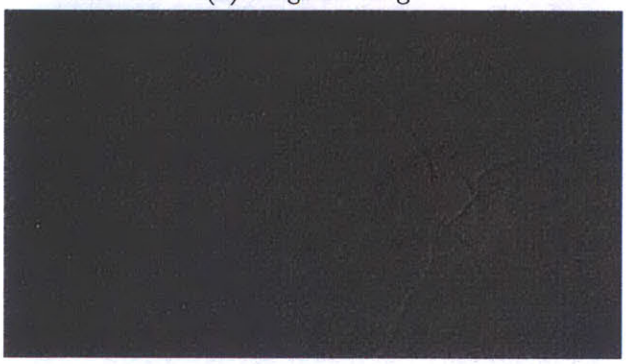
Figure 3-3: a) Often times only the retina appears on the sensor as a red spherical shape on a black background. The areas of interest are easy to detect with a simple thresholding operation. b) Occasionally there are other image artifacts such as corneal reflections or c) light leakage from around the iris that make retinal tissue registration in the image more challenging. For the latter two conditions, a frequency detection algorithm must be used.

amount of light in the eye. Low frequencies correspond to the colors in the image and the parts without much detail. The iris and sclera surrounding the pupil are extraordinarily close to the camera, before the focal length of miniature optics and so appear blurred when photographed (are made up of low frequencies). Therefore, a mid-pass filter results in an image with structures that are neither noisy nor blurry.

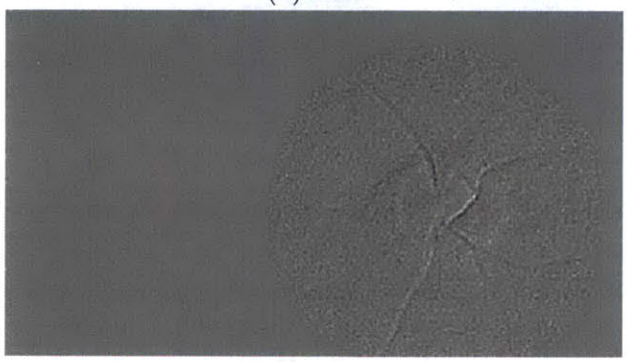
The mid-band filter of the image will give us a frequency response across the picture. Figure 3-4 shows results of mid-band filtering per image color channel (Blue, Green, and Red respectively). Notice the frequency responses in each color channel



(a) Original Image



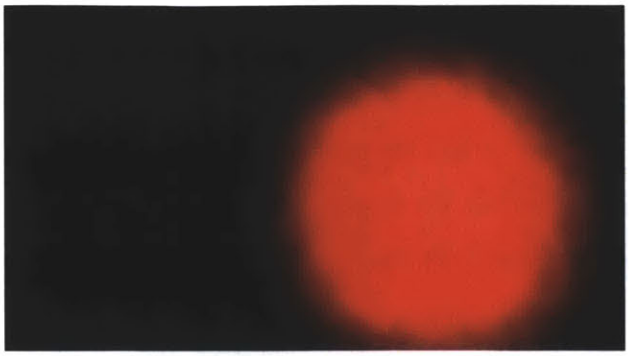
(b) Blue



(c) Green

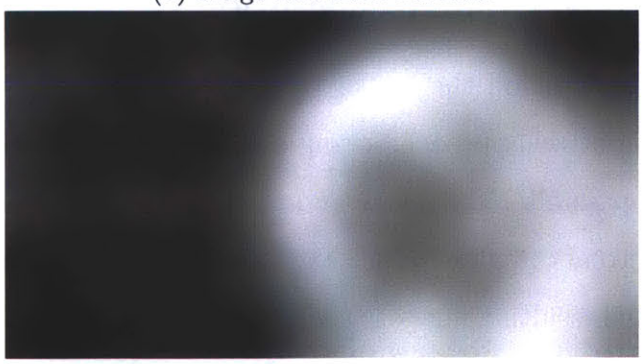
(d) Red

Figure 3-4: The resulting frequency response for the blue, green, and red color channels respectively of a retinal image. Notice that even though noise is apparent, the retinal tissue stands out against the black background



(a) Original Image

(b) Image mid-band Filtered



(c) Filtered Image Blurred

(d) High Response Regions

Figure 3-5: a) When the color channels are averaged and then low-pass filtered, we get the above filtered response. Notice that the area of the image we are interested in stands out against the dark background. b) the same image, but with a threshold response. Notice again that the computer can clearly identify which parts are retina and which are not.

greatly highlight the locations of the blood vessels.

While these images show a nice per pixel response of the frequencies present in the image in the mid-band range, they do not give us an overall picture of larger areas of interest. To correct for this, we apply a low-pass filter to the image to identify the larger regions in which there is the most contrast. A threshold operation results in a binary image clearly showing the region of the retina (Figure 3-5).

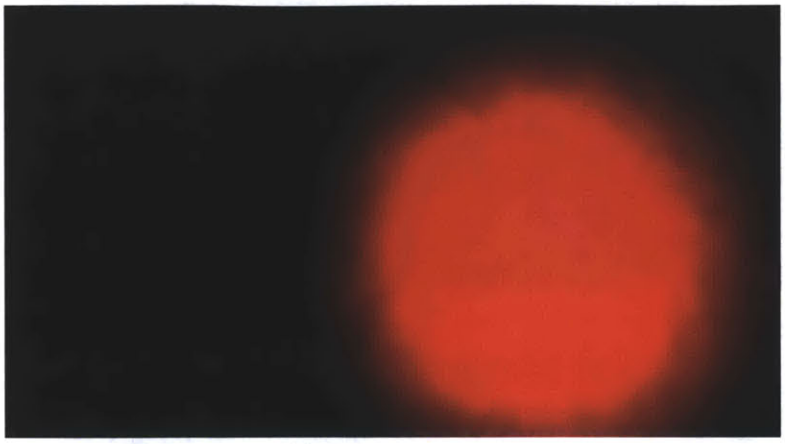
Under certain conditions this method might fail. For instance, part of the retina may not contain any vascular structures or have high-contrast regions. Examples include when the camera is over the fovea of the eye, which just contains a dark dimple and no vascular structures, or in patients with advanced diabetic retinopathy who have pools of blood in their eyes. In both of these cases there may not be a lot of detail upon which this algorithm relies.

3.2.2 Shape and Color Recognition

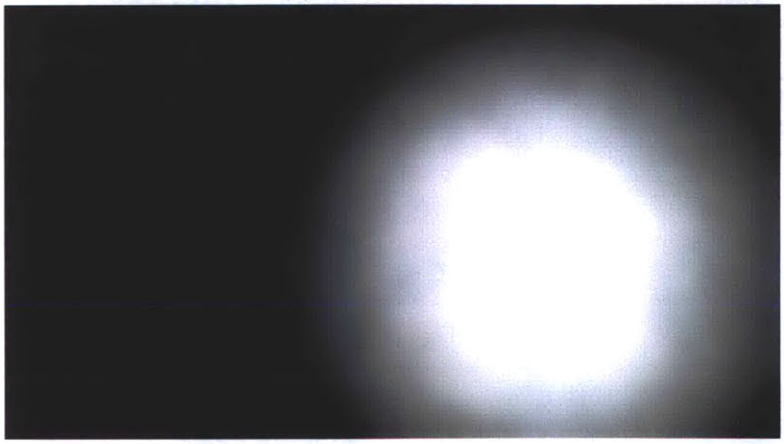
By assuming all the light on the sensor comes from the retina, we can identify the circular region with other approach. Using a similar technique to the one mentioned above, we can low-pass filter the image and then take a threshold, once again making the lower values zero and higher values one to create a binary image.

In both cases, the resulting image is a white circle on a black background indicating the region in which the retina is likely to be. To correctly reverse vignetting (the next step in the process), we must accurately identify not only the location of the circle but its size as well. Calculating the moments of the image and determining the center of mass location may fail when the circular region is partially occluded by the boundary of the image. As the eye moves around following a stimulus control this failure case tends to happen quite frequently (See Figure 1-1).

To get the accuracy we need, we will identify the boundary of the circle and then perform a circle regression. Identification of the boundary is rudimentary. A random sampling of edge points is selected and for every pair a perpendicular bisector is calculated. Then, for every two perpendicular bisectors, an intersection point is calculated and a candidate for the center of circle is created. By clustering candidates



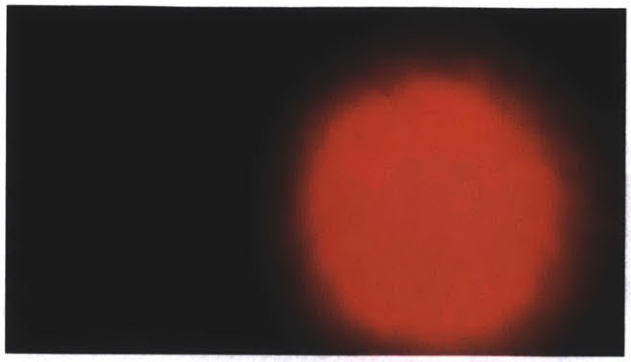
(a) Original Image



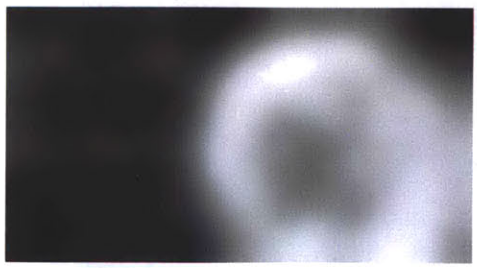
(b) Image Blurred

(c) Threshold

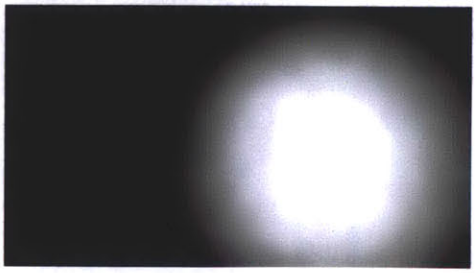
Figure 3-6: We can isolate the red channel and take a threshold to identify the area of interest in the image. Note, this only works for red circular regions on a black background, but the region appears extremely circular.



(a) Input data



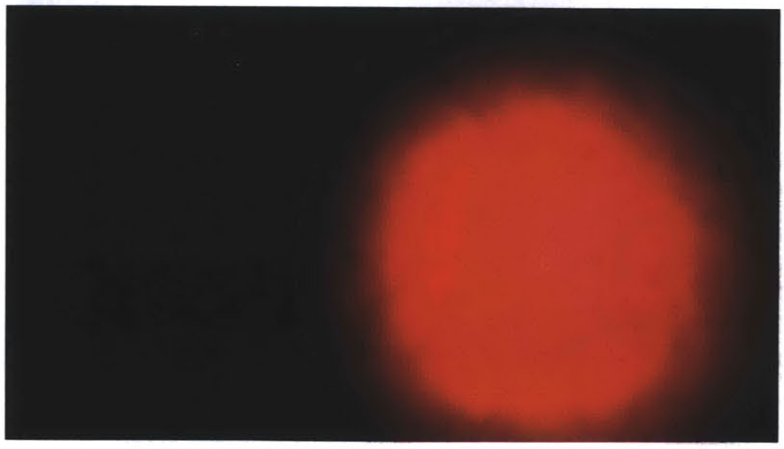
(b) Frequency response after a mid-band filter



(c) Shape and color threshold

(d) Edges Selected from previous image

(e) Edges Selected from previous image



(f) The circular region of the retina is clearly identified.

Figure 3-7: Finding the retinal region. Left column: Frequency filtered selection. Right Column: Retinal selection by shape and color. In both methods, a binary image is created to show where the retina is most likely to be.

for the circle center together, we can mitigate instrumentation error.

Once we have a group of center candidates that all fit into a single cluster, we can average them together to find a potential center of our circle. Using another random sampling of a potential edge points, and a similar clustering algorithm we can determine the radius of our circle. Figure 3-7 shows a summary and comparison of the two techniques for finding retinal tissue on the sensor, as well as edge detection.

3.2.3 Vignetting Removal

Registration of the retinal tissue allowed us to determine the center and radius of the retina shape on the center. Once a circular region of the retina has been identified, we can remove lens distortion by applying an inverse cosine falloff. Given any generic retinal image, with the retinal circular region of interest located, it was experimentally determined (see Section 4.2.2) that the falloff present was exactly cosine of the distance from the center divided by the radius of the circle normalized to extend from $[0, \frac{\pi}{2}]$. (See Figure 4-7).

$$F(d) = \cos \frac{\pi d}{2r} \tag{3.1}$$

Here $F(d)$ is the falloff at any given distance d from the origin of the circle of radius r. Once the center and radius have been obtained, we can correct for this vignetting simply by iterating through the image and dividing by $F(d)$ whenever d is less than r.

$$I_{n+1} = \begin{cases} I_n \cdot \sec(\frac{\pi d}{2r}) & d < r \\ 0 & d \geq r \end{cases} \tag{3.2}$$

Here I_n and I_{n+1} are the image before and after vignetting removal respectively. After the correction has been applied, these images appear "flat" and no longer contain a spherical shape. Figure 3-8 shows a flattened and color-corrected images. The central region of the circle will be selected and used for phase correlation in the next section. See Section A.2 for more examples of corrected vignetting.

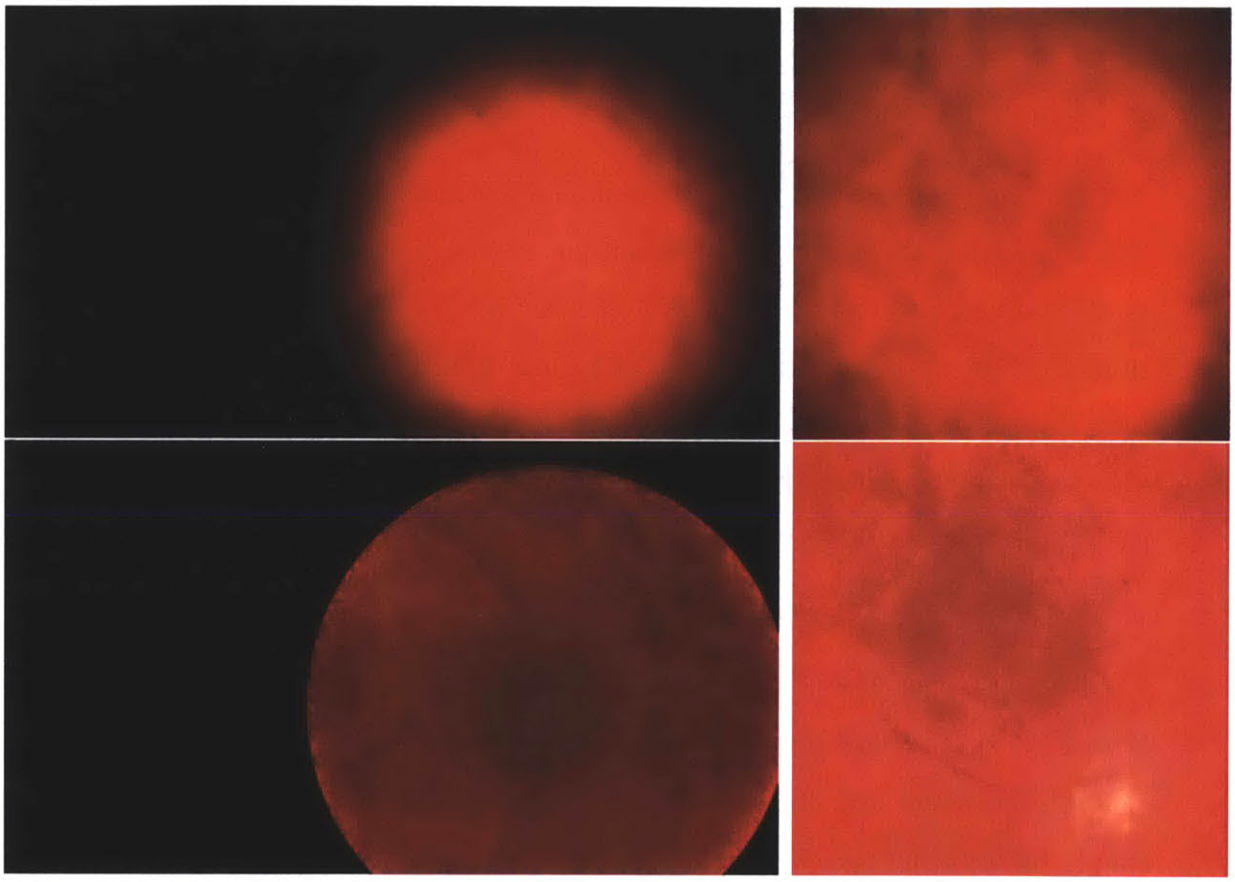


Figure 3-8: Two examples of retinal vignetting correction. The top images show the original image before vignetting correction and the bottom images show the change after the image has been corrected. Observe that the edges of the cropped image are no longer dark around the edges, and more detail becomes apparent.

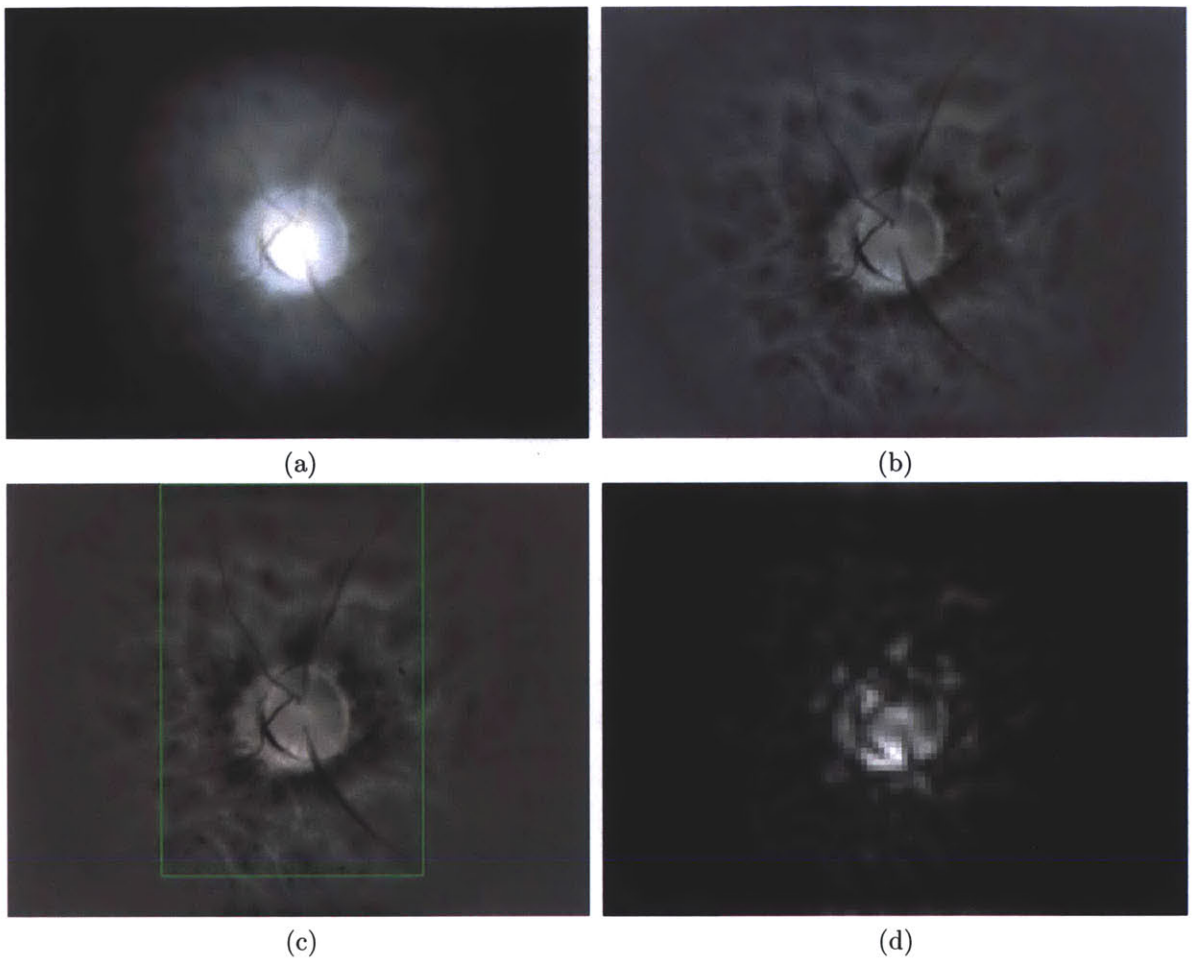


Figure 3-9: Results of high-pass-filtered data. (a) Original image as captured by our PointGrey Flea 3.0 camera. (b) The flattened image obtained by taking a wide-band high-pass filter to (a). (c) The region of interest marked based on the highest concentration of frequencies. (d) The frequency mask created by low-pass filtering the image in (b). Vignetting effect is diminished but not removed.

High-Pass Filtering

For various reasons, occasionally the two methods of registrations as discussed in Section 3.2 will fail, or might be too slow for rapid image processing. We can approximate vignetting correction blindly without identifying the location of the retina or its bounding circle. By taking a wide-band high-pass filter of the image, we can remove the low frequency vignetting effects. However, this produces undesired effects in the image such as artificial contrast enhancement and reduction in large features (such as the optic disk). Yet, this might be the only option for color correction if the

retina is too large or obscured to be identified correctly.

To blindly, wide-band high-pass filter, we first blur the image significantly and then subtract the blurred image from the original image. Results of this blind vignetting correction can be found in Figure 3-9.

The high-pass filtered image, while not as narrow-band as the images in the last section, still represents areas of most data, and therefore can also be used in identifying a region of interest. We select a region of interest based on a rectangle containing exactly 98% of the interesting (non-zero) data.

3.3 Frame Correlation with Online Clustering

3.3.1 Phase Correlation

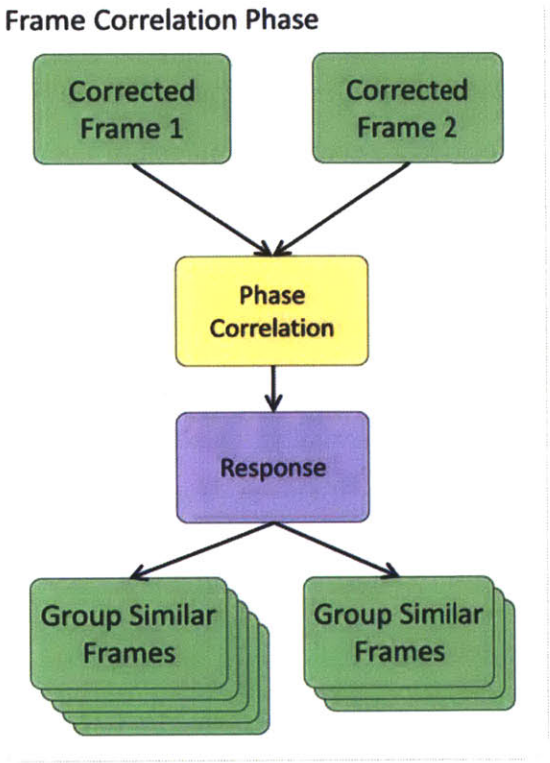


Figure 3-10: Frame Correlation with Online Clustering

Obtaining flat images is the first step in image processing, however flat images still contain a lot of noise and little contrast. By integrating images together locally, we can increase contrast and reduce noise, as well as achieving a higher resolution result. But, we must first identify images that are similar which we can integrate together. Once we have groups of similar images, we need to align them spatially so that we can integrate them temporally.

Traditional methods of image comparison use feature detection to identify similar regions in two different images. These use a combination of corners, and contrast differences, to identify points of interest in images, also called feature points. Once feature points in two images have been discovered, points from one image can be matched to corresponding points in the other.

Since the points are zero-dimensional, the mapping from one image to another using a feature matching technique allows for scaling and rotation. Therefore, not only does feature matching produce translation between images, it generates a full homography matrix giving an affine transformation.

Template matching is also a common technique to match two pieces of images together. A piece of one image, the template, is convoluted with the whole of the other, the scene, and the intensity response is mapped at all points in the image. The

peak response value on the intensity graph depicts location on the scene where the template has the highest match. Unlike feature matching, template matching only detects translation and will fail if the images have been rotated or scaled.

However in this case, both template matching and feature matching failed to produce accurate results for our data (See Section 4.2.4). The low signal-to-noise ratio in our data produced almost no robust features to match, and gave poor template matching results. Instead of trying to correlate images in the spatial domain, like template and feature matching both do, we decided to correlate images in the phase domain using a technique called phase correlation which proved to be more robust against noisy data.

Given any two images I_a and I_b the phase correlation can be found by first applying a Hanning window on both images to reduce the edge effects, and then finding the Fourier transform of each image, F_a and F_b respectively. Next, calculate the cross-power spectrum by taking the complex conjugate of the second transform and multiplying it by the Fourier transform of the first (Equation 3.3).

$$R = F_a F_b^* \tag{3.3}$$

The normalized phase correlation is obtained by dividing R by the spectrum power (Equation 3.4).

$$R = \frac{F_a F_b^*}{|F_a F_b|} \tag{3.4}$$

Now apply the inverse Fourier transform to retrieve and intensity graph of matched responses.

$$r = F^{-1}\{R\} \tag{3.5}$$

This will produce a graph as shown in Figure 3-11 The location the graph which has the highest matched response is the highest probability for the shift between the two images. A normalized matched response of an image with itself always has a magnitude of 1 at location $(0, 0)$. An image that has been shifted two pixels to the

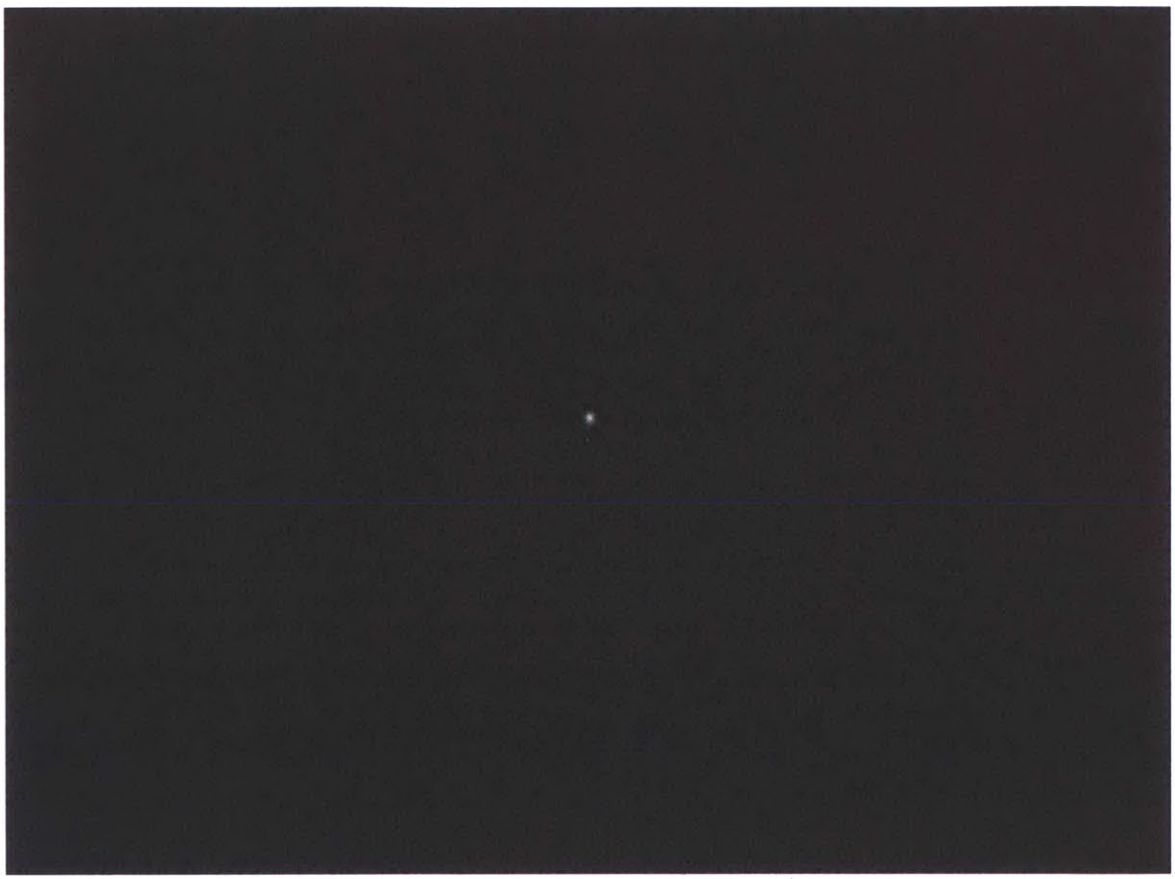


Figure 3-11: Phase correlation response graph between two flattened retinal images. The highest response, large white area, is located to the north-east of center, which is indicated by a single white pixel

right and one up from the original image will have a match response of less than one at location $(2, 1)$.

A region of interest selection can be made from flattened images obtained after vignetting correction by taking the central square region of the retina circle. Phase correlation is useful for both identifying the amount of translation from one image to another as well as the quality of correlation.

3.3.2 Online Clustering and Correlation

Phase correlation, as introduced in the last section, correlates in the frequency domain and is less affected by noise, high frequency phenomena, and blur, low-frequency phenomena. Once we have vignetting corrected and a flattened set of n images, we can find the corresponding correlation between all pairs of images, cluster them into groups, and integrate the ones with the highest corresponding values. See Figure 3-12 for correlation graph of n^2 images where the values indicate the quality of the correlation.

In theory, creating a correlation matrix of a given set of frames and using a clustering algorithm to identify groups of matches sounds like a reasonable method, but in practice, phase correlation between two images takes about 100th of a second. While that may not seem like a lot of time, given a data set of about 3000 images (100 seconds), that makes around 9 million correlations, and approximately 10 days to generate the correlation matrix for all 9 million entries. Thus an $O(n^2)$ algorithm is an unacceptable option.

We present the following online clustering algorithm to group individual frames together in an efficient manner. First we make two key observations. For one, reduction in the size of images affects neither their correlation position or relative intensity (though it does affect their absolute intensity). In Figure 3-14 you can see the correlation between 50 consecutive frames at native resolution, $2\times$ resolution, $0.25\times$ resolution, and $0.0625\times$ resolution. Notice that as the resolution decreases, the values returned from the phase correlations are increased, but their relative heights are not and that all of the graphs have the same relative shape despite their different

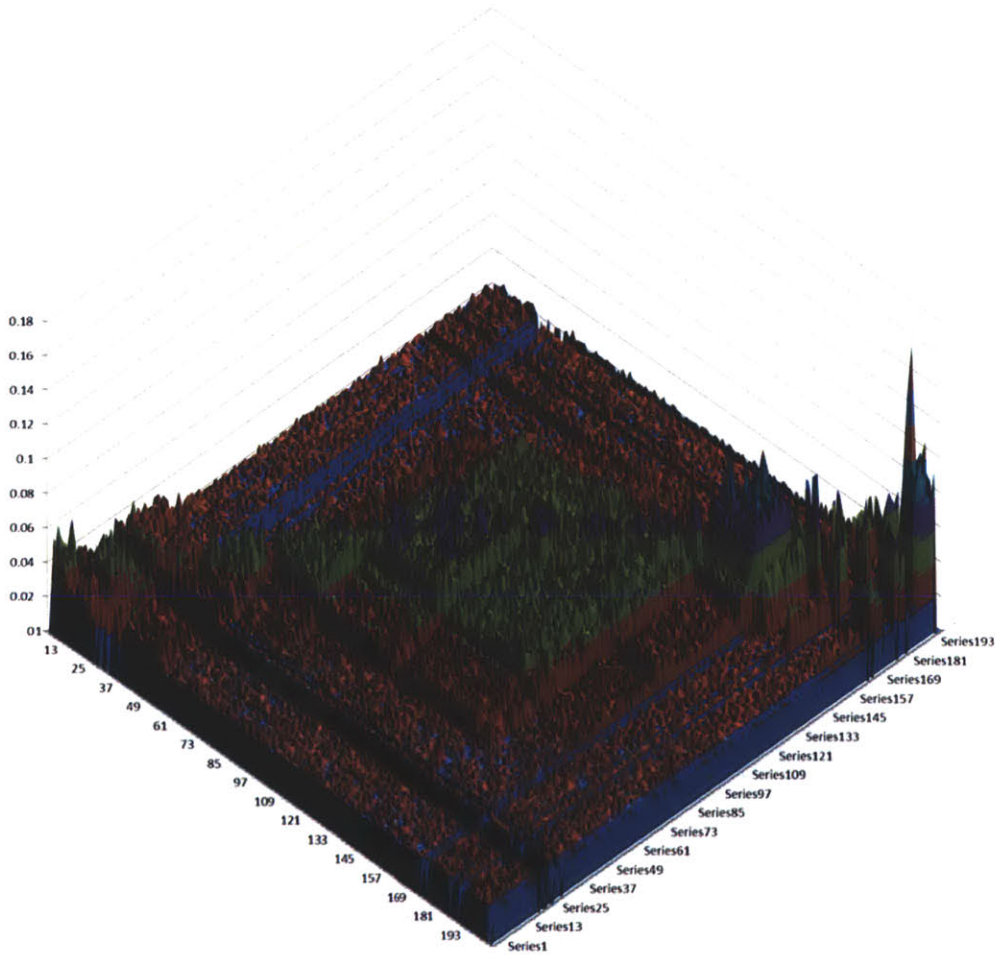


Figure 3-12: Pair-wise correlation for 200 items (40,000 correlations). Peaks indicate good correlation while troughs indicate bad correlation. The main diagonal has been removed to better illustrate contrast between similar, but not equal, elements. A normalized matched response of an image with itself always has a magnitude of 1 at location $(0, 0)$.

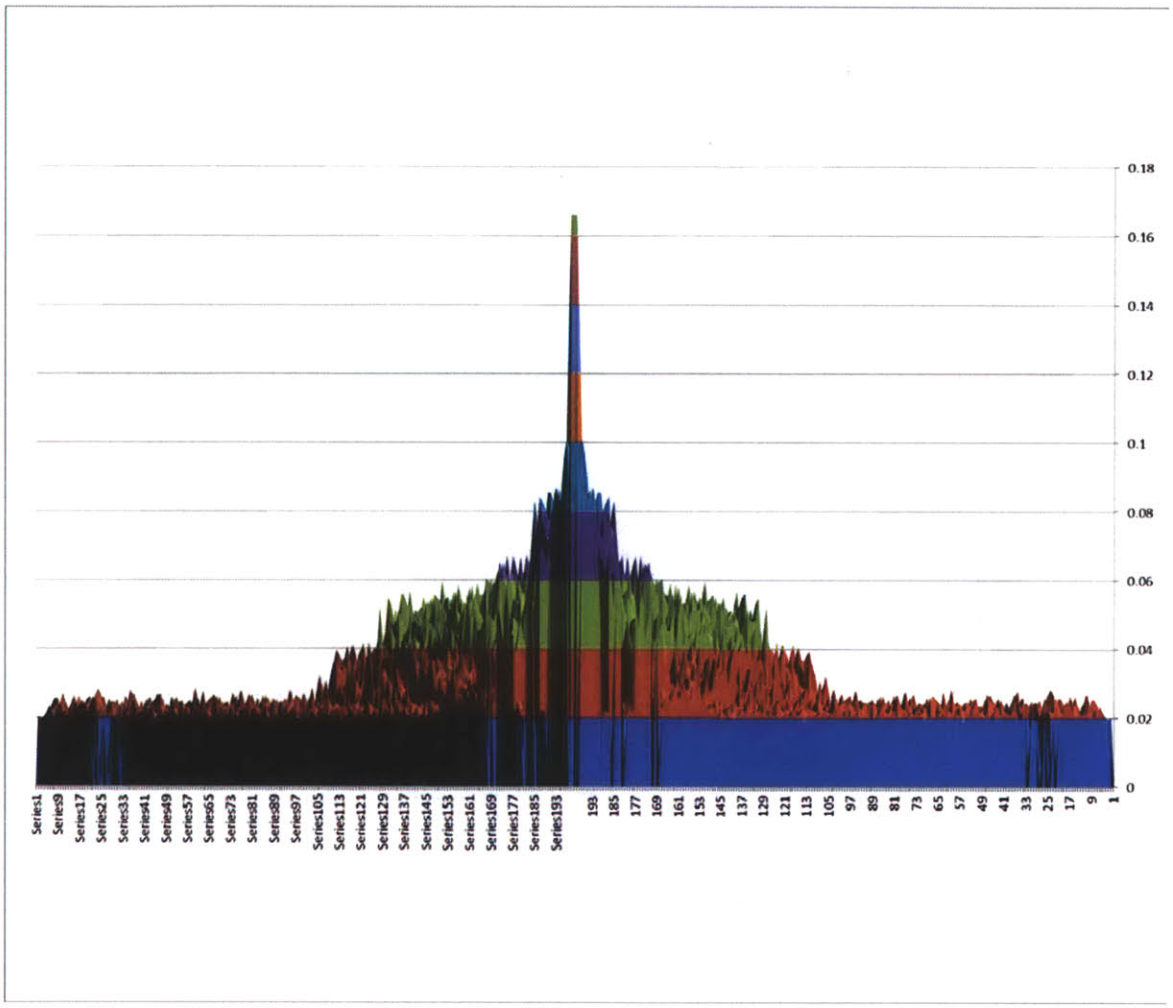
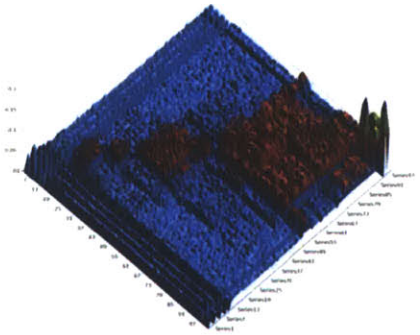
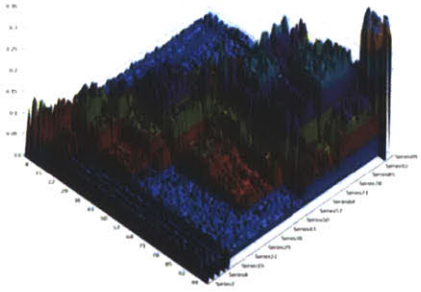


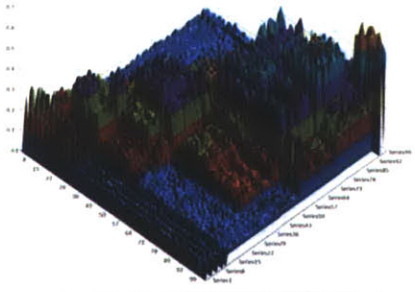
Figure 3-13: Peak Concentration. The same correlations as shown in Figure 3-12, except shown from the side with the main diagonal down the center. Most of the high correlations occur near the center of the graph and correlations decrease as images become farther apart in time.



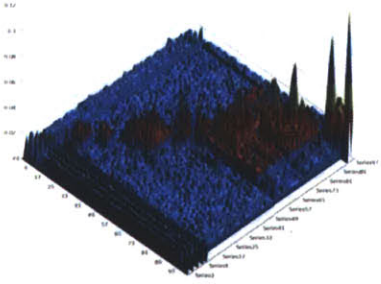
(a) No scaling



(b) Downsampling 2x



(c) Downsampling 4x



(d) Upsampling 2x

Figure 3-14: Correlations of the same dataset with different image resolutions. While the scale is different in each image, the correlations still have the same relative heights. This means we can down-sample the images before performing a phase correlation to significantly speed up processing time.

absolute values. (The main diagonal has been removed to improve the clarity of the other correlations. An image correlated with itself will always produce a response of 1).

The second key observation is in the clustering of similar frames. While not all two consecutive frames, or even three consecutive frames, have a high correlation value with each other, it is still true in general that related frames are grouped together when images are processed in the order they were recorded. This is evidenced by the red squares in Figure 3-12. Notice that at no point do red values appear anywhere but in a square or anywhere but in close proximity to the major diagonal. Figure 3-13 shows the same dataset, but from a 45 degree angle. Notice that the best correlations appear near the center of the graph, and only good correlations appear near the center of the graph.

Using these two observations we can develop the following online algorithm to identify the "square" regions of highest correlation. As frames are captured one of three things can happen: either 1) they can be part of the current square; 2) they are not part of anything and just random noise; or 3) they are part of a new square yet to be started. We can start by creating k empty groups. As a new frame is loaded, correlate it with each of the elements of the k groups. If it is a high match with every element of a single group, add it to the group. If it is not a high match, the image must either be random noise or the start of a new group. If either of those are the case, evict the group that was modified last, and add this new frame to the k as a singleton.

We present the following online clustering algorithm:

Once a group is evicted, we can decide whether or not it constitutes a "square" by looking at its number of elements. If it has more than k elements, we can integrate the elements together (as discussed in the next section) and output the resulting image. If that group has less than k members, we can throw it away as random noise (Figure 3-15).

However, this presents a problem. Even though we designed this algorithm to run in linear time, it could potentially run in $O(n^2)$ time if all elements are well correlated.

Algorithm 1 Online Clustering with Phase Correlation

```
Create k Image Groups
while frames exist in video do
  load frame
  for each k groups do
    if Correlation(frame, group[i]) > ε then
      add frame to group
      move group to front of the queue
    else
      add frame as new group
      evict last group from queue
      if Last group has > k members then
        Integrate frames together
      else
        drop group without integration
      end if
    end if
  end for
end while
```

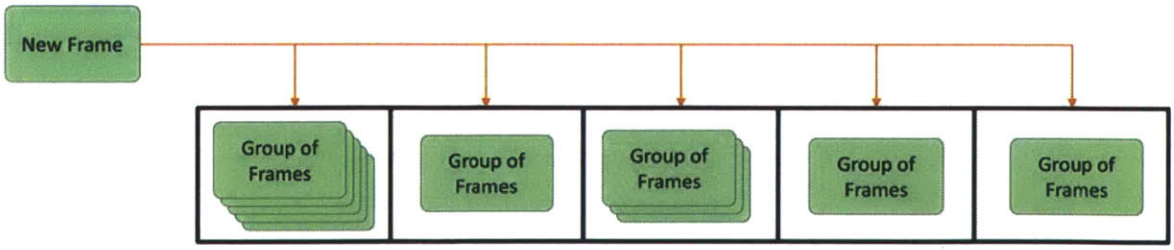


Figure 3-15: As a frame is added to the set, it is correlated with the other elements from the other sets. If it is well-correlated to all elements of a set, it is added as a member. If not, then the last group to be updated is evicted and the new frame is added as a singleton group.

To correct for this, instead of correlating a candidate frame with every member of a potential group, we can take a random sampling of k items in the group and only add the the frame if it is well-correlated with all k random elements. This limits the runtime of our algorithm $O(kn)$ where k is a small integer.

3.3.3 Local Image Integration, Upscaling, and Blending

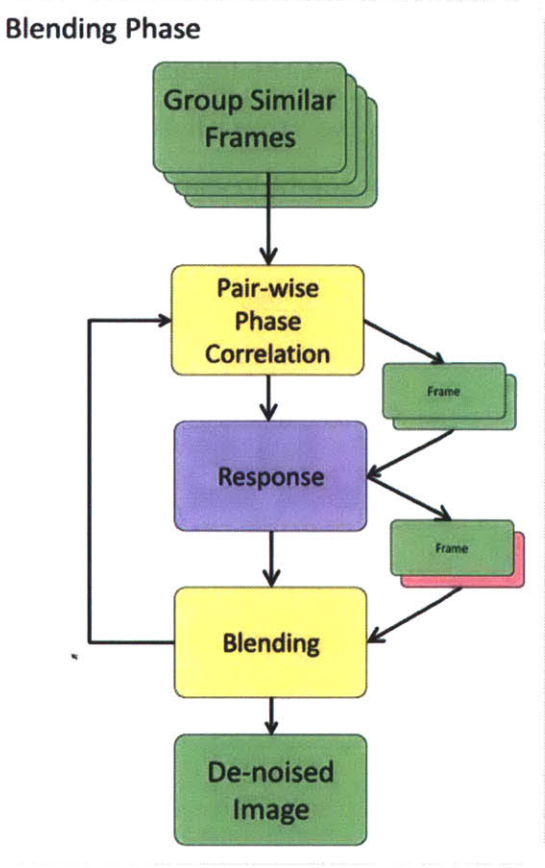


Figure 3-16: Image Integration, Upscaling, and Blending

Integrating images through time reduces noise and improves image contrast. Phase correlation produces a quality of match as well as match location. Since neither of these values are a discrete quantity this translation can be determined to a sub-pixel level [13, 16]. While individual images are limited to HD capture resolution, this sub-pixel accuracy allows us to upscale and fuse together images to increase the resolution while also increasing the contrast and decreasing noise.

Once images are grouped into sets of high correlation (per the last section), we can upscale using a bicubic approximation, sharpen with an unsharpened mask, and recorrelate at full resolution using phase correlation to generate the highest quality image. Given any two images from a group, we know they will be well-correlated since images are only placed in groups where every item is well-correlated with every other item. Therefore, given any reference image, every other image in the group will also be well-correlated to that reference image.

We can use this observation to simplify the fusing problem. Images can be correlated to one single reference image to identify the x and y translation (as shown in Section 3.3.1). Centering one image on the canvas, we calculate the translation from the reference image and then we can perform a raw integration over time on the

images.

Each color channel is blended separately and normalized between 0 and 1 after the integration is performed. In order to retrieve the most amount of data, values are added first as double floating point numbers.

Since we have already created a mask with the areas of the highest contrast, we can use them in blending images together to yield a picture with the best image quality. The mask we create from the mid-band frequency filter can be applied per color channel (See Figure 3-17). This ensures that we only obtain images of the highest contrast. The results of blending/integration can be seen in Figure 3-18.

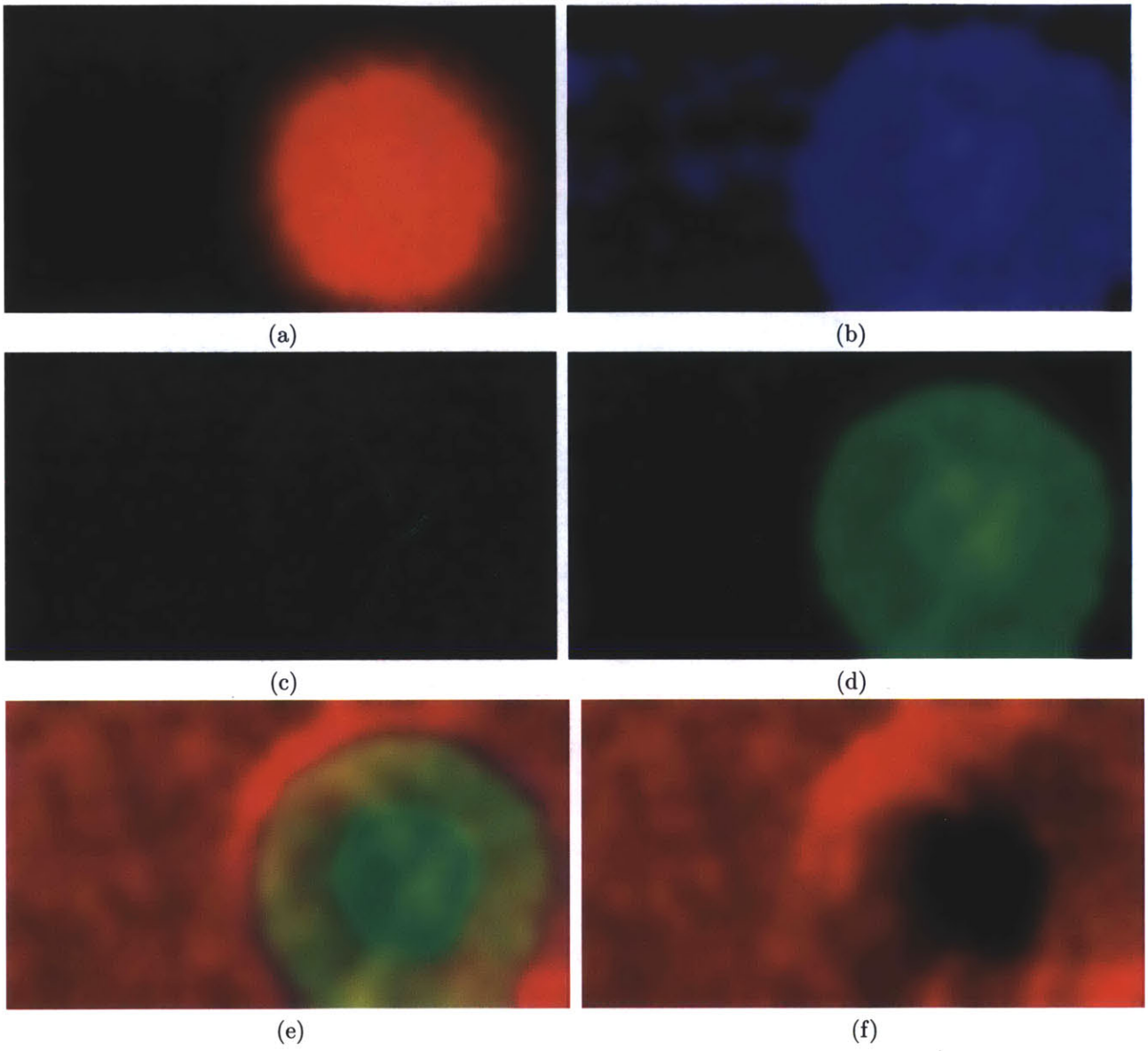
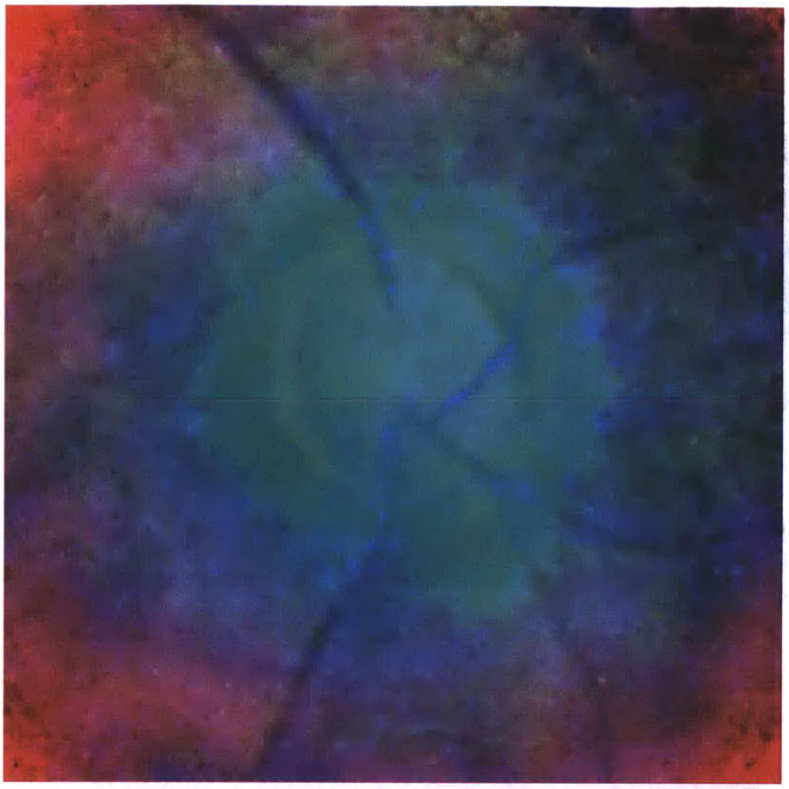
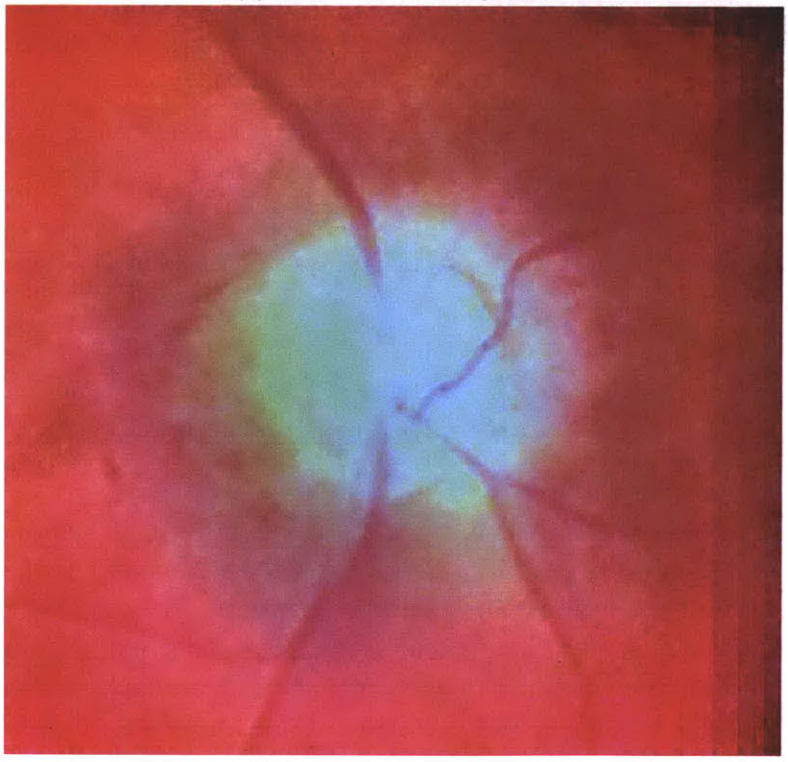


Figure 3-17: Creating a blending mask. The frequency band response calculated in the first step of the imaging process is used to merge the images. a) the original image. b) The combined filter response for all color channels. c) Low-pass filtered to bring out high contrast areas. d-f) The blue, green, and red color channels respectively. Notice that the red color channel in the bottom right is black in the center. This is because the red light is flushed and the area contains no data. This is another reason why red channels were zeroed in many of our captured images.

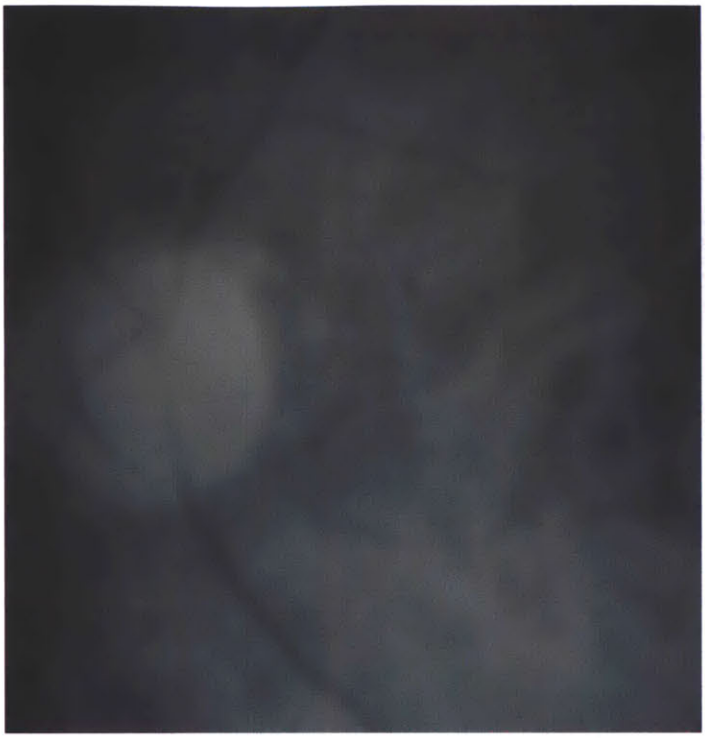


(a) Before Local Integration

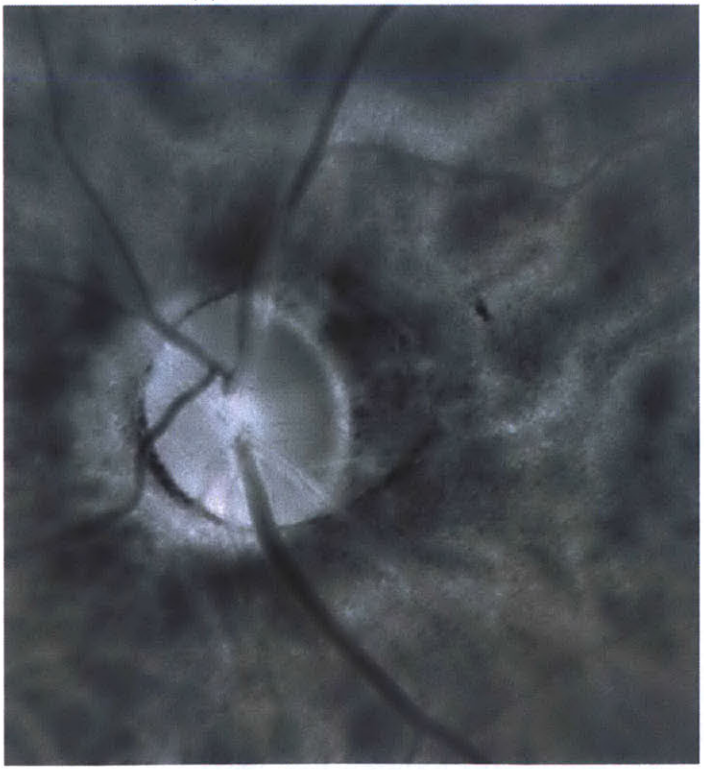


(b) After Local Integration

Figure 3-18: Results of image local integration. Notice the amount of detail that appears in (b), including micro vascular structures that were not apparent in a single individual frame (a).



(a) Before Local Integration



(b) After Local Integration

Figure 3-19: Results of image local integration with PointGrey Flea 3.0 Camera. Notice the amount of detail that appears in (b), including pigments that were not apparent in a single individual frame (a). This integration only took seven frames.

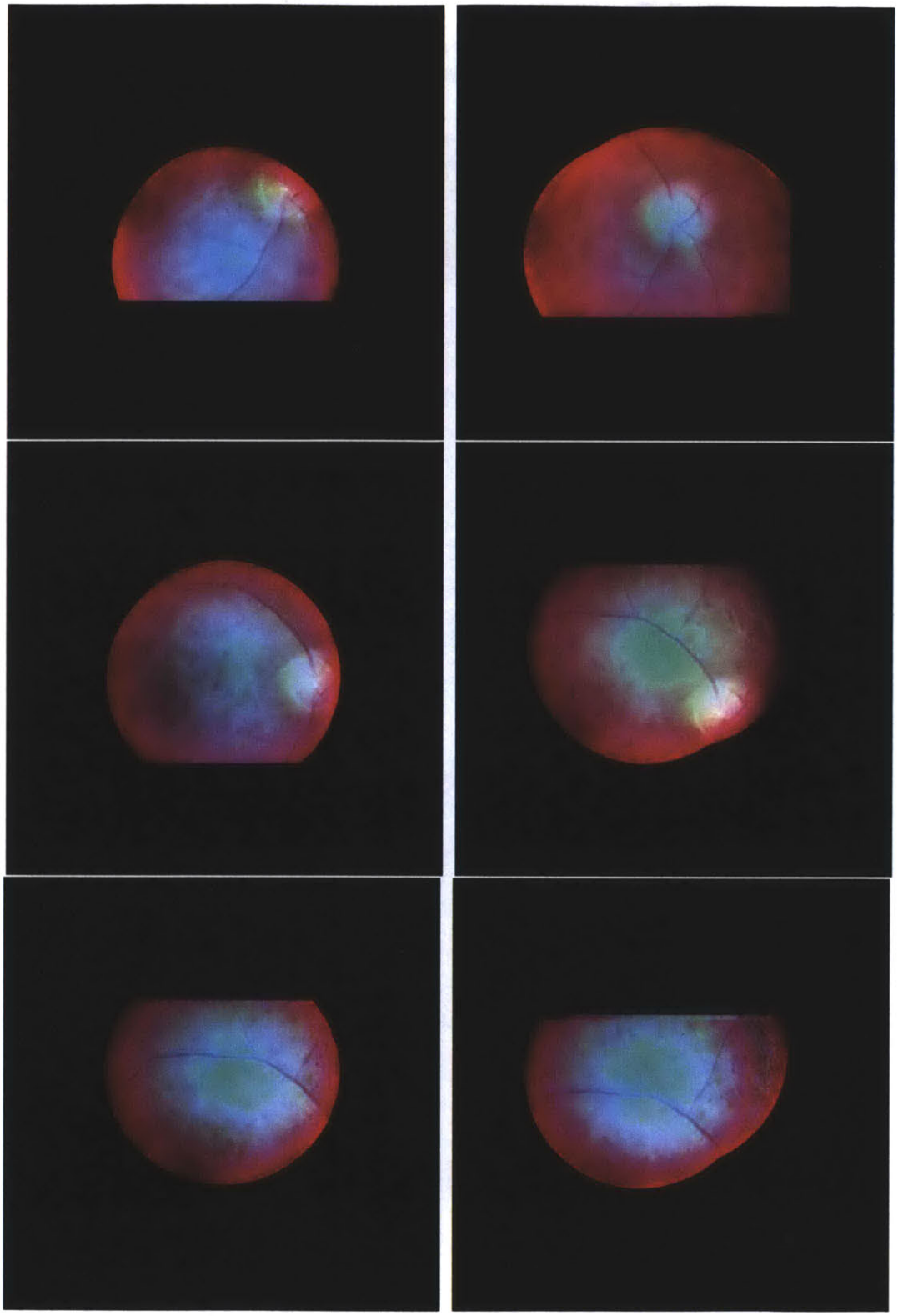


Figure 3-20: Examples of images after they have been integrated together. Red borders appear due to rounding errors at the edges of masks.

3.4 Global Panoramic Stitching and Mosaicing

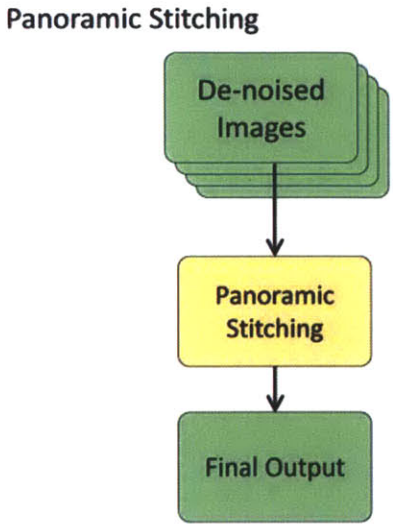


Figure 3-21: Global Panoramic Stitching and Mosaicing

After images have been upscaled and integrated, they contain enough features to be stitched together automatically into a large field of view mosaic. Since the field of panoramic stitching is well-defined and well-explored, this thesis provides only an implementation of such algorithms and not the creation of them. Areas of maximum regions of interest can be identified automatically using the regions of interest found for phase correlation, and the corresponding images can be cropped.

In order to stitch images together into a large mosaic, we utilized the open computer vision library (OpenCV) which was easy to incorporate into the current software verification system. Mosaics stitched by our system can obtain almost a 180° field of view of the retina. The following is a figure comparison of an automatic stitching from two different programs: C++ OpenCV and Hugin, a free panoramic stitching software.

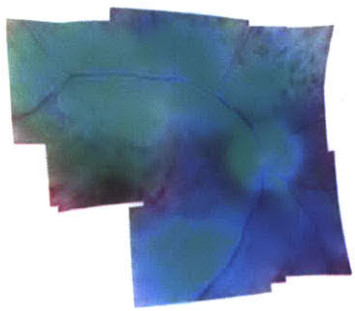


Figure 3-22: A panorama stitched automatically without user input by this imaging process

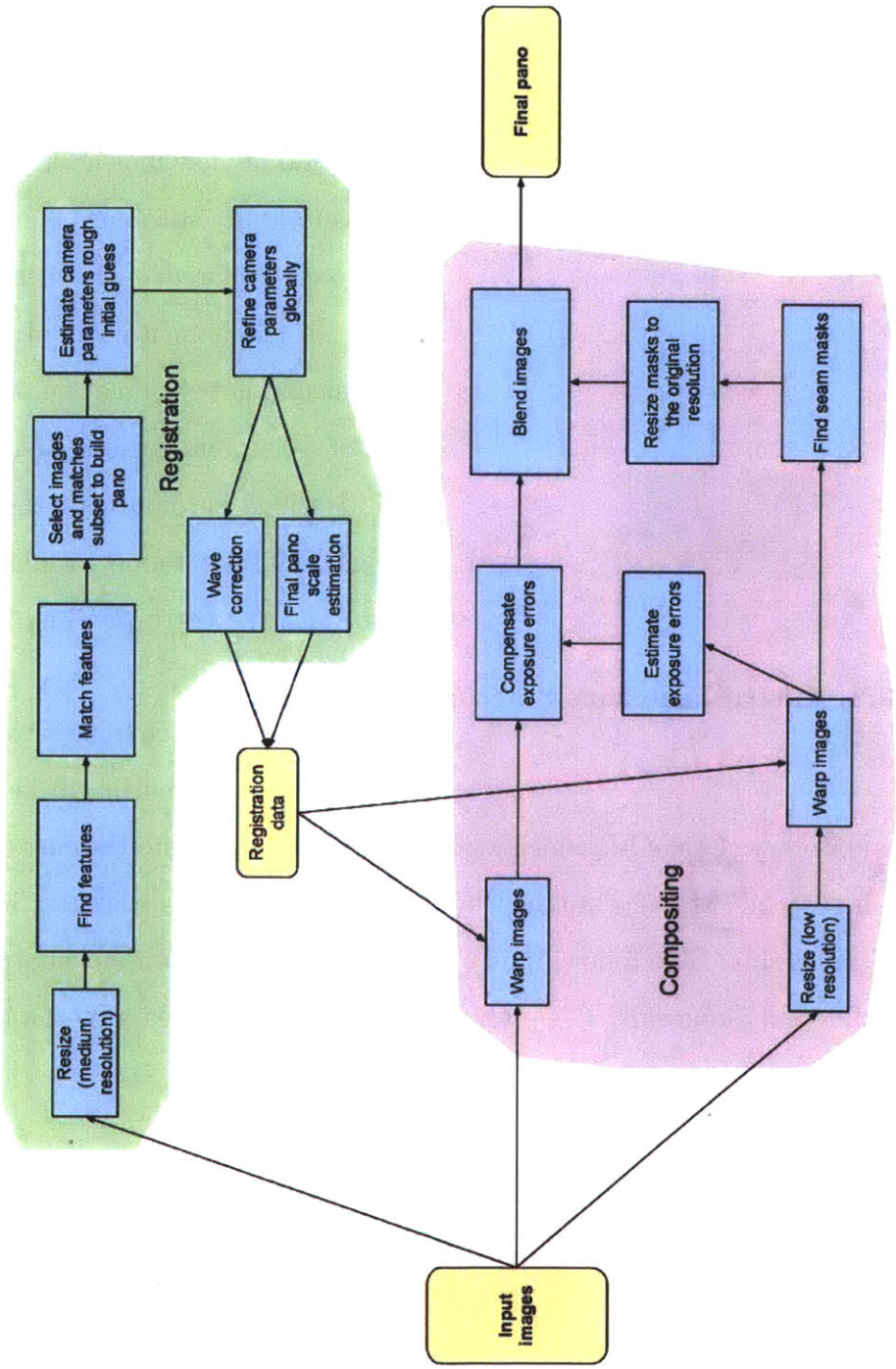


Figure 3-23: Panorama Stitching Process

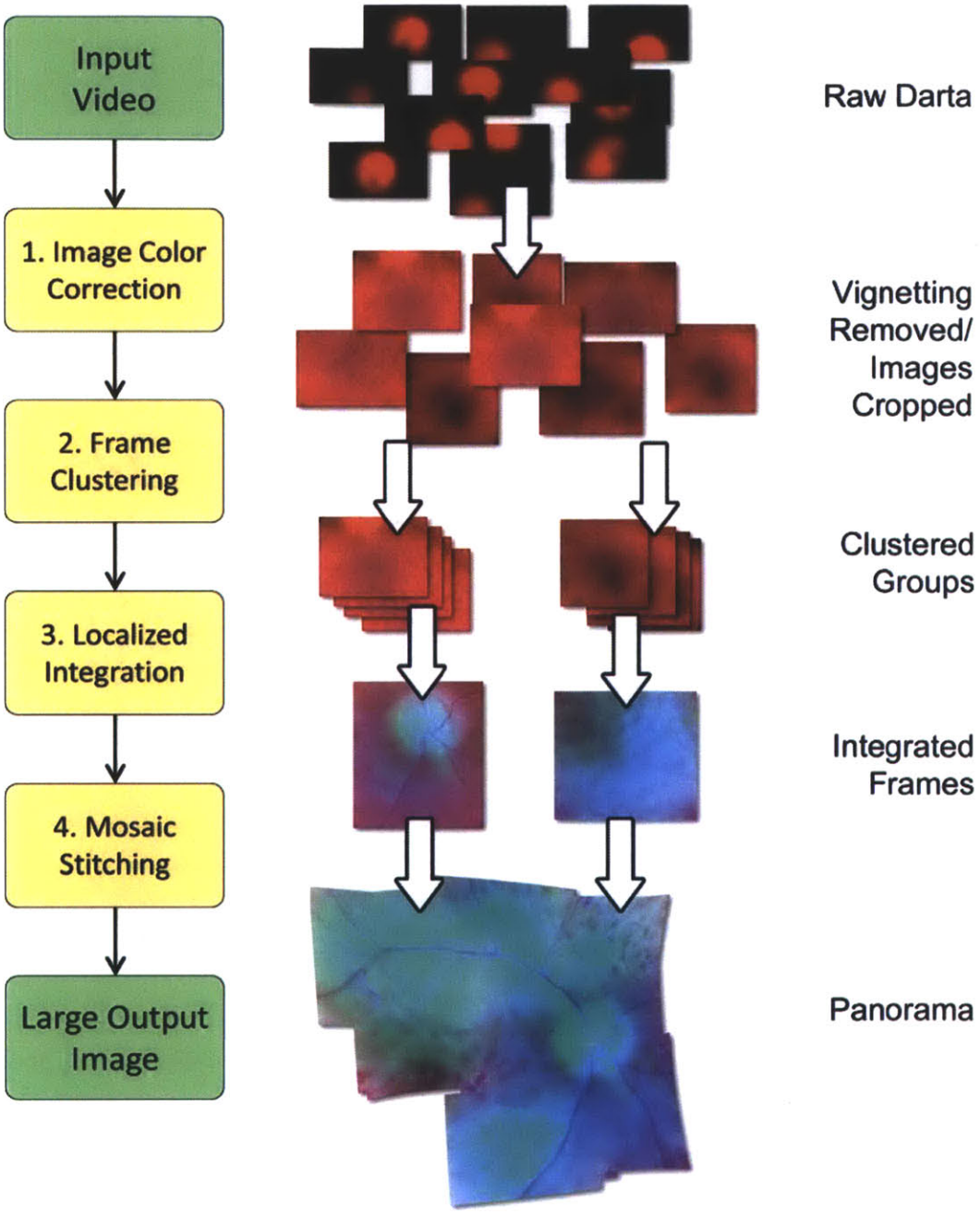


Figure 3-24: Image processing pipeline with representative sample images.

Chapter 4

Validation and Results

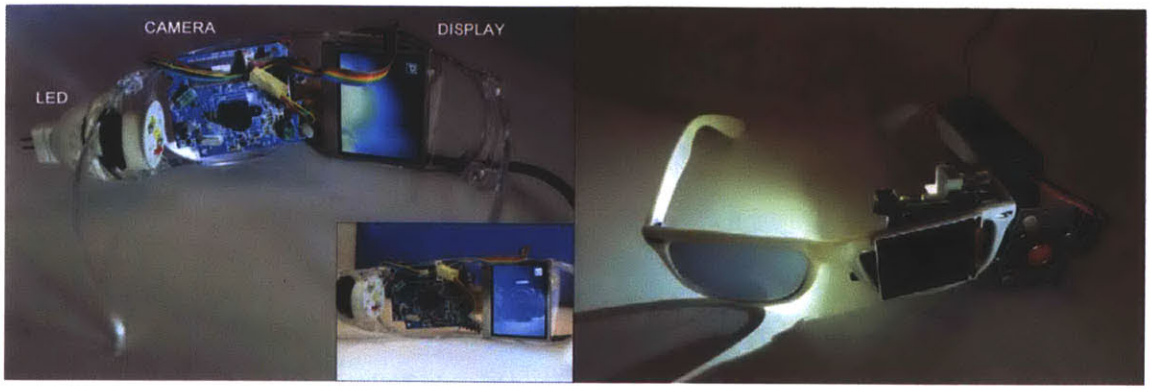
4.1 Hardware Capture Systems

Since my contributions are primarily to the software, the following sections on hardware reflect work from other students on the team. Specifically, Everett Lawson developed most of the hardware that we used in our prototyping and implementation.

4.1.1 Our Capture System

We have developed numerous working prototypes (See Figure 4-1 for two examples). Our main design is fitted into a pair of sunglasses and consists of three main components: the camera for capture and the display for stimulus are both facing the eyes, and affixed is an LED for illumination. For capture, we used components from a Flip Ultra HD (720p / 30fps) camcorder fitted into the left eye of the glasses frame. The stimulus is presented to the right eye, as it gazes through the focusing objective. A $1W$ dichromatic white LED with a luminous efficacy of $120lm/W$ is temporally placed in proximity to the posterior region of the Lateral Canthus.

For computational high dimensional capture, we use PointGrey Flea 3.0 camera. In order to test multidirectional lighting, we equipped a plastic ring with twelve evenly-spaced $1W$ white LEDs, each on a copper wire for adjustment. The LEDs



(a) Built-in display for stimulus (b) Remote display for stimulus

Figure 4-1: We present two prototypes for our imaging system. Both prototypes are based on glasses-mounted design for different methods of stimulus presentation: (a) A display fixed within the glasses, thus providing high portability. (b) View of an external display from the other eye for stimulus control. This particular prototype has the ability to involve another observer if required as is the case for clinical and research settings.

were attached to an Arduino microcontroller, controlled through a PC. We used this system in conjunction with a synthetic eye developed in the lab and a real human eye. Most literature for light safety exposures for the human eye exist for non-visible, near IR and laser-based illumination [6, 12]. Our implementation of indirect diffuse illumination using low intensity white LEDs is limited to the visible wavelengths, which are within eye safety standards.

Red and green LEDs were used to capture multi-spectrum data using a narrow band filter in front of the camera. For light field capture, we placed a $400\mu m$ 2×2 microlenslet array one focal length from the sensor, making use of the eye's lens to direct light into free space.

4.1.2 Resolution Analysis

We evaluate the resolution properties of our camera against other retinal imaging setups such as the PanOptic direct ophthalmoscope with a camera attachment based on the design by [14] and an auto-focus fundus camera [20]. In order to facilitate digital image capture for the PanOptic, we utilized a commercially available attachment [1]

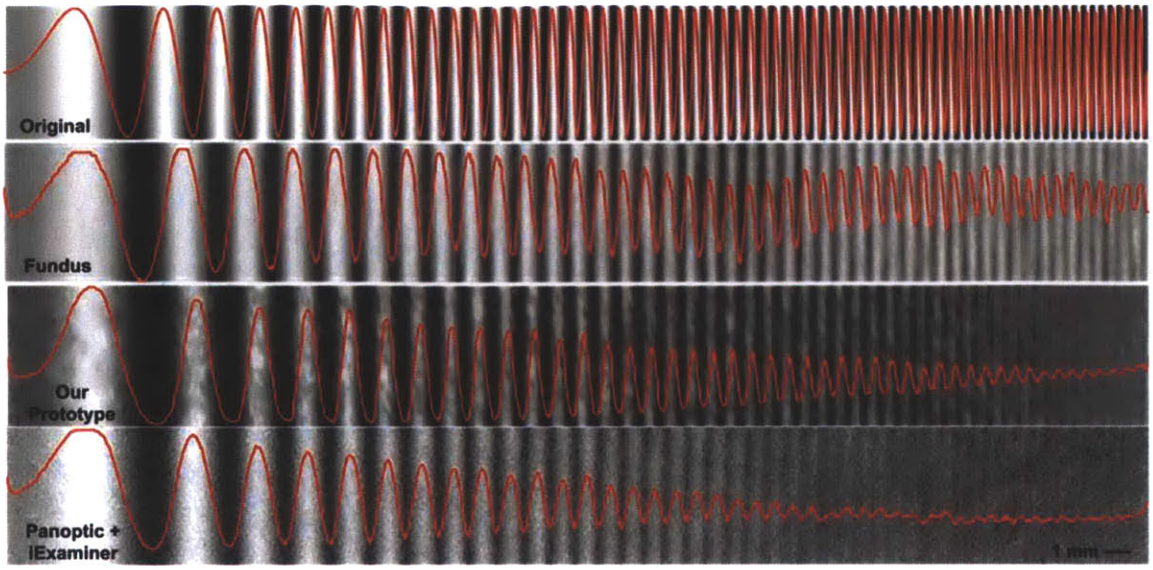


Figure 4-2: Resolution analysis using a quadratic exponential (chirp) 2-D pattern. (1) Original, (2) Fundus Camera, (3) Ours, (4) Panoptic.

and the rear camera of an Apple®iPhone4. A 2-D pattern, commonly known as the *chirp* function is used as a printed test target. This target is about $20mm$ wide with a maximum frequency of $3cyc/mm$ as shown in Figure 4-2(top). We utilized a double convex lens of $50mm$ in order to focus at the target that was placed at its focal length. This setup emulates imaging the eye, when it is focused at an object placed at infinity. All parameters were precisely measured before capturing data.

As seen from Figure 4-2, a one-dimensional profile of the corresponding test image is plotted as the mean across the first dimension. This enables visualization of each system's ability to resolve various spatial frequencies. Out of all validation setups, best performance is shown by the auto-focus fundus camera, given the advanced and precise optics used in combination with a digital SLR camera. The resolution analysis clearly shows that our prototype can comfortably resolve spatial frequencies up to $2.5cyc/mm$, this corresponds to a resolution of $\approx 200\mu m$. In comparison, the PanOptic based camera can resolve up to $2cyc/mm$ thus showing reduced ability to image closely spaced microscopic features as in the case within the eye.

4.2 Validation

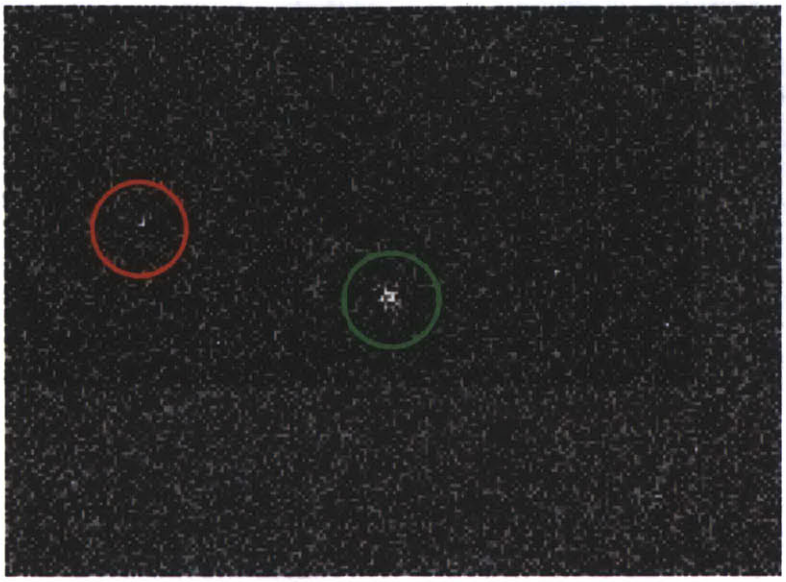
4.2.1 Software Overview

The aforementioned algorithms and procedures were coded in C++ on a standard Windows 7 PC (10GB RAM, 1TB Harddrive, 3.4GHz i7) using the OpenCV 2.3.3 libraries. The phase correlation class of the image processing library OpenCV was modified to return the correct location of the maximum-valued point. The original method simply looked for the maximum value in the phase correlation response graph, and not the maximum area (See Figure 4-3 for an explanation of the error).

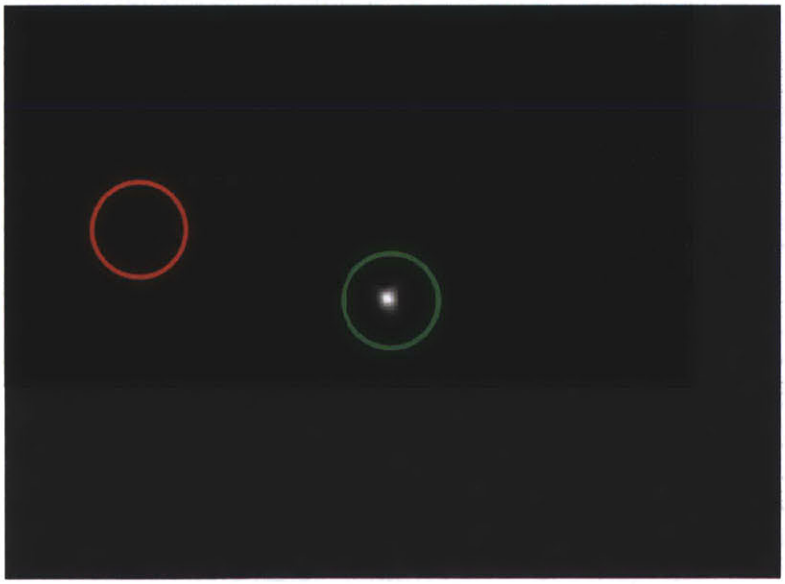
4.2.2 Vignetting Falloff Calculations

In order for phase correlation to be accurate, input images have to be "flat" i.e., not suffering from a type of vignetting, since we want the details to match and not the vignetting patterns to match. Thus it was imperative to determine the exact falloff of the spherical shape in the photographs. Simply taking a high-pass filter of the image would not give a good "flat" image as desired due to the induced contrast and loss of large details such as large dark areas caused by rupture of blood vessels due to diabetic retinopathy, or large light areas like the optic disc. At a certain point away from the center of the retinal tissue the image becomes almost black, and a high-pass filter over the image would cause strange edge artifacts. Therefore, we have to apply the inverse vignetting profile correctly to flatten a image (See Figure 4-4).

To accurately identify the pattern at which the vignetting is occurring, we took a cross-section of the retina which is shown stretched out in Figure 4-6. It was originally suggested that the falloff could be $\cos^4 d$, but after performing the regression for many different powers of cosine we determined that the falloff is exactly in the form $\cos ax + b$. The regression was performed by using a least squares approximation, first by centering the data, normalizing it between 0 and 1, taking the inverse cosine, and solving for the equation $ax + b$ where x corresponds to the physical pixel space and a and b are phase factors.



(a) Raw phase correlation as returned by the native OpenCV 2.3.3 Library



(b) Phase correlation corrected by taking a low-pass filter

Figure 4-3: The phase correlation response. Light points indicate translation. (a) The peak correlation appears in the red circle, even though there is a cluster of high intensity in the green circle. (b) After low-pass filter, a blur, has been applied to an image we can clearly see that the response we are looking for is in the green circle. Since the circle is right in the center of our image, it corresponds to a $(0, 0)$ translation.

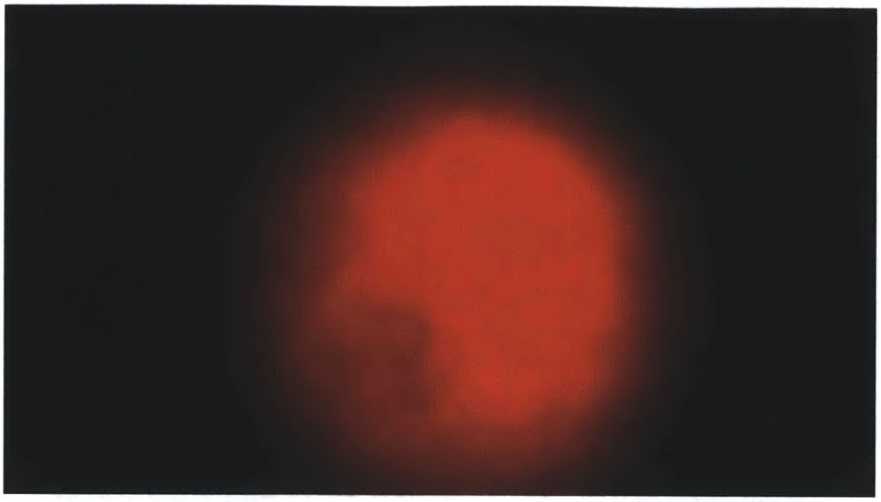


Figure 4-4: Here the vignetting is readily apparent. Changes in lens magnification cause the edges of the retinal image to go dark. Since his image has fairly smooth color throughout its entirety, we can use it to figure out the vignetting falloff.

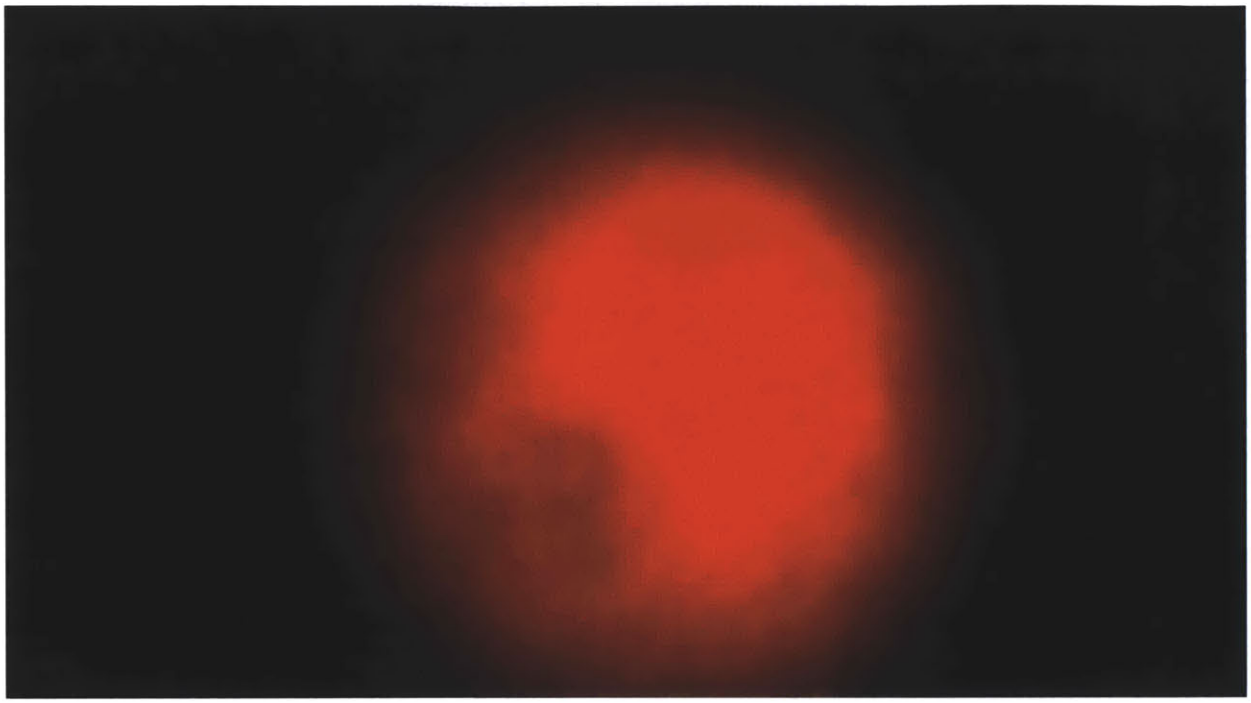
The cosine fit is shown in Figure 4-7, where the radius of the circle on the sensor corresponds to $\pi/2$ radians. Examples of corrected results using a secant scaling factor can be found in Section A.1.

4.2.3 Image Integration Verification

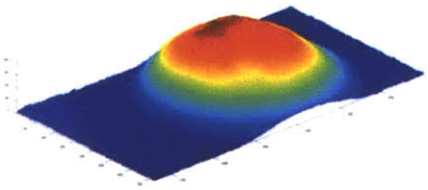
super-resolution

super-resolution is the ability to derive a larger resolution image from small resolution counterparts. We know creating a super-resolution image is possible since [34] explains how to obtain a super-resolution image from several lower resolution images and since phase correlation can be determined at a subpixel level. This means we can upscale images using a bicubic interpolation, correlate them in the higher-resolution, and then integrate them together.

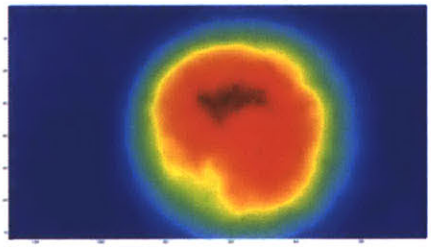
super-resolution can be easily verified by observation. Figure 4-8 shows a zoomed in version of the retina before and after data processing, where pixel sizes are denoted by squares. Notice that in the low-resolution image, Figure 4-8(a), not only is the image more blurry and contains less detail, but at a per pixel level sharp features are less apparent. In Figure 4-8(b) features appear a lot sharper and the detail is not



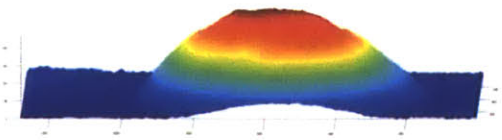
(a) Raw Data



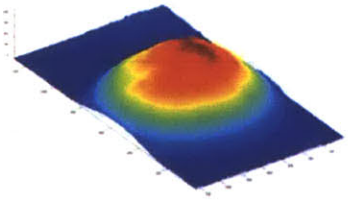
(b)



(c)



(d)



(e)

Figure 4-5: The brightness profile of the red channel of a typical image captured by our hardware. In this image, the vignetting pattern is extraordinarily apparent, with the 3-D shape relatively rotationally symmetrical about its center.

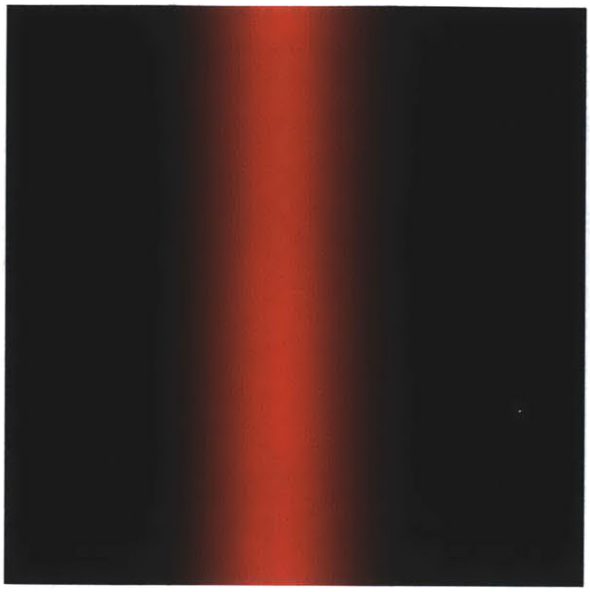


Figure 4-6: A linear slice of the retina stretched for emphasis.

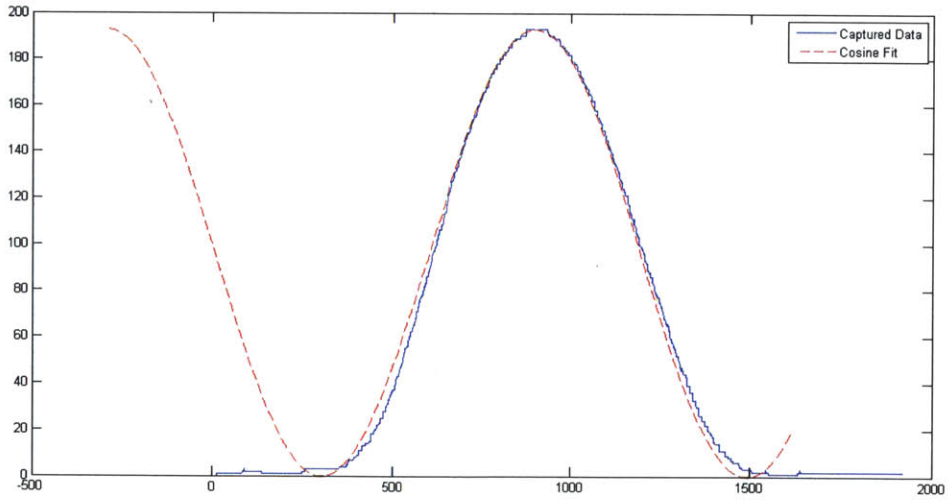
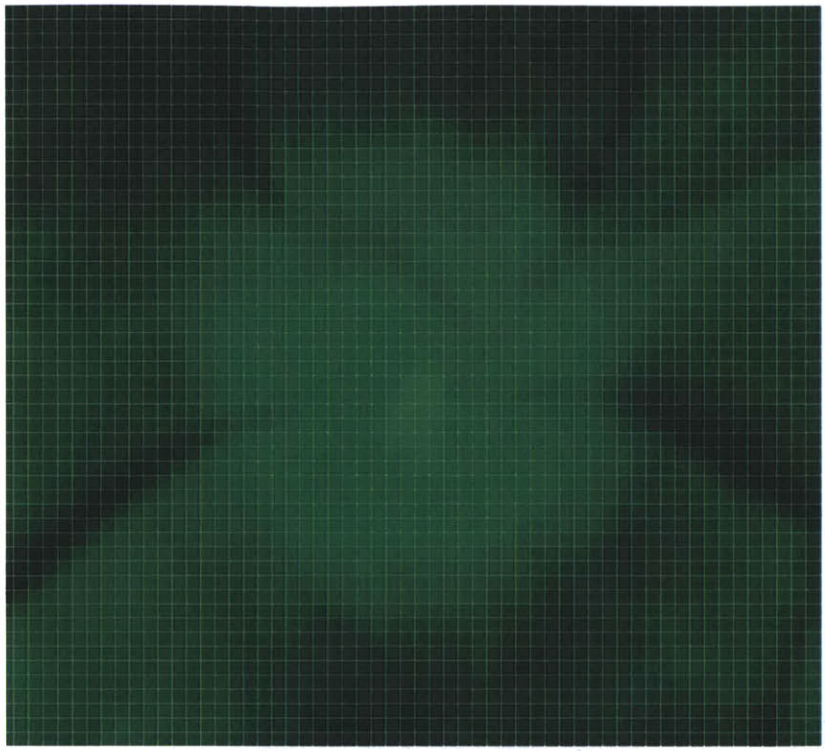
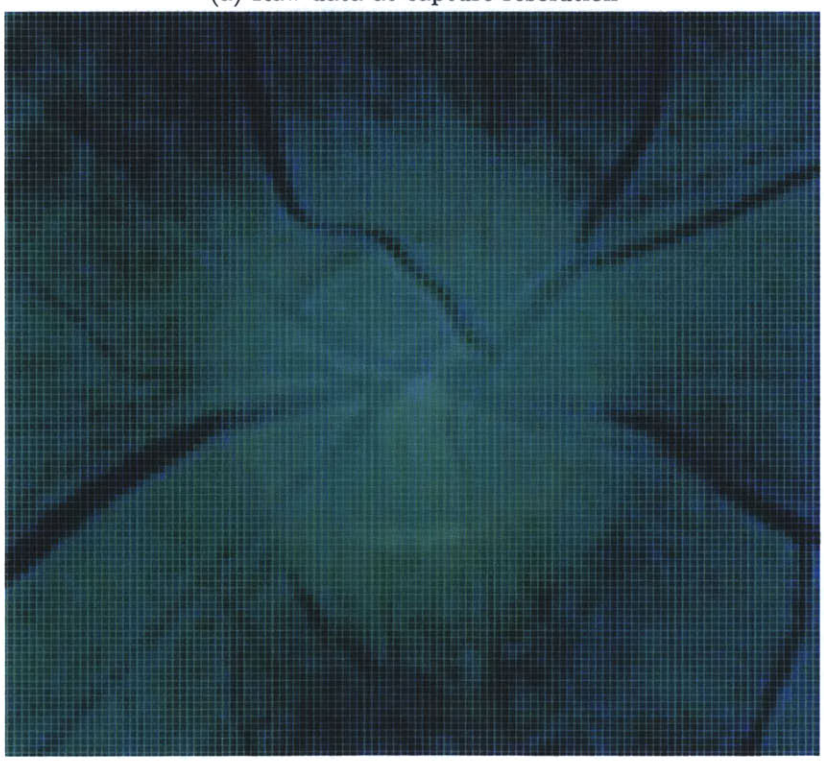


Figure 4-7: A plot of the cross-section of the retina (blue), and a cosine fit to the profile (red dash). We can observe that this is almost a perfect match.



(a) Raw data at capture resolution



(b) Integrated data at super-resolution

Figure 4-8: Comparison between resolutions of images before and after image integration. Relative pixel sizes are denoted with squares. The amount of detail in (b) is much greater than in image (a) at the same magnification

only at the high level but is apparent even on individual pixels.

Noise Reduction

One of the main issues with raw data is a high level of noise in each image, due to the low amount of light coming from the eye. We can identify noise reduction through visual inspection. Figure 3-18 shows a comparison of the frame before and after frame integration. The original image before image fusion contains a lot of noise and image artifacts, such as chromatic aberration (separation of colors in the image). However, the merged image does not suffer from as many as these artifacts. We remove the red channel to greater show the effects of image integration.

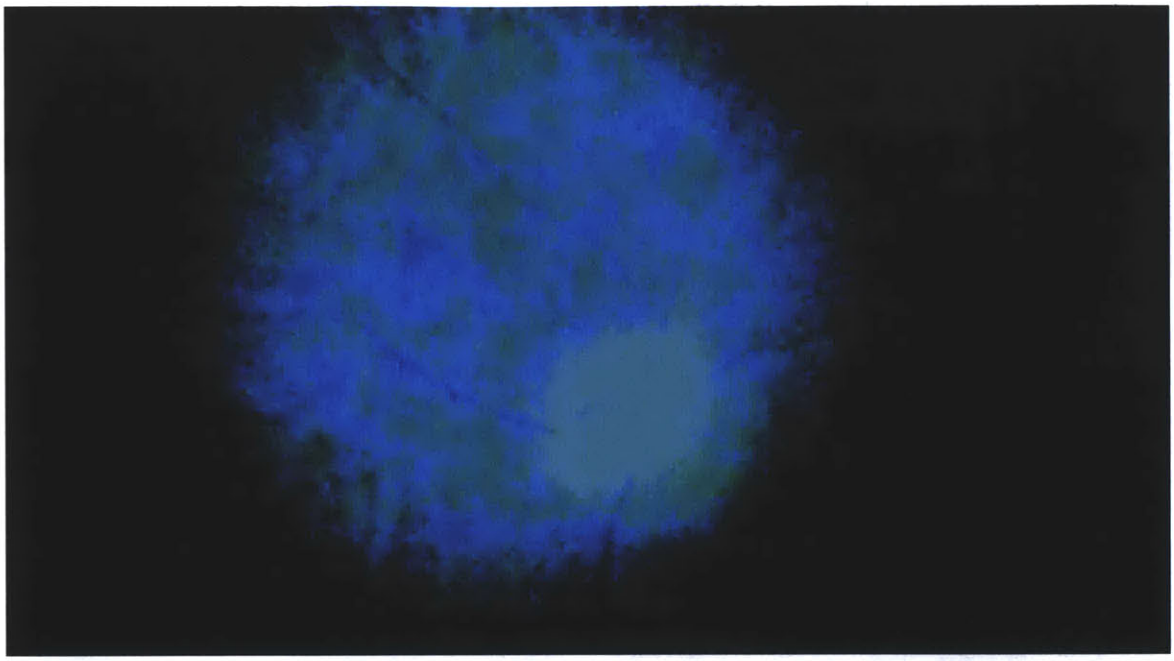
To verify the number frames needed in order to adequately remove noise, we can compare a noisy image with the ground truth. Seventy-five sequential images were averaged together to create a ground truth, de-noised image. The noise profile for one frame (raw data), three frames integrated together, and five frames integrated together can be seen in Figure 3-19. Note that these images do not have any super-resolution for simplicity.

By the fifth frame of integration, noise has been almost completely removed, which is a surprisingly small number of frames to integrate. When capturing video at 30 frames per second, this means we only need $200ms$ to sufficiently capture enough data to integrate frames at any given spot in the eye. Given our camera's field of view of about $15-30°$, we can capture a large field of view of the retina with stimulus control in about a second.

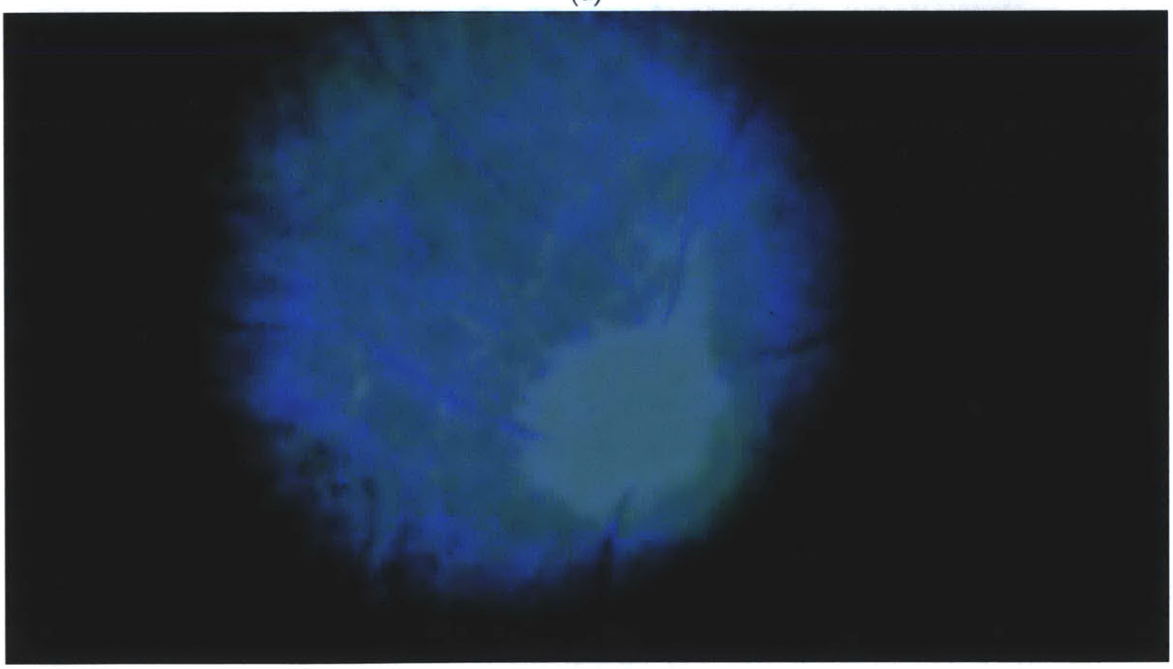
4.2.4 Frame Alignment Verification

Linear Motion

Our eye-worn design, in conjunction with stimulus computation, simplifies the number of variables in computation. By constraining rotations and translations for integration through physical form factor, we can speed up local registration by using phase correlation instead of feature matching. Stimulus control also reduces the number of



(a)



(b)

Figure 4-9: Noise removal due to image integration. The red channel has been removed to greater illustrate the effect of integration.

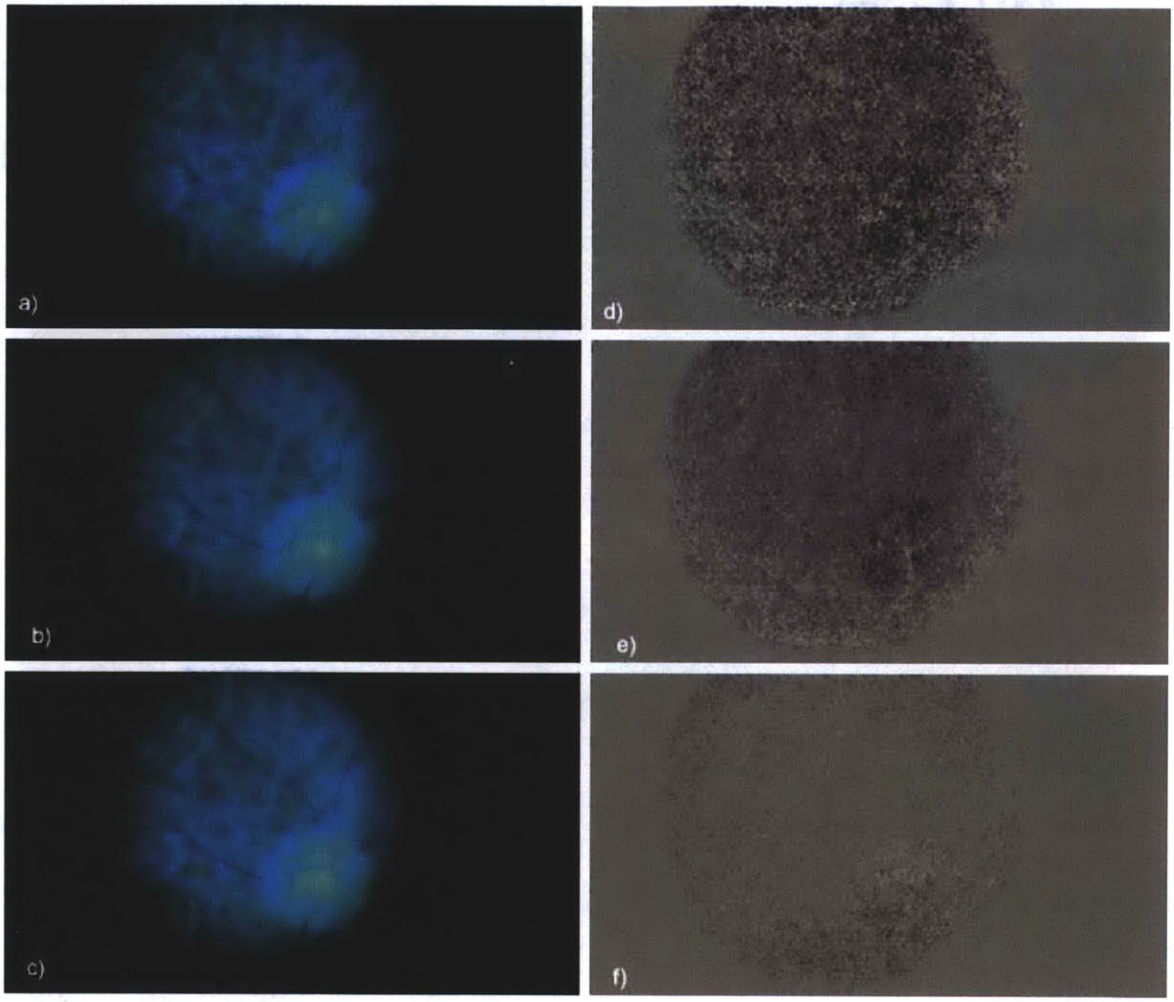
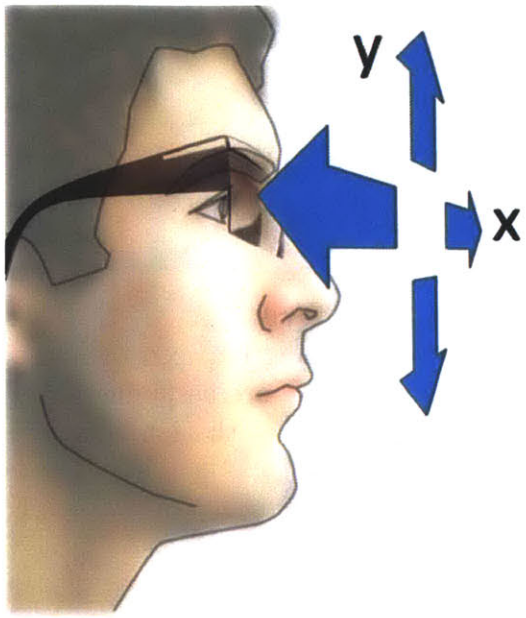


Figure 4-10: Noise removal due to image blending compared to the ground truth. The left column contains integrated images without the red channel after one frame, three frames, and five frames fused together. The right column shows the noise profile of each one of the images with respect to the ground truth. After about five images are integrated together the noise profile is almost zero.



(a)

(b)

Figure 4-11: Motion over small distances on the retina is fairly linear. (a) The rigid nature of our device, greatly constrains the degrees of freedom in our system. Instead of having six degrees of freedom (three rotation, three translation), we only have two degrees of translation as the eye moves around, up and down, left and right, behind the camera. (b) Markers show the tracking of features on the surface of the retina as the eye moves around. The distance between these markers stays constant indicating the relative magnification of observed features is nearly constant.

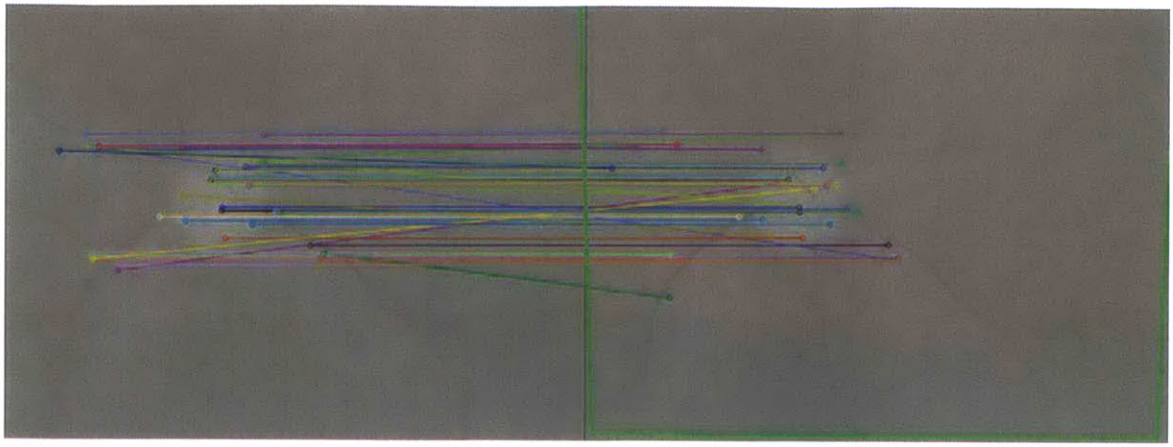


Figure 4-12: An example feature matching between two adjacent frames. Matching points are connected with lines. Since we are assuming all of our motion is translational, all lines should be parallel. Images with high noise are relatively feature-less, which results in low feature detection.

image pairs needed for registration in mosaicing.

Since magnification of objects on the retina change slowly as the camera is moved within five degrees, we can approximate the motion of frames for the integration step to be nearly linear (See Figure 4-11). This also verifies that only using translation between corresponding frames as opposed to rotation and scaling is correct.

Motion Methods Comparison

To verify phase correlation produces accurate results, we compared it to other methods for finding translation between images. State-of-the-art image matching uses feature matching to find corresponding features in two images, match them, and then calculate a homography between the two coordinate systems. This tends to work best for most real-world situations because natural images have a lot of hard features, and motion is not limited to a simple translation. Often times when tracking objects in real space, they will rotate and scale, which feature matching can detect (See Figure $4-12$).

Template matching is an extension of optical flow in two dimensions. A portion of one image, the template, is convoluted with the another image, the scene. The region of the scene that most closely corresponds to the template will result in a high match

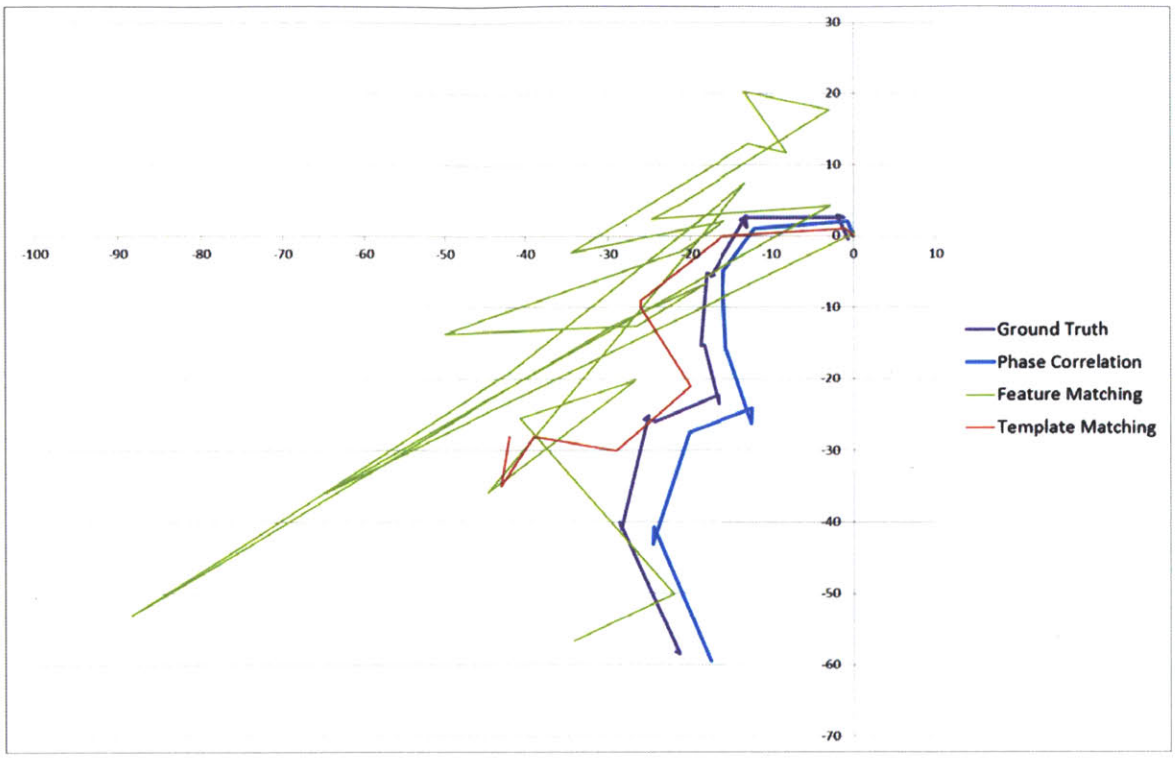


Figure 4-13: Translation #1 of ten neighboring frames

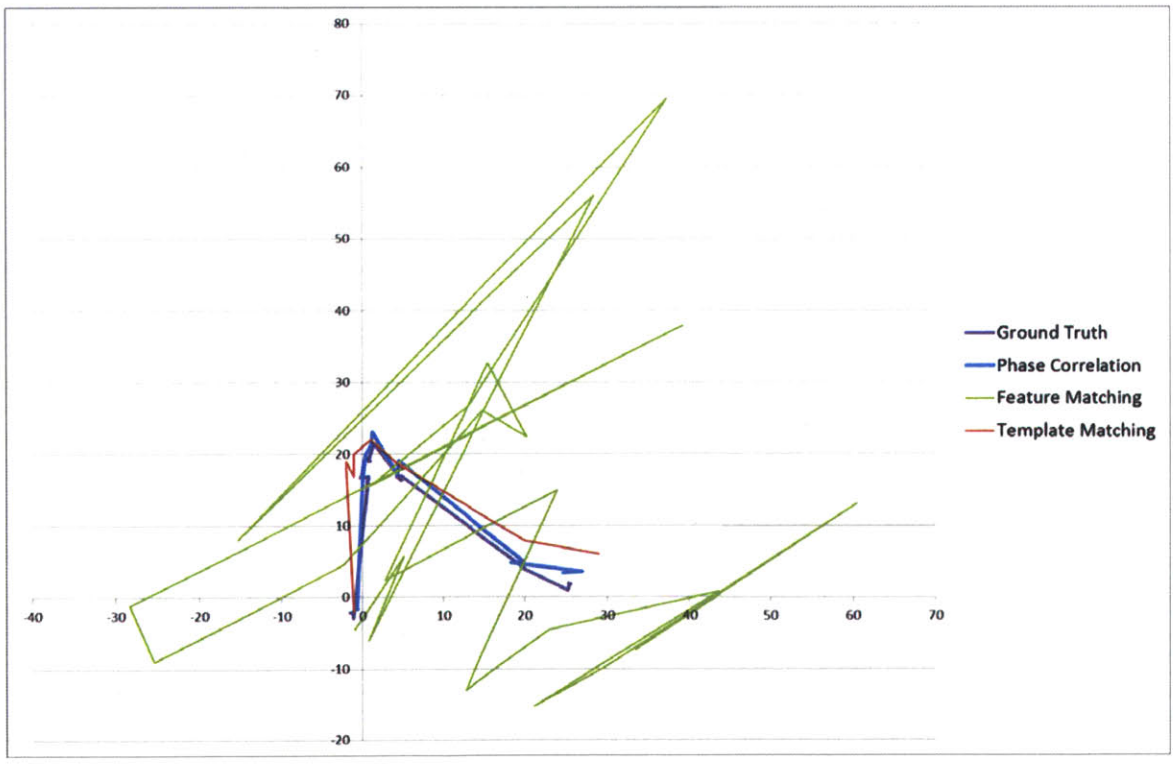


Figure 4-14: Translation #2 of ten neighboring frames

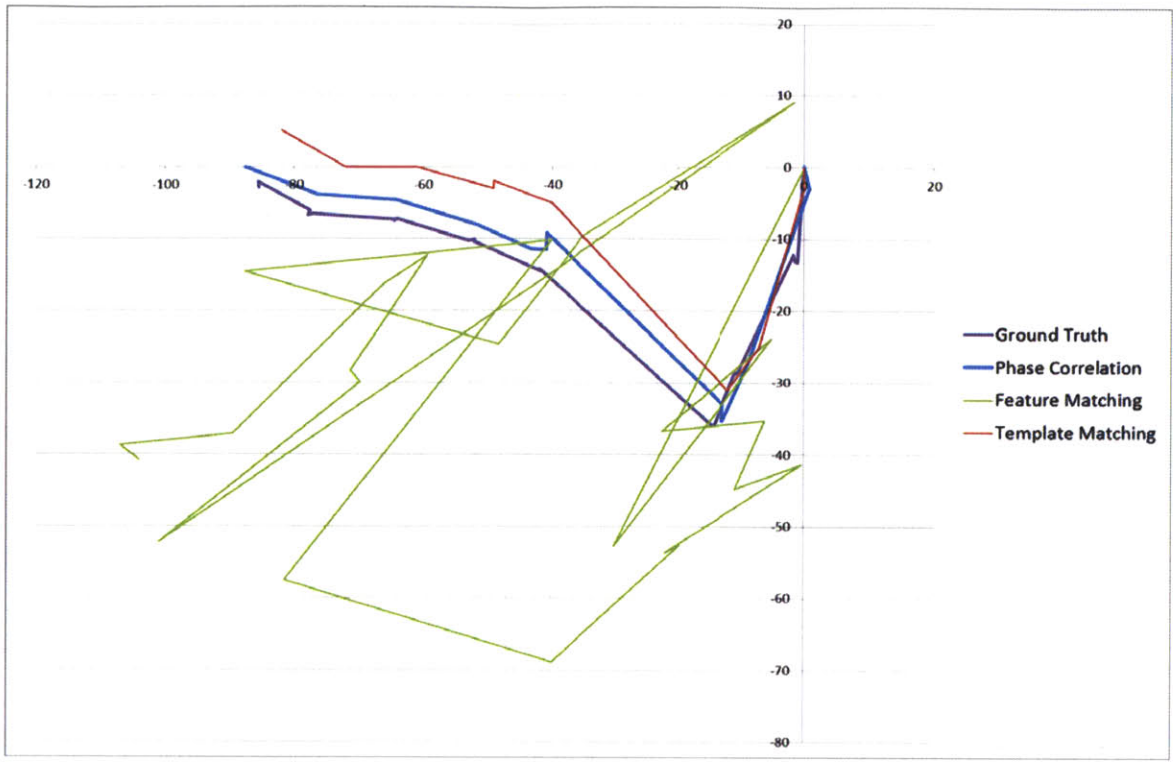


Figure 4-15: Translation #3 of ten neighboring frames

response at that location. This is the same as doing optical flow and only considering translational motion.

We developed a ground truth for tracking motion of retinal images by identifying the optic disc in the image and recording its motion. Obviously this method would not work for all portions of the retina since the optic disc is a small, but vibrant, feature. Results of the tracking methods can be seen in Figures $4-14$, $4-15$, and $4-16$. In all three test cases presented, phase correlation closely follows the ground truth. Template matching is generally a close match for relatively little motion, but does not keep the same shape or form as the ground truth like phase correlation does. Feature matching completely fails due to lack of ability to find an adequate number of features in the images.

In fact, in real life tests we have observed that feature matching has over a 99% success rate. In a set of over 1000 integrated images, so far we have only observed two failures.

4.3 Results

The following contains images of panoramas obtained by this system. For more examples of data captured and processed, please see the Appendix.

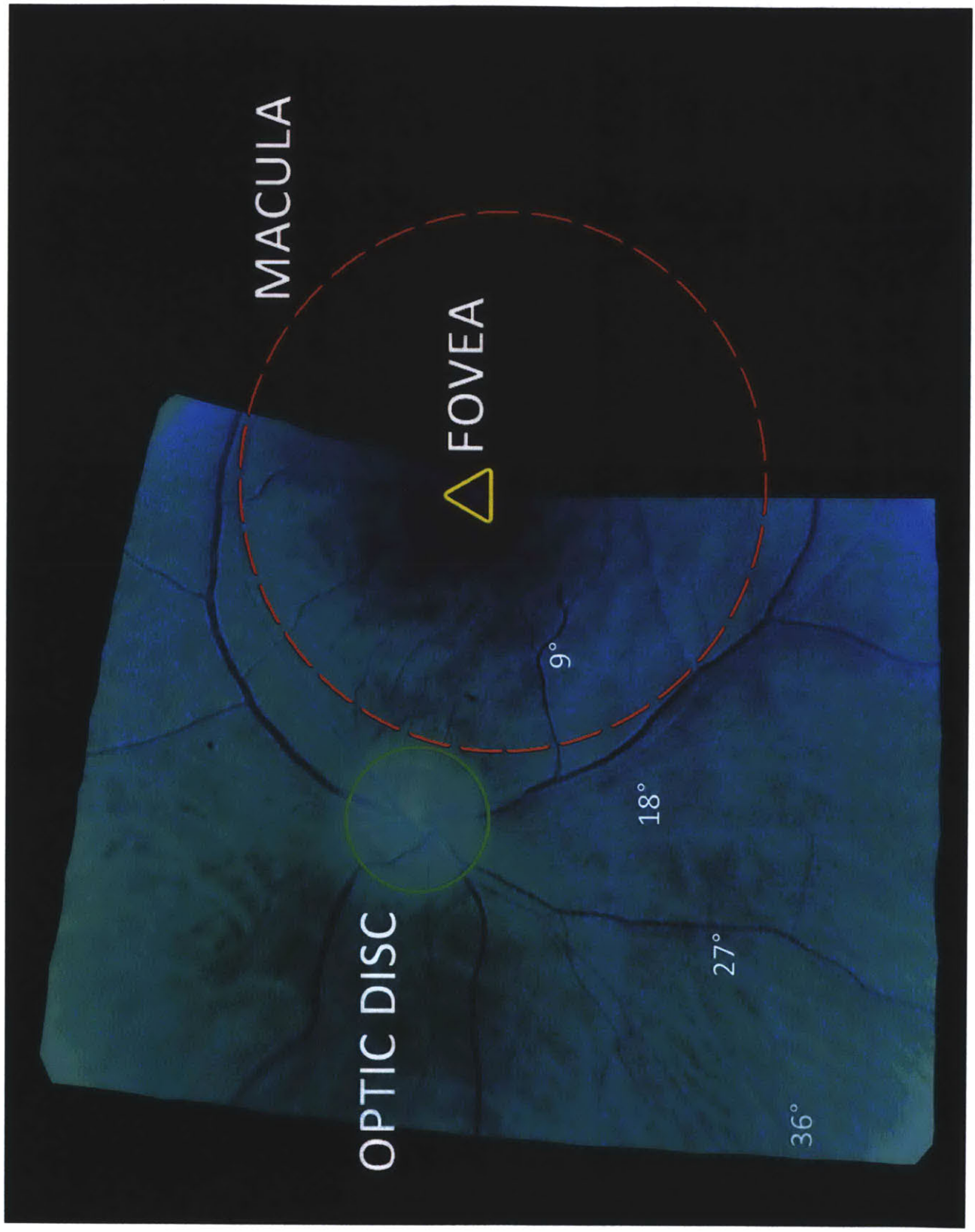


Figure 4-16: Panorama mapped to a sphere and labeled in terms of features and field of view.

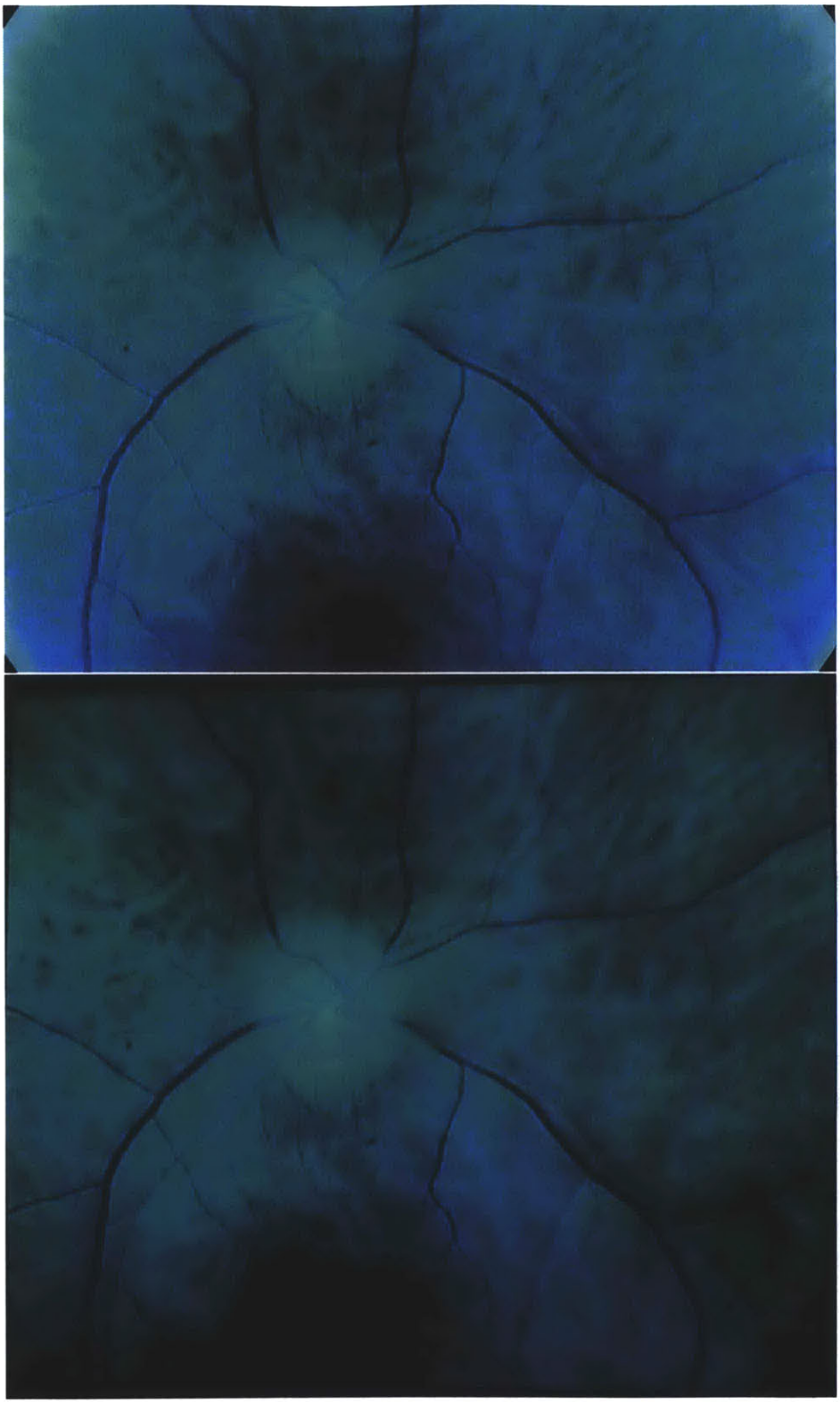


Figure 4-17: Examples of stitched images. The top image was created automatically by the free Hugin program downloadable from http://hugin.sourceforge.net/. The bottom image was created in our integrated system using C++ and the OpenCV libraries. The Hugin stitcher yields an image that is more visually pleasing, but with lower fidelity and contrast.

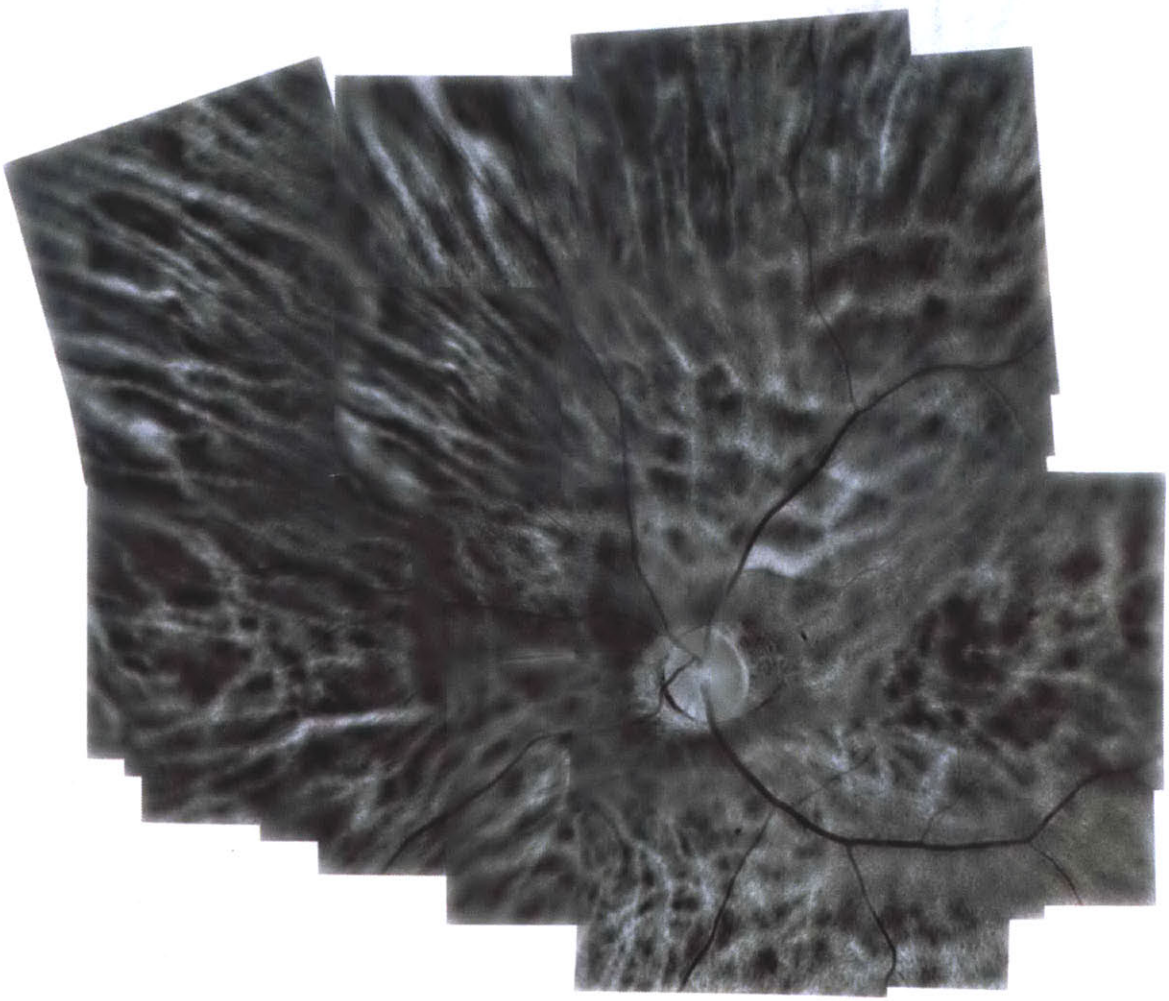


Figure 4-18: Large field of view Panorama created by combining 80 images from a PointGrey Flea 3.0 Camera.

Chapter 5

Future Research and Conclusions

5.1 Future Research

Work in this section is provided for educational use to supply possible future endeavors for this research. I have to give special thanks to Di Wu, Everett Lawson, and Ramesh Raskar who spent time helping to write the following text.

5.1.1 Light Field Capture

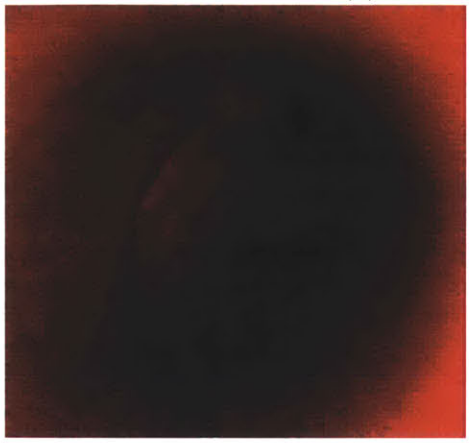
Traditional retinal imaging methods rely on manual focus in order to obtain a sharp image, however clinicians must also control the focus of the eye being examined, which could be constantly changing. Light field capture may allow for synthetic refocusing to give depth perception [26]. Refocusing will eliminate the need for auto-focus assembly and allow better focal-stack analysis but our micro-lenses are currently low quality.

5.1.2 Hyperspectral Imaging

It has been shown that different retinal structures are visible with different wavelengths [5]. We aim to confirm this with our portable setup by observing long wavelengths expose the choroid layer behind the sclera, while shorter wavelengths emphasize vascular structure. Multispectral imaging may reveal hidden features that

(a) Illumination from side 1

(b) Illumination from side 2



(c) The raw difference from when image 1 is subtracted from image 2

Figure 5-1: This is example of multidirectional lighting. Two LEDs placed closely to the eye on opposite sides are modulated so that only one LED is on at a time. By observing shadows in images we can get some depth information about the retinal tissue.

were previously unseen.

5.1.3 Multidirectional Capture

When using a direct ophthalmoscope the doctor is able to get a stereoscopic view of the eye since he is using both of his eyes to view the retina. While our data capture is currently in two dimensions, a flat image on a sensor, there are many ways of retrieving this depth information for the user. By modulating light flow in different areas around the exterior of the eye we can create shadow and height patterns that

appear on the retina. By extracting the relative difference between the shadow and the vessel itself we can infer the relative height of the vessel from the rest of the retinal tissue (see Figure 5-1).

An extension of multidirectional lighting may also allow for photometric stereo on the retinal surface to determine the surface topography. However, it is important to note that the retinal surface is extraordinarily shiny and does not follow a Lambersian model.

5.1.4 Platform integration

While mostly algorithms and methods discussed in this thesis were implemented on a PC using the C++ programming language and OpenCV,the techniques discussed are not limited to that design. Both android devices, through the Java native interface, and IOS devices, through their native interface, can execute C++ programs as well as run the OpenCV library. While processing may take a lot longer on portable devices with lesser processors, the overall results will be the same.

Furthermore, mobile devices offer a unique opportunity to allow image upload anywhere in the world. Since the problems presented in this paper do not require any user input to solve, images can be uploaded blindly, in order, and correlated together automatically into a giant panorama. Image upload can be done through the HTTP PUT protocol. Resulting images can be downloaded to user's device to give doctors on the ground a view of the retinal image, or diagnosed far away and have results returned a couple of days.

In the summer of 2012 we will implement such a system to begin clinical trials.

5.1.5 Video Overlay

Downscaling images is extraordinarily fast if you use the processing power of the graphics processing unit (GPU) to your advantage, and since phase correlation of two images takes less than 100th of a second per frame, we can overlay live video of the eye in real time on top of a panoramic image. This could allow for the doctor,

or attending physician, to view other areas of the eye in more detail. This system has already been implemented and will used in the clinical trials, but not part of this thesis.

5.2 Awards and Funding Sources

Our retinal imaging system has been to multiple different conferences and contests. We are currently finalists in the CIMIT Prize for Primary Healthcare Competition, and recently won $7,500 in the 2012 IDEAS Global Challenge.

I am funded in part by the National Science Foundation Graduate Research Fellowship Program, and appreciate all of their support throughout my years at MIT.

5.3 Conclusion

This thesis lays the groundwork for a portable computational retinal imaging system. We show how images can be enhanced and mosaiced together into a large field of view panorama of the eye by first registering the location of the retina on the sensor; correcting the image for color and vignetting; locally integrating frames together to get rid of noise and lens effects; and finally stitching together integrated images into a large field of view mosaic. We explore how motion is nearly linear on the retinal surface as the eye moves around, and show how phase correlation outperforms other alignment and correction techniques.

I hope this work in conjunction with portable retinal imaging hardware will help give eyesight to millions around the world.

As I leave this phase of my life and start the next, I am grateful for the friends I have made; opportunities I have had to explore and learn; and the research I have been able to conduct. I go forth into the world knowing more now than I did two years ago, and hope that I will put the knowledge to use in the upcoming years.

I want to thank again all of my family and friends who have supported me throughout the years, and I hope this research will inspire others to go change the world.

Appendix A

Figures

A.1 Examples of Raw Data

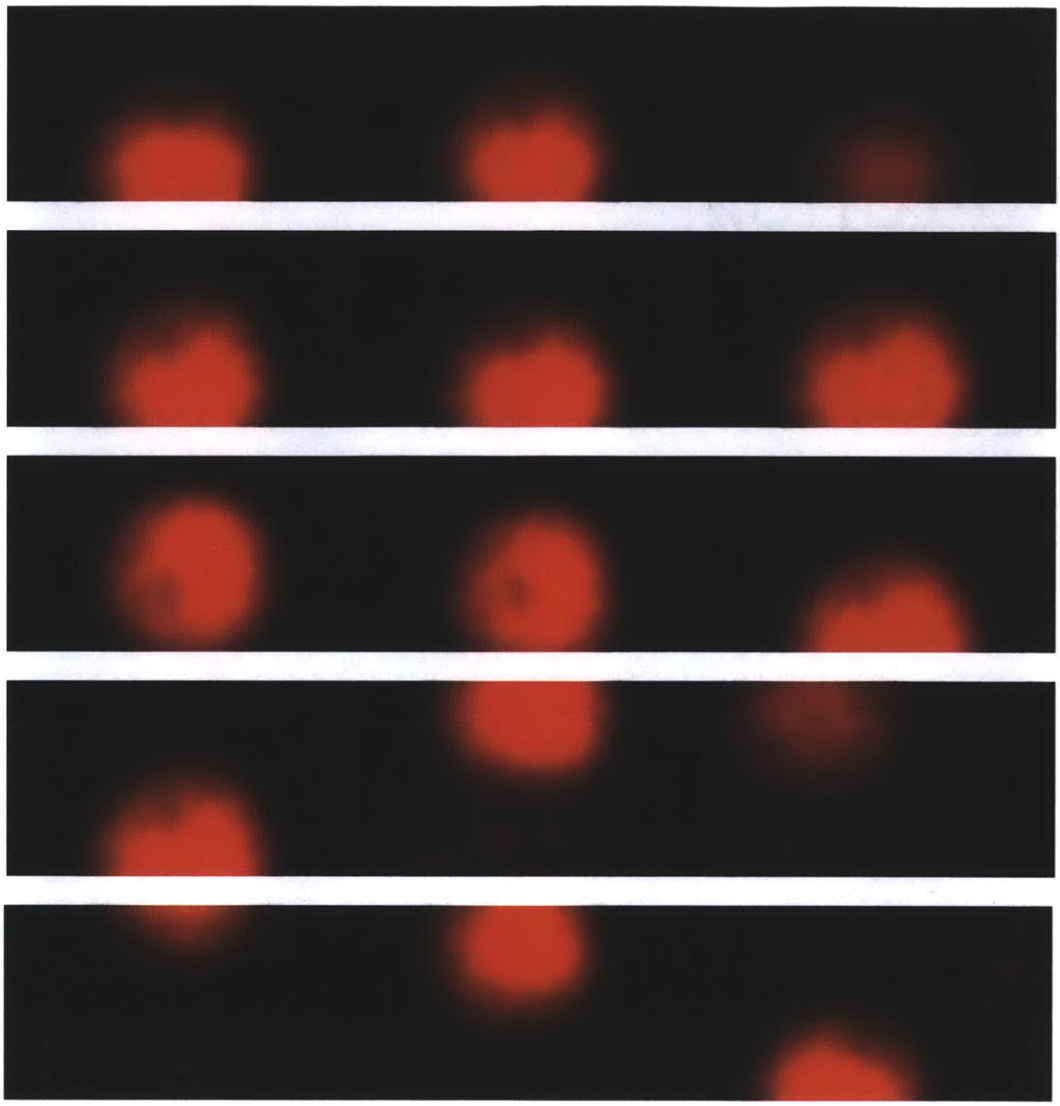


Figure A-1: Raw data from dataset 10, captured with a standard HD webcam.

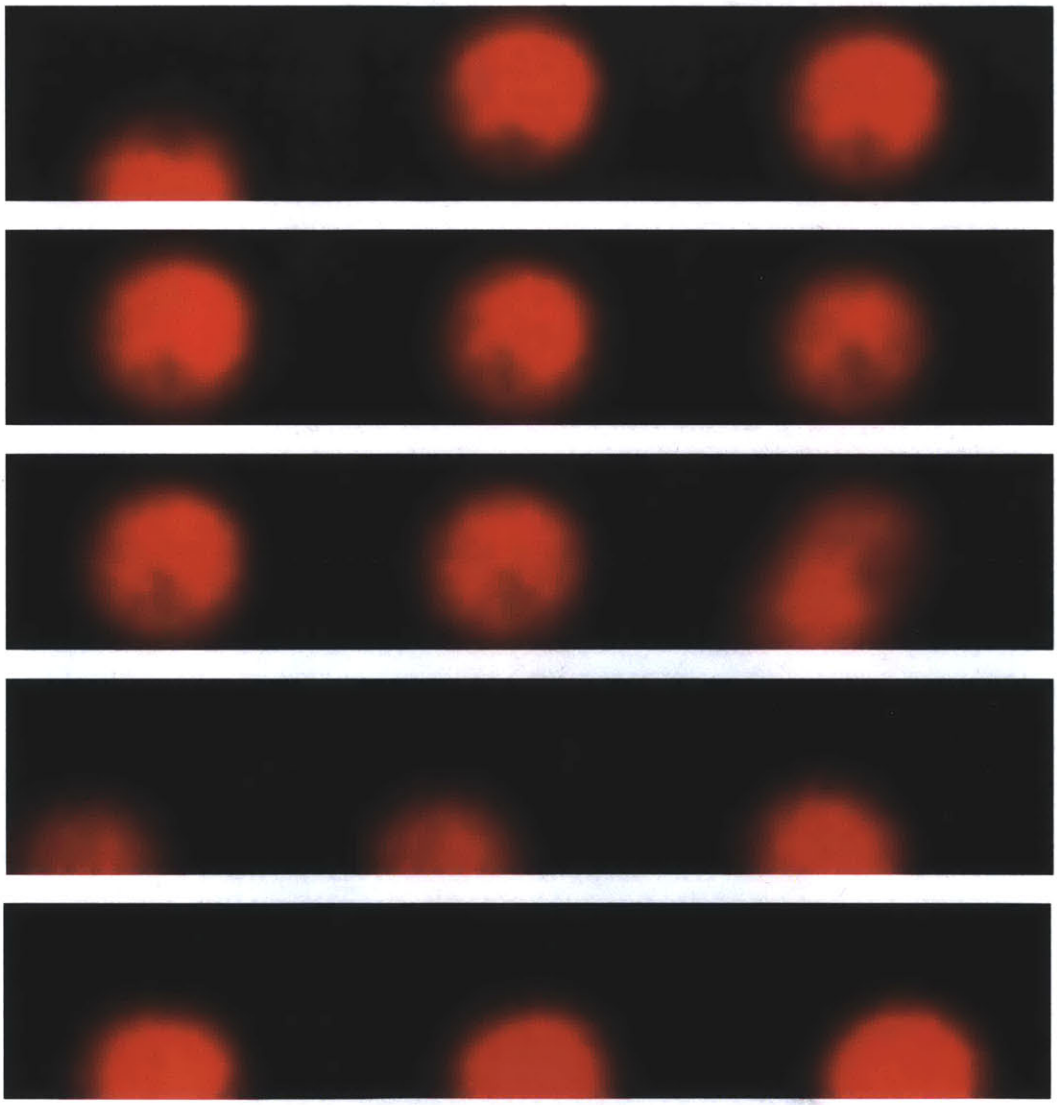


Figure A-2: Raw data from dataset 10, continued

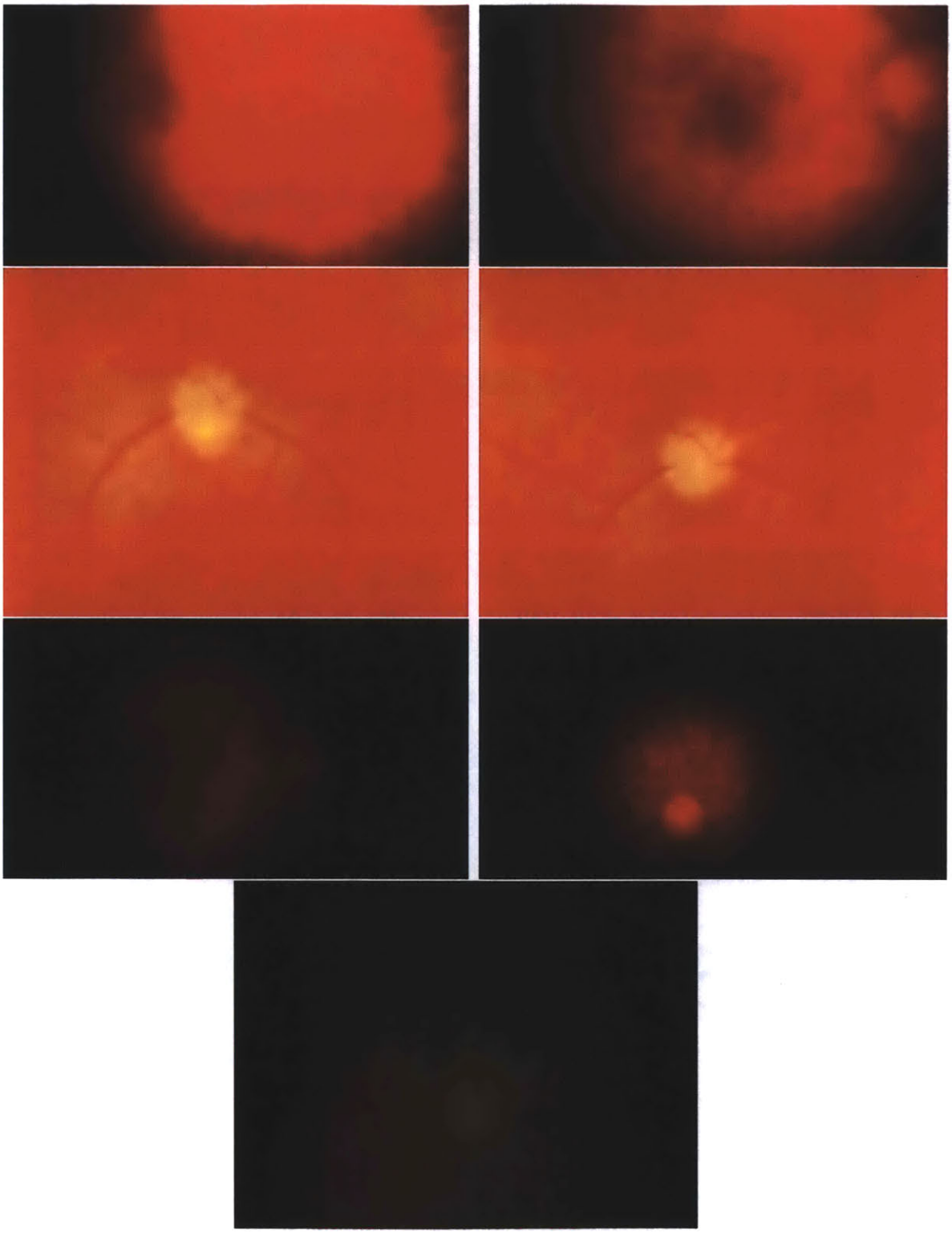


Figure A-3: Examples of raw data from a diverse set.

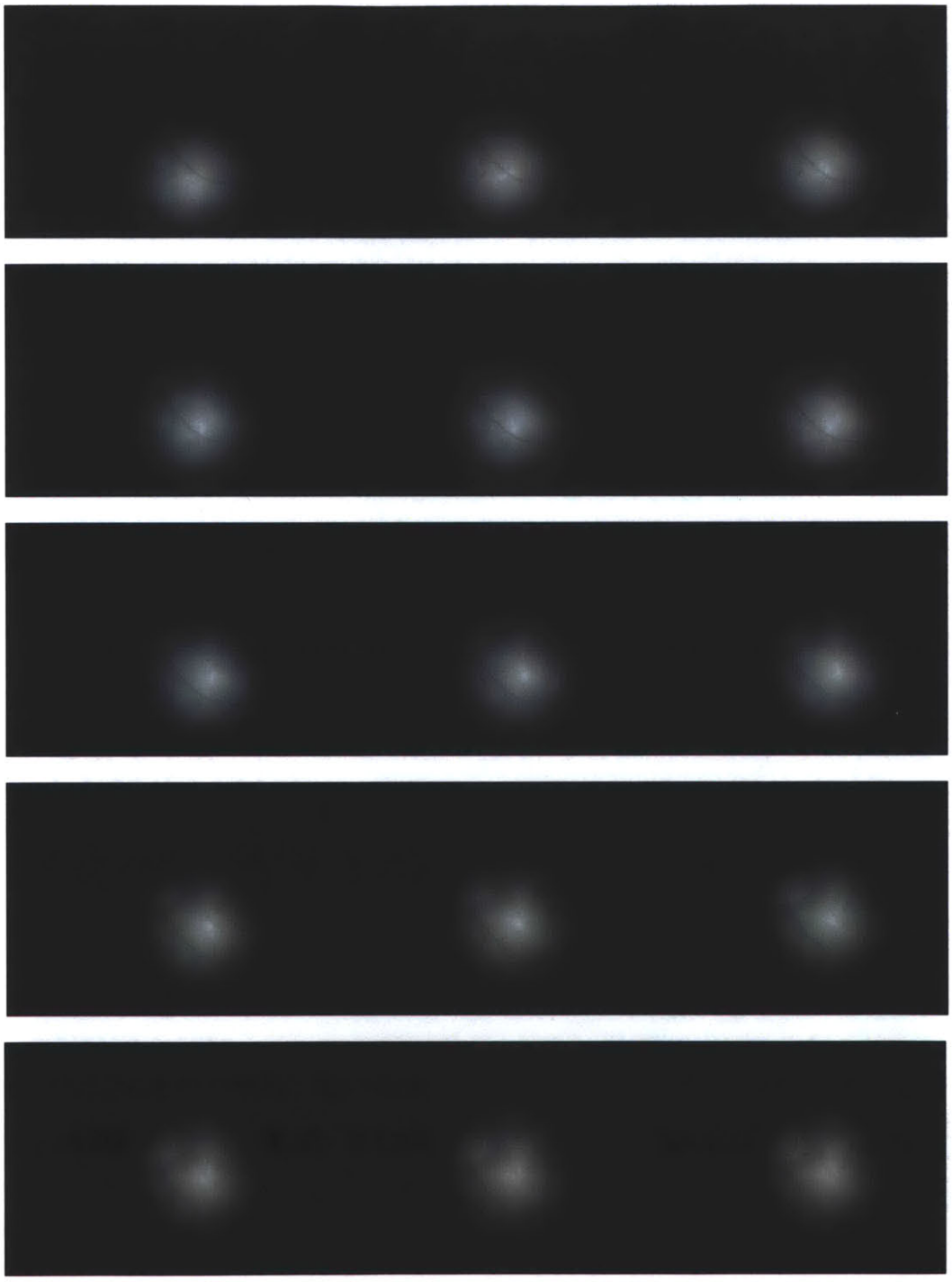


Figure A-4: Raw data from the PointGrey Flea 3.0 camera.

A.2 Examples of Data after Integration and De-Vignetting

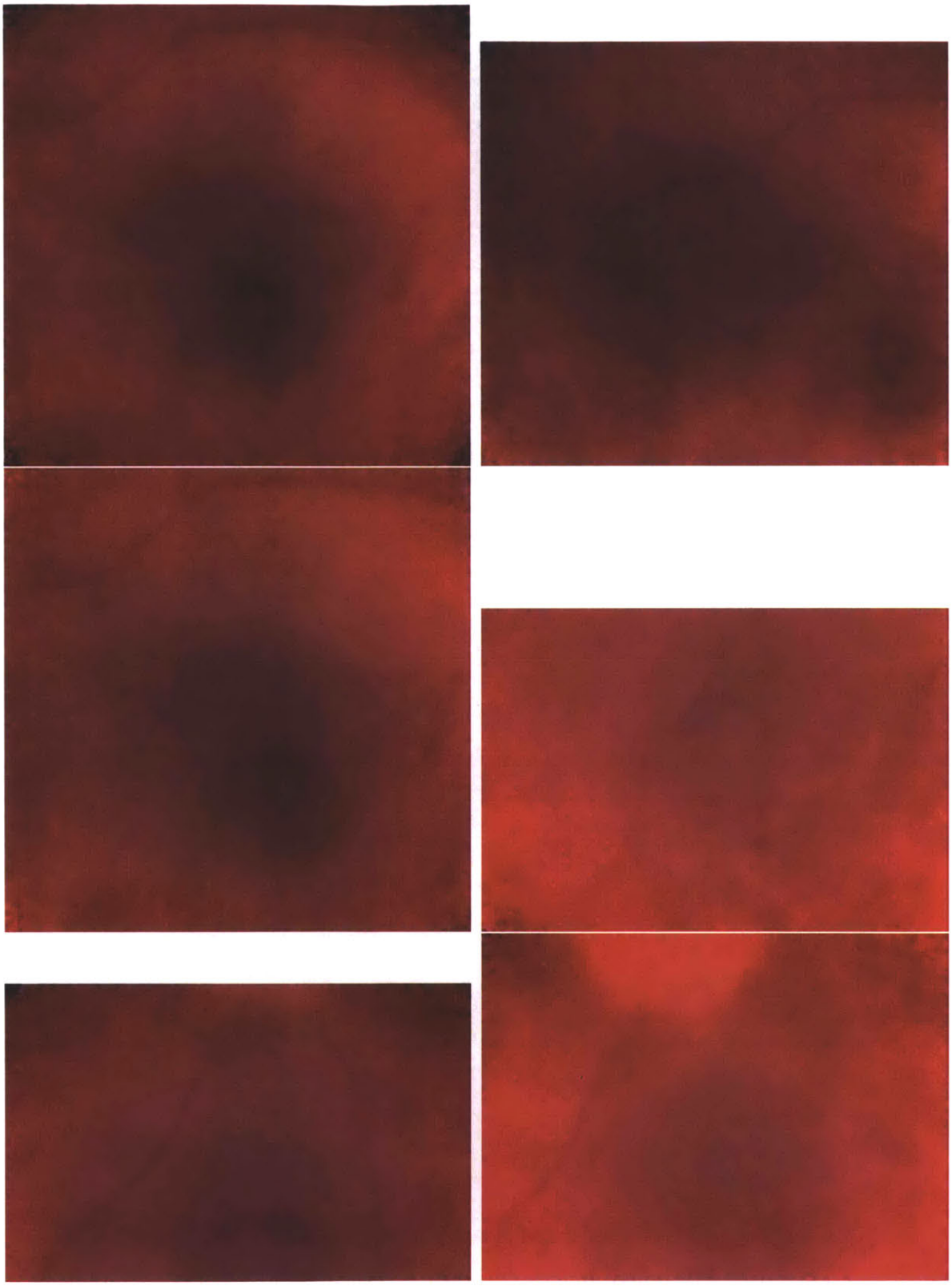


Figure A-5: Flattened images without integration.

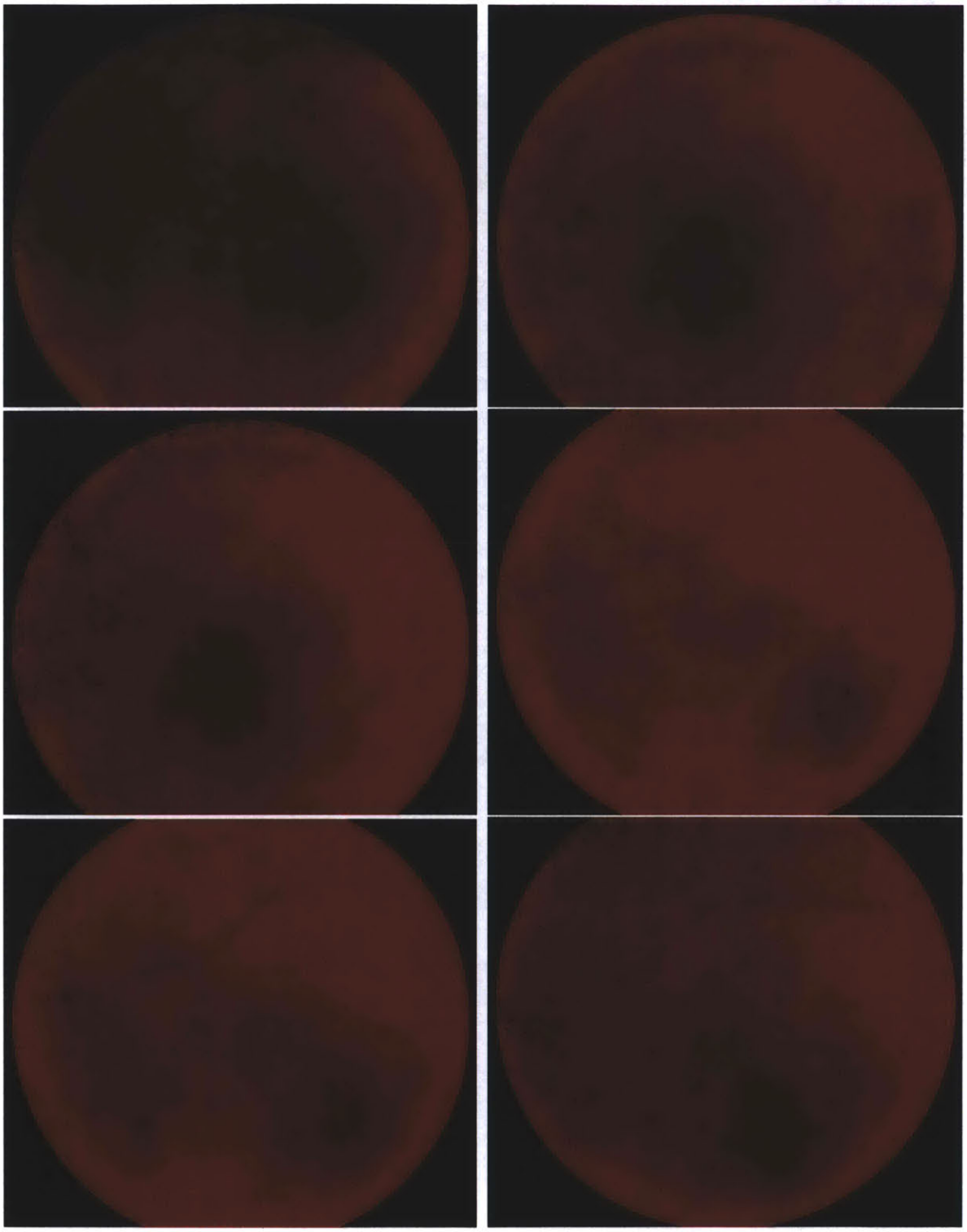


Figure A-6: Flattened images without integration and without cropping.

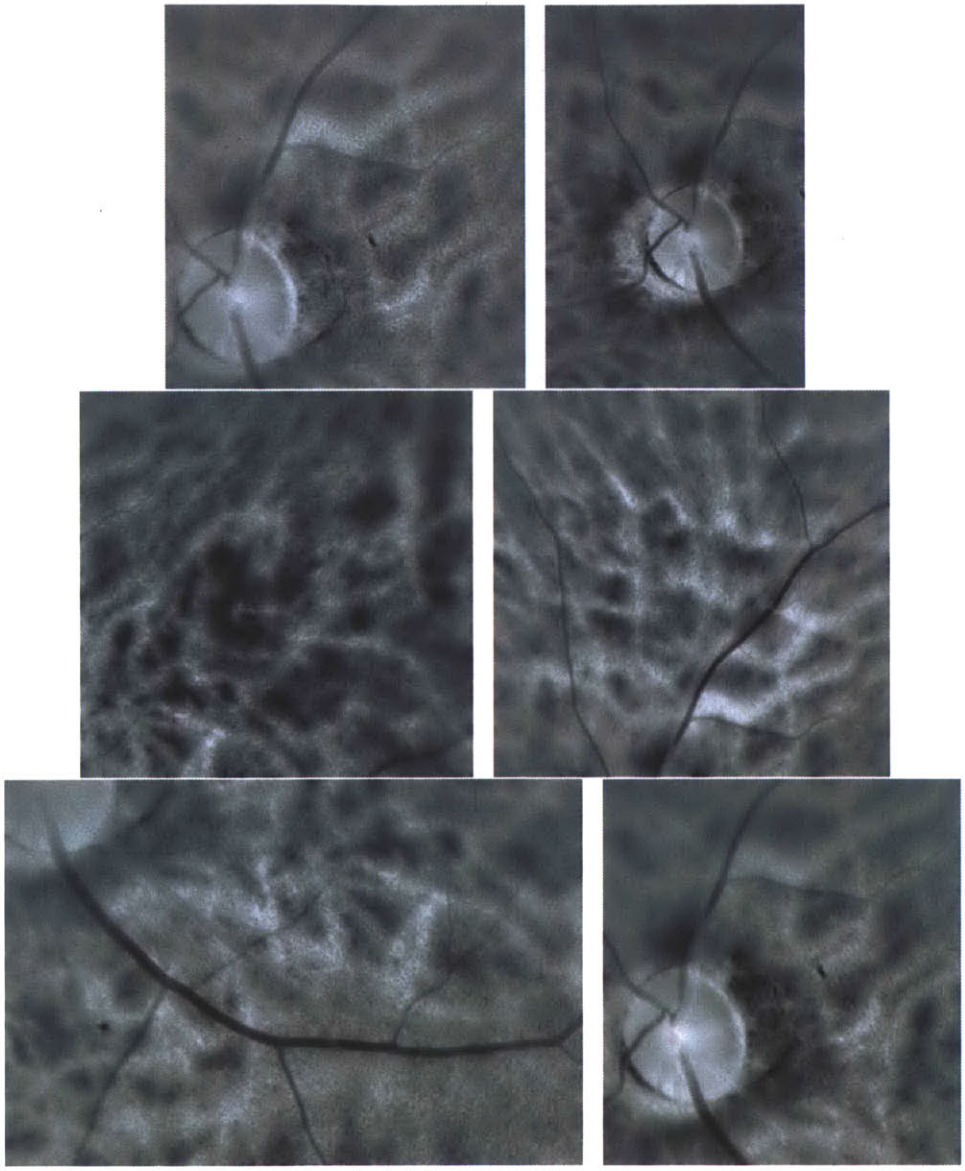


Figure A-7: Integrated images from the PointGrey Flea 3.0 camera.

A.3 Panoramas

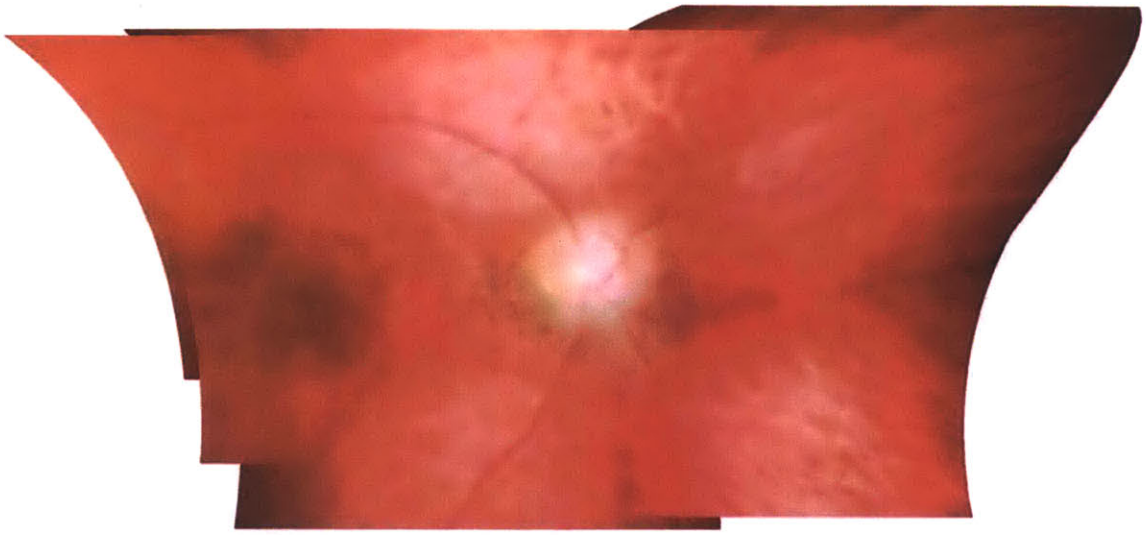


Figure A-8: Our first panorama, created by hand.

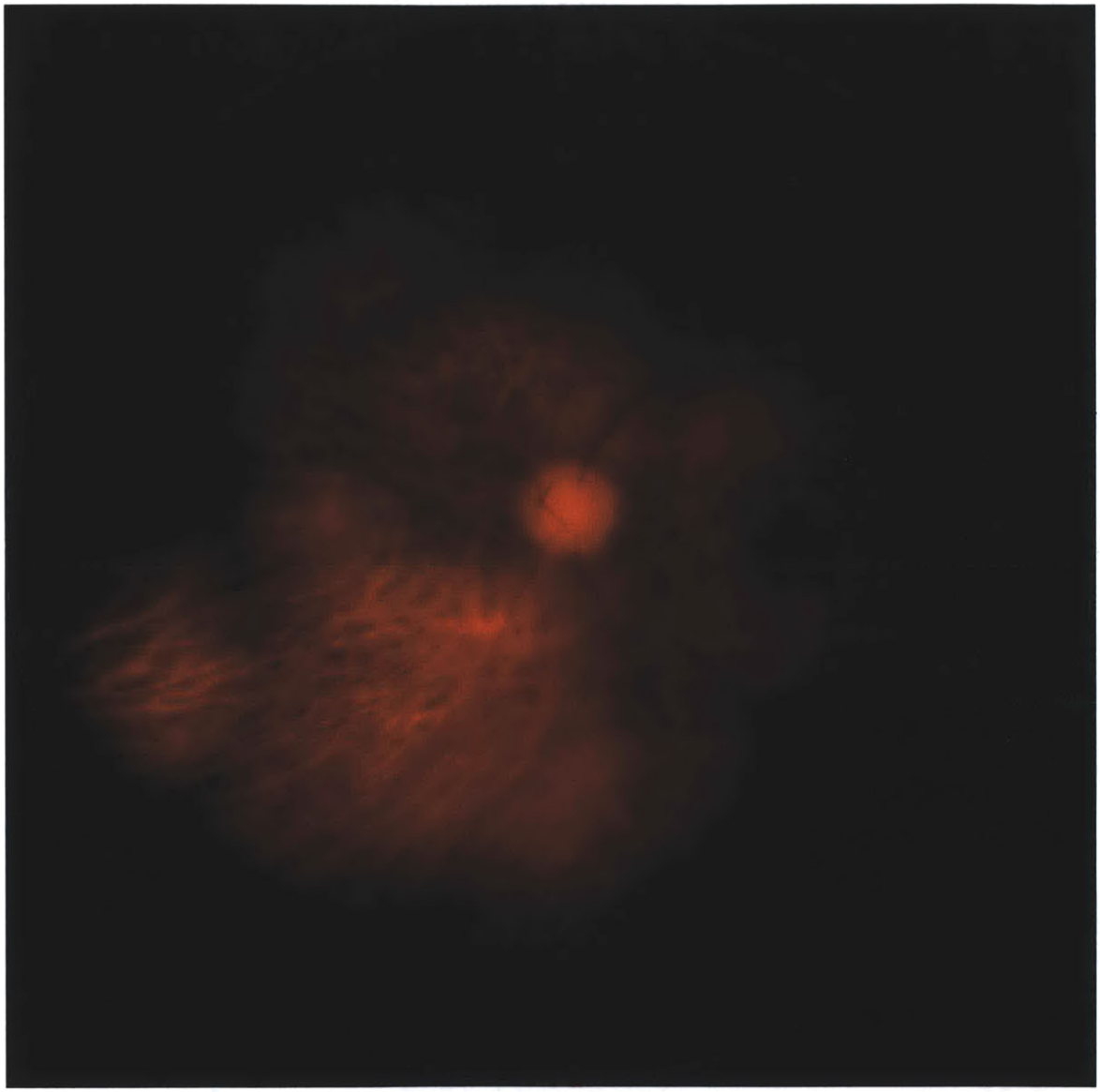


Figure A-9: Our second panorama, created by hand and rendered in 3D Studio Max.

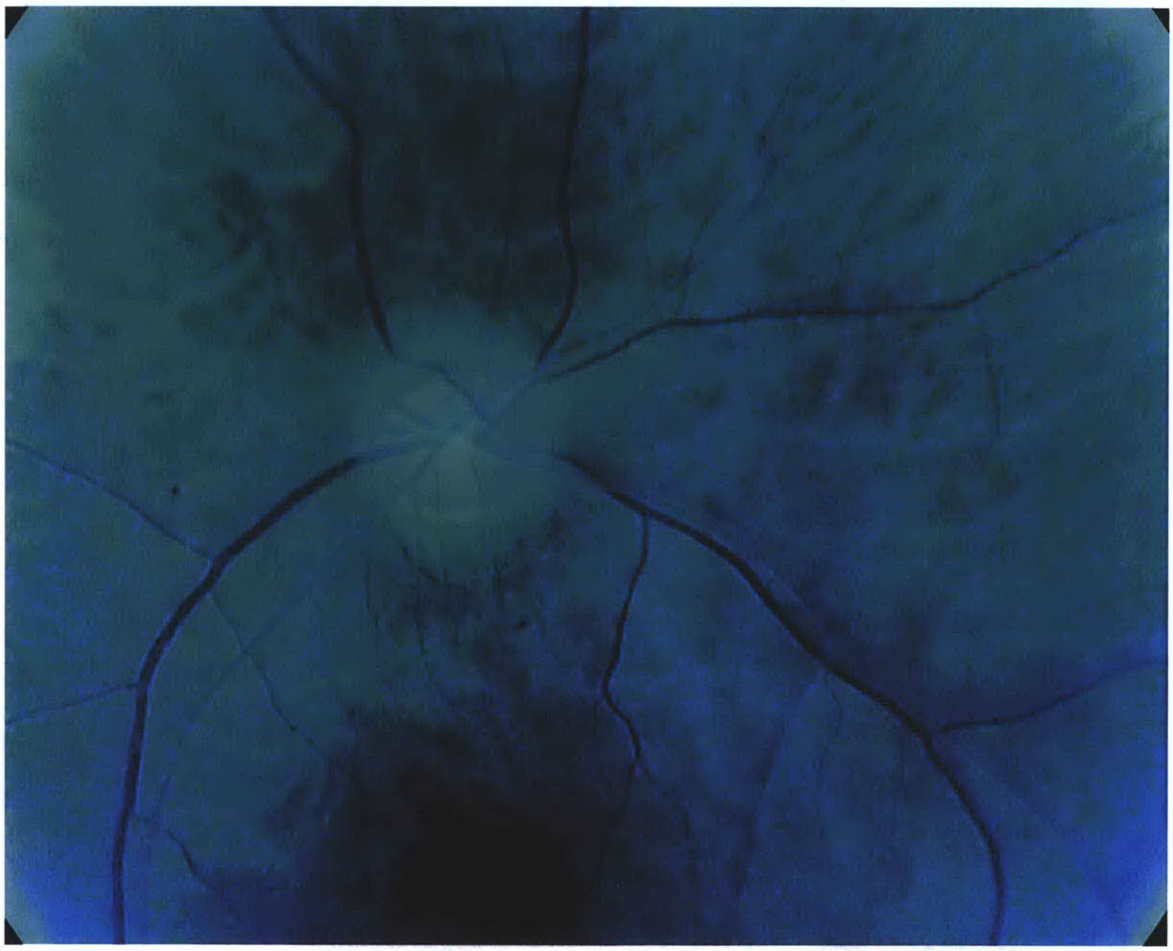


Figure A-10: First automatically stitch panorama. Created with Hugin

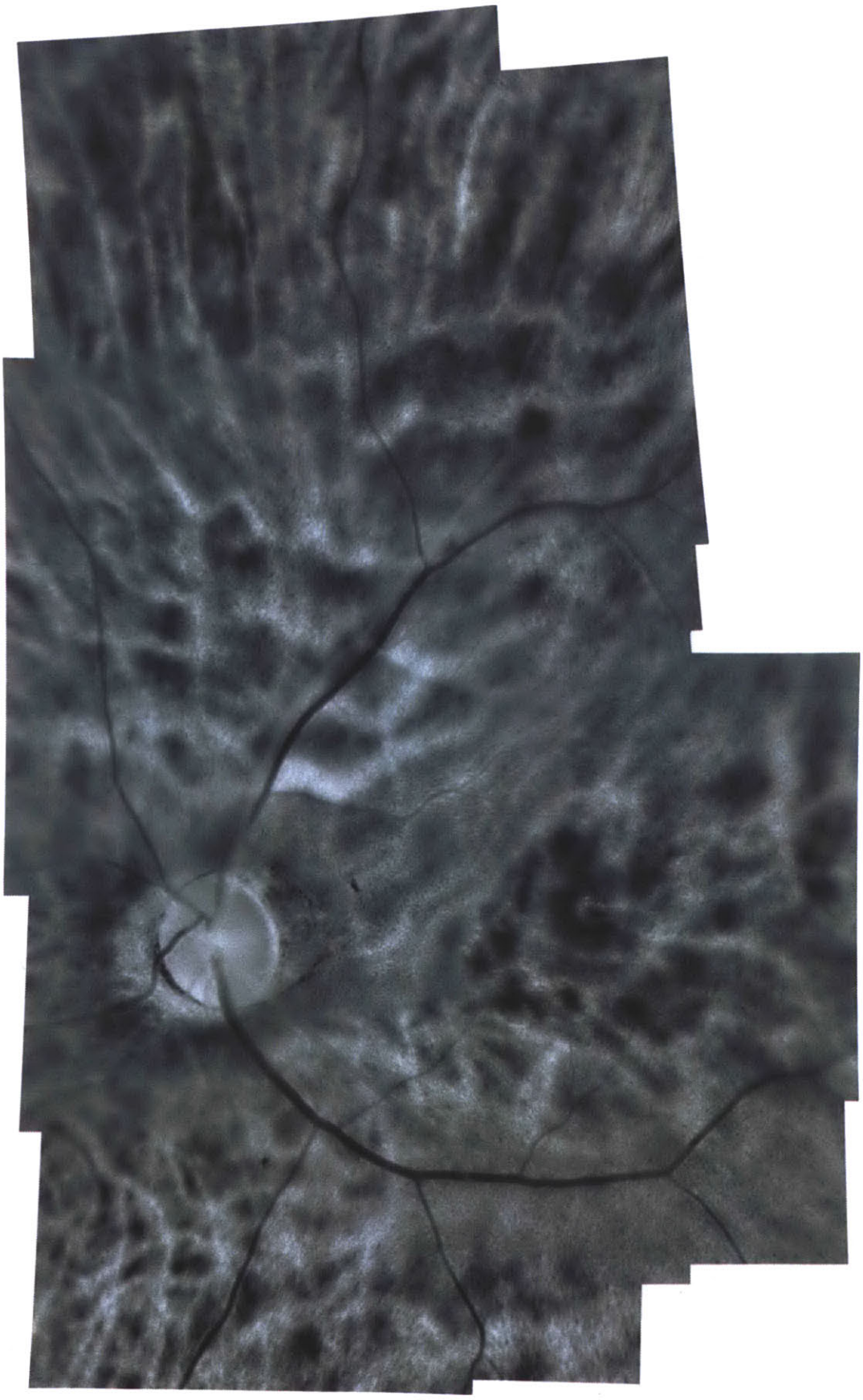


Figure A-11: Central region of data automatically stitched together from image integrated together from a PointGrey Flea 3.0 camera.

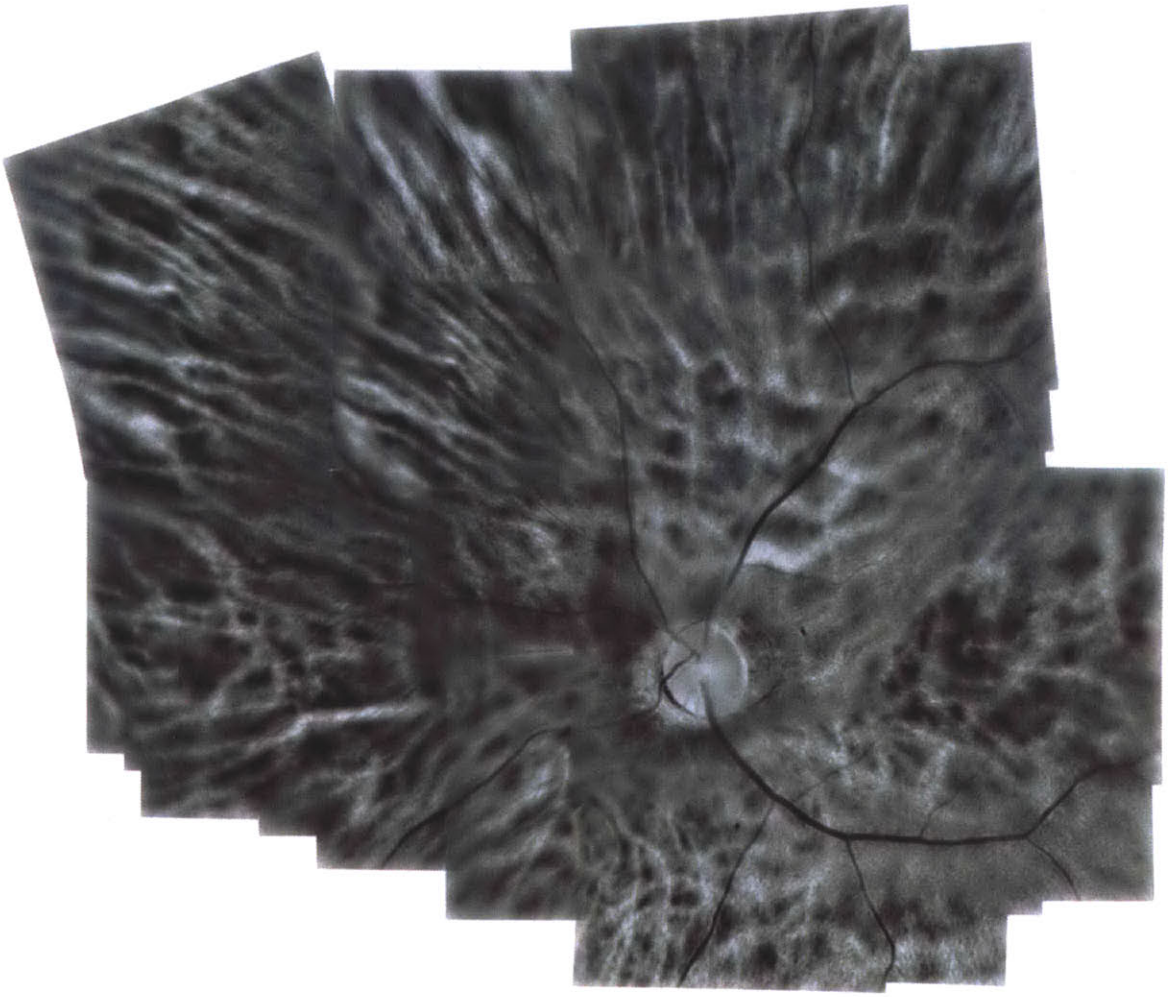


Figure A-12: Largest, most detailed panorama

Bibliography

[1] iexaminer for the panoptic, 2010.

[2] L. Allen. Ocular fundus photography: Suggestions for achieving consistently good pictures and instructions for stereoscopic photography. *American journal of ophthalmology*, 57:13, 1964.

[3] M. Brown and D.G. Lowe. Recognising panoramas. In *Proceedings of the Ninth IEEE International Conference on Computer Vision*, volume 2, page 5, 2003.

[4] M. Brown and D.G. Lowe. Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision*, 74(1):59–73, 2007.

[5] A. Calcagni, JM Gibson, IB Styles, E. Claridge, and F. Orihuela-Espina. Multispectral retinal image analysis: a novel non-invasive tool for retinal imaging. *Eye*, 25(12):1562–1569, 2011.

[6] J.L. Calkins and B.F. Hochheimer. Retinal light exposure from ophthalmoscopes, slit lamps, and overhead surgical lamps. an analysis of potential hazards. *Investigative Ophthalmology & Visual Science*, 19(9):1009, 1980.

[7] P. Cattin, H. Bay, L. Van Gool, and G. Székely. Retina mosaicing using local features. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*, pages 185–192, 2006.

[8] J. Cui, Y. Wang, J.Z. Huang, T. Tan, and Z. Sun. An iris image synthesis method based on pca and super-resolution. In *Pattern Recognition, 2004. ICPR 2004.*

Proceedings of the 17th International Conference on, volume 4, pages 471–474. IEEE, 2004.

[9] P. dàAngelo. Radiometric alignment and vignetting calibration. *Proc. Camera Calibration Methods for Computer Vision Systems*, 2007.

[10] M.F. Deering. A photon accurate model of the human eye. volume 24, pages 649–658. ACM, 2005.

[11] E. DeHoog and J. Schwiegerling. Fundus camera systems: a comparative analysis. *Applied optics*, 48(2):221–228, 2009.

[12] FC Delori, JS Parker, and MA Mainster. Light levels in fundus photography and fluorescein angiography. *Vision Research*, 20(12):1099–1104, 1980.

[13] S. Erturk. Digital image stabilization with sub-image phase correlation based global motion estimation. *Consumer Electronics, IEEE Transactions on*, 49(4):1320–1325, 2003.

[14] P.A. Filar et al. Apparatus for photographing the anterior segment and retina of the eye through the use of a camera attachment, 2011. US Patent 7,883,210.

[15] Center for Disease Control. Common eye disorders, 2009.

[16] H. Foroosh, J.B. Zerubia, and M. Berthod. Extension of phase correlation to subpixel registration. *Image Processing, IEEE Transactions on*, 11(3):188–200, 2002.

[17] W.T. Freeman, T.R. Jones, and E.C. Pasztor. Example-based super-resolution. *Computer Graphics and Applications, IEEE*, 22(2):56–65, 2002.

[18] E. Goldfain, W. Lagerway, C.R. Roberts, S.R. Slawson, and A.I. Krauter. Eye viewing device for large field viewing, 2005. US Patent 6,939,006.

[19] M.J. Hogan and JE JA Weddell. Histology of the human eye: an atlas and textbook. 1971.

[20] M. Ishihara and T. Kogawa. Fundus camera, 2008. US Patent 7,377,642.

[21] C.R. Keeler. 150 years since babbage's ophthalmoscope. *Archives of ophthalmology*, 115(11):1456, 1997.

[22] M. Lam and G. Baranoski. A predictive light transport model for the human iris. volume 25, pages 359–368. Wiley Online Library, 2006.

[23] N.M. Law, C.D. Mackay, and J.E. Baldwin. Lucky imaging: high angular resolution imaging in the visible from the ground. *Arxiv preprint astro-ph/0507299*, 2005.

[24] A. Lefohn, B. Budge, P. Shirley, R. Caruso, and E. Reinhard. An ocularist's approach to human iris synthesis. *Computer Graphics and Applications, IEEE*, 23(6):70–75, 2003.

[25] S. Makthal and A. Ross. Synthesis of iris images using markov random fields. In *Proc. 13th European Signal Processing Conf.* Citeseer, 2005.

[26] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan. Light field photography with a hand-held plenoptic camera. *Computer Science Technical Report CSTR*, 2, 2005.

[27] University of Nebraska Medical Center. Unmc our mission, 2012.

[28] Work Health Organization. Vision 2020, 2000.

[29] World Health Organization. How vision works., April 2000.

[30] World Health Organization. Action plan for the prevention of avoidable blindness and visual impairment, 2009-2013., 2010.

[31] V.F. Pamplona, A. Mohan, M.M. Oliveira, and R. Raskar. Netra: interactive display for estimating refractive errors and focal range. *ACM Transactions on Graphics (TOG)*, 29(4):77, 2010.