mobileVision: A Face-mounted, Voice-activated, Non-mydriatic "Lucky" Ophthalmoscope

Adam Samaniego Vivek Boominathan Ashutosh Sabharwal

Ashok Veeraraghavan

ABSTRACT

We present mobileVision – a portable, robust, smartphone-based ophthalmoscopy prototype intended to reduce the barriers to ocular pathology screening in developing and underserved regions. In contrast to currently available retinal imaging solutions, mobileVision provides the portability of a handheld ophthalmoscope without sacrificing retinal field-of-view or resolution. Through tight integration with a smartphone and ergonomic design, we demonstrate novel features for such a small form factor, including: automatic compensation for patient refractive error, voice-activated multi-shot retinal image acquisition without pupil dilation (non-mydriatic), and touch-gesture based control of patient fixation and accommodation. We further demonstrate a computational lucky imaging and retinal stitching pipeline which not only increases overall retinal field-of-view, but also makes the system robust to patient saccades, blinks, device jitter, and imaging artifacts such as noise or unintended scattering from ocular surfaces. We estimate through mock eye tests that the mobileVision prototype is capable of imaging the retina with 23.5µm of retinal resolution for patients with between −6 D to +13 D refractive error, and we image over ±45° of retina during an *in vivo* trial.

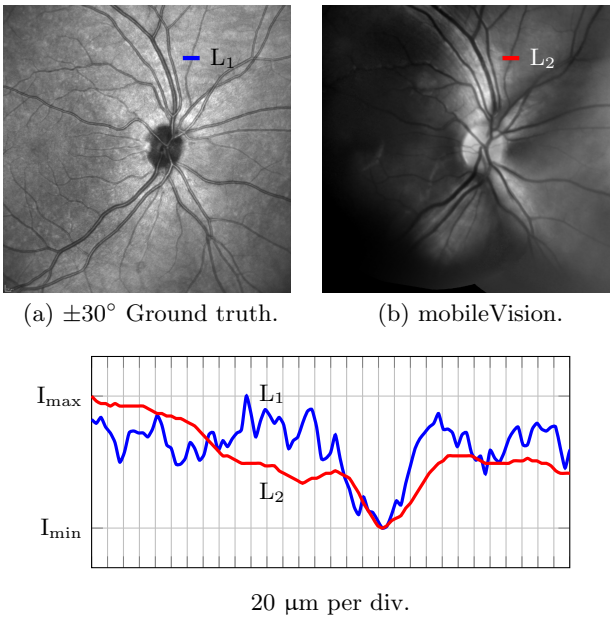
General Terms

Algorithms, Design

1. INTRODUCTION

Ocular health assessment is key to monitoring not only eye-related problems like glaucoma, cataracts, macular degeneration and refractive errors, but also many other chronic and systemic diseases such as diabetes, hypertension and neurological degeneration, all of which also impact vision. The eye is the only place where we can directly image internal tissue, like blood vessels, with a *non-invasive procedure*, making it particularly suitable for regular screening. However, while ocular imaging has advanced technically, ophthalmoscopy is still largely confined to clinical settings by design: bulky fundus cameras, which may be easy to operate, are difficult to transport and costly; handheld ophthalmoscopes, which are portable and low cost, require expert skill and dexterity for proper use. The realities of decentralized healthcare delivery in developing and underserved regions – poor infrastructure, low physician to patient ratios, the need for remote diagnosis – make screening using existing devices difficult or impractical.

Our goal is to design an ophthalmoscopy system which can be carried by a health care worker to remote screening sites, quickly and easily setup, reliably used to acquire retinal images, packed up, and carried away. When Internet access is available, the images can then be wirelessly transferred to physicians for remote review. We seek to combine the performance [14] and utility [9] of table-top, non-mydriatic fundus cameras with the portability of handheld ophthalmoscopes, without requiring an expert operator.



(a) ±30° Ground truth. (b) mobileVision.

Figure 1: (a) IR Confocal Scanning Laser Ophthalmoscope (Heidelberg Spectralis HRA+OCT). (b) Green channel of mobileVision panorama, cropped and registered. Line scans show pixel values scaled by maximum and minimum values along the line, demonstrating faithful reproduction of vasculature.

Wireless Health '14 Oct 29-31, 2014, Bethesda, MD, USA

1.1 Technical Challenges

Imaging a wide retinal field-of-view requires delivering and capturing light over a wide angle, which ordinarily leads to use of large optics. Furthermore, wide angle imaging without pupil dilation requires precise alignment between the system and the patient's eye, since the pupil constricts as soon as the retina is illuminated.

The need for precise alignment ordinarily sacrifices portability, however, as devices quickly become bulky and table mounted. In the case of mobile screening, the system must be lightweight and compact.

Finally, to not require an expert operator, the system must be easy to interact with and robust to operator error. Patient refractive error should be compensated for without operator interaction. Alignment should be intuitive and fault tolerant.

1.2 Contributions

We address the technical challenges of mobile retinal screening through the mobileVision prototype: a system comprising opto-mechanics, a smartphone with support electronics, and computational imaging algorithms. Specifically:

- We design, fabricate, and test a compact, face-mounted indirect ophthalmoscope with integrated wide-angle fixation and accommodation stimuli, which is battery powered; portable; voice or touch controlled; aligned to the patient's eye using translation knobs; and uses a smartphone to acquire images, compensate for patient refractive error, and guide the operator.
- We design and implement a robust blood-vessel matched filter response based retinal image registration technique which, by using scale invariant features, can operate reliably on small, minimally overlapping retinal fields in the presence of imaging artifacts, varying illumination, and image noise.
- We design and implement a multi-shot lucky imaging approach to retinal imaging which is capable of reducing retinal image noise while maintaining detail and automatically rejecting patient blinks, saccades, or other unwanted imaging artifacts.
- We design and implement a metric based, gradient-domain approach to retinal image stitching which is capable of avoiding imaging artifacts, without introducing color artifacts or visible seams.

The mobileVision prototype demonstrates for the first time an end-to-end computational ophthalmoscopy system capable of imaging the retina with resolution and field-of-view comparable to clinically relevant imaging modalities (fundus photography, scanning laser ophthalmoscopy), while remaining portable and user-friendly.

1.3 Related Work

1.3.1 Remote Ophthalmology

Groups such as Chen et al. [9] have demonstrated the utility of temporary eye clinics equipped with high-end tools (e.g. Canon $\pm 35^\circ$ CR6-45 non-mydriatic digital fundus camera) in remote areas for tele-ophthalmology. However, the purchase, delivery, and maintenance of such tools can be difficult and costly. These problems have inspired both us and others to create devices tailor-made for remote ophthalmology: at the time of writing, PeekVision is developing a smartphone based eye evaluation system which includes what appears to be a direct ophthalmoscope attachment for a smartphone; Forus Health's 3nethra is an eye pre-screening imaging tool, in the form of a multi-purpose table-top fundus camera; and EyeNetra's NetraG is a smartphone attachment for measuring patient refractive error. Such systems are tightly integrated, portable, and wirelessly connected to the Internet - all requirements for the effective and efficient delivery of remote ophthalmology.

1.3.2 Computational Ophthalmoscopes

In our prior work we demonstrated a table-top, Internet protocol camera turned ophthalmoscope, with minimal post-processing [16]. We now extend upon that work by making a smartphone the core of a complete computational imaging system, including algorithms. Lawson et al. [12] and Boggess [5] demonstrated a computational direct ophthalmoscope in which light is shone into the patient's sclera, rather than through the pupil. They use a computational pipeline to compensate for light loss and form images.

1.3.3 Retinal Stitching

A rich body of work exists for finding transformations from one retinal view to another, and combining the unified views. The work most similar to our own are those which base this process on finding scale invariant features [21] or blood-vessel crossover points [7]. We extend these concepts for stitching smaller fields by using scale invariant features from matched filter responses.

2. MOBILEVISION SYSTEM DESIGN

The mobileVision retinal imaging system has three parts. Section 2.1, *Opto-mechanics*, describes how light is delivered to and collected from the patient's retina at different field locations. Section 2.2, *Smartphone and Electronics*, describes how the smartphone and support electronics enable robust image acquisition while providing a simple, unified interface to the operator. Section 2.3, *Retinal Image Post-Processing*, describes how input images gathered from the smartphone are processed in order to reduce noise, increase image quality, and increase overall field-of-view.

Before detailing individual system components, we illustrate how the system works at a high-level by describing a prototypical retinal imaging session using mobileVision. Looking into the device shown in Figure 2, the patient's right eye will be imaged while the left eye views a fixation assembly. Viewing the fixation stimuli induces controllable rotation and accommodation of both eyes, allowing different retinal fields to be imaged. The operator uses knobs to translate the imaging optics and follow the patient's eye as it rotates to different positions. When the system is aligned, the operator says "snap,"(or presses the on-screen shutter button) to capture a series of consecutive retinal snapshots. The alignment and image capture process is then repeated for each of the available fixation positions. When finished imaging one eye, the entire apparatus is flipped over and the image acquisition process is repeated again to image the patient's other eye. Images are stored digitally on the phone when acquired, and are then post-processed and ready to be sent for physician review.

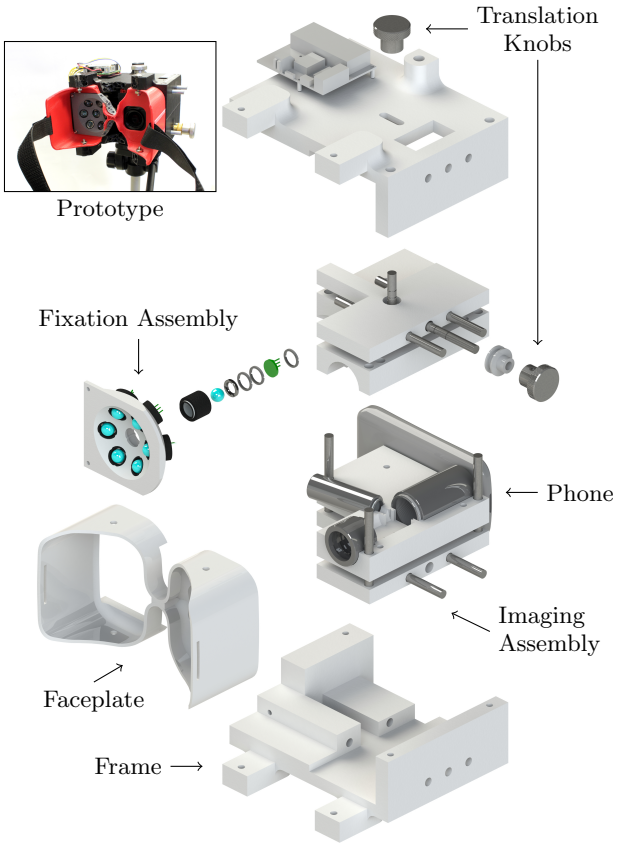


Figure 2: Mechanical system design. The imaging optics and phone translate jointly with respect to the fixation assembly. Knobs drive leadscrews to effect this translation. The faceplate (with straps) fixes the apparatus to the patient. The entire apparatus is flipped over so both eyes can be imaged.

2.1 Opto-mechanics

2.1.1 *Imaging the Retina*

The optical design is summarized in Figure 3. The imaging branch, an indirect ophthalmoscope,[1] is shown for the bottom eye. Light from a white LED is projected into a 1mm diameter homogenizing optical fiber. The fiber tip is then focused and linearly polarized before bouncing off an off-axis mirror just outside the cone visible to the camera. The reflection from this mirror is focused by the main imaging lens, and enters the patient's undilated pupil. The patient's eye defocuses the incoming light (because the fiber is conjugated nearly to patient's pupil), and in doing so $\approx \pm 25^\circ$ of retina is illuminated by polarized light.[2]

Light which reaches the retina undergoes diffuse reflection and loses its polarization state. However, light which undergoes specular reflection off of the patient's cornea or lens or other surfaces maintains its polarization state. This

[1]We repurposed components from a WelchAllyn Panoptic indirect ophthalmoscope to build this prototype.

[2]The described retinal illumination system was designed in accordance with ICNIRP broadband exposure limit guidelines [15].

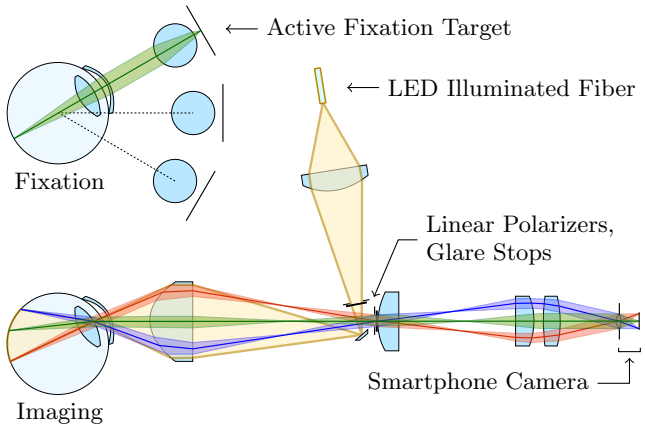


Figure 3: Optical system design. Infinite conjugate fixation (top eye) is used to guide the patient's gaze and accommodation for indirect retinal imaging (bottom eye). Patient aberration is compensated by the smartphone's built-in autofocus.

unwanted glare is removed by linearly polarizing all of the imaged light with a polarization axis orthogonal to that used in the illumination beam. The imaging optics create a nearly infinite conjugate image of the patient's retina, which is finally imaged by a smartphone.[3] We built this prototype to fit around a Samsung Nexus S smartphone.

2.1.2 *Guiding the Eyes*

While one of the patient's eyes is being imaged, the other eye views a fixation assembly, shown in Figures 2 and 3. The straps and faceplate of the device ensure this fixation assembly remains fixed relative to the patient's face for a given imaging session.

The fixation assembly is composed of 7 individual fixation tubes arranged about the eye's center of rotation. Each of the peripheral tubes is designed to induce a 30° rotation of the eye from the central position, exposing a new retinal region for imaging. Each tube comprises a single-element projector.[4] Only one of these tubes is illuminated at a time.

As the patient views each tube, his or her eyes naturally accommodate towards infinity. Although only one eye is presented with the fixation stimulus, both eyes rotate and accommodate in unison [4]. As a result, the patient's own accommodation response "pre-focuses" his or her eyes to be imaged, while simultaneously stabilizing eye orientation. Remaining defocus from refractive error is corrected by the smartphone's built-in autofocus.

2.1.3 *Eye-Camera Pupil Matching*

For each eye position, a delicate alignment between the patient's eye and the imaging optics needs to be established

[3]This design can accept any camera which can be pupil matched to the retinal imaging optics, which have an eye relief of about 10 mm. In practice, smartphone camera modules are typically only a few mm thick and have entrance pupils located within their enclosures, so nearly any smartphone will work optically.

[4]Each projects an infinite conjugate image of an LED backlit grid pattern out towards the patient's eye, using a 3/8" diameter plastic ball-lens.

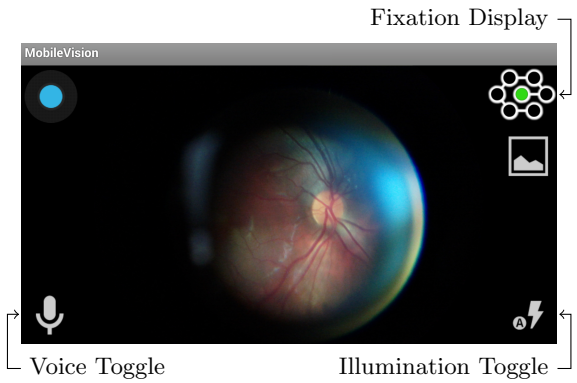
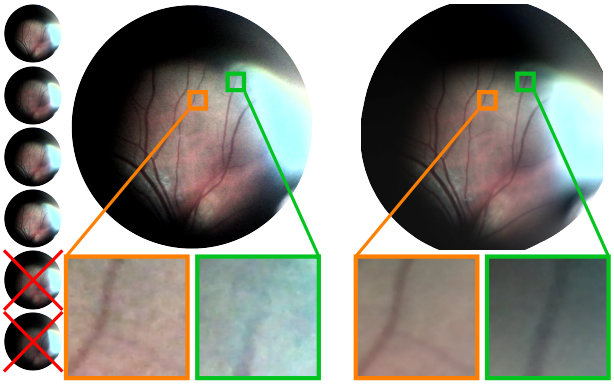


Figure 4: Android application. A live retinal video feed is shown in the center. Saying "snap" triggers 10 consecutive, full resolution snapshots. Autofocus is performed independently for each. Swiping across the screen cycles through fixation target positions.



(a) Smartphone inputs. (b) Fused view.

Figure 5: Multi-shot lucky fusing. Blurred or poorly exposed snapshots are rejected (red ×'s). Noise is reduced without sacrificing details (left insets). Artifacts are mitigated, exposing new details (right insets).

and maintained. This alignment is enabled by mechanically coupling the device to the patient's face with the faceplate and straps shown in Figure 2. The operator then rotates positioning knobs which smoothly translate the imaging assembly so that the retina can be found. Camera tripod mounting points are integrated into the frame (top and bottom) to further stabilize the system during prolonged use.

2.2 Smartphone and Electronics

2.2.1 *Android Application*

The bundled smartphone application is designed to make operation of the system as seamless and intuitive as possible. The user interface, shown in Figure 4, is functional yet approachable and visually similar to generic camera apps.

When the imaging optics are aligned with the patient's eye for a given fixation position and retina is visible on the smartphone, the operator triggers a sequence of 10 consecutive snapshots, either by touch-free "snap" voice command or by tapping an on-screen shutter button. Using repeated camera snapshot acquisitions rather than a continuous video acquisition allowed for image capture at the smartphone's full sensor resolution as well as for individual autofocus for each image acquired.

Voice control was implemented using CMU Sphinx's open-source speech recognition Android API [18]. A key feature of this API is *off-line speech recognition*: the operator need not be connected to the Internet at exam time because the command dictionary and recognition code are all stored locally on the smartphone when the app is installed. The command word "snap" was chosen for its simplicity and ease of recall. Voice control can be quickly enabled/disabled by tapping the on-screen voice toggle (useful to prevent image acquisition while talking).

The operator controls the current fixation position by performing a swiping gesture right or left across the screen. This increments or decrements, respectively, the fixation position, which cycles through all 7 positions. A display of the current position is shown at all times. This display ensures that all positions are used for imaging, and also helps orient the operator when performing pupil alignment.

The retinal illumination can be cycled on/off/auto by tapping the on-screen illumination toggle. In automatic mode the illumination is on only while focusing and exposing the smartphone camera.

2.2.2 *Support Electronics*

External circuitry necessary to drive the retinal illumination LED and fixation targets was designed and prototyped. A microprocessor (MSP430F5437) running a finite state machine and a Bluetooth 2.1 radio (RN-42) were included to enable wireless smartphone control. This circuitry is powered by 2×AAA batteries. Ni-MH rechargeable batteries can be used and recharged in-place (LTC-4060) through the included Micro-AB USB connector. The device can therefore charge itself with any smartphone Micro USB charger.

2.3 Retinal Image Post-Processing

The input images from the smartphone need to be enhanced and stitched to generate a retinal panorama. Section 2.3.1, *Multi-shot Lucky Retinal Fusing*, describes how by fusing many images of the same field, we obtain an enhanced retinal view which is better exposed, has less noise, and has less detrimental artifacts than any of the original images. Section 2.3.2, *Seamless Retinal Field Stitching*, describes how the enhanced views obtained from different fixation positions are then stitched together to increase overall field-of-view.

2.3.1 *Multi-shot Lucky Retinal Fusing*

An example 1474 × 1501 px, 24-bit input image is shown in Figure 5(a). Although arguably the best from the set of images for that view, it still suffers from degradation due to noise as well as artifacts due to scattered light off the patient's eye. Inspired by lucky imaging for astronomy [19], we reject poor quality images, register the remaining images to one another, and adaptively fuse the resulting stack to generate a superior quality image of the given retinal field.

Registration. We first attempt to align the images to one another using retinal blood vessels as guides. We apply blood vessel based matched filtering to the green channel of the RGB input images using the technique described by

Chaudhuri et al. [8]. The parameters of the matched filters were chosen based on the sizes of visible blood vessels in our acquired retinal images. This produces a matched filter response, some examples of which are shown in Figure 6(a). We perform local feature extraction on the matched filter responses using VLFeat's implementation [20] of the Scale Invariant Feature Transform (SIFT) [13]. We then use Random Sample Consensus (RANSAC) [10] to estimate the transformations from the image under consideration to a manually chosen reference image. A similarity transformation is assumed, which models lateral translation as well as scaling and rotation.

If the number of features is too low for a given image, this is indicative of a patient blink, saccade, or other image deterioration. Using a minimum SIFT feature count, poor quality images are automatically detected and discarded.

Gradient Fusing (Luminance). To obtain the enhanced retinal view, we first convert the stack of images to the YCbCr color space. In this representation, the Y channel contains the luminance of the image (which is independent of color), while the Cb and Cr channels contain chrominance (color) information. Let $Y_n(i,j)$ denote the luminance of a pixel at row i and column j, within image n of this registered stack. For each image n we calculate the pixel-wise horizontal and vertical gradients, g_x and g_y, as follows:

$$\begin{aligned} g_x(i,j)_n &= Y_n(i,j+1) - Y_n(i,j), \\ g_y(i,j)_n &= Y_n(i+1,j) - Y_n(i,j). \end{aligned} \quad (1)$$

To reduce image noise, we take stack-wise medians of the gradients at each pixel location as an initial estimate for the fused luminance gradients. Let this estimate be denoted $g^{\text{med}}_{[\cdot]}$. This is done independently for the x and y gradients, where $g_{[\cdot]}(i,j)_n$ denotes an x or y gradient for image n:

$$g^{\text{med}}_{[\cdot]}(i,j) = \underset{n}{\text{median}}(g_{[\cdot]}(i,j)_n). \quad (2)$$

Taking the medians in this way rejects outliers, however, having outliers in the final fused luminance may be desirable if one portion of the retina is well exposed in only a small subset of the captured images. To account for this, we also calculate an alternate estimate for the fused luminance gradients, g^*. This estimate relies upon a local image quality metric which uses gradient magnitudes. Let the gradient magnitude $\|g\|$ at pixel (i,j) be defined as:

$$\|g\| := \sqrt{g_x^2 + g_y^2}. \quad (3)$$

We then define a local image quality metric, $q_n(i,j)$, to be 11×11 patch-wise averages of $\|g\|$, using patches centered on (i,j). Let $n^*(i,j)$ correspond to the image with highest quality from the stack, at a given pixel:

$$n^*(i,j) = \arg\max_n(q_n(i,j)). \quad (4)$$

We find $g^*_{[\cdot]}(i,j)$ by selecting the gradients associated with image n^* from the stack, at pixel (i,j). However, to reduce susceptibility to noise in g^* we also average in the gradients from pixel (i,j) of image $\overline{n}$ if, $\overline{n} \in N$ where:

$$N = \{n^*(i+\delta_i, j+\delta_j) : \sqrt{\delta_i^2 + \delta_j^2} \leq \tau\}, \quad (5)$$

where τ is a fixed spatial radius. We term the method used to estimate g^* *trust-based averaging*. It has the effect that gradients are only averaged between images if the pixel (i,j) is in a region where other images from the stack are also marked as trustworthy (selected by n^*) – exactly as would occur in the presence of noise. If, however, n^* were to indicate that image n was the only trustworthy image over an area, then no averaging would be done, the gradients from n would be selected directly as the estimate g^*, and the region would not be deteriorated by inferior image information (say from images where that area was underexposed).

We combine the two estimates g^{med} and g^* by averaging to obtain $\hat{g}_{[\cdot]}$, the final luminance gradient estimates:

$$\begin{aligned} \hat{g}_x(i,j) &= \frac{g_x^{\text{med}}(i,j) + g_x^*(i,j)}{2}, \\ \hat{g}_y(i,j) &= \frac{g_y^{\text{med}}(i,j) + g_y^*(i,j)}{2}. \end{aligned} \quad (6)$$

We pass $\hat{g}_x$ and $\hat{g}_y$ into a 2D Poisson solver (Neumann boundary conditions) developed by Agrawal et al. [1, 2] to determine $\widehat{Y}(i,j)$, the resulting luminance image intensities.

Intensity Fusing (Chrominance). To bring color information into the fused result, we again perform trust-based averaging, but now on $\text{Cb}_n(i,j)$ and $\text{Cr}_n(i,j)$, the chrominance channel intensities of the registered image stack. Using trust-based averaging retains color information only from well exposed regions of the stack, without introducing color artifacts due to noise. We use the same quality metric and averaging scheme as was used for estimating g^*, but choose τ to be larger here. This yields fused chrominance values $\widehat{\text{Cb}}(i,j)$ and $\widehat{\text{Cr}}(i,j)$.

The final fused image, composed of intensities $\widehat{\text{Y}}, \widehat{\text{Cb}}, \widehat{\text{Cr}}$ at every pixel, is then converted back into RGB for display. The results of fusing are summarized in Figure 5(b).

2.3.2 Seamless Retinal Field Stitching

After fusing the different retinal views obtained, we stitch overlapping views together. We first register the views to one another, then adaptively stitch only the areas which are well exposed in each view. Here, however, we must take care to prevent image artifacts from being stitched, as these areas can have high gradient magnitude scores. We repeat the blood vessel based registration from Section 2.3.1 to obtain the transformations from one view to the next.

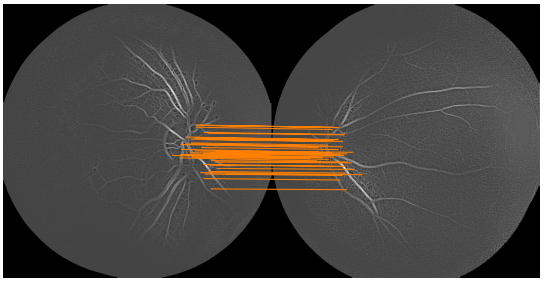
Gradient Stitching (Luminance). After registration, we have a montage of overlapping retinal areas. However, some areas are better resolved in some views than in others, and in some views, a specularity or other imaging artifact may be present. These artifacts should be avoided in the final reconstruction. For these reasons, we again use a gradient magnitude based metric to quantify local image quality, as was done in Section 2.3.1, but we now apply a weighting to penalize peripheral pixels from each view.

We first concern ourselves with just the luminance channel of the montage. We apply a weighting function $w_n(i,j)$ to $q_n(i,j)$, the gradient magnitude based metric. This penalizes a pixel's quality measure based on its distance from the center of the field-of-view which produced it. We use a 2D Gaussian to perform this penalization. This gives a new quality metric $q'_n(i,j)$:

$$q'_n(i,j) = w_n(i,j) q_n(i,j). \quad (7)$$

As before, let $n^*(i,j)$ be the image with highest quality at a given pixel:

$$n^*(i,j) = \arg\max_n(q'_n(i,j)). \quad (8)$$



(a) Matched filter response with SIFT feature matches shown.

(b) Registered, stitched and blended retinal fields.

Figure 6: Stitching of two fused retinal fields, with overlap below 22%.

We then apply trust-based averaging (see Equation (5)) using the new metric n^*, on the gradient magnitudes $g_x(i,j)_n$ and $g_y(i,j)_n$ to obtain gradient estimates $\hat{g}_x$ and $\hat{g}_y$. The estimate $\widehat{Y}(i,j)$ is found using $\hat{g}_x$ and $\hat{g}_y$ as was done in Section 2.3.1.

Intensity Stitching (Chrominance). We bring color information into the stitched result by averaging the Cb and Cr channels where possible. However, in this case, color must remain consistent from view to view to prevent the appearance of stitching artifacts.

If a pixel (i,j) is seen in more than one view, then the chrominance values assigned to it are the normalized weighted average of the chrominance values from the overlapping views at that pixel, where $\mathrm{Cb}_n(i,j)$ and $\mathrm{Cr}_n(i,j)$ are the chrominance values from view n. For a given pixel (i,j) present in multiple views M, the chrominance values are computed as follows:

$$\begin{aligned}\widehat{\mathrm{Cb}}(i,j) &= \frac{\sum_{n\in M}(w_n(i,j)\mathrm{Cb}_n(i,j))}{\sum_{n\in M}(w_n(i,j))},\\ \widehat{\mathrm{Cr}}(i,j) &= \frac{\sum_{n\in M}(w_n(i,j)\mathrm{Cr}_n(i,j))}{\sum_{n\in M}(w_n(i,j))},\end{aligned} \tag{9}$$

where w_n is the same 2D Gaussian weighting function previously used for gradient domain stitching of the luminance channel. The result is faithful color blending from view to view, shown for two views in Figure 6(b), and shown for the full panorama in Figure 8.

3. SYSTEM PERFORMANCE

Table 1: Prototype Summary

Size	19 cm × 14 cm × 20 cm
Weight	0.98 kg
Composite Field-of-view	$> \pm 45°$
Retinal Resolution	23.5 µm
Acceptable Patient Refractive Error	−6 D ↔ +13 D
Illumination Time On Single Charge	68 min

3.1 Data Collection

The retinal imagery shown in Figures 1, 4, 5, 6 and 8 was collected *in-vivo* from a single, healthy volunteer under an IRB approved research protocol, and is meant solely as system proof of concept. More comprehensive clinical trials showing the efficacy of the system for remote disease detection under varying operating conditions comprise our future work.

The apparatus was operated by a volunteer with no medical background and no prior practice capturing retinal images, under the supervision of the primary author. The smartphone and support electronics were charged prior to each imaging session, and ran solely off battery power while acquiring data. The entire system was mounted on a commodity camera tripod. The subject and the operator were seated for the imaging session. The subject was strapped into the faceplate and instructed to look at the fixation stimuli. The operator was shown the controls, and instructed to try to find the retina at each fixation position, taking images once the system was aligned. The collection of images was then transferred to a computer and processed using Matlab.[5]

3.2 Composite Field-of-view

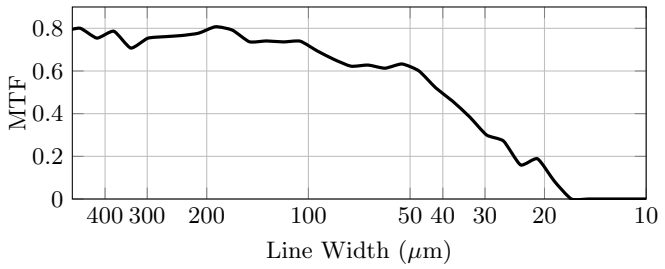
The composite field-of-view given in Table 1 was measured from the stitched output shown in Figure 8. We assume a spherical retina, then calculate the angle subtended by the optic disc. Using measurements from Jonas et al. [11] and Atchison et al. [3] we estimate that the optic disc subtends 9.8°. We use this to estimate visible field-of-view.

3.3 Retinal Resolution

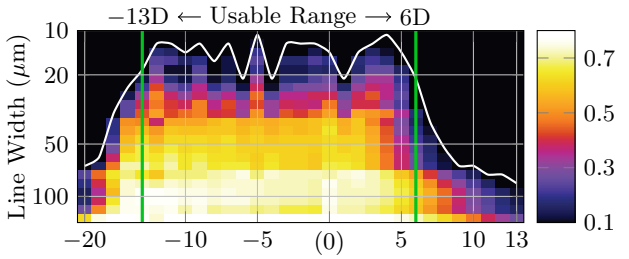
To estimate the achievable retinal resolution, a mock eye was created and imaged: A single lens ophthalmoscopy model-eye with 5mm pupil, 51.2 mm focal length, and planar retina was used. The original retina was removed and replaced by a reflective, negative USAF-1951 test plate (Edmund Optics, 36-408) placed atop a backlit white diffuser. Because the focal length of the model-eye, f_{model}, is different from the focal length of the human eye, f_{human}, we calculate the following paraxial magnification:

$$m = \frac{f_{model}}{f_{human}}. \tag{10}$$

[5]The transferring of files to a computer does not restrict the utility of the system, as cloud-based processing and wireless, encrypted file transfer are now commonplace.



(a) MTF of the system with no applied refractive error. Line widths of 23.5μm are resolvable with 16% contrast.



(b) MTF as a function of applied trial lenses. The heatmap shows contrast, determined using Equation 12. The white curve marks where retinal contrast falls below 10%. The green bars indicate the range of applied trial lens powers for which 23.5μm or smaller is still resolvable. Refractive error compensation was done for every image by the smartphone's built-in autofocus functionality.

Figure 7: Imaging performance on a magnification corrected model-eye with USAF-1951 planar retina.

And use it to map physical distances from the test plate, d_{model}, to the human retina, d_{human}:

$$\frac{d_{model}}{m} = d_{human}. \tag{11}$$

The focal length of the Gullstrand 1 Relaxed eye model, 17.185 mm [17], was used for f_{human}, meaning $m = 2.979$. This scaling factor was applied to calculate human retinal spatial frequencies from the spatial frequencies physically present on the target.

The Modulation Transfer Function (MTF) of the system was calculated by imaging this mock eye. We followed the technique used by Broxton et al. [6] to automatically calculate the MTF for a given image of the test target. Modulation or contrast is expressed as:

$$\frac{I_{max} - I_{min}}{I_{max} + I_{min}} = contrast, \tag{12}$$

where I_{max} corresponds to the 8-bit grayscale intensity of the average of the 3 maxima along a line perpendicular to and centered within the set of bars in a given Group/Element pair, and I_{min} corresponds to average of the 2 minima between those maxima. The resulting MTF is plotted in Figure 7(a). Spatial frequencies (line widths) beyond which the MTF drops below 10% we declare as not resolvable.

3.4 Acceptable Patient Refractive Error

To test how well the system is able to compensate for patient refractive error using the smartphone's built-in autofocus functionality, we repeated the test outlined in Section 3.3, but now with different trial lenses placed in front of the model-eye. These simulate different degrees of refractive error. For example, when the mock eye is augmented with a −5 D trial lens, this represents a far-sighted, or hyperopic patient, requiring +5 D corrective lenses. The result of this analysis for a variety of trial lenses is shown in Figure 7(b).

We find the range of acceptable refractive errors by finding for what trial lenses the MTF of the system falls below 10% at 23.5μm line width. We see that the system can compensate for refractive error anywhere from −6 D (myopic) to +13 D (hyperopic) on the model-eye.

4. CONCLUSIONS

In summary, we have designed, fabricated and demonstrated a portable yet robust, smartphone-based retinal imaging system. It enables the acquisition of multiple retinal field positions; takes advantage of the smartphone's autofocus to correct for patient refractive error; and utilizes computational imaging to reduce image noise, enhance image contrast, find retinal blood vessels, and seamlessly stitch together overlapping retinal fields all while rejecting unwanted imaging artifacts.

We believe this computational ophthalmoscopy system lays the foundation for mobile retinal screening systems which generate diagnostically meaningful imagery and are truly portable, robust, and easy-to-use: exactly the force-multiplier ophthalmologists need to remotely deliver healthcare in developing and underserved areas.

5. ACKNOWLEDGMENTS

Adam Samaniego and Ashok Veeraraghavan were supported by NSF-IIS grant 1116718 and a grant from the Norman Hackerman Advanced Research Program. Special thanks to: Jason Porter for the CSLO imagery, Sushain Cherivirala and Richard Latimer for Android coding, and the Rice Center for Research Support for mechanical design and fabrication.

6. REFERENCES

[1] A. Agrawal, R. Chellappa, and R. Raskar. An algebraic approach to surface reconstruction from gradient fields. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 1, pages 174–181. IEEE, 2005.

[2] A. Agrawal, R. Raskar, and R. Chellappa. What is the range of surface reconstructions from a gradient field? In *Computer Vision–ECCV 2006*, pages 578–591. Springer, 2006.

[3] D. A. Atchison, C. E. Jones, K. L. Schmid, N. Pritchard, J. M. Pope, W. E. Strugnell, and R. A. Riley. Eye shape in emmetropia and myopia. *Investigative Ophthalmology & Visual Science*, 45(10):3380–3386, 2004.

[4] R. Bhola. Binocular vision. *Department of Ophthalmology and Visual Sciences, University of Iowa*, 2006.

[5] J. R. Boggess. *Integrated computational system for portable retinal imaging*. PhD thesis, Massachusetts Institute of Technology, 2012.

[6] M. Broxton, L. Grosenick, S. Yang, N. Cohen, A. Andalman, K. Deisseroth, and M. Levoy. Wave

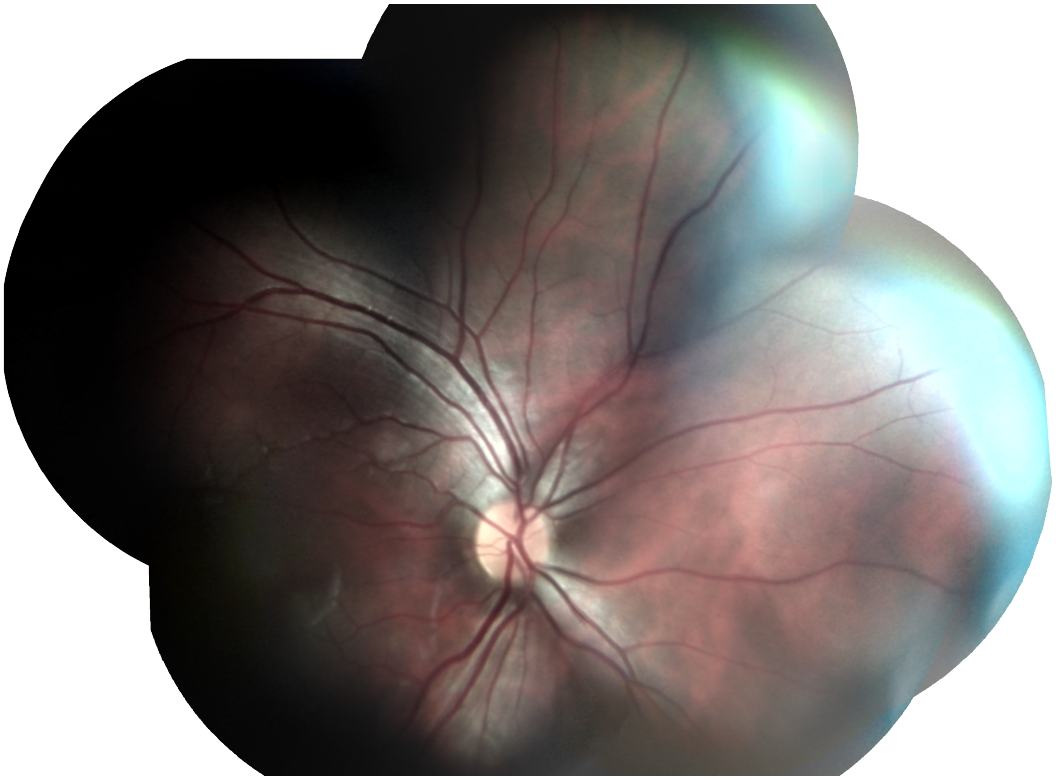


Figure 8: Acquired mobileVision retinal panorama. 70 input images, 7 views, showing over $\pm 45°$ of retina.

optics theory and 3-d deconvolution for the light field microscope. *Optics express*, 21(21):25418–25439, 2013.

[7] A. Can, C. V. Stewart, B. Roysam, and H. L. Tanenbaum. A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(3):347–364, 2002.

[8] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum. Detection of blood vessels in retinal images using two-dimensional matched filters. *IEEE transactions on medical imaging*, 8(3):263–269, 1989.

[9] L.-S. Chen, C.-Y. Tsai, T.-Y. Liu, T.-H. Tung, Y.-H. Chiu, C.-C. Chan, D.-M. Liou, and T. H.-H. Chen. Feasibility of tele-ophthalmology for screening for eye disease in remote communities. *Journal of telemedicine and telecare*, 10(6):337–341, 2004.

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

[11] J. B. Jonas, G. C. Gusek, and G. Naumann. Optic disc, cup and neuroretinal rim size, configuration and correlations in normal eyes. *Investigative ophthalmology & visual science*, 29(7):1151–1158, 1988.

[12] E. Lawson, J. Boggess, S. Khullar, A. Olwal, G. Wetzstein, and R. Raskar. Computational retinal imaging via binocular coupling and indirect illumination. In *ACM SIGGRAPH 2012 Posters*, page 51. ACM, 2012.

[13] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[14] P. Massin, A. Erginay, A. Ben Mehidi, E. Vicaut, G. Quentel, Z. Victor, M. Marre, P. Guillausseau, and A. Gaudric. Evaluation of a new non-mydriatic digital camera for detection of diabetic retinopathy. *Diabetic medicine*, 20(8):635–641, 2003.

[15] I. C. on Non-Ionizing Radiation Protection. Guidelines on limits of exposure to broad-band incoherent optical radiation (0.38 to 3μm). *Health Physics*, 73(3):539–554, 1997.

[16] A. Samaniego, J. Porter, A. Sabharwal, M. Twa, and A. Veeraraghavan. mobilevision: Towards a patient-operable, at-home, non-mydriatic retinal imaging system. *Journal of Vision*, 13(15):P28–P28, 2013.

[17] G. Smith and D. A. Atchison. *The Eye and Visual Optical Instruments*, volume 1. Cambridge University Press, 1997.

[18] C. M. U. Sphinx. Open source speech recognition.

[19] R. N. Tubbs. Lucky exposures: Diffraction limited astronomical imaging through the atmosphere. *arXiv preprint astro-ph/0311481*, 2003.

[20] A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms, 2008.

[21] L. Wei, L. Huang, L. Pan, and L. Yu. The retinal image mosaic based on invariant feature and hierarchial transformation models. In *Image and Signal Processing, 2009. CISP'09. 2nd International Congress on*, pages 1–5. IEEE, 2009.