

Synthesising Wider Field Images from Narrow-Field Retinal Video Acquired Using a Low-Cost Direct Ophthalmoscope (Arclight) Attached to a Smartphone

Keylor Daniel Chaves Viquez^a Ognjen Arandjelović^{†a} Andrew Blaikie^b In Ae Hwang^b

^aSchool of Computer Science ^bSchool of Medicine

University of St Andrews
Scotland, United Kingdom

[†]ognjen.arandjelovic@gmail.com

Abstract

Access to low cost retinal imaging devices in low and middle income countries is limited, compromising progress in preventing needless blindness. The Arclight is a recently developed low-cost solar powered direct ophthalmoscope which can be attached to the camera of a smartphone to acquire retinal images and video. However, the acquired data is inherently limited by the optics of direct ophthalmoscopy, resulting in a narrow field of view with associated corneal reflections, limiting its usefulness. In this work we describe the first fully automatic method utilizing videos acquired using the Arclight attached to a mobile phone camera to create wider view, higher quality still images comparable with images obtained using much more expensive and bulky dedicated traditional retinal cameras.

1. Introduction

It is estimated that some 285 million people in the world visually impaired, with the majority of cases potentially preventable or treatable if diagnosed promptly [9]. Unfortunately the burden of disease is greatest in low and middle income countries (East Asia, Latin America, and Africa) where access to diagnostic devices presents a major challenge [18, 7]. Observation of the retina by an eye care professional using an ophthalmoscope or through acquiring a video or photograph allows screening for key blinding diseases such as diabetic retinopathy, glaucoma and retinopathy of prematurity [9].

However, one of the key barriers to fundal screening in low and middle income countries (LMICs) emerges from the lack of access to affordable retinal imaging equipment [18]. The Arclight direct ophthalmoscope has been recently developed as a means of addressing this challenge [9]. The device, despite its simplified design and low cost, has been shown to be as effective as more expensive traditional de-



Figure 1. High quality images of the eye fundus acquired using a traditional, expensive, and bulky tabletop unit.

vices [24]. In addition, the Arclight can be attached to the camera of a mobile phone and acquire a digital recording of the fundus for the purposes of telemedicine. However, the captured images are intrinsically limited to a narrow field of view by the optics of direct ophthalmoscopy. Other limiting factors relating to the optics, sensor, and compression software of mobile phone cameras can further reduce the quality of the acquired data in comparison to more expensive dedicated traditional retinal cameras [9, 17, 10].

2. Context and practical challenges

Traditionally, retinal images (see Figure 1) are acquired using expensive and bulky tabletop units operated by trained technicians in a hospital setting [27]. Consequently such devices are impractical for use in low income areas where resources to buy and maintain traditional cameras are lim-

ited and robustness and portability is key for allowing the sharing and movement between inhospitable sites [9, 20].

However, the availability of low cost yet powerful smartphones in LMICs has led to the development of ‘m-health’ [26] whereby mobile phones can play several roles: contacting and reminding patients to take medicines and attend appointments, transferring money to pay for health care, storing guidelines and diagnostic algorithms, acquiring and archiving audio and visual information from patients, analysing data, utilizing GPS tracking, and transferring information wirelessly. These approaches [20] have quickly been adopted in low resource health care settings where the portability and connectivity of smartphones [19] has made them particularly valuable in telemedicine [20], making it possible to reach patients and communities that currently receive suboptimal care due to geographical and financial barriers [27].

2.1. The Arlight device

The Arlight is a highly portable (dimensions: 110mm \times 26mm \times 9mm, mass: 18g) low-cost (£5 if purchased in high volume¹) ophthalmoscope that has been developed in response to the need for affordable diagnostic eye care tools in LMICs [28]. Uniquely the device is illuminated by an LED allowing it to be powered by a slim lithium battery that is charged by an integrated solar panel reducing the need for expensive and hard to find consumables. Employing a small LED allows simplification of the design of the device reducing manufacturing costs and its overall size, see Figure 2 [9]. The novel features of the Arlight make it ideal for use in mobile clinics and remote low resource settings with limited access to power. Consequently over 10000 devices have been distributed, enabling thousands of health care workers worldwide to perform fundal examinations for the first time [9, 28, 24].

Studies have shown that the Arlight is just as accurate for screening for signs of diabetic retinopathy and glaucoma as more expensive traditional devices when used manually by an eye care professional [9]. However, one of the key inherent limitations of any direct ophthalmoscopy device is the narrow retinal field of view. Single still images are typically of limited utility but a sweeping video of a greater area of the fundus may offer greater clinical information, benefiting a telemedicine approach. Still, transferring video files is challenging in remote areas of a low income country where fast and reliable broadband internet may not be available. Synthesising a single wider field still image from multiple frames of a video fused together is a promising solution for overcoming many of the limitations of the optics of direct ophthalmoscopy as well as the data-heavy nature of video.

¹See <https://iapb.standardlist.org/get-arlight/>



Figure 2. Arlight: a miniature, low-cost ophthalmoscope.

3. Proposed method

In this section we describe a novel method for processing narrow field retinal video acquired using the Arlight ophthalmoscope while attached to an Apple iPhone 6s. By fusing the information contained in different frames of a video we synthesise a wider field image of higher quality, which is easier to interpret clinically and more compact to transmit and share. We start with an overview of the method and then follow up with detailed descriptions of each of the processing steps.

3.1. Overview

As we noted earlier, the quality and useful information content of different frames in a video varies greatly. Hence, in order to ensure robustness, our algorithm starts by automatically identifying the highest quality frame. Tracking, and thereafter incremental image stitching, commences both in forward and backward directions in time. Each new frame is processed in order to remove confounding appearance content, processed to extract salient information, registered with the growing synthetic image (initially the single frame which tracking begun), and stitched with it.

3.2. Detection of the region of interest

A crucial step in the pipeline outlined in the previous section concerns the detection of the region of interest (ROI) i.e. the part of the image which corresponds to the fundus and contains salient appearance content: the blood vessels and the optic disc. The parts of the frame which are outside

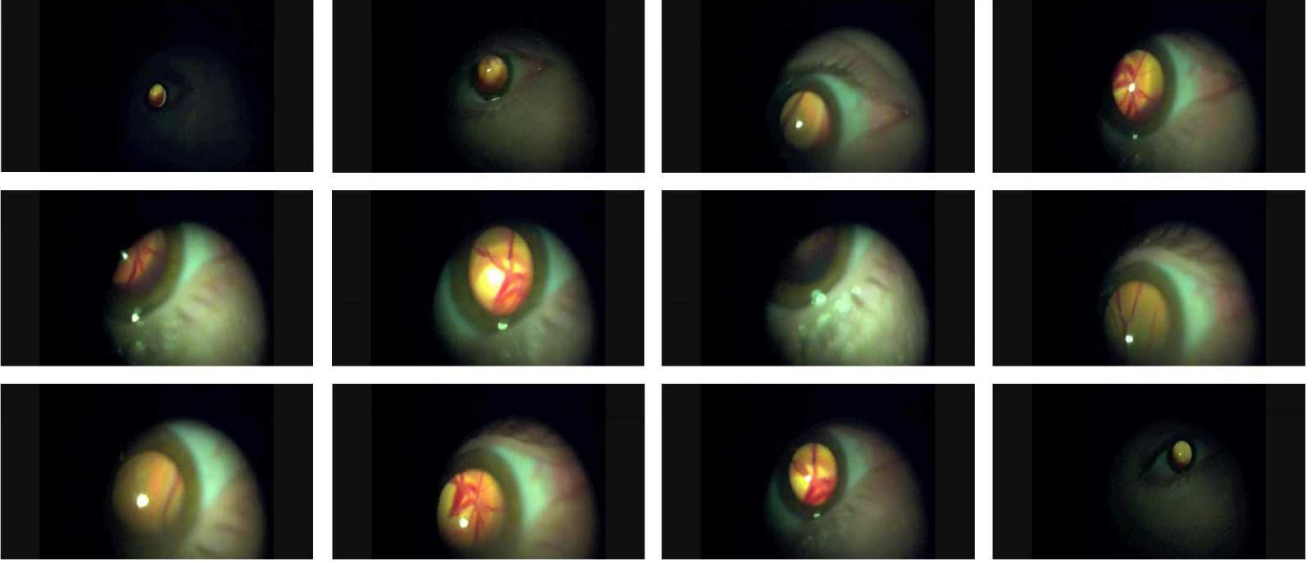


Figure 3. Typical frames extracted from a video acquired using the Arclight device. Variable scale and brightness (which is also non-uniform), frequent occlusions of the region of interest, narrow field of view, and blur are just some of the notable challenges posed both to physicians and automated methods.

of it are of no use for us and correspond to the sclera (the white, outer part of the eyeball), eyelids, etc.

As the frames in Figure 3 illustrate, the region of interest is elliptical in shape, and bright. However, colour in the image can not be used reliably as a discriminative feature for segmentation as significant variation can be observed across different individuals (e.g. there are differences between individuals of Caucasian and African descent). Variation in location, ellipse shape and orientation, as well as scale should also be noted as significant challenges.

Though the region of interest is elliptical in shape, observe that there is no sharp delineation, prohibiting the use of edge detection based approaches. Instead, we propose a two step method. In the first step, simple binary segmentation is performed using the well-known method first introduced by Otsu [25]. Here the frame is treated as a greyscale image, see Figure 4. Then the contour of the largest region used to find the best elliptical fit. We achieve this using a genetic algorithm based approach. Our methodology is inspired by the work of Conn and Arandjelović who used ellipse fitting as the first stage in the task of localizing and segmenting ancient coins in real-world images [12].

An ellipse embedded within a 2-dimensional Euclidean space is parameterized by five parameters. This can be readily seen from the implicit equation:

$$x^2 + k_{xy}xy + k_{yy}y^2 + k_x x + k_y y + k = 0. \quad (1)$$

The choice of the parameterization is crucial in making the most of our genetic search strategy. In particular, the parameterization should be such that when encoded as a ‘chromo-

some’ (in the context of a genetic algorithm), operations such as crossover, mutations, and others, are likely to effect an improvement in the fitness of a hypothesis. With this goal in mind, we parameterize an ellipse (hypothesis) using five points on its circumference and, following Conn and Arandjelović [12], use a short, non-binary chromosome comprising the coordinates of these points. As argued by Conn and Arandjelović, by enforcing the indivisibility of coordinate values we ensure that the constraint imposed by two circumference points is retained during evolutionary operations, thereby achieving a higher chance of greater generational fitness improvement.

The fitness of a specific hypothesis is evaluated in a straightforward fashion. From chromosomes formed by concatenating the coordinates of five points on an ellipse’s circumference – (x_a, y_a) , (x_b, y_b) , (x_c, y_c) , (x_d, y_d) , and (x_e, y_e) – the parameters in (1) can be obtained by solving a simple linear equation:

$$\begin{bmatrix} x_a^2 & x_a y_a & y_a^2 & x_a & y_a & 1 \\ x_b^2 & x_b y_b & y_b^2 & x_b & y_b & 1 \\ x_c^2 & x_c y_c & y_c^2 & x_c & y_c & 1 \\ x_d^2 & x_d y_d & y_d^2 & x_d & y_d & 1 \\ x_e^2 & x_e y_e & y_e^2 & x_e & y_e & 1 \end{bmatrix} \begin{bmatrix} 1 \\ k_{xy} \\ k_{yy} \\ k_x \\ k_y \\ k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \quad (2)$$

From this parameterization, the fitness of a hypothesis can be readily quantified by (i) uniformly sampling the circumference of the corresponding ellipse, and (ii) computing the number of samples which are located on edge pixels of the edge image. Equispaced samples along the circumfer-



Figure 4. Otsu's binary segmentation method based on automatic greyscale thresholding is used for the initial, rough localization of the region of interest.

ence of the corresponding ellipse are readily generated using the canonical form, i.e. using the major and minor radii (a and b) of the ellipse, its centre (x_0, y_0) , and the rotation angle θ relative to the coordinate system:

$$a = \frac{-\sqrt{2\psi\psi_1}}{k_{xy}^2 - 4k_{yy}} \quad \text{and} \quad b = \frac{-\sqrt{2\psi\psi_2}}{k_{xy}^2 - 4k_{yy}}, \quad (3)$$

$$x_0 = \frac{2k_{yy}k_x - k_{xy}k_y}{k_{xy}^2 - 4k_{yy}} \quad \text{and} \quad y_0 = \frac{2k_y - k_{xy}k_x}{k_{xy}^2 - 4k_{yy}}, \quad (4)$$

where:

$$\psi = k_{xy}^2 + k_{yy}k_x^2 - k_{xy}k_xk_y + (k_{xy}^2 - 4k_{yy})k, \quad (5)$$

$$\psi_1 = 1 + k_{yy} + \sqrt{(1 - k_{yy})^2 + k_{xy}^2}, \quad (6)$$

$$\psi_2 = 1 + k_{yy} - \sqrt{(1 - k_{yy})^2 + k_{xy}^2}. \quad (7)$$

Then, if there are n_s samples, and the samples are $\{(x_1, y_1), (x_2, y_2), \dots, (x_{n_s}, y_{n_s})\}$, the fitness becomes:

$$\phi = \frac{\sum_{i=1}^{n_s} E(x_i, y_i)}{n_s}, \quad (8)$$

where $E(x, y)$ is the value of the pixel (x, y) in the edge image \mathbf{E} .

3.3. Glare detection and removal

As mentioned earlier, one of the challenges with the input videos is the presence of significant glare. It is of critical importance that the areas of the fundus occluded by glare are detected so as not to confound the tracking process, as well as so that the relevant regions can be filled-in by exploiting information from other frames in which they are not occluded.

Glare areas are for practical purposes of unpredictable shape, often but not universally present, and their location varies. All of these characteristics make their detection difficult. Another challenge which emerges in our specific application concerns the appearance of the optic disk which is usually brighter than the rest of the fundus, and could be mistaken for glare. Yet, the appearance of the optic disk is

a crucial structure of interest to ophthalmologists, used to diagnose glaucoma and a number of other conditions [21].

Our approach to detecting glare is inspired by the work of Lange [22] on images of the uterine cervix. Broadly speaking, we detect glare areas as those regions of the image which exhibit high contrast or saturation, and are approximately white. Formally, the criterion is based on the following local measure $g(x, y)$:

$$g(x, y) = \min(R_{x,y}, G_{x,y}, B_{x,y}) - \frac{\min(R_{x,y}, G_{x,y}, B_{x,y})}{\max(R_{x,y}, G_{x,y}, B_{x,y})} \quad (9)$$

where $R_{x,y}$, $G_{x,y}$, and $B_{x,y}$ are respectively the red, green, and blue colour components of the pixel at the locus (x, y) . After computing $g(x, y)$ for all image loci the corresponding feature image is smoothed using the Alternating Sequential Filter (a sequence of morphological closing and opening operations), and the binary mask \mathbf{M} computed with $M_{x,y} = 1$ iff $G_{x,y}$ exceeds a predefined threshold, and $M_{x,y} = 0$ otherwise. Examples are shown in Figure 6.

3.4. Salient information extraction

Having removed confounding image content (glare, background, etc.) the processed frames are ready for the main part of our algorithms with concerns the tracking of salient structures present in the fundus, and the subsequent stitching. However, notwithstanding the significant processing described thus far, these tasks are still far from straightforward. For example, note that neither simple appearance nor any of the commonly used local features can be used for tracking (or, equivalently, registration). Raw appearance changes greatly due to global illumination changes, as well as local changes caused by the inherent optics of the eye. Local features – such as appearance based SIFT [23] and SURF [8], or edge based ones [2] – can also not be reliably detected as the appearance of the fundus does not contain corner-like loci.

Our approach is specifically targeted at the type of appearance content of interest in fundal images, namely blood vessels. These are line like features which form complex

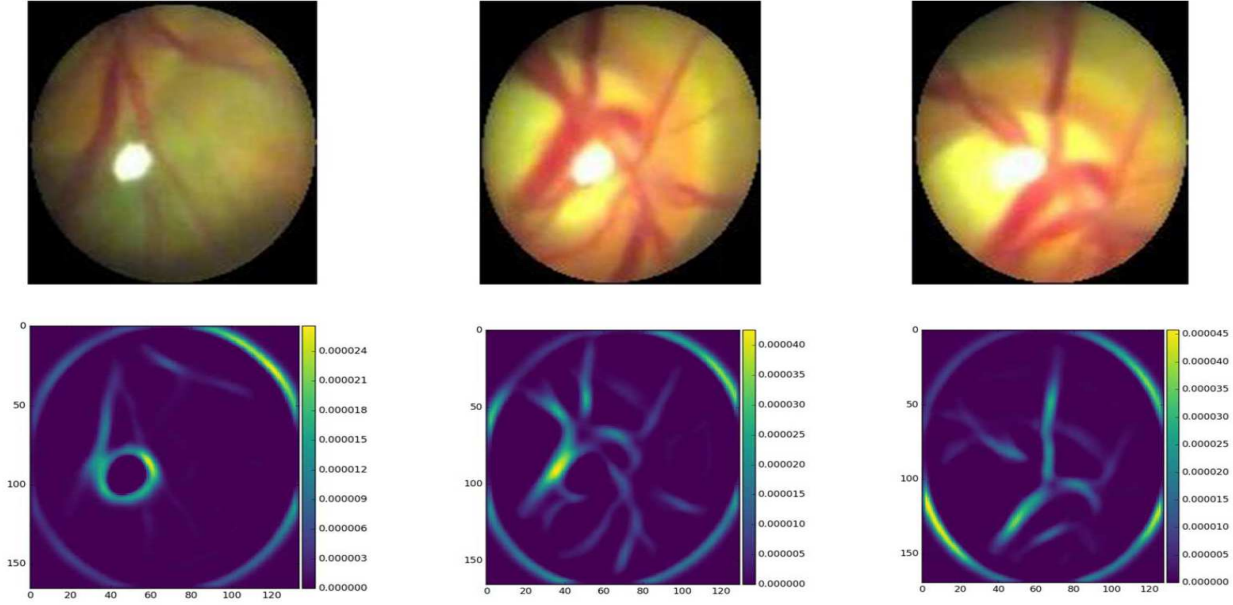


Figure 5. Examples of extracted vessel structures. Note that here only the raw output is shown: artefacts at the borders of the region of interest and the glare and removed thereafter.

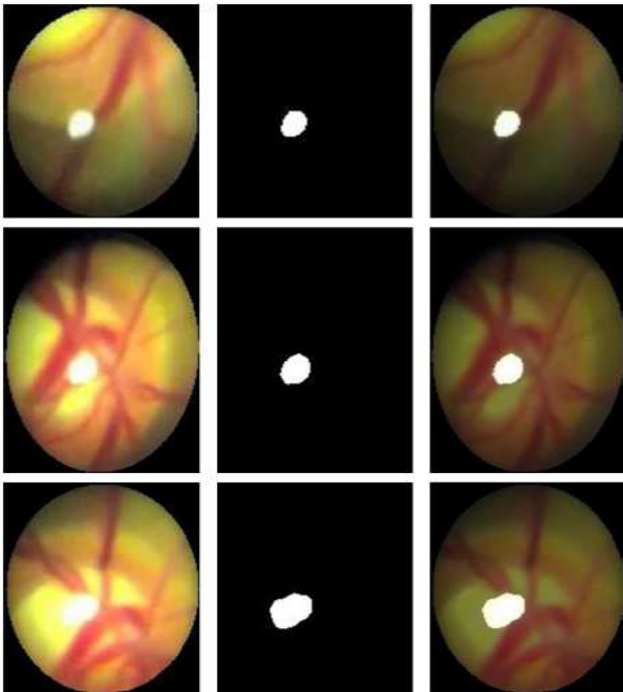


Figure 6. Glare detection result examples: original images (left), and the corresponding binary glare masks (centre) and images with glare removed (right).

and characteristic structures, and by virtue of this allow different frames to be mutually registered.

We extract characteristic vessel based features using the method introduced by Frangi *et al.* [13]. Their so-called

vesselness filter, first proposed for the use on 3D MRI data but later successfully employed in a number of different applications [14, 15, 16], extracts tubular structures from an image. For a 2D image consider the two eigenvalues λ_1 and λ_2 of the Hessian matrix computed at a certain image locus and at a particular scale. Without loss of generality let us also assume that $|\lambda_1| \leq |\lambda_2|$. The two key values used to quantify how tubular the local structure at this scale is are $\mathcal{R}_A = |\lambda_1|/|\lambda_2|$ and $\mathcal{S} = \sqrt{\lambda_1^2 + \lambda_2^2}$. The former of these measures the degree of local 'blobiness'. If the local appearance is blob-like, the Hessian is approximately isotropic and $|\lambda_1| \approx |\lambda_2|$ making \mathcal{R}_A close to 1. For a tubular structure \mathcal{R}_A should be small. On the other hand, \mathcal{S} ensures that there is sufficient local information content at all: in nearly uniform regions, both eigenvalues of the corresponding Hessian will have small values. For a particular scale of image analysis s , the two measures, \mathcal{R}_A and \mathcal{S} , are then unified into a single vesselness measure:

$$\mathcal{V}(s) = \begin{cases} 0 & \text{if } \lambda_2 > 0 \\ (1 - e^{-\frac{\mathcal{R}_A}{2\beta^2}}) \times (1 - e^{-\frac{\mathcal{S}}{2c^2}}) & \text{otherwise,} \end{cases} \quad (10)$$

where β and c are the parameters that control the sensitivity of the filter to \mathcal{R}_A and \mathcal{S} . Finally, if an image is analyzed across scales from s_{min} to s_{max} , the vesselness of a particular image locus can be computed as the maximal vesselness across the range:

$$\mathcal{V}_0 = \max_{s_{min} \leq s \leq s_{max}} \mathcal{V}(s) \quad (11)$$

We empirically set the parameter values to $\beta = 0.75$ and $c = 15$, and process frames at scales $s = 3, 4, 5$. Examples of typical results are shown in Figure 5.

3.4.1 Algorithm initialization

Recall that our algorithm starts the tracking (and thus the stitching process) from an automatically frame detected as the most reliable one i.e. one that contains sufficient salient information content to facilitate robust registration with the subsequent frames, and free of motion or out of focus blur. The vessel extraction method we just described is used to this end. In particular, following the computation of the vesselness image, the richness of salient information content is quantified by computing the entropy of this image. Input frames which do not contain many blood vessels, or which are corrupted by blur, will not produce many vessel detections, and thus result in low entropy vesselness images. Therefore, the frame with the highest entropy of the corresponding vesselness image is selected as the starting frame from which tracking commences, and as the initial synthetic image which is subsequently extended and improved through merging with newly processed frames.

3.5. Image registration

Recall that our algorithm incrementally expands the synthetic image of the fundus by expanding it by (stitching with) newly processed frames. The first step in the stitching process concerns the registration of a frame with the synthetic image. This is achieved by using the extracted vessel feature images. Specifically, we adopt the use of normalized cross-correlation as the hypothesis matching criterion [6], and exhaustively (though in a coarse to fine fashion, to speed up the process) explore the space of different scaling and translation parameters. The combination of the parameters which results in the highest normalized cross-correlation is accepted as the correct hypothesis but only if matching goodness exceeds that of the second best hypothesis (the second highest local normalized cross-correlation maximum) by a significant enough margin (herein we used a factor of 1.5). If this condition is not satisfied (which can happen, for example, when the newly acquired frame does not contain enough information to allow confident registration) the frame is discarded as unreliable.

3.6. Incremental stitching

Having registered a new frame with the growing synthetic image, the last step of our algorithm concerns the merging of the two images and with it the expansion of the synthetic result. There are several factors which make the process difficult. Firstly and as we noted before, there may be a significant illumination discrepancy between the two images. Moreover, a gradual darkening of the region of interest in the proximity of its boundary can be readily observed; note that this is a feature of the imaging process, rather than an artefact introduced by any of the steps of our algorithm. Lastly, the regions corresponding to glare areas,

void of useful information, should be taken into account.

Ideally, the result of the merging process should not only produce a seamless output image, spatially expanded (in general), but also leverage information content from both input images to increase the quality of information in the overlapping areas too. To achieve this while at the same addressing the challenges summarized above, we adopt an adaptive weighting strategy whereby the relative contributions of the two input images are adjusted depending on the particular locus under the consideration. To motivate our approach intuitively, the idea is to use preferentially data from an image in which a particular location is more central (relative to the region of interest) i.e. where the optical setup of mini ophthalmoscopies and the eye itself, produce better quality data (higher signal to noise ratio). We formalize this through the use of the distance transform [11, 1]. Specifically, after a linear illumination transform of brightness which normalizes global changes, for each pixel in an image we compute its approximate distance from the boundaries of the region of interest (the outer boundary and, if present, the boundary of the glare region). Then, when combining two corresponding pixels their relative contributions are computed by considering the inverse value of their distances from the boundaries in the original images, i.e. the output value of a RGB channel $c_{x,y}$ becomes:

$$c_{x,y} = \frac{w_f c_{x,y}^f + w_s c_{x,y}^s}{w_f w_s} \quad (12)$$

where $c_{x,y}^f$ is the value of the pixel in a new frame and $c_{x,y}^s$ in the synthetic image, and:

$$w_f = 1/d_{x,y}^f \quad w_s = 1/d_{x,y}^s \quad (13)$$

where $d_{x,y}^f$ and $d_{x,y}^s$ are the corresponding distance transform values (see Figure 7). The process can be therefore described as a form of adaptive linear fusion [4, 5, 3].

4. Evaluation

The data corpus we used in this work was collected from healthy volunteers. The acquisition process was entirely unaffected and unguided by us, guaranteeing realistic video input as would be collected in the field. Thus, we found that the sequences were highly variable in length, lasting from approximately 20 up to 90 seconds. The majority of the sequences start with a distant view of the eye (and hence little useful fundal information), and progressively close in onto the areas of actual interest such as blood vessels and the optic disc. Major motion and out of focus blur can be observed in this early phase. Once a section of the retina is clear the acquisition becomes more controlled, exhibiting purposeful motion in an effort to capture as much of the fundus as possible. Only a small section can be seen

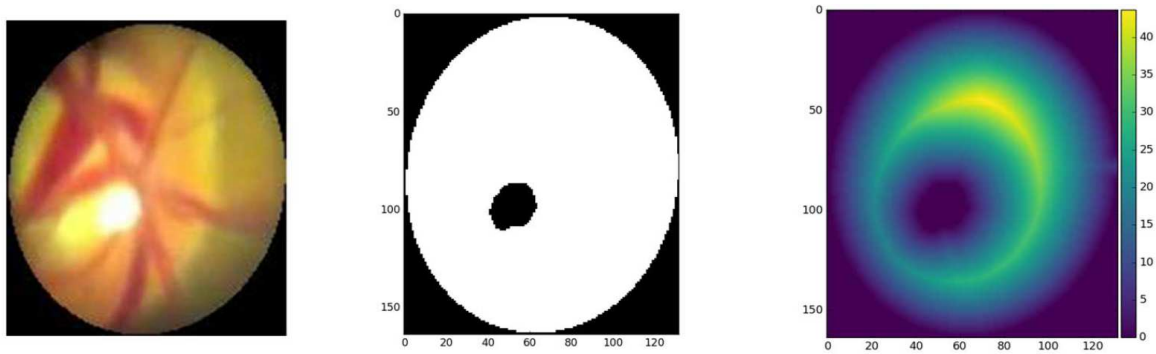


Figure 7. Original frame (left), region of interest (after the removal of the background and the glare), and the corresponding distance transform used for adaptive pair-wise image merging.

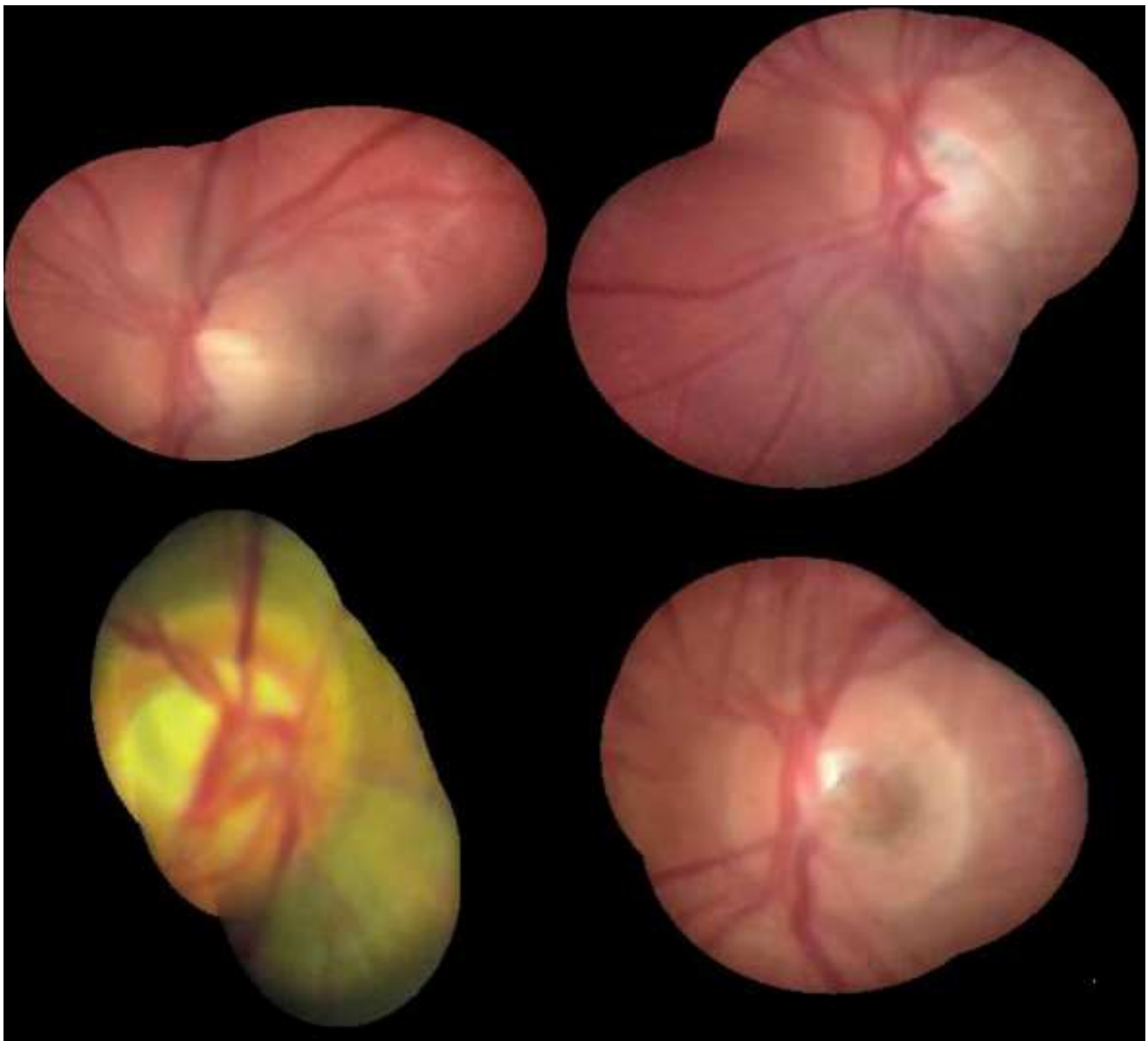


Figure 8. Examples of synthetically generated images.

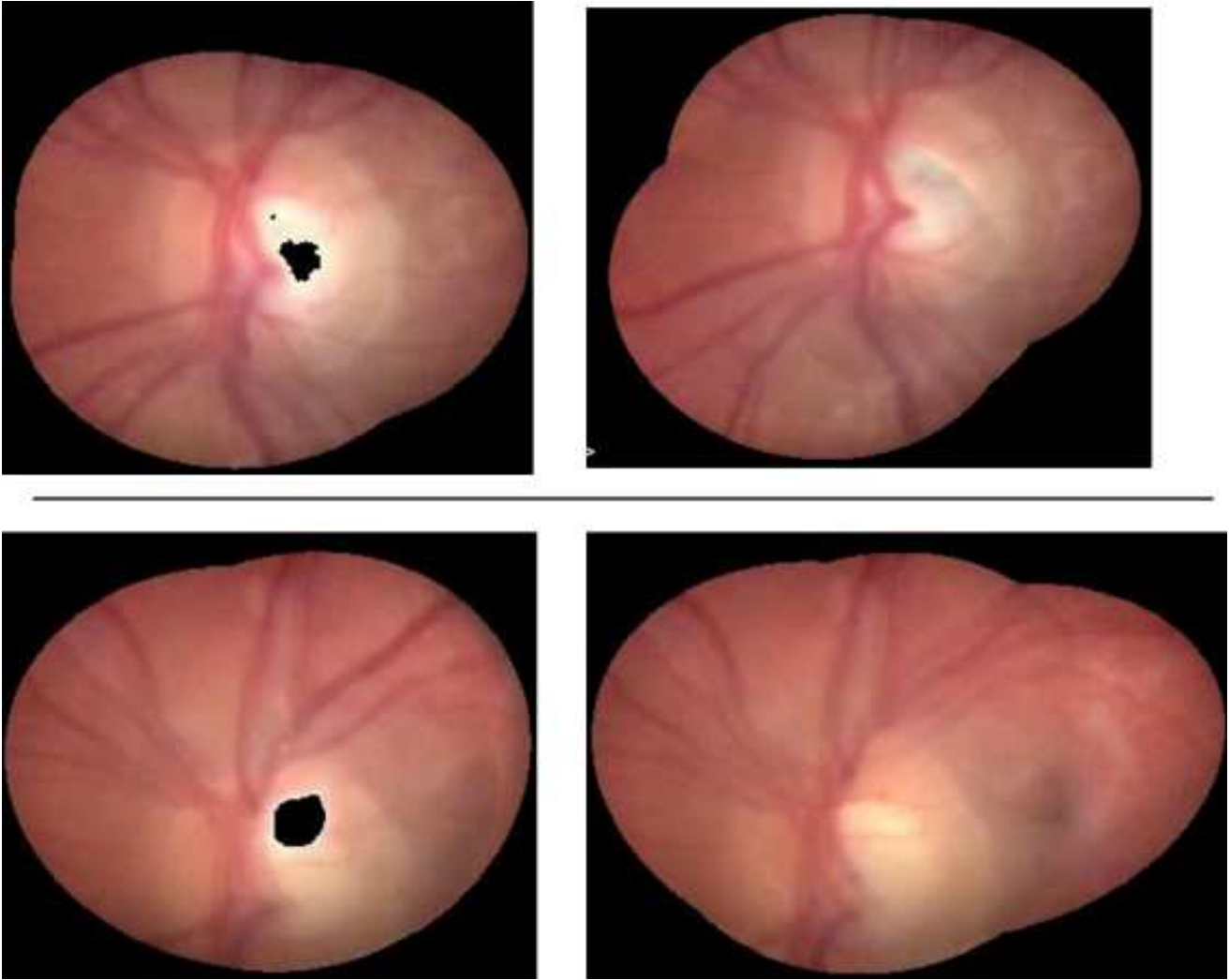


Figure 9. Examples of image fusion with the specific emphasis on the handling of glare.

at any given moment due to the narrow field of view of direct ophthalmoscopy devices. Throughout this process blur continues to pose problems, with the retina easily going out of focus, drastic changes in the direction of motion occurring, and the imaging stopping unexpectedly. Considering the nature of our data set, it was not possible to obtain objective ground truth high quality images using a traditional ophthalmoscope. Hence, at this stage we limited ourselves to a qualitative assessment of the results. Representative examples of synthetic images and the effects of specific challenges are shown in Figures 8 and 9.

5. Summary and conclusions

The Arclight is a recently developed low-cost direct ophthalmoscope that offers great potential for the prevention of blindness amongst those living in low and middle income countries. However, currently this potential is limited by the

inherent optical limitations of direct ophthalmoscopy, most notably the narrow field of view. In this paper we described the first method that uses this low quality data to create a synthetic, higher quality, wider view image comparable to one acquired using expensive and bulky traditional retinal cameras.

References

- [1] O. Arandjelović. Colour invariants under a non-linear photometric camera model and their application to face recognition from video. *Pattern Recognition*, 45(7):2499–2509, 2012. 6
- [2] O. Arandjelović. Object matching using boundary descriptors. *In Proc. British Machine Vision Conference*, 2012. DOI: 10.5244/C.26.85. 4
- [3] O. Arandjelović. Weighted linear fusion of multimodal data – a reasonable baseline? *In Proc. ACM Conference on Multimedia*, pages 851–857, 2016. 6

- [4] O. Arandjelović and R. Cipolla. A new look at filtering techniques for illumination invariance in automatic face recognition. *In Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, pages 449–454, 2006. 6
- [5] O. Arandjelović, R. I. Hammoud, and R. Cipolla. Thermal and reflectance based personal identification methodology in challenging variable illuminations. *Pattern Recognition*, 43(5):1801–1813, 2010. 6
- [6] O. Arandjelović, D. Pham, and S. Venkatesh. Efficient and accurate set-based registration of time-separated aerial images. *Pattern Recognition*, 48(11):3466–3476, 2015. 6
- [7] A. Bastawrous and B. D. Hennig. The global inverse care law: a distorted map of blindness. *British Journal of Ophthalmology*, 96(10):1357–1358, 2012. 1
- [8] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool. SURF: Speeded up robust features. *Computer Vision and Image Understanding*, 110(3):346–359, 2008. 4
- [9] A. Blaikie, J. Sandford-Smith, S. Y. Tuteja, C. D. Williams, and C. O’Callaghan. Arclight: a pocket ophthalmoscope for the 21st century. *The British Medical Journal*, 355:i6637, 2016. 1, 2
- [10] N. M. Bolster, M. E. Giardini, and A. Bastawrous. The diabetic retinopathy screening workflow. *Journal of Diabetes Science and Technology*, 10(2):318–324, 2016. 1
- [11] G. Borgefors. Distance transforms in arbitrary dimensions. *Computer Vision, Graphics and Image Processing*, pages 321–345, 1984. 6
- [12] B. Conn and O. Arandjelović. Towards computer vision based ancient coin recognition in the wild – automatic reliable image preprocessing and normalization. *In Proc. IEEE International Joint Conference on Neural Networks*, pages 1457–1464, 2017. 3
- [13] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever. Multiscale vessel enhancement filtering. *Medical Image Computing and Computer-Assisted Intervention*, pages 130–137, 1998. 5
- [14] R. S. Ghiass, O. Arandjelović, A. Bendada, and X. Maldague. Illumination-invariant face recognition from a single image across extreme pose using a dual dimension AAM ensemble in the thermal infrared spectrum. *In Proc. IEEE International Joint Conference on Neural Networks*, pages 2781–2790, 2013. 5
- [15] R. S. Ghiass, O. Arandjelović, A. Bendada, and X. Maldague. Vesselness features and the inverse compositional AAM for robust face recognition using thermal IR. *In Proc. AAAI Conference on Artificial Intelligence*, 1(0):357–364, 2013. 5
- [16] R. S. Ghiass, O. Arandjelović, A. Bendada, and X. Maldague. A unified framework for thermal IR-based face recognition. *In Proc. International Conference on Neural Information Processing*, II:335–343, 2014. 5
- [17] M. E. Giardini, I. A. T. Livingstone, S. Jordan, N. M. Bolster, T. Peto, M. Burton, and A. Bastawrous. A smartphone based ophthalmoscope. *In Proc. International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2177–2180, 2014. 1
- [18] C. Gilbert, R. Wormald, A. Fielder, A. Deorari, L. C. Zepeda-Romero, G. Quinn, A. Vinekar, A. Zin, and B. Darlow. Potential for a paradigm change in the detection of retinopathy of prematurity requiring treatment. *Archives of Disease in Childhood – Fetal and Neonatal Edition*, 101(1):6–9, 2015. 1
- [19] C. Hull. Digital imaging and screening for diabetic retinopathy. *Optometry Today*, pages 28–35, 2008. 2
- [20] Y. Kandasamy, R. Smith, I. Wright, and L. Hartley. Use of digital retinal imaging in screening for retinopathy of prematurity. *Journal of Paediatrics and Child Health*, 49(1), 2013. 2
- [21] D. Kiage, I. N. Kherani, S. Gichuhi, K. F. Damji, and M. Nyenze. The Muranga teleophthalmology study: comparison of virtual (teleglaucoma) with in-person clinical assessment to diagnose glaucoma. *Middle East African Journal of Ophthalmology*, 20(2):150–157, 2013. 4
- [22] H. Lange. Automatic glare removal in reflectance imagery of the uterine cervix. *In Proc. SPIE*, 5747:2183–2192, 2005. 4
- [23] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2003. 4
- [24] J. Lowe, C. R. Cleland, E. Mgaya, G. Furahini, C. E. Gilbert, M. J. Burton, and H. Philipppin. The Arclight ophthalmoscope: a reliable low-cost alternative to the standard direct ophthalmoscope. *Journal of Ophthalmology*, 2015, 2015. 1, 2
- [25] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979. 3
- [26] C. Z. Qiang, M. Yamamichi, V. Hausman, D. Altman, and IS Unit. Mobile applications for the health sector. *Washington: World Bank*, 2011. 2
- [27] A. Russo, W. Mapham, R. Turano, C. Costagliola, F. Morescalchi, N. Scaroni, and F. Semeraro. Comparison of smartphone ophthalmoscopy with slit-lamp biomicroscopy for grading vertical cup-to-disc ratio. *Journal of Glaucoma*, 25(9):e777–781, 2016. 1, 2
- [28] S. Tuteja and T. Young-Zvandasara. The Arclight: A pocket ophthalmoscope to revitalise undergraduate teaching? *Eye News UK*, 23(4), 2017. 2