



Unsupervised identification of cone photoreceptors in non-confocal adaptive optics scanning light ophthalmoscope images

CHRISTOS BERGELES,^{1,*} ADAM M. DUBIS,² BENJAMIN DAVIDSON,¹ MELISSA KASILIAN,² ANGELOS KALITZEOS,² JOSEPH CARROLL,³ ALFREDO DUBRA,⁴ MICHEL MICHAELIDES,² AND SEBASTIEN OURSELIN¹

¹Translational Imaging Group, Centre for Medical Image Computing, UCL, Gower Street, WC1E 6BT, UK

²NIHR Biomedical Research Centre at Moorfields Eye Hospital and Institute of Ophthalmology, UCL, 11-43 Bath St, London EC1V 9EL, UK

³Medical College of Wisconsin, 8701 W Watertown Plank Rd, Milwaukee, WI 53226, USA

⁴Stanford University, 450 Serra Mall, Stanford, CA 94305, USA

*c.bergeles@ucl.ac.uk

Abstract: Precise measurements of photoreceptor numerosity and spatial arrangement are promising biomarkers for the early detection of retinal pathologies and may be valuable in the evaluation of retinal therapies. Adaptive optics scanning light ophthalmoscopy (AOSLO) is a method of imaging that corrects for aberrations of the eye to acquire high-resolution images that reveal the photoreceptor mosaic. These images are typically graded manually by experienced observers, obviating the robust, large-scale use of the technology. This paper addresses unsupervised automated detection of cones in non-confocal, split-detection AOSLO images. Our algorithm leverages the appearance of split-detection images to create a cone model that is used for classification. Results show that it compares favorably to the state-of-the-art, both for images of healthy retinas and for images from patients affected by Stargardt disease. The algorithm presented also compares well to manual annotation while excelling in speed.

© 2017 Optical Society of America

OCIS codes: (100.0100) Image processing; (170.4470) Ophthalmology; (110.1080) Active or adaptive optics.

References and links

1. A. Roorda and J. L. Duncan, "Adaptive optics ophthalmoscopy," *Annu. Rev. Vis. Sci.* **1**, 19–50 (2015).
2. T. Y. Chui, H. Song, and S. A. Burns, "Adaptive-optics imaging of human cone photoreceptor distribution," *J. Opt. Soc. Am. A* **25**, 3021–3029 (2008).
3. M. Lombardo, S. Serrao, N. Devaney, M. Parravano, and G. Lombardo, "Adaptive optics technology for high-resolution retinal imaging," *Sensors* **13**, 334–366 (2012).
4. T. Y. P. Chui, H. Song, and S. A. Burns, "Individual variations in human cone photoreceptor packing density: variations with refractive error," *Invest. Ophth. Vis. Sci.* **49**, 4679–4687 (2008).
5. R. F. Cooper, M. Lombardo, J. Carroll, K. R. Sloan, and G. Lombardo, "Methods for investigating the local spatial anisotropy and the preferred orientation of cones in adaptive optics retinal images," *Visual Neurosci.* **33**, E005 (2016).
6. R. F. Cooper, M. A. Wilk, S. Tarima, and J. Carroll, "Evaluating descriptive metrics of the human cone mosaic," *Invest. Ophth. Vis. Sci.* **57**, 2992–3001 (2016).
7. K. Y. Li and A. Roorda, "Automated identification of cone photoreceptors in adaptive optics retinal images," *J. Opt. Soc. Am. A* **24**, 1358–1363 (2007).
8. L. Mariotti and N. Devaney, "Performance analysis of cone detection algorithms," *J. Opt. Soc. Am. A* **32**, 497–506 (2015).
9. D. M. Bukowska, A. L. Chew, E. Huynh, I. Kashani, S. L. Wan, P. M. Wan, and F. K. Chen, "Semi-automated identification of cones in the human retina using circle Hough transform," *Biomed. Opt. Express* **6**, 4676 (2015).
10. S. J. Chiu, Y. Lokhnygina, A. M. Dubis, A. Dubra, J. Carroll, J. A. Izatt, and S. Farsiu, "Automatic cone photoreceptor segmentation using graph theory and dynamic programming," *Biomed. Opt. Express* **4**, 924–937 (2013).
11. L. Mariotti and N. Devaney, "Cone detection and blood vessel segmentation on AO retinal images," in *Proceedings of Irish Machine Vision and Image Processing*, 126–128 (2015).
12. D. Scoles, Y. N. Sulai, C. S. Langlo, G. Fishman, C. Curcio, J. Carroll, and A. Dubra, "In vivo imaging of human cone photoreceptor inner segments," *Invest. Ophth. Vis. Sci.* **55**, 4244–4251 (2014).

13. E. A. Rossi, C. E. Granger, R. Sharma, Q. Yang, K. Saito, C. Schwarz, S. Walters, K. Nozato, J. Zhang, T. Kawakami *et al.*, “Imaging individual neurons in the retinal ganglion cell layer of the living eye,” *Proc. Natl. Acad. Sci. U.S.A.* **2016**:13445 (2017).
14. A. Lucchi, K. Smith, R. Achanta, V. Lepetit, and P. Fua, “A fully automated approach to segmentation of irregularly shaped cellular structures in EM images,” in *Proceedings Med. Image Comput. Comput. Assist. Interv.* **463–471** (2010).
15. S. Dimopoulos, C. E. Mayer, F. Rudolf, and J. Stelling, “Accurate cell segmentation in microscopy images using membrane patterns,” *Bioinformatics* **30**, 2644–2651 (2014).
16. J. Liu, A. Dubra, and J. Tam, “Computer-aided detection of human cone photoreceptor inner segments using multi-scale circular voting,” *Proc. SPIE Medical Imaging* **9785**, 97851 (2016).
17. J. Liu, A. Dubra, and J. Tam, “A fully automatic framework for cell segmentation on non-confocal adaptive optics images,” *Proc. SPIE Medical Imaging* **9785**, 97852J (2016).
18. D. Cunefare, R. F. Cooper, B. Higgins, D. F. Katz, A. Dubra, J. Carroll, and S. Farsiu, “Automatic detection of cone photoreceptors in split detector adaptive optics scanning light ophthalmoscope images,” *Biomed. Opt. Express* **7**, 2036–2050 (2016).
19. A. Dubra, Y. Sulai, J. L. Norris, R. F. Cooper, A. M. Dubis, D. R. Williams, and J. Carroll, “Noninvasive imaging of the human rod photoreceptor mosaic using a confocal adaptive optics scanning ophthalmoscope,” *Biomed. Opt. Express* **2**, 1864–1876 (2011).
20. R. F. Cooper, C. S. Langlo, A. Dubra, and J. Carroll, “Automatic detection of modal spacing (yellott’s ring) in adaptive optics scanning light ophthalmoscope images,” *Ophthalmic Physiol. Opt.* **33**, 540–549 (2013).
21. C. Tomasi and R. Manduchi, “Bilateral filtering for gray and color images,” in *Proceedings of IEEE International Conference on Computer Vision* pp. 839–846 (1998).
22. J. B. Zimmerman, S. M. Pizer, E. V. Staab, J. R. Perry, W. McCartney, and B. C. Brenton, “Evaluation of the effectiveness of adaptive histogram equalization for contrast enhancement,” *IEEE Trans. Med. Imaging* **7**, 304–312 (1988).
23. N. Otsu, “A threshold-selection method from gray-level histograms,” *IEEE Trans. Syst. Man. Cybern.* **9**, 62–66 (1979).
24. C. Arteta, V. Lempitsky, J. A. Noble, and A. Zisserman, “Learning to detect cells using non-overlapping extremal regions,” in *Proceedings Med. Image Comput. Comput. Assist. Interv.* pp. 348–356 (2012).

1. Introduction

Due to the slow progression of many eye diseases and their debilitating neurophysiological manifestations there is a dire need for early diagnosis, which provides greater likelihood for vision saving intervention. Ideally, cellular-level imaging would provide the earliest diagnosis of disease, and this technology would become mainstay in every practitioner’s office [1]. Precise, early detection would also be used to monitor efficacy of therapeutics such as gene replacement or stem cells [1]. To realize this level of monitoring, methods for repeatable, automated image analysis are required.

One major obstacle to achieving cellular resolution in the living human eye are monochromatic aberrations caused by the alignment of cells in the cornea and lens, as well as movement/disruption of tear film and ocular aqueous and vitreous humor. Adaptive Optics (AO) is a technique by which, through a wavefront sensor and actuated mirror, these imperfections are measured and then dynamically compensated, allowing for individual photoreceptor visualization [1–4] [see Fig. 1(a)]. AO can be embedded within any conventional imaging system, such as a fundoscope, scanning light ophthalmoscope (SLO), or optical coherence tomograph [1]. SLO is commonly utilized due to its ability to section through the layers of the retina and provide improved en-face image contrast and high lateral resolution.

Current-generation research AOSLOs are custom-built, with variable “clinical” interfaces. One trait they have in common is the ability to produce large amounts of images of the photoreceptor mosaic. The current largest challenge is the analysis of such images. Manual processing and classification for photoreceptor population studies is tedious, error prone, and time consuming. Thus, automated ways for photoreceptor counting are sought [1]. These algorithms, starting from cell identification, may then extract descriptive metrics such as density, preferred orientation, etc. [5, 6].

First-generation reflectance AOSLOs utilized confocal imaging to improve image contrast.

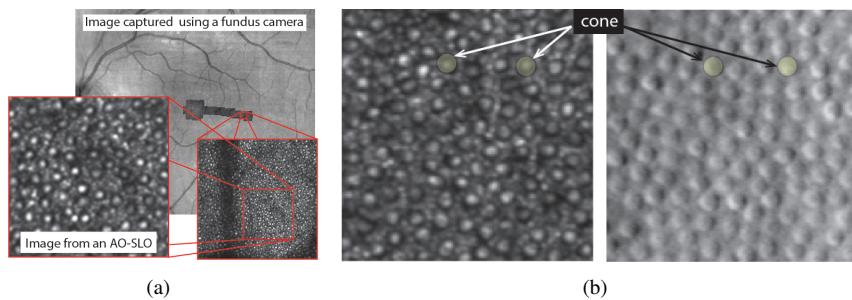


Fig. 1. (a) Illustration of the level of detail provided by AOSLO imaging. (b) Comparison of the confocal imaging approach to split-detection: (left) confocal imaging of cones, and (right) split-detection imaging of cones. The two images are matching and acquired in perfect registration. The annotated cones exemplify their different appearance on the two modalities.

This image resulted in photoreceptors appearing as a mosaic of Gaussian-like point sources [see Fig. 1(b)(left)]. Several algorithms leveraged the bright appearance of the confocal images generated by AOSLO [7, 8], including Circular Hough Transform [9], graph-cut with dynamic-programming-based methods [10], and pure intensity-based methods [11].

More recently, groups began to investigate non-confocal imaging [12, 13]. The work in [12] used a non-confocal split-detection technique to image the photoreceptor's inner segments, and improved identification of photoreceptors in retinal dystrophies [see Fig. 1(b)(right)]. Methods specific to cell detection in confocal AO images are not applicable to the new type of images arising from split detection. Further, cell-segmentation algorithms that aim to be generally applicable rely on strong cell-membrane features that do not appear in AOSLO split-detection images [14, 15]. Recently appeared work [16–18] corroborates that tailored algorithms that leverage the distinct appearance of photoreceptors are needed for AOSLO split-detection image processing, which is the approach also taken in this paper.

We present an algorithm that automatically detects cones in AOSLO split-detection images without supervision. Our algorithm is among the first that use machine learning to develop and use a photoreceptor model on-the-fly. Comparing to [18], specifically, the approach presented here can tackle both densely and sparsely populated photoreceptor images as it is independent of the spatial arrangement of cones. Further, it introduces contrast enhancement filters, which improve the quality of low signal-to-noise ratio (SNR) images.

2. Materials and methods

This study was conducted in accordance with the tenets of the Declaration of Helsinki (1983 Revision) and the applicable regulatory requirements. After approval of the study and its procedures by the ethics committees of Moorfields Eye Hospital and University College London, informed consent was obtained from all participating subjects prior to enrollment.

2.1. Non-confocal adaptive optics scanning light ophthalmoscope and protocol

Our custom AOSLO, based on [12], was used to acquire cellular-level retinal images from healthy volunteers. The use of AO allows, for the used wavelength, to approach a lateral resolution of $2 \mu\text{m}$, which is the theoretically best resolution achievable due to the numerical aperture of the eye. The next issue is sampling, and this is a balance of area, time, and pixel count. With our system we are able to achieve a nominal sampling resolution of $0.5 \mu\text{m}/\text{pixel}$, and $0.75 \mu\text{m}/\text{pixel}$,

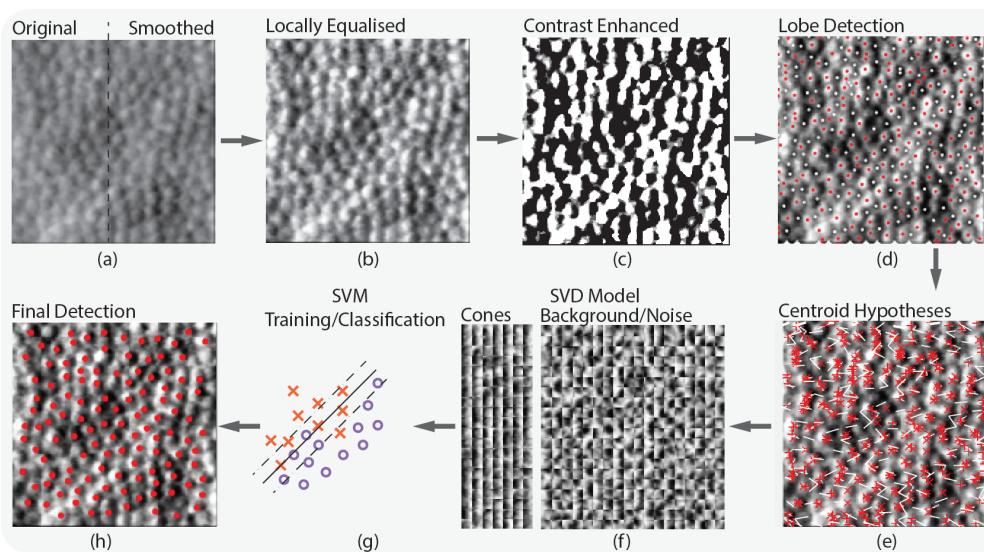


Fig. 2. Algorithm workflow on a representative non-confocal AO image from our Dataset.

for $1^\circ \times 1^\circ$, and $1.5^\circ \times 1.5^\circ$ field-of-view presets, respectively.

Details on non-confocal image formation in AO can be found in [12]. Briefly, the non-confocal (multiple-scattered) light reflected from the photoreceptors is split into two channels, I_D and I_R , which can be considered as light backscattered from the two “sides” of the cones:

$$I_{AO, S} = \frac{I_R - I_D}{I_R + I_D}. \quad (1)$$

In ophthalmoscopy, confocal and split-detection imaging are complementary to each other, with the former revealing the existence of waveguiding photoreceptor outer segments and the latter photoreceptor inner segments. In conditions that affect cones, outer segments appear dark (*i.e.* do not waveguide light) rendering their visualization impossible, whereas split-detection reveals the existence of intact inner segment structure irrespectively of whether outer segments waveguide light or not. Images generated using (1) show one side of the cone as substantially darker than the other side. The developed algorithm exploits this split-nature of the signal. First, it produces plausible cone candidates by generating cone-center hypotheses, and, subsequently, learns a model to classify these candidates as “cones” or “background”. Figure 2 shows the workflow for a sample image.

2.2. Image processing pipeline

Our image acquisition protocol follows the outline reported in [19]. Approximately 50-200 images are acquired per fixation locus, for the fields-of-view mentioned previously. Since the images are stretched due to the sinusoidal motion of the optical scanner, we use the methodology and algorithms of [19] to resample the images with equally spaced pixels laying on a grid. To remove motion artifacts, a subset of the acquired images, usually on the order of 50, is registered and then averaged. These averaged images are the input to the processing pipeline illustrated in Fig. 2 and described in the following. The rationale for selecting the algorithm’s parameters is given in Sec. 3, together with a study of its robustness on parameter changes and the effect of each “detection” step on the final performance metrics.

Photoreceptor size estimation (#1): First, the photoreceptor mosaic in the AO image under examination is analyzed in the frequency domain to extract Yellott's ring [18, 20]. Yellott's ring leads to an estimate of the photoreceptor size, c_w , in the image, provided that the photoreceptor mosaic exhibits regularity. In images of diseased retinas, especially in the case of Stargardt disease that will be examined as a pathology scenario, this value needs to be approximated by the user, and is kept constant for all images within a dataset. The filter sizes required in the following make use of the estimated photoreceptor size, and are expressed in Sec. 3 as percentages of this parameter.

Image preprocessing (#2): A bilateral filter highlights the image edges while suppressing noise [21]. The images are smoothed with a Gaussian kernel of standard deviation σ to remove spurious high-frequency components [see Fig. 2(a)]. This filtering step partially homogenizes the image. Subsequently, the image contrast is enhanced using adaptive histogram equalization (AHE) [22]. AHE operates on $N \times N$ tiles on the image and locally increases the contrast so that the histogram of each tile matches a target histogram, which, in this work, is the Rayleigh distribution (chosen after experimentation). Within AHE, all contrast-enhanced patches are fused to remove artificially introduced tile boundaries. This approach results in a substantial increase of the resolvability of the cone boundaries [see Fig. 2(b)].

Adaptive sliding-window enhancement (#3): To assist with bright/dark cone-lobe segmentation, a “voting” scheme that further increases the contrast of cell-like regions is applied. A rectangular window W of dimension c_w slides over the image with c_w increments in the x and y directions. For each image region, I_w , a grayscale threshold is automatically selected via Otsu’s algorithm [23]. This results in a preliminary classification between background and bright/dark cone pixels within I_w . The classification is stored in an accumulator, I_{acc} , which is updated as the window slides, resulting in increased confidence for pixels that are consistently classified as belonging to either side of the cones:

$$I_{\text{acc}}(i, j) = \sum_{x=i-d_w}^{i+d_w} \sum_{y=j-d_w}^{j+d_w} I_{\text{thresh}}(x, y), \text{ where } I_{\text{thresh}}(x, y) = \begin{cases} 0 & \text{if } I_w(x, y) \leq \mathcal{T}(W) \\ 1 & \text{if otherwise} \end{cases} \quad (2)$$

where $\mathcal{T}(W)$ is the Otsu threshold for window W centered around (x, y) .

This sliding-window contrast enhancement highlights the right/bright and left/dark side of the cones due to its local nature and the multiple partitions of the AO image that are examined due to its combined “voting” scheme.

Non-overlapping extremal regions (#4): Subsequently, the entire image is binarized using an aggressive threshold, t_{bin} , applied twice to extract both bright-lobe pixels and dark-lobe pixels. Aggressive thresholding results in connected components that span multiple cone lobes [see Fig. 2(c)]. Non-overlapping extremal region detection [24] is used to detect the centroids of all potential bright and dark lobes. In our approach, each connected component is iteratively eroded using a circular kernel of radius r_{er} , leading to its split into smaller constituents, which are then iteratively processed. Connected components with an area smaller than a threshold a_{\min} are discarded, while the remaining ones are placed in a tree structure denoting their hierarchy [see Fig. 3(a)]. The leaves of the tree are “maximally stable” components that correspond to cone lobes and their centroids [see Fig. 2(d)].

Cone-centroid hypotheses generation (#5): All lobes that potentially belong to a single cone, *i.e.* represent its bright and dark component in the AO image, are connected in a graph-like structure [see Fig. 2(e)]. The nodes of the graph are the centroids extracted in the previous step, and their connection respects heuristic constraints relating to their distance and the angle they form with respect to the horizontal axis. The constraints arise from the fact that the cone-lobe centroids should not be apart farther than the size of the cone (in pixels), the bright lobe should be on the left of the dark cone (for our setup), and their connecting line should generally be

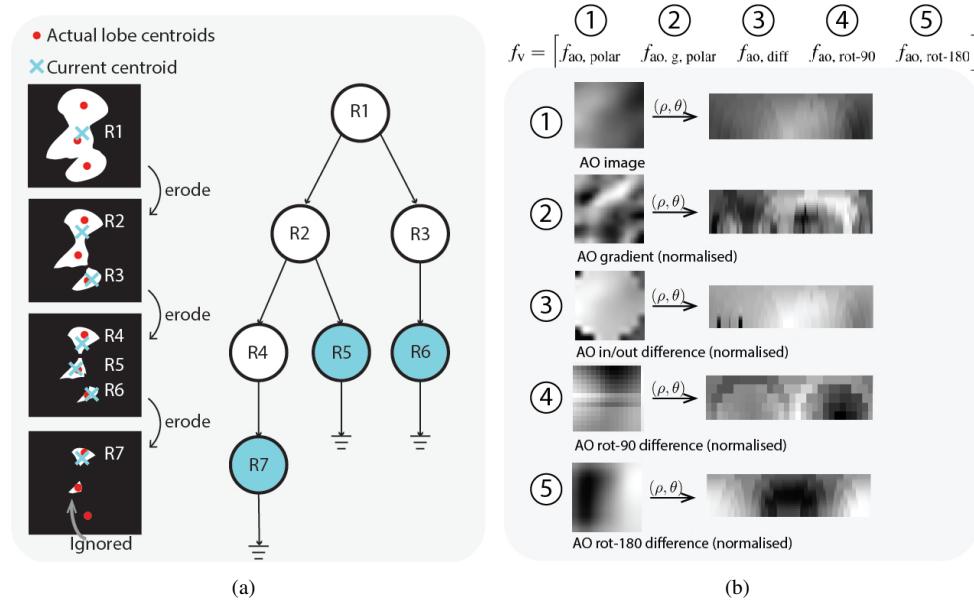


Fig. 3. (a) The non-overlapping extremal region detection algorithm. (b) Components of the feature vector for a cone. Normalization is performed for illustration purposes.

“parallel” to the horizontal image axis. Each plausible centroid couple gives rise to a single cone candidate at the centre of the connecting line, for a total of C candidates. The heuristics retain plausible cone candidates, which will be used in the cone-modelling step. The average separation distance of the bright/dark centroids updates the photoreceptor size, which is now fine-tuned for the specific image. The new value is termed d_w .

Preliminary cone-model extraction (#6): Square patches I_i^{AO} , $i \in \{1 \dots C\}$ of dimension d_w are extracted around each of the C cone-centroid candidates from the AO image. A feature vector, composed as shown in Fig. 3(b), is then generated from each AO/binary patch couple:

$$f_v = [f_{ao, \text{polar}} \quad f_{ao, g, \text{polar}} \quad f_{ao, \text{diff}} \quad f_{ao, \text{rot-90}} \quad f_{ao, \text{rot-180}}] \in \mathbb{R}^{1260 \times 1} \quad (3)$$

where:

- $f_{ao, \text{polar}} \in \mathbb{R}^{252 \times 1}$ is of I^{AO} in polar coordinates, as a vector;
- $f_{ao, g, \text{polar}} \in \mathbb{R}^{252 \times 1}$ is the gradient of I^{BW} in polar coordinates, as a vector;
- $f_{ao, \text{diff}} \in \mathbb{R}^{252 \times 1}$ is the difference between an internal region of I^{AO} and an external region of I^{AO} in polar coordinates, as a vector;
- $f_{ao, \text{rot-90}} \in \mathbb{R}^{252 \times 1}$ and $f_{ao, \text{rot-180}} \in \mathbb{R}^{252 \times 1}$ are the differences between I^{AO} , and I^{AO} rotated by 90° , and 180° , respectively, as a vector, and highlight the contrast-difference between the bright and dark lobes.

All vectors are concatenated to form a $1260 \times C$ matrix D representing all patches. Due to the heuristics, the data in D are representative cone descriptors. Singular value decomposition (SVD) is then performed on matrix D to identify basis vectors describing the cones. The basis vectors, U , corresponding to the n largest singular values form the preliminary cone model.

Training data generation (#7): The dark/bright lobe centroid candidates extracted through the non-overlapping extremal region step are revisited and shifted by $d_w/4$ to approximate their

corresponding cone centroid. Then, each new cone candidate is assigned “cone” or “background” label based on its ability to be represented by the basis vectors extracted in the previous step. In other words:

$$\text{label} = \begin{cases} \text{cone,} & \text{if } \|f_v - U(U^T f_v)\|_2 \leq e_{\text{thresh, cone}} \\ \text{background,} & \text{if } \|f_v - U(U^T f_v)\|_2 \geq e_{\text{thresh, background}} \end{cases} \quad (4)$$

The thresholds are automatically calculated from the SVD reprojection errors as the 90% quantile separator. The labels and feature vectors for the image under examination form a data corpus [see Fig. 2(f)], which expands on the heuristically extracted cones and is used to train a support vector machine (SVM), which captures variability in the cones’ appearance as well as the background [see Fig. 2(g)]. Linear SVM kernel was preferred due to the large number of features and because preliminary experiments showed that Radial Basis Functions underperformed. All feature vectors are normalized based on their mean and standard deviation.

Model refinement and final cone detection (#8): The final step relaxes the heuristic criteria of #5 to join the majority cone-lobe pairs and create cone candidates. These candidates are then classified through the SVM as “background” or “cone” [see Fig. 2(h)]. Detected centroids that are less than $\frac{d_w}{2}$ distance apart are considered to represent the same cell and are fused together.

It is worthwhile to note that the created model can be potentially transferred from image to image, or even refined based on supervised user input. Further, it is possible to consider all images within the dataset as a single montage, giving rise to a more “general” photoreceptor model. While there are several ways to process the data for photoreceptor detection, in this work, we treat each image independently.

3. Results

Volunteers had their retinas scanned, and the cones in pre-processed (averaged) AO images were annotated by two graders. The field-of-view covered by each image is approximately $80 \times 80 \mu\text{m}$; the pixel densities vary but the retinal area does not. This protocol allows us to test the developed algorithm for varying photoreceptor densities.

Six healthy volunteers were scanned and a total of 60 images of healthy retinas were used. The range of examined degrees away from the fovea covered real life clinical scenarios as to which areas would be imaged for quantification purposes. Six volunteers affected by Stargardt disease were scanned, and 31 images were used to evaluate the method. Their range is between 1° and 6° degrees away from the fovea, with approximately 3-4 images per degree across the entire dataset.

Exemplary images from different parts of the retina for healthy volunteers, and volunteers affected by Stargardt disease, are shown in Fig. 4(top), and Fig. 5(top), respectively.

The graders marked the photoreceptors in both healthy and diseased images and generated the datasets’ ground-truth detections. Only the split-detection images were used for marking. These manual detections only serve as the ground-truth, and are nowhere used within the algorithmic framework, which is unsupervised. The two graders’ performance was compared via the DICE coefficient:

$$\text{DICE} = \frac{2 \times (\#P_m)}{(\#P_1) + (\#P_2)} \quad (5)$$

where $\#P_m$ is the number of photoreceptors detected by both graders, $\#P_1$ the number detected by grader 1, and $\#P_2$ the number detected by grader 2.

For the two human graders, DICE was $92.9\% \pm 4.3\%$ for healthy retinas, and $79.7\% \pm 18.9\%$ for diseased retinas. This shows the good correspondence between the graders when healthy retinas are annotated, and suggests that either of them can be used to compare against a computer algorithm, taking into account the calculated inter-grader variation. The discrepancy in the

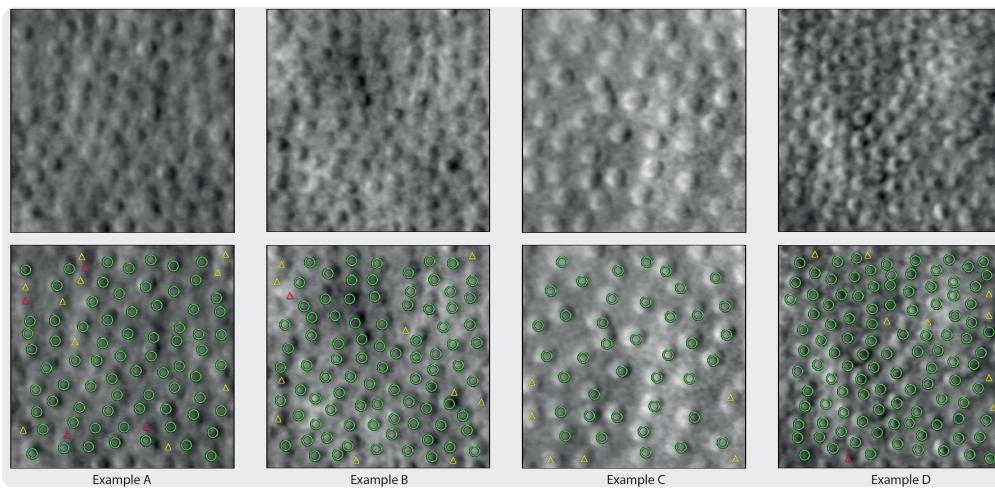


Fig. 4. Example healthy images from the AOSLO (top: raw, bottom: processed and annotated). Green circles: matched detections. Yellow triangles: unmatched human detections. Red triangles: unmatched algorithm detections.

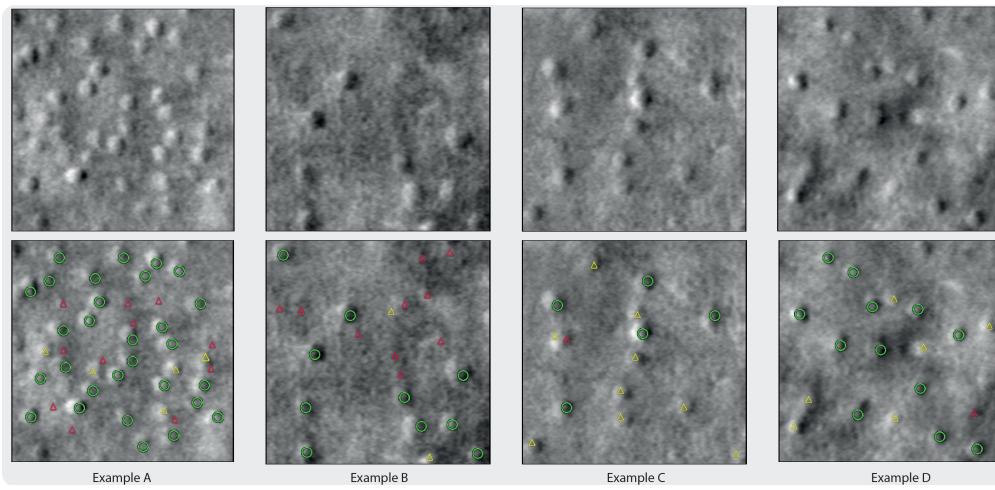


Fig. 5. Example Stargardt images from the AOSLO (top: raw, bottom: processed and annotated). Green circles: matched detections. Yellow triangles: unmatched human detections. Red triangles: unmatched algorithm detections.

metrics relating to diseased retinas provides an indication of the performance that an algorithm should achieve, and also highlights the intrinsic challenge in processing those datasets. Precision and recall will be used as additional metrics. Precision is the number of automatically detected photoreceptors that correspond to a manually annotated photoreceptor, while Recall is the number of photoreceptors detected from the set of manually annotated ones.

3.1. Performance of algorithm

Our algorithm was developed in Matlab R2015b, and was executed on a 2.7 GHz Intel Core i7 with 16 Gb RAM. The photoreceptor detection comparison and visualizations were developed in

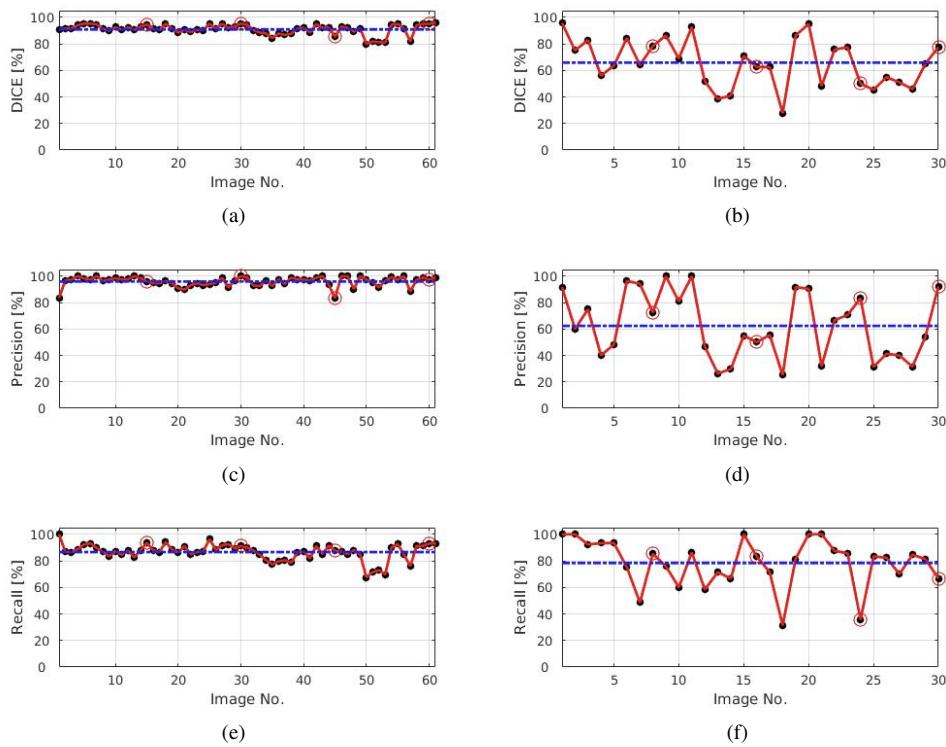


Fig. 6. Performance metrics for images of healthy retinas [(a), (c), (d)], and images of retinas affected by Stargardt disease [(b), (d), (f)]. First row: DICE coefficient; Second row: Precision metric; Third row: Recall metric.

MeVisLab. The parameters used are given in Table 1.

The algorithm estimated the expected size of the cone as input for each image, to guide the sliding window detector and the left/right centroid linking. This number need only be approximate and was estimated via Yellott's ring for healthy samples, or measured on a single image for Stargardt images. This value is updated within the algorithm after the dark/bright lobes are detected. The parameters of the bilateral and Gaussian filters are chosen experimentally to suppress spurious dark/bright regions that may be accentuated by subsequent contrast enhancement. The tiles of AHE are selected as 4 based on the image size, as it was found that a larger number of tiles causes the filter to act on very local neighborhoods, distorting the appearance of the photoreceptors. The binarisation threshold is chosen as high enough to ensure that most bright/dark centroids are segmented, even if they are fused together. The eroding element has a radius r_{er} of a single pixel to ensure maximal sensitivity to the detection of the connected components. Finally, the threshold for considering a connected component as a valid photoreceptor lobe, *i.e.* the area a_{min} , is selected based on the estimated photoreceptor size. In images from unaffected individuals, where a large number of cells is expected, it is set to a low value, as it is preferable to consider more potential lobes and discard them in the subsequent classification, rather than missing out on potential candidates. In diseased images, it is set to a larger value to disregard spurious cone-like regions.

With the parameters of Table 1, the SVM classifier used on average 260 positive and 380 negative samples for training in each healthy-image case, while 37 positive and 240 negative

Table 1. Parameters of the algorithm: fixed within the datasets.

Parameter	Samples	
	Healthy	Stargardt
Cell width, c_w [pixels]	Per [20]	12
Bilateral filter width [pixels]	$0.25c_w$	$0.25c_w$
Bilateral filter σ	$0.03c_w$	$0.1c_w$
Gaussian filter width [pixels]	$0.25c_w$	$0.25c_w$
Gaussian filter sigma σ	$0.03c_w$	$10c_w$
Adaptive histogram equalization N	4	4
Adaptive sliding window width d_w [pixels]	d_w	d_w
Image binarisation threshold t_{bin}	0.98	0.98
Extremal region - eroding element r_{er} [pixels]	1	1
Extremal region minimum area a_{min} [pixels]	$0.1c_w$	c_w
SVD quantile value [%]	90%	90%

samples were used in the diseased-retinal image scenario. In both cases, the number of negative samples is larger than the positive. This can be explained by the fact that many lobes appear as false positives, which are filtered by the SVM model after training. The large number of samples, in general, is attributed to the fact that several plausible cone centroids are in close proximity to one another, and essentially correspond to the same photoreceptor - no proximity filtering is conducted at this stage. Ten-fold cross-validation demonstrated an average of 15% misclassification for images of both healthy and diseased retinas for the trained SVM.

Figure 4(bottom) and Fig. 5(bottom) show the detected photoreceptors for the example images. In agreement with the literature, we do not consider photoreceptors that appear at the boundaries of the image as the algorithm there encounters only partial data. When less than half of the photoreceptor is visible, the algorithm cannot detect it.

Figures 6(a), 6(c), and 6(e) show the Precision, Recall, and DICE indices for images of healthy retinas. The red circles highlight the examples selected for Fig. 4. Precision is $96.2\% \pm 3.4\%$, Recall is $86.4\% \pm 6.1\%$, and DICE is $90.9\% \pm 4.0\%$. The values are close to the discrepancy also exhibited by the two human graders meaning that the algorithm may replace the human grader with confidence. The high Precision and relatively low recall indicates that the algorithm is conservative in its estimates, returning true photoreceptors and avoiding false positives at the expense of false negatives.

Figure 6(b), Fig. 6(d), and 6(f) show the Precision, Recall, and DICE for images of retina suffering from Stargardt disease. The red circles highlight the examples selected for Fig. 5. Precision is $64.8\% \pm 23.6\%$, Recall is $78.2\% \pm 17.9\%$, and DICE is $67.7\% \pm 16.9\%$. The values are close to the discrepancy exhibited by the two human graders when the standard deviations are considered. These results also indicate that more sophisticated image features may be required for successful photoreceptor detection in images of diseased retinas.

Finally, on average, images were processed within $10.45s \pm 0.69$ s per frame, which is a drastic improvement over manual annotation speed, which may take up to 2 – 3 mins for the shown images and several hours to cover the entire retina. Thus, the proposed algorithm is accurate and appreciably fast to assist in cone demographics.

3.2. Robustness to parameters

The parameters displayed in Table 1 were selected after experimentation as a set that produces reasonable results. To understand the algorithm's robustness to parameter selection, we applied a “noise” profile to a set of parameters around their selected values. For each experiment conducted, we specify the range of Precision, Recall, and DICE that was calculated to give an appreciation

of the algorithm's robustness to parameter variations. Overall, the performance metrics have their highest value in the proximity of the previously selected parameter, and drop off as the "noise" increases. Further, a trade-off between increased Precision versus increased Recall is observed, *i.e.* there is a natural trade-off between the number of false negatives versus the number of false positives that the algorithm returns. This section highlights that the algorithm can be used in clinical practice with minimal tuning, as it is generally robust to perturbations of its governing parameters.

3.2.1. Experimentation on data from healthy volunteers

The algorithm is evaluated on 10 images from the dataset acquired from healthy volunteers. The images are selected to cover all patients, and varying degrees of eccentricity. For these images, the calculated average metric ranges are presented below.

Bilateral/Gaussian filter sigma, perturbed by -15% to 15% : Varying this parameter resulted in Precision within [93.8%, 96.2%], Recall within [86.4%, 87.8%], and DICE within [90.4%, 91.2%].

Bilateral/Gaussian filter size, perturbed by -2 pixels to 2 pixels, corresponding to approximately -25% to 25% : Varying this parameter resulted in Precision within [94.1%, 95.4%], Recall within [86.3%, 88.1%], and DICE within [89.8%, 91.9%].

Adaptive histogram equalization N, perturbed by -2 to 2 tiles, corresponding to approximately -50% to 50% : Varying this parameter resulted in Precision within [95.0%, 96.2%], Recall within [85.6%, 87.8%], and DICE within [89.3%, 91.1%].

Image binarization threshold, perturbed by -15% to 0% , as the maximum value of this parameter is 1.0 : Varying this parameter resulted in Precision within [94.7%, 96.2%], Recall within [86.9%, 87.8%], and DICE within [90.4%, 90.9%].

SVD quantile inclusion threshold, perturbed by -15% to 0% , as the maximum value of this parameter is 1.0 : Varying this parameter resulted in Precision within [94.5%, 96.2%], Recall within [86.1%, 87.8%], and DICE within [90.5%, 90.9%].

The small range of the metrics indicates that the algorithm is robust to reasonable variations of all the important parameters when images of healthy retinas are considered.

3.2.2. Experimentation on data from volunteers affected by Stargardt disease

The algorithm is evaluated on 10 images from the dataset acquired from volunteers affected by Stargardt disease. The images are selected to cover all patients, and varying degrees of eccentricity. For these images, the calculated average metric ranges are presented below.

Bilateral/Gaussian filter sigma, perturbed by -15% to 15% : Varying this parameter resulted in Precision within [54.0%, 64.0%], Recall within [75.3%, 75.9%], and DICE within [57.3%, 63.8%].

Bilateral/Gaussian filter size, perturbed by -2 pixels to 2 pixels, corresponding to approximately -25% to 25% : Varying this parameter resulted in Precision within [41.8%, 60.6%], Recall within [74.1%, 87.2%], and DICE within [51.8%, 57.7%].

Adaptive histogram equalization N, perturbed by -2 to 2 tiles, corresponding to approximately -50% to 50% : Varying this parameter resulted in Precision within [58.6%, 59.9%], Recall within [73.2%, 73.4%], and DICE within [58.9%, 60.2%].

Image binarization threshold, perturbed by -15% to 0% , as the maximum value of this parameter is 1.0 : Varying this parameter resulted in Precision within [60.3%, 61.8%], Recall within [73.4%, 73.6%], and DICE within [60.5%, 61.5%].

SVD quantile inclusion threshold, perturbed by -15% to 0% , as the maximum value of this parameter is 1.0 : Varying this parameter resulted in Precision within [62.2%, 64.1%], Recall within [73.4%, 73.6%], and DICE within [61.7%, 63.0%].

Overall, it can be argued that the algorithm is robust to the examined parameters, apart from

Table 2. Contribution of each classification step.

Step	Samples and Metrics [%]					
	Healthy			Stargardt		
Precision	Recall	DICE	Precision	Recall	DICE	
Step #5	74.4 ± 8.2	84.1 ± 7.4	78.6 ± 6.2	35.7 ± 16.7	50.4 ± 14.8	40.1 ± 14.0
Step #7	76.0 ± 9.2	94.2 ± 4.8	81.7 ± 6.2	30.4 ± 18.6	58.5 ± 17.6	37.7 ± 17.6
Step #8	96.2 ± 3.4	86.4 ± 6.1	90.9 ± 4.0	64.8 ± 23.6	78.2 ± 17.9	67.7 ± 16.9

the Bilateral/Gaussian filter size. Perturbation of this parameter leads to a large discrepancy in the Precision, Recall, and DICE metrics. Larger values of σ lead to more profound smoothing of the image and suppression of the noise prevalent in the images of Stargardt-affected retinas. After all, the lack of photoreceptors creates a homogeneous retinal background wherein small variations in the signal captured by the AOSLO system can be interpreted as a dark or bright photoreceptor lobe during image processing. Therefore, smoothing the image with increased σ filters-out those regions, and leads to better cone-candidates and photoreceptor model, enabling the SVM of the final stage to perform robust classification.

3.3. Contribution of each classification step

To further evaluate the algorithm, we characterized the effect that each classification step has on the performance metrics. In the examined scenario, we assumed that the algorithm terminated after: (1) The bright/dark lobe connection (Step #5), (2) The SVD and cone-set increase (Step #7), and (3) The SVM (Step #8).

Table 2 contains the results of this evaluation, where red values indicate the maximum Precision/Recall/DICE. It can be seen that each step substantially improves the calculated metrics, and the best outcomes are delivered when the entire pipeline is considered. For images from healthy volunteers, the Precision is boosted from 74.4% to 96%, and this is echoed in the improvement of DICE from 70.6% to 90.9%. These benefits are expected as it is only at the final step that a robust cone model is extracted, allowing all plausible candidates to be revisited and robustly classified as “background” or “cone”. An exception is observed for the Recall value calculated after Step #7, which is higher than after the final SVM step. This is explained by the fact that Recall indicates how many photoreceptors from the entire manually annotated set are retrieved, and is not affected by false positives.

Similar observations hold for the dataset containing images from individuals affected from Stargardt disease, as Table 2 indicates.

3.4. Comparison with state-of-the-art

This section compares our algorithm with the method presented in [18]. As with our approach, [18] also detects and couples bright/dark lobes. Contrary to our approach, however, it is heavily based on the periodic nature of the photoreceptor mosaic and does not contain machine-learning elements that capture the intricacies of photoreceptor appearances.

We used a 3-tier approach for an exhaustive comparison. First, we compared our own implementation of [18] on the healthy and diseased retinal dataset acquired for the current paper (Moorfields Datasets). Following the error-estimation protocol, we calculated the DICE and Precision indices for [18] when grader 1 is assumed, again, as the ground truth.

Second, to reduce dataset bias, we compared our algorithm and our implementation of [18] on the dataset and manual cone annotations used for the purposes of [18] (Medical College of Wisconsin Datasets). Again, DICE and Precision indices were calculated for comparison.

Note, however, that our results are based on our own implementation of [18], based on the published information, and may lack the optimizations that delivered the outcomes reported

in [18] (the software is not available online). Therefore, as a final comparison, to reduce implementation bias we compared the results of our algorithm on the dataset of [18] (Medical College of Wisconsin Dataset), with the results reported in [18].

3.4.1. Comparison on Moorfields datasets

Our implementation of [18] was applied on the dataset gathered for this paper. The DICE and Precision indices were calculated at $70.4\% \pm 11.8\%$, and $66.7\% \pm 14.0\%$ for images of healthy retinas. For retinas suffering from Stargardt disease, however, the metrics drop to $13.6\% \pm 10.2\%$, and $9.2\% \pm 7.2\%$.

Therefore, our algorithm outperformed the state-of-the-art on both healthy and diseased datasets. This could be attributed to: a) machine learning converges to the intricate photoreceptor model that leads to improved results, especially in non-periodic photoreceptor mosaics, and b) photoreceptor mosaics in examined images of Stargardt disease have no periodic nature, which leads to erroneous filtering when [18] is used.

3.4.2. Comparison on Medical College of Wisconsin datasets

To remove dataset bias, we also compared the two algorithms on the dataset of [18]. The dataset contained a total of 80 AOSLO non-confocal images from 10 healthy volunteers, acquired under the same conditions described in our paper but with a different AOSLO system of the same specifications. Extensive details on the acquisition protocol are given in [18]. Our algorithm's DICE performance was $90.4\% \pm 4.0\%$, whereas the Precision was $91.2\% \pm 7.1\%$. The compared algorithm exhibited a DICE coefficient of $79.0\% \pm 13.2\%$ and a Precision index of $76.8\% \pm 14.1\%$. As with the previous comparison, the algorithm proposed herein appears more robust and closer to the annotations selected by graders.

The results presented in [18] for this dataset are $93\% \pm 3\%$ for the DICE coefficient, and $93\% \pm 7\%$ for the Precision. Using the results from the literature, the DICE and Precision indices are comparable. No metrics for images from affected volunteers are given in [18].

4. Conclusions and discussion

Split-detection AOSLO is an imaging modality that allows imaging of individual cone photoreceptors inner segment, and is suggested to facilitate early disease diagnosis and more precise measurement of treatment efficacy. Manually annotating and counting the individual photoreceptors is time consuming and prone to errors. Counting the photoreceptors of a single patient may take up to several hours, which makes the imaging technology impractical and limits its true potential. Hence, automated ways of detecting and counting photoreceptors are the first step towards realizing the vision of early detection of eye pathologies and meticulous therapy effect tracking.

This paper presented an algorithm for automatic detection of cones in such images, removing the labor of manual annotation and enabling further work on density quantification, mosaicking, and re-localization. The presented work leverages the split-nature of the appearance of the photoreceptors in AOSLO non-confocal images, and creates a model of the photoreceptors that is subsequently used by a classification framework. Naturally, if a different split detection of offset pinhole is used, as in [13], the algorithm would need adjustments and adaptations.

More specifically, the algorithm was first tested on 60 images collected from 6 healthy volunteers, and achieved recognition rates (DICE measure) of up to a maximum of 90.9%. We vary the eccentricity of imaging, and consider images of varying SNR. Then, we evaluated the algorithm on 31 pathological images (Stargardt disease) from 6 volunteers, reaching a maximum of 67.7% recognition (DICE measure). Finally, we evaluated our algorithm on a dataset from the literature, on 80 images collected from 10 healthy volunteers, reaching a maximum DICE of 90.4%. All results compared well with manual annotation, and were superior to the results from

implementations of state-of-the-art algorithms in healthy images and images from volunteers affected by Stargardt disease.

In addition, we examined the contribution of each classification step in the final algorithm performance, and concluded that indeed it is the combination of cstep-wise photoreceptor modelling that leads to superior performance. Finally, we demonstrated that the algorithm is generally robust to parameter selection, for perturbation of $\pm 15\%$ around the selected values. Therefore, it can be used in clinical practice with minimal fine-tuning.

Our short-term goal is to package the algorithm into a single application that can be delivered to research groups carrying out research in the cellular imaging field. Our longer-term goal to enable interactive photoreceptor detection in retinal video streams.

The developed software will be made available online at <https://goo.gl/7ZoXas> to assist with future comparisons.

Funding

National Institute for Health Research Biomedical Research Centre at Moorfields Eye Hospital National Health Service Foundation Trust and UCL Institute of Ophthalmology; Moorfields Eye Hospital Special Trustees; Moorfields Eye Charity and The Wellcome Trust [099173/Z/12/Z]; EPSRC-funded UCL Centre for Doctoral Training in Medical Imaging [EP/L016478/1]; EPSRC-funded UCL Future Leaders Award; Fight for Sight New Lecturers Grant [1728/29]; European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme [714562]; National Institutes of Health [R01EY017607 and P30EY001931].