

# Multi-surface segmentation of OCT images with AMD using sparse high order potentials

JORGE OLIVEIRA,<sup>1,5</sup> SÉRGIO PEREIRA,<sup>1</sup> LUÍS GONÇALVES,<sup>2</sup>  
MANUEL FERREIRA,<sup>3,4</sup> AND CARLOS A. SILVA<sup>1,6</sup>

<sup>1</sup>CMEMS-UMinho, University of Minho, 4800-058 Guimarães, Portugal

<sup>2</sup>Oftalmocenter, Azurém, 4800-045, Guimarães, Portugal

<sup>3</sup>Algoritmi Center, University of Minho, 4800-058 Guimarães, Portugal

<sup>4</sup>ENERMETER, Parque Industrial Celeirós, 4705-025, Braga, Portugal

<sup>5</sup>id4327@alunos.uminho.pt

<sup>6</sup>csilva@dei.uminho.pt

**Abstract:** In age-related macular degeneration (AMD), the quantification of drusen is important because it is correlated with the evolution of the disease to an advanced stage. Therefore, we propose an algorithm based on a multi-surface framework for the segmentation of the limiting boundaries of drusen: the inner boundary of the retinal pigment epithelium + drusen complex (IRPEDC) and the Bruch's membrane (BM). Several segmentation methods have been considerably successful in segmenting retinal layers of healthy retinas in optical coherence tomography (OCT) images. These methods are successful because they incorporate prior information and regularization. Nonetheless, these factors tend to hinder the segmentation for diseased retinas. The proposed algorithm takes into account the presence of drusen and geographic atrophy (GA) related to AMD by excluding prior information and regularization just valid for healthy regions. However, even with this algorithm, prior information and regularization still cause the oversmoothing of drusen in some locations. Thus, we propose the integration of local shape prior in the form of a sparse high order potentials (SHOPs) into the algorithm to reduce the oversmoothing of drusen. The proposed algorithm was evaluated in a public database. The mean unsigned errors, relative to the average of two experts, for the inner limiting membrane (ILM), IRPEDC and BM were  $2.94 \pm 2.69$ ,  $5.53 \pm 5.66$  and  $4.00 \pm 4.00$   $\mu\text{m}$ , respectively. Drusen areas measurements were evaluated, relative to the average of two expert graders, by the mean absolute area difference and overlap ratio, which were  $1579.7 \pm 2106.8$   $\mu\text{m}^2$  and  $0.78 \pm 0.11$ , respectively.

© 2016 Optical Society of America

**OCIS codes:** (100.0100) Image processing; (170.4500) Optical coherence tomography; (170.4470) Ophthalmology.

## References and links

1. A. Wood, "Retinal structure and function in age-related maculopathy," Ph.D. thesis, Cardiff University (2011).
2. Access Economics Pty Limited, "The Global Economic Cost of Visual Impairment," Tech. Rep. March (2010).
3. S. G. Schuman, A. F. Koreishi, S. Farsiu, S.-h. Jung, J. a. Izatt, and C. a. Toth, "Photoreceptor layer thinning over drusen in eyes with age-related macular degeneration imaged in vivo with spectral-domain optical coherence tomography," *Ophthalmology* **116**, 488–496 (2009).
4. G. Gregori, Z. Yehoshua, C. a. D. a. Garcia Filho, S. R. Sadda, R. Portella Nunes, W. J. Feuer, and P. J. Rosenfeld, "Change in drusen area over time compared using spectral-domain optical coherence tomography and color fundus imaging," *Invest. Ophthalmol. Vis. Sci.* **55**, 7662–7668 (2014).
5. P. T. Johnson, G. P. Lewis, K. C. Talaga, M. N. Brown, P. J. Kappel, S. K. Fisher, D. H. Anderson, and L. V. Johnson, "Drusen-Associated Degeneration in the Retina," *Invest. Ophthalmol. Vis. Sci.* **44**, 4481–4488 (2003).
6. F. A. Folgar, E. L. Yuan, M. B. Sevilla, S. J. Chiu, S. Farsiu, E. Y. Chew, and C. A. Toth, "Drusen Volume and Retinal Pigment Epithelium Abnormal Thinning Volume Predict 2-Year Progression of Age-Related Macular Degeneration," *Ophthalmology* **123**, 39–50 (2016).
7. S. J. Chiu, J. A. Izatt, R. V. O. Connell, K. P. Winter, C. A. Toth, and S. Farsiu, "Validated Automatic Segmentation of AMD Pathology Including Drusen and Geographic Atrophy in SD-OCT Images," *Invest. Ophthalmol. Vis. Sci.* **53** (2012).
8. N. Jain, S. Farsiu, A. a. Khanifar, S. Bearely, R. T. Smith, J. a. Izatt, and C. a. Toth, "Quantitative comparison of drusen segmented on SD-OCT versus drusen delineated on color fundus photographs." *Invest. Ophthalmol. Vis. Sci.* **51**, 4875–4883 (2010).

9. F. G. Schlanitz, C. Ahlers, S. Sacu, C. Schütze, M. Rodriguez, S. Schriefl, I. Golbaz, T. Spalek, G. Stock, and U. Schmidt-Erfurth, "Performance of drusen detection by spectral-domain optical coherence tomography." *Invest. Ophthalmol. Vis. Sci.* **51**, 6715–6721 (2010).
10. D. Iwama, M. Hangai, S. Ooto, A. Sakamoto, H. Nakanishi, T. Fujimura, A. Domalpally, R. P. Danis, and N. Yoshimura, "Automated assessment of drusen using three-dimensional spectral-domain optical coherence tomography." *Invest. Ophthalmol. Vis. Sci.* **53**, 1576–1583 (2012).
11. S. Farsiu, S. J. Chiu, J. a. Izatt, and C. a. Toth, "Fast Detection and Segmentation of Drusen in Retinal Optical Coherence Tomography Images," *Proc. SPIE* **6844**, 68440D1 (2008).
12. K. Yi, M. Mujat, B. H. Park, W. Sun, J. W. Miller, J. M. Seddon, L. H. Young, J. F. de Boer, and T. C. Chen, "Spectral domain optical coherence tomography for quantitative evaluation of drusen and associated structural changes in non-neovascular age-related macular degeneration." *Br. J. Ophthalmol.* **93**, 176–181 (2009).
13. Q. Chen, T. Leng, L. Zheng, L. Kutzscher, J. Ma, L. de Sisternes, and D. L. Rubin, "Automated drusen segmentation and quantification in SD-OCT images." *Med. Image Anal.* **17**, 1058–1072 (2013).
14. K. Li, X. Wu, D. Z. Chen, and M. Sonka, "Optimal surface segmentation in volumetric images—a graph-theoretic approach." *IEEE Trans. Pattern Anal. Mach. Intell.* **28**, 119–134 (2006).
15. M. K. Garvin, M. D. Abramoff, X. Wu, S. R. Russell, T. L. Burns, and M. Sonka, "Automated 3-D intraretinal layer segmentation of macular spectral-domain optical coherence tomography images." *IEEE Trans. Med. Imag.* **28**, 1436–1447 (2009).
16. Q. Song, X. Wu, and Y. Liu, "Simultaneous searching of globally optimal interacting surfaces with shape priors," in "IEEE Conf. Comput. Vis. Pattern Recognit." pp. 2879–2886 (2010).
17. P. a. Dufour, L. Ceklic, H. Abdillahi, S. Schröder, S. De Dzanet, U. Wolf-Schnurbusch, and J. Kowal, "Graph-based multi-surface segmentation of OCT data using trained hard and soft constraints." *IEEE Trans. Med. Imag.* **32**, 531–543 (2013).
18. C. Rother, V. Kolmogorov, V. Lempitsky, and M. Szummer, "Optimizing Binary MRFs via Extended Roof Duality," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* pp. 1–8 (2007).
19. C. Rother and P. Kohli, "Minimizing sparse higher order energy functions of discrete variables," in "Proc. IEEE Conf. Computer Vision and Pattern Recognition," (2009), pp. 1382–1389.
20. F. Shi, X. Chen, H. Zhao, W. Zhu, D. Xiang, E. Gao, M. Sonka, and H. Chen, "Automated 3-D Retinal Layer Segmentation of Macular Optical Coherence Tomography Images with Serous Pigment Epithelial Detachments." *IEEE Trans. Med. Imag.* **34**, 441–452 (2014).
21. M. Mayer, J. Hornegger, C. Mardin, and R. Tornow, "Retinal Nerve Fiber Layer Segmentation on FD-OCT Scans of Normal Subjects and Glaucoma Patients," *Biomed. Opt. Express* **1**, 1358–1383 (2010).
22. P. P. Srinivasan, S. J. Heflin, J. a. Izatt, V. Y. Arshavsky, and S. Farsiu, "Automatic segmentation of up to ten layer boundaries in SD-OCT images of the mouse retina with and without missing layers due to pathology," *Biomed. Opt. Express* **5**, 348 (2014).
23. S. J. Chiu, X. T. Li, P. Nicholas, C. a. Toth, J. a. Izatt, and S. Farsiu, "Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation." *Opt. Express* **18**, 19413 (2010).
24. J. M. Hernández-Lobato, M. A. Gelbart, M. W. Hoffman, R. P. Adams, and Z. Ghahramani, "Predictive Entropy Search for Bayesian Optimization with Unknown Constraints," in "Proceedings of the 32nd Int. Conf. on Mach. Learning," (JMLR.org, 2015), pp. 1699–1707.
25. J. Oliveira, S. Pereira, L. Gonçalves, M. Ferreira, and C. A. Silva, "Surface Segmentation of OCT Images with AMD using Sparse High Order Potentials: supplementary material," <http://www.dei.uminho.pt/pessoas/csilva/drusen> (2016).
26. K. B. Koozekanani, D. and C. Roberts, "Retinal thickness measurements from optical coherence tomography using a Markov boundary model," *IEEE Trans. Medical Imag.* **20**, 900–9016 (2001).

## 1. Introduction

AMD causes progressive degeneration of the central vision, being aging its main risk factor [1]. In 2010, it was estimated as the leading cause of visual impairment in developed countries [2]. With aging, the appearance of a few drusen is normal [3]. Nonetheless, an abnormal quantity of drusen is the earliest clinical feature of AMD [3], being drusen quantity and size correlated with the risk of progression of AMD to an advanced stage [4].

Drusen are extracellular deposits of material (e.g., lipids and proteins) accumulated between the retinal pigment epithelium (RPE) and the BM (Fig. 1) [5]. These lesions can have quite different shape, size, composition, color and border definition [1]. Subretinal drusen are also indicative of progression of AMD and composed of the same material as regular drusen, the main difference is the accumulation of extracellular material in the inner side of RPE [6]. In this paper, we will consider the limits of drusen as the IRPEDC and BM boundaries (Fig. 1) because these encompass both types of drusen and clinical features related to AMD, such as

hyperreflective foci or drusen remnants over GA [7].

To visualize drusen, retinography is currently the gold standard technique [8]. More recently, OCT appeared as an alternative imaging modality (Fig. 1). The OCT operating principle is based on the measurement of delay and magnitude of backscattered infrared light. Concerning drusen visualization, OCT presents some advantages over retinography, namely: better contour definition, higher number of visible drusen, higher sensitivity to size change (in follow-up exams) and no interference from eye fundus pigmentation [4, 8–10]. With these reasons in mind, OCT was the modality selected for this work.

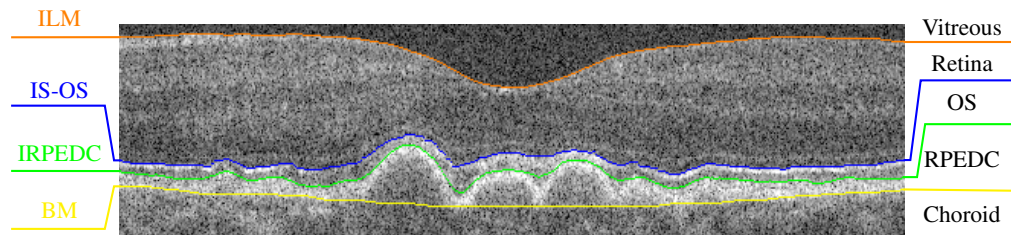


Fig. 1. Automatic segmentation with the proposed method of an OCT frame in the fovea region with the presence large drusen. IS-OS represents the boundary between the inner segment (IS) and the outer segment (OS) of the photoreceptor layer, while retina layer refers to the neurosensory retina, which excludes the RPEDC.

Presently, drusen evaluation by visual inspection is the common clinical practice. This is a subjective method that may allow changes to pass unnoticed. Alternatively, drusen can be segmented manually by an expert grader, which is a time-consuming task. To address this issue, some automatic segmentation algorithms have been proposed in the literature.

Farsiu *et al.* [11] were among the first to develop a fully automated drusen segmentation algorithm. Initially, this method detects the ILM through a peak search in the vertical gradient; this operation is necessary because ILM and RPE have similar intensities. Then, it obtains a preliminary outer RPE (ORPE) segmentation from the image intensity, which is used as the initial shape in an active contours framework based on gradient flow. The ORPE boundary is then fitted to a 2<sup>nd</sup> or 4<sup>th</sup> order polynomial to estimate the BM. The distance between the ORPE and BM corresponds to the drusen volume.

Algorithms such as [10, 12, 13] use a similar structure, i.e., ILM segmentation, coarse and fine segmentation of one of the RPE boundaries and BM estimation through smoothing. These algorithms also use similar methods, such as thresholding, gradient peak detection and polynomial fitting.

Chiu *et al.* [7] suggested a distinct approach based on graph theory and dynamic programming, particularly the Dijkstra's shortest path algorithm. The algorithm segments ILM, BM and IRPEDC sequentially. All of the segmentations use the vertical gradient to compute the weights of the graphs. The weights for the segmentations of IRPEDC and BM use additional sources of information. The graph network of each boundary is limited by the previous segmentation to reduce the computation time and to avoid interference of other boundaries.

Li *et al.* [14] proposed a graph-based method for segmenting multiple surfaces in layered tissues, which was later applied by Garvin *et al.* [15] to retinal OCT images. This method searches for the minimum cost surface, while being constrained with hard smoothness and interaction priors. The smoothness prior limits local variations of the surface, while the interaction prior restricts the distance between two surfaces.

Afterwards, this framework was extended by adding soft constraints [16, 17]. Song *et al.* [16] proposed the soft smoothness constraint, which allows to control the rigidity of the surface. While Dufour *et al.* [17] later added the soft interaction constraint, which can be seen as a force that adjusts the distance of two surfaces to a specific value. This extended framework was tested in

retinas with drusen by Dufour *et al.* [17]. They indicate that the algorithm fails to segment large drusen, pointing out as the main cause the attraction of the BM surface to the ORPE boundary. These two boundaries are indistinguishable in healthy regions, however, in drusen regions, these are disconnected due to the presence of drusen material.

The main goal of this work is to improve the segmentation of the IRPEDC and BM boundaries in regions with drusen for retinas of intermediate AMD patients and to that end, we make two contributions. The first is an algorithm capable of handling the variability of images of patients with intermediate AMD based on the framework proposed by Dufour *et al.* [17]. Our proposal takes into account the presence of clinical features related with intermediate AMD, such as drusen and GA, by excluding prior information and regularization valid just for healthy regions. Even when using this algorithm, prior information and regularization can still cause oversmoothing of drusen. To address this issue, our second contribution is the addition of local shape priors in the form of SHOPS to the framework.

The remainder of this article is organized as follows. The next section presents the multi-surface segmentation framework, as well as the proposed extension using SHOPS. Each energy term will be described in detail. Afterwards, the proposed algorithm is outlined. Section 3 describes the experimental methods. It starts with the database used for evaluating the proposed method. Then, it explains how the parameters were defined and how we implemented the algorithms proposed by Dufour *et al.* [17] for comparing results. In section 4, the results of the proposed method are presented and discussed. The last section presents the concluding remarks.

## 2. Methods

Segmenting retinal boundaries can be posed as a problem of multiple surface segmentation, which consists in searching the minimum cost surface under some regularizing constraints and spatial priors [15]. The search for a minimum cost surface can be converted into a problem of optimizing the minimum closed set of a graph that in turn can be transformed into a minimum s-t cut problem [14] and solved efficiently by methods such as the quadratic pseudo-boolean optimization (QPBO) [18].

### 2.1. Multi-surface segmentation framework

The multi-surface segmentation with SHOPS can be represented by the minimization of the following energy function:

$$E(\mathbf{S}) = \sum_{i=1}^n (E_{bound}(S_i) + E_{smooth}(S_i)) + \sum_{i=1}^{n-1} \sum_{j=i+1}^n E_{inter}(S_i, S_j) + \sum_{i=1}^n \sum_{L \in \mathbf{L}_i} E_{SHOP}(S_i, L), \quad (1)$$

where the set of  $n$  surfaces is designated by  $\mathbf{S}$  and the set  $\mathbf{L}_i$  is composed by all SHOPS favored labelings of surface  $S_i$ , each of them designated by  $L$ .

Some of the parameters of the smoothness ( $E_{smooth}$ ) and interaction ( $E_{inter}$ ) constraints can be calculated from a prior model [15]. This model is composed of height values from the manual segmentations of a training set.

For multi-surface segmentation, it is necessary to define a graph  $V_i$  as the search space of surface  $S_i$  (Fig. 2). Each node of the graph will be represented by  $V_i(x, y, z)$ , having a one-to-one correspondence to the voxels of the volumetric image  $I(x, y, z)$ . By considering the image  $I$  to be composed by columns perpendicular to the  $xy$  plane, a surface can be defined as a height function  $f_i(x, y)$  that passes exactly once through each column at  $(x, y, f_i(x, y))$ , as exemplified in Fig. 2. The next subsections will explain how each energy term operates, but for implementation details we refer the reader to [14–17].

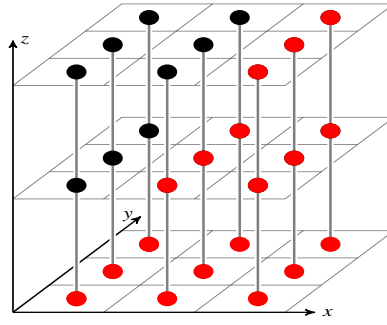


Fig. 2. Representation of the surface  $S_i$  as a height function  $f_i(x, y) = x$  in graph  $V_i$ . Nodes in red represent the closed set and its top nodes of each  $(x, y)$  column represent surface  $S_i$ . The nodes of each  $(x, y)$  column are connected by a gray line. The grid structure corresponds to the volumetric image  $I$ .

### 2.1.1. External boundary energy

$E_{bound}(S_i)$  is the external boundary energy term that pushes each surface  $S_i$  to prominent features of the image. This corresponds to searching the surface with minimum sum of weights  $\gamma_i$ , which are inversely proportional to the probability of the node belonging to the surface. Frequently, these weights depend on the image information (e.g., gradient).

### 2.1.2. Smoothness energy

The smoothness energy term  $E_{smooth}(S_i)$  is composed by hard and soft constraints that regularize surface  $S_i$ .

The hard constraints set the maximum ( $\Delta_x^{max}(x, y)$ ) and minimum ( $\Delta_x^{min}(x, y)$ ) limits of height variation (i.e., variation in the  $z$  axis) from one column to next in the  $x$  direction. The limits can be defined to include confidence interval assuming a Gaussian distribution [15]:

$$\Delta_x^{max}(x, y) = \Delta_x h_\mu^i(x, y) + \beta \cdot \Delta_x h_\sigma^i(x, y), \quad (2)$$

$$\Delta_x^{min}(x, y) = \Delta_x h_\mu^i(x, y) - \beta \cdot \Delta_x h_\sigma^i(x, y), \quad (3)$$

where  $\beta$  controls the confidence level,  $\Delta_x h_\mu^i(x, y)$  and  $\Delta_x h_\sigma^i(x, y)$  are, respectively, the mean and standard deviation of surface's  $S_i$  height derivative observed in the prior model in the  $x$  direction. For the  $y$  direction, analogous limits can be calculated.

The soft constraints impose a linear penalization to height derivatives deviating from a mean height derivative value ( $\Delta_x h_\mu^i$ ) [17]:

$$E_{smooth}(S_i) = \sum_{x \in X, y \in Y} \left( \lambda_{i_x} \left| \frac{\Delta f_i(x, y)}{\Delta x} - \Delta_x h_\mu^i(x, y) \right| \right), \quad (4)$$

where  $\lambda_{i_x}$  is the weight that controls the penalization and  $\Delta f_i(x, y)/\Delta x$  is the height derivative of surface  $S_i$  in the  $x$  direction. For the  $y$  direction, the same approach can be followed.

### 2.1.3. Interaction energy

The interaction energy  $E_{inter}(S_i, S_j)$  is also composed by hard and soft constraints.

The hard constraints limit the maximum ( $\delta_u^{i,j}(x, y)$ ) and minimum distances ( $\delta_l^{i,j}(x, y)$ ) between surfaces  $S_i$  and  $S_j$ . These can be defined to include a confidence interval assuming a Gaussian distribution [15]:

$$\delta_u^{i,j}(x, y) = d_u^{i,j}(x, y) - \beta \cdot d_\sigma^{i,j}(x, y), \quad (5)$$

$$\delta_l^{i,j}(x, y) = d_l^{i,j}(x, y) + \beta \cdot d_\sigma^{i,j}(x, y), \quad (6)$$

where  $\beta$  controls the confidence level;  $d_{\mu}^{i,j}(x,y)$  and  $d_{\sigma}^{i,j}(x,y)$  refer, respectively, to the mean and standard deviation of the distances between surfaces  $S_i$  and  $S_j$  in column  $(x,y)$  observed in the prior model.

The soft interaction constraints penalize linearly the deviation of the distance between surfaces  $S_i$  and  $S_j$  ( $d^{i,j}$ ) in relation to a mean distance value ( $d_{\mu}^{i,j}$ ) [17]:

$$E_{\text{inter}}(S_i, S_j) = \sum_{x \in X, y \in Y} \frac{\alpha_{ij}}{d_{\sigma}^{i,j}(x,y)} \left| d^{i,j}(x,y) - d_{\mu}^{i,j}(x,y) \right|, \quad (7)$$

where  $\alpha_{ij}$  is the weight that controls the linear penalization.

#### 2.1.4. SHOP energy

Higher order potentials have been used successfully for low level vision problems. Yet, the application in higher level problems (e.g., segmentation) is still limited due to the lack of efficiency in the optimization of higher order potentials [19]. To address this issue, Rother *et al.* [19] proposed a transformation of SHOPS to non-submodular pairwise potentials, which can be efficiently solved through the addition of a single auxiliary variable.

A SHOP is simply a potential  $\psi$  applied to a clique  $c$  that favors a specific labeling  $L$  over all of the others:

$$\psi_c(l_c, L) = \begin{cases} \theta_0, & \text{if } l_c = L \\ \theta_{\max}, & \text{otherwise,} \end{cases} \quad \text{or} \quad \psi_c(l_c, L) = \begin{cases} 0, & \text{if } l_c = L \\ \theta, & \text{otherwise,} \end{cases} \quad (8)$$

if  $\theta = \theta_{\max} - \theta_0$  and  $\theta > 0$ . By superimposing SHOPS, more than one labeling can be favored with the downside of an increase of auxiliary variables. Labelings that differ slightly from  $L$  should have penalties similar to  $\theta$  in practical applications [19]. Consequently, Rother *et al.* [19] proposed a compact representation of SHOPS that penalizes labelings according to their Hamming distance from  $L$ , using a deviation function  $g$ . This results in a different type of potential  $\psi_c^g(l_c)$ :

$$g(l_c, L) = \theta \sum_{v \in c} |w_v| \mathbb{1}_{l_c(v) \neq L(v)}, \quad (9)$$

$$\psi_c^g(l_c, L) = \min\{g(l_c, L), \theta\}, \quad (10)$$

where  $w_v$  controls the cost of vertex  $v \in c$  having a different label in relation to the preferred labeling  $L(v)$ . The deviation function  $g$  assigns a cost proportional to the Hamming distance between  $l_c$  and  $L$ .

The compact representation of SHOPS proposed by Rother *et al.* [18] is defined by  $\theta$  and  $w_v$ , making it unpractical to define coherent penalties for labelings of different sizes (or clique order  $|c|$ ). Thus, we propose a different formulation to parameter  $\theta$  to cope with this limitation, namely, we define:

$$\theta = p|c|, \quad w_v = \begin{cases} 1/|c|, & \text{if } L(v) = 0 \\ -1/|c|, & \text{if } L(v) = 1. \end{cases} \quad (11)$$

With this alteration, SHOPS also present a linear penalization:

$$\psi_c^g(l_c, L) = \sum_{v \in c} p \mathbb{1}_{l_c(v) \neq L(v)}, \quad (12)$$

where  $p$  is the penalty increment for each element  $l_c(v)$  that differs from  $L(v)$  (see example in Fig. 3). Therefore, the deviating costs for labelings of different sizes are standardized and the

maximum penalty remains as  $\theta$ , which is higher for larger  $|c|$ . The term  $E_{SHOP}$  is related to  $\psi_c^g$  as follows:

$$E_{SHOP}(S, L) = \psi_c^g(l_c, L), \quad \text{if } l_c = \Omega(S), \quad (13)$$

where function  $\Omega$  converts surface  $S$  (height values) to a binary labeling and restricts it to the region where the  $\psi_c^g$  is applied, i.e. clique  $c$ . In the binary label, the pixels above the surface have label 0 and the pixels on the surface itself and the pixels below it have label 1. Furthermore, if a column only has pixels with label 1, then the surface is at the top of the image for that column. Thus, the surface is composed by the first pixels with label 1 of each column when scanning downwards in vertical direction. Labelings can also be converted to surfaces using the inverse procedure, thus, applying a SHOP with a preferred labeling  $L$  means encouraging the surface that is encoded in  $L$ .



(a). Labeling  $L$ . (b). HD = 1; Cost = 5. (c). HD = 3; Cost = 15. (d). HD = 9; Cost = 45.

Fig. 3. Example of a SHOP with compact representation. Labeling  $L$  is presented in (a). The Hamming distances (HD) relative to (a) and costs of segmentations (b), (c) and (d) consider that  $p = 5$ , leading to  $\theta = 45$ .

## 2.2. Algorithm

We propose an algorithm that segments IRPEDC and BM in the presence of lesions related to AMD, such as drusen and GA. ILM and the IS-OS are used as auxiliary boundaries.

The result of each step of the algorithm is used to restrict the search spaces of the following ones. This reduces computational complexity and avoids undesired interactions between boundaries. The proposed algorithm is described in more detail in the next subsections.

### 2.2.1. Step 1: image denoising

To attenuate speckle noise, a median filter was applied to each slice. The vertical gradients of the denoised images are used as weights of the  $E_{bound}$  energy term (subsection 2.1.1). In a few exceptions, explicitly mentioned in the next subsections, a Gaussian filter was used instead.

### 2.2.2. Step 2: flattening

In this step, a joint segmentation of ILM and IRPEDC is performed. The IRPEDC segmentation is used to flatten the image and restrict the segmentation of the next step.

The joint segmentation only used hard constraints to reduce the computational complexity. The hard constraints (Eqs. 2, 3, 5 and 6) assume a Gaussian distribution for each slice (or  $y$  position), implying that all columns within a frame will have the same hard constraints limits. The hard constraints are defined in this manner because the position of the fovea is not yet known, thus the manual segmentations cannot be aligned to create a meaningful prior model for each  $x$  position. After the joint segmentation, a  $2^{nd}$  order polynomial fit of the segmentation of IRPEDC is used to flatten each slice. The flattening is accomplished by moving columns up and down until the fitted boundary is horizontal.

### 2.2.3. Step 3: ILM segmentation

After the flattening, the fovea is detected as the lowest point in the smoothed ILM boundary of the central B-scan. This information is then used to align the manual segmentations and recalculate the prior model.

Dufour *et al.* [17], defined the search space of most of the segmentations using a previously segmented boundary and the hard interaction constraints limits between that boundary and the boundary to segment ( $\delta_u^{i,j}$  and  $\delta_l^{i,j}$  from Eqs. 5 and 6). This algorithm employs the same strategy. The reference for the search space of ILM boundary is the flattened version of IRPEDC from the previous step. This boundary might include segments of the IS-OS boundary, due to the proximity of IRPEDC to IS-OS and the lack of soft constraints in the segmentation of the previous step (subsection 2.2.2). To handle this issue, the top limit of the search space of ILM is calculated using maximum distance between IRPEDC and ILM ( $\delta_l^{IRPEDC,ILM}$ ), while the bottom limit uses the minimum distance between IS-OS and ILM ( $\delta_u^{IS-OS,ILM}$ ), both assuming a Gaussian distribution and a particular confidence interval (controlled by  $\beta$ ).

#### 2.2.4. Step 4: IRPEDC preliminary segmentation

Following the same strategy as in the previous step, the search space for the individual segmentation of BM could be obtained by using ILM as reference and the hard interaction constraints limits between ILM and BM ( $\delta_u^{ILM,BM}$  and  $\delta_l^{ILM,BM}$ ). However, this approach has the limitation of often including in the search space sharp light-to-dark transitions from within the OS (Fig. 1), which would interfere with the segmentation of BM.

Hence, a preliminary joint segmentation of IRPEDC and BM is performed with the aim of acquiring a preliminary segmentation of the IRPEDC. This boundary can be used as the top limit of the search space for the individual segmentation of BM of the next step.

The search space for the joint segmentation of IRPEDC and BM is determined by using ILM as reference in conjunction with the minimum distance between ILM and IRPEDC ( $\delta_u^{ILM,IRPEDC}$ ) and the maximum distance between ILM and BM ( $\delta_l^{ILM,BM}$ ), according to the hard interaction constraints. To avoid interference from the sharp dark-to-light transitions of IS-OS and light-to-dark transitions from inside the OS, the segmentation is carried out in an image obtained by applying a 1D Gaussian filter in the vertical direction.

#### 2.2.5. Step 5: BM segmentation

After the joint segmentation of the previous step, the BM can be segmented individually. The top limit of the search space is the preliminary segmentation of IRPEDC with an unitary downward offset, while the bottom limit is the same as in the previous step.

#### 2.2.6. Step 6: IRPEDC segmentation

The dark-to-light transitions of the IS-OS hinder the segmentation of IRPEDC (Fig. 4 (c)), thus these boundaries need to be segmented in conjunction [17]. The search space is defined by two references: the ILM and the BM, for the upper and lower limits, respectively. The hard constraints limits used are  $\delta_u^{ILM,IS-OS}$  and  $\delta_l^{BM,IRPEDC}$ . This strategy is the same that Dufour *et al.* [17] used for layers between ILM and IS-OS.

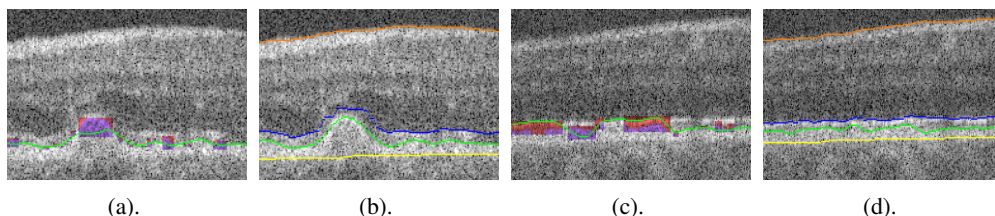


Fig. 4. Segmentation of IRPEDC (green) without SHOPs [(a), (c)] and with SHOPs [(b), (d)]. The SHOPs are presented in (a) and (c); the red and purple pixels represent labels 0 and 1 of the SHOPs, respectively. Boundaries ILM (orange), IS-OS (blue) and BM (yellow) are presented in (b) and (d).



### 2.2.7. Step 7: detection of oversmoothed areas

The IRPEDC boundary is resegmented without soft smoothness constraints and individually, i.e., without interaction constraints of the IS-OS. This segmentation is less restricted to properly segment regions where the OS is thin and large local height variations of the IRPEDC occur, such as drusen regions (Fig. 4 [(a), (b)]) [5, 17]. This was inspired by the detection of pigment epithelial detachments in [20]. The search space for this operation is particularly important because the boundaries can be attracted to the IS-OS (Fig. 4 (c)) or even to random transitions within the RPE layer. The upper bound is defined by using IS-OS with a slight offset. The offset is determined for each column by using the intensity derivative to search for an extremum a few pixels below the IS-OS. The extremum will correspond to either the maximum of the white band of the OS or the minimum of the dark band of the OS. Either way, the top limit will be below the IS-OS and above the IRPEDC. For the lower bound, the brightest pixels of the RPE are segmented by modifying the weights of the  $E_{bound}$  term (section 2.1.1) to be dependent on the intensity of the image after smoothing with a 1D Gaussian filter in the vertical direction, similar to what was done in [7]. No soft constraints were used to avoid oversmoothing of drusen. The lower bound corresponds approximately to the middle of the RPE layer. This limit excludes some transitions inside the RPE related with noise or morphological changes related to AMD, which could hinder the segmentation. Moreover, the lower bound uses intensity information, making it more robust to regions that exhibit low contrast or foci of hyperreflective drusen material.

The difference of IRPEDC segmentation of this and the previous steps results in a binary image (Fig. 5 [(b), (c)]). This image is composed by several objects whose limits are used to define a rectangular neighborhood where the SHOPS are applied, i.e. the clique  $c$  (rectangles in cyan of Fig. 5 (c)). Afterwards, the SHOPS are defined by converting the IRPEDC segmentation of this step to a labeling and restricting it to the rectangular neighborhoods (Fig. 5 (d)).

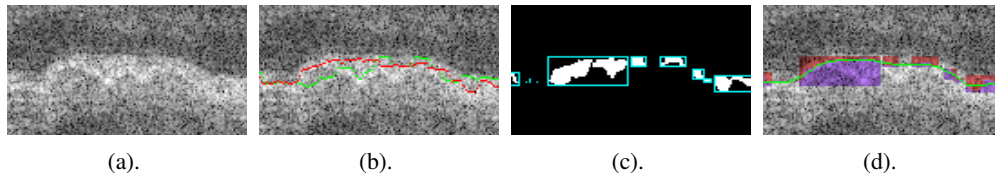


Fig. 5. Detection and application of SHOPS. The original image is presented in (a). In (b), IRPEDC from subsection 2.2.6 is in green, while IRPEDC without soft constraints from subsection 2.2.7 is in red. Image (c) refers to the difference between these two segmentations, which is used to define the cliques for each SHOP (rectangles in cyan). Image (d) shows the IRPEDC segmentation using SHOPS (subsection 2.2.8) after the post-processing in green and the favored labelings ( $L_{IRPEDC}$ ) in red and purple for the labels 0 and 1, respectively.

### 2.2.8. Step 8: IRPEDC resegmentation with SHOPS

Rother *et al.* [19] used SHOPS for a problem of texture restoration, where they applied SHOPS in every pixel with the labelings of the most repeated patterns. Consequently, the computation complexity was considerably increased. In this case, we do not have an image with reoccurring patterns anywhere in the image. Hence, we just need to apply each favored labeling ( $L_{IRPEDC}$ ) to a single location. This implies that the number of SHOPS applied in each image is reduced and that computational complexity is slightly increased relative to the framework without SHOPS.

SHOPS act as a local shape prior that encourages the boundary to follow the set of labelings  $L_{IRPEDC}$  (Fig. 4 and 5). SHOPS are applied in the form of compact representation, which encourages labelings similar to each favored labeling  $L_{IRPEDC}$  with slightly higher penalizations (Fig. 3). Moreover, SHOPS penalties are standardized for different clique sizes, as explained in section 2.1.4.

SHOPS contain non-submodular terms, which are not efficiently solvable by minimum s-t

cuts algorithms. However, the optimization can be performed by QPBO [18], which solves a generalization of the minimum s-t cut problem that includes non-submodular terms. On the downside, QPBO does not guarantee an optimal solution in the presence of non-submodular terms.

Apart from the use of SHOPS, this step is similar to the previous joint segmentation of IS-OS and IRPEDC (section 2.2.6).

### 2.2.9. Step 9: Post-processing

Manual segmentations of boundaries tend to have a slight offset from the actual boundary [21–23]. Therefore, we learned these offsets from the manual segmentations of expert 1 of the training set and correct them when testing. In the end, each boundary is smoothed with a mean filter to approximate the smoothness observed in manual segmentations.

## 3. Experimental methods

### 3.1. Database

The proposed algorithm was evaluated in a publicly available database of 20 patients with intermediate AMD [7]. The database is divided into 4 groups of patients: groups 1 and 2 present drusen, respectively, with good and poor image quality, while patients in groups 3 and 4 exhibit drusen and GA, respectively, with good and poor image quality. These volumes were acquired by 4 SD-OCT imaging systems from Bioptigen, Inc. with a mean pixel size of  $3.19 \times 6.56 \mu\text{m}$  in the axial and lateral directions, respectively. For each volume, 11 slices are available with a mean distance between them that varies from 135 to  $337.5 \mu\text{m}$ . The central B-scan is located on the fovea. The database also includes manual segmentations of the ILM, IRPEDC and BM boundaries performed by 2 expert graders. More detailed information may be found in [7]. To evaluate results in drusen regions, we use a manual detection performed by an ophthalmologist (author L.G.). The grader was asked to mark the limits of each drusen region in the  $x$  direction.

### 3.2. Evaluation

The unsigned error was the metric selected to evaluate the performance of the algorithms for each individual boundary. For evaluation of drusen areas, we computed the overlap ratio (same as in [13]) and the absolute area difference (AAD). The AAD is given by  $AAD = |A_{man} - A_{auto}|$ , where  $A_{man}$  and  $A_{auto}$  are the drusen areas obtained from the manual segmentation and the segmentation of an automated method, respectively. The automatic segmentations were evaluated by considering the manual segmentations of each expert grader and the average of experts' segmentations. The results of each A-scan were considered as a single observation for the errors, implying that the mean and standard deviation of these metrics are relative to all A-scans of all slices. For the drusen area metrics, each drusen was considered as a single observation.

### 3.3. Parameter definition

The hard constraints limits of the  $E_{smooth}$  and  $E_{inter}$  terms are calculated from the prior model once  $\beta$  is defined (Eqs. 2, 3, 5 and 6). Dufour *et al.* [17] set  $\beta$  to be always 2.6. This value results in a confidence interval around 99%, which is reasonable for healthy patients. For our problem, the 1% of the disregarded values are frequently related to abnormalities caused by large drusen. Therefore, we set another confidence interval by defining  $\beta$  to be 3.5, which corresponds to a confidence interval of 99.95%.

The parameters of the soft constraints of  $E_{smooth}$  and  $E_{inter}$  terms were also calculated from the prior model, except for the hyperparameters  $\lambda_{i_x}$  (Eq. (4)) and  $\alpha_{ij}$  (Eq. (7)). These were learned using a Bayesian optimization algorithm [24]. The same procedure was followed to learn the weight  $p$  of the SHOPS defined in Eq. (12).

Soft constraints are often responsible for oversmoothing IS-OS or IRPEDC in drusen regions. However, the proposed algorithm uses soft constraints because there is evidence that they improve segmentation in both healthy and drusen regions [17]. The results are improved possibly because the segmentation is more robust to noise, image artifacts and regions with poor contrast. Moreover, soft constraints can be disabled by the Bayesian optimization algorithm [24], if the segmentation is being impaired. In that case, the Bayesian optimization may set the hyperparameters  $\lambda_{i_x}$  (smoothness) and  $\alpha_{ij}$  (interaction) to zero.

The Bayesian optimization algorithm [24] models the objective function as a Gaussian process. According to a particular criterion, that model provides the best set of parameters for the next iteration. This avoids a blind search of parameters, as it occurs in a grid search. In this case, the criterion is the expected reduction in differential entropy [24].

The results of the proposed algorithm were evaluated using a 5-fold cross validation (4 for training and 1 for testing). The training set was further divided into 4-fold cross validation (3 for training and 1 for validation) so that the Bayesian optimization [24] could be performed. The search limits of the parameters were empirically selected. Nevertheless, it was taken into account that the search space limits should not be too restrictive to avoid excluding the optimal parameters. Furthermore, the maximum number of evaluations for each algorithm step were also defined empirically by observing when the optimization loss function stabilized. The loss function used was the mean error relative to expert 1.

### 3.4. Dufour *et al.* algorithms implementations

We have implemented the algorithms proposed in [17] for healthy retinas and retinas with drusen. These algorithms were used to assess the importance of developing algorithm that takes into account the presence of drusen and GA. For healthy retinas, the implementation has 3 steps: flattening, ILM segmentation, joint segmentation of IS-OS, IRPEDC and BM. In the third step, the original Dufour *et al.* [17] algorithm does not segment the IRPEDC, however we assume that if it did, it would perform a joint segmentation of the 3 boundaries. The algorithm for retinas with drusen has 5 steps: flattening, ILM segmentation, coarse segmentation of IS-OS and BM, BM segmentation, joint segmentation of IS-OS and IRPEDC. The first two steps of both implementations were performed exactly as the proposed algorithm to perform a fair comparison of the IRPEDC and BM boundaries, which are the main focus of this work. The search spaces were defined as in [17] and the confidence intervals of the hard constrains were 99% ( $\beta = 2.6$ ) for all steps [17], except the first two that were performed exactly as the proposed method.

## 4. Results and discussion

In this section, we performed several tests to evaluate the performance of the proposed algorithm. First, we calculated the error for the whole database to obtain an overall performance measure. Afterwards, we evaluated the performance in drusen and non-drusen regions. To evaluate drusen areas measurements, two additional metrics were computed: area difference and overlap ratio.

The evaluations included 4 methods besides the proposed: the proposed method without SHOs, to evaluate the contribution of SHOs, Dufour's methods for healthy retinas and retinas with drusen [17], to evaluate the segmentation of boundaries in a few steps instead of using a joint segmentation, and Chiu's method [7], to perform a fair comparison with a state of the art method in a common public database [7].

We also included the results of a few more tests and some complementary information in [25].

### 4.1. Analysis of overall errors

The proposed method performs similarly to an expert grader for all boundaries, as can be verified by comparing the error of the proposed method for expert 1 and 2 with the intergrader variability (Table 1).

Table 1. Overall mean unsigned error ( $\pm$  standard deviation) in  $\mu m$ . The lowest values of mean and standard deviation for each expert are shown in bold. Statistically significant results greater or lower than those of the proposed method are indicated by  $\uparrow$  and  $\downarrow$ , respectively. The absence of symbols signifies no statistical difference between results. Statistical significance was determined by p-values  $< 0.05$  computed with a two-sided Wilcoxon signed-ranked test (for paired data).

Method	Grader	ILM	IRPEDC	BM
Chiu <i>et al.</i> [7]	1	$3.39 \pm 2.93 \uparrow$	$6.32 \pm 6.91 \downarrow$	$4.19 \pm 4.48 \uparrow$
	2	$3.79 \pm 3.56 \downarrow$	$7.36 \pm 7.41 \uparrow$	$5.35 \pm 4.98 \uparrow$
	avg	$3.05 \pm 2.60 \uparrow$	$5.86 \pm 6.25 \uparrow$	$4.18 \pm 4.12 \uparrow$
Dufour <i>et al.</i> [17]	1	$3.26 \pm 2.85 \uparrow$	$8.33 \pm 10.04 \uparrow$	$4.96 \pm 6.99 \uparrow$
	2	$3.93 \pm 3.54 \uparrow$	$8.60 \pm 10.22 \uparrow$	$6.02 \pm 7.35 \uparrow$
	avg	$2.94 \pm 2.69 \uparrow$	$7.50 \pm 9.64 \uparrow$	$4.79 \pm 6.91 \uparrow$
Dufour <i>et al.</i> [17] for drusen	1	$3.26 \pm 2.85 \uparrow$	$7.24 \pm 7.92 \uparrow$	$4.37 \pm 5.47 \uparrow$
	2	$3.93 \pm 3.54 \uparrow$	$7.61 \pm 8.06 \uparrow$	$5.57 \pm 6.04 \downarrow$
	avg	$2.94 \pm 2.69 \uparrow$	$6.43 \pm 7.25 \uparrow$	$4.32 \pm 5.34 \uparrow$
Proposed without SHOPS	1	$3.26 \pm 2.85$	$7.62 \pm 8.33 \uparrow$	$4.07 \pm 4.22$
	2	$3.93 \pm 3.54$	$8.10 \pm 8.40 \uparrow$	$5.29 \pm 4.89$
	avg	$2.94 \pm 2.69$	$6.90 \pm 7.69 \uparrow$	$4.00 \pm 4.00$
Proposed	1	$3.26 \pm 2.85$	$6.39 \pm 6.56$	$4.07 \pm 4.22$
	2	$3.93 \pm 3.54$	$6.86 \pm 6.68$	$5.29 \pm 4.89$
	avg	$2.94 \pm 2.69$	$5.53 \pm 5.66$	$4.00 \pm 4.00$
Intergrader variability		$4.09 \pm 3.64$	$7.30 \pm 6.87$	$5.03 \pm 4.37$

For the ILM, the proposed algorithm performs statistically significantly better than Chiu's method for expert 1 and the average of experts, but worse for expert 2 (Table 1). Chiu's method does not include any kind of regularization for segmentation of ILM, which could explain the best result relative to expert 2. Our implementations of Dufour methods used the ILM segmented by the proposed method. Consequently, the ILM results of our and Dufour's methods are equal (Table 1). This is done to avoid the interference of the ILM segmentation on the evaluation of the results of the IRPEDC and BM, which are the main focus of this work.

The IRPEDC boundary seems to be the most difficult boundary to segment. This is supported by the largest intergrader variability and algorithms' errors of all boundaries (Table 1). For the IRPEDC, the error of the proposed method is almost always significantly better than Dufour's and Chiu's methods (Table 1). The only exception is for Chiu's method relative to expert 1, in which the error statistically significantly lower. Furthermore, SHOPS improved the segmentation of IRPEDC of the proposed algorithm in a statistically significant manner (Table 1).

As for BM, the proposed method obtains an error statistically significantly lower than all other methods, except when comparing with Dufour's method for drusen and expert 2 is the reference, in which case the proposed algorithm presents a statistically significant higher error (Table 1).

#### 4.2. Analysis of errors in drusen and non-drusen regions

Drusen cause morphological changes in the retina, in particular for the IRPEDC and BM boundaries. Empirically, we observed that those morphological changes hinder the segmentation of both boundaries. Therefore, it is important to analyze the results in drusen and non-drusen regions. These regions were defined with the manual markings of the limits of drusen in the  $x$  direction (performed by author L.G.). The information of Table 2 suggests that IRPEDC and BM are more difficult to segment in drusen regions than in non-drusen regions. This difference is in fact statistically significant for all methods and for the intergrader variability. The statistical

Table 2. Mean unsigned error ( $\pm$  standard deviation) in  $\mu\text{m}$  of drusen regions and non-drusen regions. The lowest values of mean and standard deviation for each expert are shown in bold. Statistically significant results greater or lower than those of the proposed method are indicated by  $\uparrow$  and  $\downarrow$ , respectively. The absence of symbols signifies no statistical difference between results. Statistical significance was determined by p-values  $< 0.05$  computed with a two-sided Wilcoxon signed-ranked test (for paired data). **G** refers to the expert grader.

Method	G	Drusen		Non-drusen	
		IRPEDC	BM	IRPEDC	BM
Chiu <i>et al.</i> [7]	1	9.04 $\pm$ 9.77 $\uparrow$	5.07 $\pm$ 5.16 $\uparrow$	<b>5.71</b> $\pm$ 5.91 $\downarrow$	3.99 $\pm$ 4.29
	2	10.32 $\pm$ 9.84 $\uparrow$	<b>5.89</b> $\pm$ 5.24	6.69 $\pm$ 6.56 $\uparrow$	5.22 $\pm$ 4.92 $\uparrow$
	avg	8.64 $\pm$ 8.94 $\uparrow$	4.85 $\pm$ 4.49 $\uparrow$	5.23 $\pm$ 5.25 $\uparrow$	4.03 $\pm$ 4.01 $\uparrow$
Dufour <i>et al.</i> [17]	1	13.36 $\pm$ 16.02 $\uparrow$	6.01 $\pm$ 8.45 $\uparrow$	7.18 $\pm$ 7.64 $\uparrow$	4.72 $\pm$ 6.59 $\uparrow$
	2	14.30 $\pm$ 15.97 $\uparrow$	7.12 $\pm$ 8.84 $\uparrow$	7.31 $\pm$ 7.83 $\uparrow$	5.77 $\pm$ 6.94 $\uparrow$
	avg	12.85 $\pm$ 15.70 $\uparrow$	5.87 $\pm$ 8.38 $\uparrow$	6.29 $\pm$ 7.08 $\uparrow$	4.55 $\pm$ 6.50 $\uparrow$
Dufour <i>et al.</i> [17] (drusen)	1	9.30 $\pm$ 10.65 $\uparrow$	4.52 $\pm$ 4.69	6.77 $\pm$ 7.07 $\uparrow$	4.33 $\pm$ 5.64 $\uparrow$
	2	10.03 $\pm$ 10.22 $\uparrow$	5.90 $\pm$ 5.23	7.06 $\pm$ 7.37 $\uparrow$	5.49 $\pm$ 6.20 $\downarrow$
	avg	8.57 $\pm$ 9.67 $\uparrow$	4.53 $\pm$ 4.27	5.94 $\pm$ 6.48 $\uparrow$	4.27 $\pm$ 5.55 $\uparrow$
Proposed without SHOPs	1	9.85 $\pm$ 10.72 $\uparrow$	4.52 $\pm$ 4.57	7.11 $\pm$ 7.60 $\uparrow$	3.97 $\pm$ 4.13
	2	10.56 $\pm$ 10.50 $\uparrow$	5.89 $\pm$ 5.14	7.55 $\pm$ 7.74 $\uparrow$	5.15 $\pm$ 4.83
	avg	9.11 $\pm$ 9.91 $\uparrow$	4.51 $\pm$ 4.16	6.40 $\pm$ 6.99 $\uparrow$	3.89 $\pm$ 3.96
Proposed	1	<b>7.95</b> $\pm$ <b>8.75</b>	<b>4.52</b> $\pm$ <b>4.57</b>	6.03 $\pm$ <b>5.89</b>	<b>3.97</b> $\pm$ <b>4.13</b>
	2	<b>8.83</b> $\pm$ <b>8.73</b>	<b>5.89</b> $\pm$ <b>5.14</b>	<b>6.42</b> $\pm$ <b>6.03</b>	<b>5.15</b> $\pm$ <b>4.83</b>
	avg	<b>7.17</b> $\pm$ <b>7.73</b>	<b>4.51</b> $\pm$ <b>4.16</b>	<b>5.16</b> $\pm$ <b>5.00</b>	<b>3.89</b> $\pm$ <b>3.96</b>
Intergrader variability		8.76 $\pm$ 8.13	5.66 $\pm$ 4.72	6.96 $\pm$ 6.51	4.89 $\pm$ 4.27

significance was based on p-values  $< 0.05$  of a two-sided Wilcoxon ranked sum test (for unpaired data).

#### 4.2.1. Analysis of errors in drusen regions

Considering drusen regions, the segmentation of IRPEDC of the proposed method performs statistically significantly better than all other methods (Table 2). As for the BM, the results of the proposed method in drusen regions are significantly better than the other methods or at least statistically indifferent (Table 2). The higher performance results from the algorithm being specifically designed to handle drusen and GA and from the integration of SHOPs.

Designing an algorithm that takes into account the presence of drusen and GA is very important to properly segment IRPEDC and BM in volumes of patients with AMD. This is corroborated by the worse results of Dufour's *et al.* method for healthy retinas compared to their method for drusen or our proposal without SHOPs (Table 2). Dufour's method was developed to segment boundaries in healthy retinas and does not take in account the presence of lesions that cause severe morphological changes. Therefore, this method does not perform well for large drusen or extensive GA regions (Fig. 6). This is a consequence of violations of some implicit assumptions underlying the multi-surface framework. Namely, the framework assumes that boundaries are smooth (smoothness constraints), approximately parallel and the distances between are roughly constant (interactions constraints). Moreover, the multi-surface framework also assumes that the number of boundaries is constant throughout the image. These characteristics are generally observed in healthy retinas [26]. In diseased retinas, the morphological changes caused by lesions frequently violate these assumptions [20]. For instance, large drusen disrupt the assumptions of smoothness, near constant distance and parallelism of IRPEDC and BM. Thus, the joint segmentation of IS-OS, IRPEDC and BM of Dufour's method may cause oversmoothing of the IRPEDC and the BM to follow the ORPE (Fig. 6 (b)) We incorporated this prior knowledge

into the developed algorithm by segmenting IRPEDC and BM independently, to avoid the assumptions of near constant distance and parallelism, and by setting  $\beta = 3.5$ , to allow a wider range of height variations for the smoothness constraint. With this division, the segmentation retains most of the prior knowledge useful for healthy regions and discards the prior information detrimental for lesions, such as the near constant distance between IRPEDC and BM (Fig. 6 (b)). Other authors have also incorporated the prior knowledge of morphological changes related with AMD by segregating segmentation of IS-OS, IRPEDC and BM into a few steps [7, 17, 20]. For instance, in images with drusen, Dufour et al. [17] divided the segmentation into 3 steps (subsection 3.4). The authors justify this alteration mainly because the BM segmentation would follow the ORPE in the cases of large drusen [17]. This is partially explained by the lack of contrast of BM in those regions (Fig. 6 [(a), (b)]), as referred by Dufour et al. [17].

#### 4.2.2. Analysis of errors in non-drusen regions

As for the non-drusen regions, the proposed method attains the lowest errors in most of the cases (Table 2). The proposed algorithm takes into account the presence of GA lesions, which is the main differentiating factor for other methods in these regions.

In the results of Dufour's method for healthy retinas, we observed coarse errors in extensive GA regions (Fig. 6 [(c), (d)]). This is to be expected because this algorithm was developed without considering the presence of lesions [17]. In GA regions, the smoothness, parallelism and near constant distance of the IRPEDC and BM boundaries are maintained (Fig. 6 [(c), (d)]). However, the distance between the boundaries is severely reduced, which in conjunction with the enhancement of the choroid contrast may lead to attraction of the boundaries to transitions in the choroid (Fig. 6 (d)). Furthermore, IS-OS is often missing in GA lesions, which violates the assumption of the multi-surface framework of a constant number of boundaries throughout the image. As a consequence, Dufour's method for healthy retinas segments incorrectly the three boundaries — IRPEDC is segmented in the place of IS-OS and transitions of the choroid are segmented instead of IRPEDC and BM (Fig. 6 (d)). To incorporate the information of the lacking IS-OS, the preliminary segmentation of IRPEDC (subsection 2.2.4) is performed by jointly segmenting just 2 boundaries: IRPEDC and BM. This preliminary segmentation is used for determining the top limit of the search space for independent segmentation of BM. Dufour's method for drusen also performs a preliminary segmentation of BM using 2 boundaries: IS-OS and BM (subsection 3.4). However, the search space for independent segmentation of BM is a band around the preliminary segmentation of BM [17]. We verified that in GA regions, this preliminary segmentation of BM was sometimes attracted to light-to-dark transition in the choroid and the search space would be far from the BM and would not contain it as consequence. This difference may explain the tendency for statistically significant better performance of the proposed method over the Dufour's method for drusen (Table 2).

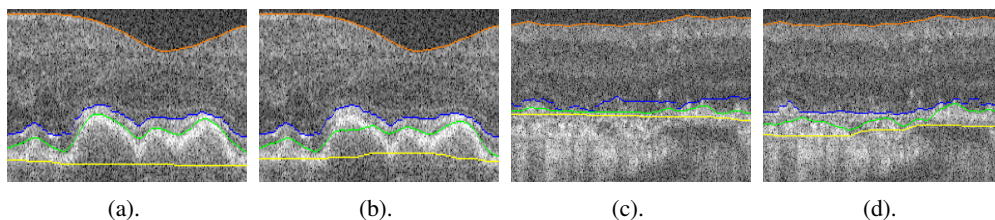


Fig. 6. Results of the proposed method without SHOPs [(a),(c)] and Dufour's method for healthy retinas [(b),(d)]. Images (a) and (b) refer to a case of large drusen, while (c) and (d) to a GA region. Depicted boundaries are: ILM (orange), IS-OS (blue), IRPEDC (green) and BM (yellow).

#### 4.2.3. Analysis of the effect of SHOPs

SHOPs improve the performance for the IRPEDC in both types of regions in a statistically significant manner (Table 2). This supports what is observed empirically: SHOPs frequently correct the oversmoothing of the IRPEDC boundary in drusen regions (Fig. 4 [(a), (b)] and 5). As for non-drusen regions, SHOPs correct IRPEDC when it is attracted to the IS-OS (Fig. 4 [(c), (d)]), because the top limit of the search space for SHOPs definition excludes pixels that compose the transition of IS-OS (subsection 2.2.7). Additionally, some transitions within the RPE are also corrected by the SHOPs. This is mainly related with the lower limit for the SHOPs definition (subsection 2.2.7), which depends on intensity information rather than the gradient, granting robustness to the definition of SHOPs to situations of low IRPEDC contrast or foci of hyperreflective drusen material.

The procedure to compute SHOPs sometimes provides incorrect labelings, which are mainly related with the less constrained segmentation of IRPEDC (subsection 2.2.7). This segmentation is sensitive to the presence of noise, artifacts and other strong image gradients (Fig. 7 (b)). SHOPs can be interpreted as local soft shape prior, which can encourage the segmentation to follow a predefined shape in a defined neighborhood without imposing it (Fig. 7). Hence, incorrectly defined SHOPs are frequently not fully followed to avoid an increase in the cost of the other constraints of the segmentation (Fig. 7 (d)). This behavior grants some immunity to the presence of incorrectly defined SHOPs.

In sum, SHOPs have stronger modeling capabilities than the other constraints of the framework. Nevertheless, SHOPs are also more complex to apply for two reasons. First, they require specific knowledge about the context in which they are applied. For example, the flexibility that SHOPs add to the segmentation may cause an undesired increase of sensitivity to noise or artifacts, so combining them with a constrained segmentation can attenuate this issue. Second, acquiring the favored labelings requires specific knowledge of the problem. If the labelings are not properly defined, they may cause coarse errors that even a constrained segmentation cannot alleviate.

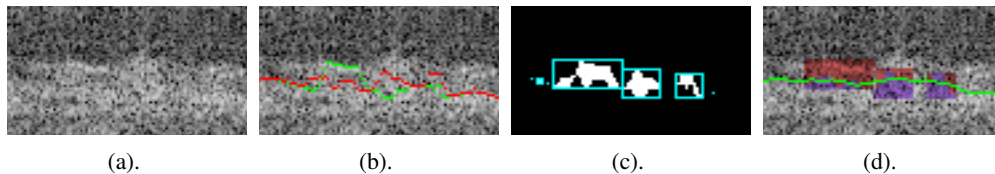


Fig. 7. Detection and application of SHOPS in a case where SHOPS are incorrectly defined. The original image is presented in (a). In (b), IRPEDC from subsection 2.2.6 is in green, while IRPEDC without soft constraints from subsection 2.2.7 is in red. Image (c) refers to the difference between these two segmentations, which is used to define the cliques for each SHOP (rectangles in cyan). Image (d) shows the IRPEDC segmentation using SHOPS (subsection 2.2.8) after the post-processing in green and the favored labelings ( $L_{IRPEDC}$ ) in red and purple for the labels 0 and 1, respectively.

#### 4.3. Analysis of drusen areas measurements

The limits of drusen are the ORPE and BM boundaries. ORPE generally exhibits low contrast when the drusen material is hyporeflexive. Thus, a reliable segmentation is rather difficult. Alternatively, we defined the limits of drusen areas to coincide with those of the RPEDC, namely IRPEDC and BM. The RPEDC is composed by RPE layer, drusen material and other structures related to AMD, such as subretinal drusen, hyperreflective foci and drusen remnants over GA [6]. Folgar *et al.* [6] presented strong evidence suggesting that the quantification of this layer in drusen regions can be used as a biomarker for the progression of AMD. Hereinafter, the area of RPEDC in drusen regions will be simply designated as drusen area.

Drusen areas were evaluated using two metrics: area difference and overlap ratio. The proposed

Table 3. Average absolute area difference in  $\mu m^2$  and overlap ratio of drusen areas. The best values of mean and standard deviation for each expert are shown in bold. Statistically significant results greater or lower than those of the proposed method are indicated by  $\uparrow$  and  $\downarrow$ , respectively. The absence of symbols signifies no statistical difference between results. Statistical significance was determined by p-values  $< 0.05$  computed with a two-sided Wilcoxon signed-ranked test (for paired data).

Method	Grader	Area difference	Overlap Ratio
Chiu <i>et al.</i> [7]	1	1792.9 $\pm$ 2349.3	0.74 $\pm$ 0.14 $\downarrow$
	2	2196.0 $\pm$ 2834.0	0.71 $\pm$ 0.14 $\downarrow$
	avg	1741.0 $\pm$ 2333.0 $\uparrow$	0.75 $\pm$ 0.13 $\downarrow$
Dufour <i>et al.</i> [17] for drusen	1	1820.8 $\pm$ 2294.0 $\uparrow$	0.74 $\pm$ 0.15 $\downarrow$
	2	2261.0 $\pm$ 2817.8 $\uparrow$	0.72 $\pm$ 0.15 $\downarrow$
	avg	1821.8 $\pm$ 2275.1 $\uparrow$	0.76 $\pm$ 0.15 $\downarrow$
Dufour <i>et al.</i> [17]	1	2902.4 $\pm$ 5334.1 $\uparrow$	0.70 $\pm$ 0.18 $\downarrow$
	2	3355.5 $\pm$ 5710.3 $\uparrow$	0.68 $\pm$ 0.19 $\downarrow$
	avg	2942.8 $\pm$ 5421.5 $\uparrow$	0.71 $\pm$ 0.19 $\downarrow$
Proposed without SHOPs	1	1933.6 $\pm$ 2373.2 $\uparrow$	0.73 $\pm$ 0.14 $\downarrow$
	2	2256.4 $\pm$ 2882.6 $\uparrow$	0.71 $\pm$ 0.15 $\downarrow$
	avg	1855.8 $\pm$ 2371.2 $\uparrow$	0.75 $\pm$ 0.14 $\downarrow$
Proposed	1	<b>1602.0 <math>\pm</math> 2134.3</b>	<b>0.77 <math>\pm</math> 0.12</b>
	2	<b>2016.7 <math>\pm</math> 2670.1</b>	<b>0.74 <math>\pm</math> 0.12</b>
	avg	<b>1549.7 <math>\pm</math> 2106.8</b>	<b>0.78 <math>\pm</math> 0.11</b>
Intergrader variability		1917.0 $\pm$ 2367.3	0.75 $\pm$ 0.10

algorithm was developed with the ultimate goal of improving the accuracy of drusen areas measurements. This algorithm performed better than other methods in both metrics in terms of mean values (Table 3). The results of the proposed algorithm are, in fact, statistically significantly better than those of other methods, except when comparing area differences with Chiu's method using expert 1 and 2 as references (Table 3). In this case, the area differences values are statistically indifferent, but the drusen area measurements obtained by the proposed method have a higher degree of overlap with the segmentations of experts.

#### 4.4. Analysis of signed errors

The global results in terms of mean signed error reveal that the boundaries present a slight offset relative to the manual segmentation (Table 4). In particular, the IRPEDC exhibits considerable offsets, which is to be expected because it also is the most difficult boundary to segment. As mentioned in subsection 2.2.9, expert graders have the tendency to segment the boundary with a small bias, which might explain the offsets observed in Table 4. To minimize this issue, the mean offsets were learned from a training set and corrected at test time (all the previous results include offset correction).

## 5. Conclusion

An algorithm for multi-surface segmentation with SHOPs was proposed with the goal of segmenting accurately boundaries that limit drusen, i.e., IRPEDC and BM, in intermediate AMD volumes. The proposed algorithm also segmented ILM and IS-OS, which were used as auxiliary boundaries. The ILM was correctly segmented and it can even be used for the retinal thickness measures. The IS-OS does not have manual segmentation in the database used for evaluation and, as such, its results were not evaluated.

The results demonstrated that the proposed algorithm segmented accurately IRPEDC and BM



Table 4. Mean signed error ( $\pm$  standard deviation) in  $\mu m$  without any offset correction.

Method	Grader	ILM	IRPEDC	BM
<b>Chiu <i>et al.</i></b>	1	$1.87 \pm 4.07$	$-1.61 \pm 5.69$	$-1.39 \pm 5.98$
	2	$1.09 \pm 5.08$	$-3.89 \pm 9.70$	$-1.51 \pm 7.15$
	avg	$1.48 \pm 3.72$	$-2.75 \pm 8.11$	$-1.45 \pm 5.69$
<b>Proposed</b>	1	$-0.66 \pm 4.35$	$-1.38 \pm 9.16$	$-2.88 \pm 5.82$
	2	$-1.44 \pm 5.25$	$-3.65 \pm 9.34$	$-3.00 \pm 7.19$
	avg	$-1.05 \pm 3.98$	$-2.51 \pm 7.86$	$-2.94 \pm 5.63$
<b>Intergrader variability</b>		$-0.78 \pm 5.42$	$-2.27 \pm 9.76$	$-0.12 \pm 6.66$

boundaries and computed trustworthy measurements of drusen areas. Several factors played key roles in achieving accurate results. One of them was the independent segmentation of IRPEDC and BM. By doing so, the proposed algorithm avoids the detrimental effect of interaction constraints in large drusen regions. Another factor was the exclusion of the IS-OS boundary from the preliminary segmentation of IRPEDC, which allowed the correct segmentation of BM in GA areas. The use of SHOPs was also very important for a successful segmentation of the IRPEDC boundary. SHOPs corrected oversmoothing in drusen regions and incorrect segmentations in non-drusen regions of IRPEDC when it was attracted to IS-OS.

In sum, we demonstrated how an algorithm based on a multi-surface framework can be developed to avoid some of the implicit assumptions associated to the framework and accurately segment IRPEDC and BM in volumes of patients of intermediate AMD. Furthermore, we demonstrated how SHOPs can be used to correct some remaining issues of the multi-surface framework, such as IRPEDC oversmoothing in drusen regions.

### Acknowledgments

This work is supported by FCT with the reference project UID/EEA/04436/2013, by FEDER funds through the COMPETE 2020 – Programa Operacional Competitividade e Internacionalização (POCI) with the reference project POCI-01-0145-FEDER-006941. The authors would like to thank the questions and suggestions of the anonymous reviewers that helped to improve this document. J.O. wishes to acknowledge the Portuguese funding institution Fundação Calouste Gulbenkian for the Ph.D. Grant (process 126501) and to ENERMETER - Sistemas de Medição, Lda for their support.