Label Propagation and Higher-order Constraint-Based Segmentation of Fluid-Associated Regions in Retinal SD-OCT Images

Tao Wang¹⁾, Zexuan Ji¹⁾, Qiang Chen¹⁾*, Quansen Sun¹⁾*, Shengchen Yu¹⁾, Wen Fan²⁾, Songtao Yuan²⁾, Qinghuai Liu²⁾

¹⁾ School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, Jiangsu, China

²⁾ Department of Ophthalmology, The First Affiliated Hospital with Nanjing Medical University, Nanjing, Jiangsu, China.

Abstract

The segmentation of the fluid-associated region in the retina plays an important role in the treatment of retinal diseases, such as diabetic retinopathy (DR) and age-related macular degeneration (AMD). The existing methods for the detection of the fluid region generally suffer from the limitation of the segmentation accuracy and the expensive time costs. To overcome these problems, in this paper, we propose an interactive segmentation method for the fluid-associated region in three-dimensional (3-D) spectral domain optical coherence tomography (SD-OCT) retinal imaging, in which only a few seeds in one SD-OCT slice are needed, after which the algorithm finishes the segmentation automatically. To improve the segmentation accuracy, the higher-order constraint is introduced into the conventional Markov random field (MRF) framework to impose the superpixel consistency. To maintain temporal coherence of the 3-D SD-OCT slices, the labeling information is propagated slice by slice via a proposed motion-estimation-based algorithm. The proposed higher-order-based energy function can be efficiently solved by the max flow algorithm on a specified graph with several auxiliary nodes. Experiments on 28 SD-OCT cubes demonstrate the competitiveness of the proposed method compared with the state-of-the-art methods.

Keywords Spectral domain optical coherence tomography, retina segmentation, graph cuts, higher-order information, motion estimation, label propagation.

1. Introduction

Optical coherence tomography (OCT) has become an increasingly important imaging modality, especially for the non-invasive management of eye diseases [31, 53]. OCT produces images with cross-sectional, micron-scale descriptions of the optical reflectance properties of the tissue,

* Corresponding author. chen2qiang@163.com; sunquansen@njust.edu.cn.

enabling clinicians to better understand retinal diseases [18, 23, 46, 55].

Many retinal diseases appear along with the production of symptomatic fluid, such as DR and AMD. These diseases are the primary causes of blindness and vision loss due to the growth of abnormal blood vessels from the choroidal vasculature into the retina, ultimately leading to leakage of fluid. Recently, treatment of these retinal diseases has become available by using anti-angiogenic intravitreal injections to cause the neovascularization to regress, resulting in the resorption of fluid. The amount of intraretinal fluid has become the primary factor for guiding the frequency of these injections, typically by subjective estimation from a limited number of SD-OCT slices [16, 17, 40], which potentially leads to substantial inconsistency in treatment. Therefore, the precise and fast segmentation of the fluid in the retina plays an important role in the treatment of these diseases.

The problem of image segmentation can be described as dividing an image into different regions [27, 34, 35], in which pixels belonging to the same region should have consistent features. Generally, the segmentation methods can be classified into unsupervised [32, 33, 36, 64], semi-supervised [7, 26, 29, 60, 62] and fully supervised [19, 47, 63, 69] approaches. The research on fluid segmentation has received considerable attention in recent years, with the majority of research efforts categorized into two fundamental strategies: slice-to-slice segmentation (2-D) and spatiotemporal analysis (3-D). Methods in the first category segment the SD-OCT slices independently [22, 48, 61, 70]. Fernandez *et al.* [22] delineated the intraretinal and subretinal fluid regions in 2-D SD-OCT B-scans using a deformable model. Quéllec *et al.* [48] identified 2-D fluid region footprints by classifying the local differences between the properties of the normal retinas and the diseased AMD under analysis. Wikins *et al.* [61] detected intraretinal fluids by a combination of thresholding and boundary tracing in individual 2-D B-scans. Zheng *et al.* [70] quantified intraretinal and subretinal fluid in individual 2-D B-scans by clicking on desired candidates produced by coarse-fine segmentations. Due to the lack of associations between neighboring slices, these 2-D-based methods cannot easily produce temporally coherent segmentations whose shape and neighborhood topology evolve smoothly over slices. The second category of approach addresses fluid segmentation as a spatiotemporal problem [5, 13, 63]. Chen *et al.* [13] segmented fluid-associated abnormalities in 3-D SD-OCT image space based on the combination of graph-search and graph-cut methods. This approach yields good segmentation

results for large, well-defined fluid pockets. However, it suffers in smaller fluid regions whose boundaries are more ambiguous and locations are less predictable. Xu *et al.* [63] detected the fluid region in a 3-D SD-OCT volume based on voxel classification by using a layer-dependent stratified sampling strategy. This method can obtain accurate segmentation results regardless of the large diversity of fluid region size, location, and shape. However, it is restricted by the expensive time cost, in which approximately 120 min are required for each SD-OCT volume [63]. Furthermore, these 3-D-based methods rely on the supervised classification technique, which requires a large number of SD-OCT image samples for training.

Due to the influence of low contrast of the retinal image, it is difficult to directly use the unsupervised segmentation methods to obtain satisfactory results. As described above, although the fully supervised segmentation method [63] can produce good results, it relies on a large number of image samples and is limited by the expensive algorithm costs. In this paper, we focus on the semi-supervised segmentation methods, in which the user is allowed to provide a few seeds to represent the label information. The semi-supervised methods are very practical in many applications because the user can add his intention during the segmentation. During recent decades, many interactive image segmentation methods have been proposed, such as graph cuts [7], random walk [20, 26], and shortest path [3]. In graph cut models, the image is represented by a weighted graph. The segmentation problem is then translated to find a minimum cut in the constructed s/t graph via a maximum flow computation [8]. In random walk models, the labels of unseeded pixels are estimated based on the seed propagation on a relationship graph. In shortest path models, the distances of the paths between unseeded and seeded pixels are measured first. One unseeded pixel is then assigned the foreground label if there is a shorter path from this pixel to a foreground seed than to any background seeds. In general, the pixel-level information is utilized by the methods described above, which is too local to capture the rich statistics of image scenes [42]. Furthermore, these conventional interactive methods are also sensitive to the seeds [37]. It is difficult to obtain satisfactory results, and the user must exert more effort when the number of seeds is limited.

Recently, higher-order information has been widely utilized for image segmentation [36, 37, 39, 42, 44, 60] and image classification [54, 66, 67, 68]. Compared with the pixel-level information, the higher-order information has the capability to model complex interactions of

random variables [42], which leads to more informative results. To solve the problems of the conventional interactive image segmentation models, methods based on the higher-order information obtained by the unsupervised segmentation algorithms (superpixels) have been proposed [37, 39]. In the Robust P^n model [39], by using the condition random field in a principled manner, the higher-order constraint is imposed for the segmentation based on the estimation of the number of pixels in the superpixel not belonging to the dominant label. In the nonparametric higher-order model [37], a nonparametric learning method is utilized to estimate the likelihood of pixels and superpixels simultaneously. The relationship between pixels and superpixels is viewed as the higher-order constraint that helps produce better performance than the conventional pixel-level-based methods. However, due to the introduction of more superpixel variables and more interactions between pixels and superpixels, this method has a heavy burden when solving the inversion of an over-large matrix [37].

Consequently, the existing methods for the detection of the fluid region generally suffer from the limitation of the segmentation accuracy and the expensive time costs. The interactive image segmentation methods can produce satisfactory results with a low algorithm burden. However, the process of interacting on each slice for the whole segmentation of the 3-D image is time consuming. To solve the problems described, in this paper, we propose an interactive segmentation method to detect the fluid region in the 3-D SD-OCT image. Only a few seeds on a key slice, which can be automatically selected, are needed, and the algorithm then finishes the whole segmentation automatically. Furthermore, two strategies are introduced in the proposed model to further improve the segmentation accuracy and maintain spatial continuity of the 3-D segmentation. First, the soft higher-order constraint is introduced into the conventional Markov random fields (MRF) model to improve the segmentation accuracy by imposing the label consistency within superpixels. Compared with the existing interactive higher-order-based methods, the proposed higher-order energy is easier to optimize and has a lower algorithm cost. Second, to maintain temporal coherence over neighboring slices, the labeling information is propagated slice by slice via a proposed motion estimation algorithm, which leads to spatially continuous segmentation results with the interactions of neighboring slices.

2. Graph cuts

The graph cuts algorithm can be used to optimize the submodular energy functions. The

energy function for segmentation based on graph cuts can be generally described as follows [6, 7, 28, 51, 59, 71]:

$$E(f) = E_{data}(f) + E_{smooth}(f) \quad (1)$$

where E_{data} is used to measure the disagreement between the labeling f and the observed data, and E_{smooth} is used to measure the extent to which f is not piecewise smooth.

The unary data term E_{data} can be typically defined as a linearity criterion:

$$E_{data}(f) = \sum_{p \in P} D_p(f_p) \quad (2)$$

and

$$D_p(f_p) = -\ln(\Pr(p | f_p))$$

where P is the set of all pixels in the image, f_p represents the label of pixel p , D_p measures the similarity between the observed data of pixel p and label f_p , and $\Pr(p | f_p)$ represents the probability of pixel p belonging to label f_p . If p matches f_p well, the penalty of D_p should be small, and vice versa.

The binary smooth term E_{smooth} can be generally described as follows:

$$E_{smooth}(f) = \sum_{(p,q) \in \aleph} V_{p,q}(f_p, f_q) \quad (3)$$

and

$$V_{p,q}(f_p, f_q) = \omega_{pq} \cdot \delta(f(p) \neq f(q))$$

where $\delta(f(p) \neq f(q)) = \begin{cases} 1 & \text{if } f(p) \neq f(q) \\ 0 & \text{otherwise} \end{cases}$, $\aleph$ is the set of interacting pairs of pixels and

typically consists of adjacent pixels, $V_{p,q}(f_p, f_q)$ is the smoothness penalty function used to describe the prior probability of a particular label configuration with the elements of the clique (p, q) , and ω_{pq} represents the similarity of the two neighboring pixels p and q . A larger value of ω_{pq} represents a higher similarity between the two pixels.

In the graph cuts model, the image is denoted by a weighted graph $G = \langle V, E \rangle$, where V represents the set of vertexes, and E represents the set of edges. V consists of all pixels in the image and two distinguished terminals that represent the two labels. E consists of two types of

edges: one is called $t-links$, which link the pixels and terminals, and the other is called $n-links$, which link the neighboring pixels. The weights of $t-links$ correspond to the unary data term, and $n-links$ correspond to the binary smooth term. The problem of minimizing the energy function is then equivalent to solving the minimum cut problem in graph G . A cut $C \subset E$ is a set of edges such that the terminals are separated in the residual graph $G(C) = (V, E - C)$. In addition, no proper subset of C separates the terminals in $G(C)$, and this can be interpreted as meaning that C is a feasible cut [7]. The cost of cut C equals the sum of its edge weights, and the minimum cut can be efficiently found by a max-flow algorithm [8].

3. Method

To detect the fluid-associated region in the 3-D retinal SD-OCT image, an interactive image segmentation method is proposed in this paper. Fig. 1 shows the sketch map of the 3-D SD-OCT image, which is produced by a fast scan of a SD-OCT device. It can be observed that the 3-D SD-OCT data cube consists of a sequence of spatially continuous gray-level image slices.

Fig. 1. Sketch map of the 3-D SD-OCT image.

Fig. 2 describes the flowchart of the proposed algorithm, which can be generally divided into the following two steps: pre-processing and segmentation. In the pre-processing step, the image denoising and layer segmentation are applied for the input SD-OCT image. In the segmentation phase, one slice is first segmented by the interactive graph cuts algorithm [7]. This labeling information is then propagated slice by slice via motion estimation. Furthermore, the higher-order

information is also utilized to impose the superpixel consistency for each slice.

Fig. 2. Flowchart of the proposed algorithm.

For a clear description, Table 1 illustrates all important notations used in the proposed method.

Table 1 Illustration of all important notations used in the proposed method.

Notation	Definition	Notation	Definition
P	Pixel set	φ	Displacement penalty function
X	Observed image	ξ	Smoothness penalty function
S	Superpixel set	θ	Gaussian parameter set
f	Labeling of the image	V	Pairwise potential function
E	Energy function	ψ	Higher-order clique potential function
ω	Weight of pairwise nodes	$\aleph$	Neighborhood system

3.1 Pre-processing

SD-OCT imaging is based on the interferometric detection of coherent optical beams, and the beams have low temporal coherence and high spatial coherence. Thus, SD-OCT images contain speckle. Speckle in SD-OCT tomograms is dependent on several factors, such as the wavelength of the imaging beam and the structural details of the imaged object [50]. The speckle size may be different in the axial and lateral dimensions, which are mainly determined by the source bandwidth and numerical aperture, respectively. In addition to speckle noise, shot noise also exists in SD-OCT images, which is additive in nature and can be adequately described by the additive white Gaussian noise process [24]. We have previously reported a modified bilateral filtering algorithm [12] for image denoising of the SD-OCT image. Compared with the conventional denoising methods, such as the patch-based methods [11, 21] or the Gaussian filtering, this method is better for edge preservation and has relatively low time complexity. We first adopt this

improved bilateral filtering algorithm [12] for image denoising.

Many layer segmentation techniques have been applied to analyze the abnormal region of the retina, for example, in [13, 63], the 11-layer surface segmentation methods [41, 48] were utilized as the initialization of their fluid detection. Although the proposed algorithm does not depend on the fine variation detection of the retina layer, we are interested in reducing the search space of the fluid region to improve the algorithm efficiency. We utilize our previous method [12] to detect the top and bottom layers of the retina, where the top layer corresponds to the inner limiting membrane (ILM), and the bottom layer corresponds to the retinal pigment epithelium (RPE). Only the region between ILM and RPE is then considered in the segmentation phase. Taking Fig. 3 for example, the left is the original slice, the middle shows the denoising result, and the right shows the layer segmentation result, where the blue and yellow lines represent the ILM and RPE layers, respectively.

Fig. 3. Image denoising and layer segmentation for the SD-OCT slice. From left to right, it shows the original slice, the denoising result, and the layer segmentation result.

3.2 Segmentation

We formulate the fluid segmentation as a pixel labeling problem by assigning each pixel $p \in P$ in SD-OCT image a label $f_p \in \{0,1\}$, where 1 represents the foreground corresponding to the fluid region, and 0 represents the background. The proposed algorithm starts from segmenting a “key-slice”. Fig. 4 shows a comparison of the normal (left) and abnormal (right) SD-OCT slices. It can be observed that the production of symptomatic fluid leads to a significant increase in the retinal thickness. For the obvious detection of the fluid, the “key-slice” is automatically selected as the one with maximum average retinal thickness among all SD-OCT slices. The interactive graph cuts algorithm [7] is then chosen to segment the starting slice, where the user needs to provide a

few seeds for the foreground and background. After obtaining the labeling of the starting slice, we propagate this prior information slice by slice to [all of the other slices](#).

Fig. 4. Comparison of the normal (left) and abnormal (right) SD-OCT slices. The blue and yellow lines represent the ILM and RPE layers, respectively. The red line denotes the [fluid-associated](#) region. The retinal thickness represents the distance between the ILM layer and the RPE layer.

For the [other](#) SD-OCT slices, we [infer](#) the labeling problem based on the Bayesian law:

$$\Pr(f | X, f') \propto \Pr(X | f) \cdot \Pr(f) \cdot \Pr(f' | f) \quad (4)$$

where X represents the observed image, f represents the labeling of the current slice, and f' represents the labeling of its neighboring slice. [Taking the](#) negative logarithm on both sides of Eq. (4), we can [then](#) transform the maximum a posteriori problem to the following energy minimization problem:

$$E(f) = E_{data} + E_{higher-order} + E_{NLC} \quad (5)$$

where E_{data} corresponds to the likelihood term used to measure the disagreement between the labeling f and the observed data, $E_{higher-order}$ corresponds to the prior term used to measure the extent to which f is not piecewise smooth, and E_{NLC} corresponds to the neighboring labeling consistent term used to encourage the spatially continuous constraint for neighboring slices.

1) Data cost function: E_{data} consists of the interactions between pixels and labels. The prior information of each label can be estimated by the seeds in many interactive segmentation methods [4, 6, 7, 26, 28, 51, 58, 59, 65, 71]. [Although](#) these interactive methods can generally obtain satisfactory results, it is time-consuming to interact for [all](#) SD-OCT slices. Recently, many label-flow [propagation-based](#) methods [14, 56, 57] have been proposed, [in which](#) the motion

estimation **between neighboring** images is utilized to propagate the **label flow** from the labeled image to the unlabeled image. Pixels belonging to the foreground/background can transfer the foreground/background prior information to their matching pixels. In this way, seeds of the current unlabeled image can be automatically obtained from its neighboring labeled image.

The popular approaches for the motion estimation include the optimal flow algorithm [10] and the SIFT flow algorithm [45]. Compared with the optimal flow estimation, the SIFT flow algorithm can provide robust matching across different object **appearances while** preserving spatial discontinuity, which allows matching of objects located **in** different parts of the image [57]. However, the cost is expensive for the SIFT flow algorithm to produce the over-accurate matching, which takes **approximately 50 s** for only a 145×105 image [45].

Fig. 5. Motion estimation by a group of local windows. p and p' are two pixels with the same position in the two neighboring slices. Nine windows $\{1, \dots, 9\}$ with the same size $R \times R$ are used to capture the motion.

Inspired by the SIFT flow algorithm [45], under the motivation of **quickly** capturing the limit motion between neighboring SD-OCT slices, in this paper, we propose a novel motion estimation algorithm to propagate the labeling information. **In contrast to** the SIFT flow algorithm, which considers **all possible** motions, we estimate each pixel's motion by a group of local windows shown in Fig. 5, where p and p' are two pixels with the same position in the two neighboring slices, and the nine windows with the same size $R \times R$ are utilized to capture the motion. The task of motion estimation is transformed into **one of** finding a matching window among the nine windows. We assume **that** the nine windows $\{1, \dots, 9\}$ are represented by nine labels $\{1, \dots, 9\}$, then the search is formulated as a discrete labeling partition problem with the following cost function:

$$E(f^{ME}) = \sum_{p \in P} \left| x_p - \bar{x}_{f_p^{ME}} \right| + \lambda \sum_{p \in P} \varphi(f_p^{ME}) + \gamma \sum_{(p,q) \in \mathcal{N}} \xi(f_p^{ME}, f_q^{ME}) \quad (6)$$

where the label $f_p^{ME} \in \{1, \dots, 9\}$ corresponds to the nine windows, x_p denotes the intensity feature at pixel p , $\bar{x}_{f_p^{ME}}$ denotes the mean intensity of the pixels **within window** f_p^{ME} , $\varphi(\cdot)$ is a penalty function used to measure the displacement, and $\xi(\cdot, \cdot)$ is a smoothness function for

restricting neighboring pixel pairs. λ and γ are two parameters used to control the influence of the functions $\varphi(\cdot)$ and $\xi(\cdot, \cdot)$, respectively. The values of $\varphi(\cdot)$ and $\xi(\cdot, \cdot)$ are shown in Fig. 6 and Table 2, respectively.

$\varphi(1)$	$\varphi(2)$	$\varphi(3)$
$\sqrt{2}$	1	$\sqrt{2}$
$\varphi(4)$	$\varphi(5)$	$\varphi(6)$
1	0	1
$\varphi(7)$	$\varphi(8)$	$\varphi(9)$
$\sqrt{2}$	1	$\sqrt{2}$

Fig. 6. Values of the function $\varphi(\cdot)$.

Table 2 Values of the function $\xi(\cdot, \cdot)$.

$\xi(\cdot, \cdot)$	1	2	3	4	5	6	7	8	9
1	0	1	2	1	$\sqrt{2}$	$\sqrt{5}$	2	$\sqrt{5}$	$\sqrt{8}$
2	1	0	1	$\sqrt{2}$	1	$\sqrt{2}$	$\sqrt{5}$	2	$\sqrt{5}$
3	2	1	0	$\sqrt{5}$	$\sqrt{2}$	1	$\sqrt{8}$	$\sqrt{5}$	2
4	1	$\sqrt{2}$	$\sqrt{5}$	0	1	2	1	$\sqrt{2}$	$\sqrt{5}$
5	$\sqrt{2}$	1	$\sqrt{2}$	1	0	1	$\sqrt{2}$	1	$\sqrt{2}$
6	$\sqrt{5}$	$\sqrt{2}$	1	2	1	0	$\sqrt{5}$	$\sqrt{2}$	1
7	2	$\sqrt{5}$	$\sqrt{8}$	1	$\sqrt{2}$	$\sqrt{5}$	0	1	2
8	$\sqrt{5}$	2	$\sqrt{5}$	$\sqrt{2}$	1	$\sqrt{2}$	1	0	1
9	$\sqrt{8}$	$\sqrt{5}$	2	$\sqrt{5}$	$\sqrt{2}$	1	2	1	0

The energy function in Eq. (6) can be solved by using the swap moves algorithm [9]. For each labeled pixel, we can approximatively determine its motion estimation as the unlabeled pixel at the center of the matching window. Pixels belonging to the foreground/background then propagate the foreground/background label information to their matching pixels. To avoid possible inaccurate matching in object boundaries, we wipe out these pixel seeds around the boundaries. Generally, any cluster algorithm can be chosen to model the seeds for each label. In this paper, the Gaussian Mixture Model (GMM) is used to learn this information. For each label $l \in \{0, 1\}$, the GMM with K components is utilized on set P_l^s , which consists of pixel seeds with label l . Therefore, we can obtain K Gaussian parameter sets $\theta^l = \{\theta_1^l, \dots, \theta_K^l\}$ and $\theta_k^l = \{\pi_k^l, \mu_k^l, \Sigma_k^l\}$ ($k \in \{1, \dots, K\}$), where π_k^l , μ_k^l and Σ_k^l denote the mixture coefficient, the mean and the covariance for the k th Gaussian component, respectively. The probability of pixel p

belonging to label l can then be calculated by

$$\Pr(p|l) = \max_{k \in \{1, \dots, K\}} \pi_k^l \frac{\exp(-\frac{1}{2} [x_p - \mu_k^l]^T (\Sigma_k^l)^{-1} [x_p - \mu_k^l])}{\sqrt{(2\pi)^{\dim} |\Sigma_k^l|}} \quad (7)$$

where $\dim$ is the dimension of x_p . The values of the probability $\Pr(p|l)$ ranges from 0 to 1 with the constraint $\sum_l \Pr(p|l) = 1$. The weight ω_{pl}^{unary} between pixel p and label l is defined as follows:

$$\omega_{pl}^{unary} = \mu \cdot (1 - \Pr(p|l)) \quad (8)$$

where the parameter $\mu \geq 0$ is used to control the influence of ω_{pl}^{unary} . If the pixel fits the label well, then the weight should be small, and vice versa. The data cost function is defined as follows:

$$E_{data} = \sum_{p \in P} \omega_{pl}^{unary} \quad (9)$$

where P denotes the set of pixels, and $f(p)$ represents the label for pixel p . The pixel likely belongs to the label with small weight.

2) Higher-order clique function: $E_{higher-order}$ consists of the interactions of pairwise pixels and higher-order cliques. The posterior distribution over the labeling of the conditional random field is a Gibbs distribution, and the corresponding Gibbs energy can be denoted as the sum of interactions over all cliques. The conventional methods restrict their attention to MRFs whose clique potentials involve pairs of neighboring pixels [42], which severely restrict their representation power because it is too local to capture rich statistics of image scenes. More recently, solving energies with higher-order cliques has received significant attention. A higher-order clique can be simply denoted by a set of pixels within superpixels. Higher-order clique potentials have the capability to model complex interactions of random variables. Compared with the pairwise model, experiments [36, 37, 39, 42, 44] showed superior results by introducing higher-order cliques.

To further improve the segmentation accuracy, we introduce the higher-order information into the conventional MRF framework. We first define the weight ω_{pq}^{binary} of two pixels p and q as follows:

$$\omega_{pq}^{binary} = \begin{cases} \exp(-\beta \cdot |x_p - x_q|) & q \in \mathbb{N}_p \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

where x_p denotes the intensity feature at pixel p , β is a constant that controls the strength of the weight, and $\mathbb{N}_p$ represents the neighborhood of p . A larger value of ω_{pq}^{binary} represents a higher similarity between the two pixels. The weight $\omega_{pc}^{higher-order}$ of pixel p and superpixel c is defined as follows:

$$\omega_{pc}^{higher-order} = \begin{cases} \eta \cdot \exp(-\beta \cdot |x_p - \bar{x}_c|) & p \in c \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

where $\bar{x}_c$ is the mean intensity of the pixels within superpixel c , and the parameter $\eta \geq 0$ is used to control the influence of $\omega_{pc}^{higher-order}$. The higher-order clique constraint function is defined as follows:

$$E_{higher-order} = \sum_{p,q \in P} V_{p,q} + \sum_{c \in S} \psi_c \quad (12)$$

where P denotes the set of pixels, and S denotes the set of superpixels. $V_{p,q}$ represents the pairwise relationships of pixels and is defined as follows by using a generalized Potts model [9]:

$$V_{p,q} = \omega_{pq}^{binary} \cdot \delta(f(p) \neq f(q)) \quad (13)$$

where $\delta(f(p) \neq f(q)) = \begin{cases} 1 & \text{if } f(p) \neq f(q) \\ 0 & \text{otherwise} \end{cases}$. If the two neighboring pixels have large weights,

then they likely belong to the same label to avoid the high penalty.

ψ_c represents the higher-order clique potential, which is used to encourage the superpixel consistency. We define ψ_c as follows:

$$\psi_c = \min\left(\sum_{p \in \{p | f_p=0\}} \omega_{pc}^{higher-order}, \sum_{p \in \{p | f_p=1\}} \omega_{pc}^{higher-order}\right) \quad (14)$$

It can be noted that if all pixels in the superpixel take the same label, then the penalty of ψ_c is zero. As opposed to higher-order-based algorithms [1, 30, 38, 43, 49, 52], which use the hard label consistency in regions assuming that all pixels constituting a particular superpixel are assigned with the same label, we consider it as a soft label consistent constraint, similar to the Robust Pⁿ model [39]. Pixels with low weights can break free from the superpixel constraint, which can efficiently remove the “bad” influence of the inaccurate superpixels. In contrast to the Robust Pⁿ model, which is based on the quality measurement of the superpixel and the number of pixels in the superpixel that do not belong to the dominant label, we impose the higher-order constraint

based on the sum of weights of all of the pixels within the superpixel that do not take the dominant label. The measurement of the relationship between pixels and superpixels in the proposed method is finer and more straightforward than the integral estimation of the superpixel quality in the Robust Pⁿ model [39]. Furthermore, it is simpler to solve the proposed higher-order energy function.

3) Neighboring labeling consistency function: E_{NLC} consists of the interactions of the neighboring slices. To maintain temporal coherence of the 3-D SD-OCT image, we utilize the proposed motion estimation algorithm to find the matching relationship from the unlabeled slice to its neighboring labeled slice. Each pixel p in the unlabeled slice can obtain a matching pixel p' in the labeled slice with the label $f(p')$. The label $f(p')$ can be regarded as the prior labeling feature for pixel p . Due to the approximate matching of the proposed motion estimation algorithm, there may exist some matching errors in the flow field. In this paper, we utilize a feature-based patch to improve the robustness of the neighboring labeling constraint and define E_{NLC} as follows:

$$E_{NLC} = \sum_{p \in P} \omega_{pp'}^{NLC} \cdot \sum_{n=1}^{r^2} |f_{\phi_n(p)} - f_{\phi_n(p')}| \quad (15)$$

$$\omega_{pp'}^{NLC} = \varepsilon \cdot \exp\left(-\frac{\beta}{r^2} \cdot \sum_{n=1}^{r^2} |x_{\phi_n(p)} - x_{\phi_n(p')}|\right) \quad (16)$$

where p' is the matching pixel of pixel p , $\phi_n(p)$ represents the n th pixel in the image patch centered at p with length r , $\omega_{pp'}^{NLC}$ measures the quality of matching pixel pair (p, p') based on the similarity of their respective image patches, and the parameter $\varepsilon \geq 0$ is used to control the influence of $\omega_{pp'}^{NLC}$. It can be seen that the labeling constraint is strong if the value of $\omega_{pp'}^{NLC}$ is large, and vice versa.

3.3 Optimization

The proposed segmentation model consists of three energy terms: E_{data} in Eq. (9), $E_{higher-order}$ in Eq. (12), and E_{NLC} in Eq. (15). E_{data} consists of unary potentials. $E_{higher-order}$ consists of pairwise potential $V_{p,q}$ and the higher-order clique potential ψ_c . E_{NLC} consists of pseudo higher-order potentials. We can transform ψ_c into a pairwise energy function and transform E_{NLC} into unary potentials. The graph cuts algorithm can then be used to find the globally optimal

solution.

Fig. 7. Construction of the graph for minimizing the higher-order energy function ψ_c in Eq. (14). Nodes $\{p_1, p_2, \dots, p_{|c|}\}$ represent all pixels within superpixel c . Node c represents an auxiliary binary variable that can be regarded as the superpixel. Each pixel $p \in \{p_1, p_2, \dots, p_{|c|}\}$ is linked to auxiliary superpixel node c with the weight $w_{pc}^{higher-order}$. The superpixel node is also linked to the two terminal nodes S and T with $w_{cs}^{unary} = 0$ and $w_{cT}^{unary} = 0$, respectively.

Theorem The higher-order energy ψ_c in Eq. (14) can be transformed into a pairwise energy function by introducing an auxiliary binary variable c as follows (the construction of the graph is shown in Fig. 7):

$$\psi_c = \min \left(\sum_{p \in \{p | f_p = 0\}} w_{pc}^{higher-order}, \sum_{p \in \{p | f_p = 1\}} w_{pc}^{higher-order} \right) = \sum_{p \in c} w_{pc}^{higher-order} \cdot \delta(f(p) \neq f(c)) \quad (17)$$

Proof. We divide the set $\{p | p \in c\}$ into two sets $P_1 = \{p | p \in c, f(p) = 1\}$ and $P_2 = \{p | p \in c, f(p) = 0\}$. We can then obtain

$$\begin{aligned} \sum_{p \in c} w_{pc}^{higher-order} \cdot \delta(f(p) \neq f(c)) \\ = \sum_{p \in P_1} w_{pc}^{higher-order} \cdot \delta(f(p) \neq f(c)) + \sum_{p \in P_2} w_{pc}^{higher-order} \cdot \delta(f(p) \neq f(c)) \end{aligned} \quad (18)$$

Because of the min-cut principle by utilizing the max-flow/min-cut algorithm on the defined graph shown in Fig. 7, we can obtain

$$\sum_{p \in c} w_{pc}^{higher-order} \cdot \delta(f(p) \neq f(c)) = \min \left(\sum_{p \in P_1} w_{pc}^{higher-order}, \sum_{p \in P_2} w_{pc}^{higher-order} \right) = \psi_c \quad (19)$$

Thus, the higher-order energy function ψ_c can be effectively solved by introducing an auxiliary binary variable. ■

We call E_{NLC} pseudo higher-order potentials because the value of $f_{\phi_n(p)}$ is known. For current pixel p , all involved matching pixel pairs can be denoted as $(\phi_n(p), \phi_n(p'))$ where $n \in \{1, \dots, r^2\}$. For each pixel pair, we transform the pseudo higher-order potential into the following

unary term:

$$\begin{cases} \omega_{\phi_n(p)0}^{unary} = \omega_{\phi_n(p)0}^{unary} + \omega_{pp'}^{NLC} & \text{if } f_{\phi_n(p')} = 1 \\ \omega_{\phi_n(p)1}^{unary} = \omega_{\phi_n(p)1}^{unary} + \omega_{pp'}^{NLC} & \text{if } f_{\phi_n(p')} = 0 \end{cases} \quad (20)$$

As described above, **all energy** terms in the proposed segmentation model can be transformed into unary and pairwise submodular functions. The graph cuts algorithm can be used to find the globally optimal labeling **that** corresponds to the final segmentation result.

Consequently, the proposed segmentation algorithm for the detection of the fluid region in the 3-D SD-OCT image can be summarized as follows.

Step 1. Pre-processing

- Step 1.1 Image denoising [12] for each slice.
- Step 1.2 ILM and RPE **layer** segmentation [12] for each slice.

Step 2. Initializing segmentation

- Step 2.1 Determining the starting slice based on the retinal thickness.
- Step 2.2 Segmenting the starting slice by the interactive graph cuts algorithm [7].

Step 3. Constructing **the** segmentation model

- Step 3.1 Propagating the seeds by the motion estimation with Eq. (6).
- Step 3.2 Constructing the data cost energy function E_{data} with Eqs. (7–9).
- Step 3.3 Generating the superpixel by the mean shift algorithm [15].
- Step 3.4 Constructing the higher-order energy function $E_{higher-order}$ with Eqs. (11–14).
- Step 3.5 Collecting the matching pixel pairs between neighboring slices **by motion** estimation with Eq. (6).
- Step 3.6 Constructing the neighboring labeling consistency energy function E_{NLC} with Eqs. (15–16).

Step 4. Optimization

- Step 4.1 Transforming the higher-order energy ψ_c into a pairwise energy function with Eq. (17).
- Step 4.2 Transforming the neighboring labeling consistency energy E_{NLC} into unary terms with Eq. (20).
- Step 4.3 Using the max-flow/min-cut algorithm [8] to obtain the globally optimal solution.

The analysis of the algorithm complexity for a slice is shown in Table 3, where $|P|$ and $|S|$ denote the number of pixels and superpixels, respectively; M represents the number of pixels between the ILM layer and the RPE layer; w represents the size of the Gaussian bilateral filter window; and r is the size of the patch.

Table 3 Computational complexity of the proposed algorithm for a slice, where $|P|$ and $|S|$ denote the number of

pixels and superpixels, respectively; M represents the number of pixels between the ILM layer and the RPE layer; w represents the size of the Gaussian bilateral filter window; and r is the size of the patch.

Algorithm step	Complexity
Image denoising [12]	$O(P \cdot w^2)$
Layer segmentation [12]	$O(P)$
Motion estimation	$O(9M)$
Producing superpixel [15]	$O(P ^2)$
Construction of data cost energy	$O(M)$
Construction of higher-order energy	$O(3M)$
Construction of neighboring consistency energy	$O(M \cdot r^2)$
Optimization [8]	$O((M+ S)^2)$

4. Results

In this study, a set of 3-D SD-OCT volumes ($512 \times 128 \times 1024$ voxels, $6 \times 6 \times 2 \text{ mm}^3$, voxel size $11.7 \times 47.2 \times 2.0 \text{ } \mu\text{m}^3$) obtained from 28 eyes of 15 patients with AMD (19 eyes) and DR (9 eyes) were analyzed. All eyes were scanned with a Cirrus SD-OCT device (Carl Zeiss Meditec, Inc., Dublin, CA) by a single experienced technician who was certified by an image reading center for OCT imaging. The scanned area covered a $6 \times 6 \text{ mm}^2$ area on the retina, and the scanning macular protocol was a $1024 \times 512 \times 128$ raster scan. The study protocol was approved by the Department of Ophthalmology of The First Affiliated Hospital with Nanjing Medical University. For the reference standard, a fellowship-trained retinal specialist manually segmented the fluid regions in the original image slices of each eye. The accuracy in terms of true positive rate (TPR) [2], which indicates the fraction of the correctly detected volume in the reference standard delineation, and false positive rate (FPR) [2], which denotes the fraction of the incorrectly detected volume in the true negative volume that were used as performance indices:

$$\text{TPR} = \frac{|V_{tp}|}{|V_t|} \quad (21)$$

and

$$\text{FPR} = \frac{|V_{fp}|}{|U_d - V_t|}$$

where V_t is defined as the total fluid volume segmented by the human expert. V_{tp} is defined as the subset of V_t that was detected by the algorithm, i.e., $V_{tp} = V_t \cap V_a$, where V_a denotes the total fluid volume segmented by the algorithm. V_{fp} is defined as the volume detected only by the

algorithm, i.e., $V_{fp} = \{V \mid (V \subset V_a) \cap (V \not\subset V_i)\}$. U_d represents the whole scene domain, which in the present case, is the volume between the ILM and RPE.

We compared the performance of our method with that of the conventional pixel-level-based methods graph cuts (GC) [7] and random walk (RW) [26], the nonparametric higher-order method (NHO) [37], graph cuts combined with the proposed higher-order energy (GC-HO), graph cuts combined with the proposed neighboring labeling consistency energy (GC-NLC), the conventional graph search method (GS) [25], the probability constrained graph cuts and graph search method (GS-GC) [13], and the stratified sampling classification method (SSC) [63]. The results of GC, RW, NHO, GC-HO, GC-NLC, and our method are obtained from our implementation, and the others are quoted directly from the best results reported in the literature. For a fair comparison, the same pre-processing step that includes the image denoising and layer segmentation is utilized for all compared methods. Furthermore, the same initial seeds are provided for all methods to segment the same starting slice, and the proposed motion estimation algorithm is utilized for all methods to propagate the seeds for the other slices.

The parameter setting of the proposed algorithm is described as follows: the controlling parameters λ and γ in Eq. (6) are set to 100 and 50, respectively; the size of window R is set to 5, the number of Gaussian components K in Eq. (7) is set to 2, the constant β in Eqs. (10–11) is fixed at 20, and the size of patch r in Eq. (16) is set to 3. The three parameters μ , η , and ε are used to control the influences of the data cost energy in Eq. (8), the higher-order energy in Eq. (11), and the neighboring labeling consistency energy in Eq. (16), respectively. We choose the three parameters based on the cross-validation, where all 28 SD-OCT cubes are divided into four groups, and each group consists of seven SD-OCT cubes. The parameters of each group are determined based on the training of the rest three groups. The best parameters of the other three groups are utilized to test the current group, and the average result of all tested groups is used to evaluate the performance of the proposed method. The algorithm was tested on an Intel Xeon CPU running at 2.0 GHz in MATLAB. For a clear description, Table 4 lists all parameter settings of the compared methods. It can be seen that the constant β is set to 20 for all methods; the number of Gaussian components K for GC, GC-HO, GC-NLC, and the proposed method are all set to 2; the three controlling parameters μ , η , ε for GC, GC-HO, and GC-NLC, respectively, are set based on their best segmentation performance, and the other parameters for

RW and NHO are set according to their original literature [26, 37].

Table 4 Parameter settings of GC [7], GC-HO, GC-NLC, RW [26], and NHO [37].

Method	Parameter	Value
GC	The constant β	20
	The controlling parameter for the data cost term μ	0.2
	The number of Gaussian components K	2
GC-HO	The constant β	20
	The controlling parameter for the data cost term μ	0.5
	The controlling parameter for the higher-order term η	0.5
	The number of Gaussian components K	2
GC-NLC	The constant β	20
	The controlling parameter for the data cost term μ	0.008
	The controlling parameter for the labeling consistency term ε	0.6
	The number of Gaussian components K	2
RW	The constant β	20
	(h_s, h_r) of mean shift algorithm [15]	$\{(10, 7), (10, 10), (10, 15)\}$
NHO	The constant β	20
	The constant λ	10^5
	The controlling parameter for pixel–superpixel term μ	0.002
	The controlling parameter for superpixel–pixel term ε	0.2

4.1 Qualitative Experiments

The pre-processing step is utilized on the SD-OCT image for image denoising and layer segmentation. We first tested the influence of this pre-processing on the segmentation results. Fig. 8 shows the comparison results of the proposed method (Row 1), the proposed method without image denoising (Row 2), and the proposed method without layer segmentation (Row 3). It can be observed that similar results are produced by the three methods, which means that the segmentation is not sensitive to the pre-processing step. However, the image denoising step is important for the layer segmentation [12]. As described above, the main motivation for using layer segmentation is to reduce the search space of the fluid and the algorithm complexity. Furthermore, because the same pre-processing step is utilized for all compared methods, this step does not affect the test of the proposed segmentation model.

Fig. 8. Segmentation results with/without the pre-processing step. Rows 1–3 are the segmentations by applying the proposed method, the proposed method without image denoising, and the proposed method without layer segmentation, respectively.

The neighboring labeling information is propagated based on the proposed motion estimation algorithm. Fig. 9 shows the comparison of the proposed motion estimation algorithm with the SIFT flow algorithm [45] on seed propagation. The first column shows the labeled slice, in which the red line denotes the boundary of the segmentation. The second column shows the neighboring unlabeled slice of the first column, in which the blue and yellow lines represent the ILM and RPE layers, respectively. The third and fourth columns show the propagated foreground seeds (the grey region) from the labeled slice to the unlabeled slice by the SIFT flow algorithm and the proposed motion estimation algorithm, respectively. The white line represents the boundary of the foreground provided by the expert. It can be seen that almost all seeds produced by the proposed algorithm are inside the foreground, similar to the SIFT flow algorithm. The time cost of the proposed motion estimation algorithm is approximately 1.8 s, which is less than that of the SIFT flow algorithm, which takes approximately 120 s for an SD-OCT slice with size 1024×512 . Compared with the SIFT flow algorithm, the proposed motion estimation algorithm is more

practical.

Fig. 9. Comparison of the proposed motion estimation algorithm with the SIFT flow algorithm [45] on seed propagation. The first column shows the labeled slice, in which the red line represents the boundary of the segmentation. The second column shows the neighboring unlabeled slice of the first column, in which the blue and yellow lines represent the ILM and RPE layers, respectively. The third and fourth columns show the propagated foreground seeds (the grey region) from the labeled slice to the unlabeled slice by the SIFT flow and the proposed motion estimation algorithm, respectively. The white line denotes the boundary of the foreground provided by the expert.

To improve the segmentation accuracy, we introduce the higher-order information to promote superpixel consistency. Fig. 10 shows the comparison of the proposed method (Row 1) with GC-NLC (Row 2). It can be seen that it is difficult to obtain satisfactory segmentation without the higher-order constraint. The superpixel consistency is helpful in determining the accurate boundary of the fluid region. To maintain temporal coherence of the 3-D segmentation, we propagate the neighboring labeling information to impose the neighboring labeling consistency. Fig. 11 shows the comparison of the proposed method (Row 1) with GC-HO (Row 2). The tested eight slices are neighborhood relations. Due to the influence of low contrast, it can be seen that many background pixels similar to the foreground are incorrectly segmented without the neighboring labeling constraint. In some cases, even the whole slice is incorrectly segmented as the background due to the lack of effective prior information to distinguish the similar foreground and background. From the result in the first row, it can be seen that all neighboring slices are correctly segmented regardless of the influence of low contrast due to the propagation of the neighboring labeling prior information.

Fig. 10. Segmentations with/without the higher-order constraint. Rows 1 and 2 show the results with/without the higher-order information, respectively.

Fig. 11. Segmentations with/without the neighboring labeling constraint. Rows 1 and 2 show the results with/without the neighboring labeling consistency, respectively.

Fig. 12 shows the comparison of the proposed method with the state-of-the-art interactive image segmentation methods. The first row shows the input SD-OCT slice with seeds for the foreground (red line) and background (green line). Rows 2–5 show the segmentation results applied by GC [7], RW [26], NHO [37], and the proposed method, respectively. It can be seen that it is difficult for the pixel-level-based methods GC and RW to produce satisfactory segmentations. Many background pixels are incorrectly segmented as the foreground by GC, and parts of foreground are incorrectly segmented as the background by RW. From the result in the fourth row, it can be seen that NHO produces better results than the pixel-level-based methods due to the introduction of the superpixel constraint. However, the boundary of the segmentation is not sufficiently accurate, and some background pixels around the boundary are incorrectly segmented as the foreground by NHO. From the results in the last row, it can be observed that the proposed method produces the best performance with accurate object details. Fig. 13 shows a

three-dimensional visualization in three directions of fluid segmentation in the whole SD-OCT volume, in which the first column shows the standard segmentation provided by the expert, and the second to the fifth columns show the results of our method, GC [7], RW [26], and NHO [37], respectively. It can be observed that the proposed method produces the most similar result to the standard result in all three directions.

Fig. 12. Comparison of our method with the state-of-the-art interactive image segmentation methods. Row 1 shows the input SD-OCT slice with seeds for foreground (red line) and background (green line). Rows 2–5 show the segmentation results by applying GC [7], RW [26], NHO [37] and our method, respectively.

Fig. 13. Three-dimensional visualization in three directions of the fluid segmentation in the whole SD-OCT volume. Column 1 shows the standard segmentation provided by the expert. Columns 2–5 show the results of our method, GC [7], RW [26], and NHO [37], respectively.

The proposed method starts by segmenting a “key-slice”. The labeling of this “key-slice” is then utilized to propagate the seeds and labeling constraint to its neighboring slices. The result of the starting slice can affect the construction of the unary energy of its neighboring slices but does not affect the binary and higher-order energies. Fig. 14 shows the influence of the starting slice on the segmentation, in which the first column shows the segmentations of two starting slices with some errors, and the second column to the seventh column show the segmentations of the neighboring slices. It can be seen that satisfactory results can also be produced regardless of the errors in the starting slices. Furthermore, the satisfactory segmentation result of the starting slice can be generally produced by the interactive graph cuts algorithm with a few seeds.

Fig. 14. Influence of the starting slice on the segmentation. The first column shows the segmentations of two starting slices with some errors, and the others are the segmentation results of the neighboring slices.

4.2 Quantitative Experiments

Quantitative experiments were conducted to test the performance of the proposed method. Table 5 summarizes the mean, standard deviation and p -value (with a significance level of 0.05) of TPR and FPR achieved by various methods. The results of GC [7], GC-HO, GC-NLC, RW [26], NHO [37], and the proposed method were obtained from our implementation, and the others were quoted directly from the best results reported in the literature. It can be seen that GC-HO and GC-NLC both produce better results than GC, which proves the effectiveness of the proposed higher-order energy and the neighboring labeling consistency energy. Compared with the pixel-level-based methods GC [7] and RW [26], the higher-order-based methods NHO [37] and the proposed method both achieve better performance. Among these state-of-the-art interactive image segmentation methods, it can be seen that the proposed method obtains the best result with the highest TPR and the lowest FPR. Because the proposed method is based on the graph cuts model, it achieves a significant improvement based on the comparison of the results between GC and the proposed method. By comparing the proposed method with GC-HO and GC-NLC, it can also be observed that the proposed higher-order energy and the neighboring labeling consistency energy both play key roles in the fluid-associated region segmentation. The TPR and FPR values achieved on each SD-OCT image via the six methods are shown in Fig. 15 and Fig. 16, respectively.

Table 5 Mean, standard deviation (Std) and p -value of TPR and FPR for quantitative comparison of the segmentation performance achieved by the proposed method and other methods.

Methods	TPR		FPR	
	Mean $\pm$ Std (%)	p -value	Mean $\pm$ Std (%)	p -value
GC [7]	83.07 $\pm$ 17.4	2.6e-03	4.20 $\pm$ 5.9	4.6e-03
GC-HO	86.11 $\pm$ 13.0	4.5e-03	4.36 $\pm$ 5.9	2.7e-03
GC-NLC	86.06 $\pm$ 15.1	4.2e-03	1.13 $\pm$ 1.1	1.9e-2
RW [26]	79.90 $\pm$ 9.3	1.2e-09	5.12 $\pm$ 4.2	2.3e-06
NHO [37]	83.41 $\pm$ 12.2	8.1e-05	1.68 $\pm$ 2.3	1.4e-2
Proposed	93.66 $\pm$ 3.6	-	0.83 $\pm$ 0.8	-
GS (reported in [13])	82.8 $\pm$ 10.5		3.2 $\pm$ 4.5	
GS-GC (reported in [13])	86.5 $\pm$ 9.5		1.7 $\pm$ 2.3	
SSC (reported in [63])	97.50 $\pm$ 3.4		0.21 $\pm$ 0.17	

Compared with the existing fluid segmentation methods, the proposed method achieves better performance than GS and GS-GC and a slightly worse result than SSC. The run times of GS-GC,

SSC and the proposed method are approximately 25 min for a $200 \times 200 \times 1024$ volume, approximately 120 min for a $512 \times 128 \times 885$ volume, and approximately 22.5 min (19 min for the pre-processing and 3.5 min for the segmentation) for a $512 \times 128 \times 1024$ volume. Thus, comprehensively considering the segmentation accuracy and the time cost, the proposed method is more competitive in segmenting the fluid-associated region in the SD-OCT image.

Fig. 15. TPR values achieved on each SD-OCT image by GC [7], GC-HO, GC-NLC, RW [26], NHO [37], and the proposed method.

Fig. 16. FPR values achieved on each SD-OCT image by GC [7], GC-HO, GC-NLC, RW [26], NHO [37], and the proposed method.

The three parameters μ , η , and ε are used to adjust the influence of the data cost function, the higher-order function and the neighboring labeling consistency function, respectively. Table 6 shows the influence of the three parameters on the results, in which the mean and standard deviation of TPR and FPR of all 28 SD-OCT volumes are used to measure the performance by varying the parameter η from 0 to 5, the parameter μ from 0.004 to 0.4, and the parameter ε from 0 to 6. It can be seen that without the higher-order information

($\eta = 0$) and the neighboring labeling information ($\varepsilon = 0$), neither performance is satisfactory, which proves the effectiveness of the two energy constraints. From the changes of the TPR and FPR values, we can find that the results are insensitive to the parameter μ but slightly sensitive to the other two parameters η and ε . It can also be noticed that the best result is achieved when $\eta = 0.5$, $\mu = 0.04$, and $\varepsilon = 0.6$.

Table 6 Mean $\pm$ standard deviation of TPR and FPR of the proposed method by varying the parameters η (up), μ (middle) and ε (down).

η	0	0.05	0.1	0.3	0.5	1	5
TPR (%)	82 $\pm$ 21	89 $\pm$ 11	90 $\pm$ 11	91 $\pm$ 8	94 $\pm$ 4	92 $\pm$ 11	86 $\pm$ 20
FPR (%)	0.90 $\pm$ 1	0.91 $\pm$ 1	0.90 $\pm$ 1	0.88 $\pm$ 1	0.83 $\pm$ 0.9	0.91 $\pm$ 1	0.95 $\pm$ 1
μ	4e-03	8e-03	0.02	0.04	0.06	0.1	0.4
TPR (%)	92 $\pm$ 6	92 $\pm$ 6	92 $\pm$ 6	94 $\pm$ 4	92 $\pm$ 7	92 $\pm$ 7	90 $\pm$ 8
FPR (%)	0.98 $\pm$ 1	0.92 $\pm$ 1	0.90 $\pm$ 1	0.83 $\pm$ 0.9	0.89 $\pm$ 1	0.86 $\pm$ 1	1.34 $\pm$ 2
ε	0	0.06	0.12	0.3	0.6	1.2	6
TPR (%)	41 $\pm$ 31	81 $\pm$ 23	85 $\pm$ 22	92 $\pm$ 10	94 $\pm$ 4	92 $\pm$ 5	91 $\pm$ 6
FPR (%)	1.26 $\pm$ 4	0.79 $\pm$ 1	0.84 $\pm$ 1	0.89 $\pm$ 1	0.83 $\pm$ 0.9	0.92 $\pm$ 1	0.96 $\pm$ 1

4.3 Statistical Correlation Analysis and Reproducibility Analysis

Fig. 17 shows the linear regression analysis comparing fluid volumes between the proposed algorithm and the expert. It can be seen that the slope of the fitting curve is 0.996, and the correlation coefficient is 0.9933. Compared with the best results of GS-GC [13] (slope 0.871, correlation coefficient 0.9447) and SSC [63] (the slope 0.905, the correlation coefficient 0.9908), the proposed method achieves better performance. Fig. 18 shows the Bland–Altman plot of fluid volume between the proposed algorithm and the expert. The x-axis is the average fluid volume of the proposed segmentation and the expert manual segmentation, and the y-axis is the difference in fluid volume between the proposed segmentation and the expert manual segmentation. Analysis of the Bland–Altman plots reveals that the 100% limits of agreement were [-0.05, 0.35] for the proposed algorithm and the expert, which is better than the result [-0.34, 0.45] of GS-GC [13] and similar to that of SSC [63]. Fig. 19 shows the segmented fluid volume of all SD-OCT images by the proposed algorithm and the expert.

Fig. 17. Linear regression analysis results comparing fluid volumes between the proposed algorithm and the expert.

Fig. 18. Bland-Altman plot of fluid volume between the proposed algorithm and the expert.

Fig. 19. Bar plot of fluid volume for each SD-OCT image.

5. Conclusion

In this paper, an interactive segmentation method for the detection of the fluid-associated region in retinal SD-OCT images has been proposed. The higher-order and neighboring labeling

information are both introduced to improve the efficiency of the segmentation: the higher-order energy imposes the superpixel consistency, and the neighboring labeling consistency energy provides temporal coherence for the 3-D segmentation. The performance of the proposed method is tested on real retinal SD-OCT images, which include 19 eyes with AMD and nine eyes with DR. The results demonstrate that the proposed method is helpful in producing high-quality segmentations for the fluid-associated region of the retina. The proposed higher-order energy helps obtain an accurate boundary of the segmentation, and the proposed neighboring labeling consistency energy helps propagate accurate labeling information over slices to overcome the influence of low contrast in the image. However, there are still some problems with the proposed method. The selection of the starting slice depends on the average retinal thickness between the ILM layer and the RPE layer. However, the variation of the thickness is not always obvious for an image with a small fluid region. Furthermore, the proposed method for seed propagation is limited in cases of discontinuous fluid regions. To solve these problems, accurate detection and analysis of the variation in the retinal multi-layer structure is needed, and this will be our work in the future.

Acknowledgments

This work was supported in part by the Graduate Innovation Project of Jiangsu Province under Grants KYZZ15_0125, in part by the six talent peaks project in Jiangsu Province under Grant 2014-SWXY-024, in part by the fundamental Research Funds for the Central Universities under Grant 30920140111004, in part by the National Science Foundation of China under Grants 61401209 and 61273251, in part by the Natural Science Foundation of Jiangsu Province, China under Grant BK20140790, and in part by China Postdoctoral Science Foundation under Grants 2014T70525 & 2013M531364.

References

- [1] P. Arbeláez, L. Cohen, Constrained image segmentation from hierarchical boundaries, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1-8.
- [2] S. Arya, D. M. Mount, N. S. Netanyahu, R. S. Verman, A. Y. Wu, An optimal algorithm for approximate nearest neighbor searching fixed dimensions, Journal of the ACM 45 (6) (1998) 891-923.
- [3] X. Bai, G. Sapiro, A geodesic framework for fast interactive image and video segmentation and matting, in: Proceedings of IEEE International Conference on Computer Vision, 2007, pp. 1-8.
- [4] A. Blake, C. Rother, M. Brown, P. Perez, P. Torr, Interactive image segmentation using an adaptive GMMRF model, in: Proceedings of European Conference on Computer Vision, 2004, pp. 428-441.
- [5] H. Bogunovic, M. D. Abramoff, L. Zhang, M. Sonka, Prediction of treatment response from retinal OCT in patients with exudative age-related macular degeneration, in: Proceedings of the Ophthalmic Medical Image Analysis, 2014, pp. 129-136.
- [6] Y. Boykov, G. Funka-Lea, Graph cuts and efficient ND image segmentation, International Journal

- of Computer Vision 70 (2) (2006) 109-131.
- [7] Y. Boykov, M. Jolly, Interactive graph cuts for optimal boundary & region segmentation of objects in ND images, in: Proceedings of IEEE International Conference on Computer Vision, 2001, pp. 105-112.
 - [8] Y. Boykov, V. Kolmogorov, An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (9) (2004) 1124-1137.
 - [9] Y. Boykov, O. Veksler, R. Zabih, Fast approximate energy minimization via graph cuts, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (11) (2001) 1222-1239.
 - [10] A. Bruhn, J. Weickert, C. Schnorr, Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods, International Journal of Computer Vision 61 (3) (2005) 211-231.
 - [11] A. Buades, B. Coll, J. M. Morel, A non-local algorithm for image denoising, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2005, pp. 60-65.
 - [12] Q. Chen, T. Leng, L. Zheng, L. Kutzscher, J. Ma, L. Sisternes, D. L. Rubin, Automated drusen segmentation and quantification in SD-OCT images, Medical image analysis 17 (8) (2013) 1058-1072.
 - [13] X. Chen, M. Niemeijer, L. Zhang, K. Lee, M. D. Abramoff, M. Sonka, Three-dimensional segmentation of fluid-associated abnormalities in retinal OCT: probability constrained graph-search-graph-cut, IEEE Transactions on Medical Imaging 31 (8) (2012) 1521-1531.
 - [14] C. Y. Chung, H. H. Chen, Video object extraction via MRF-based contour tracking, IEEE Transactions on Circuits and Systems for Video Technology 20 (1) (2010) 149-155.
 - [15] D. Comaniciu, P. Meer, Mean shift: A robust approach toward feature space analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (5) (2002) 603-619.
 - [16] F. Coscas, G. Coscas, E. Souied, S. Tick, G. Soubrane, Optical coherence tomography identification of occult choroidal neovascularization in age-related macular degeneration, American journal of ophthalmology 144 (4) (2007) 592-599.
 - [17] C. Cukras, Y. D. Wang, C. B. Meyerle, F. Forooghian, E. Y. Chew, W. T. Wong, Optical coherence tomography-based decision making in exudative age-related macular degeneration: comparison of time-vs spectral-domain devices, Eye 24 (5) (2010) 775-783.
 - [18] W. Drexler, U. Morgner, R. K. Ghanta, F. X. Kartner, J. S. Schuman, J. G. Fujimoto, Ultrahigh-resolution ophthalmic optical coherence tomography, Nature medicine 7 (4) (2001) 502-507.
 - [19] B. Du, L. Zhang, A Discriminative Metric Learning Based Anomaly Detection Method, IEEE Transactions on Geoscience & Remote Sensing 52 (11) (2014) 6844-6857.
 - [20] A. Fabijanska, J. Goclawski, The Segmentation of 3D Images Using the Random Walking Technique on a Randomly Created Image Adjacency Graph, IEEE Transactions on Image Processing 24 (2) (2015) 524-537.
 - [21] L. Fang, S. Li, Q. Nie, J. A. Izatt, C. A. Toth, S. Farsiu, Sparsity based denoising of spectral domain optical coherence tomography images, Biomedical optics express 3 (5) (2012) 927-942.
 - [22] D. C. Fernandez, Delineating fluid-filled region boundaries in optical coherence tomography images of the retina, IEEE transactions on medical imaging 24 (8) (2005) 929-945.
 - [23] J. G. Fujimoto, C. Pitris, S. A. Boppart, M. E. Brezinski, Optical coherence tomography: an emerging technology for biomedical imaging and optical biopsy, Neoplasia 2 (1) (2000) 9-25.
 - [24] M. Gargasha, M. W. Jenkins, A. M. Rollins, D. L. Wilson, Denoising and 4D visualization of OCT

- images, *Optics express* 16 (16) (2008) 12313-12333.
- [25] M. K. Garvin, M. D. Abramoff, X. Wu, S. R. Russell, T. L. Burns, M. Sonka, Automated 3-D intraretinal layer segmentation of macular spectral-domain optical coherence tomography images, *IEEE Transactions on Medical Imaging* 28 (9) (2009) 1436-1447.
 - [26] L. Grady, Random walks for image segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28 (11) (2006) 1768-1783.
 - [27] M. Gong, D. Tian, L. Su, L. Jiao, An efficient bi-convex fuzzy variational image segmentation method, *Information Sciences* 293 (2015) 351-369.
 - [28] S. Han, W. Tao, D. Wang, X. Tai, X. Wu, Image segmentation based on GrabCut framework integrating multiscale nonlinear structure tensor, *IEEE Transactions on Image Processing* 18 (10) (2009) 2289-2302.
 - [29] P. He, X. Xu, K. Hu, L. Chen, Semi-supervised clustering via multi-level random walk, *Pattern Recognition* 47 (2) (2014) 820-832.
 - [30] X. He, R. Zemel, D. Ray, Learning and incorporating top-down cues in image segmentation, in: *Proceedings of European Conference on Computer Vision*, 2006, pp. 338-351.
 - [31] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, J. G. Fujimoto, Optical coherence tomography, *Science* 254 (5035) (1991) 1178-1181.
 - [32] Q. Huang, F. Yang, L. Liu, X. Li, Automatic segmentation of breast lesions for interaction in ultrasonic computer-aided diagnosis, *Information Sciences* 314 (2015) 293-310.
 - [33] Z. Ji, J. Liu, G. Cao, Q. Sun, Q. Chen, Robust spatially constrained fuzzy c-means algorithm for brain MR image segmentation, *Pattern Recognition* 47 (7) (2014) 2454-2466.
 - [34] Z. Ji, Y. Xia, Q. Sun, G. Cao, Q. Chen, Active contours driven by local likelihood image fitting energy for image segmentation, *Information Sciences* 301 (2015) 285-304.
 - [35] W. Ju, D. Xiang, B. Zhang, L. Wang, I. Kopriva, X. Chen, Random Walk and Graph Cut for Co-Segmentation of Lung Tumor on PET-CT Images, *IEEE Transactions on Image Processing* 24 (12) (2015) 5854-5867.
 - [36] T. Kim, K. Lee, S. Lee, Learning full pairwise affinities for spectral segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35 (7) (2013) 1690-1703.
 - [37] T. Kim, K. Lee, S. Lee, Nonparametric higher-order learning for interactive segmentation, in: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 3201-3208.
 - [38] P. Kohli, M. Kumar, P. Torr, P3 & beyond: Solving energies with higher order cliques, in: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1-8.
 - [39] P. Kohli, P. Torr, Robust higher order potentials for enforcing label consistency, *International Journal of Computer Vision* 82 (3) (2009) 302-324.
 - [40] G. A. Lalwani, P. J. Rosenfeld, A. E. Fung, S. R. Dubovy, S. Michels, W. Feuer, J. L. Davis, H. W. Flynn, JR, M. Esquiabro, A variable-dosing regimen with intravitreal ranibizumab for neovascular age-related macular degeneration: year 2 of the PrONTO Study, *American journal of ophthalmology* 148 (1) (2009) 43-58.
 - [41] K. M. Lee, Segmentations of the intraretinal surfaces, optic disc and retinal blood vessels in 3D-OCT scans, Ph. D. dissertation, University of Iowa, 2009.
 - [42] J. Li, G. Zeng, R. Gan, H. Zha, L. Wang, Higher-order clique based image segmentation using evolutionary game theory, *Artificial Intelligence Research* 3 (2) (2014).

- [43] Y. Li, J. Sun, C. Tang, H. Shum, Lazy snapping, *ACM Transactions on Graphics* 23 (3) (2004) 303-308.
- [44] Z. Li, X. Wu, S. Chang, Segmentation using superpixels: A bipartite graph partitioning approach, in: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 789-796.
- [45] C. Liu, J. Yuen, A. Torralba, J. Sivic, W. T. Freeman, SIFT flow: Dense correspondence across different scenes, in: *Proceedings of European Conference on Computer Vision*, 2008, pp. 28-42.
- [46] N. Nassif, B. Cense, B. Park, M. Pierce, S. Yun, In vivo high-resolution video-rate spectral-domain optical coherence tomography of the human retina and optic nerve, *Optics Express* 12 (3) (2004) 367-376.
- [47] D. Pathak, P. Krahenbuhl, T. Darrell, [Constrained convolutional neural networks for weakly supervised segmentation](#), in: [Proceedings of Proceedings of IEEE International Conference on Computer Vision](#), 2015, pp. 1796-1804.
- [48] G. Quellec, K. Lee, M. Dolejsi, M. K. Garvin, M. D. Abramoff, M. Sonka, Three-dimensional analysis of retinal layer texture: identification of fluid-filled regions in SD-OCT of the macula, *IEEE Transactions on Medical Imaging* 29 (6) (2010) 1321-1330.
- [49] A. Rabinovich, S. Belongie, T. Lange, J M Buhmann, Model order selection and cue combination for image segmentation, in: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2006, pp. 1130-1137.
- [50] J. Rogowska, M. E. Brezinski, Evaluation of the adaptive speckle suppression filter for coronary optical coherence tomography imaging, *IEEE Transactions on Medical Imaging* 19 (12) (2000) 1261-1266.
- [51] C. Rother, V. Kolmogorov, A. Blake, Grabcut: Interactive foreground extraction using iterated graph cuts, in: *Proceedings of the ACM SIGGRAPH Conference*, 2004, pp. 309-314.
- [52] J. Shen, Y. Du, W. Wang, X. Li, Lazy random walks for superpixel segmentation, *IEEE Transactions on Image Processing* 23 (4) (2014) 1451-1462.
- [53] E. A. Swanson, J. A. Izatt, C. P. Lin, J. S. Schuman, C. A. Puliafito, J. G. Fujimoto, In vivo retinal imaging by optical coherence tomography, *Optics letters* 18 (21) (1993) 1864-1866.
- [54] W. Tao, Y. Zhou, L. Liu, K. Li, K. Sun, Z. Zhang, [Spatial adjacent bag of features with multiple superpixels for object segmentation and classification](#), *Information Sciences* 281 (2014) 373-385.
- [55] D. Thomas, G. Duguid, Optical coherence tomography—a review of the principles and contemporary uses in retinal investigation, *Eye* 18 (6) (2004) 561-570.
- [56] T. Wang, J. Collomosse, Probabilistic motion diffusion of labeling priors for coherent video segmentation, *IEEE Transactions on Multimedia* 14 (2) (2012) 389-400.
- [57] T. Wang, J. Y. Guillemaut, J. Collomosse, Multi-label propagation for coherent video segmentation and artistic stylization, in: *Proceedings of IEEE International Conference on Image Processing*, 2010, pp. 3005-3008.
- [58] T. Wang, Z. Ji, Q. Sun, Q. Chen, S. Han, Image Segmentation based on Weighting Boundary Information via Graph Cut, *Journal of Visual Communication and Image Representation* 33C (2015) 10-19.
- [59] T. Wang, Z. Ji, Q. Sun, S. Han, Combining Pixel-level and Patch-level Information for Segmentation, *Neurocomputing* 158 (2015) 13-25.
- [60] T. Wang, Q. Sun, Z. Ji, Q. Chen, P. Fu, [Multi-layer graph constraints for interactive image segmentation via game theory](#), *Pattern Recognition* (2016).
- [61] G. R. Wilkins, O. M. Houghton, A. L. Oldenburg, Automated segmentation of intraretinal cystoid

- fluid in optical coherence tomography, *IEEE Transactions on Biomedical Engineering* 59 (4) (2012) 1109-1114.
- [62] J. Wu, Y. Zhao, J. Y. Zhu, S. Luo, Z. Tu, MILCut: A Sweeping Line Multiple Instance Learning Paradigm for Interactive Image Segmentation, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp.256-263.
 - [63] X. Xu, K. Lee, L. Zhang, M. Sonka, M. D. Abramoff, Stratified Sampling Voxel Classification for Segmentation of Intraretinal and Subretinal Fluid in Longitudinal Clinical OCT Data, *IEEE Transactions on Medical Imaging* 34 (7) (2015) 1616-1623.
 - [64] S. Yang, Y. Lv, Y. Ren, L. Yang, L. Jiao, Unsupervised images segmentation via incremental dictionary learning based sparse representation, *Information Sciences* 269 (2014) 48-59.
 - [65] W. Yang, J. Cai, J. Zheng, J. Luo, User-friendly interactive image segmentation through unified combinatorial user inputs, *IEEE Transactions on Image Processing* 19 (9) (2010) 2470-2479.
 - [66] J. Yu, Y. Rui, Y. Y. Tang, D. Tao, High-order distance-based multiview stochastic learning in image classification, *IEEE Transactions on Cybernetics* 44 (12) (2014) 2431-2442.
 - [67] J. Yu, D. Tao, J. Li, J. Cheng, Semantic preserving distance metric learning and applications, *Information Sciences* 281 (2014) 674-686.
 - [68] J. Yu, D. Tao, M. Wang, Adaptive Hypergraph Learning and its Application in Image Classification, *IEEE Transactions on Image Processing* 21 (7) (2012) 3262-3272.
 - [69] L. Zhang, Q. Zhang, L. Zhang, D. Tao, X. Huang, B. Du, Ensemble manifold regularized sparse low-rank approximation for multiview feature embedding, *Pattern Recognition* 48 (10) (2015) 3102-3112.
 - [70] Y. Zheng, J. Sahni, C. Campa, A. N. Stangos, A. Raj, S. P. Harding, Computerized assessment of intraretinal and subretinal fluid regions in spectral-domain optical coherence tomography images of the retina, *American journal of ophthalmology* 155 (2) (2013) 277-286.
 - [71] H. Zhou, J. Zheng, L. Wei, Texture aware image segmentation using graph cuts and active contours, *Pattern Recognition* 46 (6) (2012) 1719-1733.