Optic Cup Segmentation for Glaucoma Detection Using Low-Rank Superpixel Representation

Yanwu Xu[1], Lixin Duan[1], Stephen Lin[2], Xiangyu Chen[1],
Damon Wing Kee Wong[1], Tien Yin Wong[3], and Jiang Liu[1]

[1] Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore
[2] Microsoft Research, P.R. China
[3] Department of Ophthalmology, National University of Singapore, Singapore

Abstract. We present an unsupervised approach to segment optic cups in fundus images for glaucoma detection without using any additional training images. Our approach follows the superpixel framework and domain prior recently proposed in [1], where the superpixel classification task is formulated as a low-rank representation (LRR) problem with an efficient closed-form solution. Moreover, we also develop an adaptive strategy for automatically choosing the only parameter in LRR and obtaining the final result for each image. Evaluated on the popular *ORIGA* dataset, the results show that our approach achieves better performance compared with existing techniques.

1 Introduction

Glaucoma is the second leading cause of blindness worldwide [2], and cannot be cured because the damage to the optic nerve is irreversible. Early detection is thus essential for early treatment and preventing the deterioration of vision.

Among the structural image cues studied for glaucoma diagnosis, the vertical optic cup-to-disc ratio CDR is a major consideration of clinicians [3]. However, clinical assessment by manually annotating the cup and disc for each image is labor-intensive, so automatic methods have been proposed to segment the disc and cup in fundus images. Recently, optic cup segmentation has attracted relatively more attention, since the disc segmentation problem is well-studied in the literature.

Various automated optic cup segmentation methods have been presented, by using traditional image segmentation methods such as level-sets [4] and active shape model (ASM) [5], or using a classification method to identify pixels [6] or regions [7] that are part of the cup or rim (the disc area outside the cup). Comparatively, classification methods show better segmentation accuracy than traditional segmentation methods, by learning prior knowledge from additional training samples. Recently, an interesting superpixel classification framework was proposed in [1], in which superpixel labeling is a key module that pre-learns a linear SVM model from additional samples, or learns a discriminative model on the test image itself using a domain prior in which superpixels near the disc center are always in the cup and ones close to the disc boundary belong to the rim (non-cup). Although the latter method is still a supervised method, no additional training samples are needed. An unsupervised similarity based propagation method is also introduced in [1], but it is lacking in accuracy.

P. Golland et al. (Eds.): MICCAI 2014, Part I, LNCS 8673, pp. 788–795, 2014.

In this work, we employ the same framework [1] of superpixel labeling except that we take an *unsupervised* approach. Based on the domain prior together with the low-rank property of the data, we model the labeling problem from another perspective. Specifically, it is formulated as a low-rank representation (LRR) problem, which can be efficiently solved in closed form. In addition, to avoid manual parameter tuning, an adaptive approach is developed based on analyzing the closed-form solution of the proposed LRR problem. Specifically, we firstly determine with some theoretical guidance a small number of parameter candidates that cover the whole range, and then obtain the final label of each superpixel by fusing the labels corresponding to each candidate parameter by majority voting.

2 Adaptive Low-Rank Reconstruction Based Superpixel Clustering

The superpixel labeling framework proposed in [1] has four major steps as follows:

1. Use the SLIC [8] toolbox to divide an input disc into superpixels;
2. Remove superpixels that lie on blood vessels, since they appear similar whether they lie in the cup or rim;
3. Label each superpixel as belonging either to cup or rim, using different methods;
4. Apply post-processing to obtain a unique ellipse representing the output cup boundary.

In this work, we follow the exact same procedure as in [1] except that we propose a new LRR based unsupervised approach for the major step of superpixel labeling.

2.1 Low-Rank Representation Formulation

Superpixel labeling is essentially a binary clustering problem, in which the two classes correspond to the optic cup and rim, respectively. As illustrated in Fig. 1, superpixels within the optic disc appear similar to each other. Moreover, in an ideal situation where superpixels are similar in the same class but are dissimilar between different classes, the superpixels in each disc image would form only two clusters (corresponding to cup and rim respectively).

Formally, we assume that the given data (feature matrix of superpixels) $\mathbf{X} \in \mathbb{R}^{d \times n}$ come from two linearly independent subspaces, where d is the feature dimension and n is the number of superpixels in a given image. Moreover, the data should have the following self-expressiveness property: each data point (*i.e.* a superpixel) $\mathbf{x}_i$ in the given data (*i.e.* feature matrix of all superpixels in a optic disc) $\mathbf{X}$ can be efficiently represented as a linear combination of all data points, *i.e.* $\mathbf{x}_i = \mathbf{X}\mathbf{z}_i$, where $\mathbf{z}_i \in \mathbb{R}^n$ is the vector of coefficients. Under such an assumption, we seek for an optimal low rank representation $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_n]$ of all the given data, and finally obtain the partition of the data points into cup or rim via spectral clustering [9] in a manner similar to [10].

In the case that the given data is noise-free and embedded in linear subspaces, the low rank representation $\mathbf{Z}$ can be obtained by solving the following optimization problem [11]:

$$\min_{\mathbf{Z}} \ rank(\mathbf{Z}), \ \ s.t. \ \mathbf{X} = \mathbf{X}\mathbf{Z}. \tag{1}$$

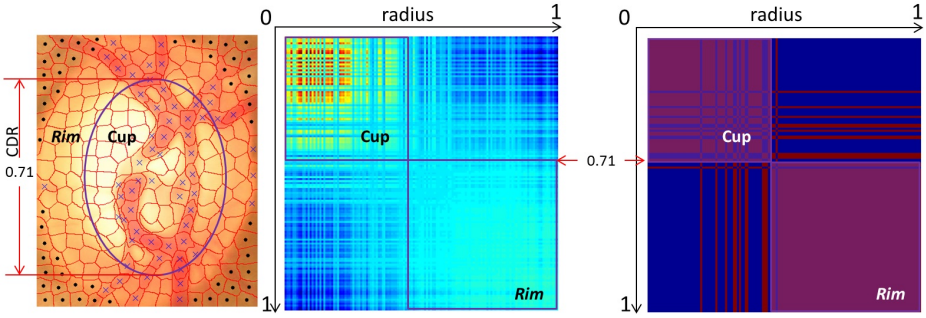

Fig. 1. Illustration of low-rank representation of superpixels in the optic disc. **Left:** 256 superpixels within the optic disc, where black dots mark superpixels outside of the disc region, blue crosses indicate superpixels lying on blood vessels, and superpixels without dots or crosses are to be labelled. **Middle:** The obtained low rank representation $\mathbf{Z}$ exhibits a block-diagonal structure, in which each column/row corresponds to a superpixel (to be classified) sorted in ascending order of distance from its center to the disc center. Warmer colors indicate higher coefficient values. **Right:** Ncut results with $\mathbf{Z}$, in which most of the superpixels in the cup are classified together in one class, while most of the remaining superpixels are labelled as rim. Purple/blue indicates that two superpixels belong to the same/different class.

However, in real-world cases, data is usually not noise-free. Supposing that the data is corrupted, we penalize the reconstruction error matrix $\mathbf{X} - \mathbf{XZ}$ using the squared Frobenius norm ($\|\cdot\|_F^2$), instead of enforcing the equation constraint in (1). Moreover, since the rank function is non-convex and difficult to optimize, we replace it by the nuclear norm, which is a convex approximation of rank. Accordingly, we arrive at the following low-rank representation (LRR) problem:

$$\min_{\mathbf{Z}} \ \lambda\|\mathbf{Z}\|_* + \|\mathbf{X} - \mathbf{XZ}\|_F^2, \tag{2}$$

where λ is a positive trade-off parameter. A nonlinear version of this problem (2) was dealt with in [12]:

$$\min_{\mathbf{Z}} \ \lambda\|\mathbf{Z}\|_* + \|\Phi(\mathbf{X}) - \Phi(\mathbf{X})\mathbf{Z}\|_F^2, \tag{3}$$

where $\Phi(\mathbf{X}) = [\phi(\mathbf{x})_1, \ldots, \phi(\mathbf{x})_n]$, with $\phi(\mathbf{x})_i$ being the nonlinear mapping of $\mathbf{x}_i$, $i = 1, \ldots, n$. Using the kernel trick, we introduce a kernel matrix $\mathbf{K} \in \mathbb{R}^{n \times n}$ for which $\mathbf{K}_{i,j} = \phi(\mathbf{x})_i' \phi(\mathbf{x})_j$, so (3) becomes

$$\min_{\mathbf{Z}} \ \lambda\|\mathbf{Z}\|_* + tr((\mathbf{I} - \mathbf{Z})'\mathbf{K}(\mathbf{I} - \mathbf{Z})), \tag{4}$$

where $tr(\cdot)$ is the trace (i.e. the sum of diagonal elements) of a square matrix. For efficient computation, a linear kernel with a low rank approximation of rank 20 is used in our implementation.

2.2 Solution

A nice property of the problem in (4) is that it can be solved in closed form, which only requires a simple process based on singular value decomposition (SVD) of the kernel matrix, as shown in the following lemma:

Lemma 1. *Suppose that the SVD of the kernel matrix* $\mathbf{K}$ *is* $\mathbf{K} = \mathbf{V}diag(\mathbf{s})\mathbf{V}'$ *, where* $\mathbf{s} = [\sigma_1, \ldots, \sigma_n]'$*, with* $\{\sigma_i\}_{i=1}^n$ *being the singular values sorted in descending order. The problem (4) is minimized by*

$$\mathbf{Z}_\lambda^* = \mathbf{V}diag(\mathbf{d}_\lambda)\mathbf{V}', \tag{5}$$

where the i*-th element of* $\mathbf{d}_\lambda$ *is* $\max(\sigma_i - \lambda, 0)/\sigma_i$.

The proof of this lemma can be found in [12].

2.3 Adaptive Superpixel Clustering

For low-rank representation based methods such as [11], parameter selection is an open problem. In previous works, the reported clustering results are often the one closest to the ground truth after manually tuning the parameters. However, in this work, we propose an adaptive approach to automatically determine candidate values of λ, followed by a majority voting process to obtain the final result based on the clustering results with respect to different candidate values of λ.

Our proposed adaptive approach is based on the following analysis of the closed-form solution in (5). Let $\mathbf{Z}_\lambda^*$ be the optimal solution of problem (4) with the parameter λ, and n_K be the total number of positive singular values of $\mathbf{K}$. We define a set of anchor points $\{\hat{\sigma}_i\}_{i=1}^{n_K+2}$, where $[\hat{\sigma}_1, \ldots, \hat{\sigma}_{n_K+2}] = [\infty, \sigma_1, \ldots, \sigma_n, 0]$. Using these anchor points, we divide the valid range for λ, namely $[0, \infty)$, into $n_K + 1$ intervals $\{\mathcal{S}_i\}_{i=1}^{n_K+1}$ where $\mathcal{S}_i = \{\lambda | \hat{\sigma}_i > \lambda \geq \hat{\sigma}_{i+1}\}$.

Based on the closed-form solution (5), we have the following conclusions: for any $\lambda \in \mathcal{S}_1$, the corresponding optimal solution $\mathbf{Z}_\lambda^*$ is a zero matrix, which provides no useful information for the subsequent clustering process. For any $\lambda \in \mathcal{S}_{n_K+1}$, the corresponding optimal solution is $\mathbf{Z}_\lambda^* = \mathbf{V}_0\mathbf{V}_0'$, where $\mathbf{V}_0 \in \mathbb{R}^{n \times r}$ is obtained by skinny SVD of $\mathbf{K}$, with r being the rank of $\mathbf{K}$. For real-world data which is generally not noise-free, the resultant matrix $\mathbf{Z}_\lambda^*$ may not be desirable for the subsequent clustering process, because it possibly contains too much noise. Therefore, we leave out the λ in the intervals $\mathcal{S}_1$ and $\mathcal{S}_{n_K+1}$.

For each remaining interval $\mathcal{S}_k = \{\lambda | \hat{\sigma}_k > \lambda \geq \hat{\sigma}_{k+1}\}$, where $2 \leq k \leq n_K$, we simply choose $\hat{\sigma}_{k+1}$ as a representative of the values inside this interval. However, under some circumstances, even if λ_1 and λ_2 come from the same interval $\mathcal{S}_{k+1}$, the resultant optimal solutions to (4) may be fairly different from each other. Specifically, we have the following lemma:

Lemma 2. *For any* λ_1, λ_2 *in the same set* $\mathcal{S}_{k+1}$ *where* $1 < k < n_K$*, and the corresponding optimal solutions* $\mathbf{Z}_{\lambda_1}^*, \mathbf{Z}_{\lambda_2}^*$ *to (4), we have*

$$\frac{\|\mathbf{Z}_{\lambda_1}^* - \mathbf{Z}_{\lambda_2}^*\|_F}{\|\mathbf{Z}_{\lambda_1}^*\|_F} \leq \frac{|\sigma_{k+1} - \sigma_k| \sqrt{\sum_{i=1}^k \frac{1}{\sigma_i^2}}}{\sqrt{\sum_{i=1}^k \frac{(\sigma_i - \sigma_k)^2}{\sigma_i^2}}}. \tag{6}$$

Proof. Since λ_1, λ_2 are in the same set $\mathcal{S}_{k+1}$, we have $\hat{\sigma}_{k+1} = \sigma_k > \lambda_i \geq \sigma_{k+1} = \hat{\sigma}_{k+2},\ i = 1, 2$. It can also be easily proved that for any vector $\mathbf{a} \in \mathbb{R}^n$, given that $\mathbf{V}$ is orthogonal, we have $\|\mathbf{V}diag(\mathbf{a})\mathbf{V}'\|_F^2 = \|a\|_F^2 = \sum_{i=1}^n a_i^2$, where a_i is the i-th element of $\mathbf{a}$. Accordingly, considering (5), we have $\|\mathbf{Z}_{\lambda_1}^* - \mathbf{Z}_{\lambda_2}^*\|_F^2 = \|\mathbf{d}_{\lambda_1} - \mathbf{d}_{\lambda_2}\|_F^2 = \sum_{i=1}^k \frac{|\lambda_1 - \lambda_2|^2}{\sigma_i^2}$, and $\|\mathbf{Z}_{\lambda_1}^*\|_F^2 = \sum_{i=1}^k \frac{(\sigma_i - \lambda_1)^2}{\sigma_i^2}$. Therefore, we have

$$\frac{\|\mathbf{Z}_{\lambda_1}^* - \mathbf{Z}_{\lambda_2}^*\|_F}{\|\mathbf{Z}_{\lambda_1}^*\|_F} = \frac{|\lambda_1 - \lambda_2|\sqrt{\sum_{i=1}^k \frac{1}{\sigma_i^2}}}{\sqrt{\sum_{i=1}^k \frac{(\sigma_i - \lambda_1)^2}{\sigma_i^2}}}.$$

Recalling that $\hat{\sigma}_{k+1} = \sigma_k > \lambda_i \geq \sigma_{k+1} = \hat{\sigma}_{k+2}, i = 1, 2$, we have $|\lambda_1 - \lambda_2| \leq |\hat{\sigma}_{k+1} - \hat{\sigma}_k|$ and $\sigma_i - \lambda_1 \geq \sigma_i - \sigma_k,\ i = 1, \ldots, k$, which leads to (6).

Based on Lemma 2, when the bound in (6) (with respect to the interval $\mathcal{S}_{k+1}$) is larger than 10%, we additionally set $0.5(\sigma_{k+1} + \sigma_k)$ as a candidate value of λ. Finally, let $\mathcal{P}$ denote all candidate values for λ.

Given the candidate value set $\mathcal{P}$ for λ, we can efficiently obtain all the corresponding optimal solutions $\{\mathbf{Z}_\lambda^* | \lambda \in \mathcal{P}\}$ by using (5). By performing Ncut [10] based on each $\mathbf{Z}_\lambda^*$, we can obtain a label vector $\mathbf{y}_\lambda \in \{0, 1\}^n$ for all superpixels. Recall that the clustering in our problem is binary, in which the two classes correspond to the cup and rim respectively. For consistent labels $\{\mathbf{y}_\lambda\}$, the class of the central superpixels is assigned as "1" (cup) and the other class as "0" (rim). Therefore, the final clustering label vector $\mathbf{y} \in \{0, 1\}^n$ based on multiple label vectors $\{\mathbf{y}_\lambda\}_{\lambda \in \mathcal{P}}$ can be obtained by majority voting, *i.e.* ,

$$\mathbf{y} = \mathbf{1}\{\frac{1}{|\mathcal{P}|} \sum_{\lambda \in \mathcal{P}} \mathbf{y}_\lambda > 0.5\}, \tag{7}$$

where $\mathbf{1}\{\cdot\}$ is an indicator function, and $|\mathcal{P}|$ is the cardinality of the set $\mathcal{P}$. Therefore, the i-th sample will be assigned to the cup class if the i-th element of $\mathbf{y}$ is equal to 1.

We note that since the candidate values of λ are related to the singular values of $\mathbf{K}$, this scheme *adaptively* selects the candidate values for *each* image.

3 Experiments

To validate the effectiveness of our unsupervised superpixel classification using LRR, the exact same feature representation and experimental settings in [1] are adopted in this work to facilitate comparisons, using the *ORIGA* dataset comprising of 168 glaucoma and 482 normal images. The four labeling methods in [1] are compared, in which only the *pre-learned* method requires a separate training set; for the other three methods, no additional training samples are needed so they can be compared fairly to our proposed approach.

For optic cup segmentation evaluation, we use the same criteria as in [1] and [7], *i.e.*, non-overlap ratio (m_1), absolute area difference (m_2) and absolute CDR error (δ), defined as

Table 1. Performance comparisons on different superpixel labeling methods. Note that our LRR and *refinement only* are unsupervised methods, while the other three are supervised.

Dataset	$S_A \& S_B$			S_C			$S_A \& S_B \& S_C$		
Method	m_1	m_2	δ	m_1	m_2	δ	m_1	m_2	δ
LRR	**0.244**	**0.275**	**0.078**	**0.268**	**0.265**	**0.080**	**0.256**	**0.270**	**0.079**
refinement only	0.331	0.341	0.105	0.325	0.318	0.112	0.328	0.329	0.109
intra-image+refinement	0.265	0.313	0.079	0.269	0.267	0.082	0.267	0.290	0.081
intra-image only	0.269	0.324	0.084	0.277	0.283	0.087	0.273	0.303	0.086
pre-learned	0.277	0.314	0.087	0.301	0.285	0.091	0.289	0.300	0.089

$$m_1 = 1 - \frac{area(E_{dt} \bigcap E_{gt})}{area(E_{dt} \bigcup E_{gt})}, \quad m_2 = \frac{|area(E_{dt}) - area(E_{gt})|}{area(E_{gt})}, \\ \delta = |CDR(E_{dt}) - CDR(E_{dt})|, \tag{8}$$

where E_{dt} denotes a detected cup region, E_{gt} denotes the ground-truth cup region, and $CDR(\cdot)$ calculates the vertical diameter of the cup. Among the three metrics, m_1 is the most commonly used for segmentation.

3.1 Comparison of Different Labeling Methods

Based on the state-of-the-art superpixel framework, using the same domain prior, our proposed LRR method is shown in Table 1 to yield improvements over all the four methods in [1]. The comparison to the unsupervised *refinement only* method shows that our unsupervised method models the problem better with the low-rank property, while the similarity based one does not fully utilize this prior information. The comparison to the other three supervised methods shows that with a proper problem formulation and solution, an unsupervised approach can achieve equal or even better performance than a supervised method, on this specific problem. From some cup segmentation results shown in Fig. 2, one can observe that most failure cases occur when the assumption of LRR does not hold, and some of the failure cases may be caused by the influence of blood vessels that occupy a large area of the cup and are removed early from labeling. This problem will be studied in our future work. For the successful cases that fit the assumption well, the cup regions are very similar with a higher intensity than the rim area.

3.2 Discussion

Validation of Adaptive Parameter Tuning. The results listed in Table 2 are all obtained by using our adaptive clustering scheme, either with or without the use of Lemma 2, to generate additional candidate parameter values. In contrast, the use of additional candidate values achieves better results with an average relative reduction of 1.8%, 2.3% and 2.7% in terms of m_1, m_2 and δ, respectively.

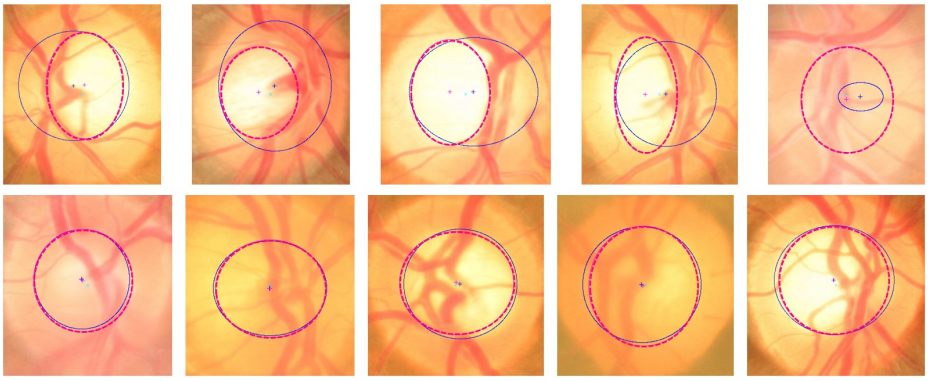

Fig. 2. Some sample results of LRR based cup segmentation with 256 superpixels: failure cases in the top row and successful cases in the bottom row. The cyan $\times$ denotes the center of the image, and the red and blue $+$ denote the center of the detected cup and the ground truth, respectively; correspondingly, the red dashed-line ellipse represents the detected cup and the blue solid-line ellipse represents the ground truth.

Table 2. Performance comparisons of our adaptive clustering scheme using different superpixel numbers

Adaptive scheme	Using Lemma 2			Without using Lemma 2		
# superpixels	m_1	m_2	δ	m_1	m_2	δ
2048	0.265	0.271	0.081	0.270	0.276	0.082
1024	0.262	0.274	0.080	0.267	0.280	0.082
512	0.263	**0.270**	**0.079**	0.269	0.277	0.083
256	**0.256**	**0.270**	**0.079**	0.259	0.278	0.081

Influence of Superpixel Number. Every superpixel labelling method in [1] has two to four parameters to be tuned besides the number of superpixels, which is the only parameter that needs to be pre-chosen manually. The cup segmentation errors with different superpixel numbers are listed in Table 2. From the table, a similar conclusion can be made as in [1], that superpixel based results are relatively stable with a proper superpixel size. However, the lowest m_1 is obtained with the smallest number (256) of superpixels per image, and the best result in [1] is obtained with the largest number (2048) of superpixels. A possible explanation is that when dividing an image into very small-sized superpixels, the information contained in each superpixel becomes less discriminative. In contrast, larger-sized superpixels (e.g., when the number of superpixels is 256) preserve more local structure, which may provide more discriminative power.

4 Conclusion

In this work, we deal with the cup segmentation problem for glaucoma detection. We extended the state-of-the-art superpixel framework [1] by modeling the binary superpixel clustering/classification task as a low-rank representation (LRR) problem in which

the domain prior and the low-rank property of the superpixels are employed. Moreover, we developed an adaptive strategy to automatically choose the candidate values of the only parameter in the LRR formulation, based on a theoretical analysis of the closed-form solution. In relation to prior art [1] [7], our method yields improved unsupervised segmentation. In addition, the proposed LRR-based unsupervised solution performs as well as or better than the supervised methods presented in [1]. The experimental results also demonstrate the effectiveness of the adaptive parameter selection strategy in producing robust results. For future work, we would like to explore the possibility of developing a similar LRR-based formulation for supervised learning. A supervised or semi-supervised LRR-based solution can benefit from utilizing discriminant information from additional training samples, while offering all the advantages of low-rank representations.

References

1. Xu, Y., Liu, J., Lin, S., Xu, D., Cheung, C.Y., Aung, T., Wong, T.Y.: Efficient Optic Cup Detection from Intra-image Learning with Retinal Structure Priors. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 58–65. Springer, Heidelberg (2012)
2. Kingman, S.: Glaucoma is second leading cause of blindness globally. Bull. World Health Organ. 82(11), 887–888 (2004)
3. Jonas, J., Budde, W., Panda-Jonas, S.: Ophthalmoscopic Evaluation of the Optic Nerve Head. Survey of Ophthalmology 43, 293–320 (1999)
4. Liu, J., Wong, D.W.K., Lim, J., Li, H., Tan, N.M., Zhang, Z., Wong, T.Y., Lavanya, R.: Argali:an automatic cup-to-disc ratio measurement system for glaucoma analysis using level-set image processing. In: IEEE Int. Conf. Engin. in Med. and Biol. Soc (2008)
5. Yin, F., Liu, J., Ong, S.H., Sun, D., Wong, D.W.K., Tan, N.M., Baskaran, M., Cheung, C.Y., Aung, T., Wong, T.Y.: Model-based Optic Nerve Head Segmentation on Retinal Fundus Images. In: IEEE Int. Conf. Engin. in Med. and Biol. Soc., pp. 2626–2629 (2011)
6. Wong, D.W.K., Lim, J.H., Tan, N.M., Zhang, Z., Lu, S., Li, H., Teo, M., Chan, K., Wong, T.Y.: Intelligent Fusion of Cup-to-Disc Ratio Determination Methods for Glaucoma Detection in ARGALI. In: Int. Conf. Engin. in Med. and Biol. Soc., pp. 5777–5780 (2009)
7. Xu, Y., Xu, D., Lin, S., Liu, J., Cheng, J., Cheung, C.Y., Aung, T., Wong, T.Y.: Sliding Window and Regression based Cup Detection in Digital Fundus Images for Glaucoma Diagnosis. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 1–8. Springer, Heidelberg (2011)
8. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S.: SLIC Superpixels Compared to State-of-the-art Superpixel Methods. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 34(11), 2274–2282 (2012)
9. Cristianini, N., Shawe-Taylor, J., Kandola, J.S.: Spectral kernel methods for clustering. In: Neural Information Processing Systems Conference, NIPS (2001)
10. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 22, 888–905 (2000)
11. Liu, G., Lin, Z., Yan, S., Sun, J., Yu, Y., Ma, Y.: Robust recovery of subspace structures by low-rank representation. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 35(1), 171–184 (2013)
12. Wang, J., Saligrama, V., Castañón, D.A.: Structural similarity and distance in learning. In: Annual Allerton Conference on Communication, Control, and Computing, Allerton (2011)