Efficient Optic Cup Detection from Intra-image Learning with Retinal Structure Priors⋆

Yanwu Xu[1], Jiang Liu[1], Stephen Lin[2], Dong Xu[3],
Carol Y. Cheung[4], Tin Aung[4,5], and Tien Yin Wong[4,5]

[1] Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore
[2] Microsoft Research Asia, P.R. China
[3] School of Computer Engineering, Nanyang Technological University, Singapore
[4] Singapore Eye Research Institute, Singapore
[5] Department of Ophthalmology, National University of Singapore, Singapore

Abstract. We present a superpixel based learning framework based on retinal structure priors for glaucoma diagnosis. In digital fundus photographs, our method automatically localizes the optic cup, which is the primary image component clinically used for identifying glaucoma. This method provides three major contributions. First, it proposes processing of the fundus images at the superpixel level, which leads to features more descriptive and effective than those employed by pixel-based techniques, while yielding significant computational savings over methods based on sliding windows. Second, the classifier learning process does not rely on pre-labeled training samples, but rather the training samples are extracted from the test image itself using structural priors on relative cup and disc positions. Third, we present a classification refinement scheme that utilizes both structural priors and local context. Tested on the $ORIGA^{-light}$ clinical dataset comprised of 650 images, the proposed method achieves a 26.7% non-overlap ratio with manually-labeled ground-truth and a 0.081 absolute cup-to-disc ratio (CDR) error, a simple yet widely used diagnostic measure. This level of accuracy is comparable to or higher than the state-of-the-art technique [1], with a speedup factor of tens or hundreds.

1 Introduction

Glaucoma is the second leading cause of blindness, with a mean prevalence of 2.4% for all age groups and 4.7% for ages 75 years and above [2]. It is critical to detect this degeneration of the optic nerve as early as possible in order to stall its progression; however, studies suggest that more than 90% of the afflicted are unaware of their condition [3,4]. To facilitate widespread testing, much recent work has focused on computer-assisted glaucoma diagnosis techniques based on inexpensive and widely used digital color fundus images.

Two major image structures used in glaucoma diagnosis are the optic disc, where optical nerve fibers join at the retina, and the optic cup, which is a depression within the optic disc where the fibers exit the retina. The cup and disc boundaries are common features used for identifying glaucoma, via measures such as the cup-to-disc ratio

---

(CDR), defined as the ratio of the vertical cup diameter to the vertical disc diameter [5]. Typically, the CDR value is determined from a manually outlined optic disc and cup. But since manual annotation is labor intensive, researchers have sought automatic methods for disc and cup segmentation.

Research in this area has primarily focused on segmentation of the optic disc, using various techniques such as intensity gradient analysis, Hough transforms, template matching, pixel feature classification [6], vessel geometry analysis, deformable models and level sets [7][8]. In this paper, we address the challenging problem of cup detection [1][9][10], using a large clinical dataset called $ORIGA^{-light}$ [11] in which the ground-truth of discs and cups is marked by a team of graders from a hospital.

Previous cup detection techniques are based either on classifying pixels as part of the cup or rim (the disc area outside the cup) [6][10] or on an analysis of sliding windows [1]. In contrast to these methods, our technique identifies a cup via classification at an intermediate superpixel scale. This allows for richer classification features than pixel-based methods, and faster processing than sliding window techniques without loss of accuracy. Our method also takes advantage of prior information on retinal structure to infer cup/rim training data directly from the test image, without needing a pre-labeled training set. This property is significant in that manual labeling of training sets is avoided, and the training data is specifically suited for the given test image. We furthermore utilize the structural priors, as well as local context, to refine the cup/rim labels. This optic cup detection approach provides accuracy comparable to or better than the current state-of-the-art optic cup detection method [1] with a substantial increase in computation speed. This detection framework indicates much promise for developing practical automated/assisted glaucoma diagnosis systems with low-cost and widespread digital fundus cameras.

2   Intra-image Learning with Retinal Structure Priors

In this work, we start with a disc image which may be obtained using methods such as [7]. Described in the remainder of this section, our method segments the input disc image into superpixels, removes superpixels that correspond to blood vessels, classifies each remaining superpixel to the cup or rim, refines the superpixel classification labels, and then determines a cup location by ellipse fitting.

2.1   Superpixel Segmentation

Superpixels are becoming increasingly popular in computer vision applications because of improved performance over pixel-based methods. In this work, we utilize the state-of-the-art SLIC (Simple Linear Iterative Clustering) algorithm [12] to segment the fundus disc image into compact and nearly uniform superpixels. Unlike dividing an image into a grid of regular patches, superpixels have the important property of preserving local boundaries, as exemplified by the typical segmentation result shown in Fig. 1. This segmentation is processed rapidly, in only 21ms for a GPU implementation or 354ms for a CPU implementation on a 640×480 image [13].
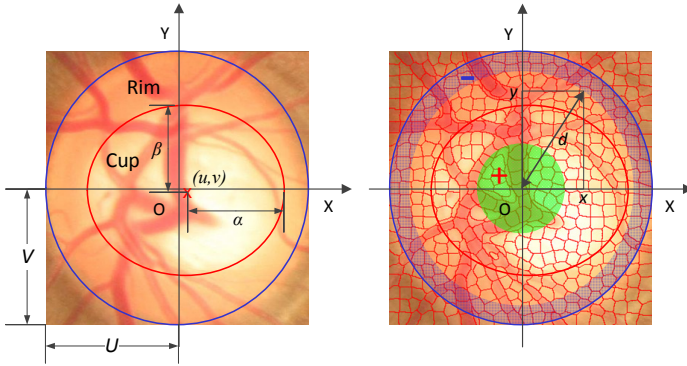
Fig. 1. Illustration of a fundus disc image segmented into 512 superpixels. Left: outline of rim and cup in the original disc image. Right: segmentation into superpixels.
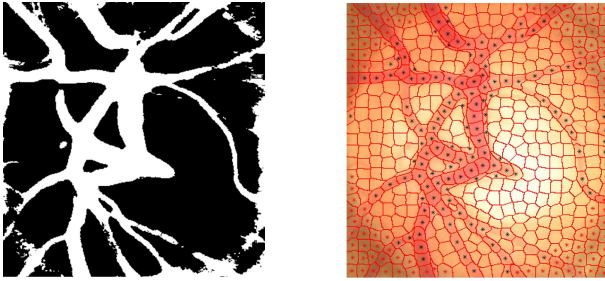

Fig. 2. Blood vessel extraction and removal. Left: blood vessel mask. Right: superpixels on blood vessels (marked with black dots) or out of disc (with red dots).

2.2 Blood Vessel Removal

Since blood vessels appear approximately the same in both the rim and cup regions, many algorithms have been proposed for blood vessel extraction in retina images to avoid their effects on rim/cup classification accuracy. In our case, extraction results need not be very precise, since images are processed at the level of superpixels. So we use the bottom-hat filtering algorithm [14], which trades off some precision for speed, to rapidly generate a rough blood vessel mask, and then identify superpixels that overlap the mask by at least 75% (Fig. 2). These superpixels, as well as those that lie outside the disc, are eliminated from further processing.

2.3 Feature Representation for Superpixels

Various features have been used for modeling superpixels, including shape, location, texture, color, and thumbnail appearance [15]. In our application, only location and color information are relevant for cup/rim classification. For the i-th superpixel, we

extract a feature vector $\mathbf{f}_i$ that consists of position information (denoted by (x_i, y_i, d_i) as shown in Fig. 1), mean RGB colors (r_i, g_i, b_i) and a 256-bin histogram (h_i^r, h_i^g, h_i^b) for each color channel. To avoid magnitude differences among the features, they are each normalized to the range of $[0, 1]$, with L_1-normalization of each histogram.

We note that for classification with pre-labeled training samples from different images, additional normalization would be needed to reduce the influence of illumination change among the images. In this case, r, g and b are divided by average RGB values of the disc image (denoted as $(\bar{r}, \bar{g}, \bar{b})$), and the histograms h^r, h^g and h^b are aligned to $\bar{r}$, $\bar{g}$ and $\bar{b}$, respectively.

2.4   Superpixel Classification

To learn a classifier without pre-labeled samples, we capitalize on prior knowledge of retinal structure. As illustrated in Fig. 1, a superpixel is essentially certain to lie in the rim region if it is very close to the disc outer boundary. On the other hand, we can be assured that a superpixel exists within the cup region if it is very close to the disc center. Based on discussions with professional graders, we have conservatively modeled this structural prior such that superpixels within 1/5 of the disc radius from the disc center (green area in Fig. 1) are considered to be definitely in the cup, while superpixels beyond 9/10 of the disc radius from the center (blue area) are definitely in the rim region.

With this structural prior we obtain cup/rim training samples from *within the test image* for classifier learning. This technique not only avoids manual labeling of training sets, but also generates training data *specific to the given retina*, without requiring any ad hoc image manipulations to conform the test data to the training data (*e.g.* illumination normalization). From these cup/rim samples, we train a classifier to label each superpixel as belonging to the cup or rim. An equal number of positive and negative samples are used for balanced training. For efficiency, the simple linear support vector machine (SVM) classifier is employed, with a weight vector ω trained to estimate the class label l_i (+1 for cup and -1 for rim) of a given superpixel with feature vector $\mathbf{f}_i$, according to $l_i = \omega^T \mathbf{f}_i$. For SVM training, we use the LIBLINEAR toolbox [16].

2.5   Classification Label Refinement

To reduce classification errors, we employ a refinement scheme that accounts for the retinal structure prior as well as local context. To reflect label confidence with respect to the structural prior, we increase/decrease classification values according to how near/far a superpixel is located from the disc center:

$$l_i' = l_i(d - [l_i > 0])^2, \tag{1}$$

where d is the radial distance from the center, and $[l_i > 0] = 1$ when $l_i > 0$, otherwise $[l_i > 0] = 0$.

Contextual information is then introduced by filtering of superpixel labels with respect to feature similarity among superpixels within a certain range (*e.g.*, 1/10 of the disc radius). This yields the final label l'' of a superpixel:

$$l_i'' = l_i' + \frac{1}{N} \sum_{j=1}^{N} l_{i,j}' \cdot s_{i,j}, \tag{2}$$

where $l'_{i,j}$ denotes the label of the j-th neighbor of the i-th superpixel and $s_{i,j}$ denotes the similarity between i-th superpixel and its j-th neighbor, defined as

$$s_{i,j} = e^{-\frac{(f_i - f_{i,j})^2}{2\sigma_f^2}} \tag{3}$$

in which σ_f controls the sensitivity to feature noise.

After obtaining the final labels of all superpixels, the minimum ellipse that encompasses all the superpixels with positive labels is computed to produce the detection result, represented by ellipse center/elongation parameters $(\hat{u}, \hat{v}, \hat{\alpha}, \hat{\beta})$.

3   Experiments

We evaluate our technique through an experimental comparison of four labeling methods based on our framework. The first is our proposed method of intra-image learning with refinement (referred to as *intra-image+refinement*). The second method learns instead from pre-labeled training samples, followed by our refinement scheme (referred to as *pre-learned+refinement*). In the third method (referred to as *refinement only*), we use the retinal structure priors to label definite cup (+1) and rim (-1) superpixels, with the others initialized to 0, and then apply our refinement scheme. The fourth method (called *intra-image only*) excludes the refinement stage from our proposed intra-image learning method. Additionally, we compare our superpixel based approach to state-of-the-art pixel [7][10] and sliding window [1] based methods. We also report how the algorithm parameters affect performance.

3.1   Cup Detection Evaluation Criteria

In this work, we use the same three evaluation criteria as in [1] to measure cup detection accuracy, namely non-overlap ratio (m_1), relative absolute area difference (m_2) and absolute CDR error (δ), defined as

$$m_1 = 1 - \frac{area(E_{dt} \bigcap E_{gt})}{area(E_{dt} \bigcup E_{gt})}, \ m_2 = \frac{|area(E_{dt}) - area(E_{gt})|}{area(E_{gt})}, \ \delta = \frac{|D_{dt} - D_{gt}|}{2} \tag{4}$$

where E_{dt} denotes a detected cup region, E_{gt} denotes the ground-truth ellipse, D_{dt} is the vertical diameter of the detected cup, D_{gt} is the vertical diameter of the ground-truth cup. The vertical diameter of the disc is set to 2, so $0 < D_{dt}, D_{gt} \leq 2$.

3.2   Experimental Setup

The settings in [1] are also adopted in this work to facilitate comparisons. For testing we use the $ORIGA^{-light}$ dataset, comprised of 168 glaucoma and 482 normal images. Besides the image sets S_A and S_B used in [1], which consist of 150 images and 175 images respectively, another set S_C comprised of 325 additional images is also tested. Among the four labeling methods based on our framework, only the pre-learned+refinement method requires a separate training set. For this, we also follow [1] by using S_A for classifier training, using the illumination-normalized feature described in Sec. 2.3.

Table 1. Performance comparisons on different image sets

Method	$S_A \& S_B$			S_C			$S_A \& S_B \& S_C$		
Evaluation criteria	m_1	m_2	δ	m_1	m_2	δ	m_1	m_2	δ
intra-image+refinement	**0.265**	**0.313**	**0.079**	**0.269**	**0.267**	**0.082**	**0.267**	**0.290**	**0.081**
pre-learned+refinement	0.277	0.314	0.087	0.301	0.285	0.091	0.289	0.300	0.089
refinement only	0.331	0.341	0.105	0.325	0.318	0.112	0.328	0.329	0.109
intra-image only	0.269	0.324	0.084	0.277	0.283	0.087	0.273	0.303	0.086
pixel based [10]	0.476	0.702	0.140	0.471	0.663	0.157	0.474	0.683	0.149
window based [1]	0.268	0.315	0.091	0.299	0.297	0.101	0.284	0.306	0.096
Error reduction relative to [10]	44.4%	55.5%	43.6%	42.9%	59.7%	47.6%	**43.6%**	**57.5%**	**45.8%**
Error reduction relative to [1]	1.3%	0.8%	13.3%	10.0%	10.0%	18.6%	**5.9%**	**5.3%**	**16.1%**

3.3   Comparison of Labeling Methods in our Framework

A comparison of the four methods demonstrates the effectiveness of our superpixel based framework. The same parameters (superpixel number $SP = 2048$, propagation range $R_p = 0.1$, regularization parameter $C = 100$ for linear SVM training, $\delta_f = 0.8$ for similarity measurement) are used for all the methods, and the results are listed in Table 1. The following observations can be made about the proposed method:

1. The comparison to *pre-learned+refinement* shows that intra-image learning with retinal structure priors has the following advantages: 1) higher accuracy, 2) no need of extra training samples, 3) no need of specific feature normalization and alignment to deal with the inconsistency between training and testing images;
2. The results of *refinement only* indicate that labeling with a trained classifier is essential for minimizing errors;
3. The comparison to *intra-image only* shows that the refinement scheme provides some reduction in error.

3.4   Comparison to Pixel Based Segmentation and Window Based Learnig

We also compared our superpixel based approach to state-of-the-art pixel and window based methods. The pixel based level-set segmentation method of [10] first identifies pixels that belong to the cup region, and then fits an ellipse to the convex hull. The results show that all four methods based on superpixel labeling lead to significant improvements in cup localization accuracy, which indicates the advantage of superpixel based features over pixel based methods, as also observed in previous work [12].

The state-of-the-art sliding window based method [1] identifies a cup as a whole, by ranking all the cup candidate regions obtained with sliding windows and then producing a single detection result. Our proposed method is shown in Table 1 to yield improvements over [1]. We note that [1] in some sense implicitly uses a structural prior as well, since it searches for cup ellipses only within a certain range of the disc. By contrast, our technique uses the structural priors for intra-image learning of a classifier specific to the retina, which we believe contributes to the higher performance.
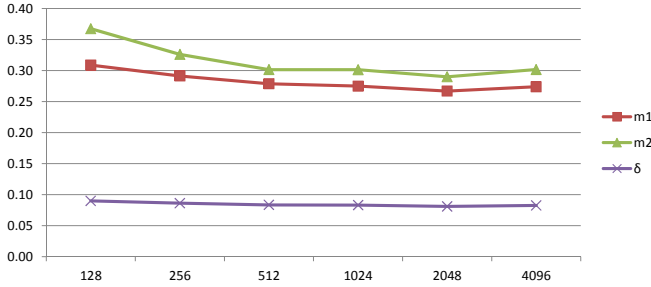
Fig. 3. Cup detection errors with different numbers of superpixels

3.5   Influence of Parameter Settings on Accuracy and Speed

We examined the performance and stability of the proposed method with different parameter settings. For superpixel based approaches, the number of superpixels (SP) is a key parameter that affects performance in terms of both accuracy and speed. As shown in Fig. 3, one can observe that the errors become almost stable when $SP \geq 512$, with the lowest errors at $SP = 2048$ (about 80 pixels per superpixel). When superpixels are too large, the resulting under-segmentation can lead to ambiguously-labeled boundary superpixels that require further segmentation. When superpixels are too small, the resulting features computed from the over-segmented regions may become somewhat less distinctive, making it more difficult to infer the correct labels.

Computation speed was also evaluated, using a four-core $3.4GHz$ PC with $12GB$ RAM. Superpixel segmentation is very fast, requiring less than 200ms for a 400×400 image. Training and testing of linear SVM classifiers is also quick. The most time consuming part is feature extraction for each superpixel, and the overall processing time increases almost linearly with the number of superpixels. For 256, 512 and 2048 superpixels, the computation takes only 1.7, 3.2 and 20.2 seconds per image, respectively. The speed is comparatively slower than that of pixel based segmentation [10], which costs about 1.5 seconds per image; however, our method is much more accurate. Compared to the window based method [1] which takes about 6 minutes per image, ours is much more efficient (tens to hundreds of times faster) with comparable or higher accuracy. We moreover note that our method was implemented with only a single thread, while [1] employs parallel computation.

We also investigated the influence of the propagation range (R_p). Little change in accuracy was found when R_p varies from 0.1 to 0.2; however, the errors become larger when R_p is set much larger (*e.g.*, 0.5). This can be expected, as long range context is less applicable for a superpixel and may introduce error.

4   Conclusion

For cup detection in glaucoma diagnosis, we proposed an intra-image learning framework based on superpixels and retinal structure priors. Tested on a large clinical dataset

with three evaluation criteria, it achieves a 26.7% non-overlap ratio (m_1) with manually-labeled ground-truth, a 29.0% relative absolute area difference (m_2) and a 0.081 absolute CDR error (δ). In future work, we plan to improve contextual label refinement by utilizing a Markov random field (MRF), and elevate system performance using an online learning algorithm which combines intra-image and pre-learned training.

References

1. 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)
2. Klein, B., Klein, R., Sponsel, W., Franke, T., Cantor, L., Martone, J., Menage, M.: Prevalence of glaucoma: the beaver dam eye study. Ophthalmology 99(10), 1499–1504 (1992)
3. Foster, P., Oen, F., Machin, D., Ng, T., Devereux, J., Johnson, G., Khaw, P., Seah, S.: The prevalence of glaucoma in Chinese residents of Singapore: a cross-sectional population survey of the Tanjong Pagar district. Arch Ophthalmology 118(8), 1105–1111 (2000)
4. Shen, S., Wong, T.Y., Foster, P., Loo, J., Rosman, M., Loon, S., Wong, W., Saw, S.M., Aung, T.: The prevalence and types of glaucoma in Malay people: the Singapore Malay eye study. Invest Ophthalmol. Vis. Sci. 49(9), 3846–3851 (2008)
5. Jonas, J., Budde, W., Panda-Jonas, S.: Ophthalmoscopic evaluation of the optic nerve head. Survey of Ophthalmology 43, 293–320 (1999)
6. Abramoff, M., Alward, W., Greenlee, E., Shuba, L., Kim, C., Fingert, J., Kwon, Y.: Automated segmentation of the optic disc from stereo color photographs using physiologically plausible features. Invest Ophthalmol. Vis. Sci. 48(4), 1665–1673 (2007)
7. Liu, J., Wong, D.W.K., Lim, J.H., 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: Int. Conf. Biomed. Eng. (2008)
8. Li, C., Xu, C., Gui, C., Fox, M.: Level set evolution without re-initialization: A new variational formulation. In: CVPR, pp. 430–436 (2005)
9. Merickel, M., Wu, X., Sonka, M., Abramoff, M.: Optimal segmentation of the optic nerve head from stereo retinal images. In: Med. Imag.: Phys., Func., and Struct. from Med. Im. (2006)
10. 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. 5777–8570 (2009)
11. Zhang, Z., Yin, F., Liu, J., Wong, D.W.K., Tan, N.M., Lee, B.H., Cheng, J., Wong, T.Y.: Origa-light: An online retinal fundus image database for glaucoma analysis and research. In: IEEE Int. Conf. Engin. in Med. and Biol. Soc., pp. 3065–3068 (2010)
12. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S.: SLIC Superpixels. EPFL Technical report (2010)
13. Ren, C.Y., Reid, I.: gSLIC: a real-time implementation of SLIC superpixel segmentation. Technical report. University of Oxford, Department of Engineering Science (2011)
14. Onkaew, D., Turior, R., Uyyanonvara, B., Akinori, N., Sinthanayothin, C.: Automatic Vessel Extraction with combined Bottom-hat and Matched-filter. In: Int. Conf. Information and Communication Technology for Embedded Systems (ICICTES), pp. 101–105 (2011)
15. Tighe, J., Lazebnik, S.: SuperParsing: Scalable Nonparametric Image Parsing with Superpixels. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part V. LNCS, vol. 6315, pp. 352–365. Springer, Heidelberg (2010)
16. Fan, R.E., Chang, K.W., Hsieh, C.J., Wang, X.R., Lin, C.J.: LIBLINEAR: A Library for Large Linear Classification. Journal of Machine Learning Research 9, 1871–1874 (2008)