Local Patch Reconstruction Framework for Optic Cup Localization in Glaucoma Detection

Yanwu Xu[1], Ying Quan[1], Yi Huang[2], Ngan Meng Tan[1], Ruoying Li[1], Lixin Duan[1], Lin Chen[1], Huiying Liu[1], Xiangyu Chen[1], Damon Wing Kee Wong[1], Mani Baskaran[3], Shamira Perera[3], Tin Aung[3,4], Tien Yin Wong[3,4] and Jiang Liu[1]

***Abstract*— Optic cup localization/segmentation has attracted much attention from medical imaging researchers, since it is the primary image component clinically used for identifying glaucoma, which is a leading cause of blindness. In this work, we present an optic cup localization framework based on local patch reconstruction, motivated by the great success achieved by reconstruction approaches in many computer vision applications recently. Two types of local patches, i.e. grids and superpixels are used to show the variety, generalization ability and robustness of the proposed framework. Tested on the *ORIGA* clinical dataset, which comprises of 325 fundus images from a population-based study, both implementations under the proposed frameworks achieved higher accuracy than the state-of-the-art techniques.**

I. Introduction

Glaucoma is a leading cause of blindness, which affects about 60 million people [1] and is responsible for approximately 5.2 million cases of blindness (15% of world total) [2]. Unfortunately, glaucoma cannot be cured since it is a serious irreversible disease which affects the optic nerve head.

Studies in [3] and [4] suggest that early detection and treatment of glaucoma is able to reduce the risk of visual field loss; thus early detection is critical to prevent blindness and aid in glaucoma management. However, glaucoma does not have any early warning signs and more than 90% of the afflicted people were unaware of their optical neurodegeneration [5][6]. Glaucoma patients usually first lose their peripheral vision, and then eventually the central vision to become fully blind.

To facilitate effective, efficient and economic glaucoma screening, several computer-assisted glaucoma detection techniques are proposed based on low-cost fundus images in recent years. These CAD systems usually follow the orphamologist to detect glaucoma by performing automatic assessment on certain image cues related to glaucoma. Among the image cues studied for glaucoma assessment, the cup-to-disc ratio (CDR) measure (see Fig. 1), the ratio between the vertical optic cup diameter and vertical optic disc diameter, is a major clinical feature used to gauge the cupping size in glaucoma. The optic disc is located where the ganglion nerve fibers congregate at the retina. The depression inside the optic disc where the fibers leave the retina via the optic nerve head (ONH) is known as the optic cup. The area between the optic nerve and cup is known as the rim. As illustrated in Fig. 2, a larger CDR value usually indicates higher glaucoma risk.

In clinical practises, the CDR value is calculated from a manually outlined optic disc and cup. However, manual annotation is labor intensive and subjective, thus automatic

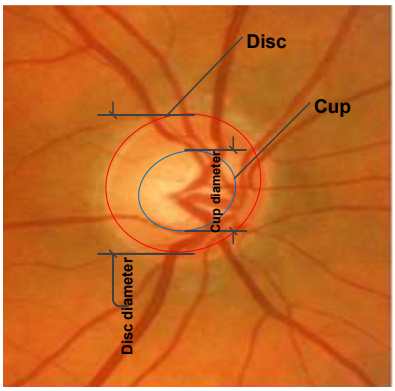

Fig. 1. Illustration of optic disc and cup.

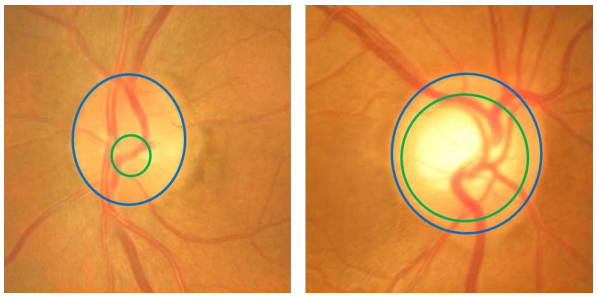

Fig. 2. Effect of glaucoma on cup-to-disc ratio (CDR): optic disc (blue) and cup (green). Left: healthy case with normal(small) cup size, Right: glaucoma case with large cup size.

[1]Y. Xu, Y. Quan, N.M. Tan, R. Li, L. Duan, L. Chen, H. Liu, X. Chen, D.W.K. Wong and J. Liu are with the Institute for Infocomm Research, Agency for Science, Technology and Research, 138632, Singapore {yaxu, quanying, nmtan, liry, duanlx, chenlin, liuhy, chenxy, wkwong, jliu} at i2r.a-star.edu.sg

[2]Y. Huang is with the Research & Innovation, SAP Asia Pte Ltd, 138602, Singapore yi.huang01 at sap.com

[3]M. Baskaran, S.A. Perera, T. Aung, and T.-Y. Wong are with the Singapore Eye Research Institute, 168751, Singapore baskaran.mani at seri.com.sg, shamira.perera at snec.com.sg, aung.tin at snec.com.sg, tien_yin_wong at nuhs.edu.sg

[4]T. Aung and T.-Y. Wong are also with the Department of Ophthalmology, National University of Singapore, 119074, Singapore

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

approaches are designed for disc and cup segmentation in recent years.

In previous works, optic disc segmentation is relatively well studied and has achieved considerable accuracy, by using various techniques such as intensity gradient analysis, Hough transforms, pixel classification [7], vessel geometry analysis, deformable models and level sets [8] [9] [10].

In this paper, we study the challenging cup segmentation problem [11]. Previous cup segmentation algorithms can be roughly categorized into 3 types: 1) classifying pixels as part of the cup or rim [7][10]; 2) evaluating the possibilities of large regions (e.g. sliding windows) to determine an optimal cup candidate [13]; 3) classifying superpixels (small regions) as part of the cup or rim [14] [15]. Currently, pixel based approaches have the lowest accuracy and superpixel based approaches are state-of-the-art.

In most of the previous methods, a discriminant classification/regression model (e.g. SVM and SVR) is prelearned to label/classify each testing object. It has recently been shown in some computer vision applications, e.g. gait recognition [16] and prostate segmentation [17], higher accuracy can be obtained through reconstruction (a.k.a. atlas approach in medical image computing area) from the samples. We demonstrate in this paper there are appreciable improvements in cup localization accuracy with the reconstruction-based approach. In many applications, reconstruction is applied to whole images represented by raw pixel values; however, good registration/alignment is needed to guarantee high accuracy, which is not applicable for some applications. In contrast to such global reconstruction approach, local patch reconstruction is used in this work which can better deal with misalignment and enlarge the reference dictionary to get better reconstruction of each test unit. It is also worth mentioning that global reconstruction can be treated as an extreme case of the local reconstruction, when each test sample has only one local patch. In addition, we also demonstrate that other image representations can be used for reconstruction, especially for the objects with different shapes and sizes.

Tested on the large *ORIGA* dataset [12] comprising of 325 image from a population-based study, four variant implementations of the proposed optic cup localization framework achieved improvement on cup localization accuracy, comparing with the current state-of-the-art optic cup segmentation methods, which demonstrate the effectiveness and robustness of the proposed framework.

II. Local Patch Reconstruction Based Cup Localization

A. Framework

In this work, we follow prior arts [14][15][13], localizing the optic cup from a given disc image which is obtained by using segmentation methods such as [8]. As illustrated in Fig. 3, the input disc image is first divided into local regions (*e.g.*, grids and superpixels), and then each patch is reconstructed with reference patches selected from training samples (reference images) to get a label for each pixel/patch, and an unique cup region is finally determined by ellipse fitting.

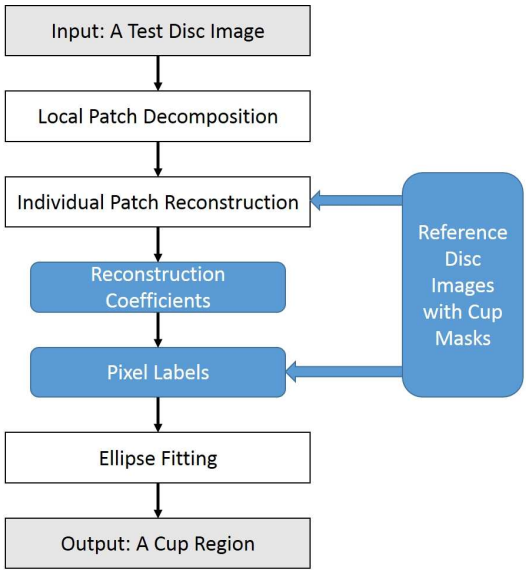

Fig. 3. Flowchart of the proposed cup localization framework using local patch reconstruction.

B. Formulation

For a given local patch $y \in \mathbb{R}^{d\times 1}$ to be reconstructed, we want to compute optimal linear reconstruction coefficients $w \in \mathbb{R}^{k\times 1}, |w| = 1$, with a dictionary consists of k reference patches, to minimize the reconstruction error $||y - Xw||^2$. The dictionary is formed by k most similar patches selected from reference discs in near location to the test patch (i.e. in our implementation, the center points of reference patches are at most $1/10$ disc diameter far from the test patch center location), denoted by $X = \{x_1, x_2, \cdots, x_k\} \in \mathbb{R}^{d\times k}$, where each column x_i is a reference patch expressed as a vector (e.g., raw pixel values or high level features).

Our objective function also includes a cost term that penalizes the use of references that are less similar to the test patch. Let us denote the costs for the reference patches in X as the vector $c = \{c_1, c_2, \cdots, c_k\}^\top \in \mathbb{R}^{k\times 1}$, where c_i is the cost of using x_i for reconstruction. The overall cost term can then be expressed as $||c \odot w||^2$ where $\odot$ denotes the Hadamard product. Combining this cost term with the reconstruction error gives the following objective function:

$$\begin{aligned} \min_{w} &||y - Xw||^2 + \lambda||c \odot w||^2, \\ &s.t.\ |w| = 1, \end{aligned} \tag{1}$$

where $\lambda > 0$ is a regularization parameter. This objective can be minimized in closed form using the Lagrange multiplier method, without the need for iterations:

$$\begin{aligned} w =& \frac{1}{\mathbf{1}^\top(\hat{X}^\top\hat{X} + \lambda C^\top C)\mathbf{1}}(\hat{X}^\top\hat{X} + \lambda C^\top C)^{-1}\mathbf{1}, \\ \hat{X} =& (\mathbf{1} \otimes y - X), \end{aligned} \tag{2}$$

where $C = \mathrm{diag}(c)$ and $\otimes$ denotes the Kronecker product. For simplicity, we define in our implementation the cost c_i

as the Gaussian distance between the test patch y and the i-th reference patch x_i, *i.e.*,

$$c_i = \exp(||y - x_i||^2/\sigma^2), \tag{3}$$

where σ is a parameter that accounts for imaging noise.

When the cost term is excluded, i.e., setting λ = 0, the reconstruction method is the well known locally linear embedding (LLE), which is also compared in the experiment.

C. Implementations

In this subsection, two implementations of this framework are presented in detail. Different processes are taken before and after the reconstruction of two types of local patches (i.e., grids and superpixels), according to the characteristics of each local patch type. As claimed in [14], [15] and [18], superpixels have the advantage of preserving local boundaries, while grids don't. Therefore superpixel label is predicted in a whole since most pixels in a superpixel have the same label, similar to the state-of-the-art supperpixel classification based cup localization approach, where the reconstruction is used as a superpixel labelling method. In addition, superpixels have different shapes and sizes, and original raw image reconstruction can not be performed directly, thus high level visual features are used to perform reconstruction. In contrast, raw image values are used for grid reconstruction, and each pixel label is predicted by applying the reconstruction coefficients on the cup masks corresponding to reference patches.

With either implementation, label of each pixel can be obtained, and then the minimum ellipse that encompasses all the pixels with positive labels (i.e. cup) is computed to produce the final detection result, represented by ellipse center/elongation parameters $(\hat{\mu}, \hat{\nu}, \hat{\alpha}, \hat{\beta})$.

1) Implementation 1 (grid reconstruction): Each disc image is resized to 256×256 and then divided to 16×16 half overlapped blocks/grids. Each block is a local patch represented with raw pixel values, thus each local patch has a 256 dimensional feature (i.e., d = 256) in this implementation. After the optimal reconstruction coefficient w of a test patch y is obtained, the weighted sum $\sum_{i=1}^{k} w_i m_i$ of the cup masks ($m_i \in \mathbb{R}^{16\times16}$) corresponding to the k reference patches is calculated, which is also a 16×16 block. After all grids are reconstructed, the label of each pixel (i.e., predicted cup mask) is obtained via consolidating all the reconstructed cup masks of each block. Since the blocks are half overlapped, most pixels are included in 4 blocks; thus the initial label of each pixel is defined as the average value of all its corresponding predictions.

2) Implementation 2 (superpixel reconstruction): When superpixel is used as local patch, the reconstruction framework can be used to label each superpixel, similar to the method in [14] and [15] under the same superpixel labelling framework. We follow these works, utilizing the SLIC (Simple Linear Iterative Clustering) algorithm [18] to segment the disc image into compact and nearly uniform superpixels (see Fig. 4). Superpixels are then presented by features used in [14] and [15]. For the i-th superpixel, we extract

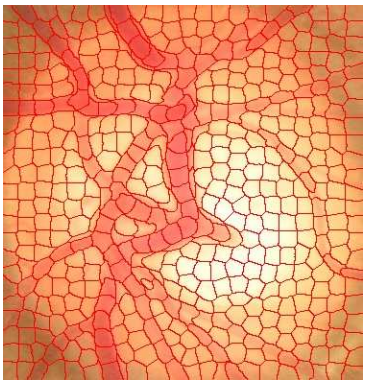

Fig. 4. Example of superpixel segementation using SLIC [18].

a feature vector $\mathbf{x}_i$ that consists of position information (denoted by (u_i, v_i, ρ_i), 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. Features are then normalized to avoid magnitude differences among dimensions. The label of a test superpixel is predicted by calculating the weighted sum $\sum_{i=1}^{k} w_i l_i$ of the cup/non-cup labels ($l_i \in \{+1, -1\}$) corresponding to the k reference superpixels. All pixels in the same superpixel share the same label. For efficiency, supepixels are segmented at only one scale (512 superpixels per image), thus superpixels are non-overlapped. After multi-scale superpixel segmentation is performed, similar averaging process can be applied to get the final label of each pixel.

III. Experiments

In this section, we first describe the evaluation criteria, and then evaluate our local patch reconstruction based cup localization framework through an experimental comparison to pixel based segmentation methods [10], sliding window method [13] and the state-of-the-art superpixel classification method [15]. For fair comparison, we use the same dataset and training/testing split method as in [14] and [15], i.e., using the *ORIGA* dataset comprised of 325 images, where the first 150 images are used as training samples and the other 175 images are used for testing. The optimal parameters are determined by using cross validation on the 150 training images, for all methods. To validate the robustness of the proposed approach, we also compared four implementations under the proposed reconstruction framework, using grid or superpixel as local patch, performing the reconstruction with or without cost term in Eq. (1).

A. Evaluation criteria

Three evaluation criteria are commonly used for cup localization/segmentation, namely non-overlap ratio (m_1), relative absolute area difference (m_2) and absolute cup-to-disc ratio (CDR) error (δ), defined as:

$$\begin{aligned} m_1 &= 1 - \frac{area(E_{dt} \bigcap E_{gt})}{area(E_{dt} \bigcup E_{gt})}, \\ m_2 &= \frac{|area(E_{dt}) - area(E_{gt})|}{area(E_{gt})}, \\ \delta &= |CDR(E_{dt}) - CDR(E_{dt})|, \end{aligned} \tag{4}$$

TABLE I

PERFORMANCE COMPARISON TO STATE-OF-THE-ART CUP LOCALIZATION/SEGMENTATION METHODS

Method	m_1	m_2	δ
Grid reconstruction w cost term	**0.258**	**0.263**	**0.077**
Pixel classification [10]	0.491	0.859	0.159
Sliding window regression[13]	0.289	0.409	0.106
Superpixel classification [15]	0.281	0.354	0.089
Relative error reduction to [10]	47.5%	69.4%	51.6%
Relative error reduction to [13]	10.7%	35.7%	27.4%
Relative error reduction to [15]	8.2%	25.7%	13.5%

TABLE II

PERFORMANCE COMPARISON OF DIFFERENT IMPLEMENTATIONS OF THE RECONSTRUCTION FRAMEWORK

Patch	Cost Term	m_1	m_2	δ
Grid	w	**0.258**	0.263	**0.077**
Grid	w/o	0.266	**0.260**	0.081
Superpixel	w	0.261	0.263	0.080
Superpixel	w/o	0.264	0.278	0.084

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 most related to cup localization accuracy and δ is most related to glaucoma diagnosis; thus these two criteria are relatively more important.

B. Comparison to prior art

We compared our local reconstruction based approach (with the implementation of *Grid reconstruction w cost term*) to state-of-the-art cup segmentation methods, namely *Pixel classification*) [10], *Sliding window regression* [13] and *Superpixel classification* [15]. The results are shown in Table I. Compared with the most advanced *Superpixel* approach [15], our method is shown to significantly improve cup localization accuracy in terms of m_1, m_2 and CDR error (δ), which are reduced by 8.2%, 25.7% and 13.5%, respectively.

C. Comparison of different implementations of the proposed framework

To show the generalizability and robustness of the proposed framework, four implementations of the proposed framework are also compared. The implementations use either grid or superpixel as local patch, with or without the cost term, respectively. From the results listed in Table I and Table II, one can observe that all of the four local reconstruction approaches outperform existing methods. Among the four implementations, grid reconstruction with cost term achieved the lowest error rates in terms of m_1 and δ, which are more important for cup localization evaluation. One can also observe that including the cost term can slightly reduce the errors; however it also increases the computing complexity.

IV. CONCLUSION

For cup localization in glaucoma detection, we proposed an effective and robust local patch reconstruction framework. Tested on a large clinical dataset with three evaluation criteria, all of the four different implementations under the proposed framework achieved higher accuracies compared to prior arts. In future work, we plan to evaluate the framework on other medical image segmentation and classification problems.

REFERENCES

[1] H.A. Quigley and A.T. Broman. The number of people with glaucoma worldwide in 2010 and 2020, *British Journal of Ophthalmology*, vol. 90, no. 3, pp. 262–7, 2006.

[2] B. Thylefors, and A. Negrel. The global impact of glaucoma, *Bull World Health Organ*, vol. 72, no. 3, pp. 323–6, 2006.

[3] A. Heijl, M. C. Leske, B. Bengtsson, L. Hyman, B. Bengtsson, and M. Hussein. Reduction of intraocular pressure and glaucoma progression: results from the Early Manifest Glaucoma Trial, *Arch. Ophthalmol.*, vol. 120, no. 10, pp. 1268–1279, 2002.

[4] G. Michelson, S. Wärntges, J. Hornegger, and B. Lausen. The papilla as screening parameter for early diagnosis of glaucoma, *Dtsch. Arztebl. Int.*, vol. 105, no. 34-35, pp. 583–9, 2008.

[5] P. Foster, F. Oen, D. Machin, T. Ng, J. Devereux, G. Johnson, P. Khaw, and S. Seah. The prevalence of glaucoma in Chinese residents of Singapore: a cross-sectional population survey of the Tanjong Pagar district, *Arch Ophthalmology*, vol. 118, no. 8, pp. 1105–11, 2000.

[6] S. Shen, T.Y. Wong, P. Foster, J. Loo, M. Rosman, S. Loon, W. Wong, S.M. Saw, and T. Aung. The prevalence and types of glaucoma in Malay people: the Singapore Malay eye study, *Invest. Ophthalmol. Vis. Sci.*, vol. 49, no. 9, pp. 3846–51, 2008.

[7] M. Abramoff, W. Alward, E. Greenlee, L. Shuba, C. Kim, J. Fingert, and Y. Kwon. Automated segmentation of the optic disc from stereo color photographs using physiologically plausible features, *Invest. Ophthalmol. Vis. Sci.*, vol. 48, no. 4, pp. 1665–73, 2007.

[8] J. Liu, D.W.K. Wong, J.H. Lim, H. Li,. N.M. Tan, Z. Zhang, T.Y. Wong, and R. Lavanya. ARGALI: an automatic cup-to-disc ratio measurement system for glaucoma analysis using level-set image processing, In *Int. Conf. IEEE Engin. in Med. and Biol. Soc.*, 2008.

[9] C. Li, C. Xu, C. Gui, and M. Fox. Level set evolution without re-initialization: A new variational formulation, In *IEEE Conf. Computer Vision and Pattern Recognition*, 2005, pp. 430–6.

[10] D.W.K. Wong, J.H. Lim, N.M. Tan, Z. Zhang, S. Lu, H. Li., M. Teo, K. Chan, and T.Y. Wong. Intelligent fusion of cup-to-disc ratio determination methods for glaucoma detection in ARGALI, In *Int. Conf. Engin. in Med. and Biol. Soc.*, 2009, pp. 5777–80.

[11] M. Merickel, X. Wu, M. Sonka, and M. Abramoff. Optimal segmentation of the optic nerve head from stereo retinal images, In *Medical Imaging: Physiology, Function, and Structure from Medical Images*, 2006.

[12] Z. Zhang, F. Yin, J. Liu, D.W.K. Wong, N.M. Tan, B.H. Lee, J. Cheng, and T.Y. Wong. Origa-light: An online retinal fundus image database for glaucoma analysis and research, In *Int. Conf. IEEE Engin. in Med. and Biol. Soc.*, 2010, pp. 3065–8.

[13] Y. Xu, D. Xu, S. Lin, J. Liu, J. Cheng, C.Y. Cheung, T. Aung, and T.Y. Wong. Sliding Window and Regression based Cup Detection in Digital Fundus Images for Glaucoma Diagnosis. In: *MICCAI 2011, Part III. LNCS,* 2013, pp. 1–8.

[14] Y. Xu, J. Liu, J. Cheng, F. Yin, N.M. Tan, D.W.K. Wong, M. Baskaran, C.Y. Cheung, and T.Y. Wong. Efficient Optic Cup Localization Using Regional Propagation Based on Retinal Structure Priors, in *Int. Conf. IEEE Engin. in Med. and Biol. Soc.*, 2012, pp. 1430–3

[15] Y. Xu, J. Liu, F. Yin, N.M. Tan, D.W.K. Wong, C.Y. Cheung, Y.C. Tham, and T.Y. Wong. Efficient Optic Cup Localization Based on Superpixel Classification for Glaucoma Diagnosis in Digital Fundus Images in : *ICPR*, 2012.

[16] D. Xu, Y. Huang, Z. Zeng, and X. Xu. Human Gait Recognition using Patch Distribution Feature and Locality-constrained Group Sparse Representation. *IEEE T. Im. Proc.* , pp. 316–26, 2012.

[17] Y. Cao, Y. Yuan, X. Li, B. Turkbey, P.L. Choyke, and P. Yan. Segmenting Images by Combining Selected Atlases on Manifold. In: *MICCAI 2011, Part III. LNCS,*, 2011, pp. 272–279.

[18] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk. SLIC Superpixels Compared to State-of-the-art Superpixel Methods, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, n.. 11, pp. 2274 - 2282, 2012.