34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Automatic localization of retinal landmarks

Xiangang Cheng[1], Damon Wing Kee Wong[1], Jiang Liu[1], Beng-Hai Lee[1], Ngan Meng Tan[1], Jielin Zhang[2], Ching Yu Cheng[3], Gemmy Cheung[3] and Tien Yin Wong[3]

Abstract— **Retinal landmark detection is a key step in retinal screening and computer-aided diagnosis for different types of eye diseases, such as glaucomma, age-related macular degeneration(AMD) and diabetic retinopathy. In this paper, we propose a semantic image transformation(SIT) approach for retinal representation and automatic landmark detection. The proposed SIT characterizes the local statistics of a fundus image and boosts the intrinsic retinal structures, such as optic disc(OD), macula. We propose our salient OD and macular models based on SIT for retinal landmark detection. Experiments on 5928 images show that our method achieves an accuracy of 99.44% in the detection of OD and an accuracy of 93.49% in the detection of macula, while having an accuracy of 97.33% for left and right eye classification. The proposed SIT can automatically detect the retinal landmarks and be useful for further eye-disease screening and diagnosis.**

I. INTRODUCTION

It is important to detect the location of retinal landmarks in color fundus images for automatic screening and diagnosis of retinal disease, as the abnormalities of an eye disease such as age-related macular degeneration(AMD) usually occur at a certain area in the retina. With the location of the retinal landmarks, the region of interest(ROI) for a disease can be easily obtained for further analysis. Among all the retinal landmarks, optic disc(OD) and macula are the most salient. Precisely, in a retinal fundus image(see Fig. 1 for illustration), the optic disc (also known as the optic nerve head) is a highly visible feature from which vessels and retinal nerve fibers exit. And the macular region is usually described as a darker region of the retina located temporally to the optic nerve head, and generally growing darker towards the macula centre, due to increased pigmentation.

Previous work try to detect the optic disc based on the characteristics of the OD structures. Among them, some methods [1] focused on the intensity based feature, which assume the OD is brightest in an retinal image. However, these methods are sensitive to illumination changes of retinal images. And they are also vulnerable to other bright lesions. Meanwhile, some methods [2], [3] detected the OD by exploring the fact that the optic disc is at the origin of the retinal vasculature and the vasculature pattern is stable over the retina. The key issue of such methods requires a prior segmentation of the vasculature. For macula detection, some researches [4] proposed algorithms based on the fact that the fovea is usually a dark-yellowish blob by searching for it directly in the image. However, the macula blob is easily effected by diseases, such as AMD. Some other recent methods [5] tried to find macula considering the position of the macula with respect to the optic disc and the vasculature.

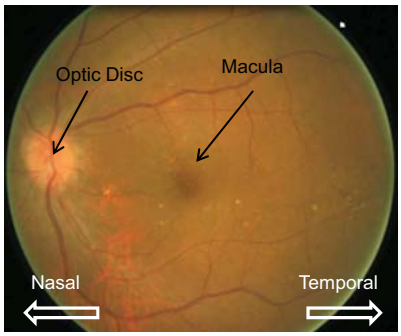


Fig. 1. Example of retinal fundus image(left eye), as well as the retinal landmarks: optic disc and macula.

The identification of different structures is highly dependent on the discriminative potential of the employed image features. Ideally, these features should be highly representative and discriminative. A single pixel with intensity or color information lacks descriptive power and is not able to distinguish different structures and lesions. In contrast to these approaches, we present a new approach to represent retinal images. Our motivation here is to boost image representation by mapping it into a new space. More precisely, motivated by the successful usage of the Bag-of-Words model [6] and semantic segmentation, we apply a local neighborhood operation for each pixel and assign it a new signature, which is called a "semantic pixel". These pixels form a "semantic image", in which each pixel embodies local characteristics of its neighborhood and is highly discriminative. In the semantic retinal image, it is easier to distinguish between different structures, such as optic disc, macula.

The rest of the paper is organized as follows: In section II we present the details of the Semantic Image Transformation(SIT). Section III gives the templates and representations of OD and macula respectively. And the experiment results are reported in section IV. Finally, we conclude the contributions of our work in Section V.

[1]X. Cheng, D.W.K. Wong, J. Liu, B.-H. Lee and N.-M. Tan, are with the Institute for Infocomm Research, Agency for Science, Technology and Research, 138632, Singapore {xcheng, wkwong, jliu, benghai, nmtan} AT i2r.a-star.edu.sg

[2]Nanyang Technological University (NTU), Singapore jzhang10 AT e.ntu.edu.sg

[3] C.-Y. Cheng, G. Cheung and T.-Y. Wong are with Singapore Eye Research Institute, 168751, Singapore ching-yu_cheng AT nuhs.edu. sg, gemmy.cheung.c.m AT snec.com.sg and tien_yin_wong AT nuhs.edu.sg

II. Semantic Image Transformation

A. Local neighborhood operation

Due to the lack of representative power, singular image pixels of intensities or rgb values are difficult to distinguish structures in retinal images. For example, both the optic disc and drusen lesion are bright (in intensity), while both the macula and vessels are dark. In addition, the pixels values are sensitive to many aspects, such as lighting and blur. In this paper, the contextures of pixels are incorporated to boost the representative power. Given an image $I :\subset \mathbb{R}^2 \rightarrow \mathbb{R}^3$, the pixel at location x can be boosted with a local neighborhood operation $\mathcal{H}$:

$$f_x = \mathcal{H}(x) \tag{1}$$

As the major anatomical retinal structures and many lesions are shown to be circular-like shapes,the SPIN feature[7] is adopted as the local neighborhood operation, which encodes the distribution of image brightness values in the neighborhood of a particular reference (center) point. It is achieved by a soft-assigned histogram of the intensity values i of pixels located at a distance d from the center x. That is,

$$f_x(d,i) = \sum_{y\in\Gamma(x)} exp\left(-\frac{(|y-x|-d)^2}{2\alpha^2} - \frac{|i_y - i|^2}{2\beta^2}\right), \tag{2}$$

where $\Gamma(x)$ is the neighbor patch of x, and α and β are the parameters representing the soft width of the two-dimensional histogram bin. In our experiments, we use 5 bins for distance d and 10 for intensity value i, thus resulting in 50-dimensional feature vector f_x.

B. Semantic pixel construction

Using the high-dimensional histogram directly is very complex and time-consuming. What's more, varying in cardinality and lacking of meaningful ordering in histograms results in difficulties of finding an acceptable model to represent the whole retinal image. To address the problems, we follow the "Bag-of-Words" method [6]. First, the k-means algorithm is applied to cluster all the features from training images into K clusters by minimizing the intra-cluster variance:

$$Minimize: \quad V = \sum_{k=1}^{K} \sum_{f_j \in C_k} \|f_j - \mu_k\|^2, \tag{3}$$

where C_k $(k = 1, 2, \ldots, K)$ is one of the sub-clusters and μ_k is the centroid of all the features $f_k \in C_k$. Given an original pixel i_x located at x with SPIN features f_x, this will be mapped to a new pixel k_x:

$$i_x \Rightarrow \{k_x = \arg \min_{k\in[1,2,\ldots,K]} \|f_x - \mu_k\|^2\}. \tag{4}$$

Usually, visual words are constructed from general clustering methods, such as K-means clustering method. However, clusters from these methods are ranged orderless and the similarity between different clusters is not considered. Here we adopt hierarchical k-means clustering method, which groups data simultaneously over a variety of scales and builds the semantic relations of different clusters. The hierarchical k-means algorithm organizes all the centers of clusters in a tree structure. It divides the data recursively into clusters. In each iteration (each node of the tree), k-means is utilized by dividing the data belonging to the node into k subsets. Then, each subset is divided again into k subsets using k-means. The recursion terminates when the data is divided into a single data point or a stop criterion is reached. Exactly, suppose the cluster tree is $L(L \geq 0)$ level and the branch numbers for each node is K, then we will get a full K-nodes tree. The number of leaf nodes is $N = K^L$. The parent and grandparent node can be computed from the leaf node easily in the full K-nodes tree. That is, given the t-th $(t \in [0, 1, N-1])$ leaf node ψ_{Lt}, we compute the index of its parent or grandparent node ψ_{Lt^*} in level l by the equation

$$t^* = \lfloor t/K^{L-l} \rfloor, \ l \in [0, 1, , L]. \tag{5}$$

One difference between k-means and hierarchical k-means is that k-means minimizes the total distortion between the data points and their assigned closest cluster centers, while hierarchical k-means minimizes the distortion only locally at each node and in general this does not guarantee a minimization of the total distortion.

As the new pixel k_x embeds the local characteristics of its surroundings, it is called "sematic pixel". Correspondingly, the construction is named as "Semantic Image Transformation(SIT)".

The new proposed semantic image has many merits (See Fig. 2 for illustration).

- The "semantic pixel" encodes local neighborhood information and refers to a specific structure of local patch. This increases its representative information and makes it easier to describe a certain retinal pattern into a machine-recognizable feature representation.
- Spatial information is embedded in semantic image and many algorithms in the original image space can be extended to the new transformation space easily.

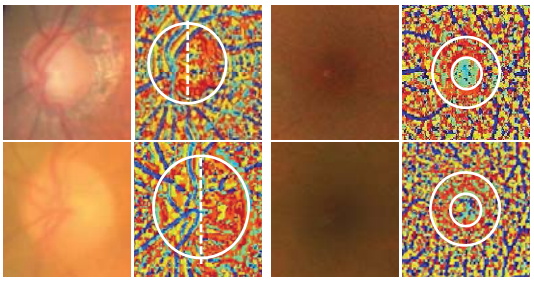


Fig. 2. Examples of retinal structures, as well as the corresponding SITs. All subimages are scaled to the same size. Please refer to the color pdf file for better view.

III. SIT-Model

A. OD detection

Optic disc(OD) is one of the major retinal structures. OD detection is usually the first step of several retinal

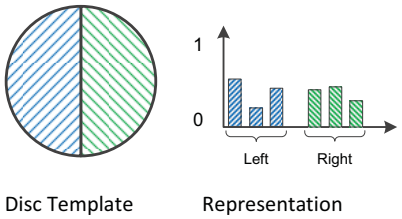


Fig. 3. SIT-OD templates and represenations

applications. Comparing to other structures, the OD is more salient and informative. It also contains the information of left or right eye [8]. In this section, we propose a new model, which can both detect the OD in the retinal image and judge if it is a left/right eye.

The SIT-OD representation should concurrently identify the optic disc and determine the eye configuration. The optic disc has the following physiological characteristics [9](see Fig. 1 for illustration):

1) Temporal intensity $>$ Nasal intensity within the optic disc;
2) Optic nerve head vessels are located towards the temporal region.

The SIT-OD model is proposed based on this(Fig. 2 (a)). We propose a circular disc template and generate a SIT histogram(Fig. 4) for each half of the model, and the SIT-OD representation is the concatenation of the two histograms.

B. Macula detection

The macula consists of light-absorbing photoreceptors and is largely darker compared to its surroundings. Furthermore, there is a physiological avascular region around the macula center. The SIT-macula model is proposed with two concentric circles. The inner circle corresponds to the darker circular region of the macula center. The outer ring exploits the avascular characteristics of the surrounding retinal region. See Fig. 2 (b) for example. Similar to optic disc, the SIT-macula representation(Fig. 4) is formed by concatenating the histograms of the two parts from the semantic image.

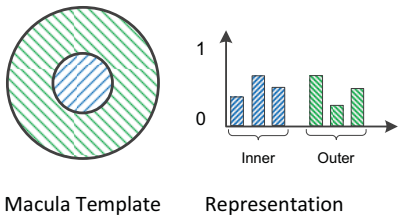


Fig. 4. SIT-Macula templates and represenations

C. Retinal Landmark Detector

In the training process, the positive instances are representations from the labeled ROIs and the negative instances are selected randomly. We use Support Vector Machines (SVM) to learn the detectors from our macular ROI representation.

TABLE I
PERFORMANCE OF ROI DETECTION AND LEFT/RIGHT EYE CLASSIFICATION IN THALIA DATASET.(MORE RESULTS WILL BE ADDED LATER)

	OD	L&R Eye	Macula
Drusen Images(%)	100	97.87	89.36
Non-drusen Images(%)	100	99.61	95.69
Overall(%)	100	99.14	93.98

For a two-class scenario, the decision function for a test sample x has the following form:

$$g(x) = \sum_i a_i y_i K(x_i, x) - b, \tag{6}$$

where $K(x_i, x)$ is the value of a kernel function for the training sample x_i and the test sample x, y_i ($+1$ or -1) the class label of x_i, α_i the learned weight of the training sample x_i, and b is a learned threshold parameter. For the detection process, the sliding window scheme is adopted and the optimal is selected. To cover the variance of the OD/macula size, multi-scale windows are detected. As the OD and macula are unique in each retinal fundus image, the window with the maximum decision value is selected as the detected OD/macula.

IV. EXPERIMENT

We tested our proposed system on images from the Singapore Malay Eye Study, which is a population-based study of an urban population in Singapore. First, the test is conducted on a batch of 350 field 2 images(THALIA dataset) with a 45 degree field of view, in which both macula and optic disc are visible. The image set consists of 96 AMD images with drusen and 254 normal images. All the images are unique and are from different patients. In our experiments, the training and testing sets are randomly chosen. Precisely, 50 drusen images as well as 50 non-drusen images are used to train the supervised classification model,of which the ROI of OD and macula are manually labeled. The rest of images are then used for testing. LibSVM toolbox [10] is used. After that, we use our learned model directly on the whole set, with a batch of 5928 images(SIMES dataset), either field 1 or field 2. In the dataset, there are images containing different diseases, such as glaucoma, age-related macular degeneration(AMD), myopia and so on.

First, we show the results of the ROI detection in THALIA dataset(Tab. IV). The optic disc detection is 100% for all subsets, which shows the effectiveness of our proposed SIT for OD detection. The overall accuracy of left/right eye recognition is 99.14% based on our SIT-OD model. From the results we can note that the performance of macula detection on the drusen set is much lower than that on the normal set when using macula detector in the whole retinal image. This shows that the drusen lesions do affect the macula structures.

Second, we conduct our proposed method on a larger dataset, which contains 5928 fundus images. The ROI detectors and left/right eye classifier are the same as the models trained on THALIA dataset. The overall accuracy for OD

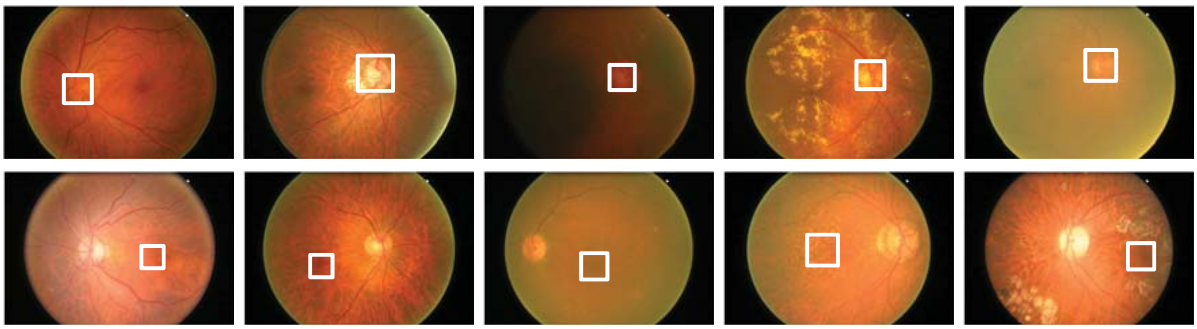
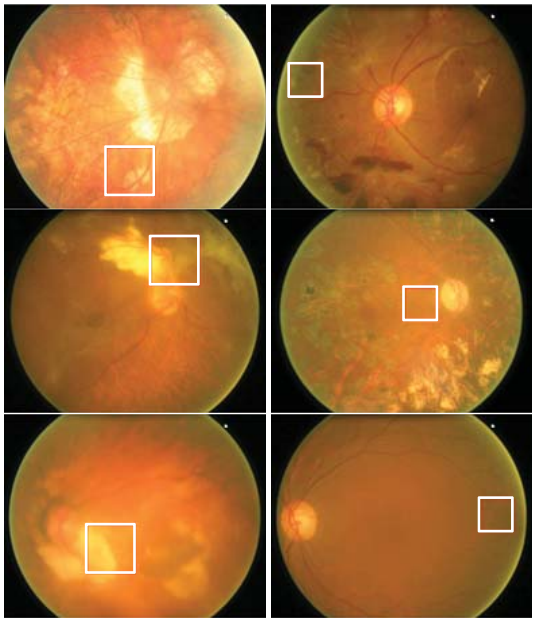


Fig. 5. Examples of ROI detections with different illuminations and diseases. The upper row is the ROIs of OD and the lower row is the ROIs of macula.

TABLE II
PERFORMANCE OF ROI DETECTION AND LEFT/RIGHT EYE CLASSIFICATION IN SIMES DATASET.

	total	OD	L&R Eye	Macula
Glaucomma	168	168/100%	164/97.62%	155/92.26%
AMD	872	871/99.89%	866/99.31%	843/96.67%
Others	4888	4856/99.35%	4740/96.97%	4544/92.96%
Overall	5928	4995/99.44%	5770/97.33%	5542/93.49%

detection is 99.44%. For the left/right eye classification, the result is 97.33%. And the result of macula detection is 93.49%. All the results show that our proposed method is general and robust, which can be used in other retinal fundus images easily and effectively. Fig. 5 shows the results of our proposed method. There are different challenges, such as bad illumination, blur, bright lesions, drusen lesions and so on. However, our algorithms can successfully detect both the OD and macula. In Fig. 6, we show some failed cases of our algorithms. For OD detection, the challenges are those images with bright lesions around blood vessels. For macula detection, our algorithm failed for some images due to the weak appearance of the macula.



(a) Optic Disc (b) Macula

Fig. 6. Examples of failed cases for retinal landmark detection. The left column is the result of OD detection and the right column is the result of macula detection.

V. CONCLUSION

This paper presents a Semantic Image Transformation for landmark detections for retinal fundus image. Such transformation improves the discriminative power of the image and boosts the visual representation of clinical signs. Experiments on two dataset demonstrate our method is effective for retinal landmark detection, such as OD and macula detection. Our proposed method can be used for large-scale population screening.

REFERENCES

[1] M Lalonde, M Beaulieu, and L Gagnon, "Fast and robust optic disc detection using pyramidal decomposition and hausdorff-based template matching.," *IEEE Transactions on Medical Imaging*, vol. 20, no. 11, pp. 1193–1200, 2001.

[2] M Foracchia, E Grisan, and A Ruggeri, "Detection of optic disc in retinal images by means of a geometrical model of vessel structure.," *IEEE Transactions on Medical Imaging*, vol. 23, no. 10, pp. 1189–1195, 2004.

[3] T Walter, J C Klein, P Massin, and A Erginay, "A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina," 2002.

[4] C Sinthanayothin, J F Boyce, H L Cook, and T H Williamson, "Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images.," *The British journal of ophthalmology*, vol. 83, no. 8, pp. 902–910, 1999.

[5] Meindert Niemeijer, Michael D Abrmoff, and Bram Van Ginneken, "Segmentation of the optic disc, macula and vascular arch in fundus photographs.," *IEEE Transactions on Medical Imaging*, vol. 26, no. 1, pp. 116–127, 2007.

[6] Josef Sivic and Andrew Zisserman, "Video google: a text retrieval approach to object matching in videos," *ICCV*, vol. 2, pp. 1470–1477, 2003.

[7] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce, "A sparse texture representation using local affine regions," *PAMI*, vol. 27, no. 8, pp. 1265–1278, Aug. 2005.

[8] Ritch Robert, Shields M. Bruce, and Krupin Theodore, "The glaucomas," *Mosby-Year Book*, vol. 1, pp. 617–626, 1989.

[9] Ngan Meng Tan, Jiang Liu, Damon W K Wong, Zhuo Zhang, Shijian Lu, Joo Hwee Lim, Huiqi Li, and Tien Yin Wong, "Classification of left and right eye retinal images," *Imaging*, vol. 7624, pp. 7624–7638, 2010.

[10] Chih-Chung Chang and Chih-Jen Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.