Coupled Sparse Dictionary for Depth-Based Cup Segmentation from Single Color Fundus Image*

Arunava Chakravarty and Jayanthi Sivaswamy

Center for Visual Information Technology,
International Institute of Information Technology Hyderabad, India

Abstract. We present a novel framework for *depth* based optic cup boundary extraction from a *single* 2D color fundus photograph per eye. Multiple depth estimates from shading, color and texture gradients in the image are correlated with Optical Coherence Tomography (OCT) based depth using a coupled sparse dictionary, trained on image-depth pairs. Finally, a Markov Random Field is formulated on the depth map to model the relative depth and discontinuity at the cup boundary. Leave-one-out validation of depth estimation on the *INSPIRE* dataset gave average correlation coefficient of 0.80. Our cup segmentation outperforms several state-of-the-art methods on the *DRISHTI-GS* dataset with an average F-score of 0.81 and boundary-error of 21.21 pixels on test set against manual expert markings. Evaluation on an additional set of 28 images against OCT scanner provided groundtruth showed an average rms error of 0.11 on Cup-Disk diameter and 0.19 on Cup-disk area ratios.

1 Introduction

Glaucoma, a sight-threatening disease, is characterized by the deformations in the optic disk (OD) in retina. The OD is a bright elliptic region with a central depression (called the optic cup) devoid of retinal nerve fibers surrounded by a neuro-retinal rim, where the nerve fibers bend into the cup region (Fig 1(a)). Glaucoma destroys optic nerve fibers causing neuro-retinal rim thinning and cup enlargement which is widely measured quantitatively via the vertical Cup-Disk diameter ratio (CDR). Deriving CDR from fundus images requires accurate OD and the cup boundaries and hence their segmentation has received much attention. Majority of the existing methods perform OD segmentation in 2 stages: OD localization using intensity and shape based template matching such as Hough transform [1][2], followed by boundary extraction using ellipse fitting [3] or specially adapted deformable models [1],[2].Though most methods report good performance in OD segmentation, cup segmentation remains a challenging problem. The optic cup boundary (OCB) is largely characterized by change in depth of the retinal surface in the OD region and hence methods for cup segmentation rely on (i) explicit measurement of depth or (ii) appearance features

* This work is partly funded by the Department of Science and Technology, Govt. of India, under Grant DST/INT/NL/Biomed/P(3)/2011(G).

P. Golland et al. (Eds.): MICCAI 2014, Part I, LNCS 8673, pp. 747–754, 2014.

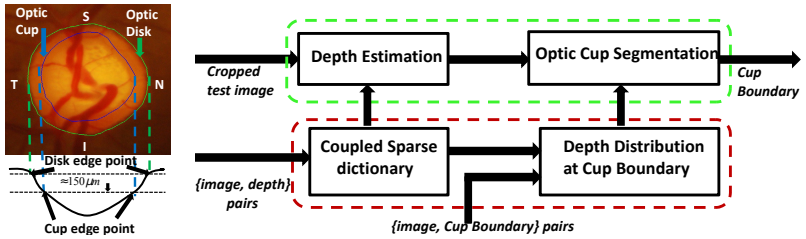

Fig. 1. (a) Sample OD sub-image with a sketch of corresponding depth. (b) System for cup segmentation. Top green box - online system; bottom red box - offline system.

which provide depth cues. In (i), depth information is obtained either from 3D imaging modalities such as OCT [4] or derived using stereo image pairs [5] [1]. Though better suited for accurate OCB extraction, the availability, portability and cost of these imaging devices inhibits their use in a glaucoma screening. In (ii) high level depth cues are extracted from single color fundus images based on pallor intensity and vessel bends [2] in addition to superpixels [3] and graph cut [6] based approaches. In absence of depth information, these features are susceptible to shape, color variabilities and indistinct OCB. Further, sparse distribution of vessels in the nasal, temporal sides and occurrence of vessel bends at non-OCB locations makes its detection challenging.

To deal with above mentioned challenges, we propose a *depth* based OCB extraction framework from *single* color fundus image per eye in which multiple depth estimates from shading, color and texture gradients are extracted from the image and correlated with OCT based depth values using a coupled sparse dictionary pre-trained on a set of image-depth pairs. Finally, OCB is extracted using a novel contour point detection based MRF formulation defined on the depth map to model the relative depth and discontinuity at OCB while reducing computation by lowering the number of sites to be labeled. We leverage the fact that in a clinical setting, OCT and fundus imaging are possible while for screening, only fundus imaging may be possible in the field. While supervised methods for estimating depth from single images have been recently used in computer vision [7] [8], such strategy remains unexplored in the medical domain.

2 Methodology

A square region around the OD center is automatically extracted using a Hough transform based OD localization [2]. The region is aligned based on symmetry in vessel density (in the nasal-temporal and superior-inferior regions) and resized to a standard size of 393×393 pixels. The proposed method shown in Fig. 1(b) comprises of 2 stages: depth map estimation from single image per eye (sec. 2.1) and OCB extraction from the estimated depth map (sec. 2.2).

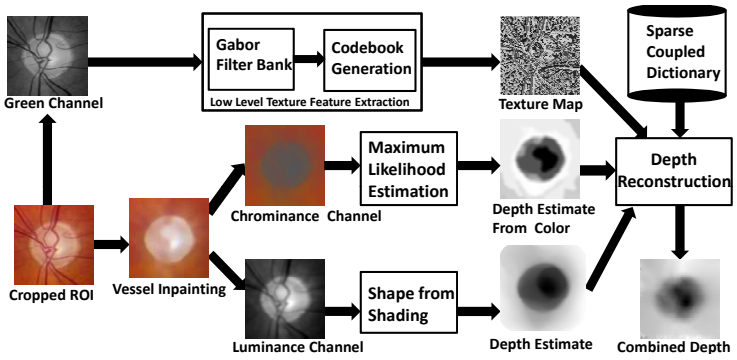

Fig. 2. Block diagram of the proposed supervised depth estimation

2.1 Depth Estimation

As noted in [1], different clinical studies have defined OCB as 50 microns below the retinal surface, $\frac{1}{2}$ or $\frac{1}{3}$ drop in depth from the OD edge to the deepest point. This indicates that *relative depth estimates rather than absolute depth values define the OCB*. Hence, we obtain depth maps defined up to an arbitrary scale factor rather than absolute depth values.

The proposed depth estimation (Fig. 2) comprises of extraction of chrominance (C), luminance (L) and texture word map (T) features, followed by individual depth map estimates d_c and d_l from C and L respectively. d_l and d_c suffer from inaccuracies due to the lack of 1-1 correspondence between C and d, simplified assumptions in shape from shading (SFS)and treating each pixel independent of its neighborhood. To obtain a more accurate and robust depth map, 8×8 image patches are extracted and represented in a feature space P_{cca} and correlated to OCT based depth estimates Q_{cca} using coupled sparse dictionaries U and V to obtain the final depth map. Details are provided below.

Feature Extraction: (a) T : 30 energy responses of a Gabor filter bank (6 orientations, 5 scales) [9] along with their 1^{st} and 2^{nd} order derivatives along two directions ($30 \times (2 + 2) = 120$) are combined to obtain a 150-D feature vector for each pixel which is clustered (during training) into 60 words. Each pixel is represented by the nearest word index [10] to obtain T. For remaining features, diffusion based inpainting is first applied to suppress vessels.(b) C : At each pixel, the color values r,g,b are normalized as $j/(r+g+b)$; j is r,g,b respectively to obtain 3-D feature C. (c) L: The luminance information is obtained by suppressing high color gradients in intensity channel using [11].

Individual Depth Estimates: d_l is obtained from L by complementing the output from a simple but fast SFS algorithm [12]. d_c is obtained from C using a supervised approach;for each depth value $d \in [0, 255]$, $P(C \mid d)$ is learnt from a training set of image-depth pairs, using a 3-D Gaussian Mixture Model with number of Gaussians selected in the range 1-6 that maximizes the Akaike

information criterion. During testing, each pixel is assigned a d that maximizes $P(d \mid C)$ using maximum a posteriori estimation.

Dimensionality Reduction: In the training phase, each pixel is represented in 2 feature spaces: (a) $F \in R^7$ obtained from the image by concatenating d_l,d_c, their gradients $\frac{\partial d_l}{\partial x}$, $\frac{\partial d_l}{\partial y}$, $\frac{\partial d_c}{\partial x}$, $\frac{\partial d_c}{\partial y}$ and T, (b) $G \in R^3$ computed from the ground truth depth maps comprising of the depth value d_{oct} along with it's gradients $\frac{\partial d_{oct}}{\partial x}$,$\frac{\partial d_{oct}}{\partial y}$ at the pixel position. The dimensionality of F and G are reduced (while maximizing the correlation between them) using Canonical Correlation Analysis [13]. This yields $P_{cca} = \phi_{img}^T.F$ and $Q_{cca} = \phi_{depth}^T.G$, where $\phi_{img} \in R^{7\times3}$ and $\phi_{depth} \in R^{3\times3}$ (x^T denotes transpose of x) are the canonical factors. Now, each pixel can be represented separately in P_{cca}, $Q_{cca} \in R^3$ spaces extracted from image and depth map respectively. Only P_{cca} is computed using pre-trained ϕ_{img} during testing.

Coupled Sparse Dictionary Training: From each image-depth pair in training dataset, 8×8 overlapping patches(1 pixel apart) are extracted and represented in 2 feature spaces: $P, Q \in R^{192}$ obtained by concatenating the 3-D features of each of the 64 pixels of the patch in P_{cca} and Q_{cca} space respectively.The objective is to learn two, overcomplete (each consisting of 1100 basis vectors), coupled sparse dictionaries U and V in the P and Q feature space, such that the same sparse code α is shared in the two representations: $P = U.\alpha$ and $Q = V.\alpha$ for all the training patches. This is done by concatenating the corresponding vectors [14][8] for each training patch, $Z = \{z_i = (p_i^T, q_i^T)^T\}$, $p_i \in P$ and $q_i \in Q$. An online dictionary learning algorithm[1] [15] was used for learning the sparse dictionary $W \in R^{384\times1100}$ from the feature set Z using batch size of 600, sparsity coefficient $\lambda = 0.6$ and max-iteration= 800. The learnt basis vectors W was split into U and V by taking the first and last 192 rows of W.

Coupled Sparse Dictionary Testing. Once U and V are learnt, given a new test image, its representation P_{test} can be extracted from the image and sparse code α^* can be estimated by solving the LASSO problem. The desired Q_{est} is then obtained by projecting α onto the depth basis V [16].

$$\alpha^* = argmin_\alpha ||U.\alpha - P_{test}||_2^2 \quad s.t. \quad ||\alpha||_1 \leq \lambda \tag{1}$$

$$Q_{est} = V.\alpha^* \tag{2}$$

After reconstructing the 3-channel D' from Q_{est}, we backproject it to obtain $D_{est} = (\phi_{depth})^{-1}.D^{'}$, D_{est} consisting of the depth value d and its gradients $\frac{\partial d}{\partial x}$ and $\frac{\partial d}{\partial y}$ at each pixel. The refined depth value is taken as the average of d and the depth estimated from $\frac{\partial d}{\partial x}$ and $\frac{\partial d}{\partial y}$ using inverse gradient methods [11].

2.2 Optic Cup Boundary Extraction

The depth map computed by stage 1 is used to extract its boundary as described next. Consider a circular ROI centered at $r_0 = (x_0, y_0)$ with radius R. We denote

[1] Implemented using SPAMS available at `http://spams-devel.gforge.inria.fr/`

the depth profile along a ray in direction θ_j from r_0 by $d_j(r)$, where $r \in [0, R]$. Let the closed curve B be the desired boundary, centred at r_0, required to be detected from the depth map. B is uniformly sampled to obtain an ordered set of boundary points $b_j; j = 0, ..J$. Note that this sampling is aligned to the orientation θ_j about r_0. We take a probabilistic approach to finding B by determining the likelihood that a point in $d_j(r)$, belongs to B. Let $B = \{b_j | 0 \leq j \leq (J)\}$, be a random field where each b_j is associated with $d_j(r)$, such that $P(b_j = r)$ represents the probability that $b_j = r$ is a boundary point on $d_j(r)$. The set of labels associated with b_j is $L = 0, 1...R$. Assuming a pairwise Markovian property that each b_i is only affected by its immediate adjacent neighbors, we define the Neighbour set $N = \{(b_j, b_{j+1}) | 0 \leq j \leq (J-1)\} \bigcup \{(b_J, b_0)\}$. We define a Markov Random Field based energy function $E(X)$ which is minimized to get the optimal labelling.

$$E(B) = \sum_{b_j \in B} D_{b_j}(b_j) + \lambda \sum_{(b_j, b_l) \in N} V_{b_j b_l}(b_j, b_l) \tag{3}$$

The data term $D_{b_j}(b_j)$ defines how well the labelling of b_j fits the probability distribution learnt from the boundary points on profile $d_j(r)$ from a set of training images. Both d and r values are normalized to $[0, 1]$ along each direction separately. We define $D_{b_j}(b_j)$ as

$$D_{b_j}(b_j = k) = 1 - P(k|b_j) \tag{4}$$

$P(k|b_j)$ represents the probability of assigning label k to the random variable b_j associated with profile $d_i(r)$. A Gaussian model is used to parameterize the distribution, since the $d_j(r)$ on cup boundaries tend to cluster around a single mode. The pairwise term $V_{b_j b_l}(b_j, b_l)$ captures the shape constraints in terms of the relative position of the boundary points in adjacent profiles.

$$V_{b_j b_l}(b_j = m_j, b_l = m_l) = 1 - P_j(|m_j - m_l|) \tag{5}$$

$|m_j - m_l|$ is the absolute difference in distance of the boundary points from r_0 in the adjacent profiles d_j and d_{j+1}. During training, the probability distribution (assumed to Gaussian) $P_j(|m_j - m_l|)$ between every adjacent boundary points is learnt from the ground truth data. Eqn.3 is solved using a multi-label graph cut approach with tree-reweighted message passing algorithm.

3 Experimental Results

Depth Estimation: The proposed depth estimation method was validated on 30 monocular color fundus images from INSPIRE dataset, obtained at a resolution of 4096×4096, cropped to a region of interest centered at OD. Corresponding depth maps obtained using SD-OCT scans in $200 \times 200 \times 1024$ mode [5] are available as ground truth. Due to limited availability of data, a leave-one-out cross validation analysis is done using correlation coefficient ρ to quantitatively measure the similarity in the overall trends of the estimated depth maps d and

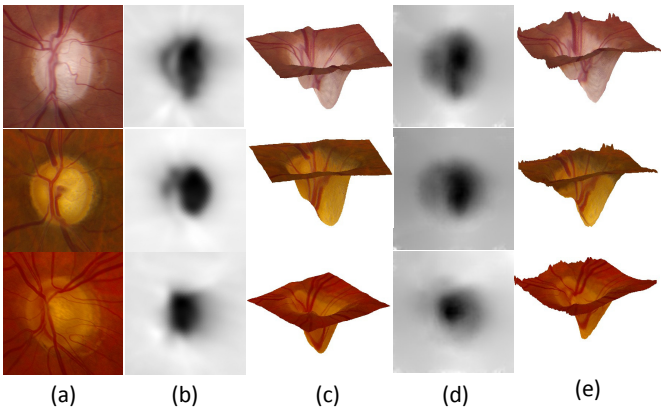

Fig. 3. OD regions from 3 sample images (column a) with corresponding depth estimates visualised as greysacle image where depth increases from white to black and topographical surface. Columns (b,c): ground truth; (d,e): computed results.

the ground truth D; $\rho(d,D) = \frac{\sum_m \sum_n (d_{m,n}-\bar{d})(D_{m,n}-\bar{D})}{\sqrt{(\sum_m \sum_n (d_{m,n}-\bar{d}))^2 (\sum_m \sum_n (D_{m,n}-\bar{D}))^2}}$ with $\bar{d}$ and $\bar{D}$ representing the mean values of d and D. m,n represents pixel locations in d and D. The mean ρ was found to be 0.80 ± 0.12.

Optic Cup Segmentation: The method for optic cup segmentation was validated on DRISHTI-GS dataset [17] which consists of 50 training and 51 test images obtained using 30 degree FOV at a resolution of 2896×1944. The ground truth OD and cup segmentation masks were obtained by a majority voting of manual markings by 4 ophthalmologists. Quantitative evaluation is based on the F-score to measure the extent of region overlap and the absolute pointwise localization error (measured in the radial direction) in the computed boundary as against the ground truth as reported in [2]. Both metrics were derived using a 10 fold cross validation approach, where for each of the 5 images, the parameters of the Gaussian distributions where learnt from the remaining 45 images. Performance evaluation on the test set is derived after training the system on an independent training set. The quantitative results have been provided in Table 1. Cup segmentation results are also reported for methods based on vessel bends [2], superpixels [3] and that provided along with the benchmark dataset in [17] for comparison. The tabulated figures show that the proposed method outperforms all the methods including [17] which relies on multiple input images. Results of cup segmentation on sample images are shown in figure 4.

Finally, we compare the results of the proposed method against OCT scanner provided values for CDR and cup to disc area ratios (CAR). Color fundus images for 28 eyes (18 Normal, 10 Glaucoma) were obtained at a resolution of 2896×1944 and 30 degree FOV along with OCT imaging whose reports provided ground truth CDR, CAR values. While the proposed method was used for cup segmentation, method in [2] was used for disc segmentation. The root mean square (rms) error for CDR was found to be 0.11 ± 0.08 for and 0.10 ± 0.05

Table 1. F-score and average boundary error in pixels

	Optic Cup			
	F-score		Boundary error(px)	
	Train	Test	Train	Test
R-bend[2]	0.74 ± 0.20	0.77 ± 0.20	33.91 ± 25.14	30.51 ± 24.80
Superpixel[3]	0.67 ± 0.12	0.63 ± 0.13	37.04 ± 16.96	41.00 ± 16.50
Multiview [17]	0.77 ± 0.17	0.79 ± 0.18	24.24 ± 16.90	25.28 ± 18.00
Proposed method	0.80 ± 0.18	0.81 ± 0.16	22.10 ± 19.47	21.21 ± 15.09

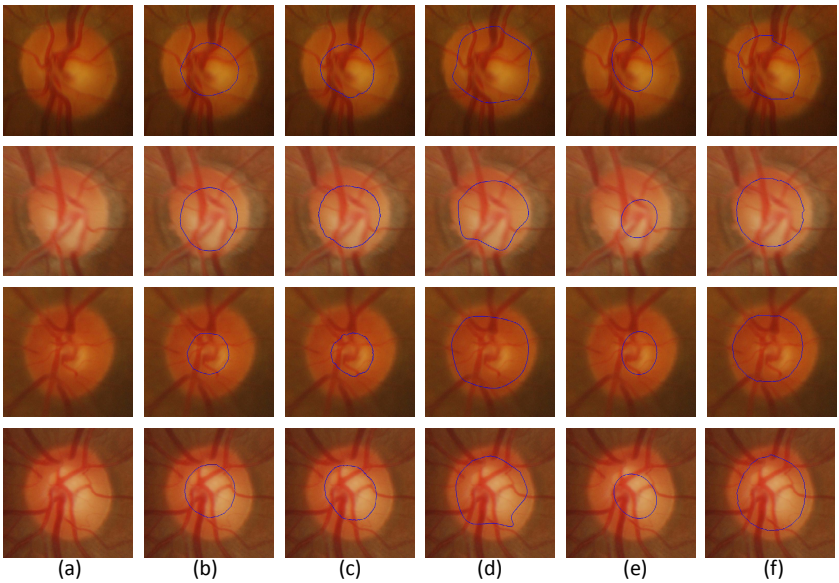

Fig. 4. Qualitative results; a: Input image ; b: ground truth cup marking; Cup boundaries computed using c: proposed method d: R-bend e: Superpixel and f: Multiview

while rms error in CAR was 0.17 ± 0.12 and 0.21 ± 0.11 for Normal and Glaucoma cases, respectively. While CDR error is uniformly low for both classes, it is marginally higher for the glaucoma class for CAR.

4 Discussion and Conclusion

Inspired by the clinical significance of depth information of the retinal surface in the OD region and its use for cup segmentation in 3D imaging modalities (like OCT, HRT) and stereo image pairs, we have proposed a novel, supervised method for depth-based cup segmentation. The method relies on a dictionary trained on fundus image-depth map pairs. Although exact estimation of depth from single view images is a highly underconstrained problem, the performance of the proposed method (avg. correlation coefficient of 0.8 against ground truth)

indicates that there is sufficient potential in the method. Since this was achieved with a moderate sized training set (30 pairs) it is possible to improve the results with a larger training set. Future work will explore ways to combine pallor and vessel kink information to further improve the reported results.

References

1. Xu, J., Chutatape, O., Sung, E., Zheng, C., Kuan, P.C.T.: Optic disk feature extraction via modified deformable model technique for glaucoma analysis. Pattern Recognition 40(7), 2063–2076 (2007)
2. Joshi, G.D., Sivaswamy, J., Krishnadas, S.R.: Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment. IEEE Trans. on Medical Imaging 30(6), 1192–1205 (2011)
3. Cheng, J., Liu, J., Tao, D., Yin, F., Wong, D.W.K., Xu, Y., Wong, T.Y.: Superpixel classification based optic cup segmentation. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part III. LNCS, vol. 8151, pp. 421–428. Springer, Heidelberg (2013)
4. Hu, Z., Niemeijer, M., Lee, K., Abràmoff, M.D., Sonka, M., Garvin, M.K.: Automated segmentation of the optic disc margin in 3-d optical coherence tomography images using a graph-theoretic approach. In: SPIE Med. Imaging (2009)
5. Tang, L., Garvin, M.K., Lee, K., Alward, W.L., Kwon, Y.H., Abràmoff, M.D.: Robust multiscale stereo matching from fundus images with radiometric differences. IEEE Pattern Anal. Mach. Intel. 33(11), 2245–2258 (2011)
6. Zheng, Y., Stambolian, D., O'Brien, J., Gee, J.C.: Optic disc and cup segmentation from color fundus photograph using graph cut with priors. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part II. LNCS, vol. 8150, pp. 75–82. Springer, Heidelberg (2013)
7. Saxena, A., Chung, S., Ng, A.: 3-d depth reconstruction from a single still image. Intl. Journal of Computer Vision 76(1), 53–69 (2008)
8. Agrawal, H., Namboodiri, A.: Shape reconstruction from single relief image. In: Asian Conf. on Pattern Recognition, pp. 527–531 (2013)
9. Kruizinga, P., Petkov, N.: Nonlinear operator for oriented texture. IEEE Trans. on Image Processing 8(10), 1395–1407 (1999)
10. Malik, J., Belongie, S., Shi, J., Leung, T.: Textons, contours and regions: Cue integration in image segmentation. In: ICCV. vol. 2, pp. 918–25 (1999)
11. Funt, B., Drew, M., Brockington, M.: Recovering shading from color images. In: Sandini, G. (ed.) ECCV 1992. LNCS, vol. 588, pp. 124–132. Springer, Heidelberg (1992)
12. Tsai, P., Shah, M.: Shape from shading using linear approximation. Image and Vision Computing 12, 487–498 (1994)
13. Weenink, D.: Canonical correlation analysis. In: Inst. of Phonetic Science, Univ. of Amsterdam, vol. 25, pp. 81–99 (2003)
14. Tang, Y., Yuan, Y., Yan, P., Li, X.: Single-image super-resolution via sparse coding regression. In: Intl. Conf. on Image and Graphics. pp. 267–272 (2011)
15. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online dictionary learning for sparse coding. In: Intl. Conf. on Machine Learning, pp. 689–696 (2009)
16. Vondrick, C., Khosla, A., Malisiewicz, T., Torralba, A.: Hoggles: Visualizing object detection features. In: Intenl. Conf. on Computer Vision, pp. 1–8 (2013)
17. Sivaswamy, J., Krishnadas, K., Joshi, G.D., Jain, M., Ujjwal, Abbas, T.S.: Drishti-GS: retinal image dataset for optic nerve head(ONH) segmentation. In: Int. Symp. on Biomed. Eng. (2014)