

# Glaucoma Classification with a Fusion of Segmentation and Image-based Features

Arunava Chakravarty and Jayanthi Sivaswamy

Center for Visual Information Technology, IIT Hyderabad, India

## ABSTRACT

Automated classification of glaucoma is of interest in early detection and treatment. Existing methods employ features which are either image-based or derived from Optic Disc (OD) and Cup (OC) segmentation. While the latter suffers from segmentation inaccuracies, the image-based features tend to overfit in limited availability of training data. We propose a solution to overcome these issues and present a classification framework that fuses both type of features within a co-training based semi-supervised setting to overcome the paucity of labelled data. A novel set of features is proposed to represent the segmented OD-OC regions. Additionally, features based on Texture of projections and color Bag of Visual Words have been designed to be sensitive to the sector-wise deformations in OD. The proposed method was trained on 386 labelled and 717 unlabelled images. It outperformed existing methods with an accuracy and AUC of 73.28%, 0.79 on a private test set of 696 unseen images and 76.77%, 0.78 when cross-tested on DRISHTI-GS1 dataset.

**Index Terms**— Co-training, Glaucoma, Retina, Fusion.

## 1. INTRODUCTION

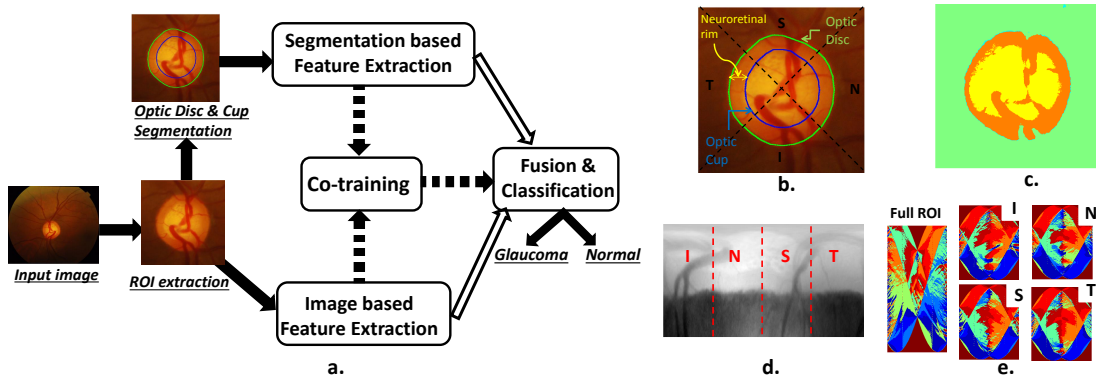
Glaucoma is an optic neuropathy caused by a gradual and progressive degeneration of the retinal nerve fibers. Large scale population screening programs are necessary to detect it at an early stage and inhibit its progression through proper treatment, which otherwise leads to irreversible blindness. An automated screening tool can aid in reducing the time and effort of ophthalmologists by filtering out a large proportion of normal images with high sensitivity. Glaucoma is primarily characterized by the structural deformations in Optic Disc (OD) of retina which appears as a bright elliptical region (Fig.1 b.) in color fundus images. The central depression in OD, called the optic cup (OC), is surrounded by an annular region known as the neuro-retinal rim where the retinal nerve fibers connect to the brain. In glaucoma, the loss of retinal nerve fibers leads to the enlargement of OC and clinical parameters such as the vertical Cup to Disk diameter ratio (CDR) and the ISNT rule [1] have been defined to quantitatively assess the OD region.

Existing work on automatic assessment of glaucoma from fundus images either employ image-based features or features

derived from segmentation of OD and OC for glaucoma classification. The latter aids in capturing structural deformation and CDR estimation is commonly employed. Most of these methods therefore report mean CDR estimation error, while a few methods [2] also report on the glaucoma classification performance. Only two methods [3], [4] have considered additional features such as the disc size, rim thickness and ISNT rule compliance. Attempts at CDR estimation have also been made based on sparse representation [5] without explicit segmentation. Though CDR is an important indicator for glaucoma, it is inadequate for accurate classification as it captures the retinal nerve fiber degeneration only along the vertical direction. Additionally, small errors in segmentation accuracy can lead to significant changes in measurements. While additional cues such as genetic information [6] have significantly improved classification performance, obtaining them is infeasible in a screening scenario.

Attempts to extract discriminative features at the pixel-level from the OD region without requiring OD-OC segmentation have also been reported. Some of the features that have been explored include: higher order spectra features [7]; energy of channels obtained from first level of wavelet decompositions [8]; and PCA coefficients of intensity, Fourier, and Gabor features [9]. A major challenge in these methods is the need to handcraft appropriate features. Preliminary results of these methods show high performance, albeit on small datasets of 60 to 120 images [8][7],[9]. Being data-driven in nature, feature selection and classification may tend to overfit due to the limited availability of annotated images. In this paper, we propose a novel classification framework that combines features derived from segmentation as well as the image as a whole. For the former, we propose to extract a comprehensive set of features from OD-OC segmentations to accurately capture the shape deformations characterizing glaucoma. For the latter, we propose to use the Texture of Projection features [10] to capture shape and textural changes and Bag of Visual words to capture color information, to represent the sector-wise structural changes in OD. In contrast to the existing methods, classification accuracy in the proposed framework is achieved by (i) combining segmentation and image-based features as they provide complementary and independent information and (ii) adopting a co-training based semi-supervised strategy to leverage the wider availability of unannotated images to automatically label a large set of unan-

Arunava is supported by the TCS Research scholar program.



**Fig. 1.** **a.** Block Diagram of the proposed method, filled arrows:common to training and testing; dashed arrows: for training only; unfilled arrows: testing only. **b.** OD and OC boundaries in a cropped fundus image; **c.** Color based clustering of **b.** used for BoW computation; **d.** polar representation of the red channel of **b.**; **e.** LBP of Radon transform of **d.** for ToP computation.

notated data during training.

## 2. METHODOLOGY

An overview of the proposed method is provided in Fig 1.a. Given, an image, illumination correction is done using a method similar to [9] followed by OD detection using Hough Transform based template matching [11] to extract a square Region of Interest (ROI) around the OD. The left-eye images are vertically flipped and all ROIs are resized to  $401 \times 401$  pixels. While the ROI is directly used to extract the image-based features, OD-OC segmentations are obtained using [12] to extract the segmentation based features. During training, the two feature spaces are used in a co-training based semi-supervised setting to annotate unlabelled images to increase the training dataset. Finally, a fusion strategy is used to combine the features and obtain the final classification. We will report on a few fusion strategies.

### 2.1. Extraction of Segmentation based features

We begin with some observations. The OD is only roughly circular. In clinical exams, disc size, variation in rim thickness and presence of localised deformation called notching are taken into account in glaucoma diagnosis. The guideline used for the former is the ISNT rule [1] by which, in the normal case, the rim is thickest in the I (inferior) and thinnest in the T (temporal) sectors. These sectors are shown on a sample image in Fig. 1b. Notching is of interest clinically, only in the Inferio-nasal( $I_N$ ), inferio-temporal( $I_T$ ), superio-nasal( $S_N$ ), superio-temporal( $S_T$ ) quadrants. Hence, we consider a wide set of measurements from a given image and construct a 14 dimensional (D) feature. In the following,  $r$  denotes the average rim thickness and  $\sigma$  denotes the standard deviation. The extracted features are as follows. a) *Rim-Disc Ratios (2D)*: vertical CDR and rim to disc area ratio; b) *Disc size (1D)*: the major axis length of OD; c) *Rim profile (7D)*:

the ratio of horizontal to vertical CDR along with 6 features computed by  $\frac{r_i - r_j}{\sigma_i + \sigma_j}$  with  $i \neq j$  and  $(i, j) \in \{(I, S, N, T)\}$  d) *Notching (4D)*:  $\sigma(r_k)$  with  $k \in \{I_N, I_T, S_N, S_T\}$ . Each feature was normalized to zero mean and unit standard deviation.

### 2.2. Extraction of Image based features

Texture of Projection (ToP)[10] and color based Bag of Words (BoW) features [13] were used to capture the color and the structural changes in the ROI. ToP is computed by extraction of Local Binary Pattern features (LBP) on the Radon transform  $R$  of an image patch. It has been shown to be a good local descriptor encoding the shape and structure of a patch. Since each projection in  $R$  is a ray sum and it is computed over a single color channel, the BoW features are included to provide additional complementary information. As both ToP and BoW are histogram based, the spatial information in the ROI will be lost. Whereas, we wish to capture the changes in bilateral symmetry and rim thickness with these features. Hence, the ROI was divided into I,S,N,T sectors (Fig. 1b) and ToP and BoW were computed for each sector as well as the entire ROI. Histograms of all sectors in ToP as well as LBP space were normalized to sum to 1. The final feature set extracted is: a) *Structural ToP features(65D)*: The red channel of the ROI was converted into polar coordinates to obtain rectangular-shaped regions for each sector (see Fig. 1d). ToP features followed by PCA based dimensionality reduction to 13D was computed on the entire ROI and each of the 4 sectors individually (Fig 1 e.) finally resulting in a  $13 \times 5 = 65D$  feature. b) *Color BoW features (25D)*: Each pixel in the ROI was represented by a 2D feature vector of its color values in the R and L\* channel of RGB and L\*a\*b\* color models respectively. The 2D features were clustered into 5 visual words using k-means (Fig 1 c.) and their histogram was computed for the entire ROI and the 4 sectors independently.

### 2.3. Co-training

Training with a large amount of expert-annotated images is desirable since training on small datasets tend to overfit. However, large training sets are not very practical in medical domain as it is expensive and time consuming to obtain annotations. Nevertheless, unannotated images are relatively easier to obtain. Assuming that segmentation and image based features provide two complementary and independent information about the ROI, with each capable of glaucoma classification independently, a solution for generating a large training set is co-training [14]. This technique permits the set of unannotated data to be automatically annotated in a semi-supervised setting. Given  $L$ , a set of retinal images with image-level diagnosis of glaucoma (-1 for normal and +1 for glaucomatous) and  $U$ , a set of unannotated retinal images, each image  $l \in L$  and  $u \in U$  can be represented in the 14D segmentation-based feature space  $\psi$  and 90D image-based feature space  $\Omega$  respectively. Co-training is a semi-supervised training framework that lets two classifiers  $C_1(s) \rightarrow \{+1, -1\}$  and  $C_2(i) \rightarrow \{+1, -1\}$  defined on  $s \in \psi$  and  $i \in \Omega$  spaces respectively to teach each other by iteratively labelling samples from  $U$ . In each iteration, these classifiers are trained on samples in  $L$  and tested on  $U$ , in  $\psi$  and  $\Omega$  feature spaces respectively. The most confident predictions of  $C_1$  and  $C_2$  (defined by samples with maximum margin from the separating hyperplane for SVM classifier) for both +1 and -1 classes are removed from  $U$  and added to  $L$  with their predictions as labels, thus labelling 4 samples from  $U$  in each iteration. The process is repeated until all samples in  $U$  have been labelled.

### 2.4. Fusion and Classification

Accurate classification of retinal images into Normal or Glaucomatous class requires an effective fusion strategy to combine the information in  $\psi$  and  $\Omega$ . The two feature spaces can be fused at the feature level (early), decision level (late) or within a classifier (intermediate). We studied all three strategies in the context of SVM classifiers whose good performance on glaucoma classification has been established in [9],[8]. In early fusion, a single classifier is learnt on a (14+90=) 104D feature space obtained by concatenation of  $\psi$  and  $\Omega$ . In late fusion, the probabilistic predictions (obtained by Platts scaling) of two separate classifiers  $C_1, C_2$  learnt on  $\psi$  and  $\Omega$  respectively, are linearly combined by the final classifier  $C(s_k, o_k) = w_1 \times C_1(s_k) + w_2 \times C_2(o_k)$  where  $o_k \in \Omega, s_k \in \psi$  describes the  $k^{th}$  sample. The training data was divided into two parts; two-thirds used to train  $C_1, C_2$  and the remaining third to learn  $w_1, w_2$  by training a separate linear SVM,  $C$  on the predictions of  $C_1, C_2$ . Multiple kernel learning (MKL) can be considered as an intermediate fusion strategy where two separate RBF kernels  $\phi_1(s)$  and  $\phi_2(o)$  defined on  $\Omega$  and  $\psi$  respectively, are linearly combined to obtain the kernel  $\phi([s_k, i_k]) = \alpha_1 \times \phi_1(s_k) + \alpha_2 \times \phi_2(o_k)$

**Table 1.** Performance of the proposed glaucoma assessment system. Best results are shown in bold.

|                      | Training     | Testing      |             |      |             |             |
|----------------------|--------------|--------------|-------------|------|-------------|-------------|
|                      | 5-fold Acc.  | Acc.         | AUC         | SN   | SP          | PPV         |
| Seg. Features alone  | 79.56        | 72.70        | 0.76        | 0.80 | 0.53        | 0.61        |
| Image Features alone | 80.56        | 69.11        | 0.77        | 0.80 | 0.59        | 0.64        |
| Early Fusion         | 85.56        | 73.28        | 0.79        | 0.80 | 0.60        | 0.65        |
| Intermediate Fusion  | <b>92.08</b> | <b>74.14</b> | 0.80        | 0.80 | 0.62        | 0.66        |
| Late Fusion          | 86.25        | 73.71        | <b>0.82</b> | 0.80 | <b>0.65</b> | <b>0.68</b> |

**Table 2.** Comparison of proposed method against existing methods.

|                              | Private Test Set |      | DRISHTI-GS1  |             |
|------------------------------|------------------|------|--------------|-------------|
|                              | Acc. %           | AUC  | Acc. %       | AUC         |
| Super-pixel Segmentation [2] | 58.48            | 0.65 | 70.13        | 0.71        |
| Wavelet features [8]         | 58.19            | 0.63 | 41.41        | 0.57        |
| Proposed Early Fusion        | 73.28            | 0.79 | <b>76.77</b> | <b>0.78</b> |

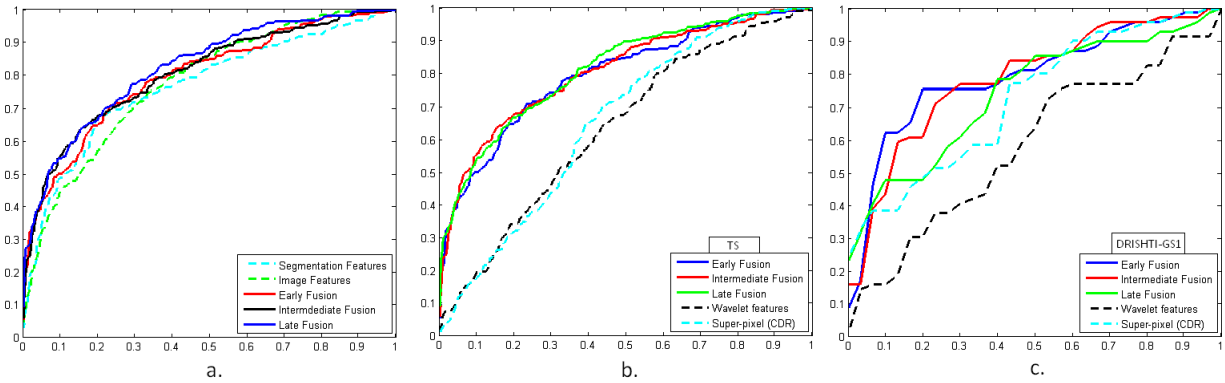
for the combined SVM classifier.  $\alpha$  was obtained using the  $L_p$ -norm multiple kernel learning in [15] with  $p=2$ .

## 3. EXPERIMENTS AND RESULTS

**Dataset:** The proposed system was trained using a set of 386 annotated images ( $L$ ) of which 201 were of Normal, 185 glaucomatous eyes and 717 unlabelled images ( $U$ ). A separate test set (TS) of 696 annotated images (360 Normal and 336 glaucomatous) was also collected. The images were acquired from the glaucoma service center of a local eye hospital. The images were 30 degree field of view, OD-centric and of size  $1900 \times 1600$  pixels. A majority consensus of image level diagnosis obtained from 3 glaucoma experts was used as the Ground Truth for the annotated training and test images. Additionally, the cross-training performance of the proposed method was also evaluated on a publicly available dataset DRISHTI-GS1 [16] comprising of 101 images (31 Normal, 70 glaucomatous) acquired at a different imaging setting than our training set with a resolution of  $2896 \times 1944$  pixels.

**Evaluation of the proposed method:** In all experiments, non-linear SVMs with RBF kernels were employed with the exception of the early fusion strategy for which the optimal performance was obtained using a linear SVM. This may be attributed to the increase in the dimensionality of the concatenated feature space. The classification performance of the segmentation and image based features and their fusion (with different strategies) is reported in Table 1. Corresponding ROC plots are presented in Fig. 2 a. Since high sensitivity (SN) is critical in disease screening, specificity (SP) and Positive Predictive Value (PPV) are reported at the operating point of SN = 0.8. The biggest improvements due to fusion are in SP (up to 23%) and PPV (10%). Accuracy is not improved much by fusion, except in the case of image features (7%).

**Comparison with existing methods** The performance of the proposed method is compared against two state of the art



**Fig. 2.** ROC plots to compare the performance of three Fusion strategies against: individual segmentation and image based features on private test set(TS) in **a.**; existing CDR [2] and wavelet [8] based methods on TS in **b.** and DRISHTI-GS1 in **c.**

methods, [2] based on CDR and [8] based on image based features in Table 2. Apart from the privately collected TS, the methods were also cross-validated on the public dataset DRISHTI-GS1 [16] *without* retraining the system. Since early fusion is the simplest strategy, we used it as a baseline to compare with other methods. The tabulated results show our method outperforms [8] and [2] on both the private and public datasets. This underscores the importance of using both segmentation and image based features in glaucoma assessment. The reported results in Tables. 1 and 2 were obtained with co-training to label the unlabelled images and training each classifier on the resultant larger training set.

The segmentation and image-based features alone had a 32% disagreement in classification indicating that they provide complementary information. Co-training marginally improved the classification accuracy in all three fusion methods: by 4.03% for Early, 1.3% for Intermediate and 5.61% for late fusion which used two-thirds of the data to learn the individual classifiers and the remaining for the final classifier.

#### 4. DISCUSSION AND CONCLUSION

While existing methods largely rely on CDR alone, we proposed a framework to combine segmentation and image based features for glaucoma classification. Our contributions are multi-fold: i) design features to describe the segmented OD and OC regions; ii) design image based features based on the ToP features and color-based BoW features. iii) co-training to leverage availability of large amount of unlabeled data, alleviating the problem of acquiring large labelled training datasets. The effectiveness of these features was seen in the classification performance: with (i) alone the improvement in accuracy over the CDR based classification [2] was  $72.70 - 58.48 = 14.22\%$  on the Test set; while with (ii) alone the improvement in accuracy over the wavelet based features [8] was  $69.11 - 58.19 = 10.92\%$ . Fusion of the proposed features along with co-training was shown to further improve on

the classification performance.

#### 5. REFERENCES

- [1] N. Harizman et al., "The isnt rule and differentiation of normal from glaucomatous eyes," *Archives of ophthalmol.*, vol. 124, no. 11, pp. 1579–1583, 2006.
- [2] J. Cheng et al., "Superpixel classification based optic disc and optic cup segmentation for glaucoma screening," *IEEE TMI*, vol. 32, no. 6, pp. 1019–1032, 2013.
- [3] J. Nayak et al., "Automated diagnosis of glaucoma using digital fundus images," *Jrnl. of medical systems*, vol. 33, no. 5, pp. 337–346, 2009.
- [4] Z. Zhang et al., "Mrmr optimized classification for automatic glaucoma diagnosis," in *IEEE EMBC*, 2011, pp. 6228–6231.
- [5] J. Cheng et al., "Sparse dissimilarity-constrained coding for glaucoma screening," *IEEE TBME*, vol. 62, no. 5, pp. 1395–1403, 2015.
- [6] J. Liu et al., "Multiple modality fusion for glaucoma diagnosis," in *ICHI*, 2014, pp. 5–8.
- [7] R.U. Acharya et al., "Automated diagnosis of glaucoma using texture and higher order spectra features," *IEEE Trans Inf Technol Biomed.*, vol. 15, no. 3, pp. 449–455, 2011.
- [8] S. Dua et al., "Wavelet-based energy features for glaucomatous image classification," *IEEE Trans Inf Technol Biomed.*, vol. 16, no. 1, pp. 80–87, 2012.
- [9] R. Bock et al., "Classifying glaucoma with image-based features from fundus photographs," in *Pattern Recognition*, pp. 355–364, 2007.
- [10] N.V. Medathati et al., "Local descriptor based on texture of projections," in *ICVGIP*, 2010, pp. 398–404.
- [11] G.D. Joshi et al., "Optic disk and cup boundary detection using regional information," in *ISBI*, 2010, pp. 948–951.
- [12] A. Chakravarty et al., "Coupled sparse dictionary for depth-based cup segmentation from single color fundus image," in *MICCAI*, pp. 747–754, 2014.
- [13] G. Csurka et al., "Visual categorization with bags of keypoints," in *Workshop, ECCV*, 2004, vol. 1, pp. 1–2.
- [14] A. Blum et al., "Combining labeled and unlabeled data with co-training," in *COLT*, 1998, pp. 92–100.
- [15] M. Kloft et al., "Lp-norm multiple kernel learning," *JMLR*, vol. 12, pp. 953–997, 2011.
- [16] J. Sivaswamy et al., "A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis," *JSM Biomedical Imaging Data Papers*, vol. 2, no. 1, 2015.