Abstract

Optic nerve head (ONH) segmentation problem is of interest for automated glaucoma assessment. Although various segmentation methods have been proposed in the recent past, it is difficult to evaluate and compare the performance of individual methods due to a lack of a benchmark dataset. The assessment involves segmentation of optic disk and cup region within the ONH. In this paper, we present a comprehensive dataset of retinal images of both normal and glaucomatous eyes with manual segmentations from multiple human experts. The dataset also provides expert opinion on an image representing a normal or glaucomatous eye and on the presence of notching in an image. Several state of the art methods are assessed against this dataset using cup to disc diameter ratio (CDR), area and boundary-based evaluation measures. These are presented to aid benchmarking of new methods. A supervised, notch detection method based on the segmentation results is also proposed and its assessment results are included for benchmarking.

ABBREVIATIONS

ONH: Optic Nerve Head; OD: Optic Disk; CDR: Vertical Cup-to-disk-Diameter Ratio; CFI: Color Fundus Image

INTRODUCTION

Glaucoma, a degenerative optic neuropathy, is the second leading cause of blindness in the world, accounting for 12.3% of the total blindness worldwide [1]. It is a chronic sight-threatening disease caused by a progressive and irreversible degeneration of optic nerve fibers. This manifests as structural changes in the optic nerve head which consists of an elliptic region called the optic disk (OD) with a central depression called the cup.

In glaucoma enlargement of optic cup occurs in all directions with a regional preference based on the stage of the disease. Inferior region is generally affected first, followed by superior then temporal and nasal region is the last to be affected [2]. An overall enlargement of cup size is called cupping while a focal enlargement of cup leads to notching. Since inferior and superior tissues are affected first, notching is typically found in one of these quadrants, often causing the optic cup to enlarge in a vertical or oblique direction with respect to the optic nerve head. Vertical cup-to-disk diameter ratio (CDR) is a common measurement used to quantify such deformations in the OD [3].

A high value of CDR indicates the presence of glaucoma in an eye. Figure 1 shows the marked difference between a normal and a glaucomatous eye which is reflected in the corresponding CDR values. Different imaging modalities such as optical coherence tomography, Heidelberg retina tomography, colour fundus image (CFI), etc are used to assess glaucoma. Of these, CFI is a low-cost, non-invasive and most preferred imaging modality for mass screening of retinal diseases. A CFI is a true color, two dimensional projection of the retina.

In order to quantify the amount of cupping and notching, indicative parameters of glaucoma such as CDR need to be estimated from OD and cup segmentation. Therefore, any automated computer aided diagnosis system for glaucoma must accurately segment the OD and the cup regions from retinal image. Irregular disk shape, ill-defined boundaries of OD region and variable imaging conditions make the problem of OD segmentation challenging. A number of methods have been proposed for OD segmentation [4-6] Techniques for OD segmentation are fairly mature and are capable of providing reliable OD boundaries. Recently, two standard datasets have...
In contrast, cup segmentation is still a challenging task as depth is the best marker (which is lost due to 2D projection in retinal image) and recently several attempts [6,9-11] have been made to improve its performance. There is also high inter-observer variability seen in manual markings of the cup boundary by human experts. To the best of our knowledge, there is no existing public dataset which caters to cup segmentation problem.

Hence, availability of a public dataset that provides cup and OD markings is highly desired. In contrast to OD and cup segmentation, automatic detection of notching has not received any attention with [12] being an exception.

In this paper, we present a reference dataset DRISHTI-GS1 consisting of 50 training and 51 testing images. For each image, manual segmentations are collected for both OD and cup region from four different human experts with varying clinical experience. Markings were collected with a dedicated marking tool which provides a fully deformable circle to enable capturing different OD and cup shapes, along with any localized shape changes such as notching.

DRISHTI-GS1 is an extension of DRISHTI-GS, a dataset recently made publicly available by us [13] for OD and cup segmentation algorithms. In addition to the structural markings provided in [13], two other expert opinions are included. These are decisions on i) the image as representing a Normal or Glaucomatous eye and ii) presence or absence of notching in the inferior and/or superior sectors of the image. Since the dataset is aimed at benchmarking different segmentation algorithms, results of OD and cup segmentation methods [6,9,11] tested on this dataset are also provided. A new supervised method for notch detection based on the OD and cup segmentations is proposed. The method was evaluated on the dataset and the results are also presented.

MATERIALS AND METHODS

In this section, we summarize the existing datasets for ONH segmentation and present a detailed description of the presented dataset.

A. Existing Datasets

There are relatively few public datasets for glaucoma assessment as against a large number of datasets for diabetic retinopathy [14-16] and vessel segmentation [17,18]. Public datasets like Rim-one [8] and Drions-db [7] cater only to OD segmentation. Rim-one has 169 images (majority of them normal) with 5 manual boundary markings for each one. Drions-db has 110 images with 2 manual markings for each image. OD and cup segmentation are both important parts of ONH segmentation and together they form a base for glaucoma assessment. A reference dataset to evaluate cup segmentation methods was announced in [19]. However, it is not available for free, public access.

B. Image Acquisition

The dataset DRISHTI-GS1 consists of a total of 101 images. It is divided into 50 training and 51 testing images. Ground truth is provided for the training set whose details are provided in Sub-section C.

All images were collected at Aravind eye hospital, Madurai from visitors to the hospital, with their consent. Glaucoma patient selection was done by clinical investigators based on clinical findings during examination. Selected patients were 40-80 years of age, with roughly equal number of males and females. Patients undergoing routine refraction test and not found to be glaucomatous were chosen to represent the normal class.

All images were taken with the eyes dilated using the following data collection protocol: centered on OD with a Field-of-View of 30-degrees and of dimension 2896 x 1944 pixels and PNG uncompressed image format. Apart from these, no other imaging constraints were imposed on the acquisition process. For each image, ground truth was collected from four glaucoma experts with experience of 3, 5, 9, and 20 years, respectively, to capture inter-observer variance in marking. Bad quality images in terms of poor contrast, positioning of OD region, etc. were discarded. In this dataset release, fundus region (image region having retinal structures) has been extracted from the original image by eliminating the surrounding non-fundus mask region, to obtain an image of approximately 2047 x 1760 pixels.
C. Ground truth Collection

A dedicated marking tool was developed to obtain boundary marking on images from human experts. This tool allows precise boundary marking for even irregular shapes of OD and cup regions. This was favored over approximating the boundary with some parametric shape such as ellipse, as local deformations in the cup are possible and need to be captured in the marking. A deformable circle with several free-to-move control points was provided to the user. These help the user to first position the circle close to the approximate region boundary and next do a better fit to the region by moving the individual control points. Thus, an accurate shape marking is possible in two steps: rough localization of the circle followed by local deformations to the circle. Figure 2 shows a screenshot of the tool wherein one point in the OD boundary is shown in pre-final position.

Additionally, diagnostic opinion on each image being Normal or Glaucomatous was obtained from the 4 glaucoma experts and a Gold standard was derived based on the majority opinion, i.e., 3 out of the 4 experts. Further, image level decision on presence of notching in the Inferior and Superior sectors was also obtained from the most experienced glaucoma expert.

The ground truth is released only for the 50 training images. In addition to the raw dataset published in this journal, a web page has been designed to allow interested research groups to register, download the test set and submit segmentation results (for OD and cup) on the same. The submitted results will be analyzed using evaluation measures presented in Subsection D and made publicly available on DRISHTI-GS web page (see Conclusion section for link).

D. Ground truth Format

The following information is provided as ground truth for each image in the dataset:

Region Boundary: Average boundaries for both OD and cup region are derived from the manually marked boundaries. The image region is divided into 80 equal angular sectors with respect to the disk center and the average boundary is obtained by averaging manual markings in the respective sector. The pixel locations \([x,y]\) corresponding to the averaged boundary of OD and cup are provided in a text file, namely, \(\text{ImageName}_{-}\text{RegionTypeAvgBoundary.txt}\). The OD center location is stored in a separate file named as \(\text{ImageName}_{-}\text{diskCenter.txt}\). Top row of Figure 3 shows four manual markings in black color and the obtained averaged boundary in red and green.

Segmentation Soft map: Manual markings obtained from four experts are fused together to generate a soft segmentation map \(B(p)\) for both the OD and cup regions. \(B \in [0,1]\) where 0 indicates no marking and 1 indicates total consensus, on \(p\) being a boundary point. Similarly, \(B \geq 0.75\) indicates agreement among at least three experts. For each image, segmentation soft maps are provided for both OD and cup region in .png image format. Bottom row of Figure 3 shows sample segmentation soft maps for OD and cup regions.

Cup-to-Disk Ratio: CDR values were determined based on the markings of the 4 experts and provided in a separate text file named as \(\text{ImageName}_{-}\text{cdrValues.txt}\). This information allows evaluation of the CDR estimation accuracy of an algorithm against each individual expert.

Image-level Decisions: A binary image-level decision on whether each image is normal or glaucomatous is provided based on the majority opinion (3 out of 4) of experts. Additionally, decision on the occurrence of notching in the superior, inferior, nasal or temporal sector of the optic nerve was assessed by a single expert. These image level decisions have been summarized.
for the training and test sets in two separate excel sheets named image_lvl_diagnosis_train.xls and image_lvl_diagnosis_test.xls respectively in which the ground truth information has been indexed by the image name. The distribution of the dataset is skewed with more stress on glaucomatous images as they present more challenges in terms of variations in the OD and Cup morphology. The distribution of Normal/Abnormal images and notching cases in the training and test set is provided in Table 1.

E. Evaluation Measures

Three different measures are proposed for the quantitative evaluation of a segmentation method. These are accuracies of a) region segmentation, b) boundary localization and c) CDR estimation.

- **Region Segmentation**
  
  Segmentation result is evaluated against binary ground truth and a single score is awarded to the method. The binary ground truth segmentation for an image can be obtained by thresholding the given segmentation soft map. In this paper, we present results against ground truth obtained with threshold set to 0.75. For a given segmentation, precision and recall values are computed as follows:

  \[
  \text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}
  \]

  where TP = count of true positive, FP = count of false positive and FN= count of false negative pixels. Next, the F-score (F), which is the harmonic mean of precision and recall, is computed. It is defined as:

  \[
  D = \frac{1}{N} \sum_{i=1}^{i=N} \sqrt{(d_{gi}^n)^2 - (d_{oi}^n)^2}
  \]

  This is also known as the F1 score, because recall and precision are evenly weighted. Maximum possible value of F-score is 1.

- **Boundary Localization**
  
  The distance between a computed region boundary and the ground truth is a metric for boundary localization. Let \( C_g \) and \( C_o \) be the ground truth and computed boundary by a method. The distance \( D \) (in pixels) between two curves is defined as:

  \[
  S_{\theta} = \frac{R_{\theta}}{r}
  \]

  where, \( d_g^n \) and \( d_o^n \) are the distance from disk center to points on \( C_g \) and \( C_o \), respectively, in the angular direction indexed by \( n \). N is taken to be 24 in our evaluation.

- **CDR Estimation**
  
  CDR is the ratio of vertical diameters of the OD and Cup regions. It ranges from 0-1 and represents the extent of glaucomatous damage in an eye. The CDR estimation error is the metric used to assess a segmentation method’s success in deriving clinically relevant disk parameters.

**BENCHMARKING RESULTS**

The DRISHTI-GS1 dataset has been used to assess some of the algorithms we have proposed earlier. These methods are summarized and results are provided for benchmarking purposes.

A. OD segmentation

The OD segmentation algorithm proposed in [6] uses a Circular Hough transform (on the gradient magnitude of the red channel of the image) to obtain an initial estimate of OD boundary. This is further refined using a novel, region-based active contour formulation. This minimizes the sum of local energy defined at each point \( b_i \) on the OD boundary. The local energy is computed in a circular neighborhood around \( b_i \), from a multi-dimensional feature representation for the neighbourhood. Further details can be found in [6]. The performance of this method on the training and test sets are presented in Table 2.

B. Cup Segmentation

Optic Cup segmentation from a retinal image is challenging as depth is the primary evidence for cup boundary which is lost in the 2-dimensional (2-D) projection of retinal images. The dataset was used to assess three recent methods [6,9,11] that use completely different strategies for cup segmentation.

First is a cup segmentation method based on vessel bends [6] detected from monocular color fundus images. The method is based on the anatomical knowledge that cup region being a depression within the OD, thin vessels while entering the optic cup tend to bend at its boundary. Though vessel bends can occur at any place within the optic disk, only a subset lie on the cup boundary. Color information is used to remove detected vessel

<table>
<thead>
<tr>
<th>Test</th>
<th>Train</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boundary localization error (pixels)</td>
<td>F-score</td>
</tr>
<tr>
<td>8.9±2.96</td>
<td>0.96±0.02</td>
</tr>
<tr>
<td>Boundary localization error (pixels)</td>
<td>F-score</td>
</tr>
<tr>
<td>8.6±1.889</td>
<td>0.96±0.05</td>
</tr>
</tbody>
</table>

The table entries are mean ± standard deviation obtained across images.
bends within the pallor region and the final cup boundary is obtained via a multi-stage robust curve fitting algorithm.

Second is a cup segmentation method which uses a relative motion framework to derive depth-discontinuity based evidences from multi-view images [9]. Multiple views of retina are first registered to obtain a displacement field. Locations of abrupt changes in the displacement field and regions with large registration errors indicate motion boundary and occlusion regions, respectively, which occur around the optic cup boundary. False positives occurring within the pallor region due to vessel occlusions are removed using color information and a robust sector-wise fitting algorithm is used to obtain the final cup boundary.

Third is a more recent dictionary-based method for depth-based cup segmentation [11]. This considers a setting where both single color fundus images and corresponding OCT based depth maps are available during training while only a single color fundus image is available while testing. During training, color, shading and depth information are extracted from the fundus images and related to the corresponding depth map using a coupled, sparse dictionary. Given an unseen color fundus image, a corresponding depth map is estimated from it using the pre-trained coupled sparse dictionary. An MRF based contour evolution strategy defined over the depth map is then used to obtain the final cup boundary. While the coupled sparse dictionary was trained over images of the INSPIRE dataset [20], the data term and pairwise potential of the MRF was trained on the training set provided in our dataset for benchmarking.

Further details on the above methods can be found in [6,9,11]. These methods were assessed on the training and test dataset and results are reported in Table 3. The CDR was determined based on OD segmentation [6] and cup segmentation [9] and is reported in Table 4. The computed CDR is compared against the individual markings of each of the 4 experts as well as the consensus based average Gold standard of the segmentations.

### C. Notch detection

We propose a new method for detection of notching from the OD and cup segmentation of the retinal images and evaluate it against the expert opinion. Let $B_\theta$ represent the OD boundary point at angle $\theta$ from the from the OD centre $c$. Similarly, let $C_\phi$ represent the cup boundary point from $c$. The rim thickness profile $R_\theta$ is computed in the range $[0, 360^\circ]$, in increments of $1^\circ$ as

$$R_\theta = d(B_\theta, c) - d(C_\phi, c)$$

where $d(a,b)$ is the Euclidean distance between $a$ and $b$.

In order to make $R_\theta$ scale invariant, it is normalized with respect to the length $r$ of the major axis of the segmented OD to obtain $S_\theta$ as

$$S_\theta = \frac{R_\theta}{r}$$

$S_\theta$ is divided into 4 sectors: $S_i$, $S_j$, $S_n$ and $S_t$ representing the

<table>
<thead>
<tr>
<th>Test</th>
<th>Boundary localization error (pixels)</th>
<th>F-score</th>
<th>Train</th>
<th>Boundary localization error (pixels)</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>30.51±24.80</td>
<td>0.77±0.20</td>
<td>33.91±25.14</td>
<td>0.74±0.20</td>
<td>[6]</td>
<td></td>
</tr>
<tr>
<td>25.28±18.00</td>
<td>0.79±0.18</td>
<td>24.24±16.90</td>
<td>0.77±0.17</td>
<td>[9]</td>
<td></td>
</tr>
<tr>
<td>21.21±15.09</td>
<td>0.81±0.16</td>
<td>22.10±19.47</td>
<td>0.80±0.18</td>
<td>[11]</td>
<td></td>
</tr>
</tbody>
</table>

The table entries are mean ± standard deviation obtained across images.

Figure 4 The Inferior (I), Superior (S), Nasal (N) and Temporal (T) sectors of the ONH region.

Figure 5 Receiver operating characteristic curve for notch detection on the training (Red) and test (Blue) sets.
Central difference rule. Inspired from the Mahalanobis distance, we define the profiles (5-D):

This feature has been designed to specifically max-min decrease in rim thickness in the sector, standard deviation and

Table 5: Error in CDR estimation using [6] for Optic disc and [9] for cup segmentation (mean±standard deviation) evaluated against different experts.

<table>
<thead>
<tr>
<th>Test</th>
<th>Train</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.18±0.14</td>
<td>0.15±0.12</td>
</tr>
<tr>
<td>0.17±0.11</td>
<td>0.13±0.10</td>
</tr>
<tr>
<td>0.12±0.12</td>
<td>0.10±0.10</td>
</tr>
<tr>
<td>0.14±0.12</td>
<td>0.11±0.11</td>
</tr>
<tr>
<td>0.16±0.02</td>
<td>0.12±0.02</td>
</tr>
</tbody>
</table>

CDR: vertical cup to disc diameter ratio

Table 5: Performance of proposed notch detection method.

<table>
<thead>
<tr>
<th>Area under curve</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Set</td>
<td>0.81</td>
<td>0.84</td>
</tr>
<tr>
<td>Test Set</td>
<td>0.79</td>
<td>0.81</td>
</tr>
</tbody>
</table>

normalized rim thickness in the inferior, superior, nasal and temporal sectors respectively (see Figure 4).

\[ S_i = \{ S_0 | 0^\circ \leq \theta \leq 45^\circ \text{ and } 316^\circ \leq \theta \leq 360^\circ \} \]

\[ S_s = \{ S_0 | 36^\circ \leq \theta \leq 225^\circ \} \]

\[ S_t = \{ S_0 | 46^\circ \leq \theta \leq 135^\circ \} \]

The distribution of rim thickness in normal images follows the ISNT rule which states that the rim thickness decreases along the sectors in the following order: Inferior>Superior>Nasal>Temporal. Due to notching, the rim thickness in the Inferior or Superior sectors decreases thereby disturbing the above distribution. Considering the above observations, we define a 17-D feature vector to represent the rim thickness profile as follows:

1. Statistical features to represent rim thickness profile in superior and inferior sectors (12-D): Maximum, minimum, mean, median, standard deviation and the difference in the maximum and minimum of the distribution is computed for \( S_i \) and \( S_s \) separately. While mean and median values capture the overall decrease in rim thickness in the sector, standard deviation and max-min values capture any focal rim thinning occurring in the sector.

2. Relative thickness of the Inferior and Superior depth profiles (5-D): This feature has been designed to specifically capture the change in thickness distribution from the ISNT rule. Inspired from the Mahalanobis distance, we define the difference \( \delta_j \) between the rim thickness in sectors \( j \) and \( k \) \((j,k \in \{i,s,n,t\})\) in terms of their means \( \mu_j \) and standard deviations \( \sigma_j, \sigma_k \) respectively as follows:

\[ \delta_{j,k} = \left| \frac{\mu_j - \mu_k}{\sigma_j + \sigma_k} \right| \]

\( \delta_{j,s}, \delta_{j,n}, \delta_{s,i}, \delta_{s,n}, \delta_{n,i}, \delta_{n,t} \) are evaluated as the feature to capture the relative thickness of the profiles in inferior and superior sectors with respect to the nasal and temporal sector.

To avoid overfitting on the training data due to small sample size, the dimensionality of the 17-D feature vector was reduced to 9-D before applying a linear SVM classifier. For evaluation, an image was considered to be a notching case if an expert had indicated notching to be present in either of the inferior and superior sectors. The proposed notch detection algorithm was assessed using segmentations of OD [6] and cup [11]. The ROC curves for the training and test data are shown in Figure 5. The sensitivity, specificity and area under the curve values are presented in Table 5.

CONCLUSION

This paper presents a dataset of retinal images for ONH evaluation. The ground truth derived from experts is comprehensive, providing boundary markings and opinion on the eye as normal/glaucomatous and presence/absence of notching. The dataset is intended to foster active research in image analysis and contribute towards developing robust solutions for automated glaucoma assessment. In addition to this open access data paper, the dataset is also available at the url: http://cvit.iiit.ac.in/projects/mip/dishti-gs/mip-dataset2/Home.php. Users can upload their results on the test images which will be evaluated and posted on the above link.

ACKNOWLEDGEMENTS

We would like to thank Dr. Kundan Karan, Dr. Prashanth R. and Dr. P.S. Vivek for providing the markings. This work was supported by the Department of Science and Technology, Government of India under Grants SR/S3/EECE/0024/2009 and ST/INT/NL/BIOMED/P[3]/2011(G).

REFERENCES


14. MESSIDOR


