Fusing Geometric and Appearance-based Features for Glaucoma Diagnosis

Kangrok Oh\textsuperscript{a} Jooyoung Kim\textsuperscript{a} Sangchul Yoon\textsuperscript{b} Kyoung Yul Seo\textsuperscript{b}
\textsuperscript{a}School of Electrical and Electronic Engineering, Yonsei University
50 Yonsei-ro, Seodaemun-gu, Seoul, 03722, Republic of Korea
kangrokoh, harrykim@yonsei.ac.kr
\textsuperscript{b}Department of Ophthalmology, Yonsei University
50 Yonsei-ro, Seodaemun-gu, Seoul, 03722, Republic of Korea
LITTLELUKE,seoky@yuhs.ac

ABSTRACT

In this paper, we propose to fuse geometric and appearance-based features at the feature-level for automatic glaucoma diagnosis. The cup-to-disc ratio and neuro-retinal rim width variation are extracted as the geometric features based on a coarse-to-fine localization method. For the appearance-based feature extraction, the principal components analysis is adopted. Finally, these features are combined at the feature-level based on the random projection and the total error rate minimization classifier. Experimental results on an in-house data set shows that the feature-level fusion can enhance the classification performance comparing with that before fusion.

KEYWORDS

Glaucoma diagnosis, feature-level fusion, cup-to-disc ratio estimation, principal components analysis, total error rate minimization

1 INTRODUCTION

Glaucoma is defined as “a group of disease that damage the eye’s optic nerve and can result in vision loss and blindness” according to [1]. Glaucoma is known as the leading cause of blindness [2], and the number of people with glaucoma in the United States is reported as about 2.7 million in 2010 and estimated to reach about 6.3 million in 2050 [3]. Since the progression of glaucoma is slow and irreversible [4], often it is called the “silent thief of sight”. Hence, diagnosing glaucoma at an early stage is important to prevent undesirable impairment or blindness caused by non-awareness.

Clinical glaucoma diagnosis is usually performed based on intra-ocular pressure measurement, visual field test, appearance analysis of the optic nerve head (ophthalmoscopy), etc. By means of technology development in the field of image processing and artificial intelligence, several attempts to diagnose glaucoma in an automatic manner using fundus images can be found in the literature recently. A brief summary of the recent works regarding automatic glaucoma diagnosis is presented in Table 1. Depending on the way of extraction features, the recent existing works related with automatic glaucoma diagnosis can be grouped into two categories: i) appearance-based feature extraction approach, and ii) geometric feature extraction approach.

The existing works in the appearance-based feature extraction approach category aim to extract features using the intensity values of an image or a sub-image. In [5], a set of appearance-based features are combined based on feature selection or ranking for accuracy performance enhancement. Similarly, a set of different features (high order spectra (HOS), discrete wavelet transform (DWT), etc.) are fused to improve the classification accuracy in [6]. In [7], texture features are extracted based on a local binary patterns (LBP) operator.

Those in the geometric feature extraction approach category attempt to localize mainly the optic disc and cup regions, and extract geometric features such as cup-to-disc ratio (CDR) and neuro-retinal RIM thickness variations. Essentially, iterative methods such as the active contour model (ACM) [8], Otsu’s threshold [9], Hough transform [10] are adopted for the localization task to improve the performance.
Table 1. A brief summary of related works.

<table>
<thead>
<tr>
<th>Index</th>
<th>Category</th>
<th>Feature Extraction</th>
<th>Classification</th>
<th>Data Set (#G\textsuperscript{a},#NG\textsuperscript{b})</th>
<th>Reported Accuracy(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acharya et al. [5]</td>
<td>Appearance-based\textsuperscript{c}</td>
<td>HOS\textsuperscript{d}, GLCM\textsuperscript{e}, RLM\textsuperscript{f}</td>
<td>SVM\textsuperscript{g}, SMO\textsuperscript{h}, NB\textsuperscript{i}</td>
<td>In-house (30,30)</td>
<td>91.00%</td>
</tr>
<tr>
<td>Krishnan and Faust [6]</td>
<td>Appearance-based</td>
<td>HOS, TT\textsuperscript{j}, DWT\textsuperscript{k}</td>
<td>SVM</td>
<td>In-house (30,30)</td>
<td>91.67%</td>
</tr>
<tr>
<td>Ali et al. [7]</td>
<td>Appearance-based</td>
<td>EBP\textsuperscript{l}</td>
<td>NN\textsuperscript{m}</td>
<td>HRF\textsuperscript{n}</td>
<td>95.10%</td>
</tr>
<tr>
<td>Fondón et al. [8]</td>
<td>Geometric\textsuperscript{b}</td>
<td>CDR\textsuperscript{o}</td>
<td>Thr\textsuperscript{p}</td>
<td>In-house (13,28)</td>
<td>78.10%</td>
</tr>
<tr>
<td>Guerre et al. [9]</td>
<td>Geometric</td>
<td>CDR, NRIM\textsuperscript{q}</td>
<td>SVM</td>
<td>In-house (15,14)</td>
<td>89.00%</td>
</tr>
<tr>
<td>Dutta et al. [10]</td>
<td>Geometric</td>
<td>CDR</td>
<td>Thr</td>
<td>HRF (15,15)</td>
<td>90.00%</td>
</tr>
</tbody>
</table>

\textsuperscript{a}: Appearance-based Feature Extraction, \textsuperscript{b}: Geometric Feature Extraction, \textsuperscript{c}: High Order Spectra, \textsuperscript{d}: Gray-Level Co-occurrence Matrix, \textsuperscript{e}: Run Length Matrix, \textsuperscript{j}: Trace Transform, \textsuperscript{k}: Discrete Wavelet Transform, \textsuperscript{l}: Local Binary Patterns, \textsuperscript{m}: Cup-to-Disc Ratio, \textsuperscript{n}: Neuro-retinal RIM width ratio, \textsuperscript{o}: Support Vector Machine, \textsuperscript{p}: Sequential Minimal Optimization, \textsuperscript{q}: Naive Bayesian, \textsuperscript{r}: Random Forests, \textsuperscript{s}: Nearest-Neighbor, \textsuperscript{t}: Threshold, \textsuperscript{u}: The number of glaumatous images, \textsuperscript{v}: The number of non-glauomatous images, \textsuperscript{w}: High Resolution Fundus image data set, \textsuperscript{x}: only left eye images are utilized.

This may result in obtaining a high computational cost. In addition, even though the geometric and appearance-based features can be easily combined, fusion of the geometric and appearance-based features is not investigated thoroughly yet.

In this paper, we propose to fuse geometric and appearance-based features at the feature-level. For the geometric feature extraction, a non-iterative coarse-to-fine localization scheme is proposed. Particularly, a matrix multiplication is designed to perform two-dimensional (2-D) mean filtering at the coarse search stage. The principal components analysis (PCA) [11] is adopted for the appearance-based feature extraction. Finally, the total error rate minimization (TER) classifier [15]. Figure 1 shows an overall flow of the proposed method.

2.1 Image Preprocessing

At the preprocessing stage, image resize, mask generation, and image cropping are sequentially performed for further feature extraction and classification. 1424 × 2144 RGB images are resized to 650 × 800 based on the bi-cubic interpolation [16]. Then, a weighted image is generated based on images in red, green, and blue channels as follows:

\[ \mathbf{W} = w_R \mathbf{R} + w_G \mathbf{G} + w_B \mathbf{B}, \]  

(1)

where \( \mathbf{R} \in \mathbb{R}^{650 \times 800} \), \( \mathbf{G} \in \mathbb{R}^{650 \times 800} \), and \( \mathbf{B} \in \mathbb{R}^{650 \times 800} \) are image matrices in red, green, and blue channels, respectively. Here, \( \mathbf{W} \in \mathbb{R}^{650 \times 800} \) denotes a weighted image matrix. \( w_R, w_G, \) and \( w_B \) denote weight values for the geometric features, the cup-to-disc ratio (CDR) and the inferior-superior rim length to nasal-temporal rim length ratio (IS-NTR) are extracted from a fundus image. The principal components analysis (PCA) [11] is adopted for the appearance-based feature extraction. The geometric and PCA features are fused at the feature-level by feature concatenation. Subsequently, the feature vector is normalized based on the min-max normalization [12], and expanded by the random projection (RP) [13, 14]. Finally, classification is performed based on the total error rate minimization (TER) classifier [15]. Figure 1 shows an overall flow of the proposed method.
2.2 Geometric Feature Extraction

The CDR and the ISNTR are adopted as the geometric features for glaucoma diagnosis. They are estimated by localizing the disc and cup regions from a fundus image. The CDR is estimated based on three different measures while the ISNTR is estimated based on a single measure. Finally, the geometric features extracted from a fundus image becomes a four dimensional vector.

2.2.1 Coarse Detection of Disc Region

In order to localize the disc and cup regions from an image, we search for coordinates of a pixel with the highest intensity value. Purpose of this search is to reduce the range for further disc and cup localization. To accomplish this, vessels are removed by means of the morphological dilation operation [16]. Next, \((2K + 1) \times (2K + 1)\) mean filtering is performed to smoothen the image. The 2-D mean filtering is implemented using the following matrix multiplication. For this operation, we define two matrices as follows:

\[
\mathbf{F}_1^L = \begin{bmatrix}
1 & \cdots & 1 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 1 & 0 & \cdots & 0 \\
0 & 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\
\vdots & \vdots & \ddots & 1 & \cdots & 1 & 0 \\
0 & 0 & \cdots & 0 & 1 & \cdots & 1
\end{bmatrix}_{R \times P},
\]

\[
\mathbf{F}_2^L = \begin{bmatrix}
1 & 0 & \cdots & 0 & 0 \\
1 & \ddots & \ddots & \ddots & \vdots \\
1 & \ddots & \ddots & 0 & 0 \\
0 & 1 & \cdots & 1 & 0 \\
0 & 0 & \ddots & \ddots & 1 \\
\vdots & \vdots & \ddots & 1 & \vdots \\
0 & 0 & \cdots & 0 & 1
\end{bmatrix}_{Q \times S},
\]

where \(\mathbf{F}_1^L \in \mathbb{I}^{R \times P}\) and \(\mathbf{F}_2^L \in \mathbb{I}^{Q \times S}\) are matrices for pre- and post- multiplications with a weighted image matrix, and \(L = 2 \cdot K + 1\). Here, \(R = P - 2 \cdot K\) and \(S = Q - 2 \cdot K\) are set to exclude the boundary pixels of the weighted image matrix for the filtering operation. We set \(w_R = 0.2\), \(w_G = 0.3\), and \(w_B = 0.5\) to generate the weighted image to obtain the brightest pixel coordinates.

From the matrices defined in (2) and (3), the output matrix from the 2-D mean filtering is obtained as follows:

\[
\mathbf{M} = \frac{1}{L^2} \mathbf{F}_1^L \cdot \mathbf{W} \cdot \mathbf{F}_2^L,
\]

where \(\mathbf{W} \in \mathbb{R}^{P \times Q}\) is a weighted image matrix (\(P = 650\) and \(Q = 800\)), and \(\mathbf{M} \in \mathbb{R}^{R \times S}\) is an output matrix from the 2-D mean filtering. Row and column coordinates of a pixel with the brightest intensity are searched from the matrix \(\mathbf{M}\). Here, we denote the row and
column coordinates as \((R, C)\). Subsequently, an image sub-region with the size of \(251 \times 251\) centered at the coordinates \((R + K, C + K)\) are cropped from the red, green, and channel image matrices \(R_c, G_c, \) and \(B_c\). The cropped red, green, and blue channel images are denoted as \(R_c \in \mathbb{R}^{251 \times 251}, G_c \in \mathbb{R}^{251 \times 251}, \) and \(B_c \in \mathbb{R}^{251 \times 251}\), respectively. Figure 2 (a) shows examples of intermediate results from the coarse detection of disc region.

2.2.2 Disc Localization

For disc localization, we utilize the cropped red channel image since it shows the most distinctive information on the disc region. To localize the disc region, a set of operations are sequentially performed on \(R_c\) as follows:

1. Vessels are removed based on the morphological dilation operations [16] from \(R_c\).

2. Histogram equalization [16] is performed on the image without vessels for contrast enhancement. The image resulting from the vessel removal and the histogram equalization is defined as \(H_c \in \mathbb{R}^{251 \times 251}\).

3. Threshold operation is performed on the histogram equalized image using two threshold values \(\tau_l\) and \(\tau_r\). From an observation regarding uneven spread of intensity values over the left and right regions of \(H_c\), we apply different threshold values on the left and right half of \(H_c\). We set \(\tau_l = 0.9 \times V\) and \(\tau_r = 0.8 \times V\) for right eye images, and \(\tau_l = 0.8 \times V\) and \(\tau_r = 0.9 \times V\) for left eye images. Here, \(V\) denotes the intensity value of the brightest pixel of \(H_c\). After the threshold operation, a binary matrix \(B_d \in \mathbb{R}^{251 \times 251}\) wherein pixels with ‘1’ values construct a candidate disc region is obtained.

4. From the binary matrix \(B_d\), a chunk of ‘1’ values whose center of mass is the closest from the coordinates \((126, 126)\) is extracted and the rest chunk of ‘1’ values are removed. Subsequently, morphological hole filling, dilation, and erosion operations [16] are performed to obtain a candidate disc region.

5. An ellipse fitting based on least squares [17] is applied to boundaries of the candidate disc region to obtain a fine disc region. The output matrix from the disc localization is defined as \(B_{disc} \in \mathbb{R}^{251 \times 251}\) where pixels inside the disc region have ‘1’ values while those in the non-disc region have ‘0’ values.

Figure 2 (b) shows examples of the intermediate results obtained from the disc localization process.

2.2.3 Cup Localization

For cup localization, a set of operations are sequentially performed on \(R_c, G_c, \) and \(B_c\) as follows:

1. Vessels are removed using the morphological dilation operations [16] from \(R_c, G_c, \) and \(B_c\). The resulting matrices from the vessel removal are defined as \(R_v \in \mathbb{R}^{251 \times 251}, G_v \in \mathbb{R}^{251 \times 251},\) and \(B_v \in \mathbb{R}^{251 \times 251}\), respectively.

2. A weighted image \(W_v \in \mathbb{R}^{251 \times 251}\) is generated from \(R_v, G_v, \) and \(B_v\) by setting the weight values as \(w_R = 0.3, w_G = 0.5,\) and \(w_B = 0.2\).

3. Element-wise matrix multiplication operations are performed on \(W_v, G_v,\) and \(G_c\) using \(B_{disc}\) to exclude the non-disc regions for cup localization process. The resulting matrices are defined as \(W_h = W_v \circ B_{disc}, G_{h1} = G_v \circ B_{disc},\) and \(G_{h2} = G_c \circ B_{disc}\) where \(\circ\) denotes the element-wise multiplication operator.

4. A threshold operation is performed on \(W_h \in \mathbb{R}^{251 \times 251}\) using a threshold value \(\tau_c = 0.9 \times W\) where \(W\) is the highest intensity value of \(W_h\). We denote the binary matrix resulting from the threshold operation as \(B_h \in \mathbb{R}^{251 \times 251}\).
5. Vessels with low intensity values are localized using a matrix difference operation, \(G_{h1} - G_{h2}\), followed by a threshold operation. The threshold value \(\tau_v\) is set as \(\tau_v = 0.8 \times J\) where \(J\) is the intensity value of the brightest pixel in \(G_{h1} - G_{h2}\). We denote the resulting binary matrix as \(B_v \in \mathbb{I}^{251 \times 251}\).

6. An element-wise OR operation is performed using \(B_h\) and \(B_v\). The resulting binary matrix is defined as \(B_c = B_h \lor B_v\) where \(\lor\) stands for the element-wise OR operator. Subsequently, the morphological dilation operation [16] is also applied to \(B_c\).

7. An ellipse fitting based on least squares is applied to boundaries of the candidate cup region to obtain a fine cup region. The output matrix from the cup localization is defined as \(B_{cup} \in \mathbb{I}^{251 \times 251}\). The elements with ‘1’ values belong to the localized cup region, and those with ‘0’ values belong to the non-cup region.

Figure 2. (c) shows examples of the intermediate results which are acquired from the cup localization process.

### 2.2.4 CDR Estimation

For estimating the value of cup-to-disc ratio (CDR), three different measures, namely i) vertical CDR \((V_{CDR})\), ii) horizontal CDR \((H_{CDR})\), and iii) area based CDR \((A_{CDR})\) are defined as follows:

\[
V_{CDR} = \frac{V_{cup}}{V_{disc}}, \quad (5)
\]

\[
H_{CDR} = \frac{H_{cup}}{H_{disc}}, \quad (6)
\]

\[
A_{CDR} = \frac{A_{cup}}{A_{disc}}, \quad (7)
\]

where \(V_{disc}\), \(H_{disc}\), and \(A_{disc}\) respectively denote a maximum value of vertical disc length, that of horizontal disc length, and the number of pixels in a disc region. Similarly, \(V_{cup}\), \(H_{cup}\), and \(A_{cup}\) stand for the corresponding values in a cup region.

### 2.2.5 ISNTR Estimation

In order to assess the neuro-retinal rim thickness variations, thicknesses of inferior, superior, nasal, and temporal rims are calculated from the disc and cup localization results. Subsequently, a ratio of the total length of nasal and temporal rims over that of inferior and superior rims \((R_{ISNT}, ISNTR)\) is obtained as follows:

\[
R_{ISNT} = \frac{L_N + L_T}{L_I + L_S}, \quad (8)
\]
where \( L_I, L_S, L_N, \) and \( L_T \) respectively denote a maximum length of inferior, superior, nasal, and temporal rims.

### 2.3 Principal Components Analysis

The principal components analysis (PCA) \([11]\) is adopted for appearance-based feature extraction. At first, Sub-images with the size of \( 181 \times 181 \) centered at the cup center coordinates are cropped from the red, green, and blue channel images (\( R_c, G_c, \) and \( B_c \)). Then, the sub-images are resized to \( 50 \times 50 \) images. For the left eye images, images are horizontally reversed. Subsequently, a weighted image is generated (\( w_R = 0.2, w_G = 0.4, \) and \( w_B = 0.4 \)) and vectorized. From the vectorized images, a training data matrix \( X_{tr} \in \mathbb{R}^{D \times N_{tr}} \) is constructed (\( N_{tr} \) = the number of training samples and \( D = 2500 \)). Here, \( X_{tr} \) concatenates column vectors of normal eye images and those of glaucomatous eye images sequentially. Similarly, a test data matrix \( X_{te} \in \mathbb{R}^{D \times N_{te}} \) is constructed (\( N_{te} \) = the number of test samples).

From \( X_{tr} \), a mean vector \( m \in \mathbb{R}^{D \times 1} \) is calculated as \( m = \frac{1}{N_{tr}} X_{tr} \cdot 1 \), where \( 1 \in \mathbb{N}^{N_{tr} \times 1} \) is a vector consisting of ‘1’ values only. Subsequently, a covariance matrix \( C \in \mathbb{R}^{D \times D} \) is calculated as follows:

\[
C = (X_{tr} - m \cdot 1^T)(X_{tr} - m \cdot 1^T)^T. \tag{9}
\]

Then, a subspace \( S \in \mathbb{R}^{d \times D} \) is obtained by concatenating \( d \) number of eigenvectors of \( C \) corresponding to the \( d \) largest eigenvalues in a row-wise manner. We note here that each row vector of \( S \) corresponds to an eigenvector. PCA features are extracted by projecting the mean subtracted training and test matrices onto \( S \) as follows:

\[
P_{tr} = S \cdot (X_{tr} - m \cdot 1^T), \tag{10}
\]

\[
P_{te} = S \cdot (X_{te} - m \cdot 1^T), \tag{11}
\]

where \( P_{tr} \in \mathbb{R}^{d \times N_{tr}} \) and \( P_{te} \in \mathbb{R}^{d \times N_{te}} \) are the resulting training and test feature matrices.

### 2.4 Feature-level Fusion

The geometric features and the PCA features are fused at the feature-level by feature concatenation. From the four dimensional geometric features and \( d \) dimensional PCA features, a feature vector \( f \in \mathbb{R}^{M \times 1} \) (\( M = d + 4 \)) is constructed where the first four elements of \( f \) are the geometric features, and the last \( d \) elements of \( f \) are the PCA features. We note here that elements of \( f \) is represented within \([0, 1]\) based on the min-max normalization \([12]\) using the entire samples. Subsequently, the feature vector \( f \) is expanded based on the random projection (RP) \([13, 14]\). An expanded feature vector \( g \in \mathbb{R}^{D_{rp} \times 1} \) is defined as

\[
g = \left[ \phi \left( r^T f + b_1 \right), \ldots, \phi \left( r^T_{D_{rp}} f + b_{D_{rp}} \right) \right]^T,
\]

where \( r_j \in \mathbb{R}^{M \times 1} \) is a random weight vector, and \( b_j \) is a random bias term for \( j = 1, \ldots, D_{rp} \). Here, \( D_{rp} \) stands for the dimension of an expanded feature vector, and \( \phi (\cdot) \) denotes a sigmoid activation function.

### 2.5 Total Error Rate Minimization

The total error rate minimization (TER) \([15]\) is adopted for classification. Firstly, a matrix containing the expanded training feature vectors is defined as follow:

\[
G_{tr} = \begin{bmatrix} G_{tr}^- \\ G_{tr}^+ \end{bmatrix}, \tag{12}
\]

where

\[
G_{tr}^- = \left[ g_1^-, \ldots, g_{N^-}^- \right]^T, \tag{13}
\]

and

\[
G_{tr}^+ = \left[ g_1^+, \ldots, g_{N^+}^+ \right]^T. \tag{14}
\]

Here, \( g_i^- \in \mathbb{R}^{D_{rp} \times 1} \) and \( g_j^+ \in \mathbb{R}^{D_{rp} \times 1} \) denote the \( i \)-th negative and the \( j \)-th positive training sample for \( i = 1, \ldots, N^- \) and for \( j = 1, \ldots, N^+ \). Here, \( N^- \) and \( N^+ \) stand for the number of negative and positive training samples, respectively.

Next, at a training phase, a weight parameter vector \( \alpha \in \mathbb{R}^{D_{rp} \times 1} \) is estimated based on the
TER [15] as follows:

\[
\alpha = \left( b + \frac{1}{N} \mathbf{G}_{tr}^T \mathbf{G}_{tr} + \frac{1}{\tau} \mathbf{G}_{tr}^{+T} \mathbf{G}_{tr}^{+} \right)^{-1} \times \left( \frac{\tau-n}{N} \mathbf{G}_{tr}^T 1_+ + \frac{\tau+n}{N} \mathbf{G}_{tr}^{+T} 1_+ \right),
\]

(15)

\[
1_- = [1, \ldots, 1]^T \in \mathbb{N}^{N \times 1} \quad \text{and} \quad 1_+ = [1, \ldots, 1]^T \in \mathbb{N}^{N+1}.
\]

We note that \( b \) is a small regularization constant (e.g., 0.0001) and \( \mathbf{I} \in \mathbb{N}^{D_{rp} \times D_{rp}} \) is an identity matrix. In (15), \( \tau \) and \( \eta \) stand for a preset threshold and an offset value. At a testing phase, the class label \( y_t \) of a test sample \( \mathbf{g}_t \) is estimated as

\[
y_t = \begin{cases} 
1 & \text{if } \mathbf{g}_t^T \alpha \geq \tau \\
0 & \text{otherwise}
\end{cases}
\]

(16)

3 EXPERIMENTS

In this section, we evaluate the classification performance of the geometric and PCA features before and after fusion. The goal of our experiments is to show whether the feature-level fusion can improve the classification accuracy comparing with that before fusion. Firstly, the data set utilized in the experiments is introduced. Next, experimental settings including evaluation protocols and parameter settings are provided. Finally, experimental results and analysis are presented.

3.1 Data Set

In our experiments, an in-house data set collected by the National Medical Center in Republic of Korea is utilized. Hereafter, the NMC data set will be used to denote the in-house data set. The NMC data set consists of 71 eye fundus images with the size of \( 1424 \times 2144 \) captured from left and right eyes. It contains 47 non-glaucomatous and 24 glaucomatous images. Figure 3 shows five non-glaucomatous and five glaucomatous images of the NMC data set.

![Non-glaucomatous Images](image1)

![Glaucomatous Images](image2)

**Figure 3.** Sample images of the NMC data set.

3.2 Experimental Settings

We evaluate the classification performance of the proposed method in terms of the accuracy which is defined as \( \frac{\mathcal{C}}{\mathcal{N}_{te}} \), where \( \mathcal{C} \) and \( \mathcal{N}_{te} \) denote the number of correctly classified test samples and the total number of test samples, respectively. For performance evaluation, stratified ten runs of five-fold cross-validation tests are performed. Additionally, ten different random values are utilized for the RP. Consequently, the classification accuracy is averaged from 500 repetitions.

For the geometric feature extraction, the weight values for the red, green, and blue channel images and threshold values for 41 images are manually adjusted to obtain the disc and cup regions more accurately. The reduced dimension \( d \) for the PCA is set to 60 which is selected from pre-training, and the dimension \( D_{rp} \) of the expanded feature vector by the RP is controlled within \( \{10, 20, \ldots, 500\} \). Both \( \tau \) and \( \eta \) of the TER classifier are set to 0.5, and the regularization constant is set as \( b = 0.0001 \). The class-specific normalization parameters are adopted as in [15] (details can be found in [15]). The best class-specific normalization parameter is selected based on training accuracy results and utilized for testing.
3.3 Results

In this section, we provide results and analysis regarding i) the geometric feature extraction, and ii) the classification accuracy performances before and after fusion. Figure 4 shows examples of correct disc and cup localization for non-glaucamatos and glaucomatous images. As shown in the figure, glaucomatous images show larger cup regions than those of non-glaucamatos images. Figure 5 shows incorrect localization of disc and cup regions for non-glaucamatos and glaucomatous images. The images in Figure 5 tend to show low contrast between the intensity values of cup and disc regions. Comparing with iterative methods for disc and cup localization [8, 9, 10], the proposed method has a lower computational complexity. However, incorrect disc and cup localization results can be obtained as shown in Figure 5. Figure 6 shows the estimated CDR values of the entire images. As shown in the figure, vertical and horizontal CDR show better results than that of area based CDR. The mean absolute difference values between the estimated CDRs and the CDR measured by an expert (ophthalmologist) are about 0.13 (vertical CDR), 0.15 (horizontal CDR), and 0.23 (area based CDR). From these values, it is expected that the classification accuracy of the geometric features are probably degraded due to the localization errors. In our experiments, we aim to improve the classification accuracy of these geometric features by means of the feature-level fusion with appearance-based features (PCA).

The average test classification accuracy (hereafter, accuracy in short) performances before and after fusion are drawn with respect to the feature dimension $D_{rp}$ in Figure 7. For the geometric features, accuracy values between about 71.20% and 74.75% are observed over the entire range of feature dimension variations. The PCA features show higher accuracy results (from about 74.50% to 75.27%) than that of the geometric features for $D_{rp} > 200$. Accuracy performance degradation is observed when the dimension of expanded feature vector by the RP ($D_{rp}$) is similar to the original PCA feature dimension ($d = 60$). After fusing the geometric and PCA features, the accuracy performance improves from 1% to 4% when $D_{rp} \geq 150$ comparing with that before fusion. This implies that the geometric and appearance-based (PCA) features provide
Figure 6. CDR values which are estimated by the proposed method and measured by an expert (ophthalmologist).

Figure 7. Average test classification accuracy performances of the geometric, PCA (before fusion), and combined features (after fusion) with respect to feature dimension variations.

complementary information for glaucoma diagnosis. The best accuracy performance after fusion the geometric and PCA features is about 78.96%.

4 CONCLUSION

In this paper, we proposed a fusion scheme based on the random projection and the total error rate minimization classifier for automatic glaucoma diagnosis. The proposed method combines the geometric and appearance-based features at the feature-level. For the geometric feature extraction, a coarse-to-fine method is proposed for optic disc and cup region localization. In particular, coarse detection of the brightest pixel coordinates is performed by a matrix multiplication which is identical to the two-dimensional mean filtering. We adopted the principal components analysis for the appearance-based feature extraction. Our experimental results showed that the classification accuracy after fusion outperforms that before fusion. Our future works include acquiring more fundus images and obtaining more reliable results based on a large set of images.

ACKNOWLEDGMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2016-31-0650).
REFERENCES


