Clustering-Based Discriminant Analysis for Eye Detection

Shuo Chen and Chengjun Liu

***Abstract*— This paper proposes three clustering-based discriminant analysis (CDA) models to address the problem that the Fisher linear discriminant may not be able to extract adequate features for satisfactory performance, especially for two class problems. The first CDA model, CDA-1, divides each class into a number of clusters by means of the k-means clustering technique. In this way, a new within-cluster scatter matrix S_w^c and a new between-cluster scatter matrix S_b^c are defined. The second and the third CDA models, CDA-2 and CDA-3, define a nonparametric form of the between-cluster scatter matrices $N - S_b^c$. The nonparametric nature of the between-cluster scatter matrices inherently leads to the derived features that preserve the structure important for classification. The difference between CDA-2 and CDA-3 is that the former computes the between-cluster matrix N-S_b^c on a local basis, whereas the latter computes the between-cluster matrix N-S_b^c on a global basis. This paper then presents an accurate CDA-based eye detection method. Experiments on three widely used face databases show the feasibility of the proposed three CDA models and the improved eye detection performance over some state-of-the-art methods.**

***Index Terms*— Discriminant analysis, k-means clustering, feature extraction, eye detection, Haar wavelets.**

I. Introduction

FISHER linear discriminant (FLD) [1] is a popular tool of discriminant analysis for feature extraction and classification. Since it was introduced, a number of its variants have been proposed and widely used in numerous fields of pattern recognition [2]–[6]. However, the FLD and most of its variants shares a major disadvantage that they may not be able to extract adequate features in order to achieve satisfactory performance, especially for two class problems. This is caused by the property that the between-class scatter matrices S_b of the FLD and its variants are generally not full rank. For any L class problem, the FLD can only derive at most $L - 1$ valid features. Thus, for two-class problems, the FLD can only derive a single valid feature, which is significantly inadequate for achieving satisfactory performance.

To address this problem, this paper proposes three clustering-based discriminant analysis (CDA) models. The first CDA model, CDA-1, divides each class into a number of clusters by means of the k-means clustering technique. In this way, a new within-cluster scatter matrix S_w^c and a new between-cluster scatter matrix S_b^c are defined. The rank of the S_b^c increases as the number of clusters increases, and therefore the CDA-1 can derive adequate features for satisfactory performance. The CDA-1 works well especially when inherent multi-models are presented in each class and the k-means clustering technique properly identifies these clusters.

Motivated by the work of nonparametric discriminant analysis (NDA) in [7], this paper further proposes another two CDA models, CDA-2 and CDA-3. The fundamental of the CDA-2 and CDA-3 is a clustering-based nonparametric form of the between-cluster scatter matrices N-S_b^c. These between-cluster scatter matrices N-S_b^c are full rank, and consequently both CDA-2 and CDA-3 can derive adequate features for classification. Furthermore, the nonparametric nature of the between-cluster scatter matrices inherently leads to the derived features that preserve the structure important for classification. The difference between CDA-2 and CDA-3 is that the former computes the between-cluster matrix N-S_b^c on a local basis whereas the latter computes the between-cluster matrix N-S_b^c on a global basis.

This paper then evaluates these three CDA models on the problem of eye detection. Accurate and efficient eye detection is important for building a fully automatic face recognition system, and the challenges for finding a robust solution to this problem have attracted much attention in the pattern recognition community [8]–[23]. Example challenges in accurate and efficient eye detection include large variations in image illumination, skin color (white, yellow, and black), facial expression (eyes open, partially open, or closed), as well as scale and orientation. Additional challenges include eye occlusion caused by eye glasses or long hair, and the red eye effect due to the photographic effect. All these challenge factors increase the difficulty of accurate and efficient eye-center localization.

This paper presents an accurate eye detection method integrating the 2D Haar transformation for image representation, the CDA for discriminatory feature extraction, and the nearest neighbor rule with some similarity measures for classification. The experiment section then fully evaluates the three CDA models and the proposed eye detection method on three widely used face databases, the Face Recognition Grand Challenge (FRGC) version 2 database [24], the BioID face database [11], and the FERET database [25]. Experiment results show the feasibility of the three CDA models and the improved performance over some state-of-the-art eye detection methods.

II. Background

The principle of the discriminant analysis is to find an optimal linear projection that is effective for reducing the feature dimensionality and preserving the class separability.

Manuscript received June 20, 2012; revised August 28, 2013; accepted October 1, 2013. Date of publication December 11, 2013; date of current version March 3, 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jesus Malo.

The authors are with the Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102 USA (e-mail: sc77@njit.edu; chengjun.liu@njit.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2013.2294548

Fisher linear discriminant (FLD) [1] is a popular tool of discriminant analysis. The FLD uses the within-class and between-class scatter matrix to formulate a criteria of the class separability. The FLD projection is then defined to maximize the criteria.

Let $\mathbf{X}$ denotes the feature vector and L denote the number of classes. Let $\omega_i, i = 1, 2, \ldots, L$ and $N_i, i = 1, 2, \ldots, L$ denote the classes and the number of feature vectors within each class, respectively. Let $M_i, i = 1, 2, \ldots, L$ and M_0 be the means of the classes and the grand mean. The within-class scatter matrix shows the scatter of feature vectors around their respective class mean vectors, which can be defined as follow:

$$S_w = \sum_{i=1}^{L} P(\omega_i) E\{(\mathbf{X} - M_i)(\mathbf{X} - M_i)^t | \omega_i\} \tag{1}$$

where $P(\omega_i)$ is a prior probability. The between-class scatter matrix shows the scatter of the class mean vectors around the grand mean, which can be defined as follow:

$$S_b = \sum_{i=1}^{L} P(\omega_i)(M_i - M_0)(M_i - M_0)^t \tag{2}$$

The FLD projection W is then defined to maximize the criteria as follow:

$$J(W) = \frac{|W^t S_b W|}{|W^t S_w W|} \tag{3}$$

and, mathematically, this criteria is maximized when W consists of the leading eigenvectors of $S_w^{-1} S_b$.

Since FLD was introduced, a number of its variants have been proposed [2]–[6], [26]–[30]. However, a major disadvantage of the FLD and most of its variants is that they may not be able to extract adequate features in order to achieve satisfactory performance, especially for two class problems. This is caused by the property that the between-class scatter matrix S_b of the FLD and its variants is generally not full rank. As indicated in Eq. 2, the rank of S_b is at most $L - 1$ for any L class problem, and consequently the rank of $S_w^{-1} S_b$ is at most $L - 1$ as well. Therefore, there are at most $L - 1$ valid eigenvectors of $S_w^{-1} S_b$, which means the FLD can only derive at most $L - 1$ valid features for any L class problem. For two class problems, the FLD can only derive a single valid feature, which is significantly inadequate for achieving satisfactory performance.

Fukunaga [7] initiated the study on nonparametric discriminant analysis (NDA) to address this problem. The NDA maintains the same within-class scatter matrix S_w with that of the FLD, but defines a nonparametric form of the between-class scatter matrix S_b using the nearest neighbor techniques. Motivated by Fukunaga's work, several improved approaches upon the NDA have been proposed [31]–[34]. All of these approaches consistently followed the idea of the nearest neighbors to define their between-class scatter matrices. One major drawback of these nearest neighbor based NDA methods is that they take the risk of suboptimal performance, since the scatter of each feature vector is measured only with respect to a very small amount of the remaining feature vectors, i.e., its k nearest neighbors.

Algorithm 1 The k-Means Clustering Algorithm Applying to the Feature Vectors Within Each Class

Input: the feature vectors $\mathbf{X}_j^{(i)}$, which denotes the jth feature vector in class ω_i, where $i = 1, 2, \cdots, L$, and $j = 1, 2, \cdots, N_i$.

Output: the mean vectors $M_p^{(i)}$, which denotes the mean vector of pth cluster in class ω_i, where $i = 1, 2, \cdots, L$, and $p = 1, 2, \cdots, k$.

FOR $i = 1, 2, \cdots, L$

- $t = 0$.
- Let $C_{p,t}^{(i)}, p = 1, 2, \cdots, k$, denotes the set of feature vectors from the pth cluster in class ω_i in the tth iteration. First, randomly assign each feature $X_j^{(i)}$ to a set $C_{p,t}^{(i)}$.
- Initialization: $M_{p,t}^{(i)} = \frac{1}{|C_{p,t}^{(i)}|} \sum_{X_j^{(i)} \in C_{p,t}^{(i)}} X_j^{(i)}$.
- REPEAT
  - Reassignment Step: reassign each feature vector to the cluster with the closest mean as follow:

    $$C_{p,t+1}^{(i)} = \{X_j^{(i)} : ||X_j^{(i)} - M_{p,t}^{(i)}|| \leq ||X_j^{(i)} - M_{p^*,t}^{(i)}|| \text{ for all } p^* = 1, 2, \cdots, k.\}$$

  - Update Step: calculate the new means to be the centroids of the feature vectors in the clusters as follow:

    $$M_{p,t+1}^{(i)} = \frac{1}{|C_{p,t+1}^{(i)}|} \sum_{X_j^{(i)} \in C_{p,t+1}^{(i)}} X_j^{(i)}.$$

  - $t = t + 1$
- UNTIL the algorithm converges when $M_{p,t}^{(i)}$ is unchanged.
- $M_p^{(i)} = M_{p,t}^{(i)}$.

END FOR

III. Clustering-Based Discriminant Analysis

This section presents the three clustering-based discriminant analysis (CDA) models, CDA-1, CDA-2, and CDA-3, to address the problem of inadequate features derived from the FLD.

A. CDA-1 Model

The CDA-1 model divides each class into a number of clusters by means of the k-means clustering technique. In this way, a new within-cluster scatter matrix S_w^c and a new between-cluster scatter matrix S_b^c are defined. The rank of the S_b^c increases as the number of clusters increases, and therefore the CDA-1 can derive sufficient valid features for achieving satisfactory performance.

Formally, the CDA-1 first use the k-means clustering technique to divide the feature vectors from each class into k clusters so as to minimize the within-cluster sum of squares.

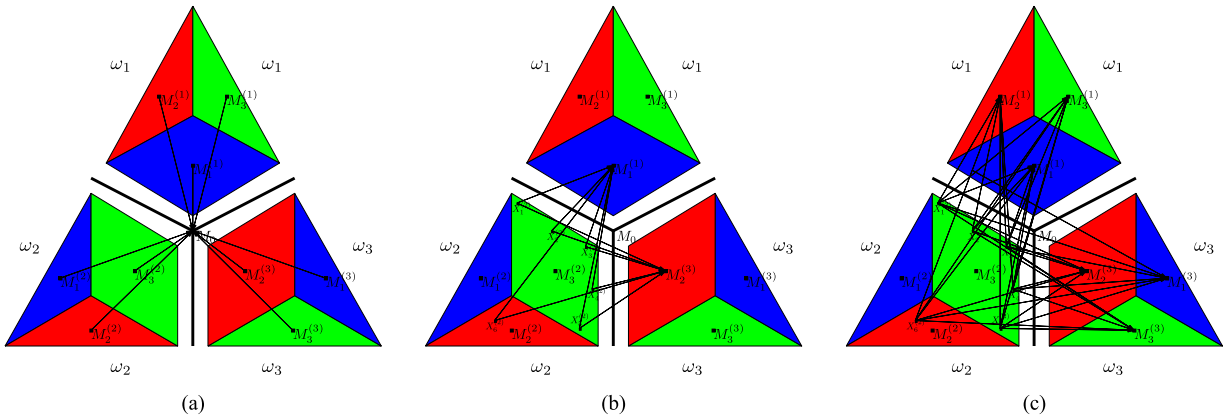


Fig. 1. The between-cluster matrices of CDA−1, −2, and −3, respectively. The figure shows a three-class problem and each class is further divided into three clusters. M_0 represents the grand mean, whereas $M_q^{(p)}$, $p, q = 1, 2, 3$, represents the mean vector of the qth cluster from class ω_p. $\mathbf{X}_j^{(2)}$, $j = 1, 2, \ldots, 6$, represents six data samples from class ω_2. (a) The between-cluster scatter matrix of CDA-1 measures the scatter of the mean vector from each cluster with respect to the grand mean. (b) The between-cluster scatter matrix of CDA-2 measures the scatter of each feature vector from one class with respect to the mean vector of its nearest cluster from otherwise classes. (c) The between-cluster scatter matrix of CDA-3 measures each feature vector from one class with respect to mean vectors of all the clusters from otherwise classes.

The algorithm of the k-means clustering applying to the feature vectors within each class is described in Alg. 1 in detail.

After we derive the mean vector of each cluster from every class, the new formulation of the within-class scatter matrix S_w^C is defined as follow:

$$S_w^c = \sum_{p=1}^{L}\{P(\omega_p)\sum_{j=1}^{k}\frac{|C_j^{(p)}|}{N_p}E\{(\mathbf{X}-M_j^{(p)})(\mathbf{X}-M_j^{(p)})^t|\omega_p\}\} \tag{4}$$

and the new formulation of the between-class scatter matrix S_b^C [Fig. 1(a)] is defined as follow:

$$S_b^c = \sum_{p=1}^{L}\{P(\omega_p)\sum_{j=1}^{k}\frac{|C_j^{(p)}|}{N_p}(M_j^{(p)}-M_0)(M_j^{(p)}-M_0)^t\} \tag{5}$$

where $|C_j^{(p)}|$ denotes the number of feature vectors in the jth cluster of the class ω_p, and $M_j^{(p)}$ denotes the mean vector of the jth cluster of the class ω_p. Note that as k decreased to one, $M_j^{(p)}$ converges to M_i. Thus, Eq. 4 and Eq. 5 is a generalization of Eq. 1 and Eq. 2, respectively.

There are two advantages of the CDA-1 model. First, the rank of the S_b^c is increased compared with S_b. Recall that the rank of the S_b of the FLD is upper bound by $L-1$ for any L class problem and thus at most $L-1$ features can be derived. In comparison, the rank of the S_b^c is upper bound by $k \times L - 1$. The S_b^c can even be full rank when enough large k is set. Therefore, more features can be derived for classification and the performance will be improved. Second, the clustering algorithm can find inherent multi-models in each class and further improve the performance. Take the task of eye detection as an example. It requires to differentiate between the eye class and the non-eye class, i.e. "the rest of the world". On one hand, the non-eye class indeed involves multi-models to represent different objects and scenes in "the rest of the world"; on the other hand, the eye class may contains multi-models as well, to represent different kinds of eyes such as open eyes, closed eyes, eyes with glasses, etc. The CDA-1 features are able to preserve the class separability among these multi-models and thus achieve better performance as indicated in Section V.

After computing S_w^c and S_b^c, the CDA-1 project matrix is the leading eigenvectors of $(S_w^c)^{-1}S_b^c$. To avoid the singularity problem, PCA is first applied before the CDA-1. Furthermore, inspired by the enhanced Fisher linear discriminant model (EFM) in [2], we decompose the CDA-1 procedure into a simultaneous diagonalization of the two within-cluster and between-cluster scatter matrices to improve the generalization performance of the CDA-1. The simultaneous diagonalization is stepwise equivalent to two operations as pointed out by Fukunaga [1]: whitening the within-cluster scatter matrix and applying PCA to the between-cluster scatter matrix using the transformed data. The CDA-1 should preserve a proper balance, during the stepwise process, between the need that the selected eigenvalues account for most of the spectral energy of the raw data (for representational adequacy), and the requirement that the eigenvalues of the within-class scatter matrix (in the reduced PCA space) are not too small (for better generalization performance) [2].

Finally, the detailed algorithm of the CDA-1 is given as follows:

1) Compute the PCA features as: $\mathbf{P} = \Phi^t\mathbf{X}$, where Φ is the leading eigenvectors of the mixture scatter matrix of $\mathbf{X}$.
2) In the feature space $\mathbf{P}$, follow the Alg. 1 to divide each classes into k clusters and compute the mean vector of each cluster.
3) In the feature space $\mathbf{P}$, compute the S_w^c as indicated by Eq. 4.
4) Whiten the S_w^c as follows:

$$S_w^c\Psi = \Psi\Lambda \quad \text{and} \quad \Psi^t\Psi = I$$
$$\Lambda^{-1/2}\Psi^t S_w^c \Psi\Lambda^{-1/2} = I \tag{6}$$

where Ψ and Λ are the eigenvectors and the diagonal eigenvalue matrices of S_w^C, respectively. Then compute the whitening transformed features $\mathbf{Y}$ with respect to S_w^c

as follow:

$$\mathbf{Y} = \Lambda^{-1/2}\Psi^t\mathbf{P} \tag{7}$$

5) In the feature space Y, compute the S_b^c as indicated by Eq. 5.
6) Diagonalize the S_b^c as follows:

$$S_b^c\Theta = \Theta\Gamma \quad \text{and} \quad \Theta^t\Theta = I \tag{8}$$

where Θ and Γ are the eigenvectors and the diagonal eigenvalue matrices of S_b^c, respectively. Then the CDA-1 features $\mathbf{Z}$ are now defined as follows:

$$\mathbf{Z} = \Theta^t\mathbf{Y} \tag{9}$$

Finally, the overall transformation matrix of the CDA-1 can be defined as:

$$T = \Phi\Psi\Lambda^{-1/2}\Theta \tag{10}$$

B. CDA-2 Model

The CDA-1 model significantly increases the number of the derived features, but the number is still upper bound by $k \times L - 1$. Even though the S_b^c can be full rank when enough large k is set, it takes the risk of impairing the inherent multi-models in each class. Motivated by the work of nonparametric discriminant analysis (NDA) in [7], this subsection further proposes the CDA-2 model. The fundamental of the CDA-2 model is a clustering-based nonparametric form of the between-cluster scatter matrix $N{-}S_b^c$. The $N{-}S_b^c$ is full rank, and consequently CDA-2 can derive adequate features for classification. Furthermore, the nonparametric nature of the between class scatter matrix inherently leads to the derived features that preserve the structure important for classification.

Specifically, the between-class scatter matrix $N{-}S_b^c$ of the CDA-2 is given on a local basis, which measures the scatter of each feature vector from one class with respect to the mean vector of its nearest cluster from otherwise classes [Fig. 1(b)]. The $N{-}S_b^c$ is defined as:

$$N\text{-}S_b^c = \sum_{i=1}^{L}\frac{P(\omega_i)}{N_i}\sum_{\substack{p=1\\p\neq i}}^{L}\sum_{j=1}^{N_i}\mathrm{w}^{(p)}(\mathbf{X}_j^{(i)}) (\mathbf{X}_j^{(i)} - M^{(p)}(\mathbf{X}_j^{(i)}))(\mathbf{X}_j^{(i)} - M^{(p)}(\mathbf{X}_j^{(i)}))^t \tag{11}$$

where $M^{(p)}(\mathbf{X}_j^{(i)})$ denotes the mean vector of the nearest cluster from the class ω_p to the feature vector $\mathrm{X}_j^{(i)}$, and $\mathrm{w}^{(p)}(\mathbf{X}_j^{(i)})$ is a weighting function, which can be defined as:

$$\mathrm{w}^{(p)}(\mathbf{X}_j^{(i)}) = \frac{\min\{d^\alpha(\mathbf{X}_j^{(i)}, M^{(i)}(\mathbf{X}_j^{(i)})), d^\alpha(\mathbf{X}_j^{(i)}, M^{(p)}(\mathbf{X}_j^{(i)}))\}}{d^\alpha(\mathbf{X}_j^{(i)}, M^{(i)}(\mathbf{X}_j^{(i)})) + d^\alpha(\mathbf{X}_j^{(i)}, M^{(p)}(\mathbf{X}_j^{(i)}))} \tag{12}$$

where α is a control parameter between zero and infinity, and $d(\mathbf{X}_j^{(i)}, M^{(i)}(\mathbf{X}_j^{(i)}))$ and $d(\mathbf{X}_j^{(i)}, M^{(p)}(\mathbf{X}_j^{(i)}))$ denote the Euclidean distance from a feature vector $\mathbf{X}_j^{(i)}$ to the mean vector of its nearest cluster from itself class ω_i and from the otherwise class ω_p, respectively. The value of the weighting function is close to 0.5 when the feature vector is near the class boundary and drops off to 0.0 as feature vector moves away from the boundary. This property allows the feature vectors near the class boundary, which preserve the classification structure, to contribute more to the between-class scatter matrix $N\text{-}S_b^c$. Note that we set $\alpha = 2$ in our experiments for the CDA-2 model.

From Eq. 11, we have the following observations. First, due to the application of the clustering technique, the CDA-2, as CDA-1 does, is capable of finding the inherent multi-models in each class and improve the performance for those case where the multi-models are indeed present.

Second, the CDA-2 can derive more features than the CDA-1. The CDA-2 makes use of all the feature vectors, in stead of only the cluster centroids, in the definition of the $N\text{-}S_b^c$. Thus, the $N\text{-}S_b^c$ is full rank. More features can be derived for the classification, and the performance may be improved as more information is included.

Third, the CDA-2 is capable of effectively preserving the classification structure. Either the FLD or the CDA-1 defines the between-class(cluster) scatter matrix only based on the mean vectors of the classes(clusters). It fails to involve the feature vectors on the class boundary which preserve the classification structure. In comparison, the CDA-2 takes account of all the feature vectors, including those on the class boundary, to define the between-cluster scatter matrix. More importantly, considering that the feature vectors far from the class boundary may distort the classification structure, the CDA-2 introduces a weighting function to emphasize the effect of those feature vectors near to the class boundary but to de-emphasize the effect of those far from the class boundary. As indicated in [Fig. 1(b)], the data samples $\mathbf{X}_1^{(2)}, \mathbf{X}_2^{(2)}, \cdots, \mathbf{X}_5^{(2)}$ fall on the class boundary and are given a bigger weight to emphasize their effect, whereas the data sample $\mathbf{X}_6^{(2)}$ is far away from the boundary and is given a smaller weight to de-emphasize its effect. In this way, the CDA-2 effectively preserves the classification structure.

Please note that the CDA-2 has two major difference from the nearest neighbor based nonparametric discriminant analysis (NDA) in [7]. First, the NDA only works on the two class problems, whereas the CDA-2 works on multi-class (equal to or bigger than two) problems. Second, the between-class scatter matrix of the NDA measures the scatter of feature vectors with respect to the mean vectors of their k nearest neighbors. This kind of measurement takes the risk of suboptimal performance, since only the local scatters with a very small amount of feature vectors are utilized and much information is lost in the learning procedure. By contrast, the CDA-2 takes advantage of the inherent multi-models in each class. The between-cluster scatter matrix of the CDA-2 measures the scatter of feature vectors with respect to the mean vectors of their nearest clusters from otherwise classes. The CDA-2 serves as a better representation of the scatter of the feature vectors with respect to the otherwise clusters(classes).

Finally, the detailed algorithm of the CDA-2 is given as follows:

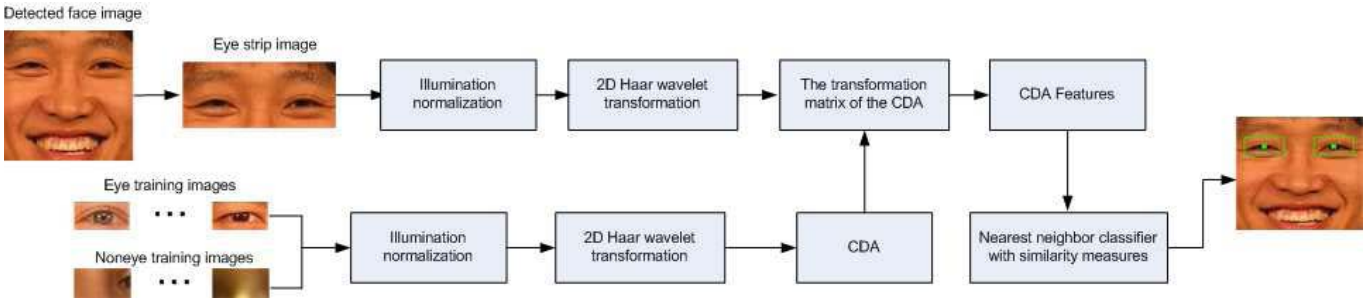


Fig. 2. System architecture of the proposed eye detection method.

1) Compute the PCA features as: $\mathbf{P} = \Phi^t \mathbf{X}$, where Φ is the leading eigenvectors of the mixture scatter matrix of $\mathbf{X}$.
2) In the feature space $\mathbf{P}$, follow the Alg. 1 to divide each classes into k clusters and compute the mean vector M of each cluster.
3) In the feature space $\mathbf{P}$, compute the S_w^c as indicated by Eq. 4.
4) Whiten the PCA features and the mean vector of each cluster with respect to S_w^c as: $\mathbf{Y} = \Lambda^{-1/2}\Psi^t\mathbf{P}$ and $M' = \Lambda^{-1/2}\Psi^t M$, where Λ and Ψ are the eigenvalues and eigenvectors of S_w^c.
5) In the feature space Y, find the nearest cluster to each feature vector from the otherwise classes. Compute the $\mathrm{w}^{(p)}(\mathbf{X}_j^{(i)})$ as indicated by Eq. 12.
6) In the feature space Y, compute $N\text{-}S_b^c$ as indicated by Eq. 11.
7) The CDA-2 features are then defined as: $Z = \Theta^t\mathbf{Y}$, where Θ are the eigenvectors of the $N\text{-}S_b^c$.

C. CDA-3 Model

The between-cluster scatter matrix $N\text{-}S_b^c$ of the CDA-2 is defined on a local basis, which measures the scatter of each feature vector from one class with respect to the mean vector of its nearest cluster from otherwise classes. One limitation of this definition is that the $N\text{-}S_b^c$ of the CDA-2 just takes account of the nearest cluster to each feature vector but ignores the contributions of other clusters. However, the fact is that the different clusters may contribute differently when we measure the scatter of each feature vector. Therefore, if all clusters are taken into account when we measure the scatter of each feature vector, the between-cluster scatter matrix may preserve the classification structure from different points of view, and hence may improve the classification performance.

Inspired by this idea, we propose the CDA-3 based on a new formulation of the between-cluster scatter matrix $N\text{-}S_b^c$. The $N\text{-}S_b^c$ of the CDA-3 is defined on a global basis, which measures the scatter of each feature vector from one class with respect to the mean vectors of all the clusters from otherwise classes [Fig. 1(c)]. The $N\text{-}S_b^c$ of the CDA-3 is defined as follows:

$$N\text{-}S_b^c = \sum_{i=1}^{L} \frac{P(\omega_i)}{N_i} \sum_{\substack{p=1 \\ p\neq i}}^{L} \sum_{q=1}^{k} \sum_{j=1}^{N_i} \mathrm{w}_q^{(p)}(\mathbf{X}_j^{(i)}) (\mathbf{X}_j^{(i)} - M_q^{(p)})(\mathbf{X}_j^{(i)} - M_q^{(p)})^t \quad (13)$$

where $M_q^{(p)}$ denotes the mean vector of the qth cluster from the class ω_p, and $\mathrm{w}_q^{(p)}(\mathbf{X}_j^{(i)})$ is a weighting function, which can be defined as:

$$\mathrm{w}_q^{(p)}(\mathbf{X}_j^{(i)}) = \frac{\min\{d^\alpha(\mathbf{X}_j^{(i)}, M_q^{(i)}), d^\alpha(\mathbf{X}_j^{(i)}, M_q^{(p)})\}}{d^\alpha(\mathbf{X}_j^{(i)}, M_q^{(i)}) + d^\alpha(\mathbf{X}_j^{(i)}, M_q^{(p)})} \quad (14)$$

where $d(\mathbf{X}_j^{(i)}, M_q^{(i)})$ and $d(\mathbf{X}_j^{(i)}, M_q^{(p)})$ denote the Euclidean distance from a feature vector $\mathbf{X}_j^{(i)}$ to the mean vector of the qth cluster from itself class ω_i and from the otherwise class ω_p, respectively. Note that we set $\alpha = 2$ in our experiments for the CDA-3 model.

The CDA-3 possesses all the advantages of the CDA-2. It is capable of finding the inherent multi-models in each class. The between-cluster scatter matrix is full rank and thus can derive sufficient features for satisfactory performance. And it can effectively preserve the classification structure due to the introduction of the weighting function.

The algorithm to derive the CDA-3 features is similar with that of the CDA-2. We don't explicitly present the detailed algorithm here.

IV. A CDA-Based Eye Detection Method

We present an accurate eye detection method based on the clustering-based discriminant analysis (CDA). The method integrates the 2D Haar transformation for image representation, the CDA for discriminatory feature extraction, and the nearest neighbor rule with some similarity measures for classification.

Fig. 2 shows the system architecture of the proposed eye detection method. We first apply the Bayesian Discriminating Features(BDF) method [35] to detect a face from an image and normalize the detected face to a predefined size (128×128 in our experiments). Then the geometric constraints are applied, which means eyes are only searched in the upper portion (eye strip) of the detected face (55×128 in our experiments). Before the 2D Haar wavelet transformation, the preprocessing of illumination normalization is used to alleviate the effect of the illumination variation. This illumination normalization consists of the Gamma correction, difference of Gaussian (DoG) filtering, and contrast equalization [36].

Next, 2D Haar wavelet transformation is applied. Since it was introduced to object detection by Papageorgiou and Oren [37], 2D Haar wavelet transformation attracted a great deal of attention. The Haar wavelets include a set of basis functions that encode the difference in average intensities between

different regions in different scales. Except for the first basis function that computes the average of the whole image, the Haar basis functions include three types of representation in the two dimensional space: (i) two horizontal neighboring rectangular regions, which compute the difference between the sum of pixels within each of them, (ii) two vertical neighboring rectangular regions, which compute the difference between the sum of pixels within each of them, and (iii) four neighboring rectangular regions, which compute the difference between diagonal pairs of the rectangles. In [9] and [10], it is shown that 2D Haar wavelets are effective for capturing the structure characteristic of eyes in different scales: centered dark pupil is surrounded by a relatively white sclera. Another advantage of 2D Haar features is that it can be efficiently computed by the Integral Image [38], in which the inner product of Haar basis functions with an image vector can be performed by just several integer additions and subtractions instead of the time-consuming floating point multiplications.

Finally, the CDA is applied for discriminant feature extraction and subsequently the nearest neighbor rule is applied for classification. All three models of the CDA are applied and compared with each other. Three kinds of similarity measures are used to measure the nearest neighbor rule. They are L_1 similarity measure δ_{L_1}, L_2 similarity measure δ_{L_2}, and cosine similarity measure δ_{cos}, which can be defined as follows:

$$\delta_{L_1}(\mathbf{X}, \mathbf{Y}) = \sum_i |\mathbf{X}_i - \mathbf{Y}_i| \tag{15}$$

$$\delta_{L_2}(\mathbf{X}, \mathbf{Y}) = (\mathbf{X} - \mathbf{Y})^t(\mathbf{X} - \mathbf{Y}) \tag{16}$$

$$\delta_{cos}(\mathbf{X}, \mathbf{Y}) = \frac{-\mathbf{X}^t\mathbf{Y}}{||\mathbf{X}||\ ||\mathbf{Y}||} \tag{17}$$

where $||\cdot||$ denotes the norm operator. A feature vector is then classified to the class of the closest mean using the similarity measures.

V. Experiment

We fully evaluate the three CDA models (CDA-1, CDA-2, and CDA-3) and the proposed eye detection method using 12,776 images from the Face Recognition Grand Challenge (FRGC) version 2 database [24]. Thus there are totally 25,552 eyes to be detected. Note that the FRGC images possess challenge characteristics, such as large variations in illumination, skin color (white, yellow, and black), facial expression (eyes open, partially open, or closed), as well as scale and orientation. Additional challenges include eye occlusion caused by eye glasses or long hair, and the red eye effect due to the photographic effect. All these challenge factors increase the difficulty of accurate eye-center localization. We also evaluate the proposed CDA-based eye detection method on the BioID face database [11] and the FERET database [25] in order to show its robustness and to compare it with some state-of-the-art eye detection methods. The experimental results on the BioID and the FERET databases are shown in the end of this section.

The training data collected from various sources contain 3,000 pairs of eyes and 12,000 non-eye patches in our experiments. The size of the eye and non-eye training images is normalized to 20×40.

We evaluate the precision of the eye detection using the normalized error γ [11]. The normalized error γ is defined as the detection pixel error normalized by the interocular distance. The advantage of the normalized error is that it measures the precision of the eye detection independent from the scale of the face and the image size.

There are two key parameters involved in the CDA: the size of extracted features (m) and the number of clusters (k). We will first evaluate the effect of these two parameters on the three CDA models. The evaluation of the best eye detection performance with the optimal parameters comes after the parameter evaluation. The comparison is made as well with PCA, FLD, and NDA methods in order to show the performance improvement of the CDA method. Please note that when we tune the parameters only 600 FRGC images are used, but the final eye detection performance with the optimal parameters are given based on the whole FRGC database.

A. Evaluation of the Size of Extracted Features

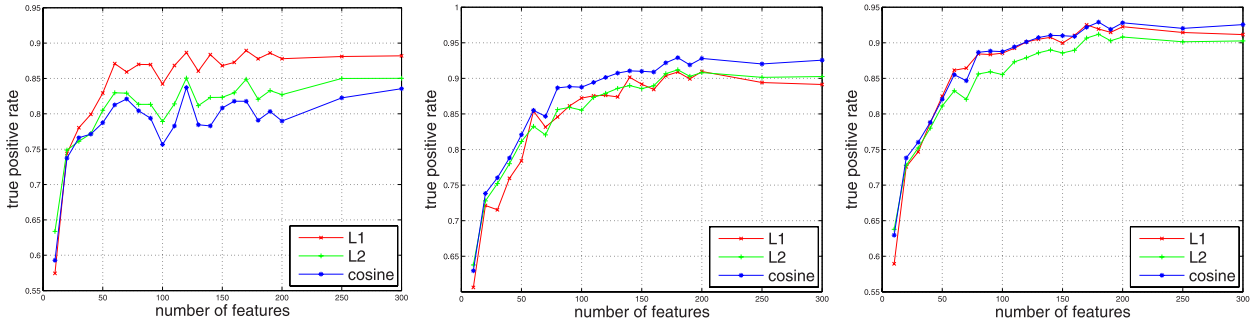
The size of the original 2D Haar wavelet features in our experiments is 1,024. There are two points involving the size determination of the features in the process of the CDA: the size of the intermediate PCA features (m_1) and the size of the final CDA features (m_2). It's hard to exhaustively evaluate all the possible combinations of m_1 and m_2. Considering the aspect of the efficiency, we only evaluate m_1 in the range of 50 and 300. For simplicity, we always choose the maximum size of the final CDA features (m_2) which can be derived as the size of the PCA features (m_1) varies. Specifically, for CDA-1, m_2 is set equal to $2 \times k - 1$; for CDA-2 and CDA-3, m_2 is set equal to m_1. Please note that we temporarily set $k = m_1/2$ when we evaluate the effect of the size of features on CDA. The complete evaluation of the effect of the number of clusters (k) on CDA will be given in the next subsection.

Fig. 3(a)–(c) show the performance comparison of the CDA-1, CDA-2, and CDA-3, respectively, as the m_1 and m_2 varies. Specifically, Fig. 3(a)–(c) show the true positive rate of the CDA-1, CDA-2, and CDA-3, respectively, at the false accept rate of 0.1 for the detection normalized error $\gamma \leq 0.07$. Note that the normalized error of 0.07 is a significantly strict criteria, which can be considered that the detected eye center is inside the eye pupil.

Fig. 3(a)–(c) reveal that the detection performance is affected by the size of the derived features. The CDA-1 reaches the best performance by using 170 features and L_1 similarity measure, while both the CDA-2 and the CDA-3 reach the best performance by using 180 features and cosine similarity measure.

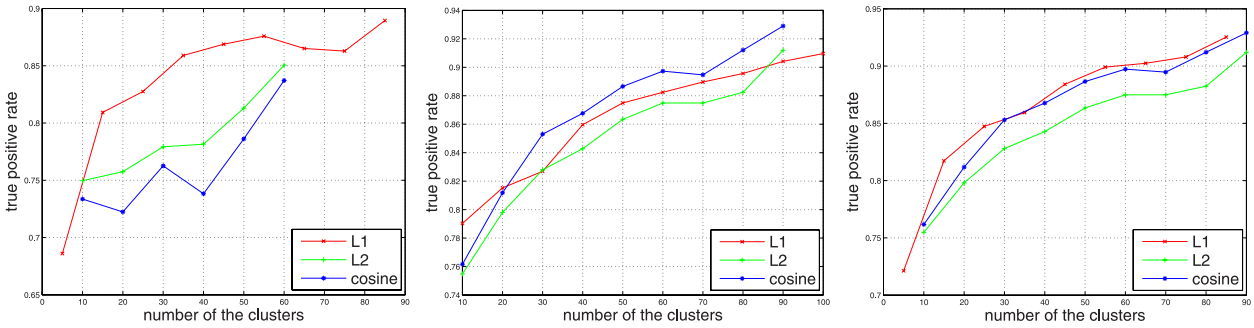
B. Evaluation of the Number of Clusters

The number of clusters (k) is evaluated between a small value ($k = 5$) and the half value of the size of PCA features ($k = m_1/2$). There are two reasons we only evaluate the number of clusters in above range: (i) the clusters inherently represent the multi-models of each class, and the number of the multi-models of each class should be neither too small



(a) (b) (c)

Fig. 3. The detection performance of the CDA-1, -2, and -3, respectively, as the size of features varies.



(a) (b) (c)

Fig. 4. The detection performance of the CDA-1, -2, and -3, respectively, as the number of clusters varies.

nor too large; and (ii) for a two class problem, the within-cluster matrix of CDA-1 is upper bound by m_1 in terms of Eq. 4, whereas the between-cluster matrix is upper bound by the smaller value of m_1 and $2 \times k - 1$ in terms of Eq. 5; both the within-cluster and the between-cluster matrices of CDA-2 and CDA-3 are upper bound by m_1 in terms of Eq. 4, 11, and 13. Therefore, even if more clusters ($k > m_1/2$) are built up, there are still at most m_1 features derived from the CDA, which means that only the first significant $m_1/2$ clusters of each class contribute to the process of the CDA.

Fig. 4(a)–(c) show the performance comparison of the CDA-1, CDA-2, and CDA-3, respectively, as the number of clusters (k) varies. For simplicity, we only evaluate the effect of the clusters based on the size of features which gives the best performance as we analyzed in the previous subsection. Specifically, we evaluate the effect for the CDA-1 using 169 ($m_1 = 170, m_2 = 169$) features for L_1 similarity measures and 119 ($m_1 = 120, m_2 = 119$) features for both L_2 and cosine similarity measures, evaluate the effect for the CDA-2 using 200 ($m_1 = 200, m_2 = 200$) features for L_1 similarity measure and 180 ($m_1 = 180, m_2 = 180$) features for both L_2 and cosine similarity measures, as well as evaluate the effect for the CDA-3 using 170 ($m_1 = 170, m_2 = 170$) features for L_1 similarity measure and 180 ($m_1 = 180, m_2 = 180$) features for both L_2 and cosine similarity measures. In addition, note that Fig. 4(a)–(c), as Fig. 3(a)–(c) do, show the true positive rate of the CDA-1, CDA-2, and CDA-3, respectively, at the false accept rate of 0.1 for the detection normalized error $\gamma \leq 0.07$, as the number of clusters (k) varies.

Fig. 4(a)–(c) reveal that the performance of all three CDA models is enhanced as the number of clusters increases and approaches to $m_1/2$. Specifically, the CDA-1 reaches the best performance by using 85 clusters of each class and L_1 similarity measure, while both the CDA-2 and the CDA-3 reach the best performance by using 90 clusters of each class and cosine similarity measure.

C. Evaluation of the Eye Detection Performance

In this subsection, we evaluate the eye detection performance of the CDA-based method in comparison with the PCA-, FLD-, and NDA-based methods. Note that the experiments are carried on the 12,776 FRGC images. For the best performance, the parameter settings are shown in Table I.

Fig. 5 shows the receiver operating characteristic (ROC) curves of the PCA-, FLD-, NDA-, and CDA-based eye detection methods. Note that the ROC curves are drawn for the detection normalized error $\gamma \leq 0.07$. Fig. 5 reveals that the CDA-based methods (CDA-1, CDA-2, and CDA-3) significant improve the eye detection performance in comparison with the PCA-, FLD-, and NDA-based methods. The CDA-3 gives the best performance, followed in order by the CDA-2, the CDA-1, the NDA, the FLD, and the PCA. Table II shows the true positive rate (TPR) of these methods at the false positive rate (FPR) of 0.1. It reveals that the CDA-3, which gives the best performance, improves the TPR of the PCA by 40.38%, the FLD by 28.52%, and the NDA by 20.60%, respectively. Even the CDA-1, which gives the lowest TPR

TABLE I

THE PARAMETER SETTINGS OF PCA-, FLD-, NDA-, AND CDA-BASED EYE DETECTION METHODS

Method	#PCA Features	#Discriminatory Features	#Nearest Neighbors	#Clusters	Similarity Measure
PCA	25	-	-	-	cosine
FLD	150	1	-	-	L_1
NDA	150	150	5	-	L_2
CDA-1	170	169	-	85	L_1
CDA-2	180	180	-	90	cosine
CDA-3	180	180	-	90	cosine

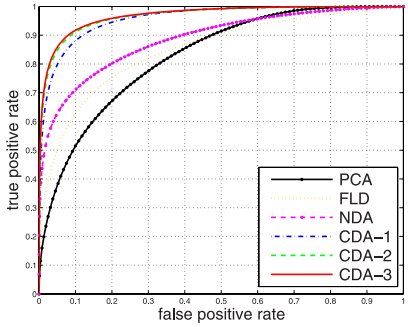


Fig. 5. The ROC curves of the PCA-, FLD-, NDA-, and CDA-based eye detection methods.

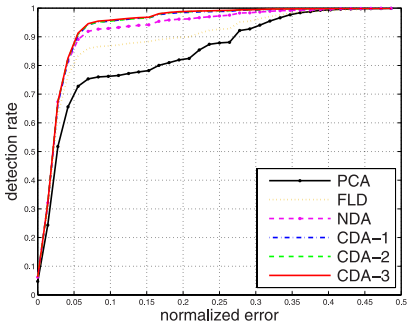


Fig. 6. The detection rate of the PCA-, FLD-, NDA-, and CDA-based eye detection methods over different normalized error.

TABLE II

COMPARISON OF THE TRUE POSITIVE RATE (TPR) OF THE PCA-, FLD-, NDA-, AND CDA-BASED EYE DETECTION METHODS AT THE FALSE POSITIVE RATE (FPR) OF 0.1

Method	TPR at FPR of 0.1
PCA	51.49%
FLD	63.35%
NDA	71.27%
CDA-1	87.76%
CDA-2	91.35%
CDA-3	91.87%

among the three CDA models, improves the TPR of the PCA by 36.27%, the FLD by 24.41%, and the NDA by 16.49%, respectively.

If we eventually choose the detected eye location as the average of the multiple detection around the pupil center (usually there are multiple detections around each pupil center), Fig. 6 show the detection rate over the different normalized detection errors. In Fig 6, the horizontal axis represents the normalized detection error, and the vertical axis represents the

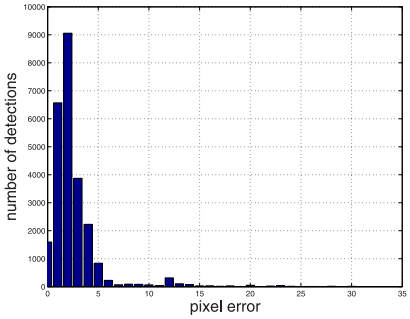


Fig. 7. Distribution of the detection pixel errors using the CDA-3 based method.

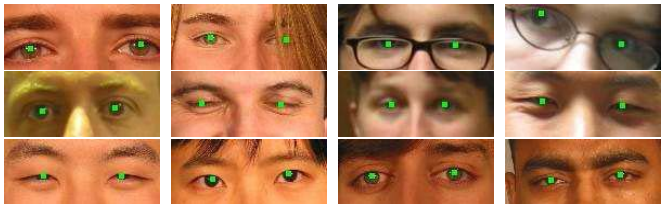


Fig. 8. Example eye detection results using the CDA-3 based method.

corresponding detection rate. Fig 6 reveals that the CDA-based methods consistently outperform the PCA-, FLD-, and NDA-based methods in terms of the detection rate. The three CDA models, CDA-1, CDA-2, and CDA-3, have comparable detection rates — the CDA-3 performs slightly better than CDA-2, and subsequently the CDA-2 performs slightly better than the CDA-1.

In order to further show the superiority of the CDA-based methods, Table III explicitly shows the detection rate (for the normalized error $\gamma \leq 0.07$) and the detection accuracy (i.e., the average detection pixel errors in the Euclidean distance) of these eye detection methods. For complete assessment, the detection accuracy on both the horizontal (x) and vertical (y) directions are shown as well in Table III. Table III show the improvement of the CDA-based methods on both the detection rate and the detection accuracy over the nonCDA-based methods. If one focuses on that more eyes are detected within a criteria (e.g., $\gamma \leq 0.07$), the CDA-3 gives the best detection rate, which is 94.58%; if one focuses on the minimum average detection pixel error, the CDA-3 still gives the best detection accuracy, which is 2.75 pixels.

Take the CDA-3 based eye detection method (which gives slightly better detection performance than CDA-1 and CDA-2) as an example. Fig. 7 shows the distribution of the detection pixel errors from the ground truth and Fig. 8 shows some example detection results.

TABLE III

THE DETECTION RATE AND DETECTION ACCURACY OF THE PCA-, FLD-, NDA-, AND CDA-BASED METHODS. NOTE THAT THE DETECTION RATE IS FOR THE NORMALIZED ERROR $\gamma \leq 0.07$. THE MEAN$(\cdot)$ AND STD$(\cdot)$ REPRESENT THE MEAN AND THE DEVIATION OF THE DETECTION PIXEL ERROR WITH RESPECT TO THE DIRECTION SPECIFIED BY THE PARAMETER, RESPECTIVELY

Method	mean(x)	std(x)	mean(y)	std(y)	mean($\sqrt{x^2+y^2}$)	Detection Rate
PCA	3.27	4.21	4.18	7.20	6.24	75.29%
FLD	2.70	3.68	2.66	5.86	4.49	85.90%
NDA	2.47	3.24	1.32	3.44	3.24	91.89%
CDA-1	2.27	2.46	1.05	2.49	2.84	94.20%
CDA-2	2.30	2.62	0.97	2.19	2.80	94.27%
CDA-3	2.27	2.56	0.94	2.10	2.75	94.58%

TABLE IV

COMPARISON OF THE EYE DETECTION PERFORMANCE WITH STATE-OF-THE-ART METHODS (NOTE THAT γ STANDS FOR THE NORMALIZED ERROR)

method	database	$\gamma \leq 0.05$	$\gamma \leq 0.10$	$\gamma \leq 0.25$
Jesorsky (2001) [11]	BioID	40.00%	79.00%	91.80%
Hamouz (2004) [12]	BioID	50.00%	66.00%	70.00%
Hamouz (2005) [15]	BioID	59.00%	77.00%	93.00%
Cristinacce (2004) [16]	BioID	56.00%	96.00%	98.00%
Asteriadis (2006) [17]	BioID	74.00%	81.70%	97.40%
Bai (2006) [18]	BioID	37.00%	64.00%	96.00%
Niu (2006) [19]	BioID	78.00%	93.00%	95.00%
Campadelli (2006) [8]	BioID	62.00%	85.20%	96.10%
Campadelli (2009) [20]	BioID	80.70%	93.20%	95.30%
Valenti (2008) [21]	BioID	84.10%	90.85%	98.49%
Campadelli (2006) [8]	FERET	67.70%	89.50%	96.40%
Monzo (2011) [22]	FERET	78.00%	96.20%	99.60%
CDA-3 (Our method)	BioID	**87.25%**	**94.87%**	**99.21%**
CDA-3 (Our method)	FERET	**85.21%**	**94.06%**	**98.52%**
CDA-3 (Our method)	FRGC	**87.79%**	**95.81%**	**99.17%**

D. Comparison with State-of-the-art Methods

In order to show the robustness of the proposed CDA-based eye detection method and to compare it with the state-of-the-art eye detection methods, we finally implement experiments on the BioID face database [11] and the FERET database [25]. The BioID database consists of 1,521 frontal face images of 23 subjects. The FERET database consists of 3,365 full frontal face images of nearly 1,000 subjects.

The methods we compare with include those used by Jesorsky et al. [11], Hamouz et al. [12], [15], Cristinacce et al. [16], Asteriadis et al. [17], Bai et al. [18], Niu et al. [19], Campadelli et al. [8], [20], Valenti et al. [21], and Monzo et al. [22]. All above methods applied the same normalized error criterion to evaluate the performance. Note that for our CDA-based eye detection method, we only list the performance using the CDA-3 model, which gives the slightly better detection performance than the CDA-1 and the CDA-2 models.

Table IV shows the performance comparisons between our method and the methods mentioned above for the normalized error of 0.05, 0.10, and 0.25, respectively. Our results are highlighted in the bold text. Note that for the performance which is in-explicitly reported by the authors, the results are estimated from the graphs in the literature.

Table IV reveals that for the normalized error of 0.05, our method outperforms all other state-of-the-art methods on both the BioID and the FERET databases. For the normalized error of 0.10, our method gives comparable performance to the best result on the BioID database and a bit lower performance than the best result on the FERET database. For the normalized error of 0.25, our method outperforms all other methods on the BioID database and gives comparable performance to the best result on the FERET database. Table IV also shows the detection performance of our method on the FRGC database. The performance of our method on all there databases is very close to each other, which indicates its robustness.

VI. CONCLUSION

This paper first proposes three clustering-based discriminatory analysis (CDA) models to address the problem of inadequate features derived from the FLD. The CDA-1 model defines a new within-cluster scatter matrix S_w^c and a new between-cluster scatter matrix S_b^c by means of the k-means clustering technique. The rank of the S_b^c increases as the number of clusters increases, and therefore the CDA-1 can derive adequate features for achieving satisfactory performance. The CDA-2 and CDA-3 models define a nonparametric form of N-S_b^c. These N-S_b^c are full rank, and consequently both CDA-2 and CDA-3 can derive adequate features for classification. Furthermore, the nonparametric nature of the between-cluster scatter matrices inherently leads to the derived features that preserve the structure important for classification.

This paper then presents an accurate CDA-based eye detection method. Experiments on three widely used face databases show that (i) the CDA significantly improves the performance of the conventional discriminant analysis methods, and (ii) the proposed CDA-based eye detection method achieves good eye detection performance and outperforms some state-of-the-art eye detection methods.

REFERENCES

[1] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. New York, NY, USA: Academic, 1990.

[2] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 467–476, Apr. 2002.

[3] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[4] M. Loog and R. P. W. Duin, "Linear dimensionality reduction via a heteroscedastic extension of LDA: The Chernoff criterion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 6, pp. 732–739, Jun. 2004.

[5] X. Wang and X. Tang, "A unified framework for subspace face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1222–1228, Sep. 2004.

[6] M. Zhu and A. M. Martínez, "Subclass discriminant analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 8, pp. 1274–1286, Aug. 2006.

[7] K. Fukunaga and J. Mantock, "Nonparametric discriminant analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 5, no. 6, pp. 671–678, Nov. 1983.

[8] P. Campadelli, R. Lanzarotti, and G. Lipori, "Precise eye localization through a general-to-specific model definition," in *Proc. Brit. Mach. Vis. Conf.*, Edinburgh, U.K., Sep. 2006, pp. 187–196.

[9] S. Chen and C. Liu, "Eye detection using color information and a new efficient SVM," in *Proc. 4th IEEE Int. Conf. Biometr., Theory, Appl. Syst.*, Sep. 2010, pp. 1–6.

[10] S. Chen and C. Liu, "A new efficient SVM and its application to a real-time accurate eye localization system," in *Proc. Int. Joint Conf. Neural Netw.*, San Jose, CA, USA, Jul./Aug. 2011, pp. 2520–2527.

[11] O. Jesorsky, K. J. Kirchberg, and R. Frischholz, "Robust face detection using the Hausdorff distance," in *Proc. 3rd Int. Conf. AVBPA*, Halmstad, Sweden, Jun. 2001, pp. 90–95.

[12] M. Hamouz, J. Kittler, J.-K. Kamarainen, P. Paalanen, and H. Kälviäinen, "Affine-invariant face detection and localization using GMM-based feature detector and enhanced appearance model," in *Proc. 6th Int. Conf. Autom. Face Gesture Recognit.*, Seoul, Korea, 2004, pp. 67–72.

[13] S. Chen and C. Liu, "Fast eye detection using different color spaces," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Anchorage, AK, USA, Oct. 2011, pp. 521–526.

[14] S. Chen and C. Liu, "Precise eye detection using discriminating HOG features," in *Proc. 14th Int. Conf. Comput. Anal. Images Patterns*, Seville, Spain, Aug. 2011, pp. 443–450.

[15] M. Hamouz, J. Kittler, J.-K. Kamarainen, P. Paalanen, H. Kälviäinen, and J. Matas, "Feature-based affine-invariant localization of faces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 9, pp. 1490–1495, Sep. 2005.

[16] D. Cristinacce, T. Cootes, and I. Scott, "A multi-stage approach to facial feature detection," in *Proc. 15th Brit. Mach. Vis. Conf.*, London, U.K., 2004, pp. 277–286.

[17] S. Asteriadis, N. Nikolaidis, A. Hajdu, and I. Pitas, "An eye detection algorithm using pixel to edge information," in *Proc. 2nd Int. Symp. Control, Commun., Signal Process.*, 2006, pp. 1–4.

[18] L. Bai, L. Shen, and Y. Wang, "A novel eye location algorithm based on radial symmetry transform," in *Proc. 18th ICPR*, Hong Kong, Aug. 2006, pp. 511–514.

[19] Z. Niu, S. Shan, S. Yan, X. Chen, and W. Gao, "2D cascaded adaboost for eye localization," in *Proc. 18th ICPR*, Hong Kong, Aug. 2006, pp. 1216–1219.

[20] P. Campadelli, R. Lanzarotti, and G. Lipori, "Precise eye and mouth localization," *Int. J. Pattern Recognit. Aritif. Intell.*, vol. 23, no. 3, pp. 359–377, 2009.

[21] R. Valenti and T. Gevers, "Accurate eye center location and tracking using isophote curvature," in *Proc. IEEE Comput. Soc. Conf. CVPR*, Anchorage, AK, USA, Jun. 2008, pp. 1–8.

[22] D. Monzo, A. Albiol, J. Sastre, and A. Albiol, "Precise eye localization using hog descriptors," *Mach. Vis. Appl.*, vol. 22, no. 3, pp. 471–480, 2011.

[23] S. Chen and C. Liu, "Discriminant analysis of haar features for accurate eye detection," in *Proc. 15th Int. Conf. Image Process., Comput. Vis., Pattern Recognit.*, Las Vegas, NV, USA, Jul. 2011, pp. 1–7.

[24] P. Phillips, P. Flynn, and T. Scruggs, "Overview of the face recognition grand challenge," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 947–954.

[25] P. Phillips, H. Moon, S. Rizvi, and P. Rauss, "The FERET evaluation methodology for face recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000.

[26] D. L. Swets and J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 831–836, Aug. 1996.

[27] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.-R. Muller, "Fisher discriminant analysis with kernels," in *Proc. IEEE Signal Process. Soc. Workshop Neural Netw. Signal Process.*, Aug. 1999, pp. 41–48.

[28] L.-F. Chen, H.-Y. M. Liao, M.-T. Ko, J.-C. Lin, and G.-J. Yu, "A new LDA-based face recognition system which can solve the small sample size problem," *Pattern Recognit.*, vol. 33, no. 10, pp. 1713–1726, 2000.

[29] X. Wang and X. Tang, "Dual-space linear discriminant analysis for face recognition," in *Proc. IEEE Comput. Soc. Conf. CVPR*, Jun./Jul. 2004, pp. 564–569.

[30] H. Kong, L. Wang, E. K. Teoh, J.-G. Wang, and R. Venkateswarlu, "A framework of 2D Fisher discriminant analysis: Application to face recognition with small number of training samples," in *Proc. IEEE Comput. Soc. Conf. CVPR*, Jun. 2005, pp. 1083–1088.

[31] Z. Li, D. Lin, and X. Tang, "Nonparametric discriminant analysis for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 4, pp. 755–761, Apr. 2009.

[32] X. Qiu and L. Wu, "Two-dimensional nearest neighbor discriminant analysis," *Neurocomputing*, vol. 70, nos. 13–15, pp. 2572–2575, 2007.

[33] X. Qiu and L. Wu, "Face recognition by stepwise nonparametric margin maximum criterion," in *Proc. 10th IEEE Int. Conf. Comput. Vis.*, Beijing, China, Oct. 2005, pp. 1567–1572.

[34] Z. Gu, J. Yang, and L. Zhang, "Push-Pull marginal discriminant analysis for feature extraction," *Pattern Recognit. Lett.*, vol. 31, no. 15, pp. 2345–2352, 2010.

[35] C. Liu, "A Bayesian discriminating features method for face detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 6, pp. 725–740, Jun. 2003.

[36] Z. Liu and C. Liu, "A hybrid color and frequency features method for face recognition," *IEEE Trans. Image Process.*, vol. 17, no. 10, pp. 1975–1980, Oct. 2008.

[37] C. Papageorgiou, M. Oren, and T. Poggio, "A general framework for object detection," in *Proc. IEEE 6th Int. Conf. Comput. Vis.*, Jan. 1998, pp. 555–562.

[38] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Kauai, HI, USA, Dec. 2001, pp. 511–518.

Shuo Chen received the Ph.D. degree in computer science from the New Jersey Institute of Technology in 2012. His research interests are in computer vision, machine learning, pattern recognition, object detection, biometrics, and image and video analysis. His present work focuses on eye detection, simplified support vector machines, new image descriptors, and nonparametric discriminant analysis.

Chengjun Liu is the Director of the Face Recognition and Video Processing Laboratory, New Jersey Institute of Technology. His research is mainly in pattern recognition, machine learning, computer vision, image and video analysis, and security. He has developed the class of new methods includes the evolutionary pursuit method, the enhanced Fisher models, the Gabor-Fisher classifier, the Bayesian discriminating features method, the kernel Fisher analysis method, new color models, new image descriptors, and new similarity measures.