Two-Dimensional Heteroscedastic Discriminant Analysis for Facial Gender Classification

Junying Gan

School of Information, Wuyi University

Jiangmen, Guangdong 529020, China

E-mail: jygan@wyu.cn, junyinggan@163.com

Sibin He

School of Information, Wuyi University

Jiangmen, Guangdong 529020 China

E-mail: hesibin123@163.com

Abstract

In this paper, a novel discriminant analysis named two-dimensional Heteroscedastic Discriminant Analysis (2DHDA) is presented, and used for gender classification. In 2DHDA, equal within-class covariance constraint is removed. Firstly, the criterion of 2DHDA is defined according to that of 2DLDA. Secondly, the criterion of 2DHDA, log and rearranging terms are taken, and then the optimal projection matrix is solved by gradient descent algorithm. Thirdly, face images are projected onto the optimal projection matrix, thus the 2DHDA features are extracted. Finally, Nearest Neighbor classifier is selected to perform gender classification. Experimental results show that higher recognition rate is obtained by way of 2DHDA compared with 2DLDA and HDA.

Keywords: Gender classification, Two-dimensional teteroscedastic discriminant analysis, Two-dimensional linear discriminant analysis

1. Introduction

Gender classification using face images is a challenging work due to the similarity between male and female face images. Thus, discriminant feature extraction is a key step to improve recognition rate. Linear Discriminant analysis (LDA) is a well-known approach for feature extraction and dimensional reduction. However, it often encounters the Small Sample Size problem (S3 problem) when the number of samples is less than the dimensionality of samples. Then, two-dimensional Linear Discriminant analysis (2DLDA) is proposed, in which discriminant features are extracted directly from 2-D images without a vectorization procedure, the computation cost is reduced and the S3 problem is overcome. However, in both of LDA and 2DLDA, it is assumed that the covariance matrices are equal for all sample classes. Thus, when the within-class covariance of each sample class is significantly unequal, optimal performances can not be gained by LDA and 2DLDA.

Heteroscedastic Discriminant analysis (HDA) is extended from LDA, in which equal within-class covariance constraint is removed. HDA can be viewed as a constrained Maximum likelihood (ML) projection, the constraint is given by the maximization of the projected between-class covariance volume and each class a single full covariance Gaussian model is satisfied. HDA is widely used in speech recognition and recognition rate is greatly increased than that of LDA. But in 1D-based approaches, the transformation matrix is difficult to calculate due to high dimensionality and extreme sparseness of the data. In this paper, based on 2DLDA and HDA, two-dimensional Heteroscedastic Discriminant analysis (2DHDA) is presented and used for gender classification. Firstly, the criterion of 2DHDA is defined, and log and rearranging terms are taken, then optimal projection matrix is solved by gradient descent algorithm. Secondly, face images are projected onto the optimal projection matrix, thus the discrimination features of face images are extracted.

Finally, Nearest Neighbor classifier is selected to perform gender classification**.** Experimental results show the validity of 2DHDA method.

2. Presented Approach

Suppose there are C sample classes, represented by $A^1, A^2, A^3, \text{L}, A^c$ respectively. The total number of samples is N and each class includes n samples, that is $nc = N$. $A_j^i \in \mathbf{R}^{m \times l}$ denotes the j th $(j = 1, 2, 3, \text{L}, n)$ sample which belongs to the i th $(i = 1, 2, 3, \text{L}, C)$ class. Thus, the mean of the i th sample class is $\overline{A^i} = \frac{1}{n} \sum_{j=1}^{n} A_j^i$, and the global mean of all samples is $\overline{A} = \frac{1}{N} \sum_{i=1}^{c} \sum_{j=1}^{n} A_j^i$.

2.1 2DLDA Approach

2DLDA's criterion is defined as

$$J(\boldsymbol{\theta}_{\text{2DLDA}}) = \frac{\left| \boldsymbol{\theta}_{\text{2DLDA}}^{\text{T}} \boldsymbol{S}_b \boldsymbol{\theta}_{\text{2DLDA}} \right|}{\left| \boldsymbol{\theta}_{\text{2DLDA}}^{\text{T}} \boldsymbol{S}_w \boldsymbol{\theta}_{\text{2DLDA}} \right|} \tag{1}$$

where $\boldsymbol{S}_w$ is called within-class covariance matrix and $\boldsymbol{S}_b$ is called between-class covariance matrix of training samples, expressed respectively as

$$\boldsymbol{S}_w = \frac{1}{N} \sum_{i=1}^{c} \sum_{j=1}^{n} \left(A_j^i - \overline{A^i} \right)^{\text{T}} \left(A_j^i - \overline{A^i} \right) \tag{2}$$

$$\boldsymbol{S}_b = \frac{1}{N} \sum_{i=1}^{c} n \left(\overline{A^i} - \overline{A} \right)^{\text{T}} \left(\overline{A^i} - \overline{A} \right) \tag{3}$$

Transformation matrix $\boldsymbol{\theta}_{\text{2DLDA}}$ is calculated by the solution of the eigenvalue and eigenvector problem of $\boldsymbol{S}_b \boldsymbol{S}_w^{-1}$.

2.2 2DHDA Approach

2DHDA is the heteroscedastic extension of 2DLDA. In 2DHDA, equal within-class covariance constraint is removed and the criterion is defined which maximizes the class discrimination in the projected subspace. The criterion of 2DHDA is defined as

$$J(\boldsymbol{\theta}_{\text{2DHDA}}) = \prod_{i=1}^{C} \left(\frac{\left| \boldsymbol{\theta}_{\text{2DHDA}}^{\text{T}} \boldsymbol{S}_b \boldsymbol{\theta}_{\text{2DHDA}} \right|}{\left| \boldsymbol{\theta}_{\text{2DHDA}}^{\text{T}} \boldsymbol{W}_i \boldsymbol{\theta}_{\text{2DHDA}} \right|} \right)^n = \frac{\left| \boldsymbol{\theta}_{\text{2DHDA}}^{\text{T}} \boldsymbol{S}_b \boldsymbol{\theta}_{\text{2DHDA}} \right|^N}{\prod_{i=1}^{C} \left| \boldsymbol{\theta}_{\text{2DHDA}}^{\text{T}} \boldsymbol{W}_i \boldsymbol{\theta}_{\text{2DHDA}} \right|^n} \tag{4}$$

where $\boldsymbol{W}_i = \frac{1}{n} \sum_{j=1}^{n} \left(A_j^i - \overline{A^i} \right)^{\text{T}} \left(A_j^i - \overline{A^i} \right)$ denotes the covariance matrix of the i th sample class. Thus, $\boldsymbol{S}_w = \frac{1}{C} \sum_{i=1}^{c} \boldsymbol{W}_i$. According to equation (1) and (4), if covariance matrix $\boldsymbol{W}_i$ of all sample classes is assumed equal, then $J(\boldsymbol{\theta}_{\text{2DHDA}}) = \left(J(\boldsymbol{\theta}_{\text{2DLDA}}) \right)^N$. However, $\boldsymbol{\theta}_{\text{2DLDA}}$ is invariant to scale transformation of $J(\boldsymbol{\theta}_{\text{2DLDA}})$, then $\boldsymbol{\theta}_{\text{2DHDA}} = \boldsymbol{\theta}_{\text{2DLDA}}$ is satisfied and 2DHDA is become 2DLDA. By taking log and rearranging terms, we get

$$H(\boldsymbol{\theta}_{\text{2DHDA}}) \square N \log \left| \boldsymbol{\theta}_{\text{2DHDA}}^{\text{T}} \boldsymbol{S}_b \boldsymbol{\theta}_{\text{2DHDA}} \right| - \sum_{i=1}^{C} n \log \left| \boldsymbol{\theta}_{\text{2DHDA}}^{\text{T}} \boldsymbol{W}_i \boldsymbol{\theta}_{\text{2DHDA}} \right| \tag{5}$$

H has two useful properties of invariance[5]. For every nonsingular matrix $\boldsymbol{\varphi} \in \mathbf{R}^{l \times l}$, $H(\boldsymbol{\varphi} \boldsymbol{\theta}_{\text{2DHDA}}) = H(\boldsymbol{\theta}_{\text{2DHDA}})$. This means that subsequent feature space transformations of the range of c will not affect the value of the criterion. The second is that the criterion is invariant to row or column scalings of $\boldsymbol{\theta}_{\text{2DHDA}}$ or eigenvalue scalings of $\boldsymbol{\theta}_{\text{2DHDA}} \boldsymbol{\theta}_{\text{2DHDA}}^{\text{T}}$. Using matrix differentiation, the derivative of H is given by

$$\frac{\partial H(\boldsymbol{\theta}_{\text{2DHDA}})}{\partial \boldsymbol{\theta}_{\text{2DHDA}}} = 2N \left(\boldsymbol{\theta}_{\text{2DHDA}}^{\text{T}} \boldsymbol{S}_b \boldsymbol{\theta}_{\text{2DHDA}} \right)^{-1} \boldsymbol{\theta}_{\text{2DHDA}} \boldsymbol{S}_b - \sum_{i=1}^{C} 2n \left(\boldsymbol{\theta}_{\text{2DHDA}}^{\text{T}} \boldsymbol{W}_i \boldsymbol{\theta}_{\text{2DHDA}} \right)^{-1} \boldsymbol{\theta}_{\text{2DHDA}} \boldsymbol{W}_i \tag{6}$$

However, there is no close-form solution for $\frac{\partial H(\boldsymbol{\theta}_{\text{2DHDA}})}{\partial \boldsymbol{\theta}_{\text{2DHDA}}} = 0$. Instead, the gradient descent algorithm is used for the optimization of H and $\boldsymbol{\theta}_{\text{2DHDA}}$ is solved. Usually, face images are projected onto the whole $\boldsymbol{\theta}_{\text{2DHDA}}$, the most discriminant features are could not extracted, thus, former d column vectors of $\boldsymbol{\theta}_{\text{2DHDA}}$ are selected as projection axes, then, the extracted features expressed as

$$Y = A\theta_{2\text{DHDA}}(:,1:d) \tag{7}$$

where $\theta_{2\text{DHDA}}(:,1:d)$ denotes the former d column vectors of $\theta_{2\text{DHDA}}$ and Y represents the extracted feature matrix of sample A.

2.3 Nearest Neighbor classifier

After a transformation of 2DHDA, Nearest Neighbor classifier is selected to perform gender classification. Suppose Y_{test} denotes the feature matrix of an arbitrary testing sample A_{test}, Y_j^i denotes the feature matrix of training sample A_j^i. Then, the distance between Y_{test} and Y_j^i can be expressed as

$$D\left(Y_j^i, Y_{\text{test}}\right) = \sum_{k=1}^{d} \left\| Y_j^i(:,k) - Y_{\text{test}}(:,k) \right\|^2 \tag{8}$$

where $Y_j^i(:,k)$ and $Y_{\text{test}}(:,k)$ are the k th column vector of Y_j^i and Y_{test} respectively. $\left\| Y_j^i(:,k) - Y_{\text{test}}(:,k) \right\|^2$ is the Euclidean Distance between $Y_j^i(:,k)$ and $Y_{\text{test}}(:,k)$. If $D\left(Y_q^p, Y_{\text{test}}\right) = \min_{i,j} D\left(Y_j^i, Y_{\text{test}}\right)$ is satisfied, the testing sample A_{test} is classified to the p th class, where Y_q^p represents the feature matrix of training sample A_q^p, and p, q are constants.

3. Experiments

3.1 Experimental objects

Experiments are based on Feret color face database and face database from University of Essex, UK. In Feret color face database, the images are varying in position, lighting and expression. We selected 10 male individuals, 10 female individuals with each individual 20 face images. Thus, there are 400 face images for experiments. In experiments, the images are chopped and resized to 100×90, then transformed to gray-scale images, as shown in Fig. 1.

In the face database from university of Essex, the images are with a resolution of 200×180, and with each individual 20 face images that vary in position, rotation, expression and lighting. We select 19 male and 19 female individuals, totally 760 face images for gender classification experiments. The original images are color images, we transformed them to gray-scale images and chopped them with a resolution of 80×70, as shown in Fig. 2.

3.2 Experimental results and analysis

In gender classification, there are only two classes, that are male and female respectively, thus in equation (4), $C = 2$. When the gradient descent algorithm is used for the optimization of H, $\theta_{2\text{DLDA}}$ is selected as the initial matrix of $\theta_{2\text{DHDA}}$ for iterations. Finally, Nearest Neighbor classifier is used for gender classification. The classification model is shown in Fig. 3.

Based on Feret color face database, firstly, the former 5 individuals of male and female, totally 200 face images are selected as training samples, the remains as testing samples. Experimental results are shown in Fig. 4. Secondly, for male and female, the former 4, 6, 8 individuals, totally 160, 240, 320 face images are selected as training samples respectively, experimental results are listed in Table 1.

Fig. 4 illustrates that, when totally 200 images are selected as training samples, the highest recognition rate of 2DHDA is 85.00%, which is 4.5% higher than that of 2DLDA. Table 1 shows that when 320 images are selected as training samples, the recognition rate of 2DHDA is 88.75%. However, the recognition rate of 2DLDA is only 83.75% and that of HDA is only 80.00%. When 160 and 240 images are selected as training samples respectively, we can know that the recognition rates of 2DHDA are also higher than that of 2DLDA and HDA. In table 1, when HDA is used for gender classification, PCA is used as a pretreatment step for dimensional reduction.

Based on face database from university of Essex, firstly, 20 individuals with 10 male and 10 female, totally 400 face images are selected as training samples and the remains as testing samples. When different numbers of feature dimension are selected, the results are shown in Fig. 5. Then, for male and female there are 9, 11 and 13 individuals for each class, totally 360, 440 and 520 face images are selected as training samples respectively, and the remains are selected as testing samples. Experimental results are listed in Table 2.

Fig. 5 demonstrates that, when 400 images are selected as training samples, the highest recognition rate of 2DHDA is 79.44%. The highest recognition rate of 2DLDA 70.89%, which is 8.55% lower than that of 2DHDA. Table 2 expresses that, the recognition rate of 2DHDA is higher than that of 2DLDA and HDA when 360, 440 and 520 images are selected as training samples respectively.

4. Conclusions and Future Work

In this paper, we presented the 2DHDA algorithm for gender classification using face images. Experimental results based on Feret color face database and face database from University of Essex show that 2DHDA is more effective than

2DLDA and HDA algorithms. However, when the images in a more complex background condition, how to improve the recognition rate need be further studied; the gradient descent algorithm is easy to run in local optimalization. Thus, in gender classification how to select the iteration factor and iteration number to gain the global optimal results need be further studied.

References

Gsles, M.J.F. (2002). Maximum Likelihood Multiple Subspace Projection for Hidden Markomodels[J]. *IEEE Transactions on Speech and Ausio Processing*, 2002, 10: 37-47.

Jing Wu, W.A.P., Smith, Eswin, R. Hancock. (2008). Facial Gender Classification Using Shape from Shading and Weighted Principal Geodesic Analysis[J]. *Springer link*, 2008, 5112: 925-934.

Marios, Kyperountas, Anastasios, Tefas, Ioannis Pitas. (2007). Weighted Piecewise LDA for Solving the Small Sample Size Problem in Face Verification[J]. *IEEE Transactions on Neural Networks*, 2007, 18: 506-519.

Modesto, Castrillon-Santana, Quoc, C. Vuong. (2007). An Analysis of Automatic Gender Classification[J]. *Springerlink*, 2007, 4756: 271-280.

Parinya, Sanguansat, Widhyakorn, Asdornwised, Somchai Jitapunkul, Sanparith Marukatat. (2006). Two-Dimensional Discriminant Analysis of Principal Component Vectors for Face Recognition[J]. *IEICE Transactions on Information and Systems*, 2006, 7: 2164:2170.

Saon, G., Padmanabhan, M., Gopinath, R., Chen. S. (2000). Maximum Likelihood discriminant Feature Spaces[J]. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2000, 2: 1129-1132.

The face database, University of Essex, UK. http://cswww.essex.ac.uk/mv/allfaces/faces94.html

The Feret face database. http://face.nist.gov/colorferet/

Table 1. Experimental results based on Feret face database when different numbers of training samples are selected

Algorithms	160	240	320
2DHDA	84.58%	86.25%	88.75%
2DLDA	80.00%	85.62%	83.75%
HDA	76.66%	81.25%	80.00%

Table 2. Correct recognition rates based on face database from universityof Essex when different numbers of training samples are selected

Algorithms	360	440	520
2DHDA	74.50%	79.06%	77.50%
2DLDA	73.75%	70.31%	67.92%
HDA	65.75%	68.33%	69.16%

Figure 1. Face images in Feret color face database

Figure 2. Face images in face database from university of Essex
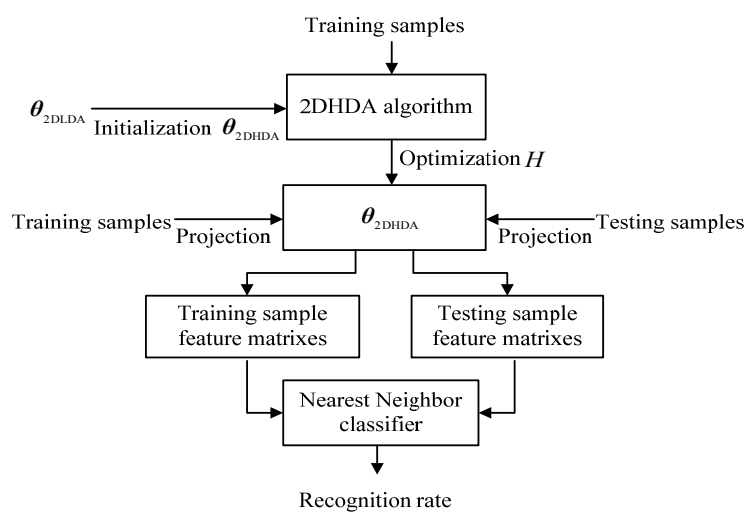
Figure 3. Gender classification model
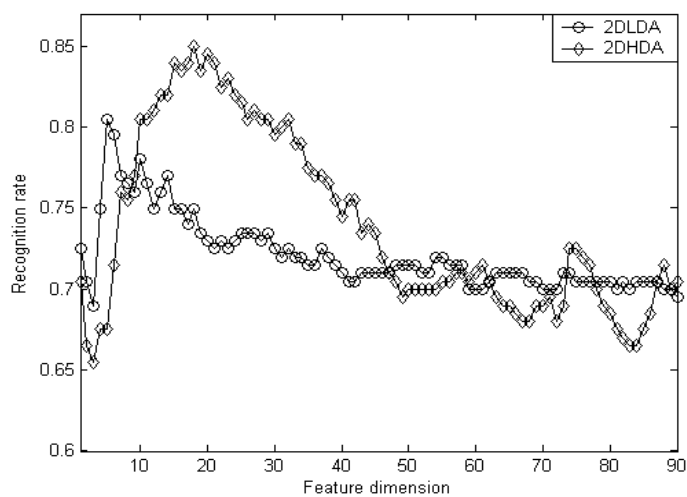
Figure 4. Experimental results based on Feret face database when 200 images selected as training samples
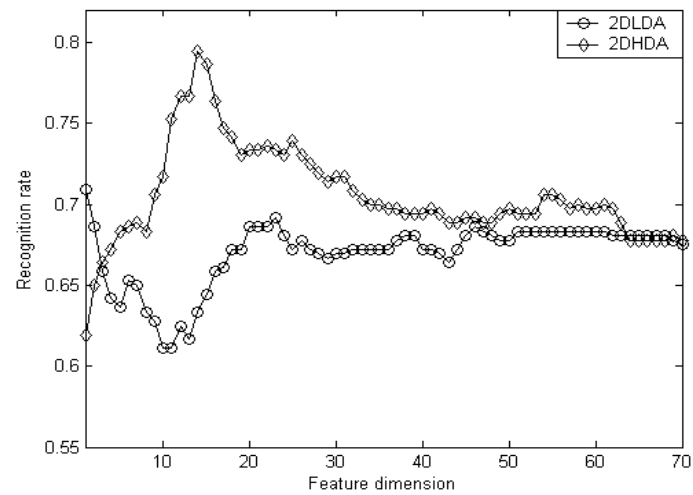
Figure 5. Experimental results with 400 training samples based on face database from university of Essex