Heteroscedastic Multilinear Discriminant Analysis for Face Recognition

Safayani M., Manzuri Shalmani M.T
Computer Engineering Department of Sharif University of Technology
safayani@ce.sharif.edu, manzuri@sharif.edu

Abstract

There is a growing attention in subspace learning using tensor-based approaches in high dimensional spaces. In this paper we first indicate that these methods suffer from the Heteroscedastic problem and then propose a new approach called Heteroscedastic Multilinear Discriminant Analysis (HMDA). Our method can solve this problem by utilizing the pairwise chernoff distance between every pair of clusters with the same index in different classes. We also show that our method is a general form of Multilinear Discriminant Analysis (MDA) approach. Experimental results on CMU-PIE, AR and AT&T face databases demonstrate that the proposed method always perform better than MDA in term of classification accuracy.

1. Introduction

One limitation of classical LDA is the implicit assumption of identical intraclass covariance matrices [1]. This assumption causes that LDA ignores the discriminative information preserved in the class covariances. Consequently, LDA cannot deal with heteroscedastic data. For solving this problem, different extensions of LDA have been proposed. A good survey of different methods can be found in [2]. However, most of these methods model a face image as a point in a high-dimensional vector space and do not consider the spatial correlation of pixels in the image. Therefore, they have high computational cost and cannot be directly applied to high-dimensional problems, such as face recognition.

On the other hand, another line of research in feature extraction considers data as a higher-order tensor [3-4]. In general, these approaches not only reduce the computational cost, by decreasing the number of projection parameters to be learned, but also preserve some implicit structural among elements of the original images. They also overcome the singularity problem of scatter matrices resulting from the high dimensionality of vectors.

There are few papers that investigate the heteroscedastic problem in the matrix-based or tensor-based discriminant analysis approaches. Recently, Zheng has investigated the heteroscedasticity of unilateral two dimensional LDA and has stated that this problem is more serious than that of the previous vector-based approaches [5]. However, he has not proposed any solution for it.

In this paper, we indicate that the main reason of the mentioned problem in multilinear-based approaches is the heteroscedasticity in the columns of the projected images. Therefore, we define different covariance matrices for the columns with the same index of image within each class and then apply the chernoff direct distance matrix for separating the class distributions. We call our method Heteroscedastic Multilinear Discriminant Analysis (HMDA) and declare that it is a general form of Multilinear Discriminant Analysis (MDA) method [4]. Experimental results on three face databases denote that our proposed method is superior to the previous tensor-based discriminant analysis approaches in term of classification accuracy.

The remaining part of the paper is organized as follows: in section 2, we describe Heteroscedastic problem in multilinear-based approaches. Section 3 introduces our algorithm. We report the experimental results on the classification accuracy in section 4. Finally, conclusions are brought in section 5.

2. Heteroscedastic Problem of Multilinear-based approaches

Zheng already showed that unilateral two-dimensional LDA had heteroscedastic problem [5]. In this section we generalize the formulation and show that this problem also exists in multilinear-based approaches such as MDA which work with higher-
order tensor data [4]. The objective function of MDA is

\[ \{U_i^*|A_i\} = \arg \max_{U_i|A_i} \sum_{c=1}^{C} \left( \prod_{j=1}^{m} p_i \right) \sum_{i \neq j}^{n} A_i \times U_i^* \times A_j \times U_j^* \right)^{-1}, \]  

(1)

where \( A_j \) is the \( j \)th sample in the dataset. \( \hat{A}_i \) is the average tensor over all the training samples, \( p_i \) is the priori probability of \( ith \) class, the class label of \( A_j \) is \( c_j \), \( N \) and \( C \) are the total number of samples and classes respectively and \( x_k U_k^|_{k=1} \) is equal to \( x_1 \times x_2 \times \ldots \times U_n \). There is not a closed-form solution for (1), so an iterative algorithm for finding the local optimal projections was proposed. In each iteration, \( U_1, \ldots, U_{k-1}, U_{k+1}, \ldots, U_n \) are assumed known, and the image samples are projected onto these projection matrices and then unfolded as follows:

\[ B_k^i = \text{mat}(A_j \times U_i^|_{i=1}^{k-1} \times A_j \times U_i^|_{i=k+1}^{n})_{k}, \]

(2)

then the optimization problem can be reformulated as a special discriminant analysis problem as:

\[ U_i^* = \arg \max_{U_i} \frac{\text{tr}(U_i^T G_i^k U_i)}{\text{tr}(U_i^T G_i^k U_i)}, \]

(3)

where \( G_i^k \) and \( G_i^k \) are \( Kth \) mode interclass and intraclass scatter matrices defined as:

\[ G_i^k = \sum_{c=1}^{C} \prod_{j=1}^{m_o} \sum_{j=1}^{m_o} p_c (b_{c,j}^k (j) - \bar{b}_c^k (j))(b_{c,j}^k (j) - \bar{b}_c^k (j))^T, \]

(4)

\[ G_i^k = \sum_{c=1}^{C} \prod_{j=1}^{m_o} \sum_{j=1}^{m_o} p_c (b_{c,j}^k (j) - \bar{b}_c^k (j))(b_{c,j}^k (j) - \bar{b}_c^k (j))^T, \]

(5)

where \( b_{c,j}^k (j) \) is the \( j \)th column of the \( B_{c,j}^k \). \( \bar{b}_c^k (j) \) is defined in the same way as \( b_{c,j}^k (j) \) with respect to the matrix \( B_{c,j}^k \). Equation (5) can be written as

\[ G_i^k = \sum_{c=1}^{C} p_c G_{i,c}^k, \]

(6)

\[ G_{i,c}^k = \sum_{j=1}^{n_c} (b_{c,j}^k (j) - \bar{b}_c^k (j))(b_{c,j}^k (j) - \bar{b}_c^k (j))^T, \]

(7)

where \( n_c \) is the total number of samples in the \( cth \) class. As can be seen from equation (7) and (8), there are two plug-in estimations for computing \( G_{i,c}^k \). First, \( G_{i,c}^k \), the covariance matrix of \( cth \) class, is estimated from the \( G_{i,c}^k \)'s, which are the sample-covariance matrix of \( jth \) column of the images in this class, then the intraclass covariance is estimated using the individual class covariances. Since, the distribution of columns of image with different indexes is substantially different i.e., \( G_{i,c}^k (i) \neq G_{i,c}^k (j) \) for \( i \neq j \), the first estimation becomes improper. The other estimation is similar to that performed in classic LDA, i.e., estimating the intraclass covariance matrix from the individual class sample-covariances which may fail due to the unequal class covariance matrices. Therefore, we can conclude that multilinear approaches such as MDA also suffer from Heteroscedastic problem. It should be noted that some recently proposed two-dimensional Heteroscedastic methods such as [6-7] only consider the second estimation and they did not deal with heteroscedasticity in the columns of the images.

### 3. Heteroscedastic Multilinear Discriminant Analysis

In the previous section, it has been shown that MDA has heteroscedastic problem. We can overcome this problem by using generalized chernoff direct distance which originally was proposed by Loog and duin [2]. This directed distance can consider the class covariances of the columns of the projected images during the optimization iterations and can extract the discriminatory information present because of heteroscedasticity of the columns of the images. We denote our method for some special cases and then generalize it to more complicated cases.

#### 3.1. Two-class case and \( U_i \in \mathbb{R}^{m_i \times 1} \)

We assume that \( U_i \in \mathbb{R}^{m_i \times 1} \) contains only one eigenvector corresponding to the leading eigenvalue, and also we assume that \( G_{i,c}^k = I \). Therefore, in this case, regarding to the equations (2) and (3), the optimization criterion becomes

\[ \text{tr}(U_i^T (\bar{b}_c^k - \bar{b}_c^k)(\bar{b}_c^k - \bar{b}_c^k)^T U_i). \]

This criterion only has one none zero eigenvalue which equals to the trace of the matrix \( G_{i,c}^k = (\bar{b}_c^k - \bar{b}_c^k)(\bar{b}_c^k - \bar{b}_c^k)^T \) and it denotes the square Euclidean distance between two-class mean. For handling heteroscedasticity of the data and keeping more discriminatory information, we replace \( G_{i,c}^k \) which its trace shows the Euclidean distance by \( G_{i,c}^k \) which its trace is equal to the chernoff distance between two-class mean.

\[ G_{i,c}^k = G_{i,c}^k + \frac{1}{P_1 P_2} \left( \frac{G_{i,c}^k}{2} \right)^{1/2} \left( \log G_{i,c}^k - P_1 \log G_{i,c}^k \right)^{1/2} \left( \log G_{i,c}^k - P_1 \log G_{i,c}^k \right)^{1/2}, \]

where \( \log(A) \) is defined as \( R \left( \log(V) \right) R^{-1} \), and \( RVR^{-1} \) is the eigenvalue decomposition of \( A \). If \( G_{i,c}^k \neq I \) we at first transform the data using \( (G_{i,c}^k)^{-1/2} \) then compute \( G_{i,c}^k \) and then apply \( (G_{i,c}^k)^{1/2} \) to transform back to the original space.
3.2. Two-class case and $U_i \in \mathbb{R}^{m_i \times m'_i}$

In this case, interclass scatter matrix $G_B^k$ can be reformulated as follows:

$$
G_B^k = \sum_{i=1}^{n} \sum_{j=i}^{m_i} \sum_{s=1}^{m'_s} p_{ij} p_s (\tilde{b}_j^k(s) - \tilde{b}_s^k(s))^T
$$

$$
= \sum_{i=1}^{n} \sum_{j=i}^{m_i} \sum_{s=1}^{m'_s} p_{ij} p_s G_{ij}^k(s),
$$

(10)

where $G_B^k(s) = (\tilde{b}_j^k(s) - \tilde{b}_s^k(s))(\tilde{b}_j^k(s) - \tilde{b}_s^k(s))^T$ is the scatter matrix which capture the difference between $sth$ column of the matrix $\tilde{B}_j^k$ and $\tilde{B}_s^k$. We generalize interclass scatter matrix by replacing $G_B^k(s)$ with chernoff scatter matrix $G_C^k(s)$.

$$
G_C^k(s) = G_B^k(s) + \frac{1}{P_1 P_2} (G_C^k)^{1/2} (\log G_C^k(s) - p_1 \log G_{C1}^k(s) - p_2 \log G_{C2}^k(s)) (G_C^k)^{1/2}.
$$

(11)

3.3. Multiclass case and $U_i \in \mathbb{R}^{m_i \times m'_i}$

According to the discussion in the previous sections and [2], interclass scatter matrix can be decomposed to:

$$
G_B^k = \sum_{i=1}^{C} \sum_{j=i}^{C} \sum_{s=1}^{m_i} \sum_{t=1}^{m'_s} p_{ij} p_s (\tilde{b}_j^k(s) - \tilde{b}_s^k(s))^T
$$

$$
= \sum_{i=1}^{C} \sum_{j=i}^{C} \sum_{s=1}^{m_i} \sum_{t=1}^{m'_s} p_{ij} p_s G_{ij}^k(s).
$$

(12)

This formula can be generalized by replacing $G_{B,i,j}^k(s)$ by $G_{C,i,j}^k(s)$ which is the chernoff scatter matrix between $sth$ column of every pair of means and defined as:

$$
G_{C,i,j}^k(s) = G_{B,i,j}^k(s) + \frac{1}{P_1 P_2} (G_{C,i,j}^k)^{1/2} (\log G_{C,i,j}^k(s) - p_1 \log G_{C1,i,j}^k(s) - p_2 \log G_{C2,i,j}^k(s)) (G_{C,i,j}^k)^{1/2},
$$

(13)

where $G_{B,i,j}^k(s) = p_1 G_{B1,i,j}^k(s) + p_2 G_{B2,i,j}^k(s)$. Therefore, $G_C^k$ obtains as follows:

$$
G_C^k = \sum_{i=1}^{C} \sum_{j=i}^{C} \sum_{s=1}^{m_i} \sum_{t=1}^{m'_s} p_{ij} p_s G_{C,i,j}^k(s).
$$

(14)

then, the new optimization formula becomes:

$$
U_k^* = \text{argmax}_{U_k^*} \frac{\text{tr}(U_k^T G_k C U_k)}{\text{tr}(U_k^T G_k W_k U_k)}. \tag{15}
$$

This optimization problem can be solved in the same way for the MDA algorithm [4]. We also regularize the within class covariance matrix as follows:

$$
\tilde{G}_{k,i,j}(s) = G_{k,i,j}(s) + \alpha I_{m_k}.
$$

(16)

The summarize procedure of HMDA is given in figure 1.

3.4 Connection to MDA

It can easily be seen that if in (14) all the covariances are the same (i.e. $G_{k,j} = \Sigma_k \forall i,s$) then $G_C^k$ reduces to $G_B^k$, equation (12), which is the interclass scatter matrix of the MDA.

Input: the sample set, $A_i \in \mathbb{R}^{m_i \times m'_i}$, $i = 1, \ldots, N$

Their class label $c_i \in \{1, 2, \ldots, C\}$, and the final lower dimensions $m_i \times m'_i \times \ldots \times m_n$.

Output: Find $U_k \in \mathbb{R}^{m_k \times m'_k} k = 1, \ldots, n$

Initialize: $U_0^k = I_{m_k} k = 1, \ldots, n$

for $t = 1, \ldots, T_{max}$

for $k = 1, \ldots, n$

$$
B_{j}^k = \text{mat}(A_j \times U_{t-1}^{[k]}), j = 1, \ldots, N
$$

Compute $G_{k,w}^k$ from (5)

$$
B_{j}^k = G_{k,w}^{1/2} B_{j}^k
$$

Compute $G_C^k$ from (14)

$$
U_k G_C^k = G_{k,w} U_k A_k, U_k \in \mathbb{R}^{m_k \times m'_k}
$$

end

Figure 1. HMDA procedure

4. Experiments

In this study, three face databases are tested. The first one is the PIE (pose, illumination, and expression) database from CMU, the second is the AR and the third is AT&T face database [8-10]. In all experiments each image is manually cropped and resized to $32 \times 32$ pixels, with 256 gray levels per pixel. The pixel values of each image is normalized to $[0, 1]$, and the resulting image is preprocessed using a histogram-equalization. The system performance is compared with Eigenface[12], Fisherface [1], DLDA [12], GLARM [13] and MDA[4] five of the most popular feature extraction methods in face recognition. Nearest neighbor has been chosen for the final classifier. We randomly select different number of images per subject ranging from 2 to 4 for training and the rest for testing. The experiments are repeated 20 times with different groups of training images, and the mean as well as standard deviation of the results are reported.

The CMU PIE face database contains 68 subjects with 41,368 face images as a whole. The subset “CMU-PIE” is established by selecting images under natural illumination for all persons from the frontal view,1/4 left/right profile and below/above in frontal view (C05, C07, C09, C27, C29). For each view, there
are three different expressions, namely natural expression, smiling and blinking. Hence there are 15 face images for each subject. Table 1 shows the average recognition accuracy of the six algorithms.

Table 1. Comparison of HMDA with other subspace algorithms on CMU-PIE face database (mean ± std) (%)

<table>
<thead>
<tr>
<th>Training Number</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenface</td>
<td>39.25±5.49</td>
<td>45.04±4.99</td>
<td>51.01±3.44</td>
</tr>
<tr>
<td>Fisherface</td>
<td>40.10±7.71</td>
<td>62.63±5.99</td>
<td>69.87±3.96</td>
</tr>
<tr>
<td>DLDA</td>
<td>14.86±6.55</td>
<td>59.81±6.15</td>
<td>68.49±4.75</td>
</tr>
<tr>
<td>GLRAM</td>
<td>54.65±6.80</td>
<td>61.46±5.78</td>
<td>66.97±2.96</td>
</tr>
<tr>
<td>MDA</td>
<td>46.99±9.79</td>
<td>58.06±6.22</td>
<td>65.56±4.28</td>
</tr>
<tr>
<td>HMDA</td>
<td>55.71±7.84</td>
<td>63.69±5.60</td>
<td>68.59±3.96</td>
</tr>
</tbody>
</table>

The third database which we used in our experiments is the AT&T face database. This database contains 650 face images of 126 different individuals (70 men and 56 women). In our experiments, we use a subset of the AR face database which contains 650 face images corresponding to 50 persons (25 men and 25 women), where each person has 13 different images. Figure 2 shows some examples from this database. The top recognition rates of different methods are shown in Table 2.

Figure 2. Samples from AR face database.

Table 2. Comparison of HMDA with other subspace algorithms on AR face database (mean ± std) (%)

<table>
<thead>
<tr>
<th>Training Number</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenface</td>
<td>44.35±12.85</td>
<td>44.03±10.78</td>
<td>52.34±11.23</td>
</tr>
<tr>
<td>Fisherface</td>
<td>63.62±5.81</td>
<td>76.86±10.78</td>
<td>88.41±10.07</td>
</tr>
<tr>
<td>DLDA</td>
<td>27.89±12.88</td>
<td>75.91±11.02</td>
<td>87.88±10.77</td>
</tr>
<tr>
<td>GLRAM</td>
<td>67.22±7.61</td>
<td>71.14±7.97</td>
<td>78.61±7.19</td>
</tr>
<tr>
<td>MDA</td>
<td>73.20±4.80</td>
<td>78.66±7.65</td>
<td>87.34±7.42</td>
</tr>
<tr>
<td>HMDA</td>
<td>76.25±6.00</td>
<td>80.53±8.32</td>
<td>89.02±7.42</td>
</tr>
</tbody>
</table>

The third database which we used in our experiments is the AT&T face database. This database contains images from 40 individuals, each providing 10 different images. Table 3 summarizes the average recognition accuracies of different algorithms.

Table 3. Comparison of HMDA with other subspace algorithms on AT&T face database (mean ± std) (%)

<table>
<thead>
<tr>
<th>Training Number</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenface</td>
<td>66.98±3.99</td>
<td>77.02±3.40</td>
<td>80.92±2.30</td>
</tr>
<tr>
<td>Fisherface</td>
<td>70.09±4.06</td>
<td>85.70±2.97</td>
<td>91.21±2.09</td>
</tr>
<tr>
<td>DLDA</td>
<td>37.53±13.37</td>
<td>85.54±2.77</td>
<td>91.98±2.27</td>
</tr>
<tr>
<td>GLRAM</td>
<td>79.36±2.76</td>
<td>87.46±2.68</td>
<td>91.42±1.82</td>
</tr>
<tr>
<td>MDA</td>
<td>78.50±2.58</td>
<td>88.66±1.95</td>
<td>92.44±1.98</td>
</tr>
<tr>
<td>HMDA</td>
<td>81.77±2.92</td>
<td>89.66±2.15</td>
<td>93.17±1.75</td>
</tr>
</tbody>
</table>

5. Conclusions

In this paper a novel approach called HMDA for solving the heteroscedastic problem of recent multilinear subspace methods was proposed. We showed that MDA had two plug-in estimations and if the data of the columns with different indexes were heteroscedastic, then those estimations would be improper. We applied the pairwise chernoff criterion for solving this problem and showed that HMDA is a general form of MDA. Experimental results on three databases showed that HMDA always perform better than MDA method no matter how many training samples per individual are used.

Acknowledgements

The support of the Iran Telecommunication Research Center (ITRC) is gratefully acknowledged.

References