1. Introduction

Face recognition is a hot research topic of biometrics and several approaches have been proposed recently. Among all these schemes, Linear Discriminant Analysis (LDA) achieves the top-level performance and it is also one of the most fundamental and important techniques for feature extraction. LDA aims to
search the directions that maximize the inter-class scatter $S_b$ and minimize the intra-class scatter $S_w$ simultaneously.

Despite successes of LDA in many applications, it still has the following intrinsic drawbacks: (1) LDA is optimal only in the cases that the data for each class is approximately Gaussian distributed with the same covariance matrix, which cannot be satisfied in real world applications; (2) the number of available projection directions is lower than the class number; and (3) the intra-class scatter $S_w$ is usually singular in face recognition.

For the third problem: in literature, lots of the research has been done. For example, Belhumeur et al.\textsuperscript{1} proposed the PCA+LDA strategy, namely LDA/Fisherface. The feature dimension can be reduced by the PCA stage, and thus, $S_w$ is no longer singular. However, LDA/Fisherface only uses the principal subspace information, while throwing away the null subspace of $S_w$. Chen et al.\textsuperscript{2} showed that the null subspace of $S_w$ contains the most discriminative information, so they choose the projection vectors by maximizing $S_b$ with the constraint that $S_w$ is zero. However, the discriminant information in the principal subspace of $S_w$ is discarded. Yu et al.\textsuperscript{16} considered the null space of $S_b$ is useless under the LDA criterion, thus, they first removed the null space of $S_b$, and then they choose the projection vectors by minimize $S_w$. Ye et al.\textsuperscript{15} proposed the LDA/Generalized Singular Value Decomposition (LDA/GSVD).

For the first two problems: in the previous research,\textsuperscript{14} the Margin Fisher Analysis (MFA) has been proposed to address them. MFA measures the intra-class compactness with the distance between each data point and its neighboring points of the same class, while measuring the inter-class separability with the class margins. Compared with the original LDA criterion, MFA has the following advantages: (1) the available projection directions are much higher than those of LDA; (2) there is no assumption on the data distribution, e.g. Gaussian distribution assumption, as a result it is more general for discriminant analysis; and (3) without the prior assumption on data distributions, the inter-class margin can better characterize the separability of different classes than the inter-class scatter in LDA. However, MFA only uses the information in the principal subspace while throwing away the information in the null space while LDA/GSVD still suffers from the Gaussian distribution assumption and the limitation of the maximum projection directions.

Thus, in this paper, we propose a new supervised scheme, which is named the Generalized Margin Fisher Analysis (GMFA) by incorporating the advantages of MFA and LDA/GSVD seamlessly. Actually, GMFA overcomes all the reported three major drawbacks of the traditional LDA algorithms, and shows superiority to LDA/Fisherface, LDA/GSVD and MFA, as demonstrated by exhaustive experiments.

The rest of the paper is structured as follows. After briefly introducing MFA and GSVD, GMFA is proposed in Sec. 2. Then, in Sec. 3, the experiments are performed on the real face data and show that the proposed GMFA outperforms the existing LDA/Fisherface, LDA/GSVD and MFA consistently. Finally, we give the conclusion remarks in Sec. 4.
2. Generalized Marginal Fisher Analysis

In this paper, we denote the sample set as \( X = [x_1, x_2, \ldots, x_N] \), \( x_i \in \mathbb{R}^m \), and \( N \) is the total number of samples. In the supervised learning problem, we denote \( l_i \) as the label of sample \( x_i \), and \( n_c \) as the total number of the \( c \)-th class. We also denote \( N_c \) as the total number of classes.

2.1. Brief review of the Marginal Fisher Analysis

The Marginal Fisher Analysis (MFA),\(^{14}\) the intra-class compactness is represented as the sum of distances between each point and its \( k_1 \)-nearest neighbors within the same class. Meanwhile, the separability of different classes is characterized as the sum of distances between the margin points and their neighboring points of different classes. More strictly speaking, the intra-class compactness is characterized by:

\[
\tilde{S}_w = \sum_i \sum_{j \in N_{k_1}^+(i) \text{ or } i \in N_{k_1}^+(j)} ||w^T x_i - w^T x_j||^2 ,
\]

where \( N_{k_1}^+(i) \) means the \( k_1 \) nearest neighbors of sample \( x_i \) of the same class. With simple deducing, Eq. (1) can be rewritten as:

\[
\tilde{S}_w = 2Tr[w^T X (D^n - W^n)X^T w] ,
\]

where \( W_{ij}^n = 1 \text{ if } j \in N_{k_1}^+(i) \text{ or } i \in N_{k_1}^+(j); 0, \text{ else} \); and \( D^n (D^n_{ii} = \sum_{j \neq i} W_{ij}^n \forall i) \) is a diagonal matrix. We also define the inter-class separability by:

\[
\tilde{S}_b = \sum_i \sum_{(i,j) \in P_{k_2}(l_i) \text{ or } (j,i) \in P_{k_2}(l_j)} ||w^T x_i - w^T x_j||^2 ,
\]

where \( P_{k_2}(c) \) is a set of data pairs that are the \( k_2 \) nearest pairs among \( \{(i,j), \ l_i = c, l_j \neq c\} \). Herein, Eq. (3) can be rewritten as

\[
\tilde{S}_b = 2Tr[w^T X (D^m - W^m)X^T w] ,
\]

where \( W_{ij}^m = 1 \text{ if } (i,j) \in P_{k_2}(l_i) \text{ or } (j,i) \in P_{k_2}(l_j); 0, \text{ else} \); and \( D^m \) is a diagonal matrix with \( D_{ii}^m = \sum_{j \neq i} W_{ij}^m \forall i \).

Figure 1 illustrates this definition, the left graph, called intrinsic graph, is for the intra-class graph connection, i.e. each sample is connected to its \( k_1 \)-nearest neighbors, and the right graph, referred to as penalty graph, is for representing the inter-class separation, i.e. we only connect the margin samples. The projection direction \( w \) can be acquired by:

\[
w^* = \arg \min_w \frac{Tr[w^T X (D^n - W^n)X^T w]}{Tr[w^T X (D^m - W^m)X^T w]} .
\]
Fig. 1. The two adjacency graphs for Marginal Fisher Analysis. Note that the left adjacency graph only plots the connection edges for one sample in each class for ease of understanding.

2.2. Brief review of the LDA/GSVD

Ye et al.\textsuperscript{15} extend the classical LDA criterion and propose the following new LDA criterion

\[
\begin{align*}
\min_G \left( \text{Tr}[\left( G^T S_b G \right)^+ (G^T S_w G)] \right) \\
\text{s.t. } \text{rank}(G^T H_b) = \delta, \quad \delta > 0,
\end{align*}
\]

where $S_w$ and $S_b$ are the intra-class scatter and inter-class scatter, respectively; $+$ means the pseudo-inverse.

\[
\begin{align*}
S_w &= \sum_{i=1}^{N} (x_i - \bar{x}_l)(x_i - \bar{x}_l)^T = H_w H_w^T, \\
S_b &= \sum_{c=1}^{N_c} n_c (\bar{x}_c - \bar{x})(\bar{x}_c - \bar{x})^T = H_b H_b^T.
\end{align*}
\]

With this modified criterion, Ye et al. applied the existing mathematics tool, namely Generalized Singular Value Decomposition, to acquire the optimal solution. More details of this algorithm can be found in Ref. 7.

2.3. Generalized Marginal Fisher Analysis

By taking the advantages of both MFA and LDA/GSVD, we propose the following Generalized Margin Fisher Analysis (GMFA) for face recognition. GMFA can be described as follows.

1. PCA projection: we project the data set into the PCA subspace by retaining $N-1$ dimensions. Let $W_{PCA}$ denote the transformation matrix of PCA. $\tilde{X} = [\tilde{x}_1, \tilde{x}_2, \cdots, \tilde{x}_N]$, is the $i$th projected low-dimensional sample;
2. Compute $\tilde{S}_w$ and $\tilde{S}_b$, as defined in Eqs. (2) and (4);
3. Compute $\tilde{H}_b$ and $\tilde{H}_w$, where $\tilde{H}_b \tilde{H}_b^T = \tilde{S}_b$ and $\tilde{H}_w \tilde{H}_w^T = \tilde{S}_w$;
(4) With $\tilde{H}_b$ and $\tilde{H}_w$, performing LDA/GSVD on $\tilde{H}_b$ and $\tilde{H}_w$; denote $G^*$ be the transform matrix;
(5) $W_{GMFA} = W_{PCA}G^*$.

3. Experiments

In this section, a set of experimental results are reported to demonstrate the effectiveness of the presented GMFA algorithm in comparison with the Fisherface, LDA/GSVD and MFA. All algorithms were tested on three different benchmark face databases, FERET,$^{10}$ ORL$^{9}$ and PIE.$^{11}$

3.1. FERET database

In this experiment, as in Ref. 6, face images of seventy persons are selected from the FERET database (each person has six different images). All images are aligned by fixing the location of the two eyes and resized to $56 \times 46$ pixels. There are facial expressions (e.g. open or closed eyes, smiling or non-smiling, etc.), illumination and pose, facial details (e.g. glasses or no glasses, hair style, etc.) variances in the images. Figure 2 displays some examples of the same person.

Two sets of experiments were conducted to compare the performance of GMFA with LDA/Fisherface, LDA/GSVD and MFA. In each set of experiments, the image set is partitioned into the gallery and probe set with different numbers. For ease of representation, $Gm/Pn$ represents $m$ images per person are randomly selected for training and the remaining $n$ images are for testing. Note that in all the experiments, the nearest neighbor classifier was employed for classification.

![Figure 2. Six sample images of one subject in the FERET database.](image)

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>G4/P2 (%)</th>
<th>G3/P3 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDA/Fisherface</td>
<td>97.1</td>
<td>91.9</td>
</tr>
<tr>
<td>LDA/GSVD</td>
<td>99.3</td>
<td>92.9</td>
</tr>
<tr>
<td>MFA</td>
<td>99.3</td>
<td>92.4</td>
</tr>
<tr>
<td>GMFA</td>
<td>100.0</td>
<td>94.3</td>
</tr>
</tbody>
</table>

Table 1. Comparison of the top-one recognition accuracy (%) of different algorithms on FERET database.
3.2. ORL database

The ORL database contains 400 images of 40 individuals. Some images were captured at different times and have different variations including expression (e.g. open or closed eyes, smiling or non-smiling, etc.) and facial details (e.g. glasses or no glasses.). The images were taken with a tolerance for some tilting and rotation of the face up to 20 degrees. We also conducted two sets of experiments with the training samples for each individual varying from 3 to 2.

3.3. CMU PIE database

The CMU PIE (Pose, Illumination and Expression) is a large face image database, which contains more than 40,000 facial images of 68 individuals. The images were acquired across different poses, under variable illumination conditions and with different facial expressions. In our experiments, we used two sub-databases. In the first sub-database PIE-I, we use five near frontal poses (C27, C05, C29, C09 and C07) and illumination 10 and 13. Thus each individual has 10 images. The flash 10 is placed at the one-quarter left profile side and flash 13 is placed at the one-quarter right profile side. Data set is randomly partitioned into gallery and probe sets with given sample numbers. We conducted two sets of experiments with the training samples for each person varying from 3 to 2. In the second sub-database PIE-II, we choose the same five poses and the illumination 08, 10, 11 and 13. The flash 08 and 11 are placed near the center and the illumination can be considered as the nearly frontal illumination. Thus each individual has 20 images. We also conducted two sets of experiments with the training samples for each individual varying from 4 to 5. Figure 4 shows four examples of one individual.

Table 2. Comparison of the top-one recognition accuracy (%) of different algorithms on ORL database.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>G3/P7 (%)</th>
<th>G2/P8 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDA/Fisherface</td>
<td>87.9</td>
<td>78.8</td>
</tr>
<tr>
<td>LDA/GSVD</td>
<td>86.8</td>
<td>81.9</td>
</tr>
<tr>
<td>MFA</td>
<td>89.3</td>
<td>79.4</td>
</tr>
<tr>
<td>GMFA</td>
<td>89.6</td>
<td>84.4</td>
</tr>
</tbody>
</table>
Fig. 4. Four sample images of one subject in the PIE face database.

Table 3. Comparison of the top-one recognition accuracy (%) of different algorithms on PIE_I database.

<table>
<thead>
<tr>
<th></th>
<th>G3/P7 (%)</th>
<th>G2/P8 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDA/Fisherface</td>
<td>80.2</td>
<td>65.8</td>
</tr>
<tr>
<td>LDA/GSVD</td>
<td>85.2</td>
<td>72.6</td>
</tr>
<tr>
<td>MFA</td>
<td>84.9</td>
<td>71.0</td>
</tr>
<tr>
<td>GMFA</td>
<td>86.5</td>
<td>74.4</td>
</tr>
</tbody>
</table>

Table 4. Comparison of the top-one recognition accuracy (%) of different algorithms on PIE_II database.

<table>
<thead>
<tr>
<th></th>
<th>G4/P16 (%)</th>
<th>G5/P15 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDA/Fisherface</td>
<td>76.9</td>
<td>83.6</td>
</tr>
<tr>
<td>LDA/GSVD</td>
<td>82.5</td>
<td>88.3</td>
</tr>
<tr>
<td>MFA</td>
<td>82.5</td>
<td>87.3</td>
</tr>
<tr>
<td>GMFA</td>
<td>84.0</td>
<td>90.6</td>
</tr>
</tbody>
</table>

From the above experiments, it can be observed that: LDA/GSVD and MFA can both improve the performance of the traditional LDA/Fisherface, and show comparable performance. Meanwhile the proposed GMFA outperforms the above three algorithms.

4. Conclusions

In this paper, we develop a new algorithm, called the Generalized Margin Fisher Analysis (GMFA), for face recognition, by seamlessly incorporating the advantages of MFA and LDA/GSVD. Compared with the existing LDA based algorithm, GMFA has the following advantages: (1) the available projection directions is much higher than that of LDA; (2) there is no assumption on the data distribution, thus it is more general for discriminant analysis; (3) without the prior assumption on data distributions, the inter-class margin can better characterize the separability of different classes than the inter-class scatter in LDA; and (4) both the principal and the null subspace information are used. The experimental results on real face databases demonstrate that: GMFA outperforms the conventional algorithms, such as LDA/Fisherface, LDA/GSVD and MFA.
References


**Dong Xu** received his BSc and PhD degrees from the Electronic Engineering and Information Science Department, University of Science and Technology of China, in 2001 and 2005, respectively. His research interests include pattern recognition, computer vision, and machine learning.
Dacheng Tao received his BEng degree from the University of Science and Technology of China, Hefei, and the M.Phil. degree from the Chinese University of Hong Kong. He is currently pursuing PhD degree at the University of London, UK.

Xuelong Li works at the University of London. He has published articles in journals (IEEE T-PAMI, T-CSVT, T-IP, T-KDE, T-MM, etc.) and conferences (IEEE CVPR, ICASSP, ICDM, etc.).

Shuicheng Yan received his BSc and PhD degrees from the Applied Mathematics Department, School of Mathematical Sciences, Peking University, China, in 1999 and 2004, respectively. His research interests include computer vision and machine learning and he is a member of the IEEE. Currently he is a Postdoc at the ECE Department of University of Illinois at Urbana-Champaign.