Fuzzy Two-dimensional Principal Component Analysis and Its Application to Face Recognition

1 Zhenyang Yu, 2 Shangbing Gao
1 Faculty of Computer Engineering, Huaiyin Institute of Technology, Huaian, Jiangsu 223003, China, 106994961@qq.com
2 Faculty of Computer Engineering, Huaiyin Institute of Technology, Huaian, Jiangsu 223003, China, 26341077@qq.com

Abstract

This paper proposes a novel method, called fuzzy two-dimensional principal component analysis (F2DPCA), which combines the two-dimensional principal component analysis (2DPCA) and fuzzy set theory. 2DPCA preserve the total variance by maximizing the trace of feature variance, but 2DPCA cannot preserve local information due to pursuing maximal variance. So, the fuzzy two-dimensional principal component analysis (F2DPCA) algorithm is proposed, in which the fuzzy k-nearest neighbor (FKNN) is implemented to achieve the distribution local information of original samples. Experimental results on ORL and Yale face databases show the effectiveness of the proposed method.

Keywords: Two-Dimensional Principal Component Analysis (2DPCA); Fuzzy Two-Dimensional Principal Component Analysis (F2DPCA); Fuzzy K-Nearest Neighbor (FKNN)

1. Introduction

Face recognition has recently attracted wide attention of the researchers in biometric authentication. PCA is a classical feature extraction and data representation technique widely used in the areas of pattern recognition and computer vision. PCA [1] aims to find a linear mapping, which preserves total variance by maximizing the trace of feature variance. The optimal mapping is the leading eigenvectors of the data’s total variance matrix associated with the leading eigenvector. But in the PCA-based face recognition technique, the 2D face image matrices must be previously transformed into 1D image vectors. The resulting image vectors of faces usually lead to a high dimensional image vector space, where it is difficult to evaluate the covariance matrix accurately due to its large size and the relatively small number of training samples.

Compared with traditional PCA, 2DPCA [2] extracts image features directly from 2D image matrices rather than 1D vectors so the image matrices do not need to be transformed into vectors. An image covariance matrix is constructed from the original image matrices for feature extraction. The optimal projection axes are its orthogonal eigenvectors corresponding to its largest eigenvalues. Due to the smaller size of image variance matrix than original variance matrix, 2DPCA requires less time to extract image features and achieves a better recognition rate. Thus, 2DPCA preserve the total variance by maximizing the trace of feature variance, but 2DPCA cannot preserve local information due to pursuing maximal variance.

But in the real world, face images are always affected by variations in illumination conditions and different facial expressions. By taking advantage of the technology of fuzzy sets [3], a number of studies have been carried out for fuzzy image filtering, fuzzy image segmentation, and fuzzy edge detection with an ultimate objective to cope with the factor of uncertainty being inherently present in many problems of image processing and pattern recognition [4]. By taking advantage of the technology of fuzzy set theory, a new feature extraction method named two-dimensional fuzzy principal component analysis (2DFPCA) is proposed in this paper.

The rest of the paper is structured as follows: In Section 2 we introduce PCA, 2DPCA. In Section 3, we propose the idea and describe 2DFPCA in detail. In Section 4, experiments on Yale and ORL face database are presented to demonstrate the effectiveness of 2DFPCA. Finally, we give concluding remarks and a discussion of future work in Section 5.
2. Outline of PCA, 2DPCA

Let us consider a set of \( N \) sample \( \{x_1, x_2, \ldots, x_N\} \) taking values in an \( n \)-dimensional image space, and assume that each image belongs to one of \( c \) classes. Let us also consider a linear transformation mapping the original \( n \)-dimensional space into an \( d \)-dimensional feature space. The new feature vectors \( y_k \in \mathbb{R}^d \) are defined by the following linear transformation:

\[
y_k = w'x_k, \quad k = 1, \ldots, N
\]

where \( w \in \mathbb{R}^{N \times d} \) is a transformation matrix.

2.1 Principal component analysis (PCA)

A fundamental unsupervised dimensionality reduction method is principal component analysis (PCA). Let \( S_T \) be the total scatter matrix:

\[
S_T = \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})^T
\]

where \( \bar{x} \) is the mean of all training sample. The PCA transformation matrix is defined as:

\[
w_{opt} = \arg \max_u [tr(w'S_Tw)] = \begin{bmatrix} w_1 & w_2 & \ldots & w_d \end{bmatrix}
\]

where \( w_i = (i = 1, 2, \ldots, d) \) is the eigenvector corresponding to the largest eigenvalue of \( S_T \).

2.2 Two-dimensional PCA

2D-PCA seeks a projection direction \( w \) which maximizes the total scatter of the resulting projected samples. Yang et al. [5] chose the following criterion:

\[
J(w) = tr(S_w) = w^T G_j w
\]

where \( S_w \) is the covariance matrix of the projected feature vectors of the training samples and \( tr(S_w) \) is the trace of \( S_w G_j \) is the image covariance (scatter) matrix:

\[
G_j = \frac{1}{N} \sum_{j=1}^{N} (X_j - \bar{X})(X_j - \bar{X})^T
\]

Then the optimal projection matrix \( w = [w_1, w_2, \ldots, w_d] \) is the orthonormal eigenvectors of corresponding to the first \( d \) largest eigenvalues. Yang et al. [5] selected features on 2D images rather than 1D vector. Their method reduced computational complexity greatly compared with traditional PCA and also improved recognition rate.

3. Fuzzy two-dimensional principal component analysis (F2DPCA) criterion

3.1 Fuzzy K-Nearest Neighbor (FKNN)

How can we completely represent the distribution of these samples and improve classification performance through extracting discriminative information from these samples? Obviously, fuzzy set theory is a good choice.

In our method, fuzzy membership degree and each class center are gained through FKNN [6] algorithm. With FKNN algorithm, the computations of the membership degree can be realized through a sequence of steps:

Step1: Compute the Euclidean distance matrix between pairs of feature vectors in training set.
Step2: Set diagonal elements of this Euclidean distance matrix to infinity.
Step 3: Sort the distance matrix (treat each of its columns separately) in an ascending order. Collect the corresponding class labels of the patterns located in the closest neighborhood of the pattern under consideration (as we are concerned with ‘\( k \)’ neighbors, this returns a list of ‘\( k \)’ integers).

Step 4: Compute the membership degree to class ‘\( i \)’ for \( j \)th pattern using the expression proposed in the literature [7].

If \( i \)th is the same as the \( j \)th label of the pattern

\[
 u_{ij} = \begin{cases} 
 0.51 + 0.49 \times \left( \frac{n_{ij}}{k} \right) & \text{if } i = j \\
 0.49 \times \left( \frac{n_{ij}}{k} \right) & \text{otherwise}
\end{cases}
\]

If \( i \)th is not the same as the \( j \)th label of the pattern. In the above expression \( n_{ij} \) stands for the number of the neighbors of the \( j \)th data (pattern) that belong to the class. As usual, \( n_{ij} \) satisfies two obvious properties:

\[
 \sum_{i=1}^{c} n_{ij} = 0 < \sum_{j=1}^{N} n_{ij} < N 
\]

Taking into account the fuzzy membership degree, the mean vector of each class is

\[
 m_i = \frac{\sum_{j=1}^{N} u_{ij} x_j}{\sum_{j=1}^{N} u_{ij}}
\]

where \( p \) is a constant which controls the influence of fuzzy membership degree. Therefore, the class center matrix and the fuzzy membership matrix \( U \) can be achieved with the result of FKNN.

\[
 U = \begin{bmatrix} u_{ij} \end{bmatrix}, i = 1, 2, Lc, j = 1, 2, LN
\]

\[
 m_i = \begin{bmatrix} m_i \end{bmatrix}, i = 1, 2, Lc
\]

### 3.2 The Idea of F2DPCA

The key step of F2DPCA is how to incorporate the contribution of each training sample into the redefine of scatter matrices. With the conception of fuzzy set theory, every sample can be classified into multi classes under fuzzy membership degree, which is different to binary classification problem. The results of the FKNN are used in the computations of the statistical properties of the patterns.

\[
 FG_i = \sum_{j=1}^{N} u_{ij}^p (X_j - \bar{m}_j)(X_j - \bar{m}_j)^T
\]

where \( p \) is same as \( p \) in Eq.(3), \( m \) is the mean of all samples. So all the scatter matrices with fuzzy set theory are redefined and the contribution of each sample is incorporated.

Our optimal fuzzy projection \( W_{F2DPCA} \) follows the expression:

\[
 W_{F2DPCA} = \arg \max W^T FG_i W
\]

Then the optimal discriminant eigenvectors matrix can be calculated with the eigensystem.

\[
 FG_i W = \lambda W
\]
After obtaining the projection axes, we can form the following linear transform for a given sample $x_i$.

$$f_i = W_F^T \cdot x_i$$

The feature vector $f_i$ is used to represent the sample $x_i$ for recognition purposes.

In summary of the preceding description, the proposed algorithm is described as follows.

Step 1: Calculate the fuzzy membership degree matrix $U$ and the new class center matrix $\omega$ of training samples using Eqs. (4), and (5) respectively.

Step 2: Calculate the fuzzy between-class scatter matrix $F_{SB}$, and fuzzy within-class scatter matrix $F_{SW}$ using Eqs.(6), and Eqs.(7) respectively.

Step 3: Calculate the transformation matrix of the proposed algorithm using Eq. (9), and project all the samples on $W_F$, then obtain the projection coefficients (feature vectors) $F = W_F^T \cdot X$.

Step 4: Classify the projection coefficients.

4. Experiments

4.1 Experiments Using FERET Database

The FERET face database is a result of the FERET program, which was sponsored by the US Department of Defense through the DARPA Program [8]. It has become a standard database for testing and evaluating state-of-the-art face recognition algorithms. The proposed method was tested on a subset of the FERET database. This subset includes 1,400 images of 200 individuals (each individual has seven images). This subset involves variations in facial expression, illumination, and pose. In our experiment, the facial portion of each original images was automatically cropped based on the location of the eyes and the cropped images was resized to $80 \times 80$ pixels. Some examples images of one person are shown in Fig.1.

![Images of one person in FERET](image)

Figure 1. Images of one person in FERET

In the experiment, we use the first $l$ images per class for training and the remaining images for testing. For feature extraction, we used, respectively, PCA [9], LDA [10], Fuzzy Fisherface [11], Complete LDA [12] and the proposed method. Note that PCA, Fisherface and Fuzzy Fisherface all involve a PCA phase. In this phase, we keep nearly 98 percent image energy and select the number of principal components, $m$, as 375 and 433. The K-nearest neighborhood parameter $K$ can be chosen as $K=l-1$. The justification for this choice is that each sample should connect with the remaining $l-1$ samples of the same class provided that within-class samples are well clustered in the observation space. The parameter $p$ in the complete fuzzy LDA is set as $p=2$. Finally a nearest neighbor classifier with cosine distance is employed. The final recognition rate of each method and the corresponding dimension are given in Table2. Table 1 show that the proposed method (CFLDA) has a better performance than others.

From Table 1, we can see that the proposed method (CFLDA) outperforms other methods. Why can CFLDA outperform other methods? In our opinion, First the overlapping sample’s distribution information is completely incorporated in the redefinition of corresponding scatter matrices by fuzzy
set theory, which is important for classification. Second, the discriminative information in the null space of fuzzy within-class scatter matrix is considered in the feature extraction.

<table>
<thead>
<tr>
<th>Table 1. Recognition Comparison On FERET</th>
</tr>
</thead>
<tbody>
<tr>
<td>$l=4$</td>
</tr>
<tr>
<td>PCA</td>
</tr>
<tr>
<td>LDA</td>
</tr>
<tr>
<td>F-Fisherface</td>
</tr>
<tr>
<td>C-LDA</td>
</tr>
<tr>
<td>Proposed</td>
</tr>
</tbody>
</table>

From Table 2, we can see that Fuzzy method improves the performance of the MSD method markedly under the minimal distance classifiers with Euclidean distance. While under the nearest neighbor classifier, no matter what distance (Euclidean distance and Cosine distance) is adopted, the performance of the Fuzzy MSD is not improved on the whole. Why this happens? As we know, the fuzzy method changes the distribution of the samples, but which does not change the class of the samples. When classification with the minimal distance classifier, we calculate the distance between test sample and the fuzzy class centers, so the performance is improved.

| Table 2. The mean and standard deviation of recognition rates (%) of PCA, LDA, MSD, Fuzzy LDA, and the proposed method (Fuzzy 2DPCA) on the FERET database under the nearest neighbor (NN) classifiers and the minimal distance (MD) classifiers with Cosine distance and Euclidean distance when the three samples ($i$, $i+1$, $i+2$) per class are used for training |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| Methods       | Euclidean       | Cosine          |
|               | MD              | NN              | MD              | NN              |
| PCA           | 42.04 ± 9.96    | 52.75 ± 7.94    | 40.3 ± 8.74     | 47.93 ± 8.32    |
| LDA           | 49.57 ± 11.45   | 44.29 ± 10.81   | 51.05 ± 10.75   | 50.2 ± 10.41    |
| MSD           | 49.52 ± 10.22   | 55.79 ± 8.63    | 57.86 ± 10.57   | 57.39 ± 11.48   |
| Fuzzy LDA     | 51.45 ± 9.8     | 45.07 ± 8.58    | 53.66 ± 9.45    | 51.48 ± 9.74    |
| Fuzzy 2DPCA   | 57.91 ± 8.54    | 55.95 ± 8.11    | 59.63 ± 9.67    | 57.89 ± 9.25    |

4.2 Experiments Using Yale Database

The Yale face database contains 165 images of 15 individuals (each person providing 11 different images) under various facial expressions and lighting conditions. In our experiments, each image was manually cropped and resized to 10. Figure 3 shows sample images of one person.

![Fig 3. Eleven images of one person in Yale](image)

The experiment was performed using the first $l(l=3,4)$ images per class for training, and the remaining five images for testing. For feature extraction, we used, respectively, PCA, LDA, Complete LDA, Fuzzy Fisherface and the proposed method. In the PCA phase of PCA, LDA and Fuzzy
Fisherface, we keep nearly 98 percent image energy and select the number of principal components, \(m\), as 34 and 43. The K-nearest neighborhood parameter \(K\) in Fuzzy Fisherface and Complete Fuzzy LDA can be chosen as \(K=1\). The parameter \(p\) in the Complete Fuzzy LDA is set as \(p=40\). Finally, the nearest neighbor (NN) classifier with cosine distance is employed for classification. The maximal recognition rate of each method and the corresponding dimensions are given in Table 2. Table 3 shows that the proposed method (CFLDA) has a better performance than others.

<table>
<thead>
<tr>
<th></th>
<th>(l=3)</th>
<th>(l=4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>0.8417(34)</td>
<td>0.8762(43)</td>
</tr>
<tr>
<td>LDA</td>
<td>0.8167(14)</td>
<td>0.8467(14)</td>
</tr>
<tr>
<td>F-Fisherface</td>
<td>0.8167(14)</td>
<td>0.8286(14)</td>
</tr>
<tr>
<td>C-LDA</td>
<td>0.8750(15)</td>
<td>0.9048(16)</td>
</tr>
<tr>
<td><strong>Proposed</strong></td>
<td><strong>0.8917(15)</strong></td>
<td><strong>0.9143(16)</strong></td>
</tr>
</tbody>
</table>

5. Summary

We have proposed a Complete Fuzzy 2DPCA method for feature extraction and face recognition by completely incorporating the distribution of the samples and considering the discriminative information in the null space of the fuzzy within-class scatter matrix. By virtue of Fuzzy K-Nearest neighbor, we can get the class membership of the binary labeled faces. This in turn allows us to compute fuzzy within-class scatter matrix and fuzzy between-class scatter matrix forming the core portion of the original 2DPCA classifier. By doing this, we could reduce the sensitivity of the method to substantial variants between face images caused by large pose, expression or illumination variations. Further, we consider the discriminative information in the null space of fuzzy within-class scatter matrix, which is important for classification.

Acknowledgements

This work was supported by the grant from Natural Science Research Project of Jiangsu Provincial Colleges and Universities, No. 11KJD520003, the Jiang Su Qing Lan project and Natural Science Research Key Project of Jiangsu Provincial Colleges and Universities, No. 11KJA460001.

References


