Orthogonal Discriminant Sparse Maximum Margin Analysis

Yu'e Lin, Jinlin Xu, Yu Zhang and Xingzhu Liang

ABSTRACT

Feature extraction is a crucial step for face recognition. In this paper, a new feature extraction method called orthogonal discriminant sparse maximum margin analysis (ODSMMA) is proposed for face recognition. ODSMMA defines two parameterless discriminant weighted matrices through the sparse representation. ODSMMA can efficiently preserve the margin between the same class and maximizes the margin between different classes simultaneously without setting any parameters. Moreover, by taking the advantage of the maximum margin criterion, ODSMMA is able to extract the orthogonal discriminant vectors in the feature space and does not suffer the small sample size problem. Experimental results on ORL database indicate the effectiveness of the proposed method.

INTRODUCTION

The goal of dimensionality reduction is to construct a meaningful low dimensional representation of high-dimensional data, which is favorable for recognition tasks. Two of the most well-known dimensionality reduction methods are principal component analysis (PCA) [1] and linear discriminant analysis (LDA)[2]. PCA is an unsupervised method and LDA is a supervised method. Recent studies have shown that high dimensional data possibly reside on a nonlinear sub-manifold. However, both PCA and LDA only preserve the global structures of the samples and fail to discover the intrinsic geometry of data. In order to preserve the local structures of the samples, a large number of manifold learning methods are proposed. The most representative manifold learning method is locality preserving projection (LPP)[3]. LPP can find an explicit linear mapping suitable for training and testing samples and preserve the intrinsic geometry of the samples. Motivated by LPP, many local discriminant approaches[4-6] have been developed for image classification, among which the most prevalent ones include margin fisher analysis (MFA)[4], locality preserving discriminant projection (LPDP)[5], local fisher discriminant analysis (LFDA)[6]. MFA considers the intrinsic geometry of the values among nearby data and seems to be more efficient than other dimensionality reduction methods for face recognition. However, there are still some limitations existed in MFA. The first shortcoming concerns the
selection of parameters. MFA need to determine the neighborhood size \( k \) of \( k \)-nearest neighbor. However, MFA is very difficult to set in advance the number of the nearest neighbors of each sample and the performance of MFA is highly dependent on the choice of the parameter value. Secondly, similar with LDA, MFA also suffers from the small sample size problem when dealing with high-dimensional data recognition task. At the same time, the basis vectors of MFA are statistically correlated. Sparse representation [7] has received considerable interest in the last few years. Sparse representation does not need to determine the model parameters such as the neighborhood size \( k \) of \( k \)-nearest neighbor. The advantage of being parameter-free makes sparse representation easy to use in practice. Based on sparse representation, Qiao et al.[8] proposed sparsity preserving projection (SPP), in which every sample was presented as a linear combination of the remaining samples. SPP tried to find a projection which can preserve the sparse reconstructive relationship. However, SPP neglects the labels information of the samples and fails to preserve the local intrinsic geometry of the samples. Motivated by MFA and SPP, we present a new method called orthogonal discriminant sparse maximum margin analysis (ODSMMA). ODSMMA can efficiently preserves the margin between the same class and maximizes the margin between the different classes simultaneously without setting any parameters. Adopting the maximum objective function, ODSMMA overcomes the small sample size problem and at the same time derives all the orthogonal optimal discriminant vectors.

THE RELATED WORK

Suppose that \( X = [x_1, \ldots, x_i, \ldots, x_N] \) is a \( n \) dimensional face sample set with \( N \) elements. There are \( C \) classes and \( N_i \) samples in the class \( i \). Thus, we have \( \sum_{i=1}^{C} N_i = N \).

MFA

The aim of marginal fisher analysis (MFA) is to find an optimal projection that can preserve intra-class compactness and inter-class separability at the same time. The intra-class scatter matrix in MFA is defined as:

\[
S_w = \sum_{i=1}^{N} \sum_{j=1}^{N} R_W(i, j) (x_i - x_j)(x_i - x_j)^T = X(D_w - R_w)X^T = XL_wX^T
\]

Where \( R_W(i, j) \) is defined as follows:

\[
R_W(i, j) = \begin{cases} 1, & \text{if } i \in N_{k_i}(j) \text{ or } j \in N_{k_i}(i) \\ 0, & \text{otherwise} \end{cases}
\]

Where \( N_{k_i}(i) \) indicates the index set of the \( k_i \) nearest neighbors of the sample \( x_i \) in the same class.

The inter-class scatter matrix in MFA is defined as:

\[
S_B = \sum_{i=1}^{N} \sum_{j=1}^{N} R_B(i, j) (x_i - x_j)(x_i - x_j)^T = X(D_B - R_B)X^T = XL_BX^T
\]

Where \( R_B(i, j) \) is defined as follows:
\[
R_g(i, j) = \begin{cases} 
1, & \text{if } (i, j) \in P_{k_2}(c) \text{ or } P_{k_2}(c_j) \\
0, & \text{otherwise}
\end{cases}
\] (4)

Where \( P_{k_2}(c) \) denotes the index set of the \( k_2 \) inter-class nearest neighbors of the sample \( x_i \). The objection function of MFA is:

\[
\min J(W) = \frac{\text{trace}(W^T S_g^T W)}{\text{trace}(W^T S_g W)}
\] (5)

**Sparsity Preserving Projection**

In sparse representation, every sample can be represented as a linear combination of the samples, and the coefficient \( s_i \) is computed by the l_1-norm optimization problem. Sparsity preserving projection aims to preserve the sparse reconstruction relationship, and its objective function is:

\[
\min J(W) = \sum_{i=1}^{n} \|W^T x_i - W^T Xs_i\|^2
\] (6)

By some simple algebra formulations, the objective function can be reduced to

\[
\min J(W) = \sum_{i=1}^{n} \|W^T x_i - W^T Xs_i\|^2 = W^T X (I - S - S^T + S^T S) X^T W
\]

\[
s.t. \quad W^T XX^T W = I
\] (7)

**ODSMMA**

In this section, we introduce the ODSMMA method based on the MFA and the SPP, which can both preserve the local similarity from same class and far away from each other form the different class without setting any parameters. Moreover, ODSMMA can overcome the small size problems and at same time derive all the orthogonal optimal discriminant vectors in the low-dimensional space.

We firstly redefine the within-class adjacent matrix \( A \) and the between-class adjacent matrix \( K \). According to SPP, every sample can be represented as a linear combination of all the training samples. In order to use the labels information, we redefine that every sample is reconstructed by the samples from the same class. Denote \( x_i^j \) belongs to the \( j \)th class. The representation coefficient \( s_i^j \) of \( x_i^j \) is defined as follows:

\[
\min_{s_i^j} \|s_i^j\|_1
\]

\[
s.t. \quad \|x_i^j - Xs_i^j\| \leq \varepsilon
\] (8)

On the other hand, every sample can be represented as a linear combination of samples from different classes. The representation coefficient \( k_i^j \) of \( x_i^j \) is defined as follows:

\[
\min_{k_i^j} \|k_i^j\|_1
\]

\[
s.t. \quad \|x_i^j - Xk_i^j\| \leq \varepsilon
\] (9)

Then we can define the column vectors \( A_i \) in weight matrix \( A \) as follows:
\[ A_i = \left[ 0, \cdots, 0, s_i^j, 0, \cdots, 0 \right]^T \]  

In the same way, we can define the column vectors \( K_i \) in weight matrix \( K \) as follows:

\[ K_i = \left[ k_i^{j,1}, k_i^{j,2}, \cdots, 0, k_i^{j,N} \right]^T \]  

Then using the within-class adjacent matrix \( A \) and the between-class adjacent matrix \( K \), the within-class scatter \( S_W^L \) and the between-class \( S_B^L \) can be expressed as:

\[
S_W^L = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij}(x_i - x_j)(x_i - x_j)^T = X(D - S)X^T = XL_wX^T 
\]  

\[
S_B^L = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} K_{ij}(x_i - x_j)(x_i - x_j)^T = X(Q - K)X^T = XL_bX^T 
\]  

ODSMMA aims to maximize Eq.13 and minimize Eq.12 simultaneously. In order to solve the small size problems, the objective function of ODSMMA can be expressed as follows:

\[
\max_{W,W=I} J(W) = \text{trace}(W^T (S_B^L - S_W^L)W) 
\]  

Finally, we apply the Lagrange multipliers to the Eq.14 and set the derivative with respect to \( W \) to zero. The objective function of ODSMMA can be converted into

\[
(S_B^L - S_W^L)w = \lambda w
\]  

Then we can obtain the eigenvectors \( w_1, w_2, \cdots, w_k \) corresponding to the first \( k \) largest eigen-values.

**EXPERIMENTAL RESULTS**

In order to test performance of our proposed method, the ORL database are used. The ORL face database contains 400 face images of 40 distinct subjects. In the experiments, the nearest distance classifier is adopted. All methods are compared on the same training sets and testing sets. In each round, \( k(3-7) \) images are randomly selected from the database. For each \( k \), 10 tests are performed and these results are averaged.

| Methods | 3    | 4    | 5    | 6    | 7    |
|---------|------|------|------|------|------|
| ODSMMA  | 90.57| 92.58| 96.25| 96.88| 97.33|
| MFA     | 88.50| 91.92| 94.50| 96.18| 96.83|
| SPP     | 86.71| 91.25| 90.60| 95.88| 96.50|

From Table 1, we find that ODSMMA is the most efficient dimensionality reduction method, and is much more efficient than SPP and MFA. The one reason may be that the ODSMMA not only overcomes the small sample size problems but also exploits the statistically orthogonal features, which is very suitable for face recognition. The other reason the parameter-free and local discriminating ability make the proposed method more efficient than the others methods.
CONCLUSIONS

In this paper, we presented a novel algorithm ODSMMA based on SPP and MFA for dimensionality reduction. Two contributions were made in this paper. (1) Combined with discriminant local structure information and the sparsity representation, ODSMMA redefines the parameterless weighted matrices, then can prevent the main geometric structure of the data and has more discriminating power without setting any parameters. (2) Adopting a difference-based optimization objective function, the optimized orthogonal transformation matrix can be computed by solving an eigen-equation, which overcomes the small size problems and further improves recognition performance. These merits make ODSMMA more robust and suitable for classification tasks.

ACKNOWLEDGEMENT

This article is supported by the Key Project of Higher Education Natural Science Foundation of Anhui Province (No. KJ2016A203) and the Master and Doctor Foundation of Anhui University of Science and Technology (No. 2010yb026).

REFERENCES

[1] M. Turk, A. Pentland. Eigenfaces for recognition. Journal of Cognitive Neuroscience, 3(1) (1991) 71 – 86.
[2] P.N. Belhumeur, J.P. Hespanha, D.J. Kiregman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence,19 (7) (1997) 711 – 720.
[3] X.F. He, S.C. Yan, Y.X. Hu, et al. Face recognition using Laplacianfaces, IEEE Transactions on Pattern Analysis and Machine Intelligence,27 (3) (2005) 328–340.
[4] D. Xu, S. Yan, D. Tao, et al. Marginal Fisher Analysis and Its Variants for Human Gait Recognition and Content-Based Image Retrieval. IEEE Transactions on Image processing, 16(11)(2007) 2811–2821.
[5] J. Gui, W. Jia, L. Zhu, et al. Locality preserving discriminant projection for face and palmprint recognition, Neurocomputing, 73(13) (2010) 2696-2707.
[6] M. Sugiyama. Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis. Journal of Machine Learning Research, 8(5), (2007) 1027–1061.
[7] J. Wright, A. Yang, A. Ganesh, et al. Robust face recognition via sparse representation. IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009)210–227.
[8] L.S. Qiao, S.C. Chen, X.Y. Tan. Sparsity preserving projections with applications to face recognition, Pattern Recognition, 43(1) (2010) 331 – 341.