

Orthogonal Discriminant Diversity and Similarity Preserving Projection for Face Recognition

Yu'e Lin^{1,a}, Fengrui Zhu¹, Kui Lu¹, Guangyu Xu¹, Xingzhu Liang¹

¹*School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan, 232001, China*

Abstract. Feature extraction is a crucial step for face recognition. In this paper, based on supervised local structure and diversity projection (SLSDP), a new feature extraction method called orthogonal discriminant diversity and similarity preserving projection (ODDSPP) is proposed for face recognition. ODDSPP defines two parameterless weighted matrices by taking into account the class label information and local structure. Thus ODDSPP could utilize both the diversity and similarity information of the data simultaneously for dimensionality reduction. Moreover, the proposed algorithm is able to extract the orthogonal discriminant vectors in the feature space and does not suffer from the small sample size problem, which is desirable for many pattern analysis applications. Experimental results on the ORL and AR databases show the effectiveness of the proposed method.

1 Introduction

Over the past few decades, a large volume of dimensionality reduction methods for face recognition have been proposed. The most well-know dimensionality reduction methods are principal component analysis (PCA)[1] and linear discriminant analysis (LDA)[2]. Recent studies have shown that large volumes of high dimensional data possibly reside on a nonlinear sub-manifold. However, both PCA and LDA fail to discover the underlying manifold structure, due to the fact that they aim only to preserve the global structures of the samples. The most representative manifold learning methods include Laplacian Eigenmap (LE)[3], Isomap [4], and Locally linear embedding (LLE)[5]. These nonlinear methods yield impressive results on some benchmark artificial datasets. However, they are difficult to find low dimensional embeddings of new test samples. Locality preserving projection (LPP)[6] is the linear approximation of LE and can compute the low dimensional embeddings for new test points. Motivated by LPP, many local discriminant approaches[7-11] have been developed for image classification, among which the most prevalent ones include margin fisher analysis (MFA)[7], locality preserving discriminant projection (LPDP)[8], local fisher discriminant analysis (LFDA)[9], local discriminant embedding(LDE)[10], locality sensitive discriminant analysis (LSDA)[11]. These methods are simple to compute and easy to operate. However, one common limitation of the above mentioned manifold-based learning algorithms is that they only pay attention on the local similarity information and neglect the diversity information of the samples. Impairing the diversity information of the samples may arise the over learning problem. In order to overcome the over learning problem, Gao et al. considered on the diversity weighted

information between points and proposed the supervised local structure and diversity projection (SLSDP)[12]. SLSDP, which explicitly considers the variation of the values among nearby data, overcomes the over learning problem successfully. However, there are still some limitations existed in SLSDP. Firstly, the diversity matrix is constructed without the guide of the class label information that is very important for recognition problem. Secondly, the neighborhood size has an impact on the learning of the manifold and is usually set in advance to the same value for all samples. Finally, similar with LDA, SLSDP also suffers from the small sample size problem (SSS) when dealing with high-dimensional data recognition task, such as face recognition.

In this paper, we present a new method called orthogonal discriminant diversity and similarity preserving projection (ODDSPP), which can directly overcome the SSS problem and at the same time derive all the orthogonal optimal discriminant vectors. The rest of this paper is organized as follows. Section 2 briefly gives an overview of related works about SLSDP. In Section 3, the dimensionality reduction method of ODDSPP is presented. In Section 4, the performance of ODDSPP is experimentally evaluated on the ORL and AR databases. At last, a brief conclusion is offered in Section 5.

2 Related work on SLSDP

The aim of SLSDP is to find an optimal projection that can preserve the similarity and the diversity information at the same time. Let S denote the local weighted matrix with elements characterizing the similarity of two close data points with same labels and B denote the local weighted matrix with elements characterizing the diversity of two close data points.

^a Corresponding author: yelin@aust.edu.cn

The elements of the local similarity weighted matrix S are defined as follows:

$$S_{ij} = \begin{cases} \exp\left(\frac{-\|x_i - x_j\|^2}{t}\right) & \text{if } x_i \in N_k(x_j) \text{ or} \\ & x_j \in N_k(x_i) \text{ and } l_i = l_j \\ 0 & \text{, otherwise} \end{cases} \quad (1)$$

Where $N_k(x_i)$ denotes the set of k nearest neighbors of x_i , l_i and l_j denote the class label of data x_i and x_j , respectively. $t \in (0, +\infty)$ is a suitable constant.

The elements of the diversity weighted matrix B are defined as follows:

$$B_{ij} = \begin{cases} \exp\left(\frac{-b}{\|x_i - x_j\|^2}\right) & \text{if } x_i \in N_k(x_j) \text{ or} \\ & x_j \in N_k(x_i) \\ 0 & \text{, otherwise} \end{cases} \quad (2)$$

Where $b \in (0, +\infty)$ is a suitable constant. B_{ij} measures the contribution of x_i relative to x_j to the diversity information.

Using the similarity matrix S , the local similarity scatter S_L can be expressed as:

$$\begin{aligned} S_L &= \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N S_{ij} (x_i - x_j)(x_i - x_j)^T \\ &= X(D - S)X^T \\ &= XLX^T \end{aligned} \quad (3)$$

Where D is a diagonal matrix whose elements on diagonal are column sum of S , i.e. $D_{ii} = \sum_j S_{ij}$.

Using the local diversity weighted matrix B , the local diversity scatter S_D can be expressed as:

$$\begin{aligned} S_D &= \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N B_{ij} (x_i - x_j)(x_i - x_j)^T \\ &= X(Q - B)X^T \\ &= XL_B X^T \end{aligned} \quad (4)$$

Where Q is a diagonal matrix whose elements on diagonal are column sum of B , i.e. $Q_{ii} = \sum_j B_{ij}$.

The objective function of SLSPP can be expressed as follows:

$$J(W) = \min_{W^T W = I} \frac{\text{trace}(W^T S_L W)}{\text{trace}(W^T S_D W)} \quad (5)$$

3 Orthogonal Discriminant Diversity and Similarity Preserving Projection

In this section, we introduce the ODDSPP method, which can preserve both the similarity and diversity information and at the same time derive all the optimal orthogonal discriminant vectors in the low-dimensional

space. We firstly redefine the discriminating similarity matrix \bar{S} and discriminating diversity matrix \bar{B} , respectively. The elements of the discriminating similarity matrix \bar{S} are defined as follows:

$$\bar{S}_{ij} = \begin{cases} \exp\left(\frac{-\|x_i - x_j\|^2}{t}\right) & \text{if } \|x_i - x_j\| \leq d_i \text{ and } l_i = l_j \\ 0 & \text{, otherwise} \end{cases} \quad (6)$$

Where d_i denotes average distance between samples with the same label. \bar{S} denotes the case that x_i and x_j have the same label.

The elements of the discriminating diversity matrix \bar{B} are defined as follows:

$$\bar{B}_{ij} = \begin{cases} \exp\left(\frac{-b}{\|x_i - x_j\|^2}\right) \left(1 - \exp\left(\frac{-b}{\|x_i - x_j\|^2}\right)\right) & \text{if } \|x_i - x_j\| \leq d_i \text{ and } l_i = l_j \\ \exp\left(\frac{-b}{\|x_i - x_j\|^2}\right) \left(1 + \exp\left(\frac{-b}{\|x_i - x_j\|^2}\right)\right) & \text{if } \|x_i - x_j\| \leq d \text{ and } l_i \neq l_j \\ 0 & \text{, otherwise} \end{cases} \quad (7)$$

Where d_i denotes average distance between samples of the same label and d denotes average distance between all samples.

From the Eq.6 and Eq.7, it is obvious to find that the discriminating similarity matrix and discriminating diversity matrix integrate both the local structure and class label information. The size of the local structure is parameterless, which is favorable for face recognition.

Then using the similarity matrix \bar{S} and diversity matrix \bar{B} , the local similarity scatter \bar{S}_L and local diversity scatter \bar{S}_D can be expressed as:

$$\begin{aligned} \bar{S}_L &= \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \bar{S}_{ij} (x_i - x_j)(x_i - x_j)^T \\ &= X(\bar{D} - \bar{S})X^T \\ &= X\bar{L}X^T \end{aligned} \quad (8)$$

and

$$\begin{aligned} \bar{S}_D &= \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \bar{B}_{ij} (x_i - x_j)(x_i - x_j)^T \\ &= X(\bar{Q} - \bar{B})X^T \\ &= X\bar{L}_B X^T \end{aligned} \quad (9)$$

Then, the objective function of ODDSP is written as follows:

$$J(W) = \min_{W^T W = I} \frac{\text{trace}(W^T \bar{S}_L W)}{\text{trace}(W^T \bar{S}_D W)} \quad (10)$$

In order to avoid the SSS problem, the objective function of ODDSP can be rewritten as:

$$J(W) = \min_{W^T W = I} \text{trace}(W^T \bar{S}_L W - W^T \bar{S}_D W) \quad (11)$$

Owing to the dimensionality of the sample very high, firstly we need to calculate the total-scatter matrix S_t eigenvectors u_1, u_2, \dots, u_r corresponding to the first r non-zero eigen-values. The total-scatter matrix S_t is defined as:

$$S_t = X(I - M)(I - M)^T X^T = H_t H_t^T \quad (12)$$

Where I is the $N \times N$ identity matrix and M is a $N \times N$ matrix with all terms equal to $1/N$.

Denote $r = \text{rank}(H_t)$, $H_t = U \Sigma E$ and $\Sigma = \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix}$.

We know that

$$S_t = U \Sigma E^T E \Sigma U^T = U \Sigma \Sigma U^T = U \begin{pmatrix} \Sigma_m^2 & 0 \\ 0 & 0 \end{pmatrix} U^T. \text{Wh}$$

ere $U = (U_1, U_2)$, $U_1 \in R^{n \times r}$, $U_2 \in R^{n \times (N-r)}$.

Denote $V = U_1^T$. Then, Eq. 11 can be rewritten as:

$$J(A) = \min_{A^T A = I} \text{trace}(A^T \bar{S}_L^V A - A^T \bar{S}_D^V A) \quad (13)$$

Where $\bar{S}_L^V = V^T \bar{S}_L V$ and $\bar{S}_D^V = V^T \bar{S}_D V$. Finally, we apply the Lagrange multipliers to the Eq.13 and set the derivative with respect to A to zero. The projection vector a that minimizes the Eq.13 is given by the minimum eigenvalue solution to the generalized eigenvalue problem

$$(\bar{S}_L^V - \bar{S}_D^V)a = \lambda a \quad (14)$$

Then we can obtain the eigenvectors a_1, a_2, \dots, a_k corresponding to the first k smallest eigen-values.

The procedure of computing ODDSP can be summarized in the following steps:

Step 1: Constructing the discriminating similarity matrix \bar{S} and the discriminating diversity matrix \bar{B} according to Eq.6 and Eq.7.

Step 2: Calculating the discriminating similarity scatter \bar{S}_L and discriminating diversity scatter \bar{S}_D according to Eq.8 and Eq. 9.

Step 3: Compute S_t according to Eq.12 and solve S_t by generalized eigenvalue decomposition, then we can obtain the basis vectors $V = U_1$;

Step4: Compute $\bar{S}_L^V = V^T \bar{S}_L V$ and $\bar{S}_D^V = V^T \bar{S}_D V$, then solve Eq.14 by generalized eigenvalue decomposition. Finally, we can obtain the optimal orthogonal transformation matrix $W = VA$.

4 Experimental Results

In order to test performance of our proposed method, the ORL and AR database are used. The ORL face database is composed of 40 distinct subjects. Each subject has 10 images under different expression and different views. Fig. 1 shows the sample images of one person in the ORL database. The AR face database contains over 4000 color face images from 126 people, including frontal views of faces with different facial expressions, lighting conditions, and occlusions. In the experiments, we choose a subset of AR database. The subset of AR database contains 100 individuals and 10 images for person. Fig. 2 shows the sample images of one person in the AR database. 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. The recognition rates of UDDPP, MFA and SLSDP are listed in Table 1 and 2.



Fig. 1. Images of one person in the ORL database



Fig. 2. Images of one person in the AR database

Table 1. The comparison of recognition rates(%) on the ORL database

methods	3	4	5	6	7
UDDPP	90.36	93.25	94.75	96.25	96.67
MFA	90.17	92.50	94.20	95.63	96.50
SLSDP	90.00	92.08	93.00	93.75	95.83

Table 2. The comparison of recognition rates(%) on the AR database

methods	3	4	5	6	7
UDDPP	92.57	94.82	95.40	95.75	96.33
MFA	91.82	94.15	94.50	95.25	95.97
SLSDP	91.71	93.66	94.00	94.50	95.66

From Table 1 and 2, we find that ODDSPP is the most efficient dimensionality reduction method, and is much more efficient than SLSDP and MFA. The one reason may be that the ODDSPP not only overcomes the small sample size problems but also exploits the orthogonal features, which is very suitable for face recognition. The other reason may be that our method can deal with the similarity and diversity information simultaneously. Furthermore, the strong structure preserving and discriminating ability make the proposed method more suitable for the recognition tasks.

5 Conclusions

In this paper, we presented a novel algorithm which is based on SLSDP, namely ODDSPP, for dimensionality reduction. Two contributions were made in this paper. (1) Combined with local structure information and the class information, ODDSPP redefines the parameterless weighted matrices, then can prevent the main geometric structure of the data and has more discriminating power. (2) Using a difference-based optimization objective function, the optimized transformation matrix can be computed by solving an eigen-equation in our method. Thus ODDSPP does not suffer from the SSS problem. These merits make our method more robust and suitable for classification tasks. In further research, we will focus on the incremental learning methods. The incremental feature set could be utilized by manifold learning to improve its recognition performance for real-time classification.

Acknowledgement

This article is supported by the Nature Science Foundation of China (No. 61471004), 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. M. Belkin, P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation, *Neural Comput*. 15 (6) (2003) 1373 - 1396
4. J. Tenenbaum, V. de Silva, J. Langford. A global geometric framework for nonlinear dimensionality reduction, *Science*. 290 (5050) (2000) 2319 - 2323
5. S.T. Roweis, L.K. Saul. Nonlinear dimensionality reduction by locally linear embedding, *Science*. 290 (5050) (2000) 2323 - 2326
6. 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
7. 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
8. J. Gui, W. Jia, L. Zhu, et al. Locality preserving discriminant projection for face and palmprint recognition, *Neurocomputing*. 73(13) (2010) 2696-2707.
9. M. Sugiyama. Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis, *Journal of Machine Learning Research*. 8(5) (2007) 1027 - 1061.
10. H.T. Chen, H.W. Chang, T.L. Liu. Local discriminant embedding and its variants, in: *Proceedings of the IEEE Computer Visual and Pattern Recognition (CVPR)*. 2005, pp. 846 - 853.
11. D. Cai, X. He, K. Zhou, et al. Locality sensitive discriminant analysis, in: *Proceedings of the 20th International Joint Conference Artificial Intelligence (IJCAI)*. 2007, pp. 708 - 713.
12. Q.X. Gao, D.Y. Xie, H. Xiu, et al. Supervised feature extraction based on information fusion of local structure and diversity information, *Acta Autom. Sin.* 36(8)(2010)1108 - 1114.