LETTER

Dual Two-Dimensional Fuzzy Class Preserving Projections for Facial Expression Recognition

Ruicong ZHI^{†a)}, Qiuqi RUAN^{†b)}, *Nonmembers*, and Jiying WU^{†c)}, *Student Member*

SUMMARY This paper proposes a novel algorithm for image feature extraction-the dual two-dimensional fuzzy class preserving projections ((2D)²FCPP). The main advantages of (2D)²FCPP over two-dimensional locality preserving projections (2DLPP) are: (1) utilizing the fuzzy assignment mechanisms to construct the weight matrix, which can improve the classification results; (2) incorporating 2DLPP and alternative 2DLPP to get a more efficient dimensionality reduction method-(2D)²LPP.

key words: *locality preserving projections, fuzzy k-nearest neighbor classifier, fuzzy class preserving projections, facial expression recognition*

1. Introduction

Facial expression plays a key role in non-verbal face-to-face communication. It is a challenging task to recognize the facial expression from a static image. Feature extraction is the most important issue for facial expression recognition because good features for representing facial expression could alleviate the complexity of the classification algorithm design. Since the intrinsic features of the facial expressions always hide in very high-dimensional space, we have to find these meaningful low-dimensional structures by dimensionality reduction.

Locality Preserving Projections (LPP) [1] is a recently proposed method for feature extraction. It is a linear dimensionality reduction method which is derived from Laplacian Eigenmap [2]. LPP aims to find an embedding that preserves local information, and obtains a face subspace that best detects the essential face manifold structure. There have been various improvements made to LPP: Orthogonal LPP [3], discriminant LPP [4], Supervised kernel LPP [5] etc. Although LPP is widely used in many domains, it suffers from the singular problem in the high dimensional image space, which makes the direct implementation of the LPP algorithm impossible. One solution is extending the vector-based LPP algorithm to matrix-based LPP, i.e. two-dimensional LPP [6], [7]. The 2DLPP algorithm can solve the singular problem, and also obtain the eigenvectors more accurately while saving time.

However, the main disadvantage of 2DLPP is that it needs many more coefficients than LPP for image representation. Thus, 2DLPP needs more memory to store its fea-

tures and takes more time to calculate distance in classification phase. Although this problem can be alleviated by using PCA after 2DLPP, it is still unclear how the dimension of 2DLPP could be reduced directly [7]. In this paper, we first indicate that 2DLPP is essentially working in the row-direction of images, and then propose an alternative 2DLPP which is working in the column-direction of images. Moreover, the weight matrix in 2DLPP is constructed on the basis that the facial expression images come fully assigned to the given classes. That is, the values of the weight matrix are nonzero if the image samples belong to the same class and zero if the image samples belong to different classes. As the face structure is complex, the differences between some facial expressions are not significant, and similar facial expressions are often confused. Consequently, we utilize a fuzzy assignment mechanism to construct the weight matrix, and this is implemented by fuzzy classification method which assigns the class labels to every sample in soft way.

According to the above analysis, we propose a novel algorithm call dual two-dimensional fuzzy class preserving projections ((2D)²FCPP). (2D)²FCPP is the combination of row and column directional 2DLPP. On the other hand, it emphasizes the membership of each sample belonging to the pattern classes, rather than the locality between samples. The new algorithm is interesting from a number of perspectives:

(1) Fuzzy weight matrix construction: adopt the fuzzy assignment mechanisms to construct the weight matrix. The weights between samples are determined by the degree of membership of each sample belonging to the given classes. The resulting fuzzy weight matrix aims to preserve the degree of each sample belonging to each class instead of preserving the relationship between samples;

(2) Dual dimensionality reduction: alternative 2DLPP is proposed to reduce the dimensionality along column directions, and (2D)²FCPP simultaneously considers the row and column directions.

As facial expressions often confused by differences between different individuals and environmental conditions, utilizing uncertain assignment can reflect all these factors in a “soft” way [8]. Fuzzy k-nearest neighbor classifier is utilized to assign the degree of membership of each sample to given classes. Furthermore, the dual dimensionality reduction scheme can reduce the number of coefficients for image representation much smaller than that of 2DLPP.

The organization of this paper is as follows: in Sect. 2, (2D)²FCPP algorithm is demonstrated in detail. In Sect. 3,

Manuscript received April 22, 2008.

Manuscript revised July 4, 2008.

[†]The authors are with Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China.

a) E-mail: 05120370@bjtu.edu.cn

b) E-mail: qqruan@center.njtu.edu.cn

c) E-mail: happywjy82@126.com

DOI: 10.1093/ietisy/e91-d.12.2880

we conduct the (2D)²FCPP on CK database. Conclusions are presented in Sect. 4.

2. Dual Two-Dimensional Fuzzy Class Preserving Projections ((2D)²FCPP)

The two important innovation of our algorithm are weight matrix construction and dual dimensionality reduction. In this section, we present the theoretical analysis of the algorithm in details.

2.1 Fuzzy Weight Matrix

In 2DLPP, the samples are assigned in a crisp way to the given classes. That means, for each sample, it fully belongs to or not belongs to the class. However, facial expressions are confused by several factors, such as individual differences, lighting conditions and intensity of the facial expression. It is more suitable to assign the samples in a “soft” way. Therefore, we construct the weight matrix by the fuzzy k -nearest neighbor classifier [9]. The weight values between two samples are determined by the degree of membership of each sample belonging to the given class. As a result, the fuzzy weight matrix aims to preserve the relationship between classes instead of preserving the relationship between samples.

Given a set of sample images $X_1, X_2, \dots, X_N$, a fuzzy “ c ”-class partition of these samples specifies the degrees of the membership of each sample to the classes [8]. The partition matrix is denoted by $U = \mu_{ij}$, for $i = 1, 2, \dots, c$, and $j = 1, 2, \dots, N$. c is the number of classes, and N is the total number of the samples. μ_{ij} denotes the membership grade of sample j belonging to class i , and it satisfies two properties:

$$(1) \sum_{i=1}^c \mu_{ij} = 1 \quad (2) \quad 0 < \sum_{j=1}^N \mu_{ij} < N$$

The main steps of constructing the fuzzy weight matrix are shown as follows:

Step1: Calculate the Euclidean distance matrix between pairs of sample images;

Step2: Set the diagonal elements of the distance matrix to infinity;

Step3: Sort each column of the distance matrix in an ascending order, and select the k -nearest neighbors. Take the class label of the k -nearest neighbors under consideration;

Step4: Using the following expression to compute the degree of membership of each sample j belonging to class i [8]:

$$\mu_{ij} = \begin{cases} 0.51 + 0.49(n_{ij}/k) & , \text{ if } gnd(j) = i \\ 0.49(n_{ij}/k) & , \text{ if } gnd(j) \neq i \end{cases} \quad (1)$$

where n_{ij} stands for the number of sample j 's neighbor which belongs to class i . $gnd(j)$ is the class label of sample j . If $n_{ij} = k$, that means all neighbors are in the same class, and then μ_{ij} returns 1.

Step5: Construct the fuzzy weight matrix $\tilde{S}$ by membership grades of each sample belonging to classes. For sample j , the weight value between sample j and sample

k is $\tilde{s}_{jk}$, then

$$\tilde{s}_{jk} = \mu_{ij} \quad (k \in \text{class } i) \quad (2)$$

As sample k belongs to class i , the weight value between sample j and sample k is denoted by the membership degree that sample j belongs to class i . We can see that, each sample is assigned a membership value for each class rather than binary decision of “belongs to” or “does not belong to”, it is assigned to classes in the fuzzy way. Therefore, the weight matrix reflects the degree of membership of each sample belonging to classes rather than the relationship between pairs of samples. The advantage of such assignment is that these membership values act as strength or confidence with which the current sample belongs to a particular class.

Step6: The fuzzy laplacian matrix is calculated as follows:

$$\tilde{L} = \tilde{D} - \tilde{S} \quad (3)$$

where $\tilde{D}$ is diagonal matrix whose entries are row sums of the fuzzy weight matrix, i.e. $\tilde{D}^{(i)} = \sum_j \tilde{S}_{ij}$.

2.2 Alternative 2DLPP

First, we indicate that 2DLPP is essentially working in the row-direction of images, and then propose an alternative 2DLPP algorithm which is working in the column-direction of images.

Let $X_k = [(X_k^{(1)})^T, (X_k^{(2)})^T, \dots, (X_k^{(m)})^T]^T$, $X_k^{(j)}$ denotes the j th row vectors of X_k . Then we have

$$\begin{aligned} X^T L X &= \sum_i X_i^T D_{ii} X_i - \sum_{i,j} X_i^T S_{ij} S_j \\ &= \sum_i D_{ii} \sum_k (X_i^{(k)})^T X_i^{(k)} - \sum_{i,j} S_{ij} \sum_k (X_i^{(k)})^T X_j^{(k)} \end{aligned} \quad (4)$$

Eq. (4) reveals that the objective function of 2DLPP can be obtained from the outer product of row vectors of images. Therefore, we claim that original 2DLPP is working in the row direction of images. Naturally, we extend the original 2DLPP to an alternative one which is working in the column direction of the images. Let V be a matrix with orthogonal columns, the image matrix X is projected to V by the transformation: $Y = V^T X$. The objective function of 2DLPP can be rewritten as

$$\begin{aligned} \sum_{i,j} \|Y_i - Y_j\|^2 S_{ij} &= \sum_{i,j} \|V^T X_i - V^T X_j\|^2 S_{ij} \\ &= V^T \left(\sum_{i,j} (X_i X_i^T - X_i X_j^T) S_{ij} \right) V \\ &= V^T \left(\sum_i X_i D_{ii} X_i^T - \sum_{i,j} X_i S_{ij} X_j^T \right) V \\ &= V^T X L X^T V \end{aligned} \quad (5)$$

Let $X_k = [(X_k^{(1)}), (X_k^{(2)}), \dots, (X_k^{(n)})]$, $X_k^{(j)}$ denotes the j th column vectors of X_k . The objective function of the alternative 2DLPP is

$$\begin{aligned}
XLX^T &= \sum_i X_i D_{ii} X_i^T - \sum_{i,j} X_i S_{ij} X_j^T \\
&= \sum_i D_{ii} \sum_k X_i^{(k)} (X_i^{(k)})^T - \sum_{i,j} S_{ij} \sum_k X_i^{(k)} (X_j^{(k)})^T \quad (6)
\end{aligned}$$

Obviously, the alternative 2DLPP utilize the outer product between column vectors of images to construct the objective function. The optimal projection matrix $V = [v_1, v_2, \dots, v_q]$ can be obtained by the following eigenvalue problem:

$$XLX^T V = \lambda XD X^T V \quad (7)$$

$v_1, v_2, \dots, v_q$ are the eigenvectors corresponding to the q smallest eigenvalues of Eq. (7).

2.3 Dual Two-Dimensional Fuzzy Class Preserving Projections

According to the discussion, 2DLPP and alternative 2DLPP learn optimal projection matrices reflecting information between rows and columns of images respectively. 2DLPP can project an $m \times n$ image to a low dimensional space of $m \times d$, while the alternative 2DLPP can project an $m \times n$ image to $q \times n$ dimensionality subspace. By combining the advantages of the two algorithms, we obtain a Dual 2DLPP algorithm which can reduce the dimensionality of image more effectively. Suppose the projection matrix on row and column directions are W and V , we project the original image matrix X to W and V simultaneously, yielding a $q \times d$ matrix Z :

$$Z = V^T X W \quad (8)$$

Z is the feature matrix used to measure the similarity between two images. When the weight matrix in Eq. (4) and Eq. (6) is the fuzzy weight matrix obtained by the scheme proposed in subsection 2.1, the Dual 2DLPP becomes Dual Two-dimensional Fuzzy Class Preserving Projections ((2D)²FCPP).

2.4 Classification

Given a test image X , first use Eq. (8) to obtain the feature matrix Z , then a nearest neighbor classifier is used for classification. If the feature matrices of training images are $Z_1, Z_2, \dots, Z_N$, the similarity between Z and Z_k is defined by

$$d(Z, Z_k) = \sqrt{\sum_{i=1}^q \sum_{j=1}^d (Z_{ij} - Z_{ij}^{(k)})^2} \quad (9)$$

If $d(Z, Z_k) = \min_j d(Z, Z_k)$ and Z_k belongs to class i , then the resulting decision is Z belongs to class i .

3. Experimental Results

In this section, experiments are carried out to show the effectiveness of our proposed (2D)²FCPP algorithm for facial

Fig. 1 Samples of face images from CK database.

Table 1 Top recognition rates of person-dependent experiment.

Methods	case 1	case 2	case 3
(2D) ² FCPP	95.14 (19 × 8)	95.69 (21 × 10)	96.33 (13 × 11)
(2D) ² LPP	93.19 (21 × 7)	93.78 (19 × 6)	95.16 (20 × 7)
2DLPP	93.54 (120 × 23)	93.24 (120 × 29)	95.67 (120 × 5)
Alternative-2DLPP	93.46 (29 × 120)	93.78 (21 × 120)	95.56 (15 × 120)

expression recognition. The experiments are conducted on the widely used Cohn-Kanade database [10]. As some subjects in CK database show less than six facial expressions, we use a subset with thirty subjects for our experiments. For each expression of a subject, the last eight frames in the videos are selected, and we treat these frames as static images for both training and testing. The images are manually cropped and resized to 120 × 120. Some of the samples in CK database are shown in Fig. 1.

3.1 Person-Dependent Recognition

In person-dependent experiments, we randomly select p images from one expression per person to construct the training set (i.e. the number of training samples of each class is $30 \times p$), and the rest of the images are used for testing. For each p , all the experiments are repeated 30 times for varying number of projection dimensions d (where $d = 1, 2, 3, \dots, 30$). Table 1 shows the comparisons of several methods on top recognition rates, where case 1, case 2 and case 3 correspond to $p = 1$, $p = 2$, $p = 3$ respectively. To verify the impact of projection dimension on (2D)²FCPP algorithm, Fig. 2 gives the recognition rate (where $p = 1$) versus different dimensions along two directions. Figure 3 reports the top recognition of each facial expression when $p = 1$. It can be seen from the illustrations that (2D)²FCPP outperforms other algorithms significantly, and it can achieve high recognition rates when the row dimension and column dimension are between 10 to 20, which has less coefficients than 2DLPP and alternative 2DLPP. Therefore, (2D)²FCPP needs less run time to compute the optimal projective matrix.

3.2 Person-Independent Recognition

The person-independent facial expression recognition is more challenging and necessary for practical applications. We use the leave-one person-out strategy to test our algorithm. It means that for each test, the images of one person are used as testing set and the remaining images are used as training set. The experiments are executed five times and the top recognition rate of (2D)²FCPP and (2D)²LPP algorithms are shown in Fig. 4. The illustration reports that (2D)²FCPP

Fig. 2 Recognition rates versus different row and column dimension.

Fig. 3 Recognition rates of six basic facial expressions on CK database.

Fig. 4 Recognition rates of person-independent experiment.

can significant improve the performance of (2D)²LPP. This is mainly because the fuzzy assignment of (2D)²FCPP can minish the impact of individual differences, and it makes

our (2D)²FCPP algorithm more efficient for facial expression recognition.

4. Conclusions

In this paper, a new algorithm called Dual Two-dimensional Fuzzy Class Preserving Projections ((2D)²FCPP) for image feature extraction was proposed. The main differences between (2D)²FCPP and 2DLPP are (a) weight matrix consists of the degree of membership of each image to the given classes; (b) projection is conducted on both row and column direction of the images. The former makes the (2D)²FCPP algorithm distinguish the confusing facial expressions correctly, and the advantage of the latter is (2D)²FCPP requires fewer number of coefficients for image representation. The features extracted by (2D)²FCPP are more efficient for dimensionality reduction and facial expression recognition.

Acknowledgments

This work is supported by the National Program 973 (No.2004CB318005) and the National Natural Science Foundation of China (No.60672062, No.60472033) and the Technological Innovation Fund of Excellent Doctoral Candidate of Beijing Jiaotong University (48036). The authors would like to thank the anonymous reviewers for their constructive advice.

References

- [1] X. He and P. Niyogi, "Locality preserving projections," Conference on Advances in Neural Information Processing System, 2003.
- [2] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," Conference on Advances in Neural Information System, 2001.
- [3] D. Cai, X. He, J. Han, and H. Zhang, "Orthogonal laplacianfaces for face recognition," IEEE Trans. Image Process., vol.15, no.11, pp.3608–3614, 2006.
- [4] W. Yu, X. Teng, and C. Liu, "Face recognition using discriminant locality preserving projections," Image Vis. Comput., vol.24, pp.239–248, 2006.
- [5] J. Cheng, Q. Liu, H. Lu, and Y. Chen, "Supervised kernel locality preserving projections for face recognition," Neurocomputing, vol.67, pp.443–449, 2005.
- [6] D. Hu, G. Feng, and Z. Zhou, "Two-dimensional locality preserving projections (2DLPP) with its application to palmprint recognition," Pattern Recognit., vol.40, pp.339–342, 2007.
- [7] S. Chen, H. Zhao, M. Kong, and B. Luo, "2D-LPP: A two-dimensional extension of locality preserving projections," Neurocomputing, vol.70, pp.912–921, 2007.
- [8] K.C. Kwak and W. Pedryca, "Face recognition using a fuzzy fisher-face classifier," Pattern Recognit., vol.38, pp.1717–1732, 2005.
- [9] J.M. Keller, M.R. Gray, and J.A. Givens, "A fuzzy k-nearest neighbor algorithm," IEEE Trans. Syst., Man Cybern., vol.15, no.4, pp.580–585, 1985.
- [10] T. Kanade, J. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," International Conference on FGR, pp.46–53, 2000.