Multiple-manifolds Discriminant Analysis for Facial Expression Recognition from Local Patches Set

Ning Zheng[1], Lin Qi[1], and Ling Guan[2(✉)]

[1] School of Information Engineering, Zhengzhou University, Zhengzhou, China
zhengning34@hotmail.com, ielqi@zzu.edu.cn
[2] Ryerson Multimedia Research Laboratory, Ryerson University, Toronto, Canada
lguan@ee.ryerson.ca

Abstract. In this paper, a novel framework is proposed for feature extraction and classification of facial expression recognition, namely multiple manifold discriminant analysis (MMDA), which assumes samples of different expressions reside on different manifolds, thereby learning multiple projection matrices from training set. In particular, MMDA first incorporates five local patches, including the regions of left and right eyes, mouth and left and right cheeks from each training sample to form a new training set, and then learns projection matrix from each expression so that maximizes the manifold margins among different expressions and minimizes the manifold distances of the same expression. A key feature of MMDA is that it can extract the discriminative information of expression-specific for classification rather than that of subject-specific, leading to a robust performance in practical applications. Our experiments on Cohn-Kanade and JAFFE databases demonstrate that MMDA can effectively enhance the discriminant power of the extracted expression features.

Keywords: Manifold learning · Facial expression recognition · Local patches · Multiple manifolds discriminant analysis

1 Introduction

Manifold learning methods have been widely applied to human emotion recognition, based on the fact that variations of expression can be represented as low dimensional manifold embedded in high dimensional data space. The original LPP [1], operated in an unsupervised manner, fails to embed the facial set in low dimensional space in which different expression classes are well clustered. Hence, supervised methods based on LPP are proposed for human emotion recognition [2]. Besides, Ptucha et al. [3] investigated the performance of combining automatic AAM landmark placement and LPP method for human emotion recognition and demonstrated the effectiveness on expression classification accuracy.

Note that the aforementioned methods assume that only one common manifold is developed from training set. However, it is difficult to determine how one manifold could well represent the structure of high dimensional data. To address this problem, Xiao et al. [4] proposed a human emotion recognition method by utilizing multiple manifolds. They claimed that different expressions may reside on different manifolds, and obtained the promising recognition performance. Lu et al. [5] presented a discriminative multimanifold analysis method to solve the single sample per person problem in face recognition, by splitting each face image into several local patches to form a training set, and sequentially learning discriminative information from each subject.

It is known that, under uncontrolled conditions, a number of specific facial areas play a more important role than the others in the formation of facial expressions and would be more robust to the variation of environmental lighting conditions. In light of the development, several methods are put forward to represent the local features. Chang et al. [7] constructed a training set of manifold from each local patch, and performed expression analysis based on local discriminant embedding method. Kotsia et al. [8] argued that local patches of facial images provide more discriminant information for recognizing emotional states.

Inspired by the aforementioned works, we propose a novel framework for feature extraction and classification of human emotion recognition from local patches set, namely multiple manifolds discriminant analysis (MMDA). MMDA first models face and obtain the landmark points of interest consisting of points from facial images based on ASM [9], and then focus on five local patches, including the regions of left and right eyes, mouth and left and right cheeks, to form a sample set for each expression. MMDA learns projection matrix of each expression so that maximizing the manifold margins among different expressions and minimizing the manifold distances of the same expression. As in [4,5], a reconstruction error criterion is employed for computing the distance of manifold-to-manifold.

2   The Proposed Method

Assume that a dataset given in R^m contains n samples from c classes x_i^k, $k = 1, 2, \cdots, c$, $i = 1, 2, \cdots, n_k$, where n_k denotes the sample size of the k-th class, $\sum_{k=1}^{c} n_k = n$ and x_i^k is the i-th sample in the k-th class. We extract five local patches from each facial image x_i^k such as the regions of two eyes, mouth and right and left cheeks, with the size of each salient patch being $a \times b$.

2.1   Problem Formation

To visually study the five local patches, we randomly pick seven facial samples with seven expressions: 'Anger' (AN), 'Sadness' (SA), 'Fear' (FE), 'Surprise' (SU), 'Disgust' (DI), 'Happiness' (HA) and 'Neutral' (NE) from Cohn-Kanade database [10]. At an intuitive level, different local patches are far apart, e.g., eyes

versus cheeks of anger, while the same local patches are very close, e.g., eyes versus eyes. Hence, it is difficult to ensure one common manifold can model the high dimensional data well and guarantee the best performance of classification. Furthermore, it is more likely that these patches of the same expressions reside on the same manifold. In this case, we can model local patches of the same expression as one manifold so that local patches with the same manifold become closer and these patches with different manifolds are far apart.

2.2   Model Formation

Let $\mathbf{M} = [M_1, \cdots, M_c] \in \Re^{d \times \hbar}$ be a set of local patches and $M_k = [P_1^k, P_2^k, \cdots, P_{n_k}^k] \in \Re^{d \times l_k}$ is the manifold of the k-th expression, where $P_i^k = [x_{i1}^k, x_{i2}^k, \cdots, x_{it}^k]$ be the patch set of the i-th facial sample in the k-th class, t is the number of local patches of each facial sample, $l_k = t \cdot n_k$ and $\hbar = \sum_{k=1}^c l_k$. The generic problem of feature extraction for MMDA is to seek c projection matrices $W_1, W_2, \cdots, W_c$ that maps manifold of each expression to low dimensional feature space. i.e., $Y_k = W_k^T M_k$, so that Y_k represents M_k well in terms of certain optimal criterion, where $W_k \in \Re^{d \times d_k}$, with d and d_k respectively denoting the dimensions of original local patch and feature space.

According to the study of Sect. 2.1, MMDA aims at maximizing the ratio of the trace of inter-manifold scatter matrix to the trace of intra-manifold scatter matrix. To achieve this goal, we formulate the proposed MMDA as the following optimization problem: (1).

$$J_1(W_1, \cdots, W_c) = \frac{\sum_{k,i,j} \sum_{\hat{x}_{ijr}^k \in N_b(x_{ij}^k)} ||W_k^T x_{ij}^k - W_k^T \hat{x}_{ijr}^k|| A_{ijr}^k}{\sum_{k,i,j} \sum_{\tilde{x}_{ijr}^k \in N_w(x_{ij}^k)} ||W_k^T x_{ij}^k - W_k^T \tilde{x}_{ijr}^k|| B_{ijr}^k} \tag{1}$$

where $N_b(x_{ij}^k)$ and $N_w(x_{ij}^k)$ denote the k_b-intermanifold neighbors and k_w-intra manifold neighbors of x_{ij}^k as well as $\tilde{x}_{ijr}^k$ denotes the rth k_b-nearest intermanifold neighbors and $\hat{x}_{ijr}^k$ represents the rth k_w-nearest intermanifold neighbors. the A_{ijr}^k, B_{ijr}^k are the weight imposed on the edge that connects x_{ij}^k with $\hat{x}_{ijr}^k \in N_b(x_{ij}^k)$ as well as that x_{ij}^k with $\tilde{x}_{ijr}^k \in N_w(x_{ij}^b)$, respectively. Just defined as in the LPP [1].

For convenience, (1) can be written in a more compact form

$$J_2(W_1, \cdots, W_c) = \frac{\sum_{k=1}^c trace(W_k^T \tilde{S}_b^k W_k)}{\sum_{k=1}^c trace(W_k^T \tilde{S}_w^k W_k)} \tag{2}$$

where $\tilde{S}_b^k = \sum_{i=1}^{n_k} \sum_{j=1}^t \sum_{\hat{x}_{ijr}^k \in N_b(x_{ij}^k)} (x_{ij}^k - \hat{x}_{ijr}^k)(x_{ij}^k - \hat{x}_{ijr}^k)^T A_{ijr}^k$, $\tilde{S}_w^k = \sum_{i=1}^{n_k} \sum_{j=1}^t \sum_{\tilde{x}_{ijr}^k \in N_w(x_{ij}^k)} (x_{ij}^k - \tilde{x}_{ijr}^k)(x_{ij}^k - \tilde{x}_{ijr}^k)^T B_{ijr}^k$ are respectively inter-manifold and intra-manifold scatter matrices of the k-th expression.

Since $(w_v^k)^T w_\varepsilon^k = \delta_{v\varepsilon}$, $\tilde{S}_b^k$ and $\tilde{S}_w^k$ are positive semi-definite matrices, it holds that $trace(W_k^T \tilde{S}_b^k W_k) \geq 0$ and $trace(W_k^T \tilde{S}_w^k W_k) > 0$, we and end up with a new optimization function from (2)

$$J_3(W_1, \cdots, W_c) = \sum_{k=1}^{c} \frac{trace(W_k^T \tilde{S}_b^k W_k)}{trace(W_k^T \tilde{S}_w^k W_k)} \tag{3}$$

without losing generality, we can easily know that $J_3(W_1, \cdots, W_c) \geq J_2(W_1, \cdots, W_c)$.

Which means that (3) can obtain more discriminating features from training set than (2). However, there is no close-form solution for simultaneously obtaining c projection matrices from (2). To address the problem, we sequentially solve each projection matrix inspired by Fisher linear discriminant criterion [11]

$$J(W_k) = \frac{trace(W_k^T \tilde{S}_b^k W_k)}{trace(W_k^T \tilde{S}_w^k W_k)} \tag{4}$$

$\tilde{S}_b^k$ can be explicitly written as shown in Eq. (5).

$$\tilde{S}_b^k = \sum_{i=1}^{n_k} \sum_{j=1}^{t} \sum_{r=1}^{k_b} (x_{ij}^k - \hat{x}_{ijr}^k)(x_{ij}^k - \hat{x}_{ijr}^k)^T A_{ijr}^k$$

$$= M_k D_k^c M_k^T - (L_b^- + L_b^{-T}) + \bar{M}_k D_k^l \bar{M}_k^T \tag{5}$$

where $L_b^- = M_k \Sigma_k \bar{M}_k^T$, Σ_k is a $l_k \times (k_b * l_k)$ matrix with entries A_{ijr}^k, $\bar{M}_k = \{\hat{x}_{ijr}^k \in N_b(x_{ij}^k)\}$, D_k^c and D_k^l are diagonal matrices with entries being the column and row sums of A_{ijr}^k, i.e., $D_k^c \leftarrow \sum_r A_{ijr}^k$ and $D_k^l \leftarrow \sum_{ij} A_{ijr}^k$.

Similarly, $\tilde{S}_w^k$ can also be reformed as shown in Eq. (6).

$$\tilde{S}_w^k = \sum_{i=1}^{n_k} \sum_{j=1}^{t} \sum_{r=1}^{k_w} (x_{ij}^k - \tilde{x}_{ijr}^k)(x_{ij}^k - \tilde{x}_{ijr}^k)^T B_{ijr}^k = 2M_k(D_k - A_k^w)M_k^T \tag{6}$$

where D_k is the diagonal matrix whose entries on the diagonal are the column sum of A_k^w and A_k^w is the matrix which is combined with entries of B_{ijr}^k.

In general, we can solve the following eigenvalue equation by Fisher discriminant criterion

$$\tilde{S}_b^k w_v^k = \lambda_v^k \tilde{S}_w^k \tag{7}$$

where $w_1^k, w_2^k, \cdots, w_{d_k}^k$ denote the eigenvectors corresponding to the d_k largest eigenvalues and $v = 1, 2, \cdots, d_k$.

Note that, for a task with high dimensional data such as facial images, (7) may encounter several difficulties. One of them is that we have to confront the issue of how to determine the feature dimension d_k for each projection matrix W_k. For this sake, we utilize a feature dimension determination method by trace ratio. In particular, because $\tilde{S}_b^k$ and $\tilde{S}_w^k$ are non-negative semi-definite matrices, we can screen out the eigenvectors corresponding to eigenvalues so that they meet the following condition

$$J_2(w_v^k) = \frac{(w_v^k)^T \tilde{S}_b^k w_v^k}{(w_v^k)^T \tilde{S}_w^k w_v^k} \geq 1 \tag{8}$$

If $J_2(w_v^k) \geq 1$, local patches reside on the same manifold (intra-manifold) are close and those patches reside on different manifolds (inter-manifold) are far apart. According to this criterion, we can automatically determine the feature dimension d_k for the k-th projection matrix W_k.

In conclusion, we summarize the steps to complete MMDA in Algorithm 1.

Algorithm 1. Multiple manifolds discriminant analysis algorithm

---

Input: $\mathbf{M} = [M_1, \cdots, M_c] \in \Re^{d \times h}$;
Output: $W_1, W_2, \cdots, W_c$, where $W_k \in \Re^{d \times d_k}$;
 1: For each local patch x_{ij}^k, calculate two weight matrices A_{ijr}^k and B_{ijr}^k respectively.
 2: For $k = 1 : c$
   1. Compute $\tilde{S}_b^k$ and $\tilde{S}_w^k$ as shown in (5) and (6), respectively.
   2. Solve the eigenvectors $[w_1^k, w_2^k, \cdots, w_{d_k}^k]$ and eigenvalues $[\lambda_1^k, \lambda_2^k, \cdots, \lambda_{d_k}^k]$ by DLDA method.
   3. Sort their eigenvectors $[w_1^k, w_2^k, \cdots, w_{d_k}^k]$ according to their associated eigenvalues: $\lambda_1^k \geq \lambda_2^k \geq \cdots \geq \lambda_{d_k}^k$. If $J_2(w_v^k) \geq 1$, simultaneously satisfy $J_2(w_{v+1}^k) < 1$, then $d_k = v$, where $v = 1, 2, \cdots, d_k - 1$.
   4. Obtain the k-th projection matrix $W_k = [w_1^k, w_2^k, \cdots, w_{d_k}^k]$.
   end

---

3   Experiments

We perform experiments on two public databases: Cohn-Kanade human emotion database [10] and Jaffe database [13], which are the most commonly used databases in the current human emotion research community.

3.1   Human Emotion Database

Cohn-Kanade database is acquired from 97 people aged from 18 to 30 years old with six prototype emotions (Anger, Disgust, Fear, Happiness, Sadness, and Surprise). In our study, 300 sequences which are selected. The selection criterion is that a sequence can be labeled as one of the six basic emotions and three peak frames of each sequence are used for processing. At last, 684 images are selected, including 19 subjects, 36 images of each subject and 6 images of each expression from each subject. Each normalized image is scaled down to the size of 128×128. Some example images in this database are depicted in Fig. 1.

JAFFE human emotion database consists of 213 images of Japanese female facial expressions. Ten subjects posed three or four examples for each of the six basic expressions. Additionally, a simple preprocessing step is applied to Jaffe database before performing training and testing. Each normalized image is scaled down to the size of 80×80. Some of the cropped face images in the Jaffe database with different human emotion are shown in Fig. 2.

Fig. 1. Six samples from Cohn-Kanade database.

Fig. 2. Six samples from Jaffe database.

3.2 Experimental Results and Analysis

In this paper, we compare the performance of MMDA with existing feature extraction and classification methods, including PCA+LDA [14], modular PCA [15], GMMSD [16], LPP [1], DLPP [17], MFA [18], Xiao's [4]. For fair comparison, we explore the performance on all possible feature dimension in the discriminant step and report the best results. The experimental results are listed in Table 1. From these results, we make several observations:

Table 1. Recognition rates of comparative methods on Cohn-Kanade and Jaffe databases

Methods	[14]	[15]	[16]	[1]	[17]	[18]	[4]	MMDA
Cohn-Kanade	65.20 %	44.08 %	78.39 %	41.95 %	59.07 %	58.72 %	84.71 %	**91.22 %**
JAFFE	50.94 %	43.45 %	61.03 %	24.85 %	43.03 %	41.73 %	68.89 %	**74.96 %**

(1) MMDA and Xiao's consistently outperform other methods, further indicating that modeling each expression as one manifold is better because the geometric structure of expression-specific can be discovered and not influenced by that of subject-specific.

(2) Comparing the performance between MMDA and Xaio's, the second best method in the comparison, reveals that MMDA encodes more discriminating information in the low-dimensional manifold subspace by preserving the local structure which is more important than the global structure for classification.

(3) It is observed that recognition performance on JAFFE database is much poorer than that on Cohn-Kanade database, likely due to the fact that there are fewer samples or subjects in the database resulting in a poor sampling of the underlying discriminant space.

In order to provide a more detailed observations, we show the corresponding mean confusion matrixes which analyze the confusion between the emotions

when applying MMDA to human emotion recognition on Cohn-Kanade and Jaffe (See Tables 2 and 3). In Table 2, we can draw the following conclusions: 'Anger', 'Happiness', 'Surprise' and 'Sadness' are better distinguished by MMDA. However, 'Disgust' obtains the worst performance in the confusion matrix. To sum up, we know that MMDA well learns expression-specific of local patches belong to 'Anger', 'Happiness', 'Surprise' and 'Sadness'. In Table 3, we see that it is very difficult to find the expression of 'Fear' accurately, which consistent with the result reported in [13].

Table 2. The confusion matrix by applying MMDA for facial expression recognition on Cohn-Kanade database

	AN	DI	FE	HA	SA	SU
AN	93.66 %	4.25 %	0	0	2.09 %	0
DI	6.52 %	78.59 %	5.66 %	0	9.23 %	0
FE	0	18.18 %	81.82 %	0	0	0
HA	0	3.06 %	0	92.42 %	0	1.52 %
SA	4.88 %	1.65 %	0	0	93.47 %	0
SU	0	0	0	2.86 %	0	90.9 %

Table 3. The confusion matrix by applying MMDA for facial expression recognition on Jaffe database

	AN	DI	FE	HA	SA	SU
AN	92.84 %	0	3.75 %	0	3.41 %	0
DI	0	85.84 %	0	14.16 %	0	0
FE	12.58 %	9.65 %	69.8 %	0	0	7.97 %
HA	0	0	18.31 %	81.69 %	0	0
SA	0	0	16.85 %	0	83.15 %	0
SU	0	0	6.6 %	2.94 %	0	90.46 %

4   Conclusions

We in this paper propose a novel model for human emotion recognition, which learns discriminative information based on the principle of multiple manifolds discriminant analysis (MMDA). Considering that local appearances can effectively reflect the structure of facial space on one manifold and provide more important discriminative information, we focus on five local patches including the regions of left and right eyes, mouth and left and right cheeks from each facial image to learn multiple manifolds features. Hence, the semantic similarity of expression from different subjects is well kept on each manifold. Extensive

experiments on Cohn-Kanade and JAFFE databases are performed. Compared with several other human emotion recognition methods, MMDA demonstrates superior performance.

References

1. He, X., Niyogi, P.: Locality preserving projections. In: NIPS, pp. 234–241 (2003)
2. Zhi, R., Ruan, Q.: Facial expression recognition based on two-dimensional discriminant locality preserving projections. J. Neurocomput. **71**(7), 1730–1734 (2008)
3. Ptucha, R., Savakis, A.: Facial expression recognition using facial features and manifold learning. In: Bebis, G., et al. (eds.) ISVC 2010, Part III. LNCS, vol. 6455, pp. 301–309. Springer, Heidelberg (2010)
4. Xiao, R., Zhao, Q., Zhang, D., Shi, P.: Facial expression recognition on multiple manifolds. J. Pattern Recogn. **44**(1), 107–116 (2011)
5. Lu, J., Peng, Y., Wang, G., Yang, G.: Discriminative multimanifold analysis for face recognition from a single training sample per person. J. Pattern Anal. Mach. Intell. **35**(1), 39–51 (2013)
6. Martinez, A.M.: Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. J. Pattern Anal. Mach. Intell. **24**(6), 748–763 (2002)
7. Chang, W.-Y., Chen, C.-S., Hung, Y.-P.: Analyzing facial expression by fusing manifolds. In: Yagi, Y., Kang, S.B., Kweon, I.S., Zha, H. (eds.) ACCV 2007, Part II. LNCS, vol. 4844, pp. 621–630. Springer, Heidelberg (2007)
8. Kotsia, I., Buciu, I., Pitas, I.: An analysis of facial expression recognition under partial facial image occlusion. J. Image Vis. Comput. **26**(7), 1052–1067 (2008)
9. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models-their training and application. J. Comput. Vis. Image Underst. **61**(1), 38–59 (1995)
10. Kanade, T., Cohn, J.F., Tian, Y.: Comprehensive database for facial expression analysis. In: 4th IEEE International Conference on Automatic Face and Gesture Recognition, pp. 46–53. IEEE (2000)
11. Fisher, R.A.: The use of multiple measurements in taxonomic problems. J. Ann. Eugen. **7**(2), 179–188 (1936)
12. Yu, H., Yang, J.: A direct LDA algorithm for high-dimensional data with application to face recognition. J. Pattern Recogn. **34**(10), 2067–2070 (2001)
13. Lyons, M., Akamatsu, S., Kamachi, M., Gyoba, J.: Coding facial expression with Gabor wavelets. In: 3th IEEE International Conference in Automatic Face and Gesture Recognition, pp. 200–205. IEEE (1998)
14. Belhumeur, P.N., Hespnha, J.P., Kriegman, D.J.: Eigenfaces vs. fisherfaces: recognition using class specific linear projection. J. Pattern Anal. Mach. Intell. **19**(7), 711–720 (1997)
15. Gottumukkal, R., Asari, V.K.: An improved face recognition technique based on modular PCA approach. J. Pattern Recogn. Lett. **25**(4), 429–436 (2004)
16. Zheng, N., Qi, L., Gao, L., Guan, L.: Generalized MMSD feature extraction using QR decomposition. In: Visual Communication and Image Processing, pp. 1–5. IEEE (2012)
17. Yu, W., Teng, X., Liu, C.: Face recognition using discriminant locality preserving projections. J. Image Vis. Comput. **24**(3), 239–248 (2006)
18. Yan, S., Xu, D., Zhang, B., Zhang, H., Yang, Q., Lin, S.: Graph embedding and extensions: a general framework for dimensionality reduction. J. Pattern Anal. Mach. Intell. **29**(1), 40–51 (2007)