

Low-resolution degradation face recognition over long distance based on CCA

Zhenyu Wang · Wankou Yang · Xianye Ben

Received: 24 September 2014 / Accepted: 22 January 2015 / Published online: 4 February 2015
© The Natural Computing Applications Forum 2015

Abstract Canonical correlation analysis (CCA) is a kind of classical multivariate analysis method. Less canonical correlation variables are used to describe the relationship between two variables completely but easily. To get high face recognition rate under low-resolution degradation over a long distance solidly, in this work, CCA is used to extract the correlation between high-resolution face images and low-resolution ones and to find the transform pair between them. Therefore, face images of the same individual with variable resolutions can be matched accurately. This is the first method that uses CCA to do low-resolution degradation face recognition over long distances. We conduct the experiments on the Extended Yale B and ORL database, and the experimental results validate the efficacy of the proposed method.

Keywords Canonical correlation analysis (CCA) · Low resolution · Correlation · Dimension matching · Degradation face recognition

1 Introduction

Face recognition is one of the most important branches of biometrics [1]. Especially in recent years, due to the governments' growing concern about the public security, face recognition has become more popular among the researchers. *Linear subspace* learning-based methods have been successfully applied on face recognition. These methods are used to extract low-dimensional features, which are more discriminant for facial image classification. Typical subspace learning-based methods include principal component analysis (PCA) [2, 3], Fisher linear discriminant analysis (LDA) [2, 3], locality preserving projections (LPP) [4] and unsupervised discriminant projection (UDP) [5]. Sparse representation and manifold learning methods are also widely exploited in face recognition [6–10].

Subspace learning-based methods could be divided into two kinds: unsupervised methods and supervised methods. LDA is a representative supervised method to learn discriminant subspace. Unfortunately, it cannot be directly applied to small sample size (SSS) problems [11] because the within-class scatter matrix is singular. Face recognition is a typical SSS problem, and many works have been proposed to use LDA for face recognition. The most popular method is Fisherface [2]. First, Fisherface uses PCA to reduce the dimensionality of the original feature; second, LDA is applied in PCA subspace. To overcome the singularity, we could add a singular value perturbation to the within-class scatter matrix [12]. Regularized discriminant analysis (RDA) is a more systematic method [13]. Another regularized version of LDA is penalized discriminant analysis (PDA) [14, 15]. PDA aims not only to address the SSS problem but also to smooth the coefficients of discriminant vectors. In face

Z. Wang · W. Yang (✉)
School of Automation, Southeast University,
Nanjing 210096, China
e-mail: wankou.yang@yahoo.com

Z. Wang · W. Yang
Key Laboratory of Measurement and Control of Complex
Systems of Engineering, Ministry of Education,
Southeast University, Nanjing 210096, China

X. Ben
School of Information Science and Engineering,
Shandong University, Jinan 250100, China

recognition, the dimension of features is often more than ten thousand. It is not practical for RDA or PDA to process high-dimensionality covariance matrices. The famous null subspace methods include LDA + PCA method [16] and direct LDA [17]. Loog [18] proposed the approximate pairwise accuracy criterion (aPAC) that used a weight to emphasize the close class pair in order to reduce the merging of close class pairs. Tao et al. [19] proposed to maximizing the geometric mean of all pairs of Kullback–Leibler (KL) divergences for subspace selection, which maximized the geometric mean of KL divergences between different class pairs.

Due to the influence of the camera hardware, surroundings and light variation, the real images are often in a very low resolution and difficult to recognize. In the criminal investigations, if we still use the traditional feature extraction strategies, then we need the image of acquisition and that of the database be normalized to the same low resolution, which severely decreases the recognition ability.

In general, low-resolution face recognition brings two problems: First, the low resolution of the face images leads to a low recognition rate; second, as for the difference between the training samples and test samples, the traditional face recognition algorithms cannot be directly applied to the low-resolution face recognition. So, it is necessary to build a reliable algorithm for the low-resolution face recognition.

Canonical correlation analysis (CCA) [20] was presented by Hotelling in 1936, which is a classical method of multivariate statistical analysis. The basic idea of the CCA is to transform the correlation between two groups to several variables. In this way, we can fully and simply depict the relationship between two variables by little canonical correlation variable. Thus, CCA is broadly used in the correlation analysis and predictive analysis between the items. Huang et al. [21] gave a face image super-resolution reconstruction using CCA. We will simply use CCA in face recognition.

If we regard the low-resolution and high-resolution images as two different groups of variable, we can use CCA to find the transform pair between them. Therefore, we can project the low- and high-resolution images to the same linear space and realize the match of images with different resolutions. To overcome the aforementioned problems, in this work, we propose a method named low-resolution degradation face recognition over long distance based on CCA. This method will connect the low- and high-resolution images by extracting their correlation and also avoid the dimension mismatch when the high-resolution images are normalized to low ones. The experiments show that the proposed method achieves promising results in several face recognition problems.

2 Low-resolution face recognition

Subspace learning methods are successfully used in face recognition. By seeking for a linear projection matrix, we project the training and testing samples to the same space, in which it is easy to classify the different kinds of samples. These projection methods require that the training samples and test samples have the same resolution, as shown in Fig. 1.

In the low-resolution degradation face recognition over long distance, it cannot meet the requirement of the same resolution between the low-resolution images and the high-resolution images. So we cannot directly apply subspace learning methods. One simple way is that we can rescale the face recognition by down-sampling the high-resolution images to the same size as the low-resolution images and employ the traditional feature extraction methods. The operation will lose much useful information, which decreases the recognition rate. We plan to seek a method that could directly classify the high-resolution images and the low-resolution degradation images.

Different from the traditional ways, we are seeking two linear transformation matrices w^h and w^l to project low- and high-resolution images to the same low-dimensional space, respectively. Then, we can make a match to recognize the face recognition between the different resolutions, as shown in Fig. 2.

As for some correlations between the different resolution face from the same person, we just need to seek a way to explore the correlation between the low- and high-resolution images. So we transform the low- and high-resolution images to a same space. The above ideas not only preserve the details of high-resolution face image, but also avoid the mismatching of dimensions.

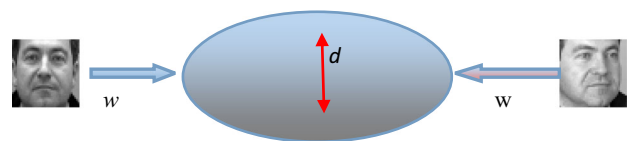


Fig. 1 Illustration of the subspace learning-based methods

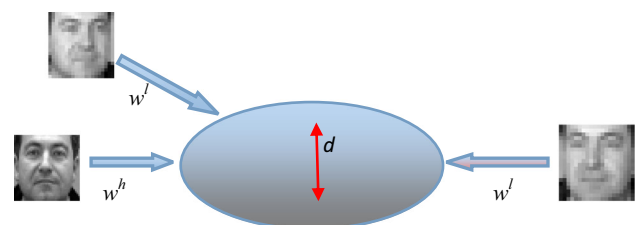


Fig. 2 Illustration of the subspace learning-based methods at different resolutions

3 CCA

CCA [20] is a kind of multivariate statistical analysis method, which is broadly applied to correlation analysis. As aforementioned, we will apply CCA to the low-resolution degradation face recognition.

3.1 The objective function

Assume that the vector dimensionality of the low- and high-resolution images is n_h and n_l , respectively. Note the m high-dimension training samples as $x^i, i = 1, 2, \dots, m$ which compose a training set X :

$$X = \begin{pmatrix} x^1 \\ x^2 \\ \vdots \\ x^m \end{pmatrix} = \begin{pmatrix} x_1^1 & x_2^1 & \cdots & x_{n_h}^1 \\ x_1^2 & x_2^2 & \cdots & x_{n_h}^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^m & x_2^m & \cdots & x_{n_h}^m \end{pmatrix}. \tag{1}$$

Similarly, m low-dimension training samples $y^i, i = 1, 2, \dots, m$ compose the sample set Y ,

$$Y = \begin{pmatrix} y^1 \\ y^2 \\ \vdots \\ y^m \end{pmatrix} = \begin{pmatrix} y_1^1 & y_2^1 & \cdots & y_{n_l}^1 \\ y_1^2 & y_2^2 & \cdots & y_{n_l}^2 \\ \vdots & \vdots & \ddots & \vdots \\ y_1^m & y_2^m & \cdots & y_{n_l}^m \end{pmatrix}. \tag{2}$$

Different from the traditional feature extract method, CCA aims to seek for a pair of projection matrix w^h and w^l and make the high-dimension samples and low-dimension samples project to the same space using the linear transformation $R^{1 \times n_d}, n_d \leq n_l$

$$\hat{x}^i = x^i w^h, \quad \hat{y}^i = y^i w^l \tag{3}$$

where $\hat{x}^i, \hat{y}^i \in R^n, i = 1, \dots, m$ can capture the maximum variation direction of the original data. But, CCA introduces the correlation between the high-dimensional samples and the low-dimensional samples, which guarantees the maximum correlation between the projected \hat{x}^i and \hat{y}^i . The CCA criterion function is as follows:

$$\arg \max_{w^h, w^l} J(w^h, w^l) = \frac{w^{hT} S_{xy} w^l}{\sqrt{w^{hT} S_{xx} w^h \cdot w^{lT} S_{yy} w^l}} \tag{4}$$

s.t. $w^{hT} w^h = I, \quad w^{lT} w^l = I$

In problem (4), S_{xx} and S_{yy} are the covariance matrix of X and Y , respectively, and S_{xy} is the cross-covariance matrix of sample set X and sample set Y , as shown below

$$S_{xx} = \frac{1}{n} \sum_{i=1}^n (x^i - \bar{x})^T (x^i - \bar{x}) \tag{5}$$

$$S_{yy} = \frac{1}{n} \sum_{i=1}^n (y^i - \bar{y})^T (y^i - \bar{y}) \tag{6}$$

$$S_{xy} = \frac{1}{n} \sum_{i=1}^n (x^i - \bar{x})^T (y^i - \bar{y}) \tag{7}$$

3.2 Solution of the target problem

According to the objective function (4), we build the following Lagrange multiplier function

$$L(w^h, w^l) = w^{hT} S_{xy} w^l - \frac{\lambda_1}{2} (w^{hT} S_{xx} w^h - 1) - \frac{\lambda_2}{2} (w^{lT} S_{yy} w^l - 1) \tag{8}$$

where λ_1 and λ_2 are Lagrange multipliers. Let the partial derivative to the projection axis w^h and w^l be zero; then, we have:

$$\frac{\partial L}{\partial w^h} = S_{xy} w^l - \lambda_1 S_{xx} w^h = 0 \Leftrightarrow w^{hT} S_{xy} w^l = \lambda_1 w^{hT} S_{xx} w^h \tag{9}$$

$$\frac{\partial L}{\partial w^l} = S_{yx} w^h - \lambda_2 S_{yy} w^l = 0 \Leftrightarrow w^{lT} S_{yx} w^h = \lambda_2 w^{lT} S_{yy} w^l \tag{10}$$

As we know $S_{xy} = S_{yx}$, thus,

$$w^{hT} S_{xy} w^l = w^{lT} S_{yx} w^h \tag{11}$$

According to Eqs. (9)–(11), we get

$$\begin{aligned} \lambda_1 w^{hT} S_{xx} w^h &= \lambda_2 w^{lT} S_{yy} w^l \Leftrightarrow w^{hT} S_{xx} w^h = \frac{\lambda_2}{\lambda_1} w^{lT} S_{yy} w^l \\ &\Leftrightarrow w^{hT} S_{xx} w^h = a w^{lT} S_{yy} w^l, \end{aligned} \tag{12}$$

where $a = \lambda_2 / \lambda_1$. By substituting Eq. (12) into Eqs. (9) and (10), we get two characteristic equations

$$S_{xy} S_{yy}^{-1} S_{yx} w^h = a \lambda_1^2 S_{xx} w^h \tag{13}$$

$$S_{yx} S_{xx}^{-1} S_{xy} w^l = \lambda_2^2 S_{yy} w^l \tag{14}$$

We set $M_{xy} = S_{xx}^{-1} S_{xy} S_{yy}^{-1} S_{yx}$, $M_{yx} = S_{yy}^{-1} S_{yx} S_{xx}^{-1} S_{xy}$; then, the projection eigenvectors w^h and w^l are their eigenvalues. By solving the eigenvectors of M_{xy} and M_{yx} , we obtain a group of optimal projection vector pair $W^h = (w_1^h, w_2^h, \dots, w_d^h)$ and $W^l = (w_1^l, w_2^l, \dots, w_d^l)$ with $d \leq \max(n_h, n_l)$. So we will obtain the best correlation between $\hat{X} = XW^h$ and $\hat{Y} = YW^l$, which is the projected sample set X and Y . The vectors \hat{x} and \hat{y} have the same dimensionality. By this, we can calculate their Euclidean distance directly.

3.3 Descriptions of algorithm flow

As we mentioned above, the low-resolution degradation face recognition over long distance in this paper is using CCA. The framework is shown in Fig. 3.

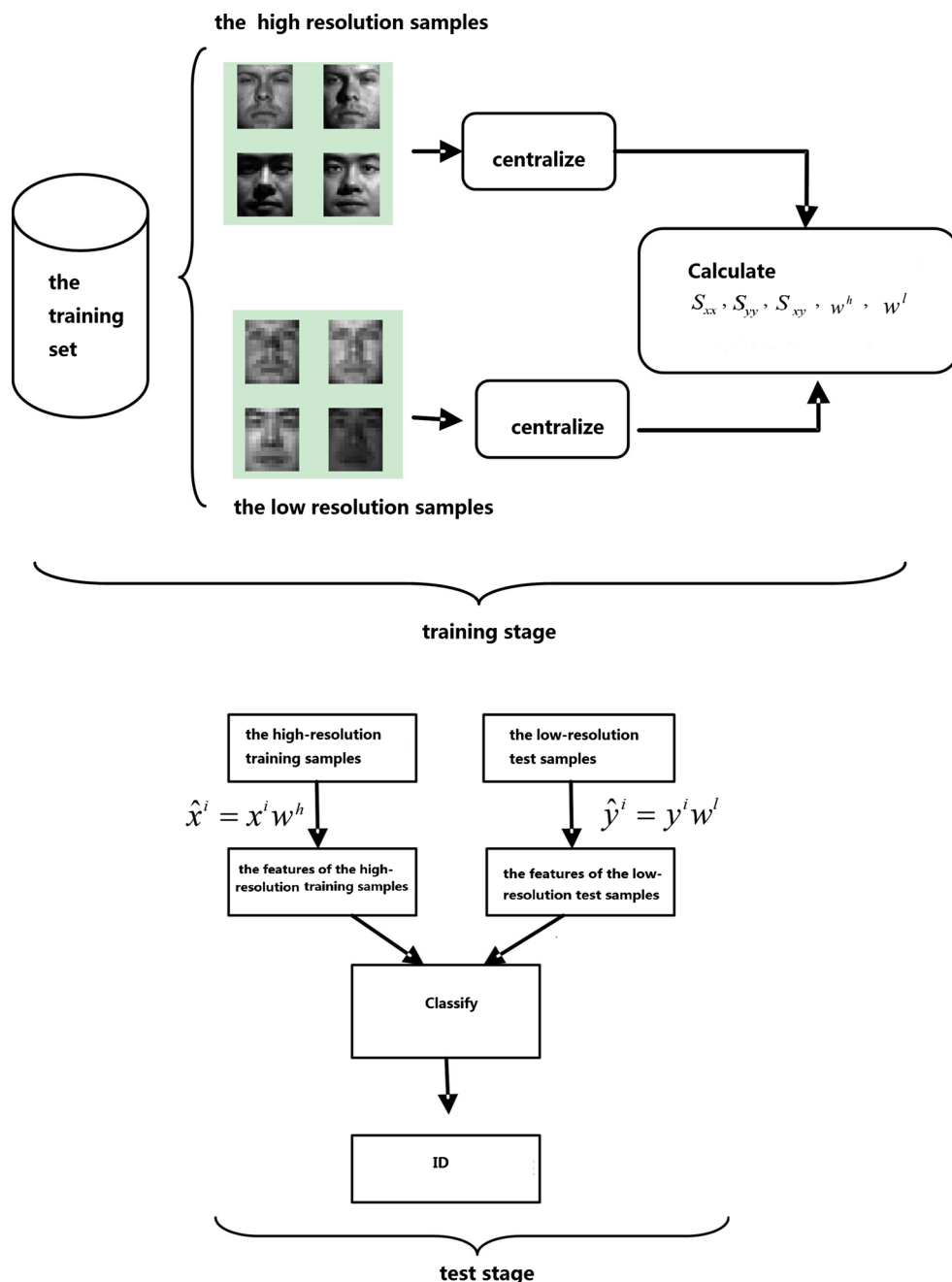
First, we centralize the sample set, and then, we train the high-dimension set X and low-dimension set Y to follow the correlation criterion of Eq. (4). The projections \hat{x}^i and \hat{y}^i which are the projections of x^i and y^i on the projection vector w^h and w^l will be in the same space R . And their correlations reach the peak. At last, in the space R , we use

the nearest neighbor classifier base on Euclidean distance to compare their minimal Euclidean distances between the projected low-resolution test samples and the high-resolution training samples and to classify the face.

4 Experiments

In this section, we conduct the low-resolution face recognition experiments on the extended Yale B, ORL and AR face databases. We compare the proposed method with

Fig. 3 Framework of the proposed method



PCA, LDA and multidimensional scaling (MDS) and LPP. Here, we exploit the nearest neighbor classifier with Euclidean distance. For PCA, LDA and LPP, we first normalize the high-resolution image to a low-resolution one and then extract the features.

4.1 Data sets

4.1.1 Yale B face database

The Extended Yale B face database includes 38 people, and each one has about 64 different images. The image resolution is 192×168 . In this paper, we choose 36 images of each person to do experiments, in which 15 of them are used as high-resolution images and the other 21 images are used as low-resolution images; especially, the quality of high-resolution image is three times over that of low-resolution image. In the training, we make two experimental sample sets with different resolution: One uses the 48×42 resolution images as the high-resolution images and the 16×14 resolution images as the low-resolution images, as is shown in Fig. 4; the other uses the

24×21 resolution images as the high-resolution images and the 8×7 resolution images as the low-resolution ones, as shown in Fig. 5.

4.1.2 ORL face database

ORL face database has 40 persons, and each one has 40 face images. The resolution of image is 112×92 . Ten images have different angels on one person. In sample set 3, we condense the first four images to the 56×46 high-resolution images and the later six images to 28×23 low-resolution images, as is shown in Fig. 6. In sample set 4, we condense the first four images to 28×23 as high-resolution images and the later six images to 14×12 as low-resolution images, as shown in Fig. 7.

4.1.3 AR face database

AR face database [22, 23] consists of 120 persons, and each one has 26 images in different sunlight, expression, shield and aging. In this paper, we select 50 persons and each one has seven images in different expression and



Fig. 4 Images in sample set 1



Fig. 5 Images in sample set 2



Fig. 6 Images in sample set 3



Fig. 7 Images in sample set 4

sunlight. In sample set 5, we condense the first three images of each one to 80×60 as high-resolution images and the later four images to 40×30 as low-resolution images, as shown in Fig. 8. In sample set 6, the first three



Fig. 8 Images in sample set 5



Fig. 9 Images in sample set 6

images of each one is resized to 40×30 as the high-resolution images and the later four images to 20×15 as low-resolution images, as shown in Fig. 9.

4.2 Experimental results

In the first test, we choose 15 low-resolution samples of each person and 15 high-resolution samples of each person in the sample set 1 as the training sample set, and the remaining six low-resolution images of each person as the test sample set. Similarly, we pick 15 low-resolution samples of each person and 15 high-resolution samples of each person in the sample set 2 as training set, and the remaining six low-resolution images of each person as test set. For PCA and LDA, we use the mentioned 15 low-resolution images to train and the six ones to test. The parentheses in the PCA test indicate the reserved dimension; LDA test uses the PCA to pretreat and retain 98 % energy. Gaussian heat kernel with a variance of $1e2$ and ten-nearest neighborhood graph are used in LPP test on set 1, and Gaussian heat kernel with a variance of $1e4$ and 25-nearest neighborhood graph are used in LPP test on set 2. In MDS, the iteration number is 60, and the results are shown in Table 1.

Similarly, the second test picks ten high-resolution images of each person and ten low-resolution images of each person as the training sample; the 11 remaining low-resolution samples of each person are regard as test sample. Other parameters are the same as that in the test 1, and the results are as shown in Table 2.

From Tables 1 and 2, the recognition rates are higher with a high resolution. For example, the recognition rate is higher in sample set 1 than the sample set 2, which indicates that the image resolution has influence on the recognition rates. Obviously, PCA is not satisfied in low-resolution field. As shown in the table, the proposed method has a better performance than the other three, whether in the first or second groups. And both the two

results imply that the number of training set will influence the recognition rate. So we employ the third test, we pick the MDS as a contrast for its stable rate variation. Kernel projection simply uses the unit projection; we make 60 times iteration with the kernel parameter 0.5. By steadily raising the number of sample (training sample from 2 to 15), we get the histograms between the recognition rate and the sample number of the two algorithms. Results of the two samples are shown in Figs. 6 and 7, respectively.

From the third test group, the recognition effect of CCA algorithm is closely related to the sample number, according to Figs. 10 and 11. When the sample number is <7 , the recognition effect of CCA will decline drastically. With more test samples, the CCA keeps a relatively perfect recognition effect. In the sample 1, when the number of training sample is 13, the recognition rate can reach 95.11 %; in the sample 2, with 8×7 low-resolution images which are hard to be recognized by naked eyes, the CCA can also reach a recognition rate of 70 %. These results demonstrate that CCA algorithm is applicable to the low-resolution images.

Table 1 Top recognition rates of the first test

	PCA	LDA	LPP	MDS	CCA
Sample set 1	0.54 (140)	0.92	0.84	0.77	0.93
Sample set 2	0.32 (50)	0.78	0.65	0.71	0.76

Table 2 Top recognition rates of the second test

	PCA	LDA	LPP	MDS	CCA
Sample set 1	0.54 (135)	0.90	0.77	0.78	0.93
Sample set 2	0.42 (45)	0.67	0.66	0.76	0.70

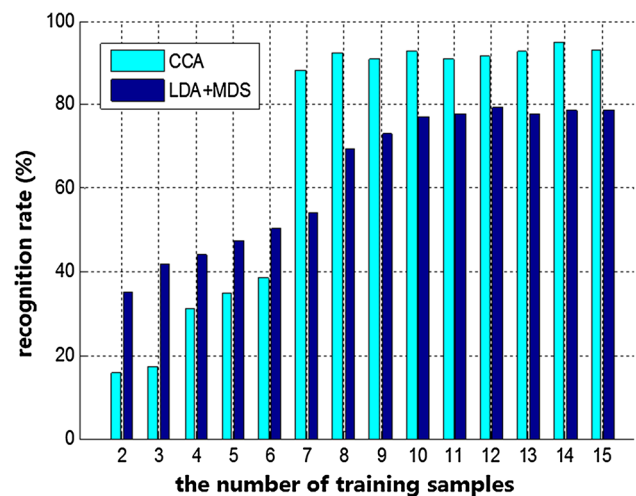


Fig. 10 Recognition rates versus the number of the training set in sample set 1

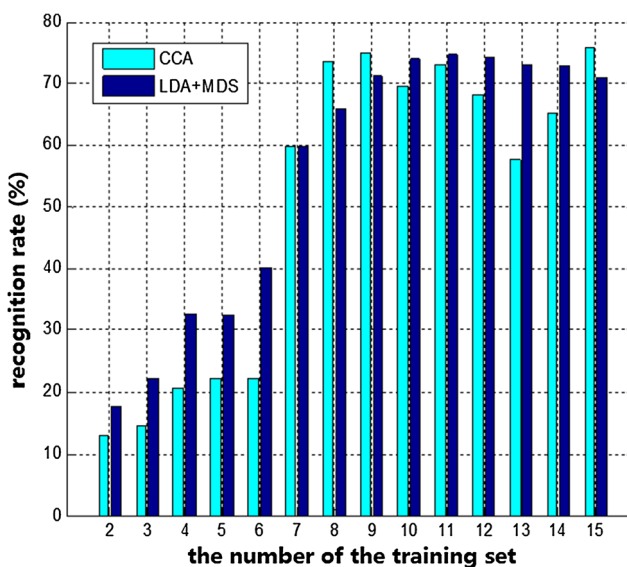


Fig. 11 Recognition rates versus the number of the training set in sample set 2

In the fourth test, as for everyone, we pick four high-resolution samples and four low-resolution samples to compose a training sample set, and the two low-resolution images remained are regard as test sample set. Similarly, we pick four high-resolution images and four low-resolution images in sample 4, and the remaining two face images are regarded as test sample. For the PCA and LDA tests, we use the mentioned four low-resolution images to train and the two remaining low-resolution images to test; the parentheses in the PCA test indicate the reserved dimension; LDA test uses the PCA to pretreat and retain 98 % energy. Consistently, LPP utilizes the PCA to pretreat and retain 98 % energy. Gaussian heat kernel with a variance of 90- and two-nearest neighborhood graph are used in LPP test on set 3, while a Gaussian heat kernel with a variance of 100- and one-nearest neighborhood graph on set 4. As for the MDA test, the number of iterations is 60, and the results are shown in Table 3.

In the fifth test, for everyone, we pick three high-resolution samples and three low-resolution samples to form the training sample set in the sample set 5, and the remaining one low-resolution face image to form the test sample set. We make similar procession on the set 6. For the PCA and LDA tests, we use the three low-resolution images to train and one low-resolution image to test. The parentheses in the PCA test

Table 3 Top recognition rates of the fourth test

	PCA	LDA	LPP	MDS	CCA
Sample set 3	0.91 (80)	0.91	0.94	0.91	0.66
Sample set 4	0.90 (65)	0.95	0.94	0.89	0.65

Table 4 Top recognition rates of the fifth test

	PCA	LDA	LPP	MDS	CCA
Sample set 3	0.3 (100)	1	0.8	1	0.98
Sample set 4	0.2 (45)	1	0.74	1	0.82

indicate the reserved dimension, and both LDA test and LPP use the PCA to pretreat and retain 98 % energy. A Gaussian heat kernel with same variance 1e4 is utilized on both set 5 and set 6. However, different nearest neighborhood graph is adopted in LPP test, 16 for set 5 and 5 for set 6. As for the MDA test, the number of iterations is 60, and the results are shown in Table 4.

5 Conclusions

In this paper, CCA is firstly applied on low-resolution degradation face recognition over long distance. We extract the most correlated component between the low-resolution and high-resolution images using CCA. CCA reduces the requirements of the same dimension of images and also avoids the mismatch between the low-resolution images and high-resolution images. CCA achieves very high recognition rates in our experiments on three face databases. The experimental results show that CCA could be successfully used on low-resolution degradation face recognition over long distance. In the future, how to apply the supervised CCA, locality preserving CCA and sparse CCA on low-resolution degradation face recognition over long distance deserves further studies.

Acknowledgments This work was supported in part by the National Natural Science Foundation of China under Grant Nos. 61201370, 61375001 and supported by the open fund of Key Laboratory of Measurement and Control of Complex Systems of Engineering, Ministry of Education (No. MCCSE2013B01).

References

1. Zhao W, Chellappa R, Rosenfel A, Phillips P (2003) Face recognition: a literature survey. *ACM Comput Surv* 35(4):399–458
2. Belhumeur V, Hespanha J, Kriegman D (1997) Eigen faces vs Fisher faces: recognition using class specific linear projection. *IEEE Trans Pattern Anal Mach Intell* 19(7):711–720
3. Fukunaga K (1990) Introduction to statistical pattern recognition, 2nd edn. Academic Press, Boston
4. He X, Yan S, Hu Y, Niyogi P, Zhang H (2005) Face recognition using Laplacian faces. *IEEE Trans Pattern Anal Mach Intell* 27(3):328–340
5. Yang J, Zhang D, Yang J, Niu B (2007) Globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics. *IEEE Trans Pattern Anal Mach Intell* 29(4):650–664

6. Lai Z, Xu Y, Yang J, Tang J, Zhang D (2013) Sparse tensor discriminant analysis. *IEEE Trans Image Process* 22(10):3904–3915
7. Lai Z, Jin J, Yang J, Sun M (2011) Locality preserving embedding for face and handwriting digital recognition. *Neural Comput Appl* 20(4):565–573
8. Lai Z, Jin Z, Yang J, Sun M (2012) Dynamic transition embedding for image feature extraction and recognition. *Neural Comput Appl* 21(8):1905–1915
9. Zhao G, Miao D, Lai Z, Gao C, Liu C, Yang J (2013) Two-dimensional color uncorrelated discriminant analysis for face recognition. *Neurocomputing* 113(3):251–261
10. Zhao C, Lai Z, Liu C, Gu X, Qian J (2012) Fuzzy local maximal marginal embedding for feature extraction. *Soft Comput* 16(1):77–87
11. Raudys SJ, Jain AK (1991) Small sample size effects in statistical pattern recognition: recommendations for practitioners. *IEEE Trans Patt Anal Mach Intell* 13(3):252–264
12. Hong Z, Yang J (1991) Optimal discriminant plane for a small number of samples and design method of classifier on the plane. *Pattern Recognit* 24(4):317–324
13. Friedman JH (1989) Regularized discriminant analysis. *J Am Stat Assoc* 84:165–175
14. Hastie T, Tibshirani R (1995) Penalized discriminant analysis. *Ann Stat* 23:73–102
15. Hastie T, Tibshirani R, Buja A (1994) Flexible discriminant analysis by optimal scoring. *J Am Stat Assoc* 89:1255–1270
16. Chen LF, Liao HYM, Ko MT, Yu GJ (2000) A new LDA-based face recognition system which can solve the small sample size problem. *Pattern Recognit* 33(1):1713–1726
17. Yu H, Yang J (2001) A direct LDA algorithm for high dimensional data—with application to face recognition. *Pattern Recognit* 34(10):2067–2070
18. Loog M (2001) Approximate pairwise accuracy criterion for multiclass linear dimension reduction: generalization of the Fisher criterion. *IEEE Trans Pattern Anal Mach Intell* 26(7):762–766
19. Tao D, Li X, Wu X, Maybank SJ (2009) Geometric mean for subspace selection. *IEEE Trans Pattern Anal Mach Intell* 31(2):260–274
20. Haroon DR, Szedmak S, Shawe-Taylor J (2004) Canonical correlation analysis: an overview with application to learning methods. *Neural Comput* 16:2639–2664
21. Huang H, He H, Fan X, Zhang J (2010) Super-resolution of human face image using canonical correlation analysis. *Pattern Recognit* 43:2532–2543
22. Martinez AM, Benavente R The AR face database. http://cobweb.ecn.purdue.edu/~aleix/aleix_face_DB.html
23. Martinez AM, Benavente R (1998) The AR face database. CVC Tech Rep 24