A New Supervised Manifold Learning Algorithm

Zhuang Zhao, Jing Han, Yi Zhang, and Lian-fa Bai^(✉)

Jiangsu Key Laboratory of Spectral Imaging and Intelligent Sense,
Nanjing University of Science and Technology, Nanjing, China
mrblf@163.com, 476336296@qq.com

Abstract. In order to overcome the shortcomings of existing manifold learning algorithm, a new supervised manifold algorithm, which improves the original algorithm and makes it more reasonably, has been proposed. Firstly, a more accurate within-class scatter matrix only with the samples belong to the same class is established to characterize the local structure of each manifold. Secondly, nearby manifolds, which can reflect the relationships of different manifolds more accurately, are selected to establish the between-class scatter matrix to characterize the discreteness of different manifolds. Finally, the Fisher criterion is used to solve the objective function and get the optimal projection direction, which can maximize the ratio of the trace of the between-class scatter matrix to the trace of the within-class scatter matrix. Experimental results demonstrate that the proposed algorithm is effective in feature extraction, leading to promising recognition performance in face recognition.

Keywords: Manifold learning · Face recognition · Feature extraction

1 Introduction

With the development of science and technology, more and more data have been obtained, but these data tend to have very high dimensions. How to extract useful data from the original high-dimensional data for the next process is a serious problem. Feature extraction is an important method to solve this problem. Dimensionality reduction has been a key problem in many fields of information processing, such as machine learning, data mining, information retrieval, and pattern recognition [1]. For face recognition, feature extraction is the key step to complete face recognition [2].

During the past decades, many feature extraction algorithms have been proposed, including linear and nonlinear algorithms, supervised and unsupervised algorithms. Among these algorithms, PCA (Principal Component Analysis) [3, 4] and LDA (Linear Discriminant Analysis) [5, 6] are two representative algorithms for linear feature extraction. PCA projects the original samples into a low-dimensional subspace, which is spanned by the eigenvectors corresponding to the largest eigenvalues of the covariance matrix of all the input samples. PCA is the optimal representation of the input samples in the sense of MSE (minimizing mean squared error). Because PCA ignores the class label information, so it is a completely unsupervised algorithm.

Unlike PCA, LDA is a supervised learning algorithm, the essential idea of LDA is to choose a direction, which can make the Fisher criterion function achieve maximum, as the optimal projection direction. This projection direction can maximize the ratio of the trace of the between-class scatter matrix to the trace of the within-class scatter matrix.

However, both PCA and LDA take the global Euclidean structure into account instead of the manifold structure of the original samples. Recent studies have shown that face images possibly reside on nonlinear sub-manifold [7–10], and faces of different people reside on different manifolds. Many manifold learning algorithms have been proposed for discovering the intrinsic low-dimensional sub-manifold embedding in the original high-dimensional manifold, ISOMAP (isometric feature mapping) [7], LLE (local linear embedding) [8], LE (Laplacian Eigenmap) [9], and LPP (Locality Preserving Projections) [10] are the most famous algorithms in these algorithms. Experiments have shown that these algorithms can find meaningful embedding for artificial and real world datasets such as face images. However, how to evaluate the effectiveness of these algorithms remains unclearly.

However, ISOMAP, LLE, LE and LPP are unsupervised manifold learning algorithms, all of these algorithms cannot use the class label information, and thus many supervised manifold learning algorithms are proposed. Li et al. proposed CMVM (constrained maximum variance mapping) [11] algorithm, which uses the within-class scatter matrix to characterize the local structure in each manifold and the between-class scatter matrix to characterize discriminant information between different manifolds, respectively. Under the Fisher criterion, the local structure in each manifold is still kept, and the discriminant information between different manifolds can be explored. However, CMVM ignores the class label information when using within-class scatter matrix to characterize the local structure in each manifold, and ignores the influences of different class distances when using the between-class scatter matrix to characterize the discriminant information between different manifolds. Yan et al. proposed MFA (Marginal Fisher Analysis) [12] algorithm, which designs an intrinsic graph to characterize the compactness of within-class and a penalty graph to characterize the discreteness of between-class. The intrinsic graph describes the within-class sample relationship, and each sample is connected to its k -nearest neighbors belong to the same class. The penalty graph describes the between-class sample relationship and the marginal point pairs of different classes are connected. However, in MFA, it is difficult to select the number of nearest neighbors for each sample and it ignores influences of different class distances too. Wankou Yang et al. proposed MMDA (Multi-Manifold Discriminant Analysis) [13] algorithm. MMDA uses the sum of the within-class weight matrix to weight the mean vectors of classes, and then regards the weighted mean vectors as the new samples, seeking for an optimal projection direction, which can maximize the ratio of the trace of the between-class scatter matrix to the trace of within-class matrix. However, using the weighted mean vectors of classes to characterize the samples exists a certain deviation, especially when the distribution of samples is irregular, this performance will be more evident.

In this paper, a new supervised manifold algorithm based on CMVM is proposed. Firstly, the within-class scatter matrix is used to characterize the local structure in each manifold and the original local structure information is kept. Then, the between-class

scatter matrix is used to characterize discriminant information between different manifolds. Finally, the within-class scatter matrix and the between-class scatter matrix are used to establish the objective function. The proposed algorithm inherits the merit of CMVM and can overcome the shortcoming of CMVM. Compared to CMVM, the proposed algorithm have two advantages over CMVM. Firstly, when establishing the within-class scatter matrix, class label information is taken into account and only the samples belong to the same class are used to establish the within-class scatter matrix. In this way, the interference caused by different classes can be reduced and make it more conform to the local linear hypothesis. Then, when establishing the between-class scatter matrix, samples, which can reflect the distribution of different classes more accurately, are selected to establish the between-class scatter matrix. And the between-class scatter matrix is used to characterize the discreteness of different classes. In this way we can avoid the shortcoming of ignoring the different class distances in the original algorithm.

The organization of this paper is as follows. In Sect. 2, the LDA and LPP are reviewed briefly. In Sect. 3, the motivation is proposed and the new algorithm is described in detail. In Sect. 4, experiments on three face databases are presented to demonstrate the effectiveness of the new algorithm. Conclusions are made in Sect. 5.

2 Review of LDA and CMVM

Suppose samples are $X = [x_1, x_2, \dots, x_N] \in R^m$ and the number of classes is C . The number of samples belong to class c_i is n_i , where $1 \leq i \leq C$. The corresponding samples after feature extraction are $Y = [y_1, y_2, \dots, y_N] \in R^d$, where $d \ll m$. The optimal projection matrix is A , and $Y = A^T * X$. Because the dimension of the original data is very high, the purpose of data dimensionality reduction is to project the original data into a meaningful d -dimensional low-dimensional space.

2.1 Introduction of LDA

LDA is a widely used algorithm for feature extraction and dimensionality reduction. The goal of LDA is seeking for a projection direction to maximize the Fisher criterion (i.e. the ratio of the trace of between-class scatter matrix to the trace of within-class scatter matrix). The between-class scatter matrix S_b , within-class scatter matrix S_w of LDA are shown in Eqs. (1) and (2), respectively.

$$S_b = \frac{1}{N} \sum_{i=1}^C n_i * (m_i - m_0)^T * (m_i - m_0) \quad (1)$$

$$S_w = \frac{1}{N} \sum_{i=1}^C \sum_{x_i \in c_i} (x_i - m_i)^T * (x_i - m_i) \quad (2)$$

where the mean vector of class c_i is denoted by m_i and the mean vector of all samples is denoted by m_0 .

The objective function of LDA is shown in Eq. (3).

$$J(W) = \arg \max_W \frac{W^T S_b W}{W^T S_w W} \quad (3)$$

The objective function can be solved by the generalized eigenvalue problem, as shown in Eq. (4).

$$S_b W = \lambda S_w W \quad (4)$$

Sort the eigenvalues λ in descending $\lambda_1 \geq \lambda_2 \dots \lambda_{m-1} \geq \lambda_m$. The optimal projection matrix W is composed by the eigenvector corresponding to the first d largest eigenvalue λ .

2.2 Introduction of CMVM

CMVM is a supervised manifold learning algorithm. During the calculation, the within-class scatter matrix and between-class scatter matrix are used to characterize the local structure in each manifold and discreteness of different manifolds, respectively. Then the projection direction is sought to maximize the ratio of the trace of the between-class scatter matrix to the trace of the within-class scatter matrix.

The calculation of CMVM is composed by two steps. Firstly, the within-class weight matrix Wc is established to characterize the local structure, this step is similar to LPP. Then the between-class weight matrix Hc is established to characterize the discriminant information based on class label information. The within-class weight matrix Wc and the between-class weight matrix Hc are shown in Eqs. (5) and (6), respectively.

$$Wc_{ij} = \begin{cases} 1 & x_i \text{ is among } k \text{ nearest neighbors of } x_j \\ 0 & \text{otherwise} \end{cases} \quad or \quad (5)$$

$$Wc_{ij} = \begin{cases} \exp\left(-\frac{\|x_i - x_j\|^2}{\beta}\right) & x_i \text{ is among } k \text{ nearest neighbors of } x_j \\ 0 & \text{otherwise} \end{cases}$$

$$Hc_{ij} = \begin{cases} 0 & \text{if sample } x_i \text{ and sample } x_j \text{ have the same class label} \\ 1 & \text{else} \end{cases} \quad (6)$$

where in Eq. (5) k and β are constants.

The objective function of CMVM is to maximize J_{Lc} and minimize J_{Dc} at the same time. The expression of J_{Lc} and J_{Dc} are shown in Eqs. (7) and (8), respectively.

$$J_{Lc} = 2A^T X(Dc - Wc)X^T A, Dc_{ii} = \sum_j^N Wc_{ij} \quad (7)$$

$$J_{Dc} = 2A^T X(Qc - Hc)X^T A, Qc_{ii} = \sum_j^N Hc_{ij} \quad (8)$$

Like LPP, the objective function of CMVM can be solved by the generalized eigenvalue problem, as shown in Eq. (9).

$$X(Qc - Hc)X^T A = \lambda X(Dc - Wc)X^T A \quad (9)$$

Sort the eigenvalues λ in descending $\lambda_1 \geq \lambda_2 \dots \lambda_{m-1} \geq \lambda_m$. The optimal projection matrix A is composed by the eigenvector corresponding to the first d largest eigenvalue λ .

3 Introduction of Proposed Algorithm

3.1 Motivation

From the above introduction in Sect. 2.2, we can get the conclusion that CMVM has two obvious shortcomings. Firstly, when establishing the within-class weight matrix, CMVM ignores the class label information. That is to say the nearby samples of one sample may contain samples belong to other classes, the local linear hypothesis will be destroyed and the classification will be influenced. Secondly, when establishing the between-class weight matrix, CMVM ignores the distances of different classes and regards all the samples belong to other classes as the nearby samples. **In fact, in many manifold learning algorithms, distance especially Euclidean distance is used to measure the** discreteness and different distance has different influence on the final result. Based on these results, the hypothesis that the power of influences of different classes rely on their distances is proposed. If the class far away from one class, its influence will be weaker, else if the class close to one class, its influence will be stronger. Experimental results demonstrate the effectiveness of the hypothesis. However, CMVM does not note this at all.

In order to overcome these shortcomings, a new algorithm based on CMVM is proposed. Firstly, when establishing the within-class matrix, only the samples belong to the same class are selected. In this way, the local linear hypothesis will be preserved. Secondly, when establishing the between-class matrix, KNN is used to select the K1 nearest samples belong to other classes and suppose these samples belong to several classes. The numbers of samples belong to these classes are $n_1, n_2, \dots, n_c$. Sort $n_1, n_2, \dots, n_c$ in ascending and suppose that the new order is $n_1^{\sim}, n_2^{\sim}, \dots, n_c^{\sim}$. The first K_n smallest $n_i^{\sim}$ are selected until the sum of these $n_i^{\sim}$ reach T percent of the total sum. The classes including in these $n_i^{\sim}$ are the nearby classes of the sample and take all the samples belong to these classes as the nearby samples. In this way, the distances of these samples have a rational scale and these samples can reflect the distribution of the nearby manifolds very well. The between-class scatter matrix established by these samples can characterize the discreteness of different classes more accurately and make it more conveniently for classification.

3.2 Implementation Process

The similar method in CMVM is used to establish the within-class weight matrix, but the class label information is taken into account. The within-class weight matrix W is shown in Eq. (10)

$$W_{ij} = \begin{cases} 1; & x_i \text{ is among } k \text{ nearest neighbors of } x_j \text{ and } x_i, x_j \text{ have the same label} \\ 0; & \text{otherwise} \end{cases} \quad (10)$$

The within-class scatter matrix J_1 is shown in Eq. (11).

$$J_1 = \sum_{i,j}^N (Y_i - Y_j)(Y_i - Y_j)^T W_{ij} = 2A^T X(D - W)X^T A, D_{ij} = \sum_j^N W_{ij} \quad (11)$$

For every sample x_i , the algorithm is used to select the nearby samples of x_i and shown in follow. Firstly, the K_1 nearest samples from other class of x_i are selected. Secondly, the number of classes are counted and the numbers of samples belong to these classes are recorded as $n_1, n_2, \dots, n_C$. Sort $n_1, n_2, \dots, n_C$ in ascending and the new order is recorded as $n_1^*, n_2^*, \dots, n_C^*$. Thirdly, the percentage of the sum of first K_n n_i^* to the total sum is calculated until the ratio reach the given threshold T , in this paper, T is set to $T = 60\%$. When the sum reach T , C_{inc} is used to record the first K_n n_i^* and all samples belong to these classes are selected as the nearby samples of sample x_i . Then these samples are used to establish the between-class weight matrix H .

$$\vec{H}_{ij} = \begin{cases} 1 & \text{if the label of } x_j \text{ include in } C_{inc} \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

The between-class scatter matrix J_2 is shown in Eq. (13)

$$J_2 = \sum_{i,j}^N (Y_i - Y_j)(Y_i - Y_j)^T H_{ij} = 2A^T X(Q - H)X^T A, Q_{ii} = \sum_j^N H_{ij} \quad (13)$$

The objective function of the proposed algorithm is shown in Eq. (14).

$$J(A) = \arg \max_A \frac{A^T X(Q - H)X^T A}{A^T X(D - W)X^T A} \quad (14)$$

Like LPP, the objective function of proposed algorithm can be solved by the generalized eigenvalue problem, as shown in Eq. (15).

$$X(Q - H)X^T A = \lambda X(D - W)X^T A \quad (15)$$

Sort the eigenvalues λ in descending $\lambda_1 \geq \lambda_2 \dots \lambda_{m-1} \geq \lambda_m$. The optimal projection matrix A is composed by the eigenvector corresponding to the first d largest eigenvalue λ .

4 Experiments

Due to the inverse of the within-class scatter matrix is needed to calculate, if the number of samples less than the dimensions of sample, the within-class scatter matrix will be singular. In order to avoid the small sample size (SSS) problem, PCA is conducted before feature extraction. Nearly 95 % energy is preserved to select the number of principal components. In the algorithm, when the nearby samples are selected, if the number of train samples n_{train} larger than 10, K is set to $K = 10$, else K is set to $K = n_{train} - 1$ and $K1$ is set to $K1 = 3 * K$, which are the sample selected to establish the between-class weight matrix. In MFA when establishing the between-class weight matrix, if the number of train samples n_{train} larger than 10, K is set to $K = 10$, else K is set to $K = n_{train} - 1$. The AR dataset, Infrared dataset and PIE dataset are used to evaluate the proposed algorithm in comparison with PCA, LDA, LPP, CMVM, MFA and MMDA. All algorithms were run 10 times and the mean training time is used as the training time of each algorithm. The experiments were implemented on a Desktops with i5 3470 CPU and 8 GB RAM under the MATLAB (R2014a) programming environment.

4.1 Experiment on AR Dataset

The AR dataset [14] contains over 4000 images of 126 people with different facial expressions, lighting conditions. A subset of the AR dataset is selected, the subset includes 1400 images of 100 people (each person has 14 images). All the samples are resized to 60×43 . In the experiment, the first L images per person are selected for training and others for testing. Finally, a nearest neighbor classifier with Euclidean distance is used to classify the extraction feature. The recognition rates and training time are shown in Tables 1 and 2, respectively. The number in brackets means the dimensions of samples when the recognition rates achieve highest (Figs. 1 and 2).

Table 1. Recognition rates on the AR dataset by different methods.

	PCA	LDA	LPP	MFA	CMVM	Proposed
$L = 7$	0.6700(61)	0.6543(26)	0.8014(73)	0.8257(72)	0.7800(69)	0.8071(72)
$L = 8$	0.7900(64)	0.8950(18)	0.9200(58)	0.9650(74)	0.9133(71)	0.9783(73)
$L = 9$	0.7760(67)	0.9440(19)	0.9220(60)	0.9620 (74)	0.9040(60)	0.9760(73)
$L = 10$	0.7575(75)	0.9375(73)	0.9375(73)	0.9600(73)	0.9350(71)	0.9825(69)
$L = 11$	0.7467(75)	0.9700(19)	0.9833(73)	0.9900(73)	0.9833(59)	1.00(71)

Table 2. Average training time on the AR dataset (s)

	PCA	LDA	LPP	MFA	CMVM	Proposed
$L = 7$	2.12	1.51	0.042	1.00	0.066	0.16
$L = 8$	2.57	1.87	0.053	1.21	0.084	0.19
$L = 9$	3.24	2.34	0.068	1.46	0.10	0.22
$L = 10$	3.82	2.81	0.083	1.72	0.13	0.26
$L = 11$	4.74	3.48	0.10	2.01	0.15	0.29

Fig. 1. Images of two people in AR dataset

Fig. 2. Recognition rates of all the algorithms on AE dataset

4.2 Experiment on Infrared Dataset

The Infrared dataset contains 1400 images of 5 people and 2 cups with different rotational angles. All of the images were captured by infrared camera. Each people or cup has 200 images and all the samples are resized to 40×40 . In the experiment, the first L images per people or cup are selected for training and others for testing. Finally, a nearest neighbor classifier with Euclidean distance is used to classify the extraction feature. The recognition rates and training time are shown in Tables 3 and 4, respectively. The number in brackets means the dimensions of samples when the recognition rates achieve highest (Figs. 3 and 4).

Table 3. Recognition rates on the infrared dataset by different methods.

	PCA	LDA	LPP	MFA	CMVM	Proposed
$L = 80$	0.6452(8)	0.6548(4)	0.7464(20)	0.7036(14)	0.6667(20)	0.7500(3)
$L = 100$	0.6071(3)	0.6143(3)	0.6414(9)	0.6743(13)	0.6643(20)	0.7500(4)
$L = 120$	0.6268(14)	0.6250(6)	0.6804(18)	0.7357(20)	0.7000(13)	0.7446(20)
$L = 140$	0.6190(12)	0.6452(1)	0.6976(2)	0.6929(19)	0.6810(19)	0.8214(5)
$L = 150$	0.5657(11)	0.6400(1)	0.7114(18)	0.7914(13)	0.7514(7)	0.8371(3)

Table 4. Average training time on the infrared dataset (s)

	PCA	LDA	LPP	MFA	CMVM	Proposed
$L = 80$	0.32	0.29	0.023	0.079	0.036	0.11
$L = 100$	0.53	0.47	0.036	0.12	0.056	0.15
$L = 120$	0.81	0.78	0.053	0.17	0.081	0.19
$L = 140$	1.17	1.07	0.08	0.20	0.14	0.23
$L = 150$	1.43	1.30	0.08	0.23	0.13	0.28

Fig. 3. Sample images of two people and one cup in infrared dataset

Fig. 4. Recognition rates of all the algorithms on infrared dataset

4.3 Experiment on PIE Dataset

The PIE dataset [15, 16] contains 41368 images of 68 people with different facial expressions, lighting conditions. These images were captured by 13 synchronized cameras and 21 flashes under varying pose, illumination and expression. 80 images for each people are selected and all of the samples are resized to 32×32 . In the experiment, the first L images per people are selected for training and others for testing. Finally, a nearest neighbor classifier with Euclidean distance is used to classify the extraction feature. The recognition rates and training time are shown in Tables 5 and 6, respectively. The number in brackets means the dimensions of samples when the recognition rates achieve highest (Figs. 5 and 6).

From the above results, we can see that the proposed algorithm has obviously higher recognition rates than other algorithms introduced in the paper. Compared to CMVM, the proposed algorithm uses a little more time.

Table 5. Recognition rates on the PIE dataset by different methods

	PCA	LDA	LPP	MFA	CMVM	Proposed
$L = 30$	0.3876(64)	0.4365(21)	0.5132(59)	0.4997(35)	0.4991(50)	0.6068(59)
$L = 40$	0.4978(60)	0.4897(19)	0.5746(57)	0.5879(59)	0.5566(50)	0.7022(59)
$L = 50$	0.5853(61)	0.5441(23)	0.5907(57)	0.6696(59)	0.6211(58)	0.7216(59)
$L = 60$	0.5772(61)	0.6140(18)	0.6890(59)	0.7404(59)	0.6816(54)	0.8162(59)
$L = 70$	0.8191(61)	0.8559(20)	0.9147(53)	0.9235(59)	0.9250(47)	0.9632(59)

Table 6. Average training time on the PIE dataset (s)

	PCA	LDA	LPP	MFA	CMVM	Proposed
$L = 30$	5.17	4.43	0.33	4.17	0.52	0.75
$L = 40$	9.77	9.02	0.58	6.88	0.90	1.19
$L = 50$	16.59	15.36	0.91	10.20	1.42	1.74
$L = 60$	27.70	25.21	1.33	14.29	2.12	2.45
$L = 70$	39.68	38.09	1.81	18.96	2.86	3.19

Fig. 5. Sample images of two people in PIE dataset

Fig. 6. Recognition rates of all the algorithms on PIE dataset

5 Conclusions

In pattern recognition, data dimensionality reduction, which enhancing the dissimilarity between different classes, is conducive for classifying. In order to overcome the shortcomings of CMVM, a new algorithm based on CMVM is proposed in this paper. In the proposed algorithm, samples, which can describe the relationship of different classes more accurately, are selected and are used to establish a more suitable between-class scatter matrix to describe the discreteness of different classes. In this way, the effectiveness of classification can be improved. Experiments on three datasets have shown that the proposed algorithm has obviously higher recognition rates than other algorithms introduced in the paper.

In the proposed algorithm, some parameters is needed to set and these parameters will influence the final result. However, how to select the parameters is a problem for the algorithm. So a non-parameter algorithm is needed to research in the future. In order to avoid the small sample size (SSS) problem, PCA is conducted. However, after PCA some useful information will be lost, so a null space based algorithm is needed to research to solve this problem in the future.

Acknowledgment. This work was financially supported by the National Natural Science Foundation of China (61231014).

References

1. Lu G.F., Zhong, L., Zhong, J.: Face recognition using discriminant locality preserving projections based on maximum margin criterion. *J. Pattern Recogn.* **43**(10), 3572–3579 (2010)
2. Jin, Z., Yang, J., Hu, Z., et al.: Face recognition based on the uncorrelated discriminant transformation. *J. Pattern Recogn.* **34**(7), 1405–1416 (2001)

3. Wold, S., Esbensen, K., Geladi, P.: Principal component analysis. *J. Chemometr. Intell. Lab. Syst.* **2**(1), 37–52 (1987)
4. Martinez, A.M., Kak, A.C.: PCA versus LDA. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(2), 228–233 (2001)
5. Fukunaga, K.: *Introduction to Statistical Pattern Recognition*, Iteration Number Initial, 2nd edn. Academic Press Inc., San Diego (1990)
6. 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)
7. Tenenbaum, J.B., De Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. *J. Sci.* **290**(5500), 2319–2323 (2000)
8. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. *J. Sci.* **290**(5500), 2323–2326 (2000)
9. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. *J. Neural Comput.* **15**(6), 1373–1396 (2003)
10. Niyogi, X.: Locality preserving projections. *Neural Inf. Process. Syst. MIT* **16**, 153–161 (2004)
11. Li, B., Huang, D.S., Wang, C., et al.: Feature extraction using constrained maximum variance mapping. *J. Pattern Recogn.* **41**(11), 3287–3294 (2008)
12. Yan, S., Xu, D., Zhang, B., et al.: Graph embedding and extensions: a general framework for dimensionality reduction. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(1), 40–51 (2007)
13. Yang, W., Sun, C., Zhang, L.: A multi-manifold discriminant analysis method for image feature extraction. *J. Pattern Recogn.* **44**(8), 1649–1657 (2011)
14. Martinez, A.M.: The AR face database. CVC Technical report, 24 (1998)
15. Sim, T., Baker, S., Bsat, M.: The CMU pose, illumination, and expression database. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(12), 1615–1618 (2003)
16. Sim, T., Baker, S., BSat, M.: The CMU pose, illumination, and expression (PIE) database of human faces. Technical report CMU-RI-TR-01-02, Carnegie Mellon University (2001)