International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.9, No.5 (2016), pp.309-320
http://dx.doi.org/10.14257/ijsip.2016.9.5.28

An Embedded Feature Extraction Algorithm with MMC for Face Recognition

Tian Liang

Shandong agriculture and engineering university, Dormitory building 1-2-702, Zhudian, Licheng district, Jinan city, Shandong Province
Zip code 250100
liangtian_ngy@sina.com

Abstract

Feature extraction is the most crucial part of face recognition system, which has important role in the field of pattern recognition. There are a lot of classic algorithms about feature extraction at present, such as the method based on linear and nonlinear. An effective classification and dimension reduction method, local embedded graph attribute selection algorithm with maximal region is to generate the inherent graph and penalty graph, to overcome the low efficiency problem of local linear embedding (LLE) method and maximum margin criterion (MMC) method. With inherent picture, the structured nonlinear can be found on the high-dimensioned space by using local geometry of the restructured linear, which leads to the same instances gathering together as more as possible. Meanwhile, different class instances are as far as possible from each other in penalty picture. Since LLE is an unsupervised method, not enhance visual clustering classification ability, so compact figure within the class to consider the sample class information, can sample the same category as compact. In this method, the smallest size instance issue was tackled by the employment of MMC and the neighborhood relationship can be better described by an adequate improvement of the adjacency matrix. The effectiveness of algorithm proposed by this paper is present by a large amount of experiences in the standard face databases of Yale, AR and ORL facial data.

***Keywords**: face recognition; feature extraction; extraction algorithm; maximum margin criterion; local graph embedding*

1. Introduction

Manifold learning algorithms can discover intrinsic low-dimensional sub-manifold embedded in the high-dimensional image space. The algorithms are unsupervised learning methods; the discriminated ability of the low-dimensional feature obtained by the methods is often lower than that of the conventional dimensionality reduction methods. Furthermore, manifold learning methods don’t have direct mapping for new example.

Face recognition [1-4] has become a main study object of pattern recognition, machine vision and computer vision. Currently, there are many linear and non-linear data dimensionality reduction method is applied to face recognition. Principal component analysis (PCA) [5] and linear discriminating analysis (LDA) [6] two classical methods for data dimensionality reduction, but they are both linear dimensionality reduction method, and did not perform well when it comes to some sort of "non-linear" data.

In recent years, among nonlinear data dimensionality reduction, the manifold learning-oriented dimensionality reduction of theoretical research and English have made great progress. The most representative methods are equidistant map [7], locally linear embedding [8-9], Laplacian Operators feature mapping [10-11], *etc* These methods can maintain the original data in the topology of the same premise; the high-dimensional data

ISSN: 2005-4254 IJSIP

is mapped to the corresponding low-dimensional space. In which, LLE is an unsupervised learning algorithm, in order to maintain the relationship between the local neighborhood approach and high-dimensional data is mapped to a low global coordinate system. LLE algorithm can see the structure of the original data and to identify the intrinsic internal structure of data points, but it is not very good for data classification. Currently based on the supervision of LLE algorithm, it is divided into two situations: First, calculating for each sample point near the Provisional point of K of the treatment LLE algorithm increased sample point class information [12-16]; Second, LLE algorithm combines with LDA algorithm, first with LLE for data dimensionality reduction, and reuse LDA algorithm for classification [16-17].

Although these nonlinear dimensionality reduction technique in theory can be found in low-dimensional embedding complex, and in the artificial test data produce good results, but in many practical applications face two problems; 1) calculating a large load; 2) can only produce defined set of points in the training data dimensionality reduction on the map, He has proposed algorithms [18] and neighborhood preserving projections remain embedding algorithm [19], and they successfully applied to face recognition; but these two algorithms and LDA as also facing a "small sample" problem [20]. Li [21] proposed another maximum margin criterion function, but for non-linear data may be not valid.

In order to solve the problems faced by the above methods, we present the main contributions as follows in this paper:

The first, a (locally graph embedded feature extraction methods with maximum margin criterion) LGE-MMC algorithm was proposed based on an effective data dimensionality reduction and classification methods—locally graph embedding feature extraction methods with maximum margin criterion, and its application to face recognition;

Second, in the MMC, the class diagrams within the compact use of existing local symmetry of the reconstructed high-dimensional data space to identify the non-linear structure, the same sample as much as possible together.

Third, Since LLE is an unsupervised method, not enhance visual clustering classification ability, so compact figure within the class to consider the sample class information, can sample the same category as compact;

Fourth, the class diagram manipulation between different categories of punishment as far as possible away from the sample, in which two optimal adjacency graph, the algorithm makes the data dimensionality reduction and as close as possible within the class between classes as far away from the ORL, Yale and AR face database standard experimental results verify its effectiveness.

The surplus of the paper is concluded in the following part. And section II represents the review on the manifold learning algorithms. Section III presents our proposed algorithm. Section IV gives the results of experiments. At length, the concluded remarks are provided in the following Section V.

2. Proposed Scheme

LLE algorithm is a nonlinear dimensionality reduction way, the basic idea is to keep the original manifold local neighborhood relationships among the high-dimensional data. And it is mapped to low-dimensional global coordinate system. MMC algorithm corresponding maximum between-class scatter matrix represents a different kind of separation between the classes, and the within-class scatter matrix represents the minimum between the same mode of sample as compact as possible, but it is not effective to keep the sample manifold local inherent structure. This article will MMC method to promote, effectively combining the inherent partial sample graph embedding structure.

That is, if x_i and x_j are adjacent to the before conversion, the conversion of y_i and y_j are also adjacent; otherwise, y_i and y_j after conversion is non-adjacent.

To deal with the efficiency problem of local linear embedding (LLE) method and maximum margin criterion (MMC) method in attribute selection, an effective classification and dimension reduction method, local embedded graph attribute selection algorithm based on maximal region is to generate the inherent graph and penalty graph, with the nearest neighbor premise preservation. With inherent picture, the structured nonlinear can be found on the high-dimensioned space by using local geometry of the restructured linear, which leads to the same instances gathering together as more as possible. Meanwhile, different class instances are as far as possible from each other in penalty picture. In this method, the smallest size instance issue was tackled by the employment of MMC and the neighborhood relationship can be better described by an adequate improvement of the adjacency matrix.

2.1. LLE and MMC

Located in a high-dimensional Euclidean space sample set $X = \{ x_1, x_2, \cdots, x_n \}$, $x_i \in R^D$, to seek a projection matrix A, hoping these samples mapped to a relatively low-dimensional feature space R^d, $d \le 0$. Thus, the sample in the new feature space representation is $Y = \{ y_1, y_2, \cdots, y_n \}$, $y_t = A^T x_i$, let the matrix $A = \{ a_1, a_2, \cdots, a_d \}$ is the best *identification of vectors* a *, constituted projection matrix.*

LLE algorithm is usually divided into three steps to achieve:

Step1. Calculated for each sample data set near the point of the k in x_i, $\{ x_{i1}, x_{i2}, \cdots, x_{ik} \}$, where k is a pre-given value.

Step2. Calculated partial reconstruction of the sample point weight matrix, an error function is defined as follows:

$$\varepsilon(W) = \sum_i |x_i - \sum_j W_{ij} x_j|^2 \tag{1}$$

By minimizing the formula (1), calculate the reconstruction of each sample point A, the weight $\sum_{j=1}^{k} W_{ij} = 1$, when $x_j \notin \{X_{i1}, \ldots, X_{in}\}$, Then $W_{ij} = 0$.

Step3. Map of all the sample points to a low dimensional space. To minimize map Conditions:

$$\varepsilon(Y) = \sum_{i=1}^{n} |y_i - \sum_{j=1}^{k} W_{ij} y_j|^2 \tag{2}$$

According to the weight W_{ik}, minimizing formula (2) the objective function, the d-dimensional projection obtained vector y_i for x_i, in order to ensure a unique solution, and must meet the $\sum_{i=1}^{n} y_i = 0$ and $\frac{1}{n} YY^T = I$. The Rayleitz-Ritz theorem to solve the minimum $M = (I - W)^T (I - W)$ $d+1$ characteristic values according to the corresponding feature vector in ascending order, drop the first eigen value corresponding eigenvectors, the rest of the d matrix composed of characteristic vectors is obtained in this low-dimensional embedding samples.

The same criteria based on Fisher linear discriminating feature extraction same purpose, MMC algorithm also aims to data from the original high-dimensional space compression to low-dimensional space, and in low-dimensional space to maintain a high separability.

In the MMC algorithm, S_b, S_W and S_t, respectively, the training sample class scatter matrix, within-class scatter matrix and total scatter matrix. Defined by its knowledge, S_b, S_W and S_t are non-negative definite matrix, and satisfies $S_t = S_b + S_W$, where

$$S_b = \sum_{i=1}^{C} l_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T$$

$$S_w = \sum_{i=1}^{c} \sum_{j=1}^{c_i} (x_i^{\,j} - \bar{x})(x_i^{\,j} - \bar{x})^T$$

Where $l_i = N_i / N$ is the prior probability of class i, x represents the mean of all samples, x denotes the mean of class i training samples, N_i is the i-th class training sample number, N is the number of all samples.

MMC function is defined as $\max J(A) = tr(S_b^{'} - S_w^{'})$. Since $S_b^{'}$ and $S_W^{'}$ denote the projection transformed as $y_t = A^T x_i$ samples in feature space between-class scatter matrix and the within-class scatter matrix, where $S^{'}_b = A^T S_b A, S^{'}_w = A^T S_w A$, so, we have

$$J(A) = tr(S^{'}_b - S^{'}_w) = tr(A^T (S_b - S_w) A)$$
$$= \sum_{i=1}^{d} a_i^T (S_b - S_w) a_i$$

When the condition is $a_i^{\,T} a_i = 1$, we can get the conclusion that

$$\begin{cases} \max \sum_{i=1}^{d} a_i^T (S_b - S_w) a_i \\ subjecttoa_i^T a_i \end{cases} \tag{3}$$

The formula (3) can draw a conclusion via Lagrange coefficients.

$$(S_b - S_w) a_i = \lambda_i a_i \tag{4}$$

Accordingly, by formula (4) can easily be solved before the d eigenvalues corresponding to eigenvectors. Where λ_i is the characteristic value of S_b-S_W, and a_i is the corresponding eigenvectors. " D " by the former largest eigenvalue eigenvector corresponding projection matrix consisting final paper will be denoted by the projection matrix $A_{MMC} = \{a_1, \ldots, a_d\}$.

2.2. MMC-Based Partial Graph Embedding

MMC algorithm corresponding maximum between-class scatter matrix represents a different kind of separation between the classes, and the within-class scatter matrix represents the minimum between the same mode of sample as compact as possible, but it is not effective to keep the sample manifold local inherent structure. This article will MMC method to promote, effectively combining the inherent partial sample graph embedding structure. That is, if x_i and x_j are adjacent to the before conversion, the conversion of y_i and y_j are also adjacent; otherwise, y_i and y_j after conversion is non-adjacent.

In this paper, LGE-MMC algorithm is aimed at maintaining local close relations under the premise that similar samples together as much as possible, and as far away from different types of samples. $X = \{x_1, \ldots, x_n\}, x_i \in R^D$ is referred to sample data, $Y = \{y_1, \ldots, y_n\}, y_i \in R^d$ is the low-dimensional data *G*={*X*, *M*} represents a weighted undirected graph, where *X* is the vertex set, $W \in R^{n \times n}$ is the similarity matrix. Diagonal matrix *D* and *G* is the Laplacian matrix L is defined as $D_{ii} = \sum_{j \neq z} W_{ij}, L = D - W, \nu i.$

LGE-MMC algorithm can be divided into three steps:

2.2.1. Within-Class Scatters Matrix Characterization: Compact within the class diagram, a data point within the class and its adjacent similar number k nearest neighbors in its implicit data is locally linear fashion, and each data point can be weighted by a neighbor data reconstruction. LGE-MMC local representation within the class similar to the LLE algorithm; the LLE algorithm only difference is that this structure is the weight matrix W, both calculated for each sample data set near the point of the k , x , have to consider the type of sample information, and LLE algorithm does not consider the data sample class information. Therefore, within-class scatters matrix characterizations with LLE algorithm to construct a similar total of three steps. Wherein the final map to minimize the objective function:

$$S_c = \sum_{i=1}^{n} |\mathrm{y}_\mathrm{i} - \sum_{j=1}^{k} W_{ij}{}^c y_j|^2 \tag{5}$$

This introduction of a linear transfer function, $y_i = A^T x_i$, then the formula (5) can be changed to $J_i(A) = \min\ tr\{YMY^T\} = \min\ tr\{A^T XMX^T A\}$; where $W_{ij}^c = \{ {0 \atop 1}$.

2.2.1. Between-Class Scatters Matrix Characterizations: Between classes punishment diagram, if the output of the two data y_i and y_j belong to different categories, to find an optimal mapping approach is to make a reasonable loss function below the maximum value of S_p. And in order to simplify the equation, the introduction of a linear transfer function $y_i = A^T x$, then there

$$\begin{aligned} S_P &= \frac{1}{2}\sum_i \sum_i \left\| y_i - y_j \right\|^2_i W_{ij}^p = \\ &\frac{1}{2}\sum_i \sum_i \left\| A^T x_i - A^T x_j \right\|^2_i W_{ij}^p = \\ &A^T X (D^P - W^P) X^T A = \\ &A^T X L^P X^T A \end{aligned} \tag{6}$$

The nature of the data is output Varimax different classes. PCA can also maintain the largest number of output variance, but it is a global method, the output is within-class and inter-class output data. The proposed inter-class method is a local approach, considering only inter-class data, ignoring the interference within the class data, more accurately classifies nonlinear high dimensional data. So this similarity matrix is constructed as $W_{ij}^p = \{ {0 \atop 1}$.

3. Experiement Result ad Analysis

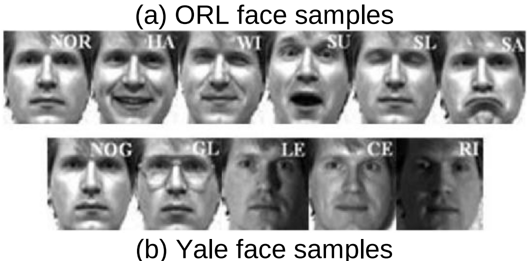
In order to verify the proposed LGE-MMC effectiveness in face recognition algorithm, we ORL, Yale and AR face image database for a full experiment and compared the LGE-MMC algorithms and LDA [6], MMC [21], LLE [8-9], DLA [22-23] and LLE+LDA [16-17] algorithm classification performance, all the algorithms are used Euclidean distance and the nearest neighbor classifier. In the experiment, it is in order to quickly get results for each algorithm with PCA pretreatment. At this time, the paper is for about 95% of the image energy. Experimental environment is as follows: Dell PC, CPU as Inter Athlon (tm) 64 Processor, 1024 MB RAM, Matlab 7.01.

Face image database and experimental design. ORL face image database from 40 people, human face scales also have up to 10% of the variation. Experiment, the image is processed into the form of dimensions 56×46, randomly selected from this experiment before l (l = 2,3,4,5,6) training images, and the remaining 10-l images used for testing. L For each selected image, are 50 times experiments, the final result is the average results of 50 times.

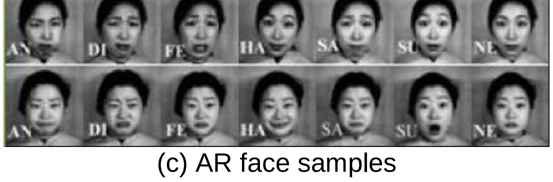
Yale face database includes 15 individual 165 grayscale face images of each person constituted by the 11 photos, these photos in different lighting conditions such as facial expressions and shoot. Test, the image is processed into a 50×40 Vader form, while randomly selected before l (l = 2, 3, 4, 5, 6) training images, and the remaining 11-l images used for testing. L For each selected image is 50 times experiments; the final result is the average result of 50.

AR face database contains 126 individuals (70 males, 56 females) of 4000 pieces of color face images, these images by different light, different expressions and different occlusion situations frontal face image, the majority of people Zhou image is separated by two hours of shooting two image sets, each set contains 13 like color images and 120 images of individuals who did not wear a scarf face image, each 20, and scale from one to 50×40 Vader grayscale images. Randomized trials before l (l = 2, 3, 4, 5, 6) image training, and the remaining 20-l image used for testing. L For each selected image, but also for 10 experiments, the final result is the average result of 10, Fig 1(a)-1(c) are three individuals face database sample image after preprocessing.

(a) ORL face samples



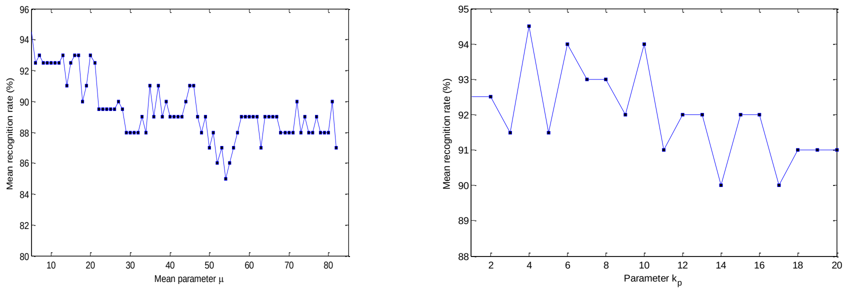
(b) Yale face samples



(c) AR face samples

Figure 1. Three Individuals Face Database Sample Images

Parameter Selection. In classical feature extraction algorithm, the choice of various parameters has been an open question. Similarly, LGE-MMC balancing algorithm parameters u and number of neighbors within the class k_r and class k_p on arguments between neighbors recognition accuracy also has a significant impact. This set the number of neighbors within the class k_r =l-1, while taking advantage of cross-validation approach balancing parameters of the algorithm and the class u choose between neighboring parameter k_p. In this paper, Yale face image database is randomly selected l (l=2) images as training samples, all other images as test samples for the experiment was repeated 50 times. Figures 2(a), 2(b) gives the parameters u and class balance between different values of parameter k_p neighbors the average recognition rate of the case, it can be seen, balanced parameters u being between 0.05 and classes the neighbors parameter k_p is 4-10, the average recognition rate is relatively stable, and the parameter u = 0.3 and the balance between-class neighbor parameters k_p = 4, the average recognition rate of a maximum of 94.79. Because different face database resulting balancing parameters u = 0.3 and between-class neighborhood parameter k_p = 4 in the other face database can be achieved when doing experiments maximum average recognition rate. Thus, in subsequent experiments, this paper balancing parameters u take 0.05-1 interval 0.05; class taking k_p between 4 and 10 neighboring parameter, the interval a cross match.



(a) (b)

Figure 2. (a) Parameters u and Class Balance Between Different Values. (b) Parameter k_p Neighbors the Average Recognition Rate

Average recognition rate of the characteristic dimensions of the relationship. The purpose of this study is to investigate the LGE-MMC arithmetic average recognition rate of change in the relationship with the characteristic dimension. In this paper, ORL, Yale and AR face image database experiments were carried out, and were randomly chosen before each image library 4,6,5 images as training samples. The library image is as all the other test samples; corresponding to the remaining images of the library as a test sample. Which, ORL and Yale face database images 50 times repeated experiments, AR face database with 10 repeated experimental, results are shown in Figure 3. This can be from Figure 3 the following conclusions: Average recognition rate of the characteristic dimensions of the relationship. The purpose of this study is to investigate the LGE-MMC arithmetic average recognition rate of change in the relationship with the characteristic dimension. In this paper, ORL, Yale and AR face image database experiments were carried out, and were randomly chosen before each image library 4, 6, 5 images as training samples. The library image is as all the other test samples; corresponding to the remaining images of the library as a test sample. Which, ORL and Yale face database images 50 times repeated experiments, AR face database with 10 repeated experimental, results are shown in Figure 3, 4 and 5. This can be concluded from Figure 3, 4 and 5:

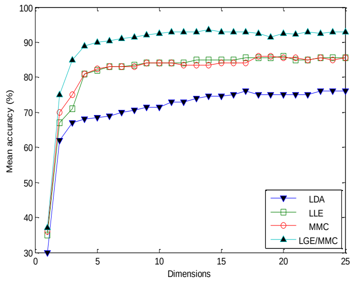


Figure 3. ORL Face Images Recognition Rate

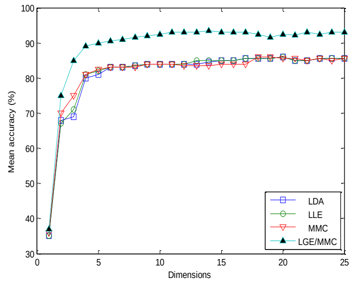


Figure 4. YALE Face Images Recognition Rate

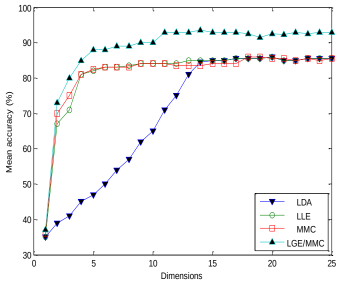


Figure 5. AR Face Images Recognition Rate

1) LGE / MMC algorithm with the characteristic dimension to increase the average recognition rate has been increased, and the relatively high number of dimensions, the average recognition rate of the algorithm is better than the other five kinds of classical algorithm, the average recognition rate.

2) LDA and LLE + LDA algorithm for feature dimension is relatively low, with an average recognition rate at LGE / MMC algorithm, the average recognition rate. This is better than the LDA algorithm to identify the optimal number of dimensions of not more than c-1, the best recognition performance. Average recognition rate of LDA algorithm process and MMC algorithm LDA algorithm and matrix singularity flip.

3) LLE algorithm, the average recognition rate is relatively low, indicating the LLE algorithm is not used for data classification. LLE + LDA algorithm average recognition rate is higher than average recognition rate of LLE algorithm LDA algorithm due to strong data classification capabilities.

4) In most cases, a supervised learning algorithm LDA, MMC and LLE + LDA average recognition rate of better than unsupervised learning algorithm LLE average recognition rate. With LDA, MMC and LLE + LDA algorithm compared to the average recognition rate, LGE / MMC best average recognition rate of the algorithm. LGE / MMC algorithm is far superior to LLE algorithm is due to the algorithm takes the data distribution within the class, taking into account the distribution of data between classes.

This section compares the training samples in different algorithms of different maximum average recognition rate of change. We were in ORL, Yale and AR face image database comparison LGE / MMC algorithm with several classical algorithms of recognition performance. In experiments on each library, the first algorithm with PCA face image preprocessing, feature extraction algorithm and then using a variety of feature extraction, and finally with the nearest neighbor classifier complete classification.

Table 1. ORL Face Recognition Rate

Algorithms	sample set				
	2	3	4	5	6
LAD	76.42(38)	84.09(38)	85.17(38)	86.23(38)	88.38(38)
LLE	71.60(44)	72.39(36)	76.46(50)	81.00(28)	88.99(50)
MMC	75.78(40)	83.44(40)	85.85(38)	87.32(38)	90.66(40)
DLA	76.86(36)	85.27(36)	88.98(46)	91.71(38)	93.39(38)
LLE+LDA	73.60(36)	86.07(40)	90.13(40)	91.40(38)	95.35(18)
LGE-MMC	76.41(40)	86.41(50)	92.98(50)	94.62(42)	96.25(50)

Table 2. YALE Face Recognition Rate

algorithm	sample set				
	2	3	4	5	6
LAD	81.83(14)	85.54(14)	88.20(14)	87.84(14)	88.36(14)
LLE	84.83(17)	84.07(11)	86.55(9)	90.00(11)	90.53(10)
MMC	81.19(21)	83.62(14)	86.89(14)	86.20(14)	87.29(14)
DLA	88.39(28)	91.37(30)	92.27(30)	94.96(28)	95.01(28)
LLE+LDA	89.44(34)	93.10(28)	93.75(16)	95.55(14)	95.45(14)
LGE-MMC	94.66(48)	95.27(28)	94.80(50)	95.58(50)	96.55(48)

Table 3. AR Face Recognition Rate

algorithm	Sample set				
	2	3	4	5	6
LAD	71.21(115)	76.24(115)	82.84(115)	87.55(115)	89.33(115)
LLE	70.29(135)	75.37(120)	83.84(130)	85.64(135)	88.38(135)
MMC	67.54(120)	76.03(120)	82.60(140)	86.11(120)	89.68(120)
DLA	70.79(150)	79.73(150)	87.87(150)	89.74(150)	91.58(150)
LLE+LDA	70.58(140)	80.02(120)	88.53(120)	90.48(125)	92.50(120)
LGE-MMC	71.98(105)	81.68(80)	90.35(140)	92.21(75)	94.40(85)

As can be seen from Table 1-3, LGE / MMC algorithm ORL, Yale and AR face database have made a very good recognition effect. Can be obtained from Table 1-3 the following conclusions:

1) In Table 1, LGE / MMC algorithm maximum average recognition rate in several other classical algorithms, especially in the case of large sample more obvious effects. This is due to the sparse samples, LGE / MMC algorithm can not accurately reflect the sample in the original space, the distribution manifold.

2) Table 2, LGE / MMC algorithm maximum average recognition rate in several other classical algorithms. But with the increase of the sample, the maximum average recognition rate of the algorithm change is not very obvious, which is due on the Yale face database greatly influenced by light.

3) In Table 3, several classical algorithms in the small-scale face database can get a better recognition result, but for large databases such as AR face database, these algorithms have to be further improved recognition performance, and LGE / MMC algorithm than the other several algorithms have a distinct advantage.

4. Conclusions

In this paper, LLE and MMC algorithm based on the LGE / MMC algorithm, and gives its derivation. The purpose of the algorithm is to maintain neighborhood premise of using

MMC to construct compact within the class diagram and within-class penalties map, use the adjusted balance parameters u, close to the number within the class k_c and class parameter within the adjacent k_p, to ensure that the vectors orthogonal to each other on the basis of partial graph embedding, so as to more effectively raise the face linear partial structure. In ORL, Yale and AR 3 kinds of people face database experiments also show that the algorithm has better performance locally maintained and used in face recognition is better than several other classical subspace learning methods. It is noteworthy that, better balancing parameters u, close to a number within the class k_c and class k_p, between adjacent parameter selection is still not a theoretical basis to determine, how to find the optimal parameters in theory to more effectively subspace face image to explore non-linear nature of high-dimensional data, internal structure, will be one of our future research directions.

References

[1] W. Zhao, R. Chellappa and P. J. Phillips, "Face recognition: a literature survey", ACM Computing Surveys, vol. 35, no. 4, **(2003)**, pp.399-458,.

[2] R. Chellappa, C. L. Wilson and S. Sirohey, "Human and machine recognition of faces: a survey ", Proceedings of the IEEE, vol. 83, no. 5, **(1995)**, pp. 705-741,.

[3] Q. S. Liu, H. Q. Lu and S. D. Ma, "A survey: subspace analysis for face recognition", Acta Automatica Sinica, vol. 29, no. 6, **(2003)**, pp. 900-911,.

[4] Y. S. Lin, "An improved kernel feature extaction method", Journal of Computer Aided Design, vol. 20, no. 1, **(2008)**, pp. 61-65,.

[5] M. Turk and A. Pentland, "Eigenfaces for recognition", Journal of Cognitive, vol. 3, no. 1, **(1991)**, pp.71-86.

[6] P. N. Belhumeur, J. P. Hespanha and D. J. Kriengman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, **(1997)**, pp.711-720.

[7] J. B. Tenenbaum, V. de Silva and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction", Sciences, vol. 290, no. 5500, **(2000)**, pp. 2319-2323.

[8] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding", Sciences, vol. 290, no. 5500, **(2000)**, pp. 2323-2326.

[9] L. K. Saul and S. T. Roweis, "Think globally, fit locally: unsupervised learning of low dimensional manifolds", JMLR, vol. 23, no. 4, **(2003)**, pp. 119-155.

[10] M. Belkin and P. Niyogi, "Laplaeian eigenmaps and spectral techniques for embedding and clustering." NIPS, vol. 14, no. 1/2, **(2002)**, pp. 585-591.

[11] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation", Neural Computation, vol. 15, no. 6, **(2003)**, pp. 1373-1396.

[12] D. Ridder and R. P. W. Duin, "Locally linear embedding for classification", Delft: Delft University of Technology, **(2002)**.

[13] D. Ridder, O. Kouropteva and O. Okun, "Supervised locally linear embedding", Lecture Notes in Computer Science, vol. 2714, no. 5, **(2003)**, pp. 333-341.

[14] X. M. Bai, B. C. Yin and Q. Shi, "Face recognition based on supervised locally linear embedding method", Journal of Information Computation Science, vol. 2, no. 4, **(2005)**, pp. 641-646.

[15] O. Kouropteva, O. Okun and M. Pietikainen, "Supervised locally linear embedding algorithm for pattern recognition", Lecture Notes in Computer Science, vol. 2252, no. 4, **(2003)**, pp.368-384.

[16] J. P. Zhang and H. X. Shen, "Unified locally linear embedding and linear discriminate analysis algorithm for face recognition", Lecture Notes in Computer Science, vol. 3338, no. 7, **(2005)**, pp.1-16.

[17] J. P. Zhang and L. He, "Ensemble-based discriminate manifold learning for face recognition", Lecture Notes in Computer Science, vol. 4221, no. 9, **(2006)**, pp. 29-38.

[18] X. F. He and S. C. Yan, "Face recognition using Laplacian face", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 3, **(2005)**, pp. 328-340.

[19] X. F. He and D. Cai, "Neighborhood preserving embedding", Proceedings of 10th International Conference of Computer Vision, Los Alamitos, pp. 1208-1213, **(2005)**.

[20] F. Keinosuke, "Introduction to statistical pattern recognition", 2nd ed, Washington D C: Academic Press, **(1991)**.

[21] H. F. Li and T. Jiang, "Efficient and robust feature extraction by maximum margin criterion", IEEE Transactions on Neural Networks, vol. 17, no. 1, **(2006)**, pp. 1157-1165.

[22] T. H. Zhang, D. C. Tao and J. Yang, "Discriminative locality alignment", Lecture Notes in Computer Science, vol. 5302, no. 6, **(2008)**, pp.725-738.

[23] T. H. Zhang, D. C. Tao and X. L. Li, “Patch alignment for dimensionality reduction”, IEEE Transactions on Knowledge and Data Engineering, vol. 21, no. 9, **(2009)**, pp. 1299-1313.
[24] J. C. Maxwell, “A Treatise on Electricity and Magnetism, Oxford: Clarendon, 3rd ed., vol. 2, **(1892)**, pp. 68–73,.
[25] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, **(1963)**, pp. 271–350.