The 2012 2nd International Conference on Circuits, System and Simulation (ICCSS 2012)
IPCSIT vol. 46 (2012) © (2012) IACSIT Press, Singapore

An Improved Algorithm for Nonlinear Dimensionality Reduction in Image Processing

Dong Li +

College of Computer and Information Engineering，Xinxiang University
Xinxiang, China

Abstract. Dimensionality reduction is the key link for high dimensional nonlinear images to improve recognition rate, and some traditional algorithms have got some results in dimensionality reduction of image processing, but also expose their own short comings. In order to achieve the ideal recognition effects to high dimensional nonlinear images based on the analysis to traditional dimension reduction algorithm and the refinement from their advantages, this paper puts forward an improved nonlinear dimensionality reduction algorithm to solve the short comings of the traditional algorithm. Respectively, the simulation experiment is done on the database images of ORL and CMUPIE, to verify the viability of the algorithm for high dimensional nonlinear image in pattern recognition.

Keywords: Linear discriminant analysis, Nonlinear dimensionality reduction, Recognition rate

1. Introduction

Image pattern recognition technology is an important branch of the information processing field, especially in reconnaissance, finance and management departments it has broad applications and exploring prospects, it has already become a key research project in pattern recognition and artificial intelligence. In order to complete the image recognition, to high dimensional image (such as facial image) dimensionality reduction has become the key technology link before the image recognition, because after the dimension reduction process the image computing complexity have significantly reduced, storage space also has significant savings. For the moment, the main methods of dimensionality reduction to high dimensional image data have linear dimensionality reduction (such as LDA [1] and PCA [2]) and nonlinear dimension reduction methods (such as LLE [3] and ISOMAP [4], etc.). Although linear dimension reduction method can realize dimensional reduction to high dimensional data through performance target looking for the linear transformation matrix, but since light, expression, posture and factors make high dimensional face image have obvious characteristics of nonlinear manifold [5], after data dimension reduction it will produce the loss phenomenon of the nonlinear characteristic of the original data, causing the image data distortion after feature extraction, which reduces the image recognition rate.

Although nonlinear dimension reduction algorithm can make the training sample manifold structure being maintained, but it is hard to through the analysis of new sample points to get low dimensional projection, from the view point of classification it is not the ideal choice. In view of the defects of dimensionality reduction method of above two kinds of high dimensional image data, this paper puts forward a new image processing method (hereinafter referred to as LDA and LLE), which merges LDA and LLE together, it is able to solve the lost defect of the nonlinear characteristics of the original data after dimensionality reduction, and can make the problems existing in nonlinear dimension in low dimensional projection to get very good solving results. The experiment shows that the image recognition rate processed through this method is much higher than the traditional linear and nonlinear dimensionality reduction method.

+ Corresponding author. Tel.: + 8613703738505.
E-mail address: itislidong@gmail.com.

2. The implementation and analyses of LDA algorithm

Linear discriminant analysis (LDA) is the classic algorithm introduced in 1996 by Belhumeur for pattern recognition and the field of artificial intelligence. In order to ensure the mode sample after projection have the best separability in new subspace, it rejects the high dimension model samples to the best identified vector space, makes the model samples in new subspace to implement spread matrix in the minimum categories and the maximum distance between categories, and to ensure the effect of compression space and the extraction of classified information. In the process of face image processing, in order to eliminate the difference between the identification information and the suppressed image, the orthogonal vector spreading in the category is taken as feature face space, which ensures the rate of discrete degree minimum between and in the categories so as to avoid the effects to face image recognition by light, expressions change, and other factors.

In order to realize LDA algorithm [7, 8], at first adopting a group $Z = \{Z_i\}_{i=1}^{C}$ as the training set, in which C represents the number of categories, in each category $Z_i = \{z_{ij}\}_j^{ci} = 1$ there are C_i face images z_{ij}, thus in training set there is face image $N = \sum_{i=1}^{C} C_1$ which uses $J = (I_W \times I_h)$ to express the image column vectors, the size of image is $I_W \times I_h$, J dimensional real space is expressed in R^J, in which there is $z_{ij} \in R^J$, and then through a set of features $M(<<J)$ and base vectors $\{\psi_m\}_m^M = 1$ to guarantee the ratio maximum among spread matrixes between and in categories, usually it uses following expression to do optimization:

$$\psi = \arg\max \psi \left(\psi^T S_b \psi * \left(\psi^T S_W \psi\right)^{-1}\right), \psi = [\psi_1, \psi_2, \cdots, \psi_m], \psi_m \in R^J \tag{1}$$

In the expression, Sb presents the matrix between categories, which can be expressed as:

$$S_b = 1/N \left(\sum_{i=1}^{C} C_i (z_i - z)(z_i - z)^T\right) = \sum_{i=1}^{C} \phi_{b,i} \phi_{b,i}^T = \phi_b \phi_b^T$$

In the expression, SW presents the matrix in category, which can be expressed as:

$$S_W = 1/N \left(\sum_{i=1}^{C} \sum_{j=1}^{ci} (z_{ij} - z_i)(z_{ij} - z_i)^T\right)$$

In the expression, $z_i = 1/C_i \left(\sum_{j=1}^{ci} z_{ij}\right), \phi_0 = [\phi_{0,1}, \cdots, \phi_{b,c}], \phi_{b,i} = (C_i/N)^{1/2} (z_i - z)$ is the mean of category z_i. It is easy to see that optimal expression (1) is equivalent to the following characteristic expression:

$$S_b \psi_m = \lambda_m S_W \psi_m, m = 1, \cdots, M \tag{2}$$

But because there exists the problem of small sample, in process of base vector ψ is searching for maximum characteristic vector corresponding to the former M characteristic values of $S_W^{-1} S_b$, when doing the recognition of face image, if SW is not the odd there usually appears degradation problems. For this kind of problems often Fisherfaces method is used to join in PCA and remove the zero space of spreading matrix in and between categories to solve. But the maximum premise condition to get expression (1) is $|\psi^T S_W \psi| = 0$ and $|\psi^T S_b \psi| \neq 0$, so that zero space of spreading matrix in S_W category containing some favorable characteristics will be discarded.

These favorable characteristics influence the image recognition effect, causing the decline of image recognition rate. Apparently through certain performance goals to look for linear transformation matrix to achieve LDA method of dimension reduction to high dimensional data, for the manifold and high dimensional nonlinear face image, it is influenced by the light, the posture, the facial expression and other factors, because through the single linear transformation the nonlinear characteristics of the original data will lose, which causes data distortion after the feature extraction, so as to influence the image recognition effect [6].

2.1. The realization and analysis of local linear embedded (LLE) algorithm

Contributions to the congress are welcome from throughout the world. Manuscripts may be submitted to Aiming to the defects of LDA algorithm, Roweis and Saul in 2000 brought forward the local linear embedded (LLE) algorithm. In LLE algorithm the adjacent points of high dimensional space are also neighboring in low dimensional manifold, i.e. it can remain good local neighborhood relationship; this is also the key properties of this method. In LLE, $X = \{x_1, x_2, \cdots, x_N \mid x_i \in R^D\}$ and $Y = \{y_1, y_2, \cdots, y_N \mid y_i \in R^d\}$ $(d<D)$ are respectively noted as input and output sets, the mapping from x_i to y_i is $f : x_i \mapsto y_i$ which can realize through the following steps:

Step 1: ensure the neighbor field of sample points in the high dimension space, there are two commonly used methods to each neighbor field of sample $x_i(i = 1,2,\cdots,N)$:

Through the definition of some distance d (. , .) to measure the distance of xi to the rest of the sample points, namely the determination of K neighbor field, thus K sample points with the smallest distance are selected as the neighbor field points; (2) through the definition of neighbor field ε, if xj is located in the neighbor field ε of xi, then one of neighbor points of xi is xj.

Step 2: through the neighbor field to the locally linear approximate center sample points, to get the weight matrix W used for the reconstruction, the target equation of local linear approximation is $\varepsilon(W) = \sum_{i=1}^{N}\left\|x_i - \sum_j w_{ij}x_j\right\|^2$, from the target equation, if conditions $x_j \notin N_k(x_i)$ is established, then $w_{ij} = 0$, this time the error of linear approximation is the smallest, therefore, the solution of the reconstructed weight matrix *W* can be converted into to the following solution of the constraint problem to following conditions:

$$\begin{cases} \min\limits_{W} \varepsilon(W) \\ s.t. \sum_{j=1}^{N} w_{ij} = 1, \quad i = 1,2,\cdots,N \end{cases} \tag{3}$$

Step 3: reconstruct the low dimensional embedded *Y*. Solving the minimum value of cost function $\varphi(Y) = \sum_{i=1}^{N}\left\|y_i - \sum_j w_{ij}y_j\right\|^2$ is in order to keep the neighborhood purpose of high dimensional manifold in the low dimensional manifold, the cost function can be simplified as $\varphi(Y) = \sum_i \sum_j M_{ij} y_i^T y_j = Trace(YMT^T)$, the large sparse matrix is *M= (I-W)* ^{T} *(I-W)* which joins the normalized constraints $YY^T/N = I$ (rotation is invariant) and centralization constraint conditions $\sum_{i=1}^{N} y_i = 0$ (translation is unchanged). To solve the conditional extreme value, with Lagrange multiplier method $(M - \wedge)Y^T = 0$ can be got, so the solution to the original problem is changed into the solving problem on matrix *M* characteristic value. Hypothesis: the smallest d+1 characteristic values of matrix M is $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_{d+1}$, the corresponding d+1 characteristic vectors are $v_1, v_2, \cdots, v_{d+1}$, $\lambda_1 = 0$ and $v_1 = [1,1,\cdots,1]^T$ can be obtained, in order to meet the centralization constraint conditions, this feature vector can be ignored, and directly fetch $Y = [v_2, v_3, \cdots, v_{d+1}]^T$.

Analyzing step 2, in LLE algorithm, the local manifold in high dimension space is come from weight value among the data points in the high dimension space. In step 3, in the condition of keeping the weight value invariant, the local linear embedded new data set in low dimensional space can be obtained. Therefore, this algorithm can complete the mapping from its original high dimensional space to low dimensional space through keeping local contact among the data. But in this step the complexity is fairly high, it is also the most time-consuming step in calculation, in the conditions of without getting the optimization, the time complexity of minimum d characteristic values is *O(dN*^{2}*)*, solving the large sparse matrix is used to lower down the time complexity.

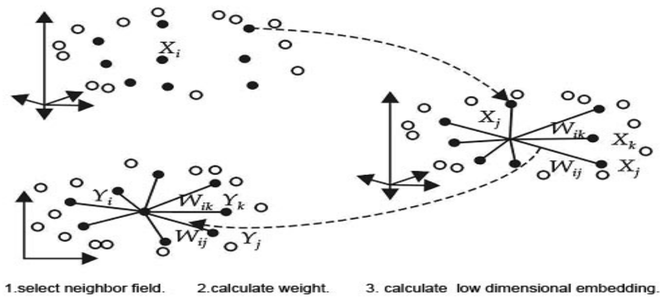

Fig. 1: LLE algorithm implementation steps

In figure 1, the weight matrix W for reconstruction is the key of the algorithm, and plays the link role between high dimensional data and low dimensional embedding. Although using this algorithm achieves the ideal effect in high dimensional data visualization, but there still exist some defects in the following:

First of all, there is not an explicit method that can realize the embedding from new sample points to low dimensional space, only using the verification method can obtain the new space dimension number, and there is not an explicit criterion to determine. Second, the algorithm does not suit for processing the sample data with class mark. In addition, although it is in the child space of structure to keep the manifold structure of training sample, but the projection in low dimensional space of new sample points cannot be obtained [9], from the aspect of classification the algorithm is not the most ideal, especially to this kind of image classification problem with high dimensional nonlinear manifold structure like face it is not the best choice.

3. The realization of improved nonlinear dimension reduction algorithm

Based on the defects of above two kinds of data dimension reduction algorithm, this paper analyzes and synthesizes their advantages, does fusion and refining, puts forward a kind of new nonlinear dimension reduction method (hereinafter referred to as LDA&LLE algorithm). This algorithm's thoughts can be summarized as: at first using the linear supervised LDA algorithm to find each neighbor point of sample points, and then through the calculation to local reconstruction weight matrix in LLE algorithm to obtain the local embedded characteristics of training set in low dimensional space, at last, according to the characteristics of high dimensional nonlinear image through the calculation of training set and its low dimensional characteristics to acquire low dimensional features of new samples. Thus through the relationship between high dimension space and low dimensional space, this algorithm is able to build the clear mapping between new sample points and its low dimensional characteristics. The realization process of this algorithm is as follows:

Hypothesis: the training set consisting of n pieces of face image $\{x_i\}_j^n = 1$ is matrix X, thereinto xi (i =1, 2, …, n) is consisted by the non-negative grey value of an image and becomes the D dimensional column vectors.

The linear discriminant analysis algorithm (LDA) is used to realize the fetch of k neighbor points of each sample point. The process of looking for the neighbor points of sample points is: first of all, LDA subspace A is obtained through the dimension reduction to training set x. And then, in high dimensional space freely choosing two points of xi and xj, and the distance between these two points is defined as Euclidean distance corresponding to its LDA feature point, such as expression (3). Finally, the new definition of European distance is used to complete the selection of k adjacent points of each sample point.

$$\text{Distance}_{ij} = \left\| A^T x_i - A^T x_j \right\| \tag{4}$$

LLE algorithm is used to calculate local reconstruction weight matrix, to get the low dimension local embedded features of training set. The 2, 3 steps in LLE algorithm are used to complete the calculation of local reconstruction weight matrix, in this step, the calculation of local reconstruction weight matrix is still completed in the original high dimensional space, thus it not only ensure the nonlinear structure of high dimensional data, but also is able to get the low dimensional local embedded features $Y = \{y_i\}_i^n = 1$ of training set X.

According to the training set X and the low dimension characteristic Y in former step, the low dimensional features corresponding to the new samples are calculated. Such as face this kind of image has obvious high dimensional nonlinear flow structure, because the relationship of local neighborhood points is constant between its high and low dimensional spaces, so as long as neighborhood relationship between new sample points and training set X is obtained, the low dimensional embedded characteristics of new sample points can be made out.

Suppose the new sample point is X_{new}, firstly new sample point doing projection to A subspace obtains the characteristics $A^T X_{new}$ of LDA, and then through the formula (3) in LLE algorithm to acquire k neighbor points of new sample point X_{new}, which are noted as x_{newj}*(j=1, 2, and k)*, finally, the formula (4) is used to find the weight matrix w_{new} between new sample points and the training sample. The calculation of low dimensional embedded features (y_{new}) to this new sample point can be completed through the following expression:

$$y_{new} = \sum_{i=1}^{k} w_i^{new} y_i, \left(y_i \in Y \in R^d\right) \tag{5}$$

Note: in formula (5) y_{new} represents the low dimensional projection of new sample point in LLE nonlinear subspace, y_i represents the low dimensional projection of neighbour points x_{newj}*(j=1, 2, and k)* of new sample points in LLE nonlinear subspace.

4. Simulation experiment and results analysis

In order to validate the effectiveness of this algorithm in nonlinear dimension reduction, ORL and CMUPIE images in face library are selected to do simulation tests, LDA, LLE and LDA&LLE algorithms are used to do image processing and contrast experiment. Before tests, firstly the images are automatically detected, posture and facial expression are positioned and the normalized processing is done, then the nearest neighbor classifier is used to do classification.

4.1. ORL face database simulation experiment and analysis

ORL face database is created by Cambridge university AT&T laboratory, the database contains different expressions and postures, there are total 40 people with slightly tilted (no more than 20 degrees) and different lighting face, each of them has 10 pieces of different face images, totally there are 400 pieces of images with resolution 92*112 of face gray[10]. This paper randomly selected five pieces of normalized picture to do training, the rest is used as test, namely 200 pieces of picture are used as the training sample, the other 200 pieces are the testing samples, so as to avoid the crossover phenomenon between training set and test set, which is shown in figure 2.

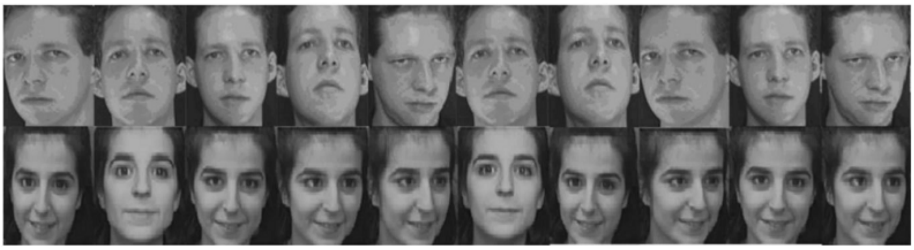

Fig. 2: ORL library image examples

In the experiment, dimension D=50 that LLE algorithm embedded in the space and the neighborhood number K=18 of each sample point are selected. After using LDA algorithm to realize dimension reduction, dimension number does not exceed its individual category number minus 1, and these dimensions are trained one by one, using the nearest neighbor classifier (NNc) to do classification, the samples belonging to different categories are used as representative points, at last, through comparing the delegate points the test samples are allocated to the category of training samples with the smallest space distance. The operation results of LDA algorithm and this proposed algorithm in the system are shown in figure 3 below.

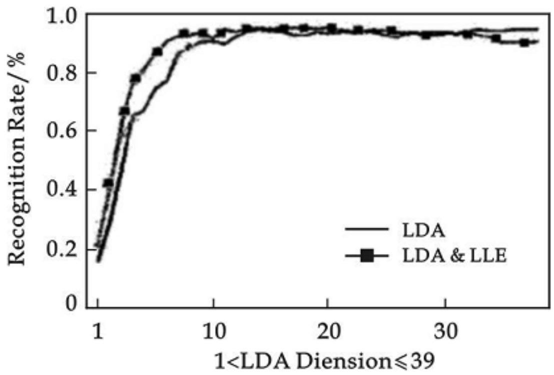

Fig.3: The simulation experiment results of two kinds of algorithm

By this experiment the algorithm (LDA&LLE) in this paper in the classification of image recognition, its advantage is significantly higher than the linear discriminant analysis (LDA) algorithm, and improves the image recognition rate average about 4.65%, the face recognition system identification is enhanced to a new height. The algorithm comprehensively uses the advantages of LDA algorithm and LLE algorithm, so as to after manifold dimension reduction the loss phenomenon of useful information for identification can be avoided, to reach the purpose of after reducing the dimension the original data still keeping the original features, the effectiveness of this algorithm is verified.

Table. 1: The experiment results of CMUPIE library face images

algorithms	Dimension of features	Recognition rate /%
LDA	11	70.58
LLE(K=80)	40	61.49
LDA&LLE(K=100)	55	88.21

4.2. CMUPIE face library simulation experiment and its analysis

CMUPIE face library is created by Carnegie Mellon University in November of 2000, coming from 68 volunteers with 41368 images, including 13 different attitude and four kinds of expression photos in 43 kinds of light conditions. In this simulation experiment, in the library random 400 photographs are selected for each person, and then from the 100 pictures randomly 50 pieces are selected as the training set, the rest 50 pieces are as a test set, part of the example pictures are shown in figure 4. Finally, respectively LDA, LLE and LDA & LLE algorithm in this paper are used to process to the selected images, the experimental results can be concluded as shown in table 1.

Fig.4: Examples of CMUPIE image library

From table 1, it is shown the nonlinear dimension reduction algorithm (LDA & LLE) improves the image recognition rate to 88.21%, it is obviously higher than linear discriminant analysis algorithm (LDA) and local linear embedded algorithm (LLE), and in the condition of light, expression and posture three factors changing together it also has certain robustness, makes the system recognition performance improved. From above two simulation experiments the conclusions can be drawn, they are:

LDA algorithm is used to acquire low dimension characteristics that are gained by LDA & LLE algorithm of neighbour point, at the same time which has the features of the minimum differences in category and the largest difference between categories, and ensures that the algorithm has strong discrimination ability.

LDA & LLE algorithm integrates classification advantages of LDA and LLE two algorithms to do face recognition, makes the image recognition rate significantly higher than other two algorithms.

In essence, LDA & LLE algorithm is nonlinear dimension reduction algorithm, it not only avoids the lost phenomenon of the nonlinear characteristic of original data in LDA algorithm, but also solves the defects that LLE algorithm is hard to get low dimensional projection through the analysis of new sample points, even if in the changeable situation of illumination, expression, posture and other factors, it still can get higher image recognition rate.

5. Last word

To sum up, the nonlinear dimension reduction algorithm (LDA&LLE) mixes together the advantages of LDA algorithm and LLE algorithm, and gets rid of their defects, not only keeps the nonlinear structure of this kind of high dimensional nonlinear image such as face, but also has strong identification ability. Using this algorithm is obvious to the reduction dimensional effect of high dimensional image, in the largely changeable situation of illumination, expressions and gesture it still has some robustness, in the aspect of increasing the image recognition rate it has extensive usability. But to the influence of recognition rate by the number of neighbor points, the determination of the most neighboring point K value is the problem to be solved.

6. References

[1] PNBelhumeur, JPHespanha, DJ Kriegman. Eigenfaces vs fisherfaces; recognition using class specific linear projection. *IEEE Trans Pattern Anal.* Machine Intell, 2008, 19(7): 711-720.

[2] Turk MA, Pentland. AP. Eigenfaces for recognition. *Cognitive Neurosc*i, 2007, 3(1): 71-86.

[3] Rowels S T, Saul LK. Nonlinear dimensionality reduction by local linear embedding. *Science*, 2000, 290(5 500): 2323-2 326.

[4] TenenbaumJ B, DeSilva V, Langford J C. Aglobal geometric framework for nonlinear dimensionality reduction. *Science*, 2000,290(5500): 2 319-2323.

[5] SamRoweis, LawrenceSaul, Geoff Hiaton. Global coordination of local linear models. *Advances in Neural Information Processing System*, 2001, 14: 452-456.

[6] SamT, Rowels, LawrenceKSaul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 2000, 12(290): 22.

[7] LIUJW, Plaitaniots KN, Venetsanopulos AN. Face recognition using lda based algorithms. *IEEE Transaction son Neural Networks*, 2003, 14(1): 195-200.

[8] LIUJW, Plataniotis KN, Vevetsanopoulos AN. Regularization Studies of linear discriminant analysis in small sample sizes cenarios with application to face recognition. *Pattern Recognition letter,* 2005, 26(2): 181-191.

[9] Zhu En, Yin Jianping, Zhang Guomin. Fingerprint Enhancement Using Circular Gabor Filter. *Proc of Intl Conf on Image Analysis and recognition*, 2004, 75—758