TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol.12, No.3, March 2014, pp. 1924 ~ 1932
DOI: http://dx.doi.org/10.11591/telkomnika.v12i3.4356

Local Application of Non Negative Matrix Factorization Algorithm in Face Recognition

Lou Xiongwei*[1,3], Huang Decai[2], Fang Luming[3], Xu Aijun[3]
[1]College of Information Engineering, Zhejiang University of Technology, Hangzhou, Zhejiang, 310032, China
[2]School of Computer Science & Tchnology, Zhejiang University of Technology, Hangzhou, Zhejiang, 310032, China
[3]College of Information Engineering, Zhejiang A & F University, Linan, Zhejiang, 311300, China
*Corresponding author, e-mail: lxwzjfc@163.com

Abstract

Face recognition is a challenging issue in the field of multi-science, the main contents of the research is how to make computer have the ability of face recognition face recognition technology involved in a lot, which is a key feature extraction and classification method, this paper focuses on the study of related theory. Non-negative matrix factorization trapped MF) algorithm and local non-negative matrix factorization (LNMF) algorithm is a feature extraction method based on local features, has been successfully used in face recognition "but NMF algorithm in face recognition rate is low, although LNMF algorithm to a certain extent, improve the recognition rate, but its price is to increase the number of iterations. In addition, the two algorithms have failed to solve good nonlinear separable problems "the kernel method combined with LNMF algorithm, the kernel local non-negative matrix factorization (KLNMF) algorithm, the first by a nonlinear transformation of the original space to high-dimensional space, making samples linearly separable, and then use the LNMF algorithm to extract face features. In the classification part, paper presents the decision rules of classification of their own, and design based on the NMF subspace classifier.

Keywords*: non-negative matrix factorization, pattern recognition, algorithm, database*

1. Introduction

With the development of society, the social parties face automatic fast identity has become increasingly urgent requirements. The biological characteristics are the inherent attributes of people, has its own stability and strong individual difference, therefore, the biological characteristics is the best basis for authentication, biometric personal identification including iris feature, retinal features, voice characteristics, the speaker's facial features, such as fingerprint features. Among them, use of facial features for identity verification is the most natural and the most direct means, compared to other biological features of human body, it has direct, friendly, convenient, low cost, reliable, and easy to be accepted by users [1]. Therefore, it is in many of the matter determination technology favored, face is the most common mode of human vision, face the vision information plays an important role and significance in the exchange of people and communication [2].

Face recognition technology has a broad application prospect, organs and units of the security and attendance, network security, bank, customs and border control, property management, military security, computer login system, can be used in face recognition technology [3]. In the organization, access control system based on face recognition not only can guarantee the safety unit, and can automatic attendance of staff. In the electronic commerce, international trade, banking and other occasions, face recognition technology can overcome the shortcomings of the instruction and the password in a certain extent [4]. In the banking system, face recognition system does not need any electronic or mechanical keys, can reduce the loss of keys, forget the password problem caused by [5]. At present, the face recognition technology in a certain range of application has been successfully trapped inside.

Face recognition technology has been used in image database retrieval technology, in large face database retrieval and indexing face images of the same or similar face like for

Received August 26, 2013; Revised September 30, 2013; Accepted October 20, 2013

example: query and management of public security departments can criminals library using face recognition technology. In daily life, people recognize the person around with the most is the face, it is non-invasive, direct, friendly, convenient features, so it is the most acceptable authentication methods. Face recognition system in general by the camera will face image acquisition, feature extraction, stored in the template library, in authentication first will face like a segmented out from the background, then the template image and database collected for comparison [6-7]. With the wide use of network technology and desktop video, image capture devices are becoming standard peripherals of the personal computer, and the use of electronic commerce and other cyber source authentication is to put forward new requirements. Therefore, face recognition has become biometric identity verification means the most potential [8].

This paper mainly focuses on theory and algorithm of feature extraction and classification of face images is studied, to improve the accuracy of face recognition, practical and further promote the technology of face recognition. Research on face recognition technology, its significance lies not only in content authentication, content retrieval based on actual needs, but also can promote the development and application of image processing, pattern recognition theory, at the same time, because of the particularity of the face pattern, conducts the research to the face recognition, for the promotion of cognitive science, also has the positive significance of physiology and other relevant subjects the.

2. Face Recognition Method

2.1. Set of Subspace Classifier

The method of feature selection and extraction of the essential characteristics of the subspace method, its main thought is in the original space (sample space) in the search for a suitable subspace (feature space), the high dimensional samples are projected to low dimension space, classification in subspace. Do at least two benefits: the reduction of high-dimensional samples hand! Compression, greatly simplifies the calculation, on the other hand, the projection of high-dimensional samples in the subspace of the can is better than in the original space separability, which we find an important standard subspace. For example, extract features of face image is shown in Figure 1.

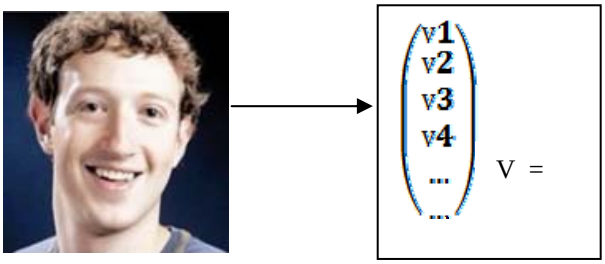

Figure 1. Extract Features of Face Image

Take a picture of the ORL face database, the size of 112*92, dimension is the original sample space is very large, is not conducive to the classifier design, more important is the original image contains too many invalid information, can't directly reflect the essence, and the correlation between the sample and the sample is very high. In order to design for the classifier, need to the original image from measuring empty mapped to the feature space of low dimension to ask, the image is studied in this feature space using the feature vector representation. Therefore, feature extraction in a broad sense refers to transform.

By calling two features from high dimensional space to one or more after mapping characteristics of low dimensional space, they are a combination of original features, usually is a linear combination of "mapping method can be divided into linear and nonlinear. But whether linear or nonlinear projection, purpose is to make the feature extraction processing for classifier, if the classifier error rate as the standard, that is the error probability of the lowest set of features, is a group of the best feature "but in reality, even in the known class conditional probability density distribution under the condition, calculation error rate is also very complex,

moreover, the distribution of the actual problem often do not know, this makes effective use of error probability as the standard to analyze the characteristics of direct comparisons difficult" therefore, validity analysis features usually use the criterion, usually have: class, class distance, dispersion, entropy function standard etc. The application of these different criteria, we can get different feature extraction methods, can be divided into: extraction methods according to the characteristics of Euclidean distance metric! Extraction method according to the characteristics of probability distance criterion.

Non-negative matrix factorization is a new kind of matrix decomposition method proposed in the world at present, namely non-negative matrix factorization is the matrix whose elements are all non-negative matrix under the constraints of the decomposition method. The method has caused the scientific researchers in various areas of attention: on the one hand, many large-scale data analysis methods in the scientific research of the need for effective treatment through the matrix form, and non-negative matrix factorization theory provides a new way for human to deal with large-scale data; on the other hand, non-negative matrix factorization algorithm compared with the traditional algorithm and a word, is simple to achieve. Decomposition and decomposition results on the interpretation, and occupy less storage space etc. Therefore, non-negative matrix factorization algorithm is one of the hot issues in the current study. Known non-negative matrix V, looking for nonnegative matrix factorization H which was shown in formula 1.

$$V = W * H \quad (1)$$

In order to find an approximate decomposition process, we must first define the objective function to guarantee the effect of approximation. The objective function that can use some distance from two nonnegative matrices B and C to obtain, one of the more useful method is to measure between the matrix B and C Euclidean distance. If and only if B=C, type a minimum value for 0.

$$||B - C||^2 = \sum_{ij} (B_{ij} - C_{ij})^2 \quad (2)$$

$$D(B||C) = \sum_{ij} \left(B_{ij} \log \frac{Bij}{Cij} - Bij + C_{ij} \right)^2 \quad (3)$$

Gradient method is perhaps the most simple, the most convenient method, but its convergence rate is very slow, and the convergence of gradient algorithm for the choice of step is very sensitive on this for practical applications, it is not very convenient. Non-negative matrix factorization in the base image study, coupled with nonnegative constraints, thus, the pixel based image training value obtained, as well as for the weight coefficient of reconstruction, are nonnegative. In this mode, only the linear superposition operation will not reduce the allowed. This will ensure that the local integral non-reducing mode. Therefore, NMF is considered as a method for extracting local features. However, NMF learning part, sometimes not as we expected, localization, and the original NMF method recognition at certain times of the rate is not very high. Local non negative matrix decomposition of the value function, which is the objective function is shown as follow:

$$D(V, WH) = ||V - WH|| + \lambda \sum_{i,j} Hij \quad (4)$$

Compared with the classical NMF method, LNMF method is based on image spatial feature localization is more obvious, and the base image to two value image, with sparse greater weight matrix, dimension reduction effect more obvious, almost every piece of basis images are only retained the basic component carrying the maximum amount of information. The LNMF algorithm through the enhancement of orthogonality as much as possible to reduce the correlation matrix between images, which makes the weight distribution is more

concentrated, the weight matrix sparsity is very good. This improved to some extent improved the feature vector.

2.2. Discussion of Classification Rules

Classification is the important process of pattern recognition, the paper has introduced several stages of face recognition, and the last stage is the classification. The design of classifier is a key problem in the whole pattern recognition. The common classification method is often discussed with SVM and nearest neighbor classifier method.

SVM is a kind of generalization performance in recent years in the field of pattern recognition as well. In theory, the algorithm will eventually translate into a two type of optimization problems, SVM is the global optimal point. It initially showed much better performance than the existing methods, and has good generalization ability, can solve small sample learning problems better, currently there are some experiments proved this conclusion.

The nearest neighbor method is the distance between samples of a classification method based on. Because the method is analyzed in theory, one of the important methods until now is still in pattern recognition. When the nearest neighbor method is provided with a threshold of K, from high dimensional space geometry angle the discrimination process of understanding: the sample space of each class of all samples J point as the center, with the threshold do hyper sphere radius, judge the samples to be tested in which super ball body, repeated samples to be tested and calculated the hyper sphere center from the toe, eventually will test sample discrimination and its recent hyper sphere center similar. Obviously, for different applications, the acquired image is different, in order to identify accurately, could lead to the decimation of different features, but the aggregation of different characteristic is not the same. In the face detection applications, we put the sample capacity of face sample set of m as the original image matrix V in the NMF algorithm, where Vj is the j side face images in a stacked array, as shown in Figure 2.

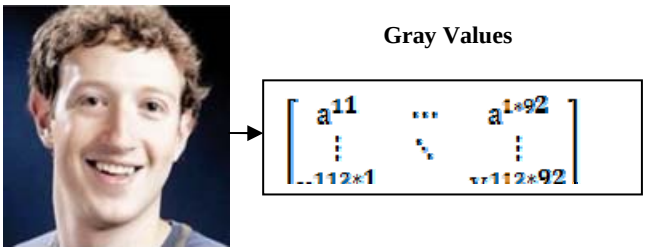

Figure 2. The Scan of Images

Thus separately on the M human face samples by line scanning, stored in columns, m column of a high-dimensional matrix V, each column is a face of sample, matrix V decomposition for matrix W and matrix H which was shown as follow:

$$\begin{bmatrix} v^{11} & \cdots & v^{1n} \\ \vdots & \ddots & \vdots \\ v^{m1} & \cdots & v^{mn} \end{bmatrix} = \begin{bmatrix} w^{1} \\ \cdots \\ w^{m} \end{bmatrix} * [h1 \ldots \ldots hm] \quad (5)$$

The corresponding teeth corresponding to the matrix decomposition of R amplitude in facial feature based image columns stacked base consisting of image matrix. H was a synthetic weight matrix corresponding to each base image of each face images. After a simple transformation that can be used conveniently, NMF and its improved algorithm depict face model and feature extraction. Each image samples for feature extract, all can be represented by a feature vector matrix W and the corresponding coefficient matrix H. The example of eigenvector matrix of a two-dimensional is shown in Figure 3.

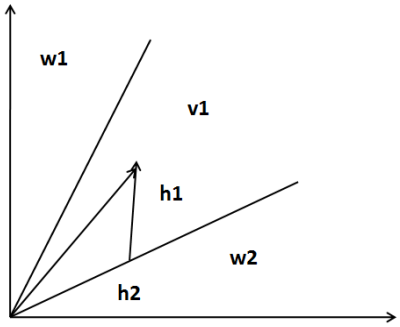

Figure 3. The Example of Eigenvector Matrix of a Two-dimensional.

Then, all the samples for each category in the can be described as shown in equation form, feature vector matrix W has become a feature of this kind of sample. This method completes the feature extraction in pattern recognition part. When extracting an arbitrary test samples, according to the reconstruction error can find the sample in which categories. Using NMF algorithm, we obtain a mapping subspace of each kind of original data and the original data in the spaces, some properties of these can reflect the original data space and structure relationship hidden in the original data. This forms the basis of "using the subspace for a newly acquired the unknown original data vector x, we can be respectively to the two categories of data subspace projection by characteristic matrix respectively, fixed the original algorithm.

In the classifier design for a while, the reconstruction error to decide the test sample classification. For example, in the two-dimensional plane, is mapped into feature space, the sample to the feature vector median distance, number of training samples but this method is only suitable for two kinds of samples of similar situation, if the number of training samples of two kinds of samples vary widely, category belongs in such a way that they can't well reflect the sample.

3. Experimental Data Analysis

Introduction to illustrate the kernel method is brought to the problem of image classification results, respectively, LNMF and KLNMF algorithms are compared in the experiment on iris plant data. Iris plant data is a classic Fisher database created in 1988, its main content is the characteristic data of some plants, is widely used in the field of pattern recognition. The iris plant data including a total of 3 categories, 4 features of each class, each class has 50 sample "the sample 1 and sample 2, sample is linearly separable 3, sample 2 and sample 3 linearly inseparable. In this experiment, we use LNMF and KLNMF algorithm, in order to make the results more intuitive display, this paper considers the sample after feature extraction is mapped to the two-dimensional plane, in order to observe the relationship between the various types of samples. Local non-negative matrix factorization algorithm, the number of feature vectors is 60. Nuclear localized non-negative matrix. Kernel function using the matrix decomposition algorithm is the Gauss kernel function, the parameter sigma of Gauss kernel function value is 11, the number of feature vectors is 50, the result is shown in Figure 4.

The study of a database is verified by using ORL face. The results show that in the feature extraction part of the KLNMF method using the kernel-based nonlinear analysis method, the nonlinear transformation to transform the original space to a high dimensional space, making samples linearly separable. This transformation will not because the dimension increases and increases the computational complexity, as long as the definition of the appropriate kernel function can now. Then, using the NMF method for feature extraction, the algorithm in the mode conversion matrix representation mode, can be respectively on pattern set for the values and eigenvectors, treatment to reduce the dimension of feature extraction, the efficiency of higher.

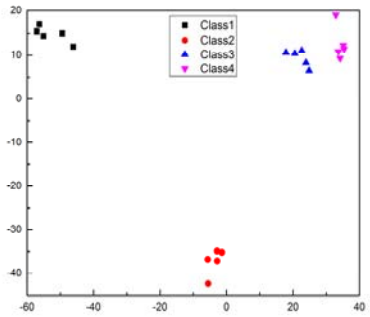

Figure 4. Classification Results of Plane Display

4. Analysis of ORL Face Database Experiment

ORL face database AT&T, the database contains 40 different 400 face images, each of the 10 picture. The image database, some people were taken at different time, and the light variation of illumination and expression, some people also wear glasses all images in the shooting back king unified uniform black, face images are the 256 gray level image frontal or near-frontal, size of 92*112 pixels, part of the image as shown in Figure 5.

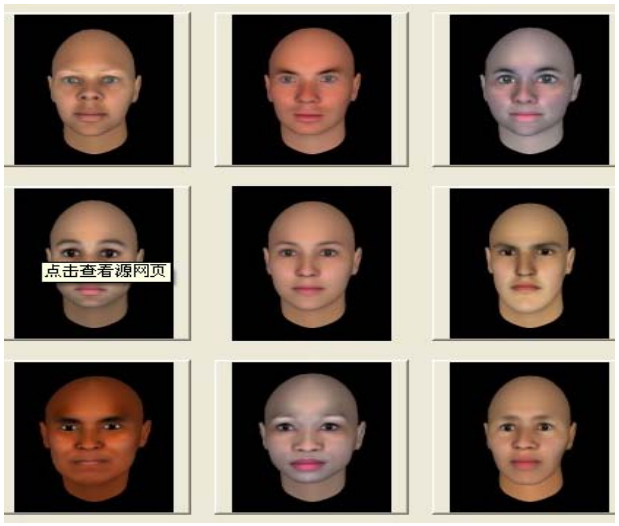

Figure 5. Database Instance

Each line is a different expression of ORL image, a total of 40 individual images database, in this paper, the face image of each individual as a sample, each sample has 10 sample data. Feature extraction method is introduced in combination with the front and classifier design, simulation algorithm for face images in the face database, and analyzes and compares.In general, the face pictures in size, pose, illumination will have different degrees of change, in order to test the recognition algorithm better, first of all need to face database for geometric correction, the image cropping, gray adjustment normalization preprocessing. In this experiment, 10 randomly selected from each picture in K images as training data, the remaining 10-K amplitude as test data. Sample selection is random, and the test and training set and do not cross. Of course, once the experiment is finished, can consider the test set and training set exchange; or the selected samples are divided into a plurality of sets, some take only one as the training set and test set. The i energy normalized of each picture was as follows:

$$Vij = \frac{Vij}{\sqrt{\sum_j Vij^2}} \quad (6)$$

$$Vij = \frac{Vij}{\max Vij} \quad (7)$$

As everyone knows, the iterative algorithm to initialize the algorithm convergence speed and the obtained results are very important to. On the one hand, select the appropriate near balance point initial value could accelerate the convergence speed; on the other hand, the equilibrium point, the choice of initial value has great influence on the final results.

For non-negative matrix factorization algorithm, initial value need to give H. The earliest methods is to use 0 and 1 between the uniform random spanning two nonnegative matrix, the characteristic of this method is no tendency, can get the objective function is sufficiently small in many experiments. This method is a local optimal solution. Another method from the Wild method based on the ball and K means clustering initialization method, its characteristic is consistent with the original data from the base had better and interpretability. As the preprocess step for expenses is larger, and the convergence when the objective function value is generally greater than the random method.

In addition, in the setting of classification rules, found that not every operation results are ideal, the results of the algorithm may be trapped in local optimal. However, when the feature matrix is initialized to a data region with its direction, each experimental result converge to the global optimum, as shown in Figure 6.

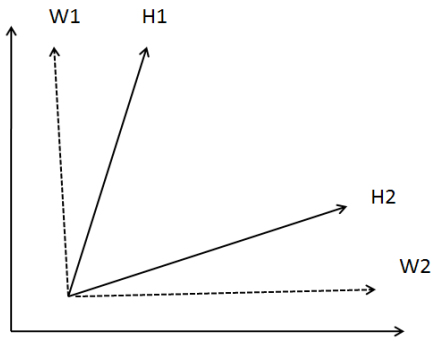

Figure 6. Initialize the Chart

In the use of nuclear localized non-negative matrix factorization algorithm, feature parameter vector r number and kernel function used has some influence on the experimental results. Therefore, the following experiment first by a number selecting different feature vectors and different kernel function parameter, the analysis of the two parameters on the recognition rate, the training data of K two in the experiment value is 5, the remaining 5 data as the test sample.Below we detailed analysis of each index of the above algorithms. First, the recognition rate is one of the most important performance indexes to measure the quality of face recognition algorithm. Secondly, one is also a key indicator of the running time of the algorithm, it relates to the algorithm can effectively be applied to the practice, which can be reflected by the number of iterations. The following gives illustration.

Face recognition using KLNMF algorithm in this experiment, the relationship between the feature vector and the number of the recognition rate, two angles, and the iterative times and the recognition rate and the relationship between NMF, LNMF algorithm in comparison to the experimental results. Kernel function with Gauss kernel function, the test sample number k = 5, sigrna = 11, Table 1 and Table 2, Table 1 shows the relationship between the feature vector and the number of the recognition rate of different subspace method, each parameter of the operating results and the average value 10 times, the number of iterations is 300.

Table 1. A relationship between the Number of Recognition Rate and the Feature Vector

Number of feature vactor	NMF(%)	LNMF(%)	KLNMF(%)
R=20	54	58	90
R=40	74	77	92
R=60	79	84	95
R=80	81	84	94
R=100	82	85	94

From Table 1 it can be seen, along with a number of feature vector increases, the recognition rate is rising, and reached a certain number in the feature vector, the recognition rate is up to a certain degree of stability, KLNMF compared with the previous two methods, to achieve this point, and the recognition rate than the other two methods of high it also shows that, KLNMF algorithm reduces the dimension effect is more obvious, in solving the problem of large data than the other two algorithms have a greater advantage.

Table 2 shows the relationship between the number of iterations and the recognition of different rate, R value is 60, the kernel function with Gauss kernel function, the test sample number k = 5, sigma = 11, each parameter of the operation results are averaged 10 times the value.

Table 2. Relationship between Number of Iteration and the Recognition Rate

Number of iteration	NMF(%)	LNMF(%)	KLNMF(%)
50	37.5	21	88
100	70	48	93
200	75	65	93
300	81	82	96
500	81	87	96

Table 2 shows, compared with the NMF method, LNMF method, LNMF method, recognition rate. The required number of stable than NMF force required for many times, can say, LNMF method with sacrifices the iterations to obtain high recognition rate. Although the KLNMF method in feature extraction process, the samples are mapped to high dimensional and infinite dimensional space, but the number of iterations did not increase, effectively solves the problem of LNMF method.

5. Conclusion

Face recognition technology, the integrated use of the pattern recognition, digital image processing, a large number of scientific knowledge in the field of computer vision, but also within the area of the important research topic. With the development of theory and recognition, face recognition technology in intelligent monitoring! Content retrieval fields based on the widely used. Because the NMF algorithm the nonnegative NMF algorithm, in recent years has been rapid developed in field of face recognition, and improve the LNMF algorithm. However, NMF, LNMF algorithm in solving nonlinear problems are still insufficient, this paper proposes the KLNMF algorithm, the kernel non-linear mapping, the data in the original space is mapped to the kernel space, and then in the kernel space for data analysis can be found and extracted from the original data characteristics of potential.

Acknowledgements

The word was supported by the Foundation of National Natural Science of China (30972361), Zhejiang Provincial Education Department Project (Y201018063), visual media, Zhejiang Key Laboratory of Intelligent Processing Technology (No: 2011E10003).

References

[1] Zheng Zhonglong, Zhao Jianmin, Yang Jie. *NMF with Log Gabor Wavelets for visualization*. Lecture Notesin Computer Scienee. 2005; 3691(2): 26-32.

[2] Bledsoe Lee, Chang Woo. *NMF-based approach to font classification of printed English alphabets for document image understanding*. Lecture Notes in Computer Scienee. 2005; 3558(7): 354-364.
[3] Liu Gang, Li XiHai. *Feature extraction of underground nuclear explosions based on NMF and KNMF. Lecture Notes in Computer Science*. Third International Symposium on Neural Networks. 2006; 8(2): 1400-1405.
[4] Li Tang, Li Yaming, Xu Kailai, Hou Xiandeng, Yi Lv. Sensitive and selective acetone sensor based on its cataluminescence from nano-La2O3 surface. *Sensors and Actuators B: Chemical*. 2008; 132(1): 243-249.
[5] Li Shutao, Wang Yaonan, Zhang Changfan. Objective evaluation and analysis of multisensorimage fusion. *Chinese Journal of Scientific Instrument*. 2002; 23(6): 651-654.
[6] JW Leonhardt. A new ppb-gas analyzer by means of GC–ion mobility spectrometry (GC-IMS). *Journal of Radioanalytical and Nuclear Chemistry*. 2003; 257(1): 133-139.
[7] Wang Jue, Wang Ren, Miao Duoqian. Data enriching based on rough set theory. *Journal of Environmental Sciences*. 2010; 29(3): 63-69.
[8] D Abbott, S Cunningham. *Development and Evaluation of Sensor Concepts for Ageless Aerospace Vehicles*. Threats and Measurands. NASA/CR-2002-211772. 2002; 11(2): 25.