# Face Recognition based on Multi-scale Singular Value Features

Ran Jin, Zhuojun Dong
Department of computer science and information technology
Zhejiang Wanli University
No.8, South Qian Hu Road Ningbo, Zhejiang, 315100
China
http://www.zwu.edu.cn

*Abstract:* - Singular value vecto r of an image is a valid feature for identification. But the recognition rate is low when only one scale singular value vector is used for face recognition. An algorithm was developed to improve the recognition rate. Many subimages are obtained when the face image is divided in different scales, with all singular values of each subimage organized and used as an eigenvector of the face image. Faces are then verified by linear discriminant analysis (LDA) under these multiscale singular value vectors. These multiscale singular value vectors include all features of an image from local to the whole, so more discriminant information for pattern recognition is obtained. Experiments were made with ORL human face image databases. The experimental results show that the method is obviously superior to the corresponding algorithms with a recognition rate of 97. 38%.

*Key-Words:* - multiscale; singular value decomposition (SVD) ; feature combination; face recognition; Fisher; ORL; subimage

## 1 Introduction

In the face recognition, extracting effective solution to identify features is the key of solving the problem. At present, people have been given a number of feature extraction methods. Thereinto principal component analysis (PCA or transform K-L) and Fisher linear discrimination analysis (FLDA) are classic method of feature extraction and are widely used. PCA was obtained the best features volume of model samples. FLDA were obtained the best identification feature volume of samples. This feature volume is more conducive to the model of classification [1].

The traditional arithmetic, such as PCA and FLDA directly operates on the original image and takes the overall features of the image which is easily affected by facial expression, posture and changes in light conditions. And because of the singular value characteristics of the matrix has a good stability, proportion of invariance and rotation invariance, and other nature, so it has been widely used in many areas such as data compression, signal processing and pattern analysis. Hong Ziquan and Yang Jingyu put singular value characteristics of the image to apply to the face recognition for the first time and achieved good results. Since then, the human being put forward a number of the face image feature algebra extraction methods based on singular value decomposition (SVD) to increase the rate of face recognition. For example, a researcher at the SVD algorithm based on a good face image feature extraction and recognition. But article [2] studies have show that above described in its image recognition only under certain constraints within the framework of an effective (such as cameras and face the relative location, direction and attitude can not be a significant change), because this image feature extraction method does not have the image of the translation, rotation and size of the non-sensitive , and thus do not have a wide range of applications[2][3][6]. In addition, the article [7, 11] studies have shown that it can not get enough information required for face recognition through singular value decomposition on only a single scale of the whole facial image, it is necessary to extract more features. The experts like Du Gan divide the face image into five parts such as the upper half, the second half, eyes, nose and mouth. extract singular value features respectively, then mix together for face recognition to increase recognition rate. However, in face recognition, on the one hand they didn’t use the type information of image samples, and it’s not the most effective from classification point of view. On the other hand, face only divided into 5 parts, and each part extracted the local characteristics of singular value independently to be similar with the whole human face. It only strengthened the local characteristics of the eyes, nose and mouth, but the facial contours, ears, eyebrows, and other local features are still not taken seriously. There will be still effective local features information which has not been extracted. In addition, when posture changes, the acquirement of five parts is more difficult. Furthermore, each part of the weight can't be determined by the non-identified method. It's randomness, the recognition rate is still not high, and the algorithm is more complex[7][8][9][10][11].

This paper put forward the LDA face recognition algorithm (multi-scale SVD linear discriminant analysis, MSVD + LDA) based on multi-scale singular value characteristics. Make multi-scale division on the image first, give the image and each division an SVD respectively to get its singular value features, combine together as the multi-scale singular value feature vector of image, gather such a characteristics vector, and then apply FLDA methods.

## 2 The Extraction of Human Face

The amount of original data in general is considerable from the images obtained directly. For example, a character image can have thousands of data, the amount of data of a remote sensing satellite images is greater. In order to identify and classify effectively, it is necessary choose or change the raw data to get the feature which can respond the classified nature best. This can constitute the feature vector. It's called process of feature extraction. Generally, we call the space composed of the raw data the measurement space. The feature space is called the space which carried out to classify and identify. Through transforming, it can change a pattern which expresses a measurement space of higher dimension to a pattern which expresses a space of lower dimension. A pattern is often called a sample in the feature space, which can often be expressed as a vector, that is, a point in the feature space.

The feature extraction is often divided into three steps: feature formation, feature extraction, and feature selection.

(1) Feature formation

According to the facial images can generate a set of basic features, and such features are known as the original features. Identifying the cells as an example, through inputting the images to get the digital images of the cells. According to the digital images of the cells, it can calculate with a total area of the cells, a total optical density, a proportion of nuclear plasma, and so on. These values is just the original features

(2) Feature extraction

The number of original features may be great. If all of the original features are sent to the classifier as the classified features, it can not only make the classifier complex, and make the discrimination number of the classified calculate large, but also the error probability is not necessarily small. Hence it need to reduce the number of features. Through the method of mapping (transform), change the feature vector of high-dimensional to the feature vector of low-dimensional. This process is known as the feature extraction. The characteristics after mapping is also known as the secondary features, which are one kind of combination of original feature. So the feature extraction is a feature mapping in the broad sense. It is expressed in formula as:

$$A: Y \to X$$

Thereinto $Y$ is measurement space, $X$ is feature space, $A$ is called feature mapping or feature classifier.

(3)Feature selection

Select the most effective feature which the number is $d(d < D)$ from a group of original features which the number is $D$. This process is known as feature selection. It can the number as (From the number of group D of the original features of a few selected It was d (d <a gong the best features of this process, known as the feature selection, which can play a role to reduce the feature space dimension.

This article uses the method of feature extraction which based on the singular value decomposition. It is an extraction method based on the overall features[12][13].

## 3 Singular Value Decomposition

Singular value decomposition (SVD) is an effective tool of the least-squares to solve the problem which is applied widely in data compression, signal processing, pattern recognition and so many ways. The use of matrix singular value decomposition characteristics of the extract is stable, shift invariance, transpose invariance and rotation invariance. Based on the SVD feature extraction methods which have got the face image of algebra features, not only weakened the impact of light and expression, but also reduced the dimensions and the computational complexity, and at the same time it retains most of the effective characteristics of a human face images, that is in order to provide a good basis to identify follow-up process. Its theorem and characterization are as follows:

If the matrix $A_{n\times r}$ of $n\times r$ dimension which the order is r expresses an image, it can exist two orthogonal matrixes:

$$U=[u_0,u_1,...,u_{r-1}]\in\Re^{n\times r} \qquad U^TU=I \qquad (1)$$

$$V=[v_0,v_1,...,v_{r-1}]\in\Re^{n\times r} \qquad V^TV=I \qquad (2)$$

and diagonal matrices

$$\Lambda=diag[\lambda_0,\lambda_1,...,\lambda_{r-1}]\in\Re^{r\times r} \qquad (3)$$

and

$$\lambda_0\geq\lambda_1\geq...\geq\lambda_{r-1}$$

appeasing

$$A=U\Lambda^{1/2}V^T \qquad (4)$$

Thereint $\lambda_i(i=0,1,...,r-1)$ is the non-zero eigenvalue of the matrix $AA^T$ and $A^TA$, $u_i$ and $v_i$ is $AA^T$ and $A^TA$ corresponding to $\lambda_i$ the eigenvector respectively. The decomposition above is called singular value decomposition of the matrix A(SVD for short). $\sqrt{\lambda_i}$ is the singular value of A. $[\lambda_0,\lambda_1,...,\lambda_{r-1}]^T$ is called the singular value vector of the matrix A. In the absence of special note, this article mentioned the back singular value of the vector have all expressed this sense.

Deducing

$$U=AV\Lambda^{1/2} \qquad (5)$$

Because it is the only singular value decomposition to any real matrix A, when it is arranged in $\lambda_0\geq\lambda_1\geq...\geq\lambda_{r-1}$, the original image A corresponds to the only one singular value vector. Thereupon singular value vector can be used to describe a numerical characteristics of the gray value matrix A. Singular value vector has a good invariance of algebra and geometry.

(1) Stability

To the description of the characteristics of the image, when the image gray becomes small , the changes of characteristics are not obvious, that is called stability. This feature has broadened the requirement of the pre-image. As the singular value vector has a good stability, Therefore, its image in different light conditions caused by the transformation does not have a sensitive character.

(2) Shift invariance

The translation of image transformation equivalent to the replacement to the image matrix for the line (or row). That is, to elementarily transform the image matrix for the exchange of the two lines (or rows). The two lines $i$ and $j$, in the exchange matrix A are equivalent to the left side of this matrix multiplies the matrix $I_{ij}$. Therefore, the original image A have the same singular value vector as the image $I_{ij}A$ after exchanging two lines. By the same token, out of the replacement has the same result. As a result, singular value vector has a shift invariance.

(3) Transpose invariance

If we make a transpose operation on the image matrix , and the singular value feature vector does not change, according to the theorem of singular value decomposition

There is :

$$AA^Tu_i=\lambda^2u_i \qquad (6)$$

$$A^TAv_i=\lambda^2v_i \qquad (7)$$

It can be seen that $A$ and $A^T$ has the same singular value, which corresponds to the same singular value vector. For the image feature extraction is concerned, it often requires the features taken with some invariance on the algebra and geometry.

(4) Rotation invariance

If making a rotation operation on the image, the singular value feature vector does not change.

(5)Singular value vector with the corresponding image brightness changes in the proportional

When the whole image of proportion to the brightness changes, the singular value vector is also proportional to change, and this change does not transform the identification information contained. So when using singular value vector to identify, adopt only simple which could eliminate the influence of the ratio coefficient. When the brightness of the image A is proportional to change (scale factor $\alpha$), it's in the equivalent of singular value to a high proportion of vector $|\alpha|$.

Supposing the gray-scale matrix of original image is A, making changes in the ratio to the gray-scale image A, that is, it can get an image matrix $\alpha A$ after multiplying a non-zero real number $\alpha$ by the matrix A. As $rank(A)=rank(\alpha A)$ , if $rank(A)=k$, moreover singular value of A and $\alpha A$ is $\lambda_0,\lambda_1,...,\lambda_k$ and $\delta_1,...,\delta_k$ respectively, thus the characteristics of a variance $(\alpha A)(\alpha A)^T$ is:

$$|(\alpha A)(\alpha A)^T-\delta^2I|=0 \qquad (8)$$

Namely

$$|AA^T-\frac{1}{\alpha^2}\delta^2I|=0 \qquad (9)$$

From this ,it can get

$$(\delta_1,...,\delta_k,0,...,0)^T=|\alpha|(\lambda_1,...,\lambda_k,0,...,0)^T \quad (10)$$

By this token, the brightness of the image A is proportional to change(scale factor), it's in the equivalent of singular value to a high proportion of vector .

The character above is the theoretical basis of the extraction of features on A, it is to ensure the invariance on algebra and geometry of the extraction characteristics on A. Therefore it can be certain that, the characteristics from SVD extraction is an ideal characteristics of algebra[14][15].

# 4 Fisher Linear Discrimination Analysis

## 4.1 Fisher Linear Discrimination Principle

Fisher linear discrimination can put the samples of d dimensional space on a straight line ,and form one dimensional space. Under normal circumstances, can always find a direction that the samples are separated each other after projecting on the straight line. Finding the projection line is a question to be resolved by Fisher.

Supposing have N training samples xi in w1/w2 two types of problem, thereinto, $N_1$ samples belong to type $w_1$, and $N_2$ samples belong to type $w_2$, the subset of training samples $X_1$ and $X_2$ are composed of $N_1$ and $N_2$ respectively, if

$$y_i = w^T x_i, i = 1,2,\cdots,N \tag{11}$$

$y_i$ is a scalar quantity through transforming $w$ by $x_i$. Actually, $y_i$ is a value of discriminant for $w$. The two subsets $Y_1$ and $Y_2$ are formed through projecting for the samples of $X_1$ and $X_2$. If $\| w \| = 1$, so $y_i$ is the projection of $x_i$ on $w$ direction, furthermore, $w$ direction is the normal direction to differentiate hyperplane.

The following studies how to gain the analysis formula of $w$ direction best. If

$$m_i = \frac{1}{N_i}\sum_{x_j \in X_i} x_j, i = 1,2 \tag{12}$$

$m_i$ is the mean vector of sample in d dimensional feature space. Kinds of averages are gained by the following formula through transforming $w$ to feature space,

$$u_i = \frac{1}{N_i}\sum_{y_j \in Y_i} y_j, i = 1,2 \tag{13}$$

After the projection, the within-class scatter matrices of kinds of samples is defined as follows,

$$S_i = \sum_{y_j \in F_1}(y_j - u_i)^2, i = 1,2 \tag{14}$$

The idea of fisher is the distance of two types , so Fisher criterion is defined as following formula,

$$J_F(w) = \frac{|u_1 - u_2|^2}{S_1^{\,2} + S_2^{\,2}} \tag{15}$$

$w^*$ is the maximal analytical solution of $J_F$, and the best solution vector, namely the linear discriminant of Fisher. Following is the solution of extremely large value. Changing formula (14) into the obvious function of $w$ first, and following formula can be gained,

$$u_i = \frac{1}{N_i}\sum_{y_j \in X_i} w^T x_j = w^T\left(\frac{1}{N_i}\sum_{x_j \in X_i} x_j\right) = w^T m, i = 1,2 \tag{16}$$

So

$$\begin{aligned} |u_1 - u_2|^2 &= \| w^T m_1 - w^T m_2 \|^2 = \| w^T (m_1 - m_2) \|^2 \\ &= w^T (m_1 - m_2)(m_1 - m_2)^T w \\ &= w^T S_b w \end{aligned} \tag{17}$$

In formula,

$$S_b = (m_1 - m_2)(m_1 - m_2)^T \tag{18}$$

$S_b$ is the within-class scatter matrices of samples in d dimensional feature space originally, and indicates the scatter size between two types of mean value vector. $S_b$ is bigger, the easier to differentiate.

$$\begin{aligned} S_i^2 &= \sum_{x_j \in X_i} (w^T x_j - w^T m_i)^2 = w^T \sum_{x_j \in X_i} (x_j - m_1)(x_j - m_2)^T w \\ &= w^T S_i w, i = 1,2 \end{aligned} \tag{19}$$

$$S_1^{\,2} + S_2^{\,2} = w^T (S_1 + S_2) w = w^T S_w w, i = 1,2 \tag{20}$$

In formula,

$$S_w = (S_1 + S_2) \tag{21}$$

$S_i$ is the within-class scatter matrices of samples in d dimensional feature space originally, and that $S_w$ is the within-class general scatter matrices of samples. the within-class scatter is smaller, the easier to classify. The obvious function of $J_F(w)$ can be gained through putting formula (11) and (15) into formula (13).

$$J_F(w) = \frac{|u_1 - u_2|^2}{S_1^{\,2} + S_2^{\,2}} = \frac{w^T S_b w}{w^T S_w w} \tag{22}$$

When making $J_F(w)$ get the maximal value, it can use Lagrange method of multiplier to solve $w$. If denominator is a nonzero constant, namely, Defining Lagrange function as following

$$L(w, \lambda) = w^T S_b w - \lambda (w^T S_w w - c) \tag{23}$$

$\lambda$ is a Lagrange multiplier in formula. Solve partial derivative to $w$, according to formula (23).

$$\frac{\partial L(w,\lambda)}{\partial w} = S_b w - \lambda S_w w$$

If partial derivative is zero, so

$$S_b w^* - \lambda S_w w^* = 0$$

Namely

$$S_b w^* = \lambda S_w w^* \quad (24)$$

$w^*$ is the extreme solution of $J_F(w)$, in other words, $w^*$ is the best projection direction from d dimensional space to one dimensional space. D dimensional sample $x_i$ ( $wi = 1,2,...N$ ) can be projected into one dimensional space using $w^*$ and formula (11). The $w^*$ that satisfies formula (23) has several different values because formula (23) is a generalized eigenvalue problem. Above conclusion is deduced based on two types of question. Fisher linear discriminant can be extended to C-1 discriminant function, thus forming the projection from d dimensional space to C-1 dimensional space. Obviously, if d>C, the within-class scatter matrices of samples can be extended to following formula:

$$S_w = \sum_{i=1}^{C} S_i \quad (25)$$

## 4.2 Fisher Linear Discrimination Extend to C types

Suppose original image $x$ is n dimensional vector and the samples have C types(have C face image samples) $w_1, w_2, ...w_c$ ,so the within-class scatter matrices of samples $S_w$, the extra-class scatter matrices of samples $S_b$ and the within-class general scatter matrices of samples $S_t$ are following formulas:

$$S_w = \sum_{i=1}^{C} P(w_i) E[(x-m_i)(x-m_i)^T / w_i] \quad (26)$$

$$S_b = \sum_{i=1}^{C} P(w_i)(m_i - m)(m_i - m)^T \quad (27)$$

$$S_t = E[(x-m)(x-m)^T] = S_w + S_b \quad (28)$$

Thereinto: $P(w_i)$ is the prior probability about the i type of samples, the mean vector about the i type of samples is $m_i = E(x / w_i)$ ( $i = 1,2,...,C$ ), and $m_i = E(x)$ is mean vector of general samples.

$$m = \sum_{i=1}^{C} p(w_i) m_i \quad (29)$$

Suppose the i type has $n_i$ samples, and all samples are recorded as $x_{ij}\,(j = 1,2,...n_i)$, so the evaluated formulas of $S_w$ and $S_b$ as following:

$$S_w = \sum_{i=1}^{C} P(w_i) \frac{1}{n_i} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)^T \quad (30)$$

$$S_b = \sum_{i=1}^{C} P(w_i)(\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T \quad (31)$$

$x_i$ and $x$ are the evaluation of $m_i$ and $m$ respectively:

$$m_i = \bar{x}_i = \frac{1}{n_i} \sum_{i=1}^{n_i} x_{ij} \quad (32)$$

$$m = \bar{x} = \sum_{i=1}^{n_i} x_{ij}\, p(w_i) \bar{x}_i \quad (33)$$

Here, the prior probability about the i type of samples is following formula:

$$p(w_i) = 1/C \quad (34)$$

The discrimination criterion function of Fisher is defined as following

$$J_F(w) = \frac{|w^T S_b w|}{|w^T S_w w|} \quad (35)$$

$w$ is the best discrimination direction when formula (35) get maximum value. Its physical meaning is to make a sample set in the projection of $w$ direction have the smallest within-class scatter and largest extra-class scatter. in other words, the same type of samples close to as much as possible through projecting, and that different types of samples as much as possible to separate[16][17].

When making $J_F(w)$ get the maximal value, it can use following method to solve $w^*$. $J_F(w)$ is a generalized Rayleigh quotient in formula (35), it can be solved by Lagrange multiplier. If denominator is nonzero constant, namely $w^T S_w w = c \neq 0$ and the definition of Lagrange is following :

$$L(w,\lambda) = w^T S_b w - \lambda(w^T S_w w - c) \quad (36)$$

$\lambda$ is Lagrange multiplier in formula (36). Solve partial derivative to $w$, according to formula (36),

$$\frac{\partial L(w,\lambda)}{\partial w} = S_b w - \lambda S_w w$$

If partial derivative is zero, it can gain following result

$$S_b w^* - \lambda S_w w^* = 0$$

Namely

$$S_b w^* = \lambda S_w w^* \qquad (37)$$

$w^*$ is the value of $w$ when formula (36) get the maximum value. Following formula can be gained when multiplying $S_w^{-1}$ on both sides of the left and $S_w$ is non-singular:

$$S_w^{-1} S_b w^* = \lambda w^* \qquad (38)$$

The discriminant function of best LDA as following:

$$d_i(x) w^{*T}(x - m_i) \qquad (39)$$

# 5 Multi-scale singular value characteristics of the face recognition algorithm

## 5.1 Algorithm Analysis and Improvement

Analysis method has been found that there are two reasons for low recognition rate. One is face recognition image has often been effected by light, facial expressions, gestures. Based on SVD of the whole image overall operates on the original image pixels directly, taking the global features of images. In fact when facial expression and the light change, only some of the face obviously changes in the region, some of the other changes a little, even without change. But according to the gray value of image pixel by taking the overall situation will significantly strengthen some of the changes, instead of neglecting natural parts which can reflect information of other types of images. It will lose a lot of valid identification information. Second, after extracting singular value features of the single measure about the whole image, many local special features information is abnegated on corresponding to zero space (zero singular value corresponding to the characteristics of space), so that to reduce the identification ability. Tian etc. also permit we can not get enough required information for face recognition on the whole face images of only a single scale[18].

As a result, we should need to improve the extraction of feature vectors. First of all, divide into multi-scale image. Make SVD on the image and each division respectively to get their singular value. Then get all the singular value together as the multi-scale image of singular value features vector. But the singular value features haven't used the type information of image sample. Therefore from the classification point of view, using singular value is not the most effective features, it needs to identify these eigenvectors using LDA method. It is called the MSVD + LDA algorithm.

Multi-scale singular value feature vector not only reflects the overall image of singular value features, but also reflects the local singular characteristics of the multi-scale image. It can reduce partial loss of information to identify effectively under single-scale, and reduce the impact of local light, noise, attitude of changes, then to improve the algorithm of generalization ability. Compared with Du Gan's methods, we don't need to pre-orientate the human eye to reduce the complexity of the algorithm. As more small-scale covers various parts of the face, the more extraction of local information, the more capacity to adapt the attitude of the changes, robustness is better.

## 5.2 Multi-scale Extraction of The Singular Value Characteristics

Divide an face image of multi-scale from overall to local, it will get different scales of the subimage $F_{i,j}$ . In which $i = 1,2,...,d$ expresses split scale, $j = 1,...,4^{i-1}$ expresses the number of fast for each of scale. Due to the limited size of the image, segmentation of the scale is limited

Segmentation method is just to see Figure 1, from left to right are 1 to 3 scale of the partition respectively. The overall image is called the No.1-scale of the subimage $F_{1,1}$, then the whole image will be sub-divided into 4 pieces $F_{2,1}, F_{2,2}, F_{2,3}, F_{2,4}$, Each subimage is called the No. 2-scale subimage, and then divide the No. 2-scale of the subimage, set up to the largest-scale so far.

(a) 第 1 尺度 (b) 第 2 尺度 (c) 第 3 尺度

Figure 1 multiscale segmentation of human face image

Multi-scale singular value to the specific feature extraction algorithm is as follows.

1) Solve the entire image $F_t$ first. Namely the singular value eigenvector $p_{1,1}, f_t = (p_{1,1})$ of subimage $F_{1,1}$.

2) Calculate from 1 to i-scale combination of multi-scale singular value eigenvector $f_t(i = 2,...,d)$.

For $j = 1,2,...,num,(num = 4^{i-2})$

a) Divide the (i-1)-scale image of each sub-block $F_{i-1,j}$, we can get four the i-scale subimage $F_{i,s+1}, F_{i,s+2}, F_{i,s+3}, F_{i,s+4}$, in which $s = 4\times(j-1)$. According to the theorem 1 , make SVD on i-scale subimage. Then calculate and get its singular value feature vector. They are: $p_{i,s+1}, p_{i,s+2}, p_{i,s+3}, p_{i,s+4}$.

b) $f_t = (f_t, p_{i,s+1}, p_{i,s+2}, p_{i,s+3}, p_{i,s+4})$ (40)

3) Change the form of vector: $f_t = f_t^T$ , that is

$$f_t = (p_{1,1}, p_{2,1},..., p_{d,k})^T, (k=4^{d-1}) \quad (41)$$

$f_t$ is the multi-scale singular value feature vector of image $F_t$. Because $f_t$ contains the singular value of images from the local to the overall features, it reflects some commonness of the overall and local images. So take this as a sample and then LDA. This is better than in the entire single-scale features on the classification of singular value, it is also better than the classic FLDA which directly puts up in the original image pixels[19].

### 5.3 Identification and Sample of The Characteristics of The Samples and Extraction.

The feature extraction procession of MSVD+LDA is as follows.

1) Receive the multi-scale singular value feature vectors of each training samples, compose the feature vector set $\{f_i^{(j)}\}$ of training samples. $f_i^{(j)}$ means the multi-scale singular value feature vector of the $j$ sample in type $i$ .

2) Give $y_i^{(j)} = H^T f_i^{(j)}$ transformation of K-L to all the samples on $\{f_i^{(j)}\}$ . Compress the dimension of the original samples from high to low. $Y = H^T F, H = (h_1,...,h_{N-c}), h_1,...,h_{N-c}$ are the feature vectors corresponding to N-c of the largest eigenvalue which $\{f_i^{(j)}\}$ general scatter the matrix $S_t$.

3) In the space $Y$ after transferring, it can use the method FLDA to extract the features. $\tilde{S}_b = H^T S_b H$ 、 $\tilde{S}_w = H^T S_w H$ express the extra-class scatter martrix and within-class scatter matrix of samples after dimension respectively. $S_b$ and $S_w$ express the extra-class scatter martrix and within-class scatter matrix of $\{f_i^{(j)}\}$ respectively. $S_w$ is non-singular, take d maximal feature values of $S_b$ and $S_w$ corresponding to the feature vector $v_1,...,v_d$ as a projection vector, that is to say $\tilde{S}_b v_j = \lambda_j \tilde{S}_w v_j$ , thereinto $\lambda_1 \ge \lambda_2 \ge ... \ge \lambda_d$ . Make $V = (v_1, v_2,..., v_d)$, use $Z = V^T Y$ to extract the feature to get the identification features of dimensional.

when recognizing, to one recognition image x, it should receive its multi-scale singular value feature vector $f$ first. By $b = V^T H^T f$, extract its identify feature vector $b$ which has been taken, each image x corresponds to one $b$. Then, use the minimum distance classifier to classify this identification feature vector. Suppose the average value of the No.i training image sample of identification feature vector is $\bar{\xi}_i, i = 1,2,...,c$ . Calculate $d(\bar{\xi}_i, b) = \| \bar{\xi}_i - b \|_2, i = 1,2,...,c$ to the test sample x. If $d(\bar{\xi}_i, b) = \min_i d(\bar{\xi}_i, b)$ , so $x \in \omega$.

## 6 Experiment Result and Analysis

Execute the experiments by utilizing the ORL(O to Rhino Laryngology) human face database and through face recognition using the minimum distance classification method. ORL face database of 400 images (40 people, 10 pictures per person). Among them, some images were taken at different periods; human face expressions and other details in the face varied to some degree. For example, laughing or not laughing, eyes open or closed, with or without glasses; Facial posture also changes somehow, it can be up to 20 degree of either plane rotation change or in depth rotation change. The dimension change on the face can be up to 10% as well. Figure 2 is the partial face images from ORL, and Ficture 3 is a series of 10 pictures of one person from ORL human face database[20].

According to article [12], there is little impact on test accuracy by reducing picture size, human face picture is still in good identification while pixel size reduced. The original picture can be reduced to 64×64 by using the insert methodology for the benefit of multi-size division of picture. To achieve the best result of picture division, multiple methods can be applied MSVD + LDA/d, d=1, ….6. For example, the MSVD + LDA/5 represents the algorithm of 1-5.

Figure 2 partial ORL face image

Figure 3 the first human face image in ORL

All samples of ORL are used in the experiment to assure the repeatability of the test, experiment result is comparable, sample selection is based on the specific 4 models(for per person): prior 5 pieces training, post 5 pieces test; post 5 pieces training, prior 5 pieces test; odd number training, even number test, even number training, odd number test. There are 200 test samples and training samples for each sample size selection. The final result is the average of test results of 4 sampling selection. Experiment result is shown in table1.

It can be seen from Table 1, with the increasing of the scale, the rate of recognition also increased, but when the multi-scale singular value of the feature vector dimension closed to the original image， the recognition rate is no longer increased. The original image on the division reaches a certain level can not be the higher the better, and this was provides a basis for how to determine the Division-level scale , as this study was carried out as long as the standard 5 divisions. In the experiment, the feature vector dimension of M SVD + LDA/5 is 1 984, and in the contrast, the original image of the feature vector dimension of the LDA is 4 096; As a result, multi-scale singular value feature extraction is essentially an effective dimension Reduction.

Table 1 experiment results in ORL face database

| 算法 | 识别率 | 鉴别特征维数 | 特征向量维数 |
|---|---|---|---|
| PCA+LDA | 91.8% | 39 | 4096 |
| SVD | 39.38% | 64 | 64 |
| SVD+LDA | 49.63% | 39 | 64 |
| MSVD+LDA/2 | 48.63% | 39 | 192 |
| MSVD+LDA/3 | 90.88% | 39 | 448 |
| MSVD+LDA/4 | 96.75% | 39 | 960 |
| MSVD+LDA/5 | 97.38% | 39 | 1984 |
| MSVD+LDA/6 | 96% | 39 | 4032 |

Table 2 Recoginition rate between the method of this paper and SVD

| 序号 | 算法来源 | 识别率 |
|---|---|---|
| 1 | Article[5] | 81.6% |
| 2 | Article[6] | 96.5% |
| 3 | Article[7] | 83.7% |
| 4 | Article[8] | 90.8% |
| 5 | Article[9] | 96.5% |
| 6 | Article[10] | 94.2% |
| 7 | This paper | 97.38 |

ORL face database of human faces and gestures scale to a certain extent there changes. Table 1 shows, FLDA recognition algorithm only 91.88 percent, but the M SVD + LDA/5 recognition of the high rate of up to 97.38 percent. That means the singular value-based multi-scale features of the image rotation and scale changes in a better adaptability.

And compared with the singular value vector of the face recognition algorithm, this article proposed by the algorithm to identify a higher rate, as shown in table 2.

Article [5-9] to only use the face of global information, article [10] also only uses some of the organs of local information, but this text uses the multi-scale singular value compositions of an image from the local to the whole, the images can be taken into account most of the local characteristics. And these local features can reflect the differences between the images better, and conducive to the classification model more.

## 7 Conclusion

Singular value vector of an image is a valid feature for identification. In this paper, we bring up an idea that combinate multi-scale singular value vector and then application of face recognition algorithm LDA. Its highlighted advantage is local features of images in multi-scale can be took out and they are better to reflect the differences between the images and identificate features in a comprehensive manner. At the same time ,we use the algorithm named LDA with good ability of classification which is better to recognize the mode. Face Recognition rate show in experiments were made with ORL human face image databases is greatly enhanced . In this sense, we understand that more effective facial feature can be took out with a method named SVD+ LDA. This method can be used to recognize other images without face

*References:*

[1] Belhumeur P N , Hespanha J P, Kriegman D. Eigenfaces vs Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), 1997, pp.711-720.

[2] HONG Ziquan. A lgebraic feature extraction of image for recognition. *Pattern Recognition*, 24(3) , 1991 , pp.211 - 219.

[3] HONG Ziquan, YANG Jingyu. Algebraic feature extraction of image for recognition. *Acta Automatic Sinica*, 18(2), 1992 : 233 - 237.

[4] HONG Ziquan, YANG Jingyu. Human facial image recognition algorithm based on singular value features and statistical model. *Computer Research and Development*, 31(3), 1994, pp.60-65.

[5] WANG Yunhong, TAN Tieniu, ZHU Yong. Face identification based on singular decomposition and data fusion. *Chinese Journal of Computer*, 23(6), 2000, pp.649-653.

[6] LIU Xiao jun, WANG Dongfeng, ZHANG Lifei, et al. An approach for face recognition based on singular value decomposition and hidden Markov model. *Chinese Journal of Computer*, 26(3), 2003, pp. 340-344.

[7] TIAN Yuan, TAN Tieniu, WANG Yunhong, et al. Do singular values contain adequate information for face recognition. *Pattern Recognition*, 36(6), 2003, pp.649-655

[8] GAN Junying, ZHANG Youwei. A new approach for face recognition based on singular value features and neural networks. *Acta Electronica Sinica*, 2004, 32(1), pp.170-173.

[9] WANG Wensheng, CHEN Fubing, YANG Jingyu. A method of feature extrxtion based on SVD. *Journal of Electronics* & *Information Technology* , 2005, 27(2), pp.294-297.

[10] DU Gan, ZHU Wenjun. Face recognition method based on singular value decomposition and fuzzy decision. *Journal of Image and Graphics*,2006, 11(10), pp.1456-1459.

[11] GAO Quanxue, LIANG Yan, PAN Quan, et al. The problem existed in face recognition using SVD and its solution . *Journal of Image and Grap hics*, 2006,11(12), pp.1784 -1791.

[12] JIN Zong, YANG Jingyu, HU Zhongshan, et al. Face recognition based on uncorrelated discriminant transformation. *Pattern Recognition*, 2001, 34(7), pp.1405 - 1416.

[13] Cheng Jun Liu , Harry Wechsler. A shape and texture based enhanced Fisher classifier for face recognition. *IEEE Transactions on Image Processing* , 2001 , 10 (4), pp.598-608

[14] Yang Jian , Tu QingHua , Yang JingYu. Fast Foley Sammon transform and face identify-cation. *Chinese Journal of Image and Graphics* , 2002, 7(1), pp.1-5

[15] BOUL GOURIS N V, TZOVARAS D, STRINNTZIS M G. Lossless image compre-ssion based on optimal prediction, adaptive lifting, and conditional arithmetic coding. *IEEE Transactions on Image Processing*, 2001,10(1), pp.1-14.

[16] MARTIN M B. New image compression techniques using multiwavelets and multi-wavelet packets. *IEEE Transactions on Image Processing*, 2001,10(4), pp.500 - 511.

[17] Moghaddam B. Probabilistic visual learning for object representation . *Proc.IEEE Conf. Computer vision and pattern recognition*，1998, pp.38-44.

[18] M .Nixon. Qualitative detection of motion by a moving observer In Proc. *DARPA Image Understanding Workshop*, 1990, pp.329-338

[19] Miao I YN BC, Wang K Q et al. A Hierarch-ica

IM ultiscale and Multiangle System for Human Face Detection in a Complex Back- ground Using Gravity-Center Template. *Pattern Recongnition*, 1999, 32(7), pp.1237-1248

[20] Belhumeur P N, Hespanha J P, Kriegman D J. Eigenfaces vs Fisherfaces Recognition Using Class Specific Linear Projection. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 1997(19), pp.711-720.