Facial Expression Recognition based on Independent Component Analysis

XiaoHui Guo
China Tobacco GuangXi Industrial CO., LTD

Xiao Zhang, Chao Deng, and Jianyu Wei
China Tobacco GuangXi Industrial CO., LTD

BeiHang University, School of Instrumentation Science and Opto-electronics Engineering

Abstract—As an important part of artificial intelligence and pattern recognition, facial expression recognition has drawn much attention recently and numerous methods have been proposed. Feature extraction is the most important part which directly affects the final recognition results. Independent component analysis (ICA) is a subspace analysis method, which is also a novel statistical technique in signal processing and machine learning that aims at finding linear projections of the data that maximize their mutual independence. In this paper, we introduce the basic theory of ICA algorithm in detail and then present the process of facial expression recognition based on ICA model. Finally, we use PCA and ICA algorithm to extract facial features, and then SVM classifier is used for facial expression recognition. Experimental results show ICA is a real effective facial expression recognition method and the recognition rate based on ICA is greater than based on PCA and 2DPCA.

Index Terms—Artificial Intelligence; Pattern Recognition; Facial Expression Recognition; Independent Component Analysis; PCA; SVM

I. INTRODUCTION

Facial expression recognition is short for computer Automatic Facial Expression Recognition is an important part of the artificial psychology theory and computer vision research. It refers to the feature extraction of facial expressions using the computer to be understood, according to human cognition and ways of thinking, classified, and then analyzes the discriminant emotions, such as disgust, surprise, anger, fear, happy, sad and so on. With the rapid development of computer vision technology, facial expression recognition has also been obtained more and more attention, and there are a lot of applications such as human-computer interaction, cognitive science and psychology and so on [1-3].

After decades of continuous research and development, facial expression recognition technology has achieved very important achievements. Face detection, expression feature extraction, facial expression classification and other aspects have a lot of mature and effective algorithms. And they have promoted related disciplines to be rapid development. However, we still see that the field still has a lot of problems to be solved. While humans have a strong ability to identify facial expressions, but the computer achieve quite easy. The main reason is because: First, the human face is a flexible body rather than rigid, hard to facial muscle movement to establish a consistent expression model. Different people different because of age, race, face and facial features, showing the same kind of expression, there will be a considerable difference. This increased the difficulty of expression modeling. Secondly, the expression of the face is a dynamic process, before and after a great deal of relevance in the process. If only extract the expression of a particular time for facial expression recognition, it will lose a lot of information. Finally, expression feature extraction is very difficult because of a wide variety of facial expressions and subtle changes in complex.

Facial expression recognition is an attractive field which is related to image processing, pattern recognition, movement tracking, physiology, psychology and so on. The facial expression recognition system mainly contains three parts: face detection, feature extraction and classification. Among them, feature extraction is the most important part which directly affects the final recognition results [4-5]. Therefore, expression feature extraction is a key step for facial expression recognition. Early facial feature research is mainly based on the geometric features [6]. The basic idea is to use a number of feature points of the person's face, the relative position and the relative distance, and then supplemented with the shape information of the face contour. But its biggest drawback is the recognition accuracy is totally dependent on the extraction of geometric features, these geometric features extraction are very sensitive to changes in illumination, facial expression, gesture, so stability is not high, low recognition rate. Recently, many extraction methods are proposed which are based on statistical characteristics. For example, template matching method [7] which trains these face images in the database as a template, the experimental results show that the performance is significantly superior to the method based on the geometric characteristics.

In face analysis, the dimension of the face image is usually very high, but the distribution of the face image in such a high dimensional space is nosebleed. Therefore, the high dimension is not conducive to the category,
what’s more, on the degree of complexity in the calculation is also very and very large. In order to reduce the dimension and extract facial feature, Kirby et al. firstly used the subspace analysis method to realize face recognition [8-9], while this method had been obtained greater success and this subspace analysis method is widely used in recent years. This method then has aroused more and more attention, and thus become one of the mainstream methods of face recognition. The main idea of subspace analysis method is based on certain performance targets to find a linear or non-linear space transform, the original signal data compression to a low-dimensional subspace and then to obtain the more compact distribution of the data in the subspace, which can describe a means of describing data subspace and reduce computational complexity. Recently, these subspace analysis methods used for facial feature extraction including as following: principal component analysis (PCA) [8-9], independent component analysis (ICA) [10], linear discriminant analysis (LDA) [11-12], and non-negative matrix factorization (NMF).

In this paper, we mainly study the facial expression recognition method based on independent component analysis, this is a subspace analysis method. Independent meta-analysis of statistics based on all bands in the solution on principal component analysis and linear discriminant analysis. ICA is as a novel statistical technique in signal processing and machine learning that aims at finding linear projections of the data that maximize their mutual independence. Firstly, we introduce the basic theory of ICA algorithm in detail and then present the process of facial expression recognition based on ICA model. Finally, we use PCA, 2DPCA and ICA algorithm to implement facial expression recognition. Experimental results show ICA is a very effective facial expression recognition method and the recognition rate based on ICA is higher than that of based on PCA and 2DPCA.

II. EXPRESSION FEATURE EXTRACTION

Facial expression recognition is to analysis a particular state for these given expression from facial images and video sequences, and to determine the objects’ psychological and emotional state. Facial expression recognition includes the following steps: image acquisition, image preprocessing, face detection, facial feature extraction, classification and recognition. Fig. 1 shows facial expression recognition system diagram.

For the nature of image processing, feature extraction methods can be divided into two categories expression: expression based on a static image feature extraction and dynamic expression video sequences based feature extraction. The former deals with single-frame still face images, the image reflects the general requirements in the expression or greatly exaggerated the state, making the extraction of features more typical expression of such methods include principal component analysis, singular value decomposition and wavelet-based methods. The latter method is used for extracting the characteristic changing process of facial expression.

PCA and 2DPCA, which are also subspace analysis method. PCA is one of widely applied technologies, which has been used in image compression and facial feature extraction. The goal of PCA is to decrease the high dimensional data space into low dimensional feature space. But in face recognition based on PCA, it needs to transform 2D image into 1D image, in order to overcome the flaw, two-dimensional principal component analysis (2DPCA) is proposed to rebuilt image and feature extraction. 2DPCA is based on the image matrix, it does not transform 2D image into 1D image, so it is very simple and more straightforward applied for features extraction. Therefore, we first introduce these two common feature extraction methods.

![Facial expression recognition system diagram](image)

Figure 1. Facial expression recognition system diagram

A. Principal Component Analysis (PCA)

The main idea of PCA is originated from the K-L transform theory, and it is to look for an optimal set of unit which is orthogonal vector-based linear transformation, and with a linear combination of them to reconstruct the original sampling set, to minimize the error between the reconstructed samples and the original samples. PCA is the solution of the eigenvalue problem to obtain the diagonal covariance matrix:

$$S = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})^T$$

(1)

$$\lambda \mathbf{w} = S \mathbf{w}$$

(2)

where $N$ is the total number of samples, $\bar{x}$ represents the mean sample form all sampling set. Calculating the eigenvalues and eigenvectors of the covariance matrix, so we can use the low-dimensional subspace projection coefficients to describe the original data:

$$y = \sum_{i=1}^{t} a_i w_i$$

(3)

where $t$ is the number of the first largest eigenvalues. This feature extraction method based on PCA is proposed by Kirby et al to use for face recognition. Latter, Turk et al. proposed Eigenfaces method for the front of the face recognition. The idea of Eigenfaces is obtained eigenvectors by PCA, the image of any given face can be approximated as a linear combination of the face images of the group characteristics, a combination of the
coefficients will be as face eigenvectors. Original Eigenfaces method is based on Euclidean distance nearest center classifier.

With many applications, there are many recognition methods based on PCA and other subspace methods. Pentland et al. used each perspective for local principal component analysis to achieve the multi-view face recognition. Some researchers used the characteristics of the two main decomposition orthogonal subspace to propose Bayesian framework based on principal component analysis. The selected main priority of Principal component analysis is usually to determine the corresponding eigenvalue according to the size, the eigenvalue is more bigger and it has the higher priority. But on the face recognition, how to choose the number of the main element is the best, the commonly used standard in two ways:

(1) When the corresponding eigenvalue and the largest eigenvalue compared to less than a certain value, to abandon it;

(2) To select eigenvalues and with the total of the characteristic value and the ratio is greater than or equal to 0.9.

Given a $M \times N$ face image, to transform the face image to a column vector firstly, then the dimension of the vector is $D = M \times N$. The mage matrix transforms to vector representation as Fig. 2. Thus, $D$ will be the dimension of the face image, which is also the number of dimension of the image space. Let us assume that there are $L$ vectors of size $D$ representing an image training set, each image is represented as:

$$x_i = [p_1, p_2, p_3, ... p_N]^T \quad i = 1, 2, ..., L$$  (4)

The sampling set is mean centered by subtracting the mean image from each image vector. So, mean image $\mu$ is represented as follows:

$$\bar{x} = \frac{1}{L} \sum_{i=1}^{L} x_i$$  (5)

Each mean centered image is computed by following equation:

$$\mu_i = x_i - \bar{x}$$  (6)

Then, we need to find a transform matrix which has the largest projection from high dimension data to low dimension onto the vector $\mu_i$. Then to obtain $M$ orthonormal vector $e_i$, this vector must meet as following:

$$\lambda_i = \frac{1}{M} \sum_{i=1}^{M} (e_i^T \mu_i)^2$$  (7)

These orthonormal vectors are constrained as:

$$e_i^T e_k = \delta_{ik}$$  (8)

In the above equation, $e_i$ and $\lambda_i$ represent the eigenvectors and the eigenvalues of the covariance matrix respectively, the covariance matrix is computed as:

$$C = WW^T$$  (9)

where $W = \sum \mu_i$. So, the size of the covariance matrix $C$ is $D \times D$, and it could be a high dimension matrix. For example, there is an image and its size is $64 \times 64$, then it can create the covariance matrix of size $4096 \times 4096$. In solving this question, it is not reasonable to compute the eigenvectors of the covariance matrix directly. A new method is proposed to solve for the eigenvectors of the covariance matrix, we can obtain the eigenvalues and the eigenvectors of the matrix $W^T W$. To assume $p_i$ and $\alpha_i$ represent the eigenvectors and the eigenvalues of the matrix $W^T W$ respectively, they meet the following equation:

$$W^T W p_i = \alpha_i p_i$$  (10)

Transform the above equation:

$$WW^T (Wp_i) = \alpha_i (Wp_i)$$  (11)

B. 2DPCA

In the face representation technology based on PCA, we need to transform the 2D face image matrices into 1D image vectors. The size of the input image vector is always very big and it is into very high dimension vector space. So, there is very difficult to compute the covariance matrix much accurately, which is due to its big size and the relatively small number of training samples. To overcome the flaws, a new method based on two-dimensional principal component analysis (2DPCA) is proposed for image texture representation and feature extraction. 2DPCA is different to original PCA, which is based on 2D matrices yet PCA is based on 1D vectors. What’s more, the original face images do not need to be previously transformed into an image vector. By contrast, 2DPCA can compute under the original image, and the covariance matrix of the training set can be built directly. 2DPCA consists of two advantages over original PCA method:

(1) 2DPCA is easier to evaluate the accuracy of the covariance matrix.

(2) 2DPCA can save some time not to transform 2D image into an image vector. So it has been widely applied rang from many fields.

Given a $M \times N$ face image, we firstly compute the mean image:

$$\bar{X} = \sum_{i=1}^{M} X_i$$  (12)

where $X_i$ represents each face image from sampling set, and it is the 2D image matrix. So, the size of $X_i$ is $M \times N$. Computing the covariance matrix:

$$C = \frac{1}{M} (X_i - \bar{X})(X_i - \bar{X})^T$$  (13)

Figure 2. Image matrix transforms to vector representation.
According to the threshold \( t \), we select the firstly \( t \) largest eigenvalues and these related eigenvectors. We note \( A \) to be the matrix as the corresponding eigenvectors. The matrix \( A \) can be used for feature extraction. For example, an unknown image projection feature is obtained:

\[ Y_k = A^T X_k \]  

(14)

where \( Y_k \) is the projection of \( X_k \). So we can obtain the principal component by using the projection technology. According to the above equation, all projection images are obtained by projection as \( Y = [Y_1, Y_2, \ldots, Y_k] \).

III. INDEPENDENT PRINCIPAL COMPONENT ANALYSIS

A. The Definition of Independent Component Analysis

Given \( n \) observed signals \( x_i (i = 1, 2, \ldots, n) \), assumed each observed signal includes a linear mixed combination with statistically independent source signals \( s_j (j = 1, 2, \ldots, m) \). That is represented as following:

\[ X = AS \]  

(15)

where \( X = [x_1, x_2, \ldots, x_n]^T \) represents the observed signal matrix, \( S = [s_1, s_2, \ldots, s_n]^T \) is source signals matrix, and \( A \) is a mixed matrix. The equation (15) represents a basic ICA model, and it describes the observed signals by means to be mixed by the independent component. As we all know, the independent component \( s_j \) can’t be direct to obtain. What’s more, the mixed matrix is also known, and the only known is the observed signals \( X = [x_1, x_2, \ldots, x_n]^T \). Therefore, the task of ICA is to estimate the mixed matrix \( A \) and independent component \( s_j \) in the case of the observed signal only to be known.

Because the mixing matrix \( A \) is unknown, source signals can’t be obtained directly from observed signals. We can obtain a separation matrix \( W \) by separating the dependent component from mixed signals, and then we can obtain as following: \( \hat{S} = WX \). The equation requires that \( \hat{S} \) is a good approximation of the true source signals \( S \). If the separation matrix \( W \) is obtained, so the mixed matrix \( A \) will be obtained by computing the inverse matrix of \( W \).

Premise without any prior knowledge, these results are certainly not unique to decompose observed signals into multiple mutually statistically independent components. So we want to add some constraints, the results are as close as possible to our expectations. Hyvarine [13] proposed three hypotheses to solve the above question. These hypotheses are as following: Between the respective components of the source signals are statistically independent; the components of the source signal must be non-Gaussian signal can only have a Gaussian signal; the number of components of the source signal is not more than the number of observed signal component, so the mixed matrix is inverse matrix or full column rank mixing matrix, and \( n \) observed signal can decompose up to the \( n \) source signal components. To satisfy the above three putative ICA model can be estimated. Among them, the first assumption is to solve the problem of independent component analysis, although this assumption looks very strict, but under normal circumstances, the source signals are sent by different physical systems, generally are able to satisfy the conditions of statistical independence. It is also for this reason, independent component analysis in order to have a wide range in many areas. For the second assumption, we know that the Gaussian distribution is completely symmetrical, and the linear combination of Gaussian distribution is still comply with the Gaussian distribution, so there are two or more components when the source signals obey the Gaussian distribution, the independent component analysis is impossible to achieve. The third is assumed to be in order to ensure the quantity of the observed signal than the number of source signals, and under normal circumstances, the mixing matrix \( A \) is assumed to remain unchanged. In most practical applications, these three are assumed to be able to satisfy.

According to above analysis, there is a certain degree of uncertainty by above method to estimate independent component. Firstly, can’t be determined independent component of variance (energy), that use the ICA estimated independent component is only true source best approximation of the signal, and between them there may be some kind of variable proportion, that is independent component may be estimated from its corresponding source signal is multiplied by the weighting coefficients derived results. Second, by the estimated independent component, we can’t determine the order in which they are. This is because the \( S \) and \( A \) are unknown. However, this uncertainty in most cases on the order is irrelevant.

B. Independent Component Analysis Independence Criterion

For independent component analysis model, the observed signal consists of a linear mixed signal of a number of independent sources, and the observed signal is closer to a Gaussian distribution, or more the former than the latter Gaussian. So we can put the separation results of non-Gaussian as a metric, which is used to detect the statistical independence between the separation results. When the non-Gaussian for the separation results are very strongest, it indicates that they have the most strongest statistically independence. Entropy is a fundamental concept in information theory, and it is a measure of the signal randomness.

To a random variable \( x \) with the probability density function: \( p(x) \), the entropy can be defined as:

\[ H(x) = -\int p(x) \log p(x) \, dx \]  

(16)

By the information theory of knowledge, we can know all random variables have the same variance, variable Gaussian distribution with maximum entropy. In the base of entropy, the negative entropy can be defined as following:

\[ J(x) = H(x_{gauss}) - H(x) \]  

(17)

where \( x_{gauss} \) and \( x \) are with the same covariance Gaussian random variable. From entropy characteristics described above, we can easily deduce that the negative
The entropy of the variable is always non-negative. When and only when the variable obeys Gaussian distribution, \( f(x) = 0 \). The value of \( f(x) \) has larger value what indicates the stronger of non-Gaussian quality, so it is more and more far away Gaussian distribution. Therefore, we can also use the negative entropy as an important indicator of the measure of non-Gaussian random variable.

The big advantage of the negative entropy as a measure of non-Gaussian is that it has a rigorous statistical theory background. Therefore, the negative entropy is a measure of non-Gaussian optimal measure. However, we can be seen to calculate negative entropy according to the probability density function of the random variables by definition, which will increase computational complexity. Generally, we use approximated method to calculate the negative entropy.

The second method is mutual information minimization criterion. Mutual information is an important concept in information theory. It is an important indicator of the measure of the independence of the independent component analysis separation results. To a \( m \) dimensional random variable \( X = \{x_1, x_2, \cdots m\} \), whose mutual information is defined as:

\[
I(x) = \sum_{i=1}^{n} H(x_i) - H(x) 
\]

where \( H(\cdot) \) is the differential entropy. According to knowledge of information theory, \( I(x) \) is non-negative. When the random components are statistical independence of each other, \( I(x) = 0 \). Therefore, mutual information can be used as an important measurement criterion to distinguish signal statistical independence. When the mutual information between signals obtains the minimum value, we can think that these signals are statistically independent.

Mutual information has an important property, so the following equation can be described for reversible linear transformation:

\[
I(y_1, y_2, \cdots y_n) = \sum_i H(y_i) - H(X) - \log|\det(W)| 
\]

By the definition of the negative entropy, and assume that the components \( y_i \) are linear irrelevant each other. Then we can obtain:

\[
I(y_1, y_2, \cdots y_n) = C - \sum_i J(y_i) 
\]

where \( C \) is constant independent of \( W \). The above equation describes the relationship between entropy and mutual information. The negative entropy is obtained the largest value, we consider that these signals are statistically independent. That mutual information reaches the minimum that negative entropy is to obtain maximum value.

The third method is maximum Likelihood criterion. Maximum likelihood estimation is common method to describe independent component analysis. Based on independent component analysis model \( X = AS \). The observed signal \( X \) is obtained by mixing the source signal \( S \), and the mixed matrix is \( A \). Therefore, the likelihood function of observed signal is defined as:

\[
L(A) = E\{\log\hat{p}(x)\} = \int p(x)\log p(A^{-1}x)dx - \log|\det(A)| 
\]

where \( \hat{p}(x) \) is the estimated value of the observed signal \( x \) with the probability density function \( p(x) \). \( L(A) \) is a function related with the matrix \( A \). When \( W = A^{-1} \), log likelihood function is as following:

\[
L(W) \approx \frac{1}{n} \sum_{i=1}^{n} \{\log p(Wx)\} + \log|\det(W)| 
\]

where \( n \) is the number of observed sample which obey independent and identically distributed variables. We can find that maximize the likelihood function to obtain the best estimate of the separation matrix \( W \) form. Then we can obtain the best estimate of independent source \( S \). From the perspective of information theory, maximum likelihood estimation criterion essentially mutual information minimization criterion is consistent.

## IV. PROPOSED SCHEME

### A. ICA Facial Expression Model

According to the aforementioned ICA model, we can obtain:

\[
x = As = \sum_{i=1}^{n} a_i s_i 
\]

where \( x = (x_1, x_2, \cdots x_m) \) is observed signals, \( s = (s_1, s_2, \cdots s_m) \) is the matrix which consists of unknown signal source, and \( A = [a_1, a_2, \cdots a_n] \) is \( m \times m \) matrix. When meet the blind separation problem under solvable conditions, the separation matrix is obtained to meet \( U = Wx \) and \( U \) is independent of each other as far as possible.

Each expression image is expanded to become a one-dimensional column vector, and we use \( X_i \) to represent the related vector of the \( i \)-th image. A training set of \( M \) expression images is described as: \( X = (X_1, X_2, \cdots X_M) \). Assume \( M \) vectors consist of \( M \) independent basis vectors \( S = (S_1, S_2, \cdots S_M) \). According to the equation (17), then we know \( X = AS \).

Then compute the separation matrix \( W \), so output is as following:

\[
U = WX = WAS 
\]

\[
W = A^{-1} 
\]

where \( U = (U_1, U_2, \cdots U_M) \). \( U \) is the estimation of statistical independent base vector, and each row represents a statistically independent basis vectors. The expression vector of the image to be identified is then projected onto the subspace, i.e. the linear combination of the set of base vectors to represent. Let \( f \) be identified expression image vector, then:

\[
f = a_1 U_1 + a_2 U_2 + \cdots a_M U_M 
\]

where \( U_1, U_2, \cdots, U_M \) represent \( M \) basis vector, and \( a_1, a_2, \cdots, a_M \) are projection coefficients. Then to choose SVM classifier to obtain recognition results.

### B. SVM Classifier

classifier is a strong generalization ability, especially in the optimization of the small sample size problem, multi-linear, non-linear and high dimensional pattern recognition problems demonstrate the unique advantage. It can be well applied to the function fitting model predictions other machine learning problem. VC theory SVM method is based on statistical learning theory and based on structural risk minimization principle above, through limited training sample information to seek the best compromise between the model complexity and generalization of learning ability, expect most good generalization ability.

The main idea of SVM is: non-linear sub-sample set, the first sample after the nonlinear transformation projection to a high space to find an optimal classification hyper-plane in high space, making the best classification results. It has a sample data set:

\[(x_1, y_1), \ldots (x_i, y_i), y_i \in \{-1, 1\}\]  (26)

Compute a nonlinear transformation \(Z = \Phi(x)\) to the sample data \(x_1\):

\[\begin{cases} w^T z_i + b \geq 1, y_i = 1 \\ w^T z_i + b \leq -1, y_i = -1 \end{cases}\]  (27)

Converse the above equation as following:

\[y_i (w^T z_i + b) \geq 1 \quad i = 1, 2, \ldots l\]  (28)

Assume that two kinds of optimal separating surface equation is:

\[w_0^T z + b_0 = 0\]  (29)

Therefore, the interval between the two categories can be expressed as:

\[\rho(w, b) = \min_{(x|y=1)} z_w^T w - \max_{(x|y=-1)} z_w^T w\]  (30)

So this time \(w_0\) should satisfy such that:

\[\rho(w_0, b_0) = \frac{2}{\|w_0\|} = \frac{2}{\|w_0 \cdot w_0\|}\]  (31)

To obtain the maximum value, we use quadratic programming to solve the optimization problem:

\[\max_{w, b} \Phi(w) = \frac{1}{2} (w^T w)\]  (32)

The formula’s constraint condition is \(y_i (w^T z_i + b) \geq 1 \quad i = 1, 2, \ldots l\). Because the above equation solving linear inseparable problems, then we construct the Lagrangian constraints to solve the minimum question:

\[\Phi(w) = \frac{1}{2} w^T W + \gamma \sum_{i=1}^l \xi_i\]  (33)

where \(\xi\) classifier error. Then we can obtain:

\[w_0 = \sum_{i=1}^l \lambda_i y_i z_i\]  (34)

Decision-making function is as following:

\[f = sgn[y \sum_{i=1}^l \lambda_i y(x^T z_i) + b]\]  (35)

In high-dimensional data transformation kernel function to solve the non-linear data conversion issues, kernel function method is rewrite the decision function in the above equation to be obtained as following:

\[f = sgn[y \sum_{i=1}^l \lambda_i y K(x, x) + b]\]  (36)

We do not need to find a mapping function from low-dimensional to high-dimensional data mapping, only need to know the output can be converted.

For the common linear inseparable, SVM can take advantage of the known nuclear function mapping low-dimensional data from low-dimensional to high-dimensional space, and can be constructed in a high-dimensional space can be divided into a linear hyperplane. Since the original classic SVM algorithm for the two types of classification and recognition algorithm, achieved by a combination of two types of facial expression recognition of multi-class problems. There are two methods:

1. "one-to-one" strategy, training multiple classifiers that separate each category twenty-two;
2. one-to-many "strategy, that is training a classifier which a separate class and all the rest of the class. This paper used the principle of "one to many", and SVM classifier with the nearest neighbor distance separation combined to achieve optimal classification performance

V. EXPERIMENTAL VERIFICATION

In this experiment, we use Jaffe facial expression database, Weizmann database and IMM face database. The database contains 213 images of 7 facial expressions (sadness, disgust, happy, fear, neutral, angry, surprise) posed by 10 Japanese female models. Each image has been rated on 6 emotion adjectives by 60 Japanese subjects. Fig. 3 shows the part of Jaffe face database.

![Figure 3. Jaffe facial expression database](image_url)

In order to reduce the position of images of human faces, the impact of the size of the human face, grayscale, size and other factors, the first facial feature points of the image to be normalized, the characteristic parts normalized to the same standards, and then the image further pretreatment, such that the image has the same size, the mean and variance after pretreatment. The normalized characteristic parts criteria: different expression images in the center of the two eyes, the mouth of the center in the same position respectively.
We use PCA and ICA algorithm to implement facial expression recognition. Tab.1 shows expression recognition rate comparison based on PCA and ICA. We can see from the table above, ICA is a real effective facial expression recognition method. The recognition rate is greater than 90%. But the large amount of computation that the ICA spend more computing time.

Next, we select the Weizmann face image database and the IMM face database to verify the recognition results based on ICA algorithm. This database has 28 subjects under 5 different poses, 3 illuminations and 3 facial expressions. We select 3 expressions for each subject to implement experiments. Fig. 4 shows the part of Weizmann database. IMM face database consists of 240 annotated images of 40 different human faces, and each image. Fig. 5 shows the part of IMM face database.

Tab. 1 shows expression recognition rate comparison based on PCA and ICA on Weizmann face database. We can see from the Tab.2, ICA is a real effective facial expression recognition method, and the recognition rate is greater than the recognition method based on PCA.

From Tab. 2, we can see that the recognition rate is biggest using ICA analysis to extract facial expression feature to implement facial expression recognition. In the Jaffe database, we can obtain 86.7% and 88.9% recognition rate by using PCA and 2DPCA respectively. Facial expression recognition can obtain much better performance by using 2DPCA than that of using PCA. Because directly based on two-dimensional sub-image matrix, which can easily reduce the dimension of the original features; in the feature extraction process can completely avoid to use the singular value decomposition method, and it is very simple. According to experimental results from Weizmann and IMM face database, the recognition rate is much higher by using 2DPCA than PCA.

From the experimental results, we also can see that the expression recognition rate is 92.6%, 93.4% and 92.7% by using ICA to extract facial feature from three face databases respectively. Facial expression recognition method based ICA is much better than PCA and 2DPCA. Because the feature extraction method based on ICA consists of principal component analysis and linear discriminant analysis. So, it will have some of their advantages and common characteristics. It can extract facial expression feature more robustness and accuracy, and which will improve the expression recognition rate.

![Figure 4. The part of Weizmann database.](image)

![Figure 5. The part of IMM database.](image)

<table>
<thead>
<tr>
<th>Face database</th>
<th>Method</th>
<th>Jaffe</th>
<th>Weizmann</th>
<th>IMM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PCA</td>
<td>86.7%</td>
<td>85.6%</td>
<td>86.8%</td>
</tr>
<tr>
<td></td>
<td>2DPCA</td>
<td>88.9%</td>
<td>88.1%</td>
<td>89.0%</td>
</tr>
<tr>
<td></td>
<td>ICA</td>
<td>92.6%</td>
<td>93.4%</td>
<td>92.7%</td>
</tr>
</tbody>
</table>

VI. CONCLUSIONS

Facial expression recognition is one of the most challenging problems in the fields of image processing, biometric identification, movement tracking, computer vision, pattern recognition, physiology, psychology and so on, and it has become a hot research topic in the field of pattern recognition and artificial intelligence recently. Facial expression recognition is all important part of effective computing and intelligent human-machine interactive, which has a wide range of applications and potential market value.

In this paper, we mainly study the facial expression recognition method based on independent component analysis, this is a subspace analysis method. ICA is a novel statistical technique in signal processing and machine learning that aims at finding linear projections of the data that maximize their mutual independence. Firstly, we introduce the basic theory of ICA algorithm in detail, and then present the process of facial expression recognition based on ICA model. Finally, we use PCA, 2DPCA and ICA algorithm to implement expression recognition on different face databases. Experimental results show ICA is a real effective facial expression recognition method and the recognition rate based on ICA is greater than based on PCA and 2DPCA.

REFERENCES


