The Method of Human Facial Expression Recognition Based on Wavelet Transformation Reducing the Dimension and Improved Fisher Discrimination

Chuang Yu
School of Information Science & Engineering
Shenyang Ligong University
Shenyang, China
Yuchuang2002@163.com

Yuning Hua
School of Information Science & Engineering
Shenyang Ligong University
Shenyang, China
sy-huayuning@163.com

Kun Zhao
School of Information Science & Engineering
Shenyang Ligong University
Shenyang, China
Zhao_y_cn@163.com

***Abstract*—The segmentation of skin color region is carried on through the mathematics morphology processing, using Hough transform to locate the eyes, and then the expression image is proceeded by geometry standardization. After using wavelet transform to reduce the dimension of images, the feature extraction of facial expression can be realized with the discrimination of improved Fishier. It can solve actual problems on dispersion matrix singular within the class .Finally ,the method of minimum Mahalanobis distance classifier is carried on facial expression recognition. It is proved in the CMU facial expression database that the method can reduce computation and enhance the recognition rate of facial expression recognition.**

Keywords-Hough transform；wavelet transform；improved Fisher discrimination；facial expression recognition;

I. INTRODUCTION

Facial expression recognition is an emerging research focus in the field of artificial intelligence, its research goal is to make some of artificial intelligence products identify the person's face automatically, and then to analyze human's emotion. It is an important component of emotion computation, intelligent human-computer interaction. It has the widespread application prospect and the potential market value. Many domestic and foreign institutions are conducting research in this area [1].

At present, most of the facial expression analysis and recognition are mainly directed against the basic expression analysis and identification. The methods used generally fall into two categories: the method based on static images (single image) and the method based on dynamic image sequence[1]. In the method of static images, they are divided into two methods: one is based on statistics, the other is based on the geometry. In the application of statistical methods to solve the issue of expression recognition, the dimension reduction has become the key to deal with practical problems. The Fisher linear discrimination is a more commonly method used for feature parameter extraction. The traditional method of Fisher linear discrimination has large amount of computation, in the request of the non-singular dispersion matrix .But in reality, the dispersion matrix within the class is usually singular. In view of this kind of situation, in this paper, the method of fisher most superior discrimination in case of dispersion matrix being singular within the class is improved, and it is applied for expression recognition[2].

II. FLOW OF EXPRESSION RECOGNITION

The flow of expression recognition in this paper is shown in Figure 1:

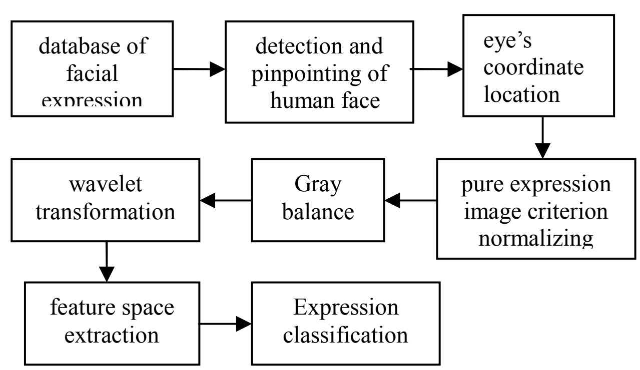

Figure 1. Flow chart of expression recognition

III. PREPROCESSING OF EXPRESSION IMAGE

A. *The Segmentation of Skin Color Regions Based on The Processing of Mathematical Morphology*

The YCbCr color model is used in this paper, and the skin color detection is carried out with chromatic aberration

978-0-7695-4249-2/10 $26.00 © 2010 IEEE
DOI 10.1109/ICINIS.2010.98

component Cb and Cr. Taking the chromatic aberration component [Cb, Cr],T is as the training samples. The definition of class label of skin color sample's is 1,and the definition of class label of non-skin color sample's is -1.The kernel function of SVM is the radial basis function, that is,

$$K(x,z) = e^{-\|x-z\|^2/\sigma^2}, x, z \in R^d$$

The chromatic aberration component of any pixel in the image to be detected x=[Cb, Cr]T, the classification function of skin color/ non-skin color

$$y = \mathrm{sgn}[\sum_{i=1}^{N_s} \alpha yiK(x_i, x) - b], x, x_i \in R^2$$

Therein, xi is for support vector, yi is the corresponding class label to xi; NS is the number of support vector. In the process of color detection, SVM carries on the classification to every pixel of image. The output result is a two value image. The pixels being judged as skin color constituted certain connected regions. The regions which are misjudged usually contain fewer pixels. The fewer pixels can be treated as noise and can be removed with the morphology of the opening operation.

This article has used the method which confirms repeatedly, after obtaining a skin color regions, the person face confirmation is carried on using of eye's characteristic.

B. Binocular Pupil's Pinpointing

The Canny operator is selected to carry on the edge detection of eyeball and to extract the edge information of eyeball. Afterwards, the improved the Hough algorithm is used to extract center coordinates of the eyeball, the steps are: First, setting the range of center coordinates (a, b) for parameter space; Second, in return, taking each spot within the range as the center of circle, the generating arc algorithm with central point is applied to extract the values of all the pixels on the circle of radius R in the image space; At last, determine whether the point is the boundary point of a normalization image; If so, the corresponding coordinates (a, b) of peak point for parameter space is the demand of the center coordinate; else, repeating the step.

C. Geometry Standardization of Expression Image

In this paper, the geometry standardization of expression image mainly includes rotating, cutting and scaling. Image is rotated and transformed, so that two eyes rotate to the same level; Image is cut according to eye's coordinate position, and the facial expression domain zone is separated from the background. The horizontal distance between the eyes is set for the d, in the vertical direction, taking 0.5d upward and 1.5d downward; In the horizontal direction, the midpoint between the eyes is as the datum, taking the length of d respectively on the left and right. The center distance of left and right eye in each image is fixed (0.5d, 1.5d), and the image size is 2d * 2d. This has been standard pure expression of face images[3]. After the standardization, the pixels of image is set for 256*256. Finally, the all images are transformed into the same size by scaling.
The pretreated images are shown in Figure 2:

Figure 2. preprocessing of expression image

IV. IMAGE PROCESSING OF DIMENSIONALITY REDUCTION USING WAVELET TRANSFORMATION

Supposing, two-dimensional image $f(x,y) \in L^2(R^2)$, then, two-dimensional approximation of image $A_j f(x,y)$ can be decomposed into:

$$A_j f = A_{j+1} f + D_{j+1}^2 f + D_{j+1}^3$$

therein, $A_{j+1} f = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} C_{j+1}(m,n) \phi_{j+1}(m,n)$;

$D_{j+1}^i = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} D_{j+1}^i (m,n) \varphi_{j+1}^i (m,n), i = 1,2,3$,and

$$\begin{cases} C_{j+1}(m,n) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} h(k-2m)h(l-2n)C_j(k,l) \\ D_{j+1}^1(m,n) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} h(k-2m)h(l-2n)C_j(k,l) \\ D_{j+1}^2(m,n) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} h(k-2m)g(l-2n)C_j(k,l) \\ D_{j+1}^3(m,n) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} h(k-2m)g(l-2n)C_j(k,l) \end{cases}$$

Making H_r and H_c represent respectively low-pass filter operator H of rows and columns of overview coefficient C_j which is j-level of wavelet decomposition to images; G_r and G_c represent respectively high-pass filter operator of rows and columns of overview coefficient which is j-level of being played role in image. The solution algorithm of two-dimensional Mal lat can be written in the form of following:

$$\begin{cases} C_{j+1} = H_r H_c C_j \\ D_{j+1}^1 = H_r G_c C_j \\ D_{j+1}^2 = G_r H_c C_j \\ D_{j+1}^3 = G_r G_c C_j \end{cases} \qquad j = 0,1,2 \ldots J$$

Supposing, $A_j f$ is the overview subgraph, of which the resolution is 2^{-j} and the wavelet decomposition is carried on it. Therein, L represents low frequency, H represents high frequency, subscript 1, 2 represents first-level and second-level of decomposition. $A_{j+1}f$ is relative to the LL part, $D_{j+1}^{1}f$ is relative to the LH part, $D_{j+1}^{2}f$ is relative to the HL part, $D_{j+1}^{3}f$ is relative to the HH part.

After a primitive image undergoing K levels of wavelet transform, the low frequency sub-image size is only the 1/2k of primitive image, therefore it can effectively achieve dimensionality reduction. But the excessive levels of wavelet transform will lead to ambiguous facial expressions, thus the important category message will be lost. It is found by the experiment that when undergoing 2 levels of wavelet decomposition, the recognition performance is the best [3]. The diagram of 2 levels of wavelet transformation decomposition is shown in Figure 3.

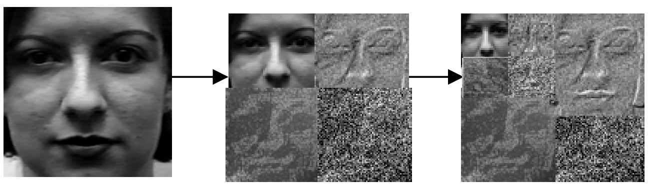

Figure 3. 2 levels of wavelet transformation to image

V. THE COMPUTATION OF EXPRESSION FEATURE SPACE BASED ON IMPROVED FISHER LINEARITY DISCRIMINATION

The dispersion matrix within the class S_W is demanded to be nonsingular for traditional fisher linear discrimination. In the actual situation, the dispersion matrix within the class S_W is usually singular. In view of this situation, in this paper, under the condition of the dispersion matrix within the class being singular, the algorithm of Fisher most superior discrimination has been improved, and the computational process is as follows:

Supposing, there are samples of C class: $\Omega_i = \left\{x_1^{(i)}, x_2^{(i)}, \ldots, x_{N_i}^{(i)}\right\}$, $x_k^{(i)} \in R^n, i = 1,2,\ldots..C, k = 1,2,\ldots,N_i$, N_i is the number of class i, then

$$
\begin{aligned}
N &= \sum_{i=1}^{C} N_i; \\
\overline{x}_i &= \frac{1}{N}\sum_{k=1}^{N_i} x_k^{(i)}; \\
\overline{x} &= \frac{1}{N}\sum_{i=1}^{C} \overline{x_i};
\end{aligned} \tag{1}
$$

the definition of the dispersion matrix within the class for sample is as follows：

$$
\begin{aligned}
S_W &= \frac{1}{N}\sum_{i=1}^{C} N_i S_W^{(i)}; \\
S_W^{(i)} &= \frac{1}{N}\sum_{k=1}^{N_i} \left(x_k^{(i)} - \overline{x}_i\right)\left(x_k^{(i)} - \overline{x}_i\right)^T;
\end{aligned} \tag{2}
$$

The definition of the dispersion matrix between the class of the sample is as follows：

$$
S_b = \frac{1}{N}\sum_{i=1}^{C} N_i \left(\overline{x_i} - \overline{x}\right)\left(\overline{x_i} - \overline{x}\right)^T; \tag{3}
$$

Suitable transformation matrix W is chosen to make the linear transformation to all samples, hoping Fisher criterion function transformed $J_F(W) = \left|W^T S_b W\right| / \left|W S_b W^T\right|$ to achieve the maximum value. Under the condition of S_W being nonsingular, W is the feature vector of $S_W^{-1} S_b$. However, if it is solved directly, the computation will be very large. Therefore, W is solved with the diagonalization method of Sb and S_W .In the practical application , S_W is often singular, in view of this situation, for two types of singular and non-singular of S_W ,they are processed as follows:

Under the condition of S_W being nonsingular, decomposing $S_W = P^T P$, making $v = Pu$, $P^{-T} S_b P^{-1} = R$, hence：

$$
\begin{aligned}
&\max_{u \in R^n} \left(u^T S_b u / u^T S_w u\right) \\
&= \max_{v \in R^n} \left(v^T R v / \|v\|^2\right)
\end{aligned} \tag{4}
$$

In the equation above, the best v* on the right is composed of the first k feature vectors corresponding to the maximum, and the best on the left u=P-1 v* .

Under the condition of S_W being nonsingular, setting rank (S_W)= $r < n$,taking the standard orthogonal basis P=(p1, p2…pr,pr+1,..pn), making PTSWP=diag (λ 1, λ 2, λ r,0,0,..0), therein λ i>0,i=1,2,….,r, making the transformation u=P v; then:

$$
\begin{aligned}
&\max_{u \neq 0} \left(u^T S_b u / u^T S_w u\right) \\
&= \max_{v \neq 0} \left(v^T R^T S_b P v / \sum_{1}^{r} \lambda_i v_i^2\right)
\end{aligned} \tag{5}
$$

Making matrix P=[P1,P2] =[(p1,p2,…..pr),(pr+1,… pn)] , then:

$$
\begin{aligned}
P^T S_b P &= \left[P_1^T P_2^T\right]^T S_b \left[P_1^T P_2^T\right] \\
&= P_1^T S_b P_1 * P_2^T S_b P_2 - P_1^T S_b P_1 * P_2^T S_b P_1
\end{aligned} \tag{6}
$$

S_W is the first type of singular matrix: $P_2^T S_b P_2 \neq 0$. As the semi-definite matrix $P_2^T S_b P_2 \neq 0$, the first $k \leq n - r$ non-zero maximum feature value of λ =diag (λ 1, λ 2,…. λ k) feature vector corresponding η =(η 1, η 2,…, η k). Constructing t to make its first r component being zero, η is the latter k component. It may result in : $t_i^T P^T S_b P t_1 1 = \lambda_1 \|\eta_1\|^2 > 0$

Therein i=r+l,….r+k, $\sum_{i=1}^{r}\lambda_i t_i=0$,then, the solution of (5)is $u=pt=p_2\eta$.

In order to reduce double counting of the expression recognition, improvement is carried out on the algorithm above as follows, the projection vector u becomes:

$u^* = u^T D_W^{-1/2}$, $D_W = u^T S_W u$. Then, the last transformation is : $y=(u^*)^T x$, x is image vector of facial expression for the input , y is the feature vector of expression image .

S_W is the second type of singular matrix:： $P_2^T S_b P_2$ =0, here $S_b P_2 = 0$, hence,(4)is transferred as :

$$\max_{u\neq 0}\left(u^T S_b u / u^T S_W u\right)$$
$$= \max_{v_1 \dots v_r}\left\{[v_1,..,v_r]P_1^T S_b P_1 [v_1,..,v_r]^T\right\} / \sum_{t}^{r} \lambda_i v_i^2 \quad (7)$$

The target value on the right of equation has no relation to vr+1,…..,vn, hence, taking ti=0,i=r+l,…n,, a further transformation vi= $\lambda_i^{1/2}\rho_i$, i=1….r and note:

$$\rho = \left[\rho_1, \rho_2, ... \rho_r\right]^T,$$
$$A = \mathrm{diag}\left(\lambda_1^{-1/2}, \lambda_2^{-1/2} ... \lambda_r^{-1/2}\right) P_1^T S_b P_1 \mathrm{diag}\left(\lambda_1^{-1/2}, \lambda_2^{-1/2} ... \lambda_r^{-1/2}\right) \quad (8)$$

Then, the right of equation(7) is $\max_{\rho\neq 0} = (\rho^T A \rho)/\|\rho\|^2$. The best ρ^* of equation (6) is feature vector of feature value corresponding A. Thus ,the best of equation (6) on the right $t_i = \lambda_i^{-1/2}\rho_i^*$, i=1,2,…r ;The solution of equality (5) is $u = Pt = P[t_1, t_2, ..., t_r, 0, ..., 0]$.

Similarly, in order to achieve better classification results, the projection vector u is made the following improvements $u^* = u^T D_W^{-1/2}$, $D_W = u^T S_W u$, the last of the transformation is: $y=(u^*)^T x$ .u.

VI. EXPERIMENTAL RESULT

The facial expression database used in this paper is Facial Expression Database of Carnegie Mellon University(The CMU-PITTSBURGH AU-Coded Face Expression Image Database),and the database is mainly composed from 18 to 50 year-old adults , of which are 69% female;31% are male;81% are Europeans and Americans;13% are black;6% are other races. Figure 4 is in part of the images in the library image.

Figure 4. part of the images in the library image of CMU

In the database of the CMU facial expression,30 individuals are selected as the sample. Each person has 7 kinds of expression(neutral face, happy, sad, surprise, anger, disappointment, fear),the training sample is composed of 4 of each person's group of expression images in random ,and the rest of the images are as the test samples. Each image is changed into 8 bit gradation images of 256×256 by normalization, the images were carried out on the 2 levels and 2-dimensional of wavelet transformation, the size of 64×64 of low frequency subgraph is obtained .In this paper, the test comparison among PCA, the traditional Fisher linear discrimination algorithm and improved recognition algorithm is carried out, using the minimum Mahalanob is distance classification[4] .The result is shown in Table 1.

TABLE I. TABLE OF EXPERIMENTAL RESULT

	recognition method		
expression	PCA	traditional Fisher	improved Fisher
neutral	77%	81%	85%
happy	79%	82%	89%
Surprise	78%	84%	89%
Sad	73%	79%	86%
anger	80%	84%	91%
disappointment	70%	72%	85%
fear	75%	78%	89%

VII. CONCLUSION

In this paper, region segmentation of skin color is carried out through mathematical morphology, using Hough transform to locate the eyes, and then the geometry standardization processing is carried out on expression image. After using wavelet transform to reduce the dimension of the image, when the data having finished wavelet transform is used as the feature vector of person facial expression recognition, the useless information is tick out, which greatly reduces computation of the feature space. The improved Fisher discrimination to achieve the feature extraction of facial expression, has solved the actual problem in the situation of dispersion matrix being singular within class, and finally the method of minimum Mahalanobis distance classification is used in facial expression recognition. The experiments through the CMU facial expression database show that on the base of reducing computations, the recognition rate of the improved Fisher algorithm surpasses PCA method[5] and the traditional Fisher linearity discrimination.

REFERENCE:

[1] Wang Zhiliang and Chen Fengjun, "A survey of facial expression recognition," Computer Applications And Software, vol. 20(12), pp.27-30, 2003.

[2] He Hualiang, Zhou Chairong and Bao Yongqiang, "The research advance of facial expression recognition," Journal Of Circuits And Systems,vol. 10(1), pp.30-34, 2005.

[3] Wang Hongbo and Wu Yan, "Face Recognition Based on Wavelet Transform and BP Neural Network," Computer Engineering And Applications , vol. 40 (24), pp.51-53,2004.

[4] Qu Zhiyi and Huang Heming, "Mahalanobis distance based facial expression recognition," Journal Of Lanzhou University, vol.41(6), pp.13-15, 2005.

[5] Zhang Nan, "Facial expression recognition based on modular 2DPCA," Journal of Shandong Institute Of Light Industry, vol. 21(1), pp. 15-16 2007.

[6] Chuang C,"Automatic extraction of head and face boundaries and facial," Information Sciences, vol. 158(1),pp. 117-130,2004.

[7] M Buciu I, Kotropoulos C and Pitas I, "ICA and Gabor representation for facial expression recognition," International Conference on Image Processing, pp. 855-858,2003.