Facial Expression Recognition Based on Local Vector Model

Yan Wang[a], Guoqing He[a]
[a] School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China.
Email: wangyan@lut.cn
Email:hegq.731@163.com

***Abstract*— Texture feature extraction is an important step in the facial expression recognition system. The traditional LBP method ignored the statistical characteristics of the texture change direction in the process of feature extraction, and we can extract more detailed texture information by the LDP method based on LBP, but the computational complexity is greatly increased. In order to extract more detailed texture information with the computational complexity is not increased, we proposed a method named Local Vector Model (LVM). In this method, modulus value and direction of the local texture changes are extracted as the features of classification. Furthermore, in order to improve the robustness that the algorithm to the subtle deformation of expression image, the Image Euclidean Distance is introduced and embedded in LVM. Finally, the even decreasing function is used to get the neighbor classification distance. Experiments on JAFFE facial expression databases with different resolution demonstrated that the method proposed in this paper is better than other modern methods.**

***Index Terms*— facial expression recognition, feature extraction, local vector mode, image Euclidean distance**

I. INTRODUCTION

THE/technology of facial expression recognition has taken a very important position in the man-machine interactive system, and it has attracted more and more attention in recent years [1], [2]. To realize the identification of the person's expression, not only the face facial texture variation is needed, but also physical characteristics such as ECG, Pulse Features and other physiological characteristics are needed. In generally, what we most concerned about is how to extract the expression features from images. At present, the main method for facial expression recognition can be divided into three main categories: the active appearance model(AAM) [3], the methods based on Gabor wavelet [4], [5] and the local binary pattern (LBP) [6] method. The methods based on AAM model can obtain more reliable expression characteristics but the calculation is complex and the initial parameters is difficult to determined. The method based on Gabor wavelet has higher requirements to internal memory and time, which is bad for the establishment of the real-time system. The LBP method is being researched and applied widely by researchers for its advantages such as strong performance of anti-jamming and convenient calculation. Compared with the AAM method and Gabor wavelet method, extract the texture features using LBP method does have obvious advantages. However, LBP approach ignored statistics for the change of direction, and cannot have a more detailed description for the strength and direction of the local texture changes, there are still some difficulties to extract the texture directly using the method to achieve a better classification. Thus, some researchers proposed the method of LDP [7], compared with LBP, the LDP method can extract more detailed texture feature to some extent, but the computing complexity is greatly higher than LBP. It is particularly necessary to find a method which can extract more detailed texture feature and less computing complexity for the establishment of the human-computer interaction system based on facial expression recognition. In this paper, we proposed a method of Local Vector Mode(LVM), in which we can get more detailed texture feature and lower computing complexity than LBP and LDP. The LVM is the vector of the local texture variation and have poor robustness for small deformation, so the Image Euclidean distance (IMED) [8] is introduced and embedded into the LVM method in order to extract more reasonable and effective expression features. Eventually, the method we proposed has the following advantages: Firstly, it can represents the strength and angles of local texture changes effectively. Secondly, it has a certain resistance noise ability and a stronger discernment. Thirdly, which has a better robustness for small deformation. Experiments on JAFFE facial expression databases with different resolution demonstrated that the method we proposed is better than the method of LBP and LDP.

Manuscript received August 21, 2013; revised December 20, 2013; accepted February 18, 2014.

This research is supported by the National Natural Science Foundation of China (No: 61263019), the Fundamental Research Funds for the Gansu Universities (No.1114ZTC144), the Natural Science Foundation of Gansu province (No: 1112RJZA029) and the Doctoral Foundation of LUT.

II. LOCAL VECTOR MODE

In the traditional LBP method, each pixel in the image is marked to get a binary number [9], and then convert the binary number into decimal number, which is called LBP code of the pixel. Eventually the histogram composed of the LBP codes of all pixels is regarded as LBP feature of the image. As the LBP formula shown in equation (1):

$$LBP(P,R) = \sum_{i=1}^{p-1} S(g_m - g_c)2^m \qquad (1)$$

doi:10.4304/jsw.9.9.2241-2245

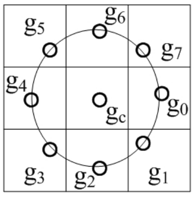

Figure 1. Local image pixels

In the equation, P and R respectively represent the number of points and radius in the template. g_m is neighbor of g_c. $S(x)$is defined as follows:

$$S(x) = \begin{cases} 1 & \text{if } x \geq 0, \\ 0 & \text{if } x < 0. \end{cases} \tag{2}$$

The proposed Local Directional Pattern (LDP) [7] is based on LBP method and can better describe the texture in the change of the direction. In generally, it is an eight bit binary code assigned to each pixel of an input image, which calculates eight directional edge response value of a particular pixel using Kirsch masks in eight different orientations centered on its own position. Although LBP and LDP can extract variation characteristics of facial texture to some extent, but the LBP ignores statistics of the changes in the direction and the computing complexity of the LDP is greatly higher. Based on the considerations of the change intensity, direction and time-consuming, this paper proposes the Local Vector Model method(LVM).

Local Vector Model can obtain the vector of local texture feature using a template, which utilized the strength of the local texture variation as the modulus of vector and used the texture change of angle as the direction of vector. For the template of $3 \leq 3$, g_c is the central pixel, $g_m(m = 0, 1, ..., 7)$ is a neighborhood of the center point as shown in fig.1 The calculation steps of the Local Vector Mode are as follows:

1) Calculate to get the matrix of texture intensity changes $A = [a_0, a_1, a_2, a_3]$. The values of A represents the degree of texture changes of each direction in the template, the element in which calculated by the change of the texture degree according to the counterclockwise direction from the horizontal direction to the right. Firstly, we calculate the template threshold c with the equation(3), and then a_i be calculated by the equation(4).

$$c = \frac{max(G) - min(G)}{9} \tag{3}$$

$$a_i = \frac{(g_i - g_{i+3})}{c} \tag{4}$$

where $G = [g_0, g_1, g_2, g_3, g_4, g_5, g_6, g_7, g_c]$. The choice of threshold c is critical in the step1). If the same threshold were used in different templates, the effect of the texture extraction will be greatly reduced. In LVM method, threshold c is determined by the gray value of all pixels within template, and then thresholding the change of each direction. In this way, it can demonstrate the variation of the local texture in each template and get a good classification results.

2) Plural transformation of texture variation. Calculate the biggest value of modulus and the corresponding serial number s using equation (5). s represents the corresponding serial number of the biggest change direction of the texture variation. Obtain the complex value of texture vector variation using the transform matrix. The row vector order in matrix M is consistent with the statistical order of the texture intensity change in step 1), so the maximum texture template change information can be represented as a complex number by the equation (6):

$$\begin{aligned} mod(s) = max\left\{abs(a_i)\right\}, \\ i = 0, 1, 2, 3ands = i \end{aligned} \tag{5}$$

$$A0 = A(s) \leq M(s + i) \tag{6}$$

where$M = [1, 1 + i, i, -1 + i]$, in which the real part represented the horizontal changes and the imaginary part expressed the direction of vertical changes. How to express both the strength and the direction of changes in texture with one number is a crucial step in LVM method. If a vector was expressed with two scalars, the computation will be greatly increased in the comparison of vector distance. Therefore the texture changes adopted the complex number to solve the problem. Here, there are only four complex numbers in M, we can determine which one is expressed in the eight direction of texture changes according to the plus or minus of a_i and the value of i.

3) Calculate the modulus matrix md and angle matrix ag. The md and matrix ag defined as follows:

$$S(x) = \begin{cases} ag = angle(A0) \\ md = module(A0) \end{cases} \tag{7}$$

The function $angle$ returns the phase angle of vector $A0$ and the function $module$ returns the module value of vector $A0$. md and ag are the important features using LVM to realize expression classification. In this paper, the classification is based on the comparison of md and ag in LVM characteristics of each image, which will be fully detailed in Section III.

An example of LVM feature extraction is shown in Fig.2, Extracting LVM features from a template of $3 \leq 3$. According to the procedure of LVM, we obtain $G = [120, 195, 180, 60, 160, 83, 140, 72]$ and $c = (195 - 60)/9 = 15$ from 1), and then according to step 2),we know $mod = 7$ and $s = 1$, that is $A0 = 7 + 7i$. Finally, we get $ag = 0.7854$ and $md = 10$, here ag is radians converted into angle$45°$. Thus the template texture changes can be ultimately defined by a complex number.

III. Classification Method Based on LVM

A. Image Euclidean Distance

Due the same individual expression intensity maybe is different, so the different textures generated. The feature

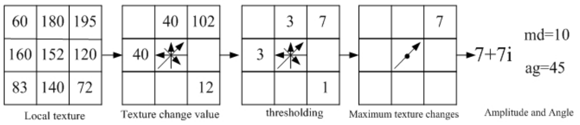

Figure 2. Examples of LVM Feature Extraction

Figure 3. Image embedded IMED

extracted using the texture variation will have subtle changes, and the changes will affect the accuracy of the classification. So it is necessary to take measures that is not obvious to the subtle deformation, Liwei Wang et al proposed a method called Image Euclidean Distance [8](Image Euclidean distance, IMED) which is provided with the above requirement. In Ref. [10]and [11], researchers used IMED to improve the related algorithms, and the experimental results have shown that the method has good robustness for enhanced features. Compared with the traditional Euclidean distance, the IMED takes the spatial relationship of the pixels into consideration, which is not sensitive to minor disturbance, computing simple and easy to be embedded in image recognition technology. Fig.3 shows the comparison of the original image with the embedding IMED image. Assume there are two images x, y, given the size of $M \leq N$,the traditional Euclidean Distance d_E^2 is given by:

$$d_E^2 = \sum_{i=1}^{MN} (x^i - y^i)^2 = (X - Y)^T (X - Y) \tag{8}$$

The IMED is written by:

$$\begin{aligned} d_{IE}^2 &= \sum_{i=1}^{MN} (x^i - y^i) g_{ij} (x^i - y^i) \\ &= (X - Y)^T G (X - Y) \end{aligned} \tag{9}$$

where the $G = g(ij)_{MN*\,NM}$ is symmetric positive semi-definite matrix. The Image Euclidean Distance is equal to the traditional Euclidean distance when the G is identity matrix. According to the requirements of IMED method, the element g_{ij} in the matrix G about pixel site can be calculated by Gaussian function, which is written as:

$g_{ij} = f(|P_i - P_j|) = \frac{1}{2\pi\sigma^2} exp\{-|P_i - P_j|^2 \frac{1}{2\sigma^2}\}$

where $\sigma = 1$ and $|P_i - P_j|$ is the distance between P_i and P_j on the image lattice. From the above description, we know that the value of G only depends on the distance between P_i and P_j, and has nothing to do with gray levels of pixels. So G could be calculated in advance. To improve the efficiency of calculation, Liwei Wang et al proposed a method named ST(Standardizing transform).

Decomposition of matrix G, $G = A^T A = G^{\frac{1}{2}} G^{\frac{1}{2}}$, where the symmetric matrix $G^{\frac{1}{2}}$ is defined uniquely as $G^{\frac{1}{2}} = \Gamma \Lambda^{\frac{1}{2}} \Gamma$. The images x, y are converted into u and v.

$$u = G^{\frac{1}{2}} x, v = G^{\frac{1}{2}} y \tag{10}$$

Therefore, the IMED between x and y can be written as:

$$\begin{aligned} d_{IE}^2 &= (X - Y)^T G (X - Y) \\ &= (X - Y)^T G^{\frac{1}{2}} G^{\frac{1}{2}} (X - Y) \\ &= (U - V)^T (U - V) \end{aligned} \tag{11}$$

The transformed image can be used as the input of other algorithms, that is to embed IMED variables automatically in these algorithms.

B. The Distance Between Two LVM Features

In this paper, the extent of the texture changes and the direction of the change are embedded to IMED separately, and then compare the test samples with the standard samples, finally get the classification of the facial expression according to the distance. The features of a facial expression image I is extracted by LVM, and get the modulus matrix MD and the angle matrix AG. The size of MD and AG are equal to I, and the angle matrix AG is expressed by radians. According to equation (10), we get $md = G^{\frac{1}{2}} MD$ and $ag = G^{\frac{1}{2}} AG$. Thus, the module distance d_{md} and the angle distance d_{ag} between x and y can be defined as:

$$d_{md}^2 = (md_x - md_y)^T (md_x - md_y) \tag{12}$$

$$d_{ag}^2 = (ag_x - ag_y)^T (ag_x - ag_y) \tag{13}$$

For each image, texture changes not only reflected on the modulus and angle independently, but also reflected on the relationship between the modulus and angle. That is to say, not only the distance between the blocks of different textures was calculated, but also the relationship between module and angle cannot be neglected. Therefore, calculate the distance between two LVM features should follow the following properties: If the two textures are same, reduce the distance between them; If the two texture are different, increases the distance between them. The angle of the vector θ in a certain range, $\theta \in [0, \pi)$. So the even decreasing function $f(d_{ag})$ should satisfy the following characteristics:

- $f(d_{ag})$ is an even function and satisfied with $f(d_{ag}) = f(-d_{ag})$.
- $f(d_{ag})$ should be a monotone decreasing function. If the angle between two vectors is small, the distance should be increased by $f(d_{ag})$, otherwise, the distance should be reduced by $f(d_{ag})$.

Here, we choose the cosine function to calculate the distance LVM features d_{LVM}^2:

$$d_{LVM}^2 = d_{md}^2 \cos(d_{ag}) \tag{14}$$

TABLE I.
DIFFERENT SITUATION BASED ON DIFFERENT METHODS

size	content	Embed IMED	No embedded IMED	Embed IMED	No embedded IMED	Embed IMED	No embedded IMED
$16 * 16$	Recognition rate%	69.05	63.81	68.57	64.29	72.86	72.38
$16 * 16$	Recognition times	0.0039	0.0063	0.1104	0.1122	0.0215	0.0247
$32 * 32$	Recognition rate%	71.90	69.52	80.00	73.81	80.59	76.67
$32 * 32$	Recognition times	0.0515	0.0058	0.4938	0.4386	0.1249	0.0796
$64 * 64$	Recognition rate%	69.05	68.10	76.67	73.33	77.14	74.76
$64 * 64$	Recognition times	2.4887	0.0173	4.2268	1.7139	2.7759	0.3677

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The JAFFE database was used for the experiments in this paper, we compared the LVM with the LBP and LDP, and also compared the recognition rate of the algorithm after embedded IMED, the experiments proved that we can embed the IMED algorithm to LVM. All the algorithms in this paper were programmed in Matlab R2009b, and the harder is Intel Core i5-2450M2.5GHz processor and 4G memory computer.

There are 213 images altogether that contain 10 subjects with 7 expressions in JAFFE database. We select 210 images for our experiments. Before the images were used to extract features, we detected each facial expression of the image and normalized to the size of $64 \leq 64$ pixels, $32 \leq 32$ pixels and $16 \leq 16$ pixels with bilinear method. First step, extracting the expression features by LBP, LDP and LVM method respectively, then embedding the IMED into the feature images, and finally using the nearest neighbor classification method to obtain the expression classification. The three different situations of different resolutions obtained by each method are showed in Table 1.

From the data in table 1, we know:

- Take the recognition rate as the standard. No matter the methods were used, when the image size is $32 \leq 32$, the recognition rate of our proposed method is the highest.
- With the facial expression recognition at the size of $32 \leq 32$ as an example. Take the recognition rate and time-consuming into consideration, the algorithm obtained the highest recognition rate when the LVM is embedded with IMED, the recognition rate is 80.59% and the time is 0.1249s. Also with LDP method can get a higher recognition rate, the recognition rate in this method is 80.00% and the recognition time is 0.4938s. Compared with LVM method, LDP method spend more than four times than LVM on computing time. The LBP method spending the shortest time, but the recognition rate is greatly reduced.

According to the above analysis, the LVM method get better expression features than LDP and LBP method in recognition rate and time consuming. That is because the extent of local texture variations and the direction of change information are all taken into consideration as a local texture information. Moreover, from table 1, we known that the computational complexity of LVM method is lower than the LDP method when extracted texture features. And also we can easily recognized that our method achieved a higher recognition rate than the method which not embed IMED from table 1, it demonstrated that LVM can also be embedded with IMED.

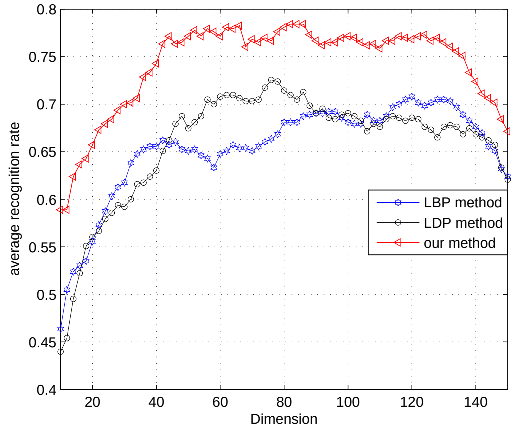

Figure 4. $32 * 32$ resolution expression recognition rate

To illustrate the advantages of our method further, we embed the high-dimensional feature vectors that obtained by LVM into the low-dimensional space by LLE [12] in the process of expression feature extraction. The recognition rates obtained by different methods are shown in Fig.4. From Fig.4, we can observe clearly that recognition rate obtained by LVM method is greatly higher than other two methods.

V. CONCLUSIONS

In this paper, we proposed a new method for facial texture feature extraction named Local Vector Mode(LVM), which can obtain more detailed texture information and spend less time to extract texture features. Moreover, compared with other methods, the recognition rate is greatly increased when use LVM expression method to extract texture characteristics. The experiment results showed that the LVM is a very effective approach and have a better performance than other approaches.

REFERENCES

[1] Lang L Y, Jing X K. "Application of Non-negative sparse matrix factorization in occluded face recognition," *Journal of Computers*, 2011, 6(12): 2675-2679.

[2] Wang S, Yang H, Li H. " Facial Expression Recognition Based on Incremental Isomap with Expression Weighted Distance," *Journal of Computers*, 2013, 8(8).

[3] Kun-Long ZuoWen-Yao Liu. “Facial Expression Recognition Using Active Appearance Models,” *Journal of optoelectronics.Laser*, 2004, 15(7): 853-857.
[4] Shui-Shi LIUYan-Tao TIANChuan WAN. “Facial Expression Recognition Method Based on Gabor Multi-orientation Features Fusion and Block Histogram,” *ACTA AUTOMATICA SINICA*, 2011, 37(12): 1455-1463.
[5] Jian M, Guo H, Liu L. “Texture image classification using visual perceptual texture features and Gabor wavelet features,” *Journal of Computers*, 2009, 4(8): 763-770.
[6] Wei-Hong YanHai-Feng ZhaoBin Luo. “Face recognition based on local binary pattern feature,” *Computer Engineering and Applications*, 2011, 47(18): 170-172.
[7] Jabid TKabir M HChae O S. “Local directional pattern(LDP)for face recognition,” *Consumer Electronics (ICCE), 2010 Digest of Technical Papers International Conference on*, 2010: 239-330.
[8] WANG LZHANG YFENG J. “On the Euclidean distance of image,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(8): 1334-1339.
[9] T.OjalaM.Pietikainen and T.Maenpaa.“Multiresolution Gray Scale and Rotation Invariant Texture Analysis with Local Binary Patterns,” *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2002, 24(7): 971-987.
[10] Rui-Ping Wang Jie Chen Shi-Guang Shan. “Enhancing Training Set for Face Detection Based on SVM,” *Journal of Software*, 2008, 19(11): 2921-2931.
[11] Xiao-Feng Fu Wei WEI. “Facial expression recognition based on multi-scale centralized binary patter,” *Control Theory & Applications*, 2009, 26(6): 629-633.
[12] Roweis SSaul L. “Nonlinear dimensionality reduction by locally linear embedding,” *Science*, 2000, 290(5500):2323-2326.

Yan Wang was born in China 1971. She received the BS, MS in computer since from Xi’an University of Science and Technology, ShanXi,China, in 1996,1999 respectively. She become a lecture, an associate professor, an professor of computer in the College of computer and communication at Lan Zhou University of technology,China, in2001, 2007and2013, respectively. She is an professor in the College of computer and communication at Lan Zhou University of technology, China, her research interests include Database theory and application data mining, pattern recognition, image processing and knowledge engineering.

Guoqing He was born in China 1989. He received the BS, in computer since from Lan Zhou University of technology, Gansu, China in 2011.He is currently a postgraduate in the College of computer and communication at Lan Zhou University of technology. His research interests include, pattern recognition,image processing and Communications Engineering.