Multiscale Adaptive Local Directional Texture Pattern for Facial Expression Recognition

Zhengyan Zhang[1,2], Jingjie Yan[1], Guanming Lu[1], Haibo Li[1], Ning Sun[3], and Qi Ge[1]
[1] College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing, 210003, Jiangsu – P. R. China
[e-mail: zhangzhengyan@just.edu.cn, yanjingjie1212@163.com, lugm@njupt.edu.cn, lihb@njupt.edu.cn, geqi@njupt.edu.cn]
[2] School of Electronics and Information, Jiangsu University of Science and Technology, Zhenjiang, 212003, Jiangsu – P. R. China
[3] Engineering Research Center of Wideband Wireless Communication Technology, Ministry of Education, Nanjing University of Posts and Telecommunications, Nanjing, 210003, Jiangsu – P. R. China
[e-mail: sunning@njupt.edu.cn]
*Corresponding author: Guanming Lu

Received December 7, 2016; revised April 19, 2017; accepted May 28, 2017; published September 30, 2017

Abstract

This work presents a novel facial descriptor, which is named as multiscale adaptive local directional texture pattern (MALDTP) and employed for expression recognition. We apply an adaptive threshold value to encode facial image in different scales, and concatenate a series of histograms based on the MALDTP to generate facial descriptor in term of Gabor filters. In addition, some dedicated experiments were conducted to evaluate the performance of the MALDTP method in a person-independent way. The experimental results demonstrate that our proposed method achieves higher recognition rate than local directional texture pattern (LDTP). Moreover, the MALDTP method has lower computational complexity, fewer storage space and higher classification accuracy than local Gabor binary pattern histogram sequence (LGBPHS) method. In a nutshell, the proposed MALDTP method can not only avoid choosing the threshold by experience but also contain much more structural and contrast information of facial image than LDTP.

This work was supported in part by the National Natural Science Foundation of China under Grant No. 61501249, No. 61071167, No. 41601601 and No. 61471206, the Key Research and Development Program of Jiangsu Province under Grant No. BE2016775, the Natural Science Foundation of Jiangsu Province under Grant No. BK20150855 and No. BK20141428, the Natural Science Foundation for Jiangsu Higher Education Institutions under Grant No. 15KJB510022 and the Postgraduate Innovation Project of Jiangsu Province (No. KYLX16_0660).

https://doi.org/10.3837/tiis.2017.09.020 **ISSN : 1976-7277**

Keywords: Facial expression recognition, multiscale adaptive local directional texture pattern, facial descriptor, Gabor filters

1. Introduction

Facial expression can convey much more information than verbal and vocal cues in human daily interactions [1-4]. With the swift and violent development of computer technology such as the growing popularity of intelligent device, it is essential to enable a machine to understand human emotions and intentions by itself [5, 6]. The facial expression recognition (FER) plays a significant role in establishing a harmonious and friendly man-machine environment, and exhibits a broad application prospects in human-computer interaction (HCI), such as clinical psychology, pain assessment, online distance education and emotion analysis [5, 7, 8]. As a result, the FER has been becoming one of research hotspots in computer vision field during the past decades. Hence, many related efforts have been done by researchers [2, 3, 9-12]. However, pursuing a high-efficiency approache to improve recognition rate of facial expression is still a challenging issue so far.

An automatic FER system usually contains three aspects: detecting and segmenting the face portion, extracting feature from facial expression image and classification for facial expressions [4, 13]. In general, extracting an appropriate feature with a high between-class variance and a lower within-class variance from facial image largely determines the recognition accuracy in the FER [11, 14-16]. There are two ways to extract facial features. One is utilizing deep learning [17-19] to obtain the descriptor of facial image, and the other is directly extracting features from facial image by conventional image processing methods. The former sets up a deep architecture to learn features at multiple levels of representation automatically [20, 21]. However, it takes much time to adjust a large number of parameters and requires large-scale training data. On the other hand, the latter is simple and efficient due to the pixel-level operation, and thus it remains appealing to researches.

Appearance-based method is one of conventional feature extraction approaches, which extracts appearance changes from either the entire face or some particular regions [22-25]. Local Binary Pattern (LBP) [26-28], initially proposed for analyzing image texture, has been successfully adopted in facial expression recognition [29-32]. By comparing the central pixel with its neighborhood pixels, LBP encodes the result into a bit string. Although LBP has the advantage of efficient calculation and robustness against the monotonic illumination change, it is intolerable to non-monotonic illumination variations and random noise [33, 34]. Local Directional Pattern (LDP) [35, 36] encodes the top k directional information computed by Kirsch masks into a 2^k bit string. LDP has better performance than LBP owing to using edge responses instead of intensity, nonetheless, it overlooks the contrast information in the feature descriptor [34]. Gabor Filters [37-41], which extract the detailed features from facial image in multi-scale and multi-orientation, have been widely used in facial image analysis. Local Gabor Binary Pattern Histogram Sequence (LGBPHS) [42] achieves a high recognition rate via combining Gabor filters and LBP to extract the appearance features of facial image, and it is insensitive to noise and facial appearance variation. However, the shortcoming of LGBPHS is the feature vectors have very high dimensions [42]. Recently, local directional texture pattern (LDTP) [43] has been presented for facial expression recognition and scene recognition. LDTP encodes not only directional information but also intensity information of image, so it obtains a better performance than LBP and LDP. Nevertheless, to a certain extent, the recognition rate of LDTP is affected by a threshold value, which is determined by experience in the process of coding.

In this paper, we present a novel facial descriptor, which is named as multiscale adaptive local directional texture pattern (MALDTP), for facial expression recognition. We convolve the facial image with Gabor filters to generate Gabor magnitude response images (GMRIs) in different scales and different orientations. By employing an adaptive threshold value to replace the empirical one of LDTP, we encode facial image by adaptive LDTP (ALDTP) in each scale, and finally concatenate a series of histograms based on the MALDTP to generate the facial descriptor. In this way, our proposed approach not only avoids choosing threshold value by experience but also contains much more structural and contrast information than LDTP, and thus it is robust to noise and illumination variations. Furthermore, in order to evaluate the performance of the MALDTP method, we conduct some person-independent experiments by using Support Vector Machine (SVM) [19] for classification on two remarkable facial expression databases, namely the extended Cohn-Kanade (CK+) database [44] and Japanese Female Facial Expression (JAFFE) database [37].

The rest of paper is organized as follows: Some related works are briefly introduced in Section 2. The proposed MALDTP method is presented in detail in section 3. Section 4 reports the experimental setup. Extensive experiments are conducted and the results are discussed in Section 5. Finally, Section 6 draws a conclusion.

2. Related Work

2.1 Gabor Filters

With good performance in extracting the local spatial and frequency domain information of the object, the Gabor filters have been one of powerful tools for facial image analysis [37-39]. A 2-D Gabor filter can be seen as a complex sinusoidal plane wave multiplied by a Gaussian envelop, which is defined in spatial domain as [45, 46]:

$$\Psi_{u,v}(x,y) = \frac{{f_u}^2}{\pi ab} \mathrm{e}^{-(\frac{{f_u}^2}{a^2}x'^2 + \frac{{f_u}^2}{b^2}y'^2)} \mathrm{e}^{j2\pi f_u x'} \tag{1}$$

$$x' = x\cos\theta_v + y\sin\theta_v \tag{2}$$

$$y' = -x\sin\theta_v + y\cos\theta_v \tag{3}$$

where (x, y) represents the pixel coordinate, a and b are standard deviations of the elliptic Gaussian envelop along the x-axis and the y-axis, and central frequency and orientation of the complex plane wave are determined by f_u and θ_v, respectively [47, 48]. The representation of the Gabor filters is shown in **Fig. 1**.

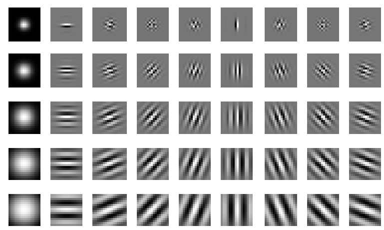

Fig. 1. Representation of Gabor filters. The first column shows the magnitudes of Gabor filter and its real parts are shown in the rest columns.

2.2 Local Directional Texture Pattern (LDTP)

The LDTP encodes intensity difference of the image in the first and second maximum directions, and thus it contains not only intensity information but also directional information.

To obtain the LDTP code, firstly computing the eight absolute edge response values G_i of each pixel by Kirch masks [43, 49]:

$$G_i = |I * M_i|, \quad i = 0,1,\ldots,7 \tag{4}$$

where I is the original image, M_i is the ith Kirch mask, and $*$ denotes the convolution operation.

Then sorting the values G_i to determine the first and second maximum directions. The first maximum direction number D^1 is defined as:

$$D^1 = \arg\max_i \{G_i, 0 \le i \le 7\} \tag{5}$$

The second maximum direction number D^2 is computed in the same way. In each of the two principal directions, calculating the intensity difference of each pixel in its Moore neighborhood by:

$$d^i = \begin{cases} P_{D^i} - P_{D^i+4}, & D^i \in \{0,1,2,3\} \\ P_{D^i} - P_{D^i-4}, & D^i \in \{4,5,6,7\} \end{cases}, \quad i = 1,2 \tag{6}$$

where P_i is the gray value of the original image, and the Moore neighborhood is shown in **Fig. 2**.

P_3	P_2	P_1
P_4	P_c	P_0
P_5	P_6	P_7

Fig. 2. Moore neighborhood. P_c is the gray value of the central pixel and P_0, P_1, …, P_7 are the gray values of its neighborhood pixels.

Then each intensity difference is encoded as:

$$C(d^i) = \begin{cases} 0, & if\ d^i \le |\varepsilon| \\ 1, & if\ d^i < -\varepsilon, \quad i = 1,2 \\ 2, & if\ d^i > \varepsilon \end{cases} \tag{7}$$

where C is the encoded intensity difference, and ε is the empirical threshold value.

Finally, the LDTP code is given by:

$$\mathrm{LDTP}(x, y) = 16D^1(x, y) + 4C(d^1(x, y)) + C(d^2(x, y)) \tag{8}$$

where LDTP(x, y) corresponds to the code for the pixel coordinate (x, y), $D^1(x, y)$ is its maximum directional number, $C(d^1(x, y))$ and $C(d^2(x, y))$ are the max and second encoded intensity differences, respectively.

3. Facial Descriptor Based on the MALDTP

The general framework of the MALDTP is shown in **Fig. 3**, and the concrete procedure of the MALDTP method for facial representation is as follows.

Step 1. Preprocessing for facial image.

Step 2. Obtaining GMRIs by Gabor filters.

Step 3. Encoding the facial image by ALDTP code in each scale.

Step 4. Generation of the MALDTP facial descriptor for classification.

In addition, the following subsections go into detail about the procedure mentioned above.

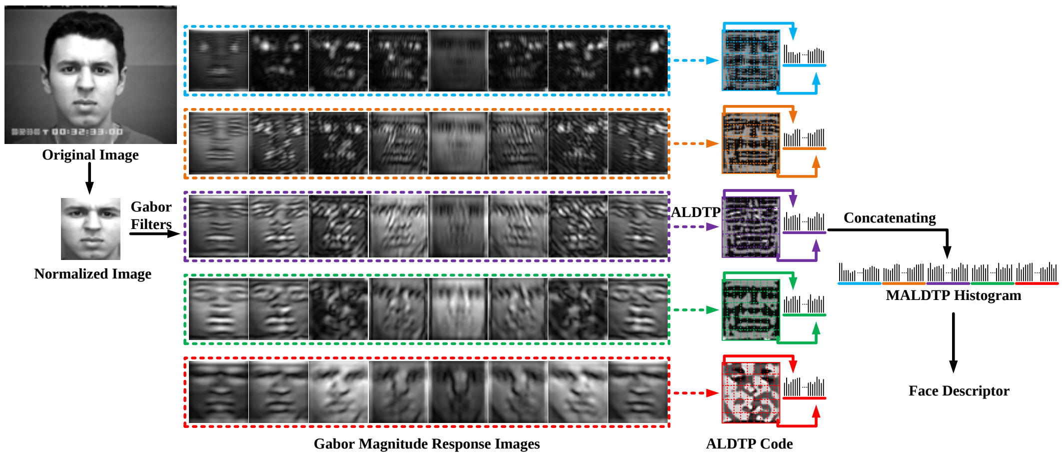

Fig. 3. General framework of the MALDTP method.

3.1 Preprocessing for facial image

In order to minimize the effects caused by background, it is indispensable to preprocess facial images. We utilize the Viola-Jones algorithm [50] to detect the face portion, crop the face region for each facial image from database, and then normalize all the cropped images to 100×100 pixel. This mechanism is beneficial to extract more effective facial features.

3.2 Gabor Magnitude Response Images (GMRIs)

In order to extract proper features, a bank of Gabor filters with five scales and eight orientations in our experiment is designed by the following parameters [41]:

$$a = b = \sqrt{2} \tag{9}$$

$$f_u = (\sqrt{2})^{-u} f_{\max},\ u = 0,1,\cdots,4 \tag{10}$$

$$\theta_v = \frac{v}{8}\pi,\ v = 0,1,\cdots,7 \tag{11}$$

Accordingly, Eq. (1) can be further simplified as [51]:

$$\Psi_{u,v}(x,y) = \frac{f_u^{\,2}}{2\pi} e^{-\frac{f_u^{\,2}}{2}(x^2+y^2)} e^{j2\pi f_u(x\cos\theta_v + y\sin\theta_v)} \tag{12}$$

Consequently, the Gabor response image $G_{u,v}(x, y)$ can be calculated by [40, 50, 51]:

$$G_{u,v}(x,y) = g(x,y) * \Psi_{u,v}(x,y) = \left|G_{u,v}(x,y)\right| e^{i\varphi_{u,v}(x,y)} \tag{13}$$

where $g(x, y)$ is a facial image, $*$ represents the convolution operation, $|G_{u,v}(x, y)|$ and $\varphi_{u,v}(x, y)$ denote the magnitude and phase responses respectively. As the phase responses change drastically while the magnitude responses vary slowly, we compute the MALDTP by the aid of the magnitude responses.

3.3 Adaptive LDTP

After obtaining the 40 GMRIs, we employ an adaptive threshold value instead of empirical one to compute the ALDTP code in each scale with eight directional numbers.

To obtain ALDTP code, we sort the directional values by [43]:

$$D_u^{\,1} = \arg\max_{v}\{|G_{u,v}(x,y)|, 0 \le v \le 7\} \tag{14}$$

where $D_u^{\,1}$ means the maximum directional number in the uth scale, $|G_{u,v}(x, y)|$ is the magnitude response mentioned above. In the same way, we separately compute the other top three directional numbers $D_u^{\,2}$, $D_u^{\,3}$, and $D_u^{\,4}$.

Then, we respectively calculate the intensity difference of each pixel in its Moore neighborhood in the top four directions by:

$$d_u^{\,i} = \begin{cases} P_{D_u^{\,i}} - P_{D_u^{\,i}+4}, & D_u^{\,i} \in \{0,1,2,3\} \\ P_{D_u^{\,i}} - P_{D_u^{\,i}-4}, & D_u^{\,i} \in \{4,5,6,7\} \end{cases}, \quad i = 1,2,3,4 \tag{15}$$

where $d_u^{\,i}$ represents the intensity difference in the ith direction of the uth scale, and P_i is the gray value of the original image.

And then we encode the difference of the first and second maximum directional numbers by [42]:

$$C(d_u^{\,i}) = \begin{cases} 0, & if\ d_u^{\,i} \le |\xi_u| \\ 1, & if\ d_u^{\,i} < -\xi_u \\ 2, & if\ d_u^{\,i} > \xi_u \end{cases}, \quad i = 1,2 \tag{16}$$

where C denotes the encoded intensity difference of each pixel, and ξ_u is the adaptive threshold value in the uth scale, which is defined as:

$$\xi_u = < \frac{1}{4}\left|\sum_{i=1}^{4} d_u^{\,i}\right| > \tag{17}$$

where < > rounds the element to the nearest integer toward zero.

Consequently, the ALDTP code can be calculated by:

$$\mathrm{ALDTP}_u(x,y) = 16D_u^{\,1}(x,y) + 4C(d_u^{\,1}(x,y)) + C(d_u^{\,2}(x,y)) \tag{18}$$

where $\mathrm{ALDTP}_u(x, y)$ corresponds to the code for the pixel coordinate (x, y), $D_u^{\,1}(x, y)$ is its maximum directional number, $C(d_u^{\,1}(x, y))$ and $C(d_u^{\,2}(x, y))$ are the max and second encoded intensity differences, respectively.

As can be seen from the whole coding process, the threshold value is not artificial selected by experience but an adaptive one determined by the average of intensity differences in the top four directions. Simultaneously, the code is more robust to illumination and noise owing to combining the structural information and the contrast information of facial image [42].

3.4 Generation of the MALDTP facial descriptor

As mentioned before, after obtaining the ALDTP code in five scale, we divide encoded image of each scale into 5×9 regions, and extract the histogram $\hbar_u^{\,i}$ from each region by employing each code as a bin [15]. Thereby, the histogram sequence $\hbar_u$ in uth scale can be calculated as:

$$\hbar_u = \bigcup_{i=0}^{44} h_u^{\,i} \tag{19}$$

where U represents the concatenation operation, $h_u^{\,i}$ is the histogram of ith region in uth scale.

Finally, we concatenate the five histogram sequences to generate the MALDTP histogram, **H,** as the facial descriptor:

$$\mathbf{H} = \bigcup_{u=0}^{4} \hbar_u \tag{20}$$

4. Experimental Setup

To evaluate the performance of the MALDTP method, we conduct some experiments on two well-known databases (CK+ [44] and JAFFE [37]) by using LIBSVM [51] with Linear kernel and RBF kernel to classify the facial expressions, where the parameter C for RBF kernel is set to 100. We divide the dataset into training set and test set by person-independent way, which means one individual's expression once belonging to the training set should not appear in the test set and vice versa [52]. According to the literatures, the recognition rate of facial expression in person-independent way is usually lower than person-dependent way. Notwithstanding, the former has the vital practical significance in that human not only can identify expressions of familiar person, but also can recognize the unfamiliar and even unseen person's expressions [53, 54]. In addition, we adopt 10-fold cross-validation testing strategy in our experiment. To be more specific, we divide the dataset into ten group, meanwhile, make sure one person's expressions can not be separated into different groups. Accordingly, we make each group serve as test set once for classification by turn, and the average value of the ten recognition results serves as the final recognition rate.

5. Experimental Results

To test the performance of the MALDTP method, we conducted some experiments of comparison with several other approaches. Moreover, Principal Component Analysis (PCA) [55] was employed to reduce the dimensionality of the features, which is advantageous to decrease the quantity of calculation and improve the recognition rate.

5.1 Results on CK+ Database

The CK+ database [44] consists of 593 sequences from 123 persons with different race, age and gender. However, only 327 sequences in the database carry the expression labels (Anger, Contempt, Fear, Sadness, Disgust, Surprise and Happy). We selected the most expressive images from 325 sequences with correct labels from 118 subjects to construct our experimental database, which contains 1482 expression images with 7 types of facial expression for classification [44].

Table 1 shows the recognition rates of different approaches. Evidently, our proposed method outperforms the others. With linear kernel, the recognition rate is as high as 96.0488% in the 6-class classification problem and 94.9058% in the 7-class classification problem. Compared with LDTP, our proposed method with RBF kernel improves the recognition rate by approximately 2.7% in 6-class problem and 2.4% in 7-class problem. One of the reasons is that our approach takes advantage of Gabor filters to extract more detailed directional and intensity information from different scales. Moreover, we present the confusion matrixes in 6-class and 7-class expression classification problems, as listed in **Table 2** and **Table 3** respectively. As can be seen that with the inclusion of Contempt expression, the recognition rate of Fear expression is from 82.2034% down to 77.1186%.

Table 1. Comparison of recognition rate (%) on CK+ database using person-independent cross-validation

Method	6-class		7-class	
	Linear	RBF (C=100)	Linear	RBF (C=100)
Gabor	89.8657	89.8672	88.7972	88.8575
LBP	90.3632	90.3574	89.1892	89.7881
LDP	90.9617	91.2489	90.8638	90.6610
LDTP	93.1812	93.2496	92.0453	92.3212
LGBPHS	95.0757	95.0747	93.5488	93.4159
MALDTP	**96.0488**	**95.9720**	**94.9058**	**94.7044**

Table 2. Confusion matrix of classification accuracy on CK+ database using SVM with RBF kernel based on the MALDTP in 6-class expression classification problem

	Anger (%)	Disgust (%)	Fear (%)	Happy (%)	Sadness (%)	Surprise (%)
Anger	**97.5490**	1.9608	0	0	0.4902	0
Disgust	2.4691	**97.5309**	0	0	0	0
Fear	2.5424	0.8475	**82.2034**	5.9322	3.3898	5.0847
Happy	0	1.2012	0	**98.7988**	0	0
Sadness	4.3478	3.6232	1.4493	0	**86.9565**	3.6232
Surprise	0	0	0.2801	0	0	**99.7199**

Table 3. Confusion matrix of classification accuracy on CK+ database using SVM with RBF kernel based on the MALDTP in 7-class expression classification problem

	Anger (%)	Disgust (%)	Fear (%)	Happy (%)	Sadness (%)	Surprise (%)	Contempt (%)
Anger	**95.5882**	1.9608	0	0	2.4510	0	0
Disgust	2.4793	**97.5207**	0	0	0	0	0
Fear	3.3898	0	**77.1186**	6.7797	0.8475	7.6271	4.2373
Happy	0	1.2012	0	**98.7988**	0	0	0
Sadness	0.7246	0	2.1739	0	**89.8551**	3.6232	3.6232
Surprise	0	0	0.2801	0	0.0000	**99.7199**	0
Contempt	2.2472	0	5.6180	5.6180	2.2472	1.1236	**83.1461**

5.2 Results on JAFFE Database

The JAFFE database [37] is a free and non-commercial facial expression database, which includes 213 images from ten Japanese female with seven expressions, namely Anger, Neutral, Surprise, Disgust, Fear, Sadness and Happy. Each image in the database has the same resolution of 256×256 pixels with 8-bit grayscale [44].

We compared our proposed method MALDTP with the same approaches used in the CK+ database, as shown in **Table 4**. It is clear that the MALDTP method has the better performance, and the recognition rate with linear kernel achieves 80.7274% in 6-class classification problem and 77.8854% in 7-class classification problem. However, the recognition rate is lower than that in the CK+ database in that some expressions are incorrectly labeled in JAFFE database. Moreover, the confusion matrixes in 6-class and 7-class classification problems are presented in **Table 5** and **Table 6** respectively. The classification accuracy of some facial expressions, such as Anger, Sadness and Surprise, decreases owing to the confusion with the Neutral expression.

Table 4. Comparison of recognition rate (%) on JAFFE database using person-independent cross-validation

Method	6-class		7-class	
	Linear	RBF(C=100)	Linear	RBF (C=100)
Gabor	72.3875	72.4202	71.1430	71.6191
LBP	76.1459	76.1459	71.6603	72.1148
LDP	74.9685	74.9685	72.3207	72.3405
LDTP	78.2922	77.7040	75.6617	75.6617
LGBPHS	78.4143	78.9699	75.3970	75.3970
MALDTP	**80.7274**	**79.6748**	**77.8854**	**77.4308**

Table 5. Confusion matrix of classification accuracy on JAFFE database using SVM with RBF kernel based on the MALDTP in 6-class expression classification problem

	Anger (%)	Disgust (%)	Fear (%)	Happy (%)	Sadness (%)	Surprise (%)
Anger	**93.3333**	0	0	3.3333	3.3333	0
Disgust	13.7931	**65.5172**	6.8966	0	13.7931	0
Fear	9.3750	9.3750	**59.3750**	3.1250	15.6250	3.1250
Happy	0	0	0	**93.5484**	6.4516	0
Sadness	6.4516	6.4516	0	3.2258	**83.8710**	0
Surprise	0	0	6.6667	3.3333	0	**90.0000**

Table 6. Confusion matrix of classification accuracy on JAFFE database using SVM with RBF kernel based on the MALDTP in 7-class expression classification problem

	Anger (%)	Disgust (%)	Fear (%)	Happy (%)	Sadness (%)	Surprise (%)	Neutral (%)
Anger	**86.6667**	10.0000	0	3.3333	0	0	0
Disgust	17.2414	**65.5172**	0	0	17.2414	0	0
Fear	12.5000	9.3750	**59.3750**	3.1250	9.3750	3.1250	3.1250
Happy	0	0	0	**100.0000**	0	0	0
Sadness	0	9.6774	6.4516	3.2258	**74.1935**	3.2258	3.2258
Surprise	0	0	6.6667	3.3333	0	**80.0000**	10.0000
Neutral	6.6667	0	10.0000	0	3.3333	0	**80.0000**

5.3 Further Discussion

In the experiments using RBF kernel for classification, we arbitrarily set a fixed the parameter C rather than choose an optimal one for each approach, in which case we can evaluate the performance of each method by the same classifier. Hence, it can be seen from **Table 1** and **Table 4** that the recognition rate by RBF kernel is sometimes higher, sometimes lower and sometimes equal in comparison with which by linear kernel.

Furthermore, we report the average classification accuracy of single facial expression on CK+ databases in **Table 7**. The average classification accuracy of our proposed method is highest, which achieves 93.7931% in 6-class problem and 91.6782% in 7-class problem. The corresponding experimental results on the JAFFE database are presented in **Table 8**. From these two tables, it can be observed that although the MALDTP proposed method fails to achieve highest recognition rate on all of the single facial expression, the average classification accuracy is higher than that of the others. In contrast to LDTP, our proposed method makes the average classification accuracy increased by approximately 3.6% on CK+ database and 2.4% on JAFFE database respectively.

Additionally, taking into account both our proposed method and LGBPHS are based on the Gabor filters, we implement a further performance evaluation between them，as shown in **Table 9**. We record the execution time of feature extraction on two databases, and calculate the average value by using an Intel® Core™ processor with 3.4 GHz and non-optimized MATLAB code. From **Table 9**, it is obvious that the MALDTP method with less execution time, fewer memory space and higher classification accuracy undoubtedly outperforms than LGBPHS.

Table 7. Average classification accuracy (%) using SVM (RBF) on CK+ database

		Gabor	LBP	LDP	LDTP	LGBPHS	MALDTP
6-class	Anger	83.3333	84.8039	84.3137	87.2549	93.6275	97.5490
	Disgust	93.4156	84.7737	92.5926	97.5309	97.9424	97.5309
	Fear	71.1864	92.3729	82.2034	90.6780	74.5763	82.2034
	Happy	98.1982	98.1982	98.4985	98.7988	98.7988	98.7988
	Sadness	66.6667	65.9420	68.8406	67.3913	87.6812	86.9565
	Surprise	98.5994	98.5994	98.0392	99.1597	100.0000	99.7199
	Average	85.2333	87.4484	87.4147	90.1356	92.1043	**93.7931**
7-class	Anger	79.4118	83.8235	84.8039	87.2549	94.6078	95.5882
	Disgust	95.0413	86.3636	95.8678	95.8678	97.9339	97.5207
	Fear	72.8814	89.8305	84.7458	92.3729	76.2712	77.1186
	Happy	98.4985	97.8979	98.7988	98.7988	98.7988	98.7988
	Sadness	73.9130	65.2174	71.0145	71.7391	86.9565	89.8551
	Surprise	98.8796	97.7591	98.0392	98.8796	100.0000	99.7199
	Contempt	59.5506	77.5281	70.7865	70.7865	66.2921	83.1461
	Average	82.5966	85.4886	86.2938	87.9571	88.6943	**91.6782**

Table 8. Average classification accuracy (%) using SVM (RBF) on JAFFE database

		Gabor	LBP	LDP	LDTP	LGBPHS	MALDTP
6-class	Anger	90.0000	83.3333	90.0000	93.3333	86.6667	93.3333
	Disgust	68.9655	62.0690	65.5172	79.3103	82.7586	65.5172
	Fear	65.6250	68.7500	59.3750	62.5000	56.2500	59.3750
	Happy	80.6452	90.3226	80.6452	93.5484	93.5484	93.5484
	Sadness	48.3871	64.5161	58.0645	51.6129	67.7419	83.8710
	Surprise	83.3333	90.0000	96.6667	90.0000	86.6667	90.0000
	Average	72.8260	76.4985	75.0448	78.3842	78.9387	**80.9408**
7-class	Anger	83.3333	86.6667	80.0000	83.3333	80.0000	86.6667
	Disgust	72.4138	51.7241	72.4138	72.4138	89.6552	65.5172
	Fear	53.1250	56.2500	46.8750	65.6250	59.3750	59.3750

	Happy	90.3226	83.8710	93.5484	93.5484	87.0968	100.0000
	Sadness	48.3871	70.9677	61.2903	58.0645	54.8387	74.1935
	Surprise	76.6667	86.6667	86.6667	86.6667	83.3333	80.0000
	Neutral	76.6667	66.6667	66.6667	70.0000	76.6667	80.0000
	Average	71.5593	71.8304	72.4944	75.6645	75.8522	**77.9646**

Table 9. Comparison the execution time and feature vector length between LGBPHS and MALDTP

Method	Execution Time (*s*)			Feature Vector Length
	JAFFE (213 images)	CK+ (1482 images)	Average (one image)	
LGBPHS	113.1048	809.5622	0.5386	59×40×4×4=37760
MALDTP	34.732	253.7409	0.1671	72×5×5×9=16200

6. Conclusion

In this paper, we present a novel facial descriptor, MALDTP, for facial expression recognition. The MALDTP method uses an adaptive threshold value instead of empirical one to encode the facial image, and it combines much more directional information and the intensity information in different scales. Thereby, the descriptor is more robust against the illumination changes and noise. In addition, we conduct some experiments to evaluate the performance of the MALDTP method. The experimental results show that the MALDTP method achieves higher recognition rate than the others in the tested databases, such as Gabor, LBP, LDP and LDTP. Moreover, compared with the LGBPHS, the MALDTP method owns the advantages of lower computational complexity, fewer storage space and higher classification accuracy. In our future work, we will dedicate to design more robust and discriminative facial descriptors and further improve the recognition rate.

References

[1] Mehrabian, A. and J.A. Russell, "An approach to environmental psychology," *MIT Press*, Cambridge, MA, USA, 1974.

[2] Khatri, N.N., Z.H. Shah, and S.A. Patel, "Facial expression recognition: A Survey," *International Journal of Computer Science & Information Technologies*, vol. 5, no. 1, pp. 149–152, Jan., 2014.

[3] Sumathi, C.P., T. Santhanam, and M. Mahadevi, "Automatic facial expression analysis a survey," *International Journal of Computer Science & Engineering Survey*, vol. 3, no. 6, pp. 47-59, Dec., 2012. Article (CrossRef Link)

[4] Suthar, J. and N. Limbad, "A literature survey on facial expression recognition techniques using appearance based features," *International Journal of Computer Trends & Technology*, vol. 17, no. 4, pp. 161-165, Nov., 2014. Article (CrossRef Link)

[5] Zavaschi, T.H.H., et al., "Fusion of feature sets and classifiers for facial expression recognition," *Expert Systems with Applications*, vol. 40, no. 2, pp. 646–655, Feb., 2013. Article (CrossRef Link)
[6] Tian, Y., T. Kanade, and J.F. Cohn, "Recognizing Action Units for Facial Expression Analysis," *IEEE Transactions on Pattern Analysis & Machine Intelligence*. vol. 23, no. 2, pp. 97-115, Feb., 2001. Article (CrossRef Link)
[7] Tian, Y., T. Kanade, and J.F. Cohn, "Facial expression recognition," in *Proc. of Handbook of face recognition* (*Second Edition*), Springer London, London, UK, pp. 487-519, 2011. Article (CrossRef Link)
[8] Sung, E. and R.E. Mayer, "Five facets of social presence in online distance education," *Computers in Human Behavior*, vol. 28, no. 5, pp. 1738-1747, Sept., 2012. Article (CrossRef Link)
[9] Fasel, B. and J. Luettin, "Automatic facial expression analysis: a survey," *Pattern Recognition*, vol. 36, no. 1, pp. 259-275, Jan., 2003. Article (CrossRef Link)
[10] Samal, A. and P.A. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: a survey," *Pattern Recognition*, vol. 25, no. 1, pp. 65-77, Jan., 1992. Article (CrossRef Link)
[11] Li, Z., et al., "Robust Structured Subspace Learning for Data Representation," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 37, no.10, pp. 2085-2098, Oct., 2015. Article (CrossRef Link)
[12] Z, L. and T. J, "Unsupervised Feature Selection Via Nonnegative Spectral Analysis and Redundancy Control," *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, vol. 24, no. 12, pp. 5343-5355, Dec., 2015. Article (CrossRef Link)
[13] Youssif, A.A.A. and W.A.A. Asker, "Automatic facial expression recognition system based on geometric and appearance features," *Computer & Information Science*, vol. 4, no. 2, pp. 115-124, Mar., 2011. Article (CrossRef Link)
[14] Kan, M., et al., "Multi-view discriminant analysis," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 38, no. 1, pp. 188-194, Jan., 2016. Article (CrossRef Link)
[15] Peltonen, J. and S. Kaski, "Discriminative components of data," *IEEE Transactions on Neural Networks*, vol. 16, no. 1, pp. 68-83, Jan., 2005. Article (CrossRef Link)
[16] Lee, S.H., K.N. Kostas Plataniotis, and M.R. Yong, "Intra-class variation reduction using training expression images for sparse representation based facial expression recognition," *IEEE Transactions on Affective Computing*, vol. 5, no. 3, pp. 340-351, July-Sept., 2014. Article (CrossRef Link)
[17] LeCun, Y., Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436-444, May, 2015. Article (CrossRef Link)
[18] Schmidhuber, J., "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85-117, Jan., 2014. Article (CrossRef Link)
[19] Tang, Y., "Deep learning using linear support vector machines," in *Proc. of International Conference on Machine Learning*, Atlanta, USA, June 16-21, 2013. Article (CrossRef Link)
[20] Bengio, Y., "Learning deep architectures for AI," *Foundations and trends® in Machine Learning*, vol. 2, no. 1, pp. 1-127, Jan., 2009. Article (CrossRef Link)
[21] Cheng, Y., B. Jiang, and K. Jia, "A Deep Structure for Facial Expression Recognition under Partial Occlusion," in *Proc. of Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Kitakyushu, Japan, Aug. 27-29, 2014. Article (CrossRef Link)

[22] Aroussi, Mohamed El, et al., "Local appearance based face recognition method using block based steerable pyramid transform," *Signal Processing*, vol. 91, no. 1, pp. 38-50, Jan., 2011. Article (CrossRef Link)
[23] Zhang, S., X. Zhao, and B. Lei, "Facial expression recognition based on local binary patterns and local fisher discriminant analysis," *Wseas Transactions on Signal Processing*, vol. 8, no. 1, pp. 21-31, Jan., 2012. Article (CrossRef Link)
[24] Shan, C. and T. Gritti. "Learning Discriminative LBP-Histogram Bins for Facial Expression Recognition," in *Proc. of British Machine Vision Conference*, Leeds, UK, Sept. 1-4, 2008. Article (CrossRef Link)
[25] Zhang, Z., "Feature-based facial expression recognition: sensitivity analysis and experiments with a multi-layer perceptron," *International Journal of Pattern Recognition & Artificial Intelligence*, vol. 13, no. 6, pp. 893-911, Sept., 2011. Article (CrossRef Link)
[26] Ojala, T., M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern recognition*, vol. 29, no. 1, pp. 51-59, Jan., 1996. Article (CrossRef Link)
[27] Ojala, T., M. Pietikainen, and D. Harwood, "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions," in *Proc. of* 12*th International Conference on Pattern Recognition*, vol. 1, pp. 582-585, Jerusalem, Israel, Oct. 9-13, 1994. Article (CrossRef Link)
[28] Ojala, T., M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on pattern analysis and machine intelligence*. vol. 24, no. 7, pp. 971-987, July, 2002. Article (CrossRef Link)
[29] Shan, C., S. Gong, and P.W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803-816, May, 2009. Article (CrossRef Link)
[30] Zhang, S., X. Zhao, and B. Lei, "Facial expression recognition based on local binary patterns and local fisher discriminant analysis," *WSEAS Transactions on Signal Process*, vol. 8, no. 1, pp. 21-31, Jan., 2012. Article (CrossRef Link)
[31] Hablani, R., N. Chaudhari, and S. Tanwani, "Recognition of facial expressions using local binary patterns of important facial parts," *International Journal of Image Processing* (*IJIP*), vol. 7, no. 2, pp. 163-170, Apr., 2013.
[32] Shan, C., S. Gong, and P.W. McOwan, "Robust facial expression recognition using local binary patterns," in *Proc. of IEEE International Conference on Image Processing*, Genoa, Italy, Sept. 11-14, 2005. Article (CrossRef Link)
[33] Zhou, H., R. Wang, and C. Wang, "A novel extended local-binary-pattern operator for texture analysis," *Information Sciences*, vol. 178, no. 22, pp. 4314-4325, Nov, 2008. Article (CrossRef Link)
[34] Hasanul Kabir, T.J. and O. Chae, "Local directional pattern variance (LDPv): A robust feature descriptor for facial expression recognition," *International Arab Journal of Information Technology*, vol. 9, no. 4, pp. 382-391, July, 2010.
[35] Jabid, T., M.H. Kabir, and O. Chae, "Local directional pattern (LDP) for face recognition," in *Proc. of the IEEE International Conference on Consumer Electronics*, pp. 329 - 330, Nha Trang, VIETANAM, Aug. 11-13, 2010. Article (CrossRef Link)
[36] Jabid, T., M.H. Kabir, and O. Chae, "Facial expression recognition using local directional pattern (LDP)," in *Proc. of IEEE International Conference on Image Processing*, Hong Kong, China, Sept. 26-29, 2010. Article (CrossRef Link)

[37] Lyons, M., et al., "Coding facial expressions with gabor wavelets," in *Proc. of IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 200-205, Nara, Japan, Apr. 14-16, 1998. Article (CrossRef Link)

[38] Deng, H.-B., et al., "A new facial expression recognition method based on local gabor filter bank and pca plus lda," *International Journal of Information Technology*, vol. 11, no. 11, pp. 86-96, Nov., 2005.

[39] Lyons, M.J., J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 12, pp. 1357-1362, Dec., 1999. Article (CrossRef Link)

[40] Tai, S.L., "Image Representation Using 2D Gabor Wavelets," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 18, no. 10, pp. 959-971, Oct., 1996. Article (CrossRef Link)

[41] Shen, L. and L. Bai, "MutualBoost learning for selecting Gabor features for face recognition," *Pattern Recognition Letters*, vol. 27, no. 15, pp. 1758-1767, Nov., 2006. Article (CrossRef Link)

[42] Zhang, W., et al., "Local Gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition," in *Proc. of IEEE International Conference on Computer Vision*, vol. 1, no. 786-791, Beijing, China, Oct. 17-20, 2005. Article (CrossRef Link)

[43] Rivera, A.R., J.R. Castillo, and O. Chae, "Local directional texture pattern image descriptor," *Pattern Recognition Letters*. vol. 51, no. 1, pp. 94-100, Jan., 2015. Article (CrossRef Link)

[44] Lucey, P., et al., "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 94 - 101, San Francisco, California, USA, June 13-18, 2010. Article (CrossRef Link)

[45] Kamarainen, J., V. Kyrki, and H. Kalviainen, "Fundamental frequency Gabor filters for object recognition," in *Proc. of the* 16*th International Conference on Pattern Recognition*, vol.1, pp. 628–631, Quebec, Canada, Aug. 11-15, 2002. Article (CrossRef Link)

[46] ŠTruc, V., and Pave, Nikola, "Gabor-based kernel partial-least-squares discrimination features for face recognition," *Informatica*, vol. 20, no. 1, pp. 115-138, Jan., 2009. Article (CrossRef Link)

[47] Haghighat, M., S. Zonouz, and M. Abdel-Mottaleb, "CloudID: Trustworthy cloud-based and cross-enterprise biometric identification," *Expert Systems with Applications*, vol. 42, no. 21, pp. 7905-7916, Nov., 2015. Article (CrossRef Link)

[48] Shen, L. and L. Bai, "A review on Gabor wavelets for face recognition," *Pattern Analysis and Applications*, vol. 9, no. 2, pp. 273-292, Oct., 2006. Article (CrossRef Link)

[49] Zeng, H., et al., "Compact Local Directional Texture Pattern for local image description," *Advances in Multimedia*, vol. 2015, no. 8, pp. 1-10, Setp., 2015. Article (CrossRef Link)

[50] Viola, P. and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1-I, pp. I-511-518, Kauai, HI, USA, Dec. 8-14, 2001. Article (CrossRef Link)

[51] Chang, C.-C. and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology* (*TIST*), vol. 2, no. 3, pp. 389-396, Apr., 2011. Article (CrossRef Link)

[52] Xue, M., W. Liu, and L. Li, "Person-independent facial expression recognition via hierarchical classification," in *Proc. of IEEE Eighth International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, pp. 449 - 454, Melbourne, Australia, Apr. 2-5, 2013. Article (CrossRef Link)
[53] Du, Y. and X. Lin, "Mapping Emotional Status to Facial Expressions," in *Proc. of 16th International Conference on Pattern Recognition*, vol. 2, pp. 524- 527, Quebec, Canada, Aug. 11-15, 2002. Article (CrossRef Link)
[54] Du, Y. and X. Lin, "Emotional facial expression model building," *Pattern Recognition Letters*. vol. 24, no. 16, pp. 2923-2934, Dec., 2003. Article (CrossRef Link)
[55] Abdi, H. and L.J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews*: *Computational Statistics*, vol. 2, no. 4, pp. 433–459, July, 2010. Article (CrossRef Link)

Zhengyan Zhang received the BS degree in electronic information engineering and the MS degree in signal and information processing from Jiangsu University of Science and technology, Zhenjiang, Jiangsu, China, in 2004 and 2007, respectively. He is currently pursuing the Ph.D. degree at the College of Telecommunications and Information Engineering in Nanjing University of Posts and Telecommunications. His current research interests include pattern recognition, machine learning and computer vision.

Jingjie Yan received the B.E. degree in electronic science and technology and the M.S. degree in signal and information processing from the China University of Mining and Technology, Beijing, China, in 2006 and 2009, respectively, and the Ph.D. degree in signal and information processing from Southeast University, Nanjing, China, in 2014. Since January 2015, he has been with the College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing, China, as a Lecturer. His current research interests include pattern recognition, affective computing, computer vision, and machine learning.

Guanming Lu received the B.E. degree in radio engineering and the M.S. degree in communication and electronic systems from the Nanjing University of Posts and Telecommunications (NUPT), Nanjing, China, in 1985 and 1988, respectively, and the Ph.D. degree in communication and information systems from Shanghai Jiao Tong University, Shanghai, China, in 1999. He is currently a Professor with the College of Communication and Information Engineering, NUPT. His current research interests include image processing, affective computing, and machine learning.

Haibo Li received the B.E. degree in wireless engineering and the M.S. degree in communication and electronic systems from the Nanjing University of Posts and Telecommunications (NUPT), Nanjing, China, in 1985 and 1988, respectively, and the Ph.D. degree in information theory in 1993 from Linöping University, Linöping, Sweden. He is a Professor of Innovative Media Technology with the KTH Royal Institute of Technology, Stockholm, Sweden. His research interests include mainly media signal processing, including facial and hand gesture recognition and invisible interaction technology.

Ning Sun received the B.S., M.S. and Ph.D. degrees from Guilin University of Electronic Technology, Nanjing Institute of Electronic Technology and Southeast University, in 2000, 2004 and 2007, respectively. Since 2012, he has been with Nanjing University of Posts and Telecommunications, Nanjing, China, where he is currently an Associate Professor in the Engineering Research Center of Wide Band Wireless Communication Technology, Ministry of Education. His current research interests include deep learning, pattern recognition and embedded platform based video analysis.

Qi Ge received Ph.D. degree in Pattern Recognition and Intelligent Systems in Nanjing University of Science and Technology, Nanjing, 2013. Her research interests include pattern recognition, image processing, and image segmentation. She is now working with Nanjing University of Posts and Telecommunications.