Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2018) 26: 1236 – 1248

doi:10.3906/elk-1711-2

Compact local Gabor directional number pattern for facial expression recognition

Zhengyan ZHANG[1,2], Guanming LU[1,*], Jingjie YAN[1], Haibo LI[1], Ning SUN[3], Xia LI[1,4]

[1]School of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing, P.R. China

[2]School of Electronics and Information, Jiangsu University of Science and Technology, Zhenjiang, P.R. China

[3]Ministry of Education, Engineering Research Center of Wideband Wireless Communication Technology, Nanjing University of Posts and Telecommunications, Nanjing, P.R. China

[4]School of Mathematics and Physics, Anhui University of Technology, Maanshan, P.R. China

Received: 01.11.2017 • **Accepted/Published Online:** 06.02.2018 • **Final Version:** 30.05.2018

Abstract: This paper explores a novel method to represent face images for facial expression recognition; it is named compact local Gabor directional number pattern (CLGDNP). By convolving the face images with Gabor filters, we encode the magnitude and phase response images in each scale, and calculate the histograms in several nonoverlapping regions of each encoded image. Finally, we obtain two spatial histogram sequences by the aid of the mean pooling technology and concatenate them to form the facial descriptor. Moreover, for evaluating the performance of the proposed method, we employ a support vector machine to conduct some extensive classification experiments on the Radboud faces database, the extended Cohn-Kanade database, and the Japanese Female Facial Expression database. The experimental results demonstrate that the proposed CLGDNP method achieves better performance in classification.

Key words: Facial expression recognition, compact local Gabor directional number pattern, Gabor filters, feature, support vector machine

1. Introduction

Over the past two decades, as facial expressions can express human emotions and convey their intentions, facial expression recognition (FER) has been widely investigated by researchers [1,2]. Recently, automatic FER has been gradually applied to many fields, such as human–computer interaction (HCI), medical treatment, automated tutoring systems, and human emotion analysis [3,4]. However, as the facial expression is hard to describe accurately, achieving an ideal recognition result remains a difficult problem. Hence, it still leaves much to be desired.

Currently, although deep learning has made great achievements in pattern recognition and artificial intelligence, it is still significantly dependent on high-performance computer hardware and massive training samples [5]. Owing to the small number of images available, we could not follow deep learning-based approaches [6–8]. We intend to pursue a proper and effective facial representation by traditional feature extraction methods without the limitations mentioned above.

Generally, there are two types of facial expression features: geometric and appearance features [4]. The former usually indicates the shapes and locations of facial components, which can be mostly extracted from the geometric relationships between them, such as position, angle, and distance [3]. However, it is hard to

*Correspondence: lugm@njupt.edu.cn

accurately detect and locate these facial components [9]. On the other hand, the appearance features are usually extracted from the changes in face image by filters on the whole face or in some specific regions [3]. There are many existing methods based on appearance features, such as Gabor feature [10–12], local binary pattern (LBP) [9,13], local directional pattern (LDP) [14,15], local directional number pattern (LDN) [16], and local Gabor binary pattern histogram sequence (LGBPHS) [17]. According to the literature, Gabor filters have been widely adopted in face image analysis; they extract the local appearance changes in specific facial regions from different scales and orientations [10]. Moreover, with simple and fast computing capability as well as robustness to the monotonic illumination variations, LBP has been also applied in FER [13]. However, it fails in cases of nonmonotonic illumination changes and random noise [15]. LGBPHS combines Gabor and LBP to extract features and achieves a satisfactory accuracy [17]. Nonetheless, its dimension is so high that it is a heavy burden of computation. LDN proves to be insensitive to noise and has superior performance over Gabor and LBP in that it encodes the directional information as well as facial intensity changes [16]. Nevertheless, the basic length of the LDN code is high, and it fails to consider the relationship between adjacent pixels and lacks more details of the facial texture.

Inspired by Gabor feature and LDN, we propose a novel facial expression descriptor for FER; it is named compact local Gabor directional number pattern (CLGDNP). We encode not only the Gabor magnitude response images (GMRIs) but also Gabor phase response images (GPRIs) in different scales by using a compact LDN. This mechanism brings much more powerful and discriminative features in less space and makes our method robust against noise due to LDN and to illumination due to Gabor filters.

The remainder of this paper is organized as follows. Firstly, we briefly introduce the related work in Section 2. Secondly, we elaborate the proposed CLGDNP method in Section 3. Subsequently, we conduct some experiments and analyze the results in Section 4. Finally, we draw conclusions in the last section.

2. Related work

2.1. Gabor filters

Gabor filters prove to be a useful tool for FER. They are defined in the spatial domain by the following equations [10]:

$$\Psi_{u,v}(x,y) = \frac{f_u^2}{2\pi}\exp\left[-\frac{f_u^2}{2}(x^2+y^2)\right]\exp[j2\pi f_u(x\cos\theta_v + y\sin\theta_v)] \tag{1}$$

$$f_u = (\sqrt{2})^{-u} f_{\max} \tag{2}$$

$$\theta_v = \frac{v}{8}\pi, \tag{3}$$

where f_u and θ_v denote the center frequency and orientation of the complex plane sine wave, respectively. In our experiments, we employ a filter bank with five scales and eight orientations for extracting features [10], that is, $u = 0, 1, \ldots, 4$ and $v = 0, 1, \ldots, 7$.

2.2. Local directional number (LDN) pattern

LDN encodes the directional information and the intensity variations of face images [16]. To obtain the LDN code, firstly we use Kirsch masks to compute the eight edge response values G_i of each pixel by the following

equation:

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

where I is a face image, M_i means the ith Kirsch mask, as shown in Figure 1, and $*$ denotes the convolution operation.

$$\underset{m_0}{\begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix}} \quad \underset{m_1}{\begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix}} \quad \underset{m_2}{\begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}} \quad \underset{m_3}{\begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}}$$

$$\underset{m_4}{\begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{bmatrix}} \quad \underset{m_5}{\begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix}} \quad \underset{m_6}{\begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{bmatrix}} \quad \underset{m_7}{\begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{bmatrix}}$$

Figure 1. Kirsch masks in eight directions.

Then we sort the values G_i to calculate two directional numbers D_1 and D_2, which are followed by

$$D_1(x, y) = \arg\max_i \{G_i \,|\, 0 \le i \le 7\} \tag{5}$$

$$D_2(x, y) = \arg\min_i \{G_i \,|\, 0 \le i \le 7\}, \tag{6}$$

where (x, y) is the pixel position, and $D_1(x, y)$ and $D_2(x, y)$ are the directional numbers of the maximum positive and minimum negative responses, respectively.

Finally, the LDN code is computed as

$$LDN(x, y) = 8 \times D_1(x, y) + D_2(x, y) \tag{7}$$

3. Facial description based on CLGDNP

The framework of the proposed CLGDNP method is shown in Figure 2. In general, we firstly use Gabor filters to filter the face images to obtain 40 GMRIs and 40 GPRIs. Secondly, we encode the GMRIs and GPRIs in each scale to generate CLGDNP_Mag and CLGDNP_Pha maps, respectively. Then we divide each map into small nonoverlapping regions to calculate their local histograms, and employ the mean pooling technology to build the CLGDNP_Mag and CLGDNP_Pha histogram sequences. Finally, we concatenate these two histogram sequences to form the facial descriptor.

3.1. Compact local Gabor directional number pattern (CLGDNP)

For a long time, the Gabor magnitude response has been used as a more useful facial description than the phase response, owing to its stability and slower changes [11]. However, many studies indicate that the Gabor phase response is a powerful discriminative feature and very robust to illumination changes [11,12]. Consequently, we intend to employ the Gabor magnitude and phase information for facial description.

To obtain the CLGDNP code, we employ Gabor filters instead of Kirsch masks to convolve the face images, which extracts much more detailed features in different scales. Given a face image, the Gabor response image $F_{u,v}(x, y)$ is calculated by

$$F_{u,v}(x, y) = I(x, y) * \Psi_{u,v}(x, y) = A_{u,v}(x, y) e^{i\Phi_{u,v}(x,y)} \tag{8}$$

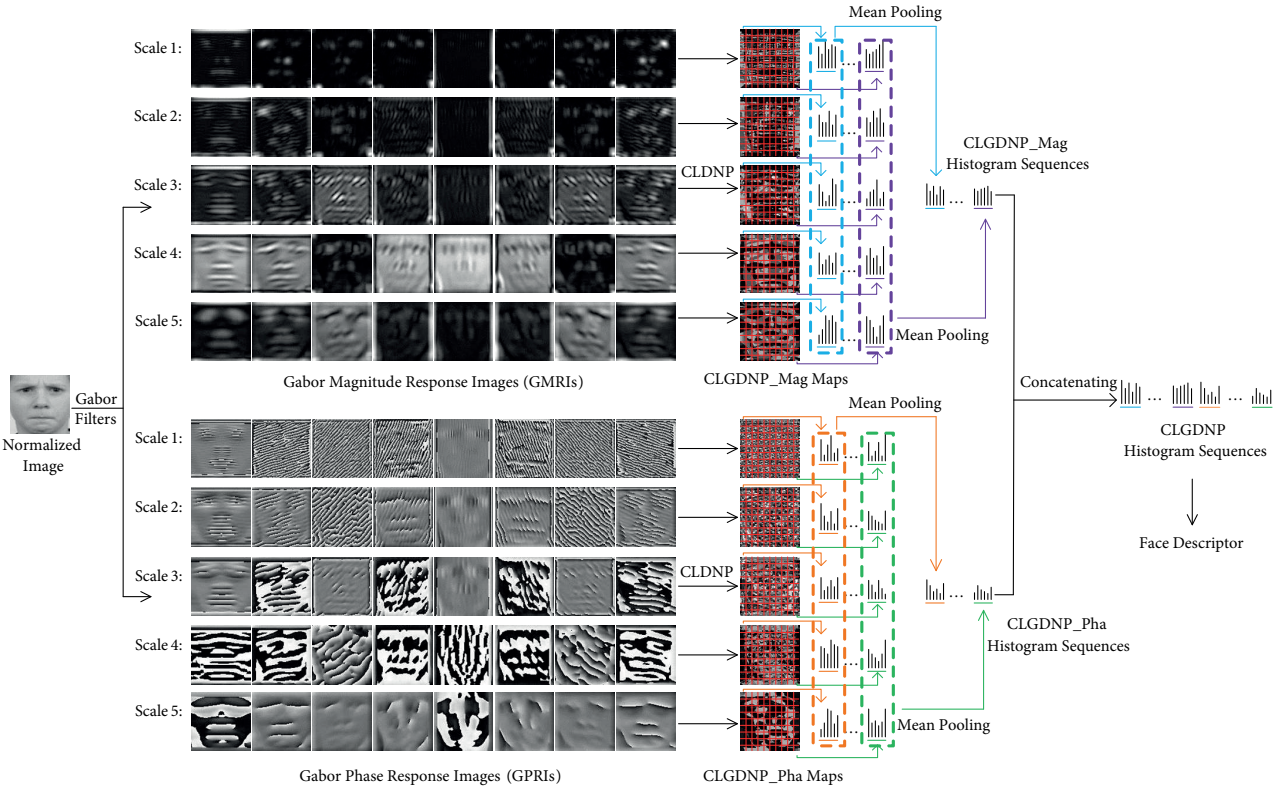


Figure 2. The framework of the proposed CLGDNP method.

$$A_{u,v}(x,y) = \sqrt{Re^2[F_{u,v}(x,y)] + Im^2[F_{u,v}(x,y)]} \tag{9}$$

$$\Phi_{u,v}(x,y) = \arctan\left(\frac{Im[F_{u,v}(x,y)]}{Re[F_{u,v}(x,y)]}\right), \tag{10}$$

where $I(x, y)$ represents a face image, $A_{u,v}(x, y)$ and $\Phi_{u,v}(x, y)$ denote the GMRI and GPRI, and $*$ is the convolution operation.

Then we separately calculate the maximum and minimum directional numbers of the GMRIs and GPRIs in each scale, which are followed by

$$D^1_{A,u}(x,y) = \arg\max_v \{A_{u,v}(x,y) \,|\, 0 \le v \le 7\} \tag{11}$$

$$D^2_{A,u}(x,y) = \arg\min_v \{A_{u,v}(x,y) \,|\, 0 \le v \le 7\} \tag{12}$$

$$D^1_{\Phi,u}(x,y) = \arg\max_v \{\Phi_{u,v}(x,y) \,|\, 0 \le v \le 7\} \tag{13}$$

$$D^2_{\Phi,u}(x,y) = \arg\min_v \{\Phi_{u,v}(x,y) \,|\, 0 \le v \le 7\}, \tag{14}$$

where $(x,\ y)$ is the pixel coordinate, $D^1_{A,u}(x,y)$ and $D^2_{A,u}(x,y)$ denote the maximum and minimum directional numbers in the uth scale of GMRIs, and $D^1_{\Phi,u}(x,y)$ and $D^2_{\Phi,u}(x,y)$ represent the maximum and minimum directional numbers in the uth scale of GPRIs.

In order to reduce the dimension while still contain the structural information of the facial features, we encode the GMRIs and GPRIs in each scale by using the two neighboring pixels with the maximum and minimum values but ignore their sequences. Therefore, the CLGDNP code is defined as

$$CLGDNP\_Mag_u(x,y) = \min\left\{8 \times D^1_{A,u}(x,y) + D^2_{A,u}(x,y),\ D^1_{A,u}(x,y) + 8 \times D^2_{A,u}(x,y)\right\} \tag{15}$$

$$CLGDNP\_Pha_u(x,y) = \min\left\{8 \times D^1_{\Phi,u}(x,y) + D^2_{\Phi,u}(x,y),\ D^1_{\Phi,u}(x,y) + 8 \times D^2_{\Phi,u}(x,y)\right\}, \tag{16}$$

where CLGDNP_Mag$_u(x,\ y)$ denotes the encoded image in the uth scale of GMRIs, and CLGDNP_Pha$_u(x,\ y)$ is the encoded image in the uth scale of GPRIs.

As can be seen from Eqs. (15) and (16), the basic length of the CLGDNP code is only 28, which is much shorter than that of other similar approaches. For instance, the basic length of uniform LBP [9] is 59, of LDP [14] is 56, and of LDN [16] is 56. Moreover, we encode the responses rather than directly use their values, which makes the feature much more stable.

3.2. Facial description

Considering that facial expression variations usually appear in some specific regions of the face, we use a histogram sequence to represent each face, which makes the description contain fine to coarse facial information. We divide CLGDNP_Mag and CLGDNP_Pha maps in each scale into K regions, $\{R_1,\ R_2,\ \ldots,\ R_K\}$. We employ each code as a bin to extract the histograms from the kth region in each scale, which are calculated by

$$h^k_{M,u} = \bigcup_{i=1}^{28} h^k_{M,u,i} = \left\{h^k_{M,u,1},\ h^k_{M,u,2},\ \ldots,\ h^k_{M,u,28}\right\} \tag{17}$$

$$h^k_{P,u} = \bigcup_{i=1}^{28} h^k_{P,u,i} = \left\{h^k_{P,u,1},\ h^k_{P,u,2},\ \ldots,\ h^k_{P,u,28}\right\} \tag{18}$$

Here

$$h^k_{M,u,i} = \sum_{(x,y)\in R_k} \delta\left(CLGDNP\_Mag_u(x,y), C_i\right),\ \ i = 1,\ 2,\ \cdots,\ 28 \tag{19}$$

$$h^k_{P,u,i} = \sum_{(x,y)\in R_k} \delta\left(CLGDNP\_Pha_u(x,y), C_i\right),\ \ i = 1,\ 2,\ \cdots,\ 28 \tag{20}$$

$$\delta(s, C_i) = \begin{cases} 1,\ s = C_i \\ 0,\ otherwise \end{cases}, \tag{21}$$

where $h^k_{M,u}$ and $h^k_{P,u}$ denote the histograms from the kth region in the uth scale on CLGDNP_Mag and CLGDNP_Pha maps, respectively, $\bigcup$ is the concatenation operation, $(x,\ y)$ is the pixel coordinate in the kth region, C_i represents a code value, and $\delta(\bullet)$ is an accumulation function.

For further reducing the dimensions of the final facial descriptor and retaining the detailed information, we adopt the mean pooling technology to build the CLGDNP_Mag and CLGDNP_Pha histogram sequences. Specifically, we calculate the average values of the histograms in the corresponding regions on five scales, and concatenate them to form the histogram sequences of the CLGDNP_Mag and CLGDNP_Pha maps as follows:

$$\hbar_M = \bigcup_{k=1}^{K}\left(\frac{1}{5}\sum_{u=0}^{4} h_{M,u}^{k}\right) = \left\{\frac{1}{5}\sum_{u=0}^{4} h_{M,u}^{1}, \frac{1}{5}\sum_{u=0}^{4} h_{M,u}^{2}, \cdots, \frac{1}{5}\sum_{u=0}^{4} h_{M,u}^{K}\right\} \tag{22}$$

$$\hbar_P = \bigcup_{k=1}^{K}\left(\frac{1}{5}\sum_{u=0}^{4} h_{P,u}^{k}\right) = \left\{\frac{1}{5}\sum_{u=0}^{4} h_{P,u}^{1}, \frac{1}{5}\sum_{u=0}^{4} h_{P,u}^{2}, \cdots, \frac{1}{5}\sum_{u=0}^{4} h_{P,u}^{K}\right\}, \tag{23}$$

where $\hbar_M$ and $\hbar_P$ denote the histogram sequences of CLGDNP_Mag and CLGDNP_Pha maps, respectively, $\bigcup$ is the concatenation operation, and K is the number of the regions.

Finally, we concatenate the two histogram sequences, $\hbar_M$ and $\hbar_P$, to form the feature vectors $\mathbf{H}$ as the facial descriptor:

$$\mathbf{H} = \hbar_M \bigcup \hbar_P \tag{24}$$

4. Experiments and results

4.1. Experimental setup

To evaluate the performance of the proposed method, we conduct some extensive classification experiments by LIBSVM [18] on three facial expression databases, namely the Radboud faces database (Rafd) [19], the extended Cohn-Kanade (CK+) database [20], and the Japanese Female Facial Expression (JAFFE) database [21]. We adopt a 10-fold cross-validation testing mechanism in a person-independent way [22]. To be specific, we partition the database into ten groups, and simultaneously ensure one subject's expressions belong to the same group. Each group serves as the test set once by turns.

In our experiments, we crop the face portion detected by the Viola–Jones algorithm [23], normalize all the cropped images to 100 × 100 pixels, and divide each image into 9 × 10 nonoverlapping regions. Additionally, we use PCA [24] with 99% of variance retained to decrease the dimension of the feature vectors and employ a support vector machine (SVM) with RBF kernels to classify the facial expressions, where the parameter C of RBF kernel is 100.

4.2. Experiments on Rafd database

The Rafd [19] includes 8040 face images of 67 models. Each model shows eight expressions with three gaze directions, as shown in Figures 3a and 3b. Moreover, each image in the database has five camera angles (0°, 45°, 90°, 135°, and 180°), as shown in Figure 3c. In our experiments, we selected the 45°-camera, 90°-camera, and 135°-camera images as three subdatabases, described as Rafd (45°), Rafd (90°), and Rafd (135°). Accordingly, each subdatabase contains 1608 facial expression images.

For testing the effect of the combination of magnitude and phase information, we separately use CLGDNP_Mag and CLGDNP_Pha histogram sequences as the facial descriptor for classification and report the results in Table 1. It is clear that the magnitude information and phase information are complementary and their combination effectively improves the recognition rates.

Figure 3. Examples from the Rafd. (a) Examples of the eight expressions. From the top left: surprise, anger, sadness, disgust, fear, happiness, neutral, and contempt. (b) Examples of three gaze directions. From the left: looking left, looking frontal, looking right. (c) Examples of five camera angels in the order: 0°, 45°, 90°, 135°, and 180°.

Table 1. Recognition rates (%) of CLGDNP_Mag, CLGDNP_Pha, and CLGDNP on the three subdatabases of Rafd.

Method	Rafd (45°)		Rafd (90°)		Rafd (135°)	
	6-class	8-class	6-class	8-class	6-class	8-class
CLGDNP_Mag	93.25	87.30	97.05	93.75	94.23	88.58
CLGDNP_Pha	92.95	84.78	98.33	94.53	94.07	86.73
CLGDNP	96.10	91.39	99.09	97.01	96.27	92.53

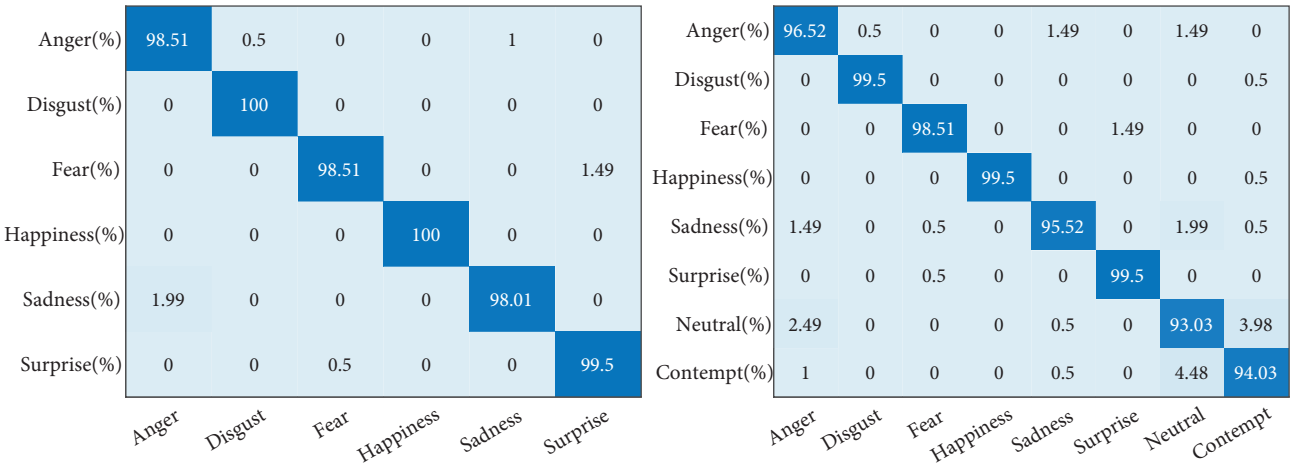
Moreover, we compare the classification performance with several other approaches, as listed in Table 2. The local descriptor, such as LBP, LDP, and LDN, extracts the facial feature from a single resolution. Gabor feature and LGBPHS consider the Gabor magnitude information but ignore the phase information. However, by a compact LDN, the proposed CLGDNP method extracts the features in different scales, which are not only from the Gabor magnitude but also from the Gabor phase. Thus, the proposed CLGDNP method gives much more structural information of the facial texture and achieves the highest recognition rates both in 6-class and 8-class classification problems. Moreover, we observe that the recognition rates of Rafd (90°) are higher than those of Rafd (45°) and Rafd (135°). One of the reasons is that the face images in Rafd (90°) are frontal and have much more facial details.

Additionally, it can be observed that the recognition rate in the 8-class classification problem is lower than that in the 6-class classification problem owing to the addition of neutral and contempt expressions. For example, we present the confusion matrices of the 6-class and 8-class classification problems on Rafd (90°) in Figure 4. It is clear that the recognition rates of some basic expressions (such as anger, disgust, happiness, and sadness) decline in that these expressions are confused with the neutral and contempt expressions.

Furthermore, we conduct some dedicated experiments on three subdatabases to investigate the effect of PCA on the recognition rates, and the results are depicted in Figure 5. We intend to retain a different number of variance in PCA, varying from 85% to 99%. It can be obviously seen from Figure 5 that the proposed method outperforms the other approaches both in 6-class and 8-class problems on three subdatabases.

Table 2. Comparison of recognition rates (%) on the three subdatabases of Rafd.

Method	Rafd (45°)		Rafd (90°)		Rafd (135°)	
	6-class	8-class	6-class	8-class	6-class	8-class
Gabor	87.92	79.06	95.57	89.78	85.61	80.17
LBP	91.96	84.18	93.15	88.24	91.67	84.71
LDP	88.10	80.43	95.15	86.27	90.69	82.14
LDN	92.33	85.24	96.26	93.02	91.81	84.03
LGBPHS	95.69	91.26	98.55	94.84	94.13	90.19
CLGDNP	96.10	91.39	99.09	97.01	96.27	92.53



(a) 6-class problem (b) 8-class problem

Figure 4. Confusion matrices of the 6-class and 8-class problems on the Rafd (90°).

In addition, we also report the average results of these fifteen recognition rates in Table 3. It clearly indicates that the CLDGNP method achieves the highest average recognition rate. Especially, compared with LGBPHS, the proposed CLDGNP method improves the average recognition rate by approximately 3% in the 6-class problem and 5% in the 8-class problem. Moreover, we compare the execution time for extracting features of each image between them by using an Intel i3 CPU with 3.4 GHz in that they are both based on Gabor filters. The proposed method takes roughly 0.21 s on average while the LGBPHS takes about 0.53 s. Consequently, our proposed method has high performance and low computation cost.

4.3. Experiments on the CK+ database

The CK+ database [20] consists of 593 sequences from 123 persons. There are six basic facial expressions (anger, disgust, fear, happiness, sadness, surprise) and one contempt expression in the database. The facial expressions in each sequence are from the neutral to the peak expression [20]. There are only 325 of 593 sequences with correct labels from 118 subjects in the CK+ database, and thus we choose the most expressive images of them, which resulted in 1482 face images in our experiment.

According to the recognition rates listed in Table 4, the proposed method performs better in both the 6-class and 7-class classification problems. Moreover, we present the confusion matrix of the 7-class expression

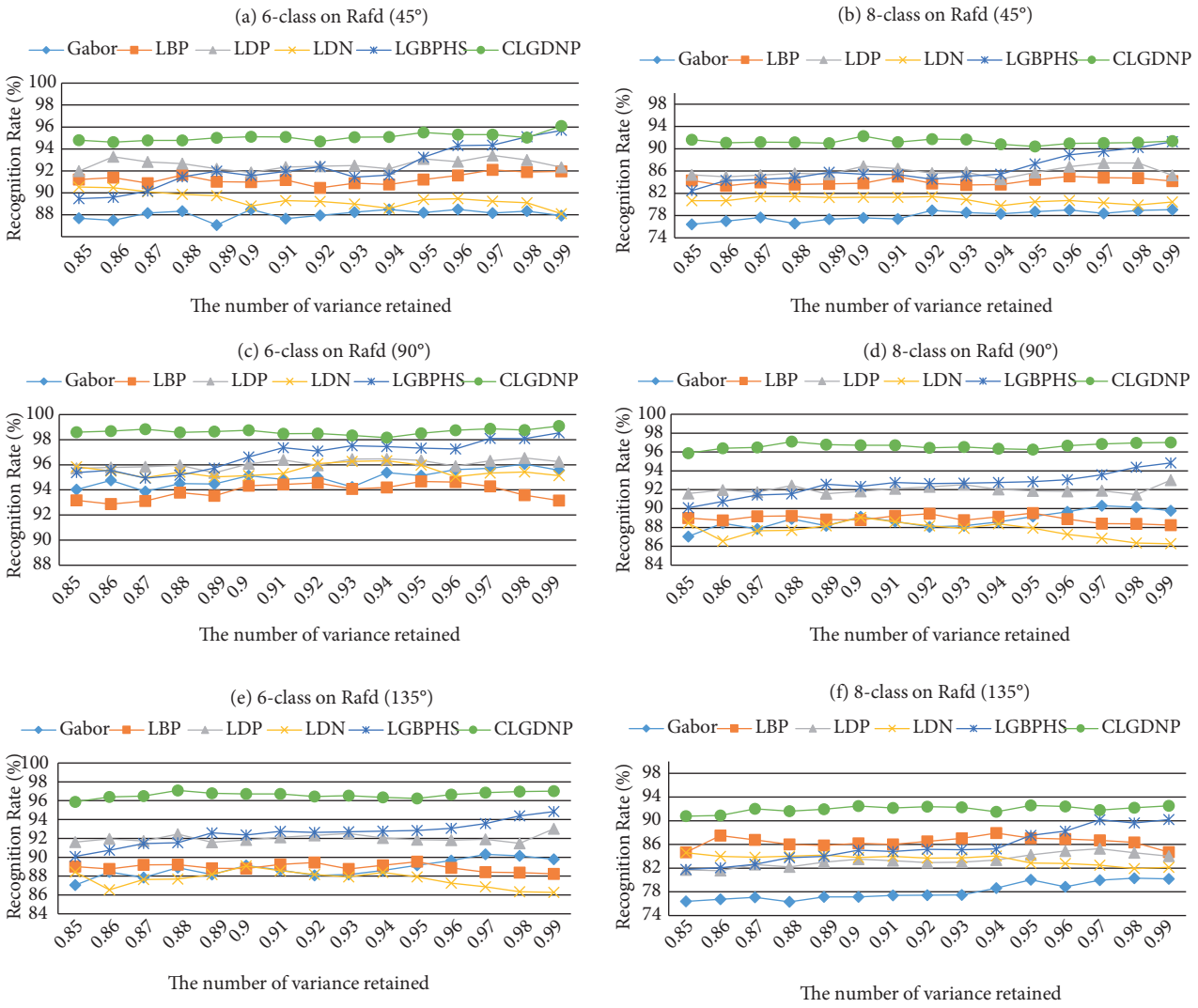


Figure 5. Recognition rates using PCA with a different number of variance retained.

Table 3. The average recognition results (%) using PCA with a different number of variance retained on the three subdatabases of Rafd.

Method	Rafd (45°)		Rafd (90°)		Rafd (135°)	
	6-class	8-class	6-class	8-class	6-class	8-class
Gabor	88.04 ± 0.42	77.98 ± 0.90	94.96 ± 0.65	88.80 ± 0.91	85.74 ± 1.07	78.07 ± 1.45
LBP	91.27 ± 0.48	84.10 ± 0.58	93.89 ± 0.62	88.92 ± 0.38	92.12 ± 0.55	86.40 ± 0.90
LDP	92.59 ± 0.47	85.90 ± 0.89	96.09 ± 0.34	92.02 ± 0.41	91.00 ± 0.73	83.35 ± 1.10
LDN	89.39 ± 0.67	80.79 ± 0.55	95.52 ± 0.46	87.68 ± 0.86	90.97 ± 0.52	83.48 ± 0.82
LGBPHS	92.28 ± 1.91	86.32 ± 2.53	96.81 ± 1.67	92.56 ± 1.25	91.68 ± 1.66	85.69 ± 2.82
CLGDNP	95.09 ± 0.37	91.23 ± 0.45	98.64 ± 0.23	96.61 ± 0.32	96.39 ± 0.33	91.97 ± 0.56

classification problem in Figure 6. It can be seen that the recognition rates of sadness and contempt expressions are slightly below 80%.

Table 4. Recognition rates (%) on the CK+ and JAFFE database.

Method	CK+ database		JAFFE database	
	6-class	7-class	6-class	7-class
Gabor	89.87	88.86	72.42	71.62
LBP	90.36	89.79	76.15	72.11
LDN	91.76	89.90	77.41	74.22
LDP	91.25	90.66	74.97	72.34
LGBPHS	95.07	93.42	78.97	75.40
CLGDNP	95.27	94.27	79.04	77.88

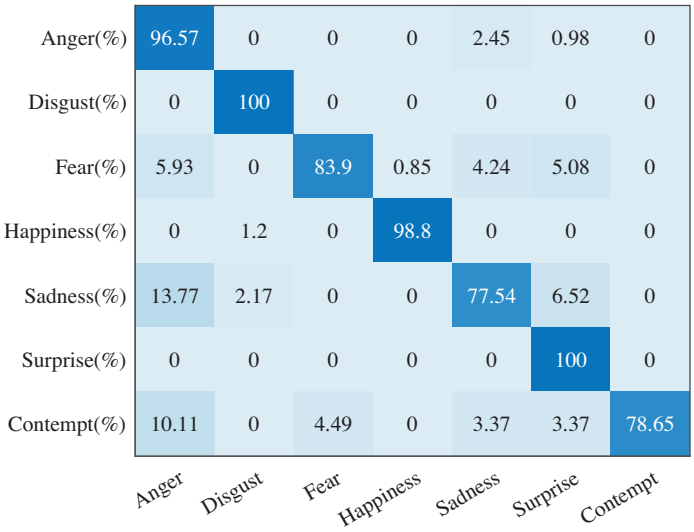


Figure 6. Confusion matrix of the 7-class problem on the CK+ database.

Additionally, the comparison between our proposed method and several other state-of-the-art methods on the CK+ database is shown in Table 5. We choose these methods owing to their similar testing protocols (e.g., person-independent classification). It can be observed that the proposed method has a comparable or even better performance compared to the state-of-the-art methods.

Table 5. Comparison of recognition rate with the state-of-the-art methods on the CK+ database.

Year	Method	Cross validation	6-class	7-class
2014	3DCNN+DAP [6]	15-fold	–	92.4%
2014	STM-ExpLet [24]	10-fold	–	94.19%
2016	deep neural networks [7]	5-fold	93.2%	–
2017	Intra-class variation reduction [25]	10-fold	–	89.6%
2017	3D Inception-ResNet+landmarks [8]	5-fold	–	93.21%
	CLGDNP	10-fold	95.27%	94.27%

4.4. Experiments on the JAFFE database

The JAFFE database [21] includes 213 images from ten different Japanese women, which are labeled by six basic expressions and one neutral expression. For exposing the expressive zones on the face, everyone's head is frontal and the hair is tied back [21].

It can be seen from Table 4 that the proposed method still has better performance than the others. Moreover, we observe that the recognition rate on the CK+ database is higher than that on the JAFFE database. One of the reasons is that the expression images selected from the CK+ database are all correctly labeled, while some expression images from the JAFFE database are probably labeled by mistake, as shown in Figure 7.

(a) (b)

Figure 7. Examples of original expression images from the JAFFE database. (a) and (b) are titled "KR.HA1.74.tiff" and "KR.SA3.79.tiff" respectively. The expressions in the two images are so similar that there is almost no obvious difference between them. However, they belong to different types. (a) is labeled happiness while (d) is labeled sadness.

5. Conclusion

In this work, we propose a novel facial representation, CLGDNP, that encodes not only the GMRIs but also the GPRIs in different scales to extract the structural information of the facial texture. In order to reduce the dimension of the features, we use the mean pooling technology to concatenate the feature vectors. Thereby, it is nonsusceptible to noise and illumination change. The experimental results on three public databases manifest that our proposed method achieves higher classification accuracy.

In our method, we simply concatenate the CLGDNP_Mag and CLGDNP_Pha histogram sequences to form the facial descriptor. A better combination method might improve the classification performance. In the future, we will dedicate further study to the data fusion methods and improve the recognition accuracy and efficiency, especially in the nonfrontal facial expression condition.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant No. 61501249, No. 61071167, 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 and No. KYLX15_0827).

References

[1] Zhang S, Wu ZY, Meng HM, Cai LH. Facial expression synthesis based on emotion dimensions for affective talking avatar. In: Nishida T, Jain LC, Faucher C, editors. Modeling Machine Emotions for Realizing Intelligence. Berlin, Germany: Springer, 2010. pp. 109-132.

[2] Calvo MG, Nummenmaa L. Perceptual and affective mechanisms in facial expression recognition: an integrative review. Cogn Emot 2016; 30: 1081-1106.

[3] Zavaschi THH, Britto AS Jr , Oliveira LES, Koerich AL. Fusion of feature sets and classifiers for facial expression recognition. Expert Syst Appl 2013; 40: 646-655.

[4] Tian Y, Kanade T, Cohn JF. Facial expression recognition. In: Li SZ, Jain AK, editors. Handbook of Face Recognition. London, UK: Springer, 2011. pp. 487-519.

[5] Wang M, Hu ZP, Sun Z, Zhu M, Sun M. Sample group and misplaced atom dictionary learning for face recognition. Turk J Elec Eng & Comp Sci 2017; 25: 4421-4430.

[6] Liu MY, Li SX, Shan SG, Wang RP, Chen XL. Deeply learning deformable facial action parts model for dynamic expression analysis. In: Asian Conference on Computer Vision; 1–5 November 2014; Singapore. Springer. pp. 143-157.

[7] Mollahosseini A, Chan D, Mahoor MH. Going deeper in facial expression recognition using deep neural networks. In: 2016 IEEE Winter Conference on Applications of Computer Vision; 7–10 March 2016; Lake Placid, NY, USA. New York, NY, USA: IEEE. pp. 1-10.

[8] Hasani B, Mahoor BH. Facial expression recognition using enhanced deep 3D convolutional neural networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops; 21–26 July 2017; Honolulu, HI, USA. New York, NY, USA: IEEE. pp. 2278-2288.

[9] Shan CF, Gong SG, McOwan PW. Facial expression recognition based on local binary patterns: a comprehensive study. Image Vision Comput 2009; 27: 803-816.

[10] Shen LL, Bai L. MutualBoost learning for selecting Gabor features for face recognition. Pattern Recogn Lett 2006; 27: 1758-1767.

[11] Bereta M, Pedrycz W, Reformat M. Local descriptors and similarity measures for frontal face recognition: a comparative analysis. J Vis Commun Image R 2013; 24: 1213-1231.

[12] Qing LY, Shan SG, Chen XL, Gao W. Face recognition under varying lighting based on the probabilistic model of Gabor phase. In: 18th International Conference on Pattern Recognition (ICPR); 20–24 August 2006; Hong Kong, China. New York, NY, USA: IEEE. pp. 1139-1142.

[13] Ahonen T, Hadid A, Pietikainen M. Face description with local binary patterns: Application to face recognition. IEEE T Pattern Anal 2006; 28: 2037-2041.

[14] Jabid T, Kabir MH, Chae O. Robust facial expression recognition based on local directional pattern. ETRI J 2010; 32: 784-794.

[15] Kabir H, Jabid T, Chae O. Local directional pattern variance (LDPv): a robust feature descriptor for facial expression recognition. Int Arab J Inf Techn 2012; 9: 382-391.

[16] Rivera AR, Castillo JR, Chae O. Local directional number pattern for face analysis: face and expression recognition. IEEE T Image Process 2013; 22: 1740-1752.

[17] Zhang WC, Shan SG, Gao W, Chen XL, Zhang HM. Local Gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition. In: IEEE International Conference on Computer Vision; 17–21 October 2005; Beijing, China. New York, NY, USA: IEEE. pp. 786-791.

[18] Chang CC, Lin CJ. LIBSVM: a library for support vector machines. ACM T Intel Syst Tec 2011; 2: 1-27.

[19] Langner O, Dotsch R, Bijlstra G, Wigboldus DH, Wigboldus DHJ. Presentation and validation of the Radboud faces database. Cogn Emot 2010; 24: 1377-1388.

[20] Lucey P, Cohn JF, Kanade T, Saragih J, Ambadar Z, Matthews I. The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops; 13–18 June 2010; San Francisco, CA, USA. New York, NY, USA: IEEE. pp. 94-101.

[21] Lyons M, Akamatsu S, Kamachi M, Gyoba J. Coding facial expressions with Gabor wavelets. In: IEEE International Conference on Automatic Face and Gesture Recognition; 14–16 April 1998; Nara, Japan. New York, NY, USA: IEEE. pp. 200-205.

[22] Xue ML, Liu WQ, Li L. Person-independent facial expression recognition via hierarchical classification. In: IEEE 8th International Conference on Intelligent Sensors, Sensor Networks and Information Processing; 2–5 April 2013; Melbourne, VIC, Australia. New York, NY, USA: IEEE. pp. 449-454.

[23] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 8–14 December 2001; Kauai, HI, USA. New York, NY, USA: IEEE. pp. I-511-518.

[24] Abdi H, Williams LJ. Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics 2010; 2: 433-459.

[25] Liu MY, Shan SG, Wang RP, Chen XL. Learning Expressionlets on Spatio-temporal Manifold for Dynamic Facial Expression Recognition. In: IEEE Conference on Computer Vision and Pattern Recognition; 23–28 June 2014; Columbus, OH, USA. New York, NY, USA: IEEE. pp. 1749-1756.

[26] Lee SH, Yong MR. Intra-class variation reduction using training expression images for sparse representation based facial expression recognition. IEEE T Affect Comput 2017; 5: 340-351.