RECOGNITION OF FACE EXPRESSIONS USING LOCAL PRINCIPAL TEXTURE PATTERN

Adin Ramirez Rivera, Jorge A. Rojas Castillo, and Oksam Chae

Department of Computer Engineering, Kyung Hee University, South Korea

{adin, jarc, oschae}@khu.ac.kr

ABSTRACT

Deriving an effective facial feature from original face images is a vital step for a successful automatic facial expression recognition. In this paper, we proposed a new feature descriptor, Local Principal Texture Pattern (LPTP), for expression recognition. We compute the LPTP feature, at each pixel, by extracting the principal directions of the local neighborhood, and we code the intensity differences on these directions. The mixture of direction and contrast information makes our descriptor robust against rotation and illumination changes. Consequently, we represent each expression as a distribution of LPTP codes. Our experiments demonstrate the superiority of the proposed feature, on two facial expression databases, over the existing methods.

Index Terms— Image representation, face expression recognition, local principal texture pattern, feature extraction

1. INTRODUCTION

Facial expressions are the more natural and immediate way of communication of human beings. Hence, several fields, such as, human-computer interaction, data driven animation, and video indexing, need automatic facial expression recognition to detect and analyze human emotions and intentions. Due to the extended necessity, facial expression recognition has gained much importance in the computer vision field [1, 2]. Regardless of this progress, achieving a high recognition accuracy is still a challenging task.

The objective of facial feature extraction is to acquire face characteristics which provide an efficient and robust recognition. There are two common approaches to extract facial features: geometric feature-based and appearance-based methods [2]. The former [3, 4] encodes the shape and locations of different facial components, which are combined into a feature vector that represents the face. However, the geometric feature-based methods usually requires accurate and reliable facial feature detection and tracking, which is difficult to accommodate in many situations. The appearance-based methods [5, 6] use image filters, either on the whole-face, to create holistic features, or some specific face-region, to create local features, to extract the appearance change in the face image. The performance of the appearance-based methods is excellent in constrained environment but its performance degrades in environmental variation [7].

Local Binary Pattern (LBP) [8] has been successfully applied as a local feature extraction method in facial expression recognition [9]. Despite the robustness to monotonic illumination of LBP, it is sensitive to non-monotonic illumination variation, and shows poor performance in presence of random noise [10]. A directional pattern (LDP) [11] has been proposed to overcome the limitation of LBP. However, it suffer in noisy conditions and is sensitive to rotations. Moreover, it cannot detect different transitions in the intensity regions. In this paper, we proposed a novel feature descriptor, LPTP, that takes advantage of both directional and intensity information. The combination of both features outperforms the singled-feature counterparts, LBP and LDP. Additionally, the proposed method extracts the information from the principal textures of the neighborhood. The performance of proposed LPTP feature is evaluated with a machine learning method, Support Vector Machine (SVM), on two different databases.

2. LOCAL PRINCIPAL TEXTURE PATTERN

LBP feature labels each pixel by thresholding a set of sparse points of its circular neighborhood, and encodes that information in a string of bits. Similarly, LDP encodes the principal directions of each pixel's neighborhood into a eight bit string. However, these methods choose a fixed start position for the code creation. This *ad hoc* construction overlooks the prominent information in the neighborhood. In contrast, we create a code from the principal directions of the local neighborhood, and extract the contrast information from these directions. Consequently, we code the principal direction and the intensity difference of the two principal directions into one number. This approach allows us to encode the important texture information of the neighborhood that is revealed by its prominent directions. Figure 1 shows an abstraction of the proposed encoding scheme.

We calculate the principal directions of the neighborhood using the Kirsch compass masks—in eight different directions as shown in fig. 2. Therefore, we compute the absolute value of the eight Kirsch mask's responses, $\{M_0, \ldots, M_7\}$, applied to a particular pixel, and take the two greatest responses. These directions indicate the principal axis of the

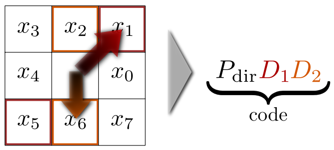

Fig. 1. LPTP codes the principal direction, P_{dir}, of the local texture, and the differences in the principal, D_1, and secondary direction, D_2.

$$\underset{M_0}{\begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix}} \underset{M_1}{\begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix}} \underset{M_2}{\begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}} \underset{M_3}{\begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}}$$

$$\underset{M_4}{\begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{bmatrix}} \underset{M_5}{\begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix}} \underset{M_6}{\begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{bmatrix}} \underset{M_7}{\begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{bmatrix}}$$

Fig. 2. Kirsch compass masks.

local texture. In each of the two principal directions, we compute the intensity difference of the opposed pixels in the neighborhood. This local difference, is equivalent to the local threshold that LBP does. Unlike the LBP binary encoding, we encode the difference using three levels (negative, equal, and positive), which create a more distinctive code for the neighborhood. Then each difference is encoded as

$$D = \begin{cases} 0, & \text{if } -\varepsilon \le d \le \varepsilon \\ 1, & \text{if } d < -\varepsilon \\ 2, & \text{if } d > \varepsilon, \end{cases} \tag{1}$$

where D is the encoded intensity difference, d is the actual intensity difference, ε is a threshold value (in our experiments we use $\varepsilon = 15$).

Consequently, the code is created by concatenating the binary form of the principal direction, and the two differences. This concatenation can be represented by the following operation

$$\text{LPTP}(x, y) = 16 \times P_{\text{dir}}^{(x,y)} + 4 \times D_1^{(x,y)} + D_2^{(x,y)}, \tag{2}$$

where $\text{LPTP}(x, y)$ is the code for the pixel (x, y), $P_{\text{dir}}^{(x,y)}$ is the index of the principal direction (from 0 to 7) of the neighborhood of the pixel (x, y), and $D_1^{(x,y)}$ and $D_2^{(x,y)}$ are the first and second coded differences (using Eq. (1)) of the neighborhood of the pixel (x, y), respectively.

For example, consider the neighborhood shown in fig. 3a, we compute the Kirsch mask responses in the neighborhood, and we show them in their respective orientation in fig. 3b. The principal, M_1, and the secondary, M_6, directions are shown in red and blue, respectively. Then, we compute the intensity difference of the corresponding pixel intensities in these directions (as shown by the colored pairs in fig. 3a). In this case, the differences are: $d_1 = 143 - 133 = 10$, and $d_2 = 137 - 141 = -4$, which are transformed with Eq. (1) into $D_1 = 0$, and $D_2 = 0$, assuming a threshold $\varepsilon = 15$.

$$\underset{\text{(a)}}{\begin{bmatrix} 139 & 141 & 143 \\ 134 & 140 & 145 \\ 133 & 137 & 133 \end{bmatrix}} \quad \underset{\text{(b)}}{\begin{bmatrix} -3 & 69 & 117 \\ -67 & x & 53 \\ -83 & -91 & 5 \end{bmatrix}} \quad \underset{\text{(c)}}{\overbrace{001}^{P_{\text{dir}}}\ \overbrace{00}^{D_1}\ \overbrace{00}^{D_2}}$$

Fig. 3. Example of the LPTP code computation. (a) shows a neighborhood with intensity values, and (b) its responses after applying the masks, $\{M_0, \ldots, M_7\}$. (c) shows the LPTP code formed by the coded differences (shown in (a) as colored pairs), over the principal directions (shown in (b)), and the principal direction. (Red denotes the principal direction and blue the secondary direction.)

Finally, we create the LPTP code by concatenating the binary form of the principal direction index, and the two differences as shown in fig. 3c, which is equivalent to apply Eq. (2).

3. FACIAL EXPRESSION RECOGNITION USING LOCAL PRINCIPAL TEXTURE PATTERN

In this section we describe the proposed expression recognition method using LPTP features. We use the LPTP feature to represent the facial expressions, and we use Support Vector Machines (SVM) to classify those expressions.

3.1. Facial expression representation

We represent the facial expressions through a vector of histograms of LPTP features. Consequently, we generate the LPTP coded face using Eq. (2). Each code contains micro patterns of the face, which represent certain information of each neighborhood. However, the histogram looses the spatial information of the coded face. Hence, we divide the face into several regions, $\{R_0, \ldots, R_N\}$, and compute a histogram, H_i, from each region, R_i, where each bin corresponds to a different pattern value; in the LPTP feature we have 72 different possible code values. Furthermore, to construct the final descriptor we concatenate the histograms of each region, R_i, into a single feature vector. The different regions and the histogram concatenation are shown in fig. 4.

3.2. Facial expression recognition

We perform a person-independent facial expression recognition, by using a machine learning technique (SVM) to evaluate the performance of the proposed method. SVM [12] is a supervised machine learning technique that implicitly maps the data into a higher dimensional feature space. Consequently, it finds a linear hyperplane, with a maximal margin, to separate the data in different classes in this higher dimensional space.

Given a training set of M labeled examples $T = \{(x_i, y_i) \mid i = 1, \ldots, M\}$, where $x_i \in \mathbb{R}^n$ and $y_i \in \{-1, 1\}$, the test

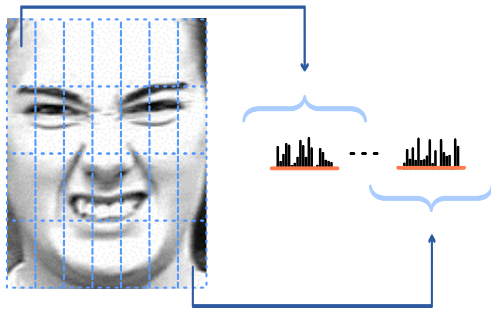

Fig. 4. Face representation using combined LPTP histogram.

data is classified by

$$f(x) = \text{sign}\left(\sum_{i=1}^{M} \alpha_i y_i K(x_i, x) + b\right), \qquad (3)$$

where α_i are Lagrange multipliers of dual optimization problem, b is a bias, and $K(\cdot, \cdot)$ is a kernel function. Note that SVM allows domain-specific selection of the kernel function. Although many kernels have being proposed, the most frequently used kernel functions are the linear, polynomial, and Radial Basis Function (RBF) kernels.

Given that SVM makes binary decisions, multi-class classification can be achieved by adopting the one-against-one or one-against-all techniques. In our work, we opt for one-against-one technique, which constructs $k(k-1)/2$ classifiers, that are trained with data from two classes [13]. We perform a grid-search on the hyper-parameters in a 10-fold cross-validation scheme for parameter selection, as suggested by Hsu *et al*. [14]. The parameter setting producing the best cross-validation accuracy was picked.

4. EXPERIMENTAL RESULTS

We evaluated the performance of the proposed method with the images from the Cohn-Kanade Facial Expression (CK) database [15] and the Japanese Female Facial Expression (JAFFE) database [16]; and against other three methods: Local Binary Pattern (LBP) [5], Local Directional Pattern (LDP) [11], and Gabor features [17]. The former database consists of hundred university students, in age from 18 to 30 years, of which 65% were female, 15% were African-American, and 3% were Asian or Latino. The subjects were instructed to perform a series of 23 facial displays, six of which were based on descriptions of prototypical emotions (*i.e.*, anger, disgust, fear, joy, sadness, and surprise). The last database contains 213 images of seven facial expressions (six basic facial expressions plus one neutral) posed by ten Japanese female models. Each image has been rated on six emotion adjectives by 60 Japanese subjects.

For our experiments, we cropped all the images to 110×150 pixels, based on the ground truth positions of the eyes and mouth, and partitioned the images into 4×7

Table 1. Comparison against others methods, in CK and JAFFE databases.

Method	CK		JAFFE	
	6 class (%)	7 class (%)	6 class (%)	7 class (%)
LBP	92.6 ± 2.9	88.9 ± 3.5	86.7 ± 4.1	80.7 ± 5.5
LDP	98.5 ± 1.4	94.3 ± 3.9	85.8 ± 1.1	85.9 ± 1.8
Gabor	89.8 ± 3.1	86.8 ± 3.1	85.1 ± 5.0	79.7 ± 4.2
LPTP	$\mathbf{99.4 \pm 1.1}$	$\mathbf{95.1 \pm 3.1}$	$\mathbf{90.2 \pm 1.0}$	$\mathbf{88.7 \pm 0.5}$

Table 2. Average recognition rate of LPTP in CK and JAFFE databases.

Kernel	CK		JAFFE	
	6 class (%)	7 class (%)	6 class (%)	7 class (%)
Linear	98.9 ± 0.9	94.0 ± 2.6	91.3 ± 1.1	91.5 ± 0.7
Polynomial	99.3 ± 1.8	96.4 ± 2.0	92.9 ± 0.6	92.0 ± 0.7
RBF	99.4 ± 1.1	95.1 ± 3.1	90.2 ± 1.0	88.7 ± 0.5

regions. Since LPTP detects the principal components of the textures, no further alignment of facial features was performed in our data, and since it is robust against illumination changes, no attempts were made for pre-processing in order to remove such changes. Note that these actions were taken to demonstrate the robustness of the proposed method against the corresponding changes. Moreover, we achieve person-independent classification by dividing the databases into several partitions and by ensuring that one person's expressions are not distributed into two partitions. In our experiments, we randomly divide the images into ten partitions, and we performed a leave-one-out cross-validation, which is equivalent to a 10-fold cross validation. Additionally, for the SVM we used a polynomial kernel of degree one, and the standard deviation for the RBF kernel was 2^{11} and 2^{13}, for six- and seven-class recognition.

The average recognition rate is shown in table 1, which demonstrate the superiority of LPTP feature in person-independent expression recognition over other existing facial features. This high accuracy is because LPTP extracts the information from the principal texture of each local neighborhood. Thereby encoding the prominent information, which makes it a more discriminant feature. In contrast, other methods use fixed coding points, which may be sub-optimal or may be influenced by noise. Additionally, the use of the principal directions makes LPTP robust against rotations. Given that we did not perform any pre-processing to remove these problems, other methods, that are not robust against rotation, were not able to detect correctly some expressions. Furthermore, in case of six and seven class recognition problem, LPTP feature achieves excellent accuracy with SVM, and the detailed results, for different kernels, are shown in table 2.

Additionally, the confusion matrix of 6-class recognition in the CK database is shown in table 3, which shows that the classification accuracy of each expression is over 96%.

Table 3. Confusion matrix of 6-class recognition using SVM (RBF), in CK database.

(%)	Anger	Disgust	Fear	Joy	Sadness	Surprise
Anger	99.5	0	0	0	0	0.5
Disgust	0	100.0	0	0	0	0
Fear	0	0	100.0	0	0	0
Joy	0	0	0	100.0	0	0
Sadness	3.1	0	0	0	96.9	0
Surprise	0	0	0	0	0	100.0

Table 4. Confusion matrix of 7-class recognition using SVM (RBF), in CK database.

(%)	Anger	Disgust	Fear	Joy	Sadness	Surprise	Neutral
Anger	87.3	0	0	0	0	0	12.7
Disgust	0	96.2	0	0	0	0	3.8
Fear	0	0	97.9	0	0	0	2.1
Joy	0	0	1.1	98.1	0	0	0.8
Sadness	0	0	0	0	95.2	0	4.8
Surprise	0	0	0	0	0	100.0	0
Neutral	3.9	1.0	0.8	0	1.8	0	92.5

The lowest accuracy occurs with the sadness expression being confused with an anger expression. As well, the confusion matrix of 7-class recognition, on the same database that is shown in table 4, ensures that other than the anger expression all other expressions can be classified with reasonable higher accuracy, over 92%. Despite most expression being confused minimally with the neutral expression, only the anger expression had a confusion over 10%.

5. CONCLUSION

In this paper, we proposed a novel local face-feature based on LPTP code for person-independent facial expression recognition. LPTP extracts the principal texture information in each neighborhood, and encodes the prominent characteristics of the neighborhood; instead of trying to accommodate all available information, which sometimes may introduce error into the code, as existing methods do. Moreover, the proposed LPTP descriptor is insensitive to noise, non-monotonic illumination, and slight rotation variations. Therefore, the facial expression recognition based on LPTP can achieve higher recognition accuracy over existing methods.

6. REFERENCES

[1] B. Fasel and J. Luettin, "Automatic facial expression analysis: a survey," *Pattern Recognition*, vol. 36, no. 1, pp. 259–275, 2003.

[2] S. Z. Li, A. K. Jain, Y. L. Tian, T. Kanade, and J. F. Cohn, "Facial expression analysis," in *Handbook of Face Recognition*, pp. 247–275. Springer New York, 2005.

[3] H. Hong, H. Neven, and C. von der Malsburg, "Online facial expression recognition based on personalized galleries," in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, apr 1998, pp. 354 –359.

[4] I. Kotsia and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 172–187, jan. 2007.

[5] C. F. Shan, S. G. 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, 2009.

[6] Y. L. Tian, "Evaluation of face resolution for expression analysis," in *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW '04. Conference on*, june 2004, p. 82.

[7] M. Pantic and L.J.M. Rothkrantz, "Automatic analysis of facial expressions: the state of the art," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 12, pp. 1424–1445, dec 2000.

[8] T. Ojala, M. Pietikainen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, Jan. 1996.

[9] C. F. Shan, S. G. Gong, and P. W. McOwan, "Robust facial expression recognition using local binary patterns," in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, sept. 2005, vol. 2, pp. II – 370–3.

[10] H. Zhou, R. W., and Cheng Wang, "A novel extended local-binary-pattern operator for texture analysis," *Information Sciences*, vol. 178, no. 22, pp. 4314–4325, 2008.

[11] T. Jabid, M. H. Kabir, and O. Chae, "Facial expression recognition using local directional pattern (ldp)," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, sept. 2010, pp. 1605 –1608.

[12] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[13] C. W. Hsu and C. J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–425, mar 2002.

[14] C. W. Hsu, C. C. Chang, and C. J. Lin, *A practical guide to support vector classification*, Taipei, Taiwan, 2003.

[15] T. Kanade, J. F. Cohn, and Y. L. Tian, "Comprehensive database for facial expression analysis," in *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, 2000, pp. 46 –53.

[16] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, apr 1998, pp. 200 –205.

[17] M. S. Bartlett, G. Littlewort, I. Fasel, and J. R. Movellan, "Real time face detection and facial expression recognition: Development and applications to human computer interaction.," in *Computer Vision and Pattern Recognition Workshop, 2003. CVPRW '03. Conference on*, june 2003, vol. 5, p. 53.