FACIAL EXPRESSION RECOGNITION BASED ON LOCAL SIGN DIRECTIONAL PATTERN

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

Department of Computer Engineering, Kyung Hee University, South Korea

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

ABSTRACT

In this paper, we propose a novel local feature descriptor, Local Sign Directional Pattern (LSDP), for face expression recognition. LSDP encodes the directional information of the face's textures—*i.e.*, the texture's structure—in a compact way, producing a more discriminating code than other state-of-the-art methods. The structure of each micro-pattern is encoded by using its prominent directions and sign—which allows it to distinguish among similar structural patterns that have different intensity transitions. We divide the face into several regions, from which we extract the distributions of the LSDP features. These features are concatenated into a feature vector, and used as a face descriptor, and the expression recognition is obtained with the aid of Support Vector Machine classifiers.

Index Terms— Local Sign Directional Pattern, Face descriptor, Local patterns

1. INTRODUCTION

Facial expression recognition is an interesting and challenging problem, and it impacts important applications in many areas, such as human-computer iteration and data-driven animation. Given that this topic is widely studied nowadays, several applications for facial expression recognition have gained attention [1,2]. And there are two common approaches to extract facial features: geometric features-based methods and appearance-based methods [3]. The geometric features represent the shape and location of facial components, which are extracted to form a feature vector that represents the face geometry. Moreover, major face components and/or feature points are detected in the images. In the latter, principal component analysis and multi-layer neural networks are extensively used to obtain a low-dimensional representation of the face.

Furthermore, video cameras have recently become an integral part of many consumer devices [4], and can be used for capturing facial images for recognition of people and their emotions. This ability to recognize emotions can enable customized applications [5, 6]. Though many work has been done on automatic facial expression recognition [1], higher accuracy in natural environments still remains a great challenge [7]. Consequently, a robust facial expression recognition system is needed to support these applications.

In this paper, we propose a novel face expression descriptor based on Local Sign Directional Pattern (LSDP), for robust person-independent face expression recognition, that encodes the structural information and intensity variations of the face's textures. LSDP encodes the directional information by extracting the edge responses in the neighborhood, in eight different directions using a Kirsch mask [8], and encoding the most positive and negative responses. These responses produce a meaningful descriptor that is able to differentiate similar structure with different intensity transitions. This approach allows us to distinguish intensity changes, *e.g.*, from dark to bright intensities and vice versa, in the texture that otherwise will be missed. We study the effectiveness of facial image representation based on LSDP for recognizing human expressions, and evaluate the performance of this representation using support vector machines (SVM). Consequently, we extensively experimented with two widely-used expression databases, namely, Cohn-Kanade (CK) facial expression [9] and the Japanese female facial expression (JAFFE) database [10], which demonstrate that LSDP feature is more robust in extracting the facial features in different environments. Moreover, we evaluate the performance of our approach against three different state-of-the-art methods: Local Binary Pattern (LBP) [11], Local Directional Pattern (LDP) [12], and Gabor wavelet features [13] approaches.

2. LOCAL SIGN DIRECTIONAL PATTERN

LBP [11] uses a thresholded sparse sample to encode the neighborhood intensity changes. However, the few pixels used reduces its accuracy. Also, it discards most of the information in the neighborhood, which makes the method sensible to noise. Furthermore, these drawbacks are more evident for bigger neighborhoods. Hence, to avoid these problems all the neighborhood's pixels can be used, as LDP [12] does. Although the use of more information makes LDP stable, it still encodes the information in the same way as LBP does, by marking the maximum absolute directions in a bit string. This encoding scheme, however, misses some directional information—the responses' sign—by treating all directions equally. Moreover, LDP is sensible to noise and illumination

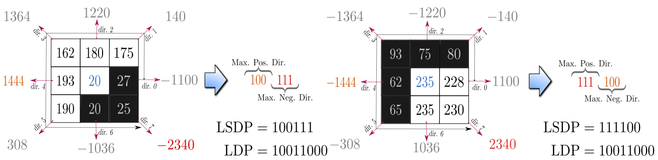

Fig. 1. LSDP code computation. The Kirsch compass mask is applied to a neighborhood to extract the edge responses. From those responses, we choose the prominent (positive and negative) directions to encode the texture in the neighborhood. LSDP can detect changes in the intensity regions, by producing a different code, while other directional patterns (LDP) cannot, as they produce the same code for different textures.

changes. To avoid these problems we propose the Local Sign Directional Pattern (LSDP).

The Local Sign Directional Pattern (LSDP) is a six bit binary code assigned to each pixel of an input image, that represents the structure of the texture and intensity transitions. Consequently, the pattern is created by computing the edge response of the neighborhood using a Kirsch mask, and conveying the most positive and negative directions of these edge responses. The positive and negative responses provide valuable information of the structure of the neighborhood, as they reveal the gradient direction of bright and dark areas in the neighborhood. This codding scheme is illustrated in fig. 1. Furthermore, these distinctions, between bright and dark responses, allows LSDP to differentiate between blocks with the positive and the negative direction swapped—which is equivalent to swap the bright and dark areas of the neighborhood. For example, the LSDP code shown in fig. 1, changes with the intensity transitions, producing 100111 and 111100. However, the LDP code remains the same, 1001100, for the different textures. Hence, LSDP distinguishes between two blocks that are the intensity complement of each other, while LDP does not.

2.1. Coding scheme

In our coding scheme, we generate the LSDP code by analyzing the edge response in eight directions using Kirsch masks, which are rotated $45°$ apart, and by combining the dominant information. Note that each response represents the edge significance in its respective direction, and given that the responses are not equally important, the presence of high negative or positive value signals a prominent bright or dark area. Hence, to encode the sign information we use a fixed position for the positive value, as the three most significant bits in the code, and the three least significant bits are the negative value, as shown in fig. 1. Therefore, we define the code as:

$$\mathrm{LSDP}(x_c, y_c) = 8 i_{x_c,y_c} + j_{x_c,y_c}, \tag{1}$$

where (x_c, y_c) is the central pixel of the neighborhood being coded, i_{x_c,y_c} is the direction number of the maximum positive response, and j_{x_c,y_c} is the direction number of the minimum negative response define by:

$$i_{x_c,y_c} = \arg\max_i \{M^i(x_c, y_c) \mid 0 \le i \le 7\}, \tag{2}$$

$$j_{x_c,y_c} = \arg\min_j \{M^j(x_c, y_c) \mid 0 \le j \le 7\}, \tag{3}$$

where M represents the responses of the ith and jth directions.

3. FACE DESCRIPTOR

Each face is represented by a LSDP histogram, H, as shown in fig. 2. The LSDP code contains fine to coarse information of an image, such as edges, corners, spots, and other local texture features. Given that the histogram only encodes the occurrence of certain micro-patterns without location information, in order to aggregate the location information we divide the face in smaller regions, $\{R^1, ..., R^N\}$. Finally, the histogram, H, is computed by concatenating each histogram of the regions by:

$$\mathrm{H} = \prod_{i=1}^{N} H_i, \tag{4}$$

where $\prod$ represents the concatenation operation, N is the number of the regions of the divided face, and H_i is the histogram of the ith region of the divided face. The spatially combined histogram, H, plays the role of a global face feature for the face. Moreover, we adopted Support Vector Machine (SVM) as classifier for expression recognition. As a powerful machine learning technique for data classification, SVM performs an implicit mapping of the data into a higher dimensional feature space, and then finds a linear separating hyperplane with the maximal margin to separate data, in that higher dimensional space. Given a training set of labeled samples $\{(x_i, y_i), i = 1, \ldots, h\}$, where $x_i \in \mathbb{R}^n$ and $y_i \in \{-1, 1\}$, a new test example x is classified by:

$$z(x) = \mathrm{sign}\left(\sum_{i=1}^{h} \varphi_i y_i K(x_i, x) + c\right), \tag{5}$$

where φ_i are Lagrange multipliers of a dual optimization problem that describe the separating hyperplane, $K(\cdot,\cdot)$ is a kernel function, and c is the threshold parameter of a hyperplane. The training sample x_i with $\varphi_i > 0$ is called a support vector, and SVM finds the hyperplane that maximizes the distance between the support vector and the hyperplane.

4. EXPERIMENTS

We perform experiments to evaluate the performance of the proposed algorithm under six, and seven facial expressions using a person-independent classification. We test our method in two different databases: CK and JAFFE. Moreover, we

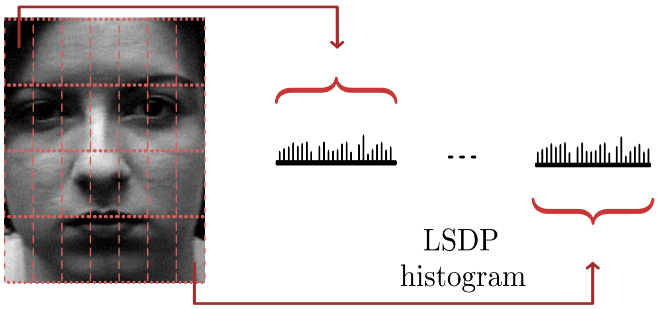

Fig. 2. Face descriptor LSDP.

cropped and normalized all the images to 110 $\times$ 150 pixels, based on the ground truth positions of the eyes and mouth. In our experiments, every image is partitioned into 4 $\times$ 7 regions. We compared the performance of the proposed LSDP based method against three state-of-the-art local encoding schemes: Local Binary Pattern (LBP) [11], Local Directional Pattern (LDP) [12], and Gabor wavelet features [13]. To evaluate the generalization performance to novel subjects, we adopted a 10-fold cross-validation testing scheme in our experiments. More specifically, we partitioned the datasets randomly into ten groups of roughly equal numbers of subjects. Nine groups were used as the training data to train the classifiers, while the remaining group was used as the test data. The above process was repeated ten times for each group in turn to be omitted from the training process. We reported the average recognition results on the test sets. In this section we discuss the evaluation process and the results.

4.1. Databases

The CK [9] database consist of 100 university students aged from 18 and 30 years, of which 65% were females, 15% were African-American and 3% were Asian or Latino. Subjects were instructed to perform a series of 23 facial displays, six of which were based on description of prototype emotions. Image sequences from neutral to target display were digitized into 640 $\times$ 490 pixel arrays with eight-bit precision for grayscale values. In our setup, we selected 408 image sequence from 96 subjects, each of which was labeled as one of the six basic emotions. For 6-class prototypical expression recognition, the three most expressive image frames were taken from each sequence that resulted into 1224 expression images. In order to build the neutral expression set, the first frame (neutral expression) from all 408 sequences was selected to make the 7-class expression dataset (1632 images).

The JAFFE [10] contains only 213 images of female facial expression expressed by 10 subjects. Each image has a resolution of 256 $\times$ 256 pixels with almost the same number of images for each categories of expression. The head in each image is usually in frontal pose, and the subject's hair was tied back to expose all the expressive zones of her face.

Table 1. Recognition performance of LSDP based SVM with different kernels, using CK database.

	6-Class recognition (%)	7-Class recognition (%)
SVM (Linear)	98.4 ± 1.4	92.3 ± 3.0
SVM (Polynomial)	99.1 ± 0.7	95.1 ± 4.1
SVM (RBF)	99.2 ± 0.8	94.8 ± 3.1

Table 2. Confusion matrix of 6-class facial expression recognition using SVM (RBF), using CK database.

	Anger (%)	Disgust (%)	Fear (%)	Joy (%)	Sadness (%)	Surprise (%)
Anger	99.02	0	0	0	0.49	0.49
Disgust	0	100.0	0	0	0	0
Fear	0	0.51	99.49	0	0	0
Joy	0	0	0	100.0	0	0
Sadness	3.16	0	0	0	96.84	0
Surprise	0	0	0	0	0.42	99.58

4.2. Results on Cohn-Kanade and JAFFE databases

SVM is a well-devised machine learning technique that provides excellent classification accuracy in pattern recognition. Therefore, we conducted the recognition using SVM with different kernels to classify the facial expressions. The comparative generalized performances with the SVM classifier based on different features are shown in tables 1 and 5, where the degree of the polynomial kernel is one, and the standard deviation for the RBF kernel is 2^{13} for 7-class recognition and 2^{11} for 6-class recognition. So far, we have discussed the average recognition accuracy of several prototypical expressions. To get a better picture of the recognition accuracy of individual expression types, the confusion matrices for 6-class and 7-class expression recognition with support vector machine using the CK database are given in tables 2 and 3, respectively. As we include the neutral expression in the 7-class recognition problem, the accuracy of other five expressions decrease because some facial expression samples are confused with a neutral expression. However, the surprise expression maintains the recognition rate without being confuse with the neutral expression.

We observed that the recognition accuracy in JAFFE database is relatively lower than CK database. One of the main reasons behind this accuracy is that some expressions in the JAFFE database are very similar with other expressions. Thus, depending on whether these expression images are used for training or testing, the recognition result is influenced. Furthermore, we compared the proposed method against three state-of-the-art methods, and we show their recognition rates on the tables 4 and 6. As observed, our approach outperforms the others methods.

Table 3. Confusion matrix of 7-class facial expression recognition using SVM (RBF), using CK database.

	Anger (%)	Disgust (%)	Fear (%)	Joy (%)	Sadness (%)	Surprise (%)	Neutral (%)
Anger	87.67	0	0	0	0	0	12.33
Disgust	0	95.56	0	0	0	0	4.44
Fear	0	0	97.42	0	0	0	2.58
Joy	0	0	0	98.88	0	0	1.12
Sadness	0	0	0	0	95.16	0	4.84
Surprise	0	0	0	0	0	100.0	0
Neutral	3.82	0.76	1.53	0	2.29	0.51	91.09

Table 4. Comparisson with others state-of-the-art methods, using CK database.

	6-Class recognition (%)	7-Class recognition (%)
LBP [11]	92.6 ± 2.9	88.9 ± 3.5
LDP	98.5 ± 1.4	94.3 ± 3.9
Gabor [13]	89.8 ± 3.1	86.8 ± 3.1
LSDP	$\mathbf{99.2 \pm 0.8}$	$\mathbf{94.8 \pm 3.1}$

5. CONCLUSION

In this paper, we introduced a novel encoding scheme, LSDP, that takes advantage of the structure of the face's textures and that encodes them efficiently into a compact code, for person independent facial expression recognition. LSDP uses directional information, that is more stable against noise than intensity, to code the different patterns from the face's texture. The code scheme that we presented, inherently, uses the sign information of the directions which allows it to distinguish similar texture's structures with different intensity transitions—*e.g.*, from dark to bright and vice versa. Then, we used this information to represent the facial expressions and machine learning techniques to classify the expressions.

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] 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.

[3] 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, 10.1007/0-387-27257-7-12.

[4] M. Z. Uddin, J. J. Lee, and T. S. Kim, "An enhanced independent component-based human facial expression recognition from video," *Consumer Electronics, IEEE Transactions on*, vol. 55, no. 4, pp. 2216–2224, november 2009.

Table 5. Recognition performance of LSDP based SVM with different kernels, using JAFFE database.

	6-Class recognition (%)	7-Class recognition (%)
SVM (Linear)	92.9 ± 1.7	90.1 ± 3.0
SVM (Polynomial)	93.4 ± 2.2	91.1 ± 3.0
SVM (RBF)	92.3 ± 1.7	89.2 ± 2.8

Table 6. Comparisson with others state-of-the-art methods, using JAFFE database.

	6-Class recognition (%)	7-Class recognition (%)
LBP [11]	86.7 ± 4.1	80.7 ± 5.5
LDP	85.8 ± 1.1	85.9 ± 1.8
Gabor [13]	85.1 ± 5.0	79.7 ± 4.2
LSDP	$\mathbf{92.3 \pm 1.6}$	$\mathbf{89.2 \pm 2.8}$

[5] P. Corcoran, C. Iancu, F. Callaly, and A. Cucos, "Biometric access control for digital media streams in home networks," *IEEE Trans. Consum. Electron.*, vol. 53, no. 3, pp. 917–925, aug. 2007.

[6] M. C. Hwang, L. T. Ha, N. H. Kim, C. S. Park, and S. J. Ko, "Person identification system for future digital tv with intelligence," *IEEE Trans. Consum. Electron.*, vol. 53, no. 1, pp. 218–226, february 2007.

[7] G. Zhao and M. Pietikäinen, "Boosted multi-resolution spatiotemporal descriptors for facial expression recognition.," *Pattern Recognition Letters 30(12):1117-1127*, 2009.

[8] R. A. Kirsch, "Computer determination of the constituent structure of biological images," *Computers & Biomedical Research*, pp. 315–328, 1970.

[9] 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.

[10] M. J. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 21, no. 12, pp. 1357–1362, dec 1999.

[11] C. Shan, 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, 2009.

[12] T. Jabid, M. H. Kabir, and O. Chae, "Robust facial expression recognition based on local directional pattern," *ETRI Journal*, vol. 32, pp. 784–794, 2010.

[13] 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.