



Double δ -LBP: A Novel Feature Extraction Method for Facial Expression Recognition

Fang Shen^(✉), Jing Liu, and Peng Wu

School of Artificial Intelligence, Xidian University, Xi'an, China
f.shen@qq.com

Abstract. The Local Binary Pattern (LBP) is a widely used descriptor in facial expression recognition due to its efficiency and effectiveness. However, existing facial expression recognition methods based on LBP either ignore different kinds of information, such as details and the contour of faces, or rely on the division of face images, such as dividing the face image into blocks or letting the block centering on landmarks. Considering this problem, to make full use of both detail and contour face information in facial expression recognition, we propose a novel feature extraction method based on double δ -LBP (D δ -LBP) in this paper. In this method, two δ -LBPs are employed to represent details and the contour of faces separately, which take different kinds of information of facial expression into account. Experiments conducted on both lab-controlled and wild environment databases show that D δ -LBP outperforms the original LBP method.

Keywords: Facial expression recognition · Local binary patterns
Feature extraction · Principal component analysis · Support vector machine

1 Introduction

The purpose of automatic facial expression recognition is to make the machine recognize different expressions of people, which has been widely used in the area of human-machine interaction. Many algorithms have been proposed for automatic facial expression recognition, in which the most crucial part is feature extraction. Recent successful features in facial expression recognition have been either handcrafted or learned from data. The handcraft feature focuses on constructing informative features manually. A good low-level feature should be both discriminative for inter-expression difference and invariant to intra-expression variations, such as lighting and the same expression of different people.

The Local Binary Pattern (LBP) descriptor has been widely used to both face recognition and facial expression recognition [1]. Existing feature extraction approaches using LBP can be generally grouped into two approaches: dividing face images into regular non-overlapping patches [2–4] and extracting patches centering at key points [5]. The regular non-overlapping dividing method first divides each face image into N overlapping patches with the same size, then applies the LBP to each patch. However, this method may separate the same information into two parts and this is not good for classification. The centering at key points patches method first uses the face alignment method to get landmarks, and then extracts feature centering at the key

points. It relies largely on the accurate facial landmarks. However, most existing methods of this category present an expression image ignoring different kinds of information like details and the contour, which is crucial for facial expression recognition in reality.

This paper proposes a novel Double δ -LBP (D δ -LBP) based facial expression recognition approach which applies the key points based method and also takes different kinds of information into consideration. After get the dense landmarks by the face alignment method, we extract patches centering at each landmark, and then apply two δ -LBP to get the detail and contour information respectively.

The performance of D δ -LBP is validated on four databases: the Extended Cohn-Kanade (CK+) database [6, 7], the Japanese Female Facial Expression (JAFPE) database [8, 9], the MMI database [10] and the Real-world Affective Face Database (RAF-DB) [11]. Experimental results illustrate that D δ -LBP achieves superior performance in comparison with the single δ -LBP method.

The main contributions of this paper are summarized as follows:

- (1) We employ two δ -LBP with two different parameters to obtain two parts of representation for facial expressions, which takes both detail and contour information into consideration and higher accuracies are achieved compared with single δ -LBP methods;
- (2) Different feature extraction methods of LBP are compared, which can be taken as a reference in further research;

The organization of the rest of this paper is as follows. Section 2 gives a review on LBP feature extraction methods in facial expression recognition and δ -LBP. Section 3 presents the proposed Double δ -LBP approach. Section 4 shows the experimental results. The conclusion is drawn in Sect. 5.

2 Related Work

In this section, we will first review existing feature extraction methods using LBP. Then, δ -LBP, which is the improved form of LBP is introduced. Also, we explain how we came up with the new feature extraction method D δ -LBP. The form of LBP and the feature extraction method are two important factors in facial expression recognition using LBP. From the weakness of previous feature extraction methods, combined with advantages of the improved LBP, the D δ -LBP is proposed.

2.1 Feature Extraction Methods Using LBP

We briefly review feature extraction methods using LBP in facial expression recognition in aforementioned two categories: regular non-overlapping dividing method and centering at key points method.

Considering the problem that an LBP histogram computed over the whole face image encodes only the occurrences of the micro-patterns without any indication about their locations, the regular non-overlapping dividing method equally divides face images into small regions R_0, R_1, \dots, R_m to extract LBP histograms. Shan *et al.* [2]

divide the 110×150 pixels face images into 18×21 pixels regions. That is, face images are divided into $42(6 \times 7)$ regions and represented by the LBP histograms with the length of 2478, giving a good trade-off between recognition performance and feature vector length. Ahmed *et al.* [12] partition each image into a number of regions and the proposed CLBP histograms are generated from each of those regions. The histograms of all regions are concatenated to obtain the extended LBP histogram. And the number of regions divided is also estimated in the experiment.

The centering at key points based patches method extracts the feature of facial images centering at landmarks which need to use the face alignment method to get landmarks first. Chen *et al.* [5] constructed the feature by extracting multi-scale patches centered at dense facial landmarks. After using recent face alignment method, they extract multi-scale image patches centered around each landmark. Each patch is divided into a grid of cells and codes each cell by a certain descriptor. Finally, they concatenate all histograms to form the high-dimensional feature.

2.2 δ -LBP

The original LBP operator was introduced by Ojala *et al.* [15, 16] and was proved to be a powerful texture description as it can detect even a tiny change of the grayscale value. LBP operator is defined as follows:

$$LBP_{P,R}(x, y) = \sum_{p=0}^{P-1} s(g_p - g_c)2^p, s(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \quad (1)$$

where g_c stands for the grayscale value of the center pixel and g_p ($p = 0, 1, \dots, P - 1$) represents the neighbor of the center pixel on a circle of radius R , and P denotes the number of the neighbors. In conclusion, the LBP value of a pixel is computed by comparing this pixel with its neighbors.

One fatal weakness of the original LBP operator is that it is sensitive to noise, especially in the near-uniform facial image regions since the thresholds are set exactly to the value of central pixel. To address this problem, Lu *et al.* [13] proposed the δ -LBP operator. δ -LBP which has 2-valued codes by comparing twice can be considered as the simplified LTP which has 3-valued codes by comparing three times. δ -LBP cut back a formula and in this way, it can greatly reduce the computational burden. δ -LBP is defined as follows:

$$\delta - LBP_{P,R}(x, y) = \sum_{p=0}^{P-1} s(g_p - g_c)2^p$$

$$s(x) = \begin{cases} 0, & x \leq \delta th \\ 1, & x > \delta th \end{cases}, \delta th \geq 0 \quad (2)$$

Compared with (1), (2) introduces a parameter δ_{th} to describe the difference between the peripheral pixel value and the intermediate pixel value and we can select different values of δ_{th} to achieve different effects. Figure 1 shows the encoding process of δ -LBP. Obviously, when δ_{th} is set to 0, δ -LBP equals to the original LBP.

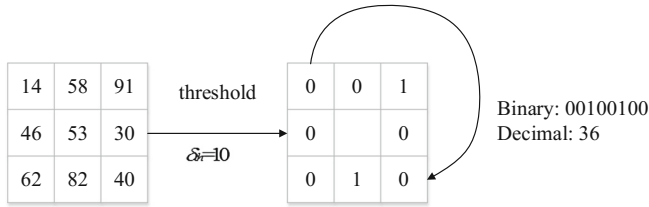


Fig. 1. The encoding process of δ -LBP.

2.3 Motivation of Proposing $D\delta$ -LBP

As mentioned in Sect. 2.1, most existing feature extraction methods using LBP in facial expression recognition did not take different kinds of information like details and the contour into account but relied on selecting patches on face images. In order to more fully extract the information of facial expression, we use two δ -LBP instead of single LBP to represent the features of facial expressions. One is used to extract the features of details, the other to extract the features of contours.

3 Facial Expression Representation Based on $D\delta$ -LBP

The framework of the proposed approach to represent the facial expression based on $D\delta$ -LBP is illustrated in Fig. 2. In this approach, a facial expression image is modeled as a combination of two histograms by applying two δ -LBP with two different δ_{th} . The approach can be described by the following procedure: (1) After pretreatment, e.g. face alignment, patches which have the size of 20×20 are generated centering around the landmarks of an input image. The number of patches equals to the number of landmarks; (2) The first histogram is formed by applying δ -LBP with the smaller parameter δ_{th1} to each patch and concatenate them. This histogram represents the detail information of the facial expression; (3) The second histogram is formed by applying δ -LBP with the larger parameter δ_{th2} to the patches and concatenate them. The second histogram represents the contour information of the facial expression; (4) The first and the second histograms are concatenated to form the final histogram as the representation of the facial expression. The more detailed procedure and the selection of parameters are described below.

3.1 Facial Expression Representation Based on Single δ -LBP

The parameter δ_{th} occupies the most significant position in the δ -LBP. In addition, the choice of δ_{th} affects the representation result. When the value of δ_{th} is small, the texture map obtained by applying δ -LBP operator presents more details of the face. When the value of δ_{th} is large, the texture map shows more contour information for the reason that δ -LBP with larger δ_{th} emphasizes the contrast of the surrounding pixels and the middle pixel value—Only those grayscale value contrasts between the surrounding and the middle pixel are obvious (e.g. edge regions like eyebrows, eyes, and mouth etc.)

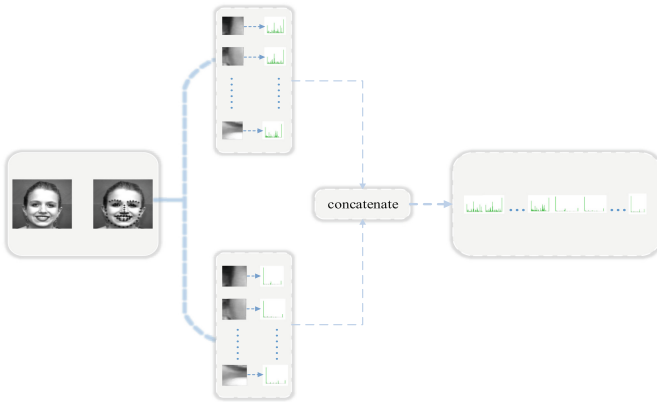


Fig. 2. Representation for the facial expression based on $D\delta$ -LBP.

may be set to 0, while other gentle areas is set to 1. Therefore, the selection of δ_{th} depends on the problem you need to solve. Figure 3 shows the texture map using δ -LBP with different values of δ_{th} .

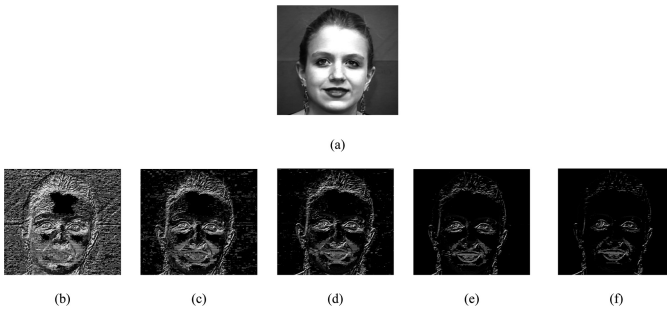


Fig. 3. Examples of texture map using δ -LBP with different δ_{th} . (a) The original face image. (b) The texture map using δ -LBP with $\delta_{th} = 0$. (c) The texture map using δ -LBP with $\delta_{th} = 3$. (d) The texture map using δ -LBP with $\delta_{th} = 5$. (e) The texture map using δ -LBP with $\delta_{th} = 10$. (f) The texture map using δ -LBP with $\delta_{th} = 15$.

3.2 Facial Expression Representation Using Double δ -LBP

The information of facial expression is reflected in two aspects: One is the change of facial features like eyebrows, eyes and mouth, etc., the other is the creation of wrinkles. In this case, we can use a pair of δ -LBP to represent these two different types of information. As illustrated in Fig. 3, δ -LBP with smaller δ_{th} reflects more details of the face, so we can use a δ -LBP with smaller δ_{th} to stand for wrinkles. δ -LBP with larger δ_{th} reflects contours of the face, so a δ -LBP with larger δ_{th} is used to demonstrate facial features. The detailed procedure is described below.

Firstly, we employ the 3000 fps [14] to get 68 landmarks and align faces. Then, the 68 patches which have the size of 20×20 are centering on landmarks obtained in the previous step and the boundary points that do not fall in the center of pixels can be estimated by bilinear interpolation. The feature vector of a face image is achieved by concatenate two histograms, each of which has the length of 68×59 . The first histogram is obtained by δ -LBP with smaller δ_{th} to describe detail information. The secondary histogram is obtained by δ -LBP with the larger δ_{th} to describe contour information.

The most crucial part of the procedure above is the selection of δ_{th1} and δ_{th2} . It derives from the result with the single δ -LBP. The single δ -LBP is similar to the D δ -LBP and the only different step is that the feature vector is formed with one histogram obtained from δ -LBP with one δ_{th} . We increase the value of δ_{th} from 0 to 40 in the step of 1 to get 41 results of recognition rate using the single δ -LBP method. Then we found that the result has a significant phenomenon: The highest recognition rate is achieved by a smaller δ_{th} and the second highest recognition rate is achieved by a larger δ_{th} . It turns out that smaller δ_{th} preserves more information including details and the contour at the same time. The larger δ_{th} cuts off most noise and can preserve the information of contours of the face, which are the most discriminating features of facial expressions. Chen *et al.* [5] found that high dimensionality leads to high performance in face recognition because it contains the amount of discriminative information for inter-person difference. To construct informative feature, we form the higher dimensional feature containing both details information with δ_{th1} achieved the best performance using the single δ -LBP and the contour information with δ_{th2} achieved the second-best performance using the single δ -LBP.

3.3 Dimension Reduction and Classification

We construct the high dimensional feature, so before the feature is input to the classifier, the high dimensional feature should be compressed. We use the joint of supervised and unsupervised subspace learning methods, joining Principal Component Analysis (PCA) [17] and Linear Discriminant Analysis (LDA) [18] to compress the high dimensional feature. After the compression, we send the feature into the support vector machine (SVM) [19] to recognize the expression.

4 Experiments

4.1 Database

The Extended Cohn-Kanade (CK+) Database. The database [6, 7] includes 593 sequences from 123 subjects posed or nonposed by 210 people from different area and different genders. The image sequence varies in duration and incorporate the onset to peak formation of the facial expressions.

The Japanese Female Facial Expression (JAFFE) Database. The database [8, 9] contains 213 photos of seven classes of facial expressions (six classes of basic facial expressions + 1 neutral faces) posed by ten Japanese females.

The MMI Database. The MMI database [10] consists of 30 subjects of both sexes (44% female) aged from 19 to 62, including either Asian, European or South American ethnic background and 213 sequences have been labeled with six expressions.

The Real-World Affective Face Database (RAF-DB). The RAF-DB [11] is a large and diverse real-world database that contains 29,672 static face images uploaded by Flickr users worldwide and provides multi-tagged emotional annotations.

4.2 Experiments on Single δ -LBP Method

Before conducting the experiment on D δ -LBP, we first investigate the single δ -LBP method to determine the two crucial parameters: δ_{th1} and δ_{th2} . In this experiment, we extract image patches centered at 68 landmarks and the patch size is fixed to 20×20 . Each patch is further divided into 3×3 cells then each cell is encoded with δ -LBP. 68 histograms calculated from each patch constitute the feature of one image. The dimension of the features is reduced by joint PCA (reserving 90% energy) and LDA. We apply 41 δ_{th} which range from 0 to 40 to evaluate the performance of the single δ -LBP method.

Experiment on CK+ Database. Six basic emotions (remove all “contempt” sequences) and neutral face are used to compare with other methods. For each sequence, the first image (neutral face) and the three peak images are used for prototype facial expression recognition. We construct 10 person-independent subsets by ascending ID order with the step size of 10 based on the subject ID in the dataset. Also, 10-fold cross validation is adopted. The average recognition rate of 7 classes with δ_{th} varying from 0 to 40 is shown in Fig. 4(a). We can find that δ_{th} which performs the best has the value of 2 and δ_{th} which performs the second-best has the value of 11. One has a smaller value and the other one has a larger value.

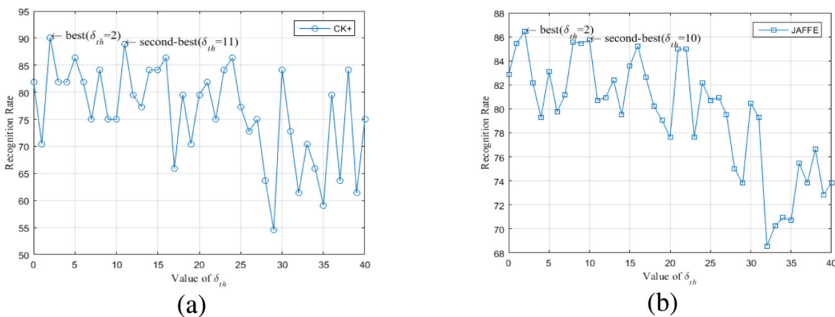


Fig. 4. (a) The average recognition rate of 7 classes with δ_{th} varying from 0 to 40 on CK+. (b) The average recognition rate of 7 classes with δ_{th} varying from 0 to 40 on JAFFE.

Experiment on JAFFE Database. All the 213 images are used for 7-class expression recognition. We adopt person-independent facial expression recognition and 10-fold cross-validation. Specifically, we use all images of one person as the validation set and the remaining images as the training set and the experiment is repeated 10 times so that each person is used for testing. The recognition rate with δ_{th} varying from 0 to 40 is shown in Fig. 4(b). We can also easily find that δ_{th} which performs the best has the value of 2 and δ_{th} which performs the second-best has the value of 10. One has a smaller value and the other one has a larger value.

4.3 Experiment on Different Face Region Selection Methods

To decide which face region selection method to use in $D\delta$ -LBP, we test the single δ -LBP in two selection methods: dividing the face into regular non-overlapped patches and patches centering at the key points. In the first method, we divide the face into 6×7 regular patches. In the second method, we use 3000 fps to get the 68 key points. As shown in Table 1, the centering at the key points method performs better than the other method. So we choose centering at the key points method in $D\delta$ -LBP.

Table 1. The comparison between different face region selection methods.

Database	Method	
	Regular patches	Centering at key points
CK+	70.53%	93.18%
JAFFE	80.95%	95.24%
MMI	52.86%	56.22%
RAF-DB	65.79%	70.83%

4.4 Experiment on Double δ -LBP Method

We conduct $D\delta$ -LBP experiment on both lab-controlled and wild environment databases. δ_{th1} is set to 2 and δ_{th2} to 11 on four databases.

We also extract the feature centered at each 68 landmarks and each patch is fixed to 20×20 . First, we apply δ -LBP with δ_{th1} to all patches and concatenate them together to form the $68 \times 59 = 4012$ dimension feature. Second, we apply δ -LBP with δ_{th2} to all patches and concatenate them together to form the second 4012-dimension feature. Finally, we concatenate the two 4012-dimension features to form the final 8024-dimension feature. This high dimensional feature is reduced by joint PCA (reserving 90% energy) and LDA before it is input to the SVM classifier.

The comparison of recognition rate between single δ -LBP and $D\delta$ -LBP on two databases is reported in Table 2. As shown in Table 2, the improvement of recognition rates due to $D\delta$ -LBP are 2.27%–7.84% on four databases, which shows the effectiveness of the $D\delta$ -LBP method.

Table 2. The comparison between single δ -LBP and D δ -LBP.

Database	Single δ -LBP	D δ -LBP
CK+	93.18%	95.45%
JAFFE	95.24%	97.62%
MMI	70.83%	78.67%
RAF-DB	56.22%	62.84%

5 Conclusions

A double δ -LBP based facial expression recognition method (D δ -LBP) is proposed in this paper. D δ -LBP employs two δ -LBP to represent facial expression with different scales of information into consideration. Considering the most important property of facial expressions, we use two δ_{th} to represent the detail and the contour information separately. Experiments are conducted on four databases to illustrate the effectiveness of the proposed method. Compared with the single δ -LBP method, D δ -LBP achieves better performance in terms of facial expression recognition accuracy. The key advantages of D δ -LBP method is that it takes both details and the contour of faces into account, which can fully extract the information of facial expressions. The proposed method can also be applied to other fields, such as face recognition, object detection, and so on.

References

1. Huang, D., Shan, C., Ardabilian, M., Wang, Y., Chen, L.: Local binary patterns and its application to facial image analysis: a survey. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **41**(6), 765–781 (2011)
2. Shan, C., Gong, S., Mcowan, P.W.: Facial expression recognition based on local binary patterns: a comprehensive study. *Image Vis. Comput.* **27**(6), 803–816 (2009)
3. Ahonen, T., Hadid, A., Pietikäinen, M.: Face description with local binary patterns: application to face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(12), 2037–2041 (2006)
4. Kumari, J., Rajesh, R., Pooja, K.: Facial expression recognition: a survey. *Int. Symp. Comput. Vis. Internet* **58**, 486–491 (2015)
5. Chen, D., Cao, X., Wen, F., Sun, J.: Blessing of dimensionality: high-dimensional feature and its efficient compression for face verification. *Comput. Vis. Pattern Recognit.* **9**(4), 3025–3032 (2013)
6. Kanade, T., Cohn, J.F., Tian, Y.: Comprehensive database for facial expression analysis. In: 2000 IEEE International Conference on Automatic Face and Gesture Recognition, pp. 484–490 (2000)
7. Lucey, P., Cohn, J.F., Kanade, T., Saragih, J.: 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, vol. 36, no. 1, pp. 94–101 (2010)

8. Lyons, M., Akamatsu, S., Kamachi, M., Gyoba, J.: Coding facial expressions with Gabor wavelets. In: 1998 IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200–205 (1998)
9. Lyons, M.J., Budynek, J., Akamatsu, S.: Automatic classification of single facial images. *IEEE Trans. Pattern Anal. Mach. Intell.* **21**(12), 1357–1362 (1999)
10. Pantic, M., Valstar, M., Rademaker, R., Maat, L.: Web-based database for facial expression analysis. In: Proceedings of the 2005 IEEE International Conference on Multimedia and Expo, vol. 14 (2005)
11. Deng, W., Hu, J., Zhang, S., Guo, J.: DeepEmo: real-world facial expression analysis via deep learning. In: Visual Communications and Image Processing, pp. 1–4 (2016)
12. Ahmed, F., Hossain, E., Bari, A.S.M.H, Shihavuddin, A.: Compound local binary pattern (CLBP) for robust facial expression recognition. In: IEEE International Symposium on Computational Intelligence and Informatics, pp. 391–395 (2011)
13. Lu, S., Yang, J.H., Zhang, B., Zhang, J.Q.: Infrared target detection based on LBP. *J. Changchun Univ. Sci. Technol.* **32**(1), 22–24 (2009)
14. Ren, S., Cao, X., Wei, Y., Sun, J.: Face alignment at 3000 fps via regressing local binary features. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1685–1692 (2014)
15. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(7), 971–987 (2002)
16. Ojala, T., Pietikäinen, M., Mäenpää, T.: A generalized local binary pattern operator for multiresolution gray scale and rotation invariant texture classification. In: Singh, S., Murshed, N., Kropatsch, W. (eds.) ICAPR 2001. LNCS, vol. 2013, pp. 399–408. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-44732-6_41
17. Jolliffe, I.T.: Principal Component Analysis, vol. 87, pp. 41–64. Springer, Berlin (1986). <https://doi.org/10.1007/978-1-4757-1904-8>, no. 100
18. Altman, E.I., Marco, G., Varetto, F.: Corporate distress diagnosis: comparisons using linear discriminant analysis and neural networks. *J. Bank. Financ.* **18**(3), 505–529 (1994)
19. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2**(3), 27 (2011)