Extracting Local Binary Patterns from Image Key Points: Application to Automatic Facial Expression Recognition

Xiaoyi Feng[1], Yangming Lai[1], Xiaofei Mao[1], Jinye Peng[1], Xiaoyue Jiang[1], and Abdenour Hadid[2]

[1] College of Electronics and Information
Northwestern Polytechnic University
710072 Xian, China
fengxiao@nwpu.edu.cn

[2] Center for Machine Vision Research
Department of Computer Science and Engineering
University of Oulu, Finland
hadid@ee.oulu.fi

Abstract. Facial expression recognition has widely been investigated in the literature. The need of accurate facial alignment has however limited the deployment of facial expression systems in real-world applications. In this paper, a novel feature extraction method is proposed. It is based on extracting local binary patterns (LBP) from image key points. The face region is first segmented into six facial components (left eye, right eye, left eyebrow, right eyebrow, nose, and mouth). Then, local binary patterns are extracted only from the edge points of each facial component. Finally, the local binary pattern features are collected into a histogram and fed to an SVM classifier for facial expression recognition. Compared to the traditional LBP methodology extracting the features from all image pixels, our proposed approach extracts LBP features only from a set of points of face components, yielding in more compact and discriminative representations. Furthermore, our proposed approach does not require face alignment. Extensive experimental analysis on the commonly used JAFFE facial expression benchmark database showed very promising results, outperforming those of the traditional local binary pattern approach.

Keywords: Local Binary Patterns, Facial expression Recognition, Key Points.

1 Introduction

Automatic facial expression recognition has attracted significant attention in computer vision research. Compared to methods using face video sequences, static-image based methods for facial expression recognition are more challenging due to the lack of the dynamic information (i.e. facial movements). Many

J.-K. Kämäräinen and M. Koskela (Eds.): SCIA 2013, LNCS 7944, pp. 339–348, 2013.

previous works have however pointed out the feasibility of facial expression recognition even from static images. Most of these works have been reported on the Japanese Female Facial Expression (JAFFE) Database which is one of the most popular benchmark database for evaluating static-image based recognition methods [1,2,3,4,5,6,7,8,9,10].

LBP-based facial expression analysis is a very popular method for recognizing facial expressions from static images. For instance, Feng et al. [3] exploited a coarse-to-fine classification scheme with LBP features for facial expression recognition obtaining good results. He et al. [4] used LBP on Gabor coefficients of images for facial expression recognition Yielding better performance than extracting LBP from the original images. To consider multiple cues, Liao et al. [5] extracted LBP features in both intensity and gradient maps, and computed the Tsallis entropy of the Gabor filters responses as the first feature set and selected null-space LDA for the second feature set. Feng et al. [6] extracted the local texture features by applying LBP to facial feature points, the direction between each pair of feature points was also considered as shape information, and subject-independent recognition rate of 72% was reported on the JAFFE dataset. Cao et al. [7] combined LBP with embedded hidden Markov model to recognize facial expressions by using an active shape model (ASM), and achieved an accuracy of 65% on the JAFFE dataset. Fu and Wei [8] proposed centralized binary patterns embedded with image Euclidean distance for facial expression recognition. Ahmed et al. [9] presented the compound local binary patterns codes for facial expression recognition.

It appears that in all these previous works using LBP representations for facial expression recognition, the features are extracted from each image pixel and face alignment is very crucial. This undoubtedly limits the use of such methods in real-world applications. To overcome these limitations, we propose in this work a novel method based on extracting local binary patterns (LBP) from only key points. The face region is first segmented into six facial components (left eye, right eye, left eyebrow, right eyebrow, nose, and mouth). Then, local binary patterns are extracted only from the edge points of each facial component. Finally, the local binary pattern features are collected into a histogram and fed to an SVM classifier for facial expression recognition. Compared to the traditional LBP methodology extracting the features from all image pixels, our proposed approach extracts LBP features only from a set of points of face components, yielding in more compact and discriminative representations. Furthermore, our proposed approach does not require face alignment. Extensive experimental analysis on the commonly used JAFFE facial expression benchmark database showed very promising performance, outperforming those of traditional LBP based approaches.

The general scheme of our proposed approach consists of four main steps which are (1) face detection, (2) facial component segmentation, (3) key points detection and feature extraction, and (4) facial expression classification. Since face detection and facial component segmentation problems have been thoroughly studied in the literature, our work mainly focuses on the two challenging issues of key points detection and feature extraction and expression classification from static images.

The rest of this paper is organized as follows. Section 2 briefly describes traditional face representations using LBP. The proposed key point detection and feature extraction are then introduced in Section 3. Section 4 presents the adopted classification scheme. Section 5 provides extensive experiments and comparative analysis on the effectiveness of the proposed method. A conclusion is drawn in Section 6.

2 Face Representation Using Local Binary Patterns

The LBP texture analysis operator, introduced by Ojala et al. [16], is defined as a gray-scale invariant texture measure, derived from a general definition of texture in a local neighborhood. It is a powerful means of texture description and among its properties in real-world applications are its discriminative power, computational simplicity and tolerance against monotonic gray-scale changes.

The original LBP operator forms labels for the image pixels by thresholding the 3×3 neighborhood of each pixel with the center value and considering the result as a binary number. Fig. 1 shows an example of an LBP calculation. The histogram of these $2^8 = 256$ different labels can then be used as a texture descriptor.

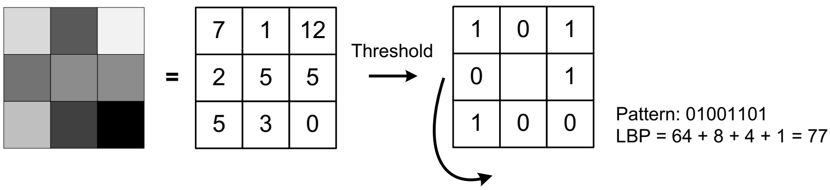


Fig. 1. The basic LBP operator

The operator has been extended to use neighborhoods of different sizes. Using a circular neighborhood and bilinearly interpolating values at non-integer pixel coordinates allow any radius and number of pixels in the neighborhood. The notation (P, R) is generally used for pixel neighborhoods to refer to P sampling points on a circle of radius R. The calculation of the LBP codes can be easily done in a single scan through the image. The value of the LBP code of a pixel (x_c, y_c) is given by:

$$\mathrm{LBP}_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p, \tag{1}$$

where g_c corresponds to the gray value of the center pixel (x_c, y_c), g_p refers to gray values of P equally spaced pixels on a circle of radius R, and s defines a thresholding function as follows:

$$s(x) = \begin{cases} 1, \text{ if } x \geq 0; \\ 0, \text{ otherwise.} \end{cases} \tag{2}$$

Another extension to the original operator is the definition of so called *uniform patterns*. This extension was inspired by the fact that some binary patterns occur more commonly in texture images than others. A local binary pattern is called uniform if the binary pattern contains at most two bitwise transitions from 0 to 1 or vice versa when the bit pattern is traversed circularly. In the computation of the LBP labels, uniform patterns are used so that there is a separate label for each uniform pattern and all the non-uniform patterns are labeled with a single label. This yields to the following notation for the LBP operator: $\text{LBP}_{P,R}^{u2}$. The subscript represents using the operator in a (P, R) neighborhood. Superscript $u2$ stands for using only uniform patterns and labeling all remaining patterns with a single label.

Each LBP label (or code) can be regarded as a micro-texton. Local primitives which are codified by these labels include different types of curved edges, spots, flat areas etc. The occurrences of the LBP codes in the image are collected into a histogram. The classification is then performed by computing histogram similarities. For an efficient representation, facial images are first divided into several local regions from which LBP histograms are extracted and concatenated into an enhanced feature histogram.

3 Proposed Key Point Detection and Feature Extraction

As preprocessing, our proposed approach starts by detecting the faces (using the method proposed in [11]) and segmenting them into six facial components (left eye, right eye, left eyebrow, right eyebrow, nose, and mouth) using a geometrical face model approach. Results of face detection and facial component segmentation are shown in Fig. 2.

It is known that facial expressions can be discriminated based on some specific expression-sensitive features such as the shapes of facial regions (eyebrows, eyes, mouth) or facial wrinkles. Local feature based approaches are usually effective in recognizing expressions but are relatively sensitive to the localization of the facial points. On the other hand, the global feature based approaches are easier to implement but at the cost of lower performance. Our proposed approach inherits the advantages of both local and global based methods for expression recognition.

The traditional approach to facial expression recognition using LBP [3] follows the methodology described in Section 2. First, the face image is divided into several overlapping or non-overlapping blocks. Then, LBP histograms are extracted from each block and concatenated into a single histogram which is used a feature vector for classification. In such an approach, the facial expressions are

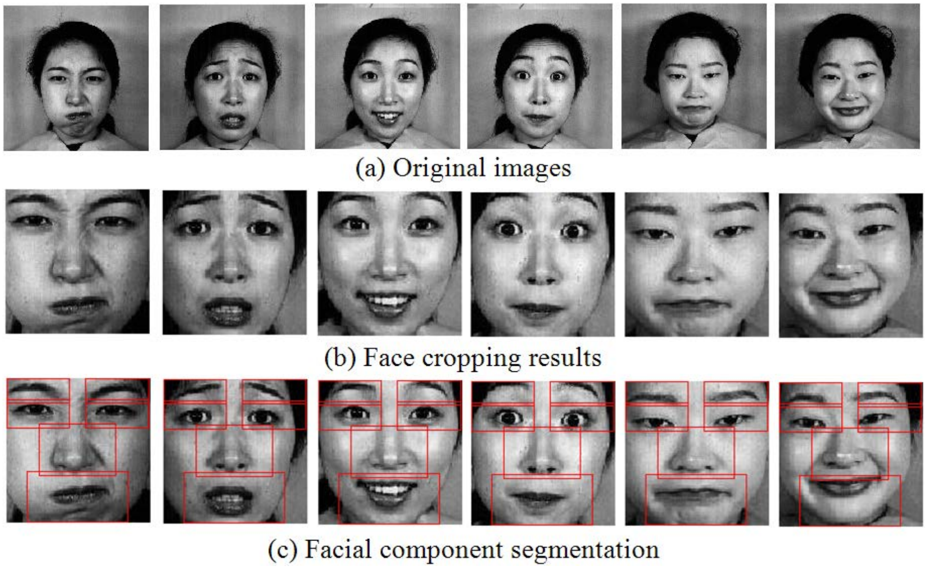
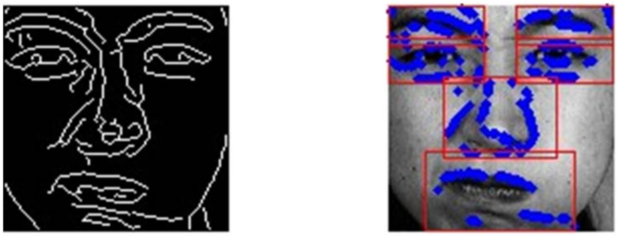


Fig. 2. Examples of face cropping and component segmentation results on the JAFFE database

represented by LBP histograms whereas the shape of the face is encoded by the concatenation of the local histograms. Experiments have shown that such an approach can encode some typical features which are effective in recognizing some intense and typical expressions [14].

However, this traditional approach to facial expression recognition using LBP suffers from several limitations. First, such an approach is quite sensitive to face registration. Second, the LBP patterns may be sensitive to noise. Third, the approach uses global features extracted from local regions and hence does not represent facial expressions in details. Forth, the LBP operator is applied to all image pixels yielding in long and sparse histograms which are inadequate for classification.

To overcome these limitations, our proposed approach computes the edges in the face images and then extracts the LBP features but only from some key edge points. More specifically, the scale space is generated firstly; then the LBP operator is applied to the images in scale space; then edges are extracted at different scales with canny edge detector. The intersection of these edge points across different scales results in a sub set of edge points, of which points with same LBP value at different scale are referred to scale-invariant LBP key points. The facial expression features are formed by the LBP values of these key points. The histograms of these features in each block are contacted and used as a feature vector for classification as done in the traditional LBP approach.



(a) Edges of facial components (b) Key points

Fig. 3. Example of key point feature detection

Our proposed scheme for feature extraction can be summarized in the followings steps:

1. As suggested in [15], face images are smoothed with different scale of Gaussian filters.This results in a group of same size images, which is refered to scale space.
2. Key points are extracted with the following steps:
   - Extract edge points of each scale image using canny operator.
   - LBP feature extration of each scale image with the uniform patterns LBP operator.
   - Select points which are both common edge points and scale-invariant LBP points as key points.
3. Each selected key point is represented by a 72 binary code corresponding to the key point and its 8 neigbording pixels, each represented by an 8 binary code.

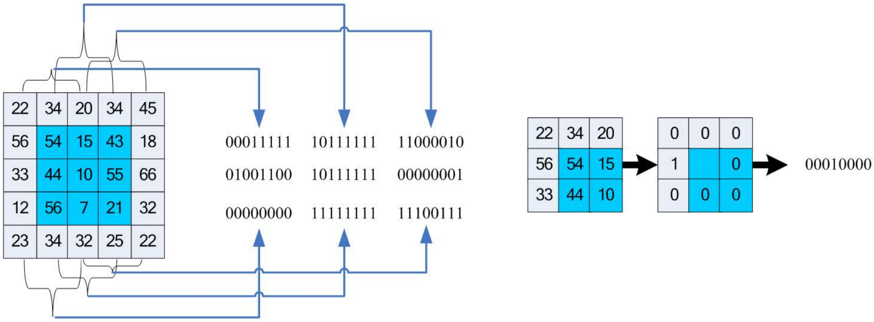
The proposed approach does not require face registration. The selected key points are shown to be relatively robust to noise. An example of key point detection results is shown in Fig.3 whereas Fig. 4 illustrates an example feature calculation.

4 Classification Scheme

To determine the expression of a give face image, we extracted LBP histograms from the selected key points and used them as inputs to SVM classifiers. The choice of SVM is motivated by its proven performance in various object detection and recognition tasks in computer vision. SVM [13,14] constructs a higher dimensional feature space from an original data, and then finds a linear separating hyperplane with the maximal margin to separate data in this higher dimensional space.

Let $\{(x_i, y_i),\ \ i = 1, ..., l\}$ be a set of l training samples, each data $x_i \in \Re^n$, n being the dimension of the input space, belongs to a class labeled by $y_i \in \{-1, 1\}$. A new test example x is classified by the following function:

$$f(x) = sgn(\sum_{i=1}^{l} \alpha_i y_i K(x_i, x) + b)$$



(a) Example of key point feature extraction result

(b) LBP value of one pixel

Fig. 4. Example of key point feature extraction

where α_i are Lagrange multipliers, $K(u, v)$ is a kernel function, and b is the threshold parameter of the hyperplane.

SVMs are basically designed for binary classification. In our proposed approach, we adopted one-against-one scheme and built an SVM classifier for each pair of facial expression (e.g. happiness versus anger, happiness versus sadness, anger versus sadness etc.). The classification decision is done by a voting strategy in which every classifier assigns the test sample to one of the two classes. The class with the most votes determines the facial expression of the face in the given image.

5 Experimental Analysis

To assess the effectiveness of our proposed face key point detection and feature extraction for facial expression recognition, we conducted extensive experiments evaluating the performance of the proposed representation and comparing the results against those of using traditional local binary pattern approaches (Section 2).

The experiments were conducted on the popular Japanese Female Facial Expression (JAFFE) benchmark Database [10]. The database contains 213 images of ten expressers posed 3 or 4 times of each of the seven basic expressions (happiness, sadness, surprise, anger, disgust, fear, and neutral). Sample images from the database are shown in Fig. 2. To report the results, we used cross-validation technique over the 213 images for person-independent facial expression recognition.

We divided the database into 10 subsets, each corresponds to one expresser. Then, we trained the SVM classifiers using data from nine subsets. The remaining set is used for testing. This process is repeated 10 times so that each of the ten sets is used once as the test set. The reported results are the average of 10 permutations.

Our proposed method is compared against traditional LBP approaches with (Fig.5.a) and without (Fig.5.b) face alignment.

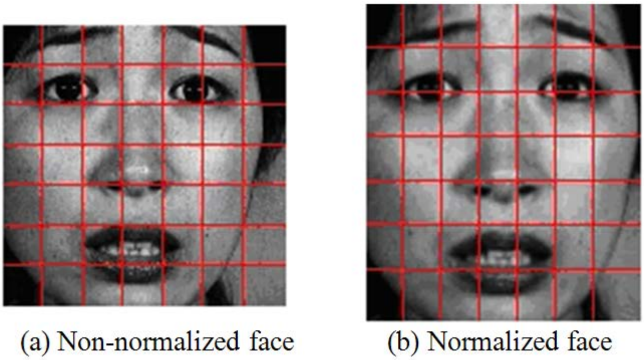


Fig. 5. Traditional LBP approach with and without face alignment

Table 1. Facial expression recognition results: our proposed approach versus traditional LBP approaches

	Anger %	Disgust %	Fear %	Happiness %	Sadness %	Surprise %	Neutral %	Average %
LBP without normalisation	56.7	40.8	30.0	66.7	53.3	30.8	73.3	50.2
LBP with normalisation	70.0	40.0	57.5	73.3	63.3	50.8	60.0	59.3
Our proposed approach	73.3	39.2	45.0	64.2	53.3	44.2	70.0	55.7

The recognition rates obtained using different methods are shown in Table 1. As one can expect, the performance of traditional LBP approach with non-normalized face images is much lower than that of LBP with normalized face images. This confirms the relative sensitivity of traditional LBP approaches to face alignment.

Our proposed method outperforms its LBP counterpart with non-normalized face images. It is also worth noting that the results of our approach which does not require alignment are comparable to those of traditional LBP with alignment. This clearly demonstrates the robustness of our approach against misalignment. Furthermore, our approach yields in much more compact feature vectors than in traditional LBP. The main raison is that, in our approach, the features are extracted only from selected key points while traditional LBP approach processes each pixel.

6 Conclusion

From the observations that local binary pattern based approaches for facial expression recognition from still images have significant limitations as they extract features from each pixel and are sensitive to misalignment and to noise, we proposed in this work a novel method based on extracting local binary patterns (LBP) from only some selected key points. The combination of edge detection and scale-invariant point detection is used to determine these key points. The face region is first segmented into six facial components and then local binary patterns are extracted from the edge points of each facial component. Finally, the LBP features are collected into a histogram and fed to SVM classifiers for facial expression recognition. Compared to the traditional LBP methodology which extracts the features from all image pixels, our proposed approach extracts LBP features only from a set of points of face components, yielding in more compact and discriminative representations. Furthermore, our proposed approach does not require face alignment. Extensive experimental analysis on the commonly used JAFFE facial expression benchmark database showed very promising performance, outperforming those of traditional LBP based approaches without alignment.

This work is by no means complete. Our goal was primary to show that an approach based on features extracted from some fiducial points of the face should always be preferred since detailed information about facial expression are better exploited. Our proposed methodology is not limited to the single problem of facial expression recognition but can easily be adapted to other face related problems such as face recognition, gender classification, age estimation and face spoofing detection.

It is also of interest to expand the proposed methodology to not only focus on static images but also to encode the dynamics of the fiducial points in the face over the video for more accurate face analysis. This is inspired by psychophysical and neural studies indicating that behavioural characteristics are person-specific and do also provide valuable information to face analysis in the human visual system.

Acknowledgment. The authors thank Dr. M. Lyons for providing the Japanese Female Facial Expression (JAFFE) Database. This work is partially supported by the National Science Foundation of China (under Grant No.61075014 and 60875016). The support of the EU project TABULA RASA (grant agreement No: 257289) is also acknowledged.

References

1. Zhao, G., Pietikäinen, M.: Dynamic Texture Recognition Using Local Binary Patterns with An Application to Facial Expressions. IEEE Trans. Pattern Anal. Mach. Intell. 29(6), 915–928 (2007)
2. Huang, X., Zhao, G., Zheng, W., Pietikäinen, M.: Spatiotemporal Local Monogenic Binary Patterns for Facial Expression Recognition. IEEE Signal Processing Letters 19(5), 243–246 (2012)
3. Feng, X., Pietikäinen, M., Hadid, A.: Facial Expression Recognition with Local Binary Patterns and Linear Programming. Pattern Recognition and Image Analysis 15(2), 546–548 (2005)

4. He, L., Zou, C., Zhao, L., Hu, D.: An enhanced LBP feature based on facial expression recognition. In: Proc. Ann. Int. Conf. Eng. Med. Biol. Soc., pp. 3300–3303 (2005)
5. Liao, S., Fan, W., Chung, A.C.S., Yeung, D.Y.: Facial expression recognition using advanced local binary patterns, tsallis entropies and global appearance features. In: Proc. IEEE Int. Conf. Image Process., pp. 665–668 (October 2006)
6. Feng, X., Lv, B., Li, Z., Zhang, J.: A novel feature extraction method for facial expression recognition. In: Proc. Joint Conf. Inform. Sci. Issue Adv. Intell. Syst. Res., Kaohsiung, Taiwan, pp. 371–375 (2006)
7. Cao, J., Tong, C.: Facial expression recognition based on LBP-EHMM. In: Proc. Congr. Image Signal Process. (2008)
8. Fu, X., Wei, W.: Centralized binary patterns embedded with image Euclidean distance for facial expression recognition. In: Proc. Int. Conf. Neural Comput., pp. IV: 115–119 (October 2008)
9. Ahmed, F., Hossain, E., Bari, A., Shihavuddin, A.: Compound Local Binary Pattern (CLBP) for Robust Facial Expression Recognition. In: International Conference of Soft Computing and Pattern Recognition (SoCPaR), pp. 391–395 (2011)
10. Lyons, M., Akamastu, S., Kamachi, M., Gyoba, J.: Coding facial expressions with Gabor wavelets. In: Proc. Third IEEE Conf. Face and Gesture Recognition, Nara, Japan, pp. 200–205 (April 1998)
11. Nilsson, M., Nordberg, J., Claesson, I.: Face Detection Using Local SMQT Features and Split up SNOW Classifiers. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP (2007)
12. Valstar, M.F., Pantic, M.: Biologically vs. logic inspired encoding of facial actions and emotions. In: Proc. of IEEE Intl. Conf. on Multimedia and Expo (ICME), pp. 325–328 (2006)
13. Bartlett, M.S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., Movellan, J.: Recognizing facial expression: machine learning and application to spontaneous behavior. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR (2005)
14. Shan, C., Gong, S., McOwan, P.W.: Facial expression recognition based on Local Binary Patterns: A comprehensive study. Image and Vision Computing 27, 27803–27816 (2009)
15. David, D.G.: Lowe, Distinctive Image Features from Scale-Invariant Points. International Journal of Computer Vision 60(2), 91–110 (2004)
16. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(7), 971–987 (2002)