JAIDM
*Journal of **AI** and **D**ata **M**ining*
Vol. 2, No .1, 2014, 33-38.

Local gradient pattern - A novel feature representation for facial expression recognition

M. Shahidul Islam

Department of Computer Science, School of Applied Statistics, National Institute of Development Administration, Bangkok, Thailand.

Received 23 April 2013; accepted 16 June 2013
**Corresponding author: suva.93@grads.nida.ac.th (M.Shahidul Islam)*

Abstract
Many researchers adopt Local Binary Pattern for pattern analysis. However, the long histogram created by Local Binary Pattern is not suitable for a large-scale facial database. This paper presents a simple facial pattern descriptor for facial expression recognition. Local pattern is computed based on local gradient flow from one side to another side through the center pixel in a 3x3 pixels region. The center pixel of that region is represented by two separate two-bit binary patterns, named as Local Gradient Pattern (LGP) for the pixel. LGP pattern is extracted from each pixel. Facial image is divided into 81 equal sized blocks and the histograms of local LGP features for all 81 blocks are concatenated to build the feature vector. Experimental results prove that the proposed technique along with Support Vector Machine is effective for facial expression recognition.

Keywords: *Facial Expression Recognition, Local Feature Descriptor, Pattern Recognition, CK+, LIBSVM.*

1. Introduction

Facial expression is very important in daily interactions. It adds some more meaning to the verbal communication. Facial expression contributes 55% to the meaning of the communication whereas verbal expression contributes only 7% **[1]**. Therefore, a developing automatic facial expression recognition system has become a research issue in computer vision. Few important applications of this system are video surveillance for security, driver state monitoring for automotive safety, educational intelligent tutoring system (ITS), clinical psychology, psychiatry and neurology, pain assessment, image and video database management, and lie detection. Due to these applications, it attracts much attention of the researchers in the past few years [**2**]. Most of the facial expression recognition systems (FERS) available in the market are based on the Facial Action Coding System (FACS) [**3**] which involves more complexity due to facial feature detection and extraction procedures. Shape based models have problem with on plane face transformation. Gabor wavelets [**4**] are very popular but the feature vector length is huge and has a computational complexity. Another appearance-based approach, LBP-local binary pattern [**5**], which is adopted by many researchers also has disadvantages, for example: (a) it produces long histograms, which slows down the recognition speed, and (b) Under some certain circumstances, it misses the local feature, as it does not consider the effect of the center pixel. To overcome all the above problems, this paper proposed a novel appearance-based local feature descriptor LGP - Local Gradient Pattern, which has a tiny feature vector length of 8. Appearance based methods are less dependent on initialization and can encode patterns from either local or full facial area. However, appearance features do not generalize across individuals, as they encode specific appearance information. Ahonen *et al*. [**6**] implemented a facial expression recognition system using Local Binary Pattern. LBP was first proposed by Ojhala *et al*., [**5**] in 1996 for texture analysis. Slowly it has become popular, due to its unique power to differentiate textures using local 8-bit binary pattern. An LBP value from a local 3x3 pixels region is computed using the following equation (1).

$$\mathrm{LBP(i,j)} = \sum_{p=1}^{P} 2^{p-1} * f(\mathrm{g_c} - \mathrm{g_p}) \qquad (1)$$

$$f(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

Where g_c and g_p are the gray color intensity of the center pixel (i, j) and p neighboring pixel respectively. A detailed example of obtaining LBP value is shown in Figure 1.

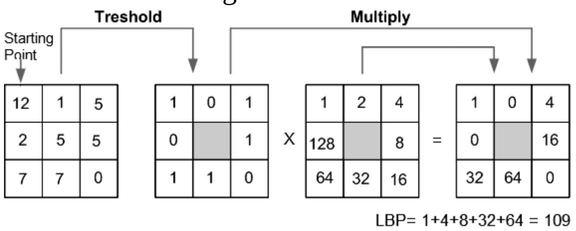

Figure 1. Example of obtaining LBP value.

Zhao et al. [**7**] applied both facial dynamics and Local Binary Pattern on the Three-Orthogonal-Planes (LBP-TOP) and Volume Local Binary Patterns (VLBP) to combine motion and appearance. Xiaohua et al. [**8**] also used LBP-TOP on eyes, nose and mouth separately. They formulated an equation to find the weight for those areas. An extension to the original LBP operator called LBP_{RIU2} was proposed in [**9**]. A binary pattern differs from different angles. According to $LBP_{RIU2,}$ these entire eight directional pattern would be counted as a single bin. For example, an LBP pattern 00000001 can be seen as 00000010 from 45-degree angle; therefore, they will be counted in the same bin. Although LBP features achieved high accuracy rates for facial expression recognition but it is time consuming due to its long feature dimension. Ojansivu et al. [**10**] proposed LPQ (Local Phase Quantization), another appearance-based facial feature extraction method that is blur insensitive. Yang et al. [**11**] used a local binary pattern and local phase quantization together and achieved very good result in case of person independent environment. Li et al (2012) used RI-LPQ (rotation invariant Local Phase Quantization) along with SRC (Sparse Representation-based Classification) classifier and obtained better accuracy than LBP. Kabir et al. [**12**] proposed LDPv (Local Directional Pattern - Variance) by computing weight from local region using local variance and applying it to the feature vector. He used support vector machine, as a classifier and found his method to be effective for facial expression recognition.

The proposed feature representation method captures crucial texture information from local 3x3 pixels area. The referenced pixel is surrounded by eight pixels. Each of these 8 pixel's gray color intensity is used to build two 2-bit binary patterns. It is robust to monotonic gray-scale changes caused, for example, by illumination variations as it uses the magnitude of color difference instead of direct gray value. The feature vector length is only 8, which is suitable for large-scale dataset. In comparison with LBP [**5**] or LPQ[**10**], the proposed method performs better both in time and classification accuracy.

The rest of the paper is organized to explain the proposed local feature representation method in section 2, proposed framework in section 3, data collection and experimental setup in section 4, results and analysis in section 5 and conclusion in section 6.

2. Methodology - Local Gradient Pattern

The proposed representation method is based on facial color gradient differences of a local 3x3 pixels region (see Figure 2).

a1	b1	a2
b4	C	b2
a4	b3	a3

Figure 2. Local 3x3 pixels region

The representations operate on gray scale facial images. The pixel C in Figure 2 is represented by two 2-bit binary patterns e.g., pattern-1 and pattern-2 as shown in Figure 3. a1 to a4 and b1 to b4 represents the gray color intensity of the corresponding pixels. Bit-1 for Pattern-1 is 0 if $a1 \leq a3$ else it is 1. Similar way other bits are also calculated using the formula shown in Figure 3. Pattern-1 can have 2^2=4 different combinations e.g., 00, 01, 10 and 11. Therefore four bins are needed to represent pattern-1. Similarly, another four bins are needed for pattern-2 (see Figure 4).

This new representation is called LGP. An LGP is computed for each pixel of the image. An example of obtaining LGP is shown in Figure 5.

Therefore, for the pixel '25', both bin-3 and bin-6 will be increased by one.

Pattern-1		Pattern-2	
Bit-1	Bit-2	Bit-1	Bit-2
$=\begin{cases}1, & a1 > a3\\ 0, & a1 \le a3\end{cases}$	$=\begin{cases}1, & a2 > a4\\ 0, & a2 \le a4\end{cases}$	$=\begin{cases}1, & b1 > b3\\ 0, & b1 \le b3\end{cases}$	$=\begin{cases}1, & b2 > b4\\ 0, & b2 \le b4\end{cases}$

Figure 3. Single pixel representation using two separate patterns

Pattern-1				Pattern-2			
00	01	10	11	00	01	10	11
Bin-1	Bin-2	Bin-3	Bin-4	Bin-5	Bin-6	Bin-7	Bin-8

Figure 4. Bin representations for the patterns

18	25	14	pattern-1=10
85	25	86	pattern-2=01
45	65	14	

Figure 5. Example of obtaining LGP

3. Proposed Framework for Facial Expressions Recognition

The framework consists of the following steps:

I. For each of training images, convert it to gray scale if in a different format.
II. Detect the face in the image, resize it to 180x180 pixels and divide it into equal sized 20x20 pixels block.
III. Compute feature value for each pixel using LGP.
IV. Construct the histogram for each of 81 blocks.
V. Concatenate the histograms of each block to get the feature vector for the whole image.
VI. Build a multiclass Support Vector Machine for face expression recognition using feature vectors of the training images.
VII. Do step 1 to 5 for each of testing images and use the Multiclass Support Vector Machine from step 6 to identify the face expression of the given testing image.

4. DATASET and Experimental setup

The extended Cohn-Kanade dataset (CK+) [**13**] is used for experiments to evaluate the effectiveness of the proposed method. In CK+, there are 326 peak facial expressions from 123 subjects. Seven emotion categories are there. They are 'Anger', 'Contempt', 'Disgust', 'Fear', 'Happy', 'Sadness' and 'Surprise'.

Figure 6 shows the numbers of instances for each expression in the CK+ dataset. No subject with the same emotion has been collected more than once. All the facial images in the dataset are posed and they are taken in a controlled environment. Some samples from the dataset are shown in Figure 7.

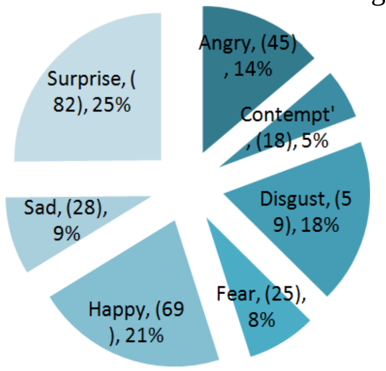

Figure 6. CK+ Dataset, 7 class expressions, number of instances of each expression and percentage.

Figure 7. Cohn-Kanade (CK+) sample dataset

The steps of face detection, preprocessing and feature extraction are illustrated in Figure 8. *fdlibmex* library, free code available for Matlab is used for face detection. The library takes a gray color image as input and returns the frontal face. The returned face is square in size and the resolution varies from 170x170 to 190x190. Therefore, all the detected faces are normalized to 180x180 pixels dimension for the experimental purpose. The face is then masked using an elliptical shape as shown in Figure 9.

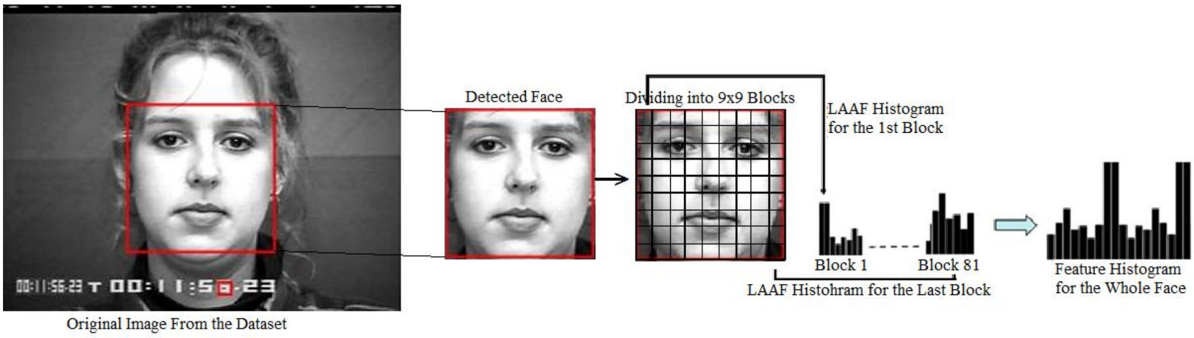

Figure 8. Steps of Facial Feature Extraction

Figure 9. Elliptical shaped cropped sample faces from CK+ dataset.

The input face is divided into several equal sized blocks, because a block can give more location information. In the experiment, the 180x180 sized face is equally divided into 9x9=81 blocks of 20x20 pixels each. An LGP is extracted for each pixel from each block. Concatenating feature histograms of all the blocks produces a unique feature vector for the given image. In the experiments, a non-overlapping ten-fold cross validation is applied. LIBSVM [**14**], a multiclass support vector machine is used for classification. 90% of 326 images (90% from each expression) are used in each fold as the training input and the remaining 10 % of images are used as testing.
Ten rounds of this experiment are conducted and the average confusion matrix is reported. The kernel parameters for the classifier are set to: s=0 for SVM type C-Svc, t=1 for polynomial kernel function, c=100 is the cost of SVM, g= 1/ (length of feature vector), b=1 for a probability estimation. This setting of LIBSVM is found to be suitable for CK+ dataset with seven classes of data.

5. Experimental results and analyses

The feature vector length and feature dimension for 81 blocks are compared with some other methods in Table 1.
Face dimension, block dimension and the number of blocks are calculated before experiments. According to Table 2 and Table 3, the number of blocks considered for further experiments is 9x9=81. Combinations like 9x8, 8x9, 6x9, 9x6, 8x10, 10x8 are also tried but 9x9=81 is found to be the best among all combinations.
The confusion matrix for LGP is shown in

Table **4**. Confusion matrix makes it easy to see if the system is confusing two classes or not.

Table 1. Feature Vector Dimension of proposed methods and some other popular methods.

Method	Feature vector Length	Feature vector Dimension used in our experiment	
LBP (Local Binary pattern)[**5**]	256	256x9x9	=20736
LPQ (Local Binary pattern)[**10**]	256	256x9x9	=20736
LDPv (Local Directional Pattern Variance)[**12**]	56	56x9x9	=4536
LGP (Proposed Method)	8	8x9x9	=648

Table 2. Block Dimension vs. Classification Accuracy

Blocks	Block Dimension (Pixels)	Classification Accuracy (%) LGP
6x6	30x30	88.52%
9x9	20x20	91.90%
10x10	18x18	90.75%
12x12	15x15	89.42%
15x15	12x12	89.34%
18x18	10x10	89.71%

Table 3. Face Dimension vs. Classification Accuracy

Face Dimension (Pixels)	Blocks	Classification Accuracy (%) LGP
240x240	12x12	89.28%
220x220	11x11	89.10%
200x200	10x10	89.80%
180x180	9x9	91.90%
160x160	8x8	89.25%
140x140	7x7	89.63%

Table 4. Confusion matrices for FERS using LGP (Accuracy= 91.9%)

C :		Actual						
		Angry	Contempt	Disgust	Fear	Happy	Sad	Surprise
prediction	Angry	**84.4**	4.4	4.4	0.0	0.0	6.7	0.0
	Contempt	5.6	**83.3**	0.0	5.6	0.0	0.0	5.6
	Disgust	4.9	0.0	**93.4**	1.6	0.0	0.0	0.0
	Fear	4.0	8.0	4.0	**60.0**	12.0	0.0	12.0
	Happy	0.0	0.0	0.0	0.0	**100.0**	0.0	0.0
	Sad	7.1	3.6	7.1	3.6	0.0	**78.6**	0.0
	Surprise	1.2	0.0	0.0	0.0	0.0	0.0	**98.8**

From the confusion matrix, it is clear that the most confusing expression classes are anger, contempt, fear and sad. The highest accuracy for class anger is obtained by the proposed method i.e., LGP. It should be noted that the results are not directly comparable due to different experimental setups, different versions with different emotion labels, preprocessing methods, and the number of instances used, but they still point out the discriminative power of each approach. **Error! Reference source not found.** compares LGP with some other static analysis methods in terms of the number of sequences and classification accuracy. An LGP clearly outperforms the others in almost all cases. It takes 0.0057 second to extract features from face of 180x180 resolutions, which cannot be compared with other methods, as feature extraction time is not cited in their papers.

Table 6 shows that proposed texture based expression recognition is better than previous texture based methods and compares favorably with state-of-the-art procedures using shape or shape and texture information combined.

Table 5. Comparison with Different Approaches, Number of expression classes=7 and non-dynamic.

Author	Number of subjects (Dataset)	Number of sequences	Classification accuracy (%)
LGP	123(CK+)	326	92%
[15]	123(CK+)	327	90%
[16]	96(--)	320	88%
[17]	90(--)	313	87%
[18]	123(CK+)	327	87%
[19]	123(CK+)	327	82%

Table 6. Results comparison, using different methods on (CK+/CK) dataset with seven main facial expressions. S: shape based method, T: texture based method. An. = Anger, Co. = Contempt, Di. = Disgust, Fe. =Fear, Ha. =Happy, Sa. =Sad, Su. =Surprise, Ne.=Neutral)

Method	T/S	Dataset	An.	Co.	Ne.	Di.	Fe.	Ha.	Sa.	Su
***LGP**	T	CK+	84%	83%	-	93%	60%	100%	79%	99%
***LBP**	T	CK+	79%	94%	-	97%	72%	100%	79%	99%
***LPQ**	T	CK+	62%	67%	-	78%	52%	93%	43%	96%
LDPv	T	CK	79%	-	90%	93%	93%	100%	92%	99%

*Experiments using LGP, LBP and LPQ are conducted using same experimental setup.

6. Conclusion

The proposed method in this paper achieved classification accuracy of 92%, which is better than other available methods in the person dependent environment. It computes the feature pattern for a single pixel using the gray color value differences of its surrounding pixels, which is independent of grayscale change caused by monotonic illumination changes. The facial region is divided into 81 equal sized blocks and histogram of LGP codes computed for each block is concatenated to build the feature vector, which uniquely represents the face. The best thing is that it has a very tiny feature vector length. Therefore, the method is suitable for a large-scale dataset consists of high dimensional photos. The classification accuracy obtained using multiclass support vector machine (LIBSVM) outperforms

most of the appearance-based methods proposed in the last few decades. In future, some boosting algorithm can be incorporated with the classifier. Face alighnment is also a big issue for classification. In CK+ dataset, pictures are taken in a controlled environment. Thus, it is not natural and also the instances for different classs of expressions are not equal. Experiments on equally distributed data might increase the accuracy level.

References

[1] Mehrabian, A. (1968). Communication without words. Psychological today, 2, 53-55.

[2] Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A survey of affect recognition methods: Audio,

visual, and spontaneous expressions. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 31(1), 39-58.

[3] Tong, Y., Liao, W., & Ji, Q. (2007). Facial action unit recognition by exploiting their dynamic and semantic relationships. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 29(10), 1683-1699.

[4] Zhang, Z., Lyons, M., Schuster, M., & Akamatsu, S. (1998, April). Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron. InAutomatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on (pp. 454-459). IEEE.

[5] Ojala, T., Pietikäinen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. Pattern recognition, 29(1), 51-59.

[6] Ahonen, T., Hadid, A., & Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 28(12), 2037-2041.

[7] Zhao, G., & Pietikainen, M. (2007). Dynamic texture recognition using local binary patterns with an application to facial expressions. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 29(6), 915-928.

[8] Expression recognition in videos using a weighted component-based feature descriptor. In Image Analysis (pp. 569-578). Springer Berlin Heidelberg.

[9] Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 24(7), 971-987.

[10] Ojansivu, V., & Heikkilä, J. (2008). Blur insensitive texture classification using local phase quantization. In Image and Signal Processing (pp. 236-243). Springer Berlin Heidelberg.

[11] Yang, S., & Bhanu, B. (2012). Understanding discrete facial expressions in video using an emotion avatar image. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 42(4), 980-992.

[12] Kabir, H., Jabid, T., & Chae, O. (2012). Local Directional Pattern Variance (LDPv): A Robust Feature Descriptor for Facial Expression Recognition. The International Arab Journal of Information Tecnology,9(4), 382-391.

[13] Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010, June). The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on (pp. 94-101). IEEE.

[14] Chang, C. C., & Lin, C. J. (2011). LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST), 2(3), 27.

[15] Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., & Bartlett, M. (2011, March). The computer expression recognition toolbox (CERT). In Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on (pp. 298-305). IEEE.

[16] Shan, C., Gong, S., & McOwan, P. W. (2005, September). Robust facial expression recognition using local binary patterns. In Image Processing, 2005. ICIP 2005. IEEE International Conference on (Vol. 2, pp. II-370). IEEE.

[17] Bartlett, M. S., Littlewort, G., Fasel, I., & Movellan, J. R. (2003, June). 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 (Vol. 5, pp. 53-53). IEEE.

[18] Jeni, L. A., Lőrincz, A., Nagy, T., Palotai, Z., Sebők, J., Szabó, Z., & Takács, D. (2012). 3D shape estimation in video sequences provides high precision evaluation of facial expressions. Image and Vision Computing, 30(10), 785-795.

[19] CL, S. N., Jha, S. S., Das, P. K., & Nair, S. B. (2012). Automatic Facial Expression Recognition Using Extended AR-LBP. In Wireless Networks and Computational Intelligence (pp. 244-252). Springer Berlin Heidelberg.