Person-Independent Facial Expression Recognition Based on Compound Local Binary Pattern (CLBP)

Faisal Ahmed[1], Hossain Bari[2], and Emam Hossain[3]
[1]Department of Computer Science and Engineering, Islamic University of Technology, Bangladesh
[2]Samsung Bangladesh R & D Center Ltd, Bangladesh
[3]Department of Computer Science and Engineering, Ahsanullah University of Science and Technology, Bangladesh

Abstract: *Automatic recognition of facial expression is an active research topic in computer vision due to its importance in both human-computer and social interaction. One of the critical issues for a successful facial expression recognition system is to design a robust facial feature descriptor. Among the different existing methods, the Local Binary Pattern (LBP) has been proved to be a simple and effective one for facial expression representation. However, the LBP method thresholds P neighbors exactly at the value of the center pixel in a local neighborhood and encodes only the signs of the differences between the gray values. Thus, it loses some important texture information. In this paper, we present a robust facial feature descriptor constructed with the Compound Local Binary Pattern (CLBP) for person-independent facial expression recognition, which overcomes the limitations of LBP. The proposed CLBP operator combines extra P bits with the original LBP code in order to construct a robust feature descriptor that exploits both the sign and the magnitude information of the differences between the center and the neighbor gray values. The recognition performance of the proposed method is evaluated using the Cohn-Kanade (CK) and the Japanese Female Facial Expression (JAFFE) database with a Support Vector Machine (SVM) classifier. Experimental results with prototypic expressions show the superiority of the CLBP feature descriptor against some well-known appearance-based feature representation methods.*

Keywords: *Facial expression recognition, feature descriptor, LBP, SVM, texture encoding.*

Received April 26, 2012; accepted December 31, 2012; published online January 29, 2013

1. Introduction

Facial expression provides a non-verbal form of communication that facilitates the cognition of human emotions and intensions [1]. Automated facial expression analysis is an interesting task that has attracted much attention in the recent years due to its potential applicability in various areas, such as human-computer interaction, data-driven animation, and customized applications for consumer products [2]. Deriving an efficient and effective feature representation that can minimize the within-class variations while maximizing the between-class variations is the fundamental component for any successful facial expression recognition system [14]. However, the inherent variability of facial images caused by different factors like variations in illumination, pose, alignment, and occlusions makes expression recognition a challenging task. Therefore, the aim of the ongoing research in automated expression recognition is to increase the robustness of the underlying feature descriptor against these factors.

Some surveys on facial feature representations for face recognition [28] and expression analysis [9] addressed these challenges and possible solutions in detail. Based on the type of features used, expression recognition approaches can be broadly divided into two categories, namely geometric feature-based methods and appearance-based methods [23]. Early methods for facial feature extraction were mostly based on the geometric relationships (e.g., positions, distances, and angles) between different facial components. Facial Action Coding System (FACS) introduced by Ekman and Friesen [7] is one of the most popular geometric feature-based methods that represents facial expression using a set of Action Units (AU), where each action unit corresponds to the physical behavior of a specific facial muscle. Later, Zhang [27] proposed a feature extraction method based on the geometric positions of 34 manually selected fiducial points. A similar representation was adopted by Guo and Dyer [11], where they employed linear programming in order to perform simultaneous feature selection and classifier training. Recently, Valstar *et al.* [25, 26] have studied facial expression analysis based on tracked fiducial point data and reported that, geometric features provide similar or better performance than appearance-based methods in action unit recognition. However, geometric methods are difficult to accommodate in many situations as they rely on accurate detection of facial components [14].

Appearance-based methods employ image filter or filter bank on the whole face or some specific regions of the facial image in order to extract changes in facial appearance. Principal Component Analysis (PCA) [19] and Independent Component Analysis (ICA) [5, 9] are the common appearance-based methods. PCA utilizes only the holistic information of an image, where ICA can also be used to extract local information. In addition, other local appearance-based methods, such as Gabor-wavelets [17, 22] and local feature analyses [6] are also explored in the literature. Recently, facial expression analyses based on Local Binary Pattern (LBP) [20], and its variants have gained much popularity for their superior performances. The LBP operator was originally introduced for texture analyses [18] and later this method has been successfully applied in face authentication and facial expression recognition. The LBP method extracts local texture information by thresholding P neighbors at the value of the central pixel in a local neighborhood, which is computationally efficient and robust to monotonic illumination variation. Although LBP provides a theoretically simple and efficient approach to facial expression analyses, it has some limitations. Firstly, it shows poor performance in the presence of random noise [29]. To address this issue, Local Ternary Pattern (LTP) [21] has been presented with one additional discrimination level than LBP in order to increase the robustness against noise in uniform and near-uniform regions. Secondly, LBP method only considers the sign of the difference between two gray values and thus, discards the magnitude of the difference which is very important texture information. To exploit the magnitude information, Local Directional Pattern (LDP) [14] and LDPv [15] were introduced. Instead of considering gray level values, both LDP and LDPv employ the magnitude of the edge response values in different directions in order to encode the texture information of a local region. However, LDP still generates inconsistent codes in uniform and smooth regions and heavily depends on the number of prominent edge directions [2].

In this paper, we present a robust feature descriptor constructed with the Compound Local Binary Pattern (CLBP), an extension of the LBP for person-independent facial expression recognition. Unlike the original LBP operator that uses P bits to encode only the signs of the differences between the center pixel and the P neighbor gray values, the proposed method employs $2P$ bits, where the additional P bits are used to encode the magnitude information of the differences between the center and the neighbor gray values using a threshold. The motivation behind the proposed encoding scheme is to increase the robustness of the feature descriptor by incorporating additional local information that is discarded by the original LBP operator. The performance of the CLBP feature representation is evaluated in terms of classification rate using Support Vector Machine (SVM). Experiments with the Cohn-Kanade (CK) [16] and the Japanese Female Facial Expression (JAFFE) database [17] demonstrate that, the proposed CLBP operator is more robust in extracting facial information and provides higher classification rate compared to some existing feature representation techniques.

2. Local Binary Pattern

LBP is a gray-scale and rotation invariant texture primitive that describes the spatial structure of the local texture of an image. The LBP operator selects a local neighborhood around each pixel of an image, thresholds the P neighbor gray values with respect to the center pixel and concatenates the result binomially. The resulting binary value is then assigned to the center pixel. Formally, the LBP operator can be described as:

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(i_p - i_c)2^p \quad (1)$$

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (2)$$

Here, i_c is the gray value of the center pixel (x_c, y_c), i_p is the gray value of its neighbors, P is the number of neighbors and R is the radius of the neighborhood. The basic LBP encoding process is illustrated in Figure 1.

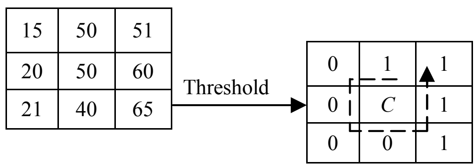


Figure 1. Illustration of the basic LBP operator. Here, the LBP code = 10000111 for pixel C.

In practice, the LBP operator considers the signs of the differences of the gray values of P equally spaced neighbors with respect to the central pixel, which is then represented using a P-bit binary number. If any neighbor does not fall exactly on a pixel position, then the value of that neighbor is estimated using bilinear interpolation. The histogram of the encoded image block obtained by applying the LBP operator is then used as a texture descriptor for that block.

One extension to the original LBP operator, known as the uniform LBP, exploits certain LBP patterns, which appear more frequently in a significant area of the image. These patterns contain very few bitwise transitions from 0 to 1 or vice versa in a circular sequence of bits. One example of a uniform pattern is 00011111. It has only one transition from 0 to 1. Ojala *et al.* [18] observed that, uniform LBP patterns are the fundamental properties of texture, which provide a vast majority of all the LBP patterns present in any texture image. Therefore, uniform patterns are able to describe significant local texture information, such as bright spot, flat area or dark spot, and edges of varying positive and negative curvature [18].

The basic LBP operator discards the magnitude information of the differences between the center and the neighbor gray values in a local neighborhood. As a result, the LBP method tends to produce inconsistent codes in many cases. One example is shown in Figure 2. Here, the 8-bit uniform LBP code (11111111) corresponds to a flat area or a dark spot at the center pixel [18], which is not consistent with the local region.

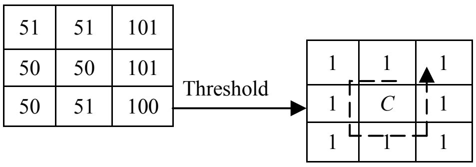


Figure 2. Generation of inconsistent binary pattern in the LBP encoding approach.

3. Compound Local Binary Pattern

3.1. Basic CLBP Encoding Method

The original LBP encoding scheme considers only the sign of the difference between two gray values and thus, it often fails to generate binary codes consistent with the texture property of a local region. Being motivated by this, we propose CLBP, an extension of the original LBP operator that assigns a *2P*-bit code to the center pixel based on the gray values of a local neighborhood comprising *P* neighbors. Unlike the LBP operator that employs one bit for each neighbor to represent only the sign of the difference between the center and the corresponding neighbor gray values, the proposed method uses two bits for each neighbor in order to encode the sign as well as the magnitude information of the difference between the center and the neighbor gray values. Here, the first bit represents the sign of the difference between the center and the corresponding neighbor gray values like the basic LBP encoding. The other bit is used to encode the magnitude of the difference with respect to a threshold value, which is the average magnitude M_{avg} of the difference between the center and the neighbor gray values in the local neighborhood of interest. The CLBP operator sets this bit to 1 if the magnitude of the difference between the center and the corresponding neighbor is greater than the threshold M_{avg}. Otherwise, it is set to 0. Thus, the indicator $s(x)$ of equation 2 is replaced by the following function:

$$s(i_p, i_c) = \begin{cases} 00 & i_p - i_c < 0, \quad |i_p - i_c| \le M_{avg} \\ 01 & i_p - i_c < 0, \quad |i_p - i_c| > M_{avg} \\ 10 & i_p - i_c \ge 0, \quad |i_p - i_c| \le M_{avg} \\ 11 & otherwise \end{cases} \tag{3}$$

Here, i_c is the gray value of the center pixel, i_p is the gray value of a neighbor p, and M_{avg} is the average magnitude of the difference between i_p and i_c in the local neighborhood. The basic CLBP operator is illustrated in Figure 3.

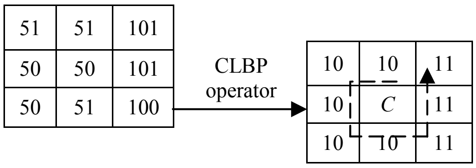


Figure 3. Illustration of the basic CLBP operator. Here, the CLBP code = 1010101010111111 for pixel *C*.

From Figure 3, it can be observed that, unlike the LBP encoding shown in Figure 2, the proposed CLBP method discriminates the neighbors in the north-east, east, and south-east directions as they have higher gray values than the other neighbors and thus, produces a pattern consistent with the local texture property.

3.2. Generation of Sub-CLBP Codes

In a 3×3 neighborhood, the proposed CLBP method encodes an image by operating on the 8 neighbors around the central pixel and assigning a 16-bit code to that pixel. As 16-bit codes are used to label the pixels, the number of possible binary patterns is 2^{16}. To reduce the number of features, He and Cercone [12] proposed to consider less number of neighbors while forming the binary patterns. Thus, the length of the feature vector can be reduced by discarding some degree of neighborhood information. In this paper, we have presented a different approach where all the CLBP binary patterns are further split into two sub-CLBP patterns. Each sub-CLBP pattern is obtained by concatenating the bit values corresponding to $P/2$ neighbors, where P is the number of neighbors. Formally, in a local neighborhood, the two sub-CLBP patterns are formed by concatenating the corresponding values of the bit sequence (1, 2, 5, 6, …, $2P$-3, $2P$-2) and (3, 4, 7, 8, …, $2P$-1, $2P$), respectively of the $2P$-bit original CLBP code.

In other words, a 16-bit CLBP pattern is split into two 8-bit sub-CLBP patterns, where the first one sub-$CLBP_1$ is obtained by concatenating the bit values corresponding to the neighbors in the north, east, south, and west directions, respectively and the second sub-CLBP pattern sub-$CLBP_2$ is obtained by concatenating the bit values corresponding to the neighbors in the north-east, south-east, south-west, and north-west directions, respectively. Thus, this method reduces the number of possible patterns significantly, which results in a total of 2^8 distinct sub-CLBP patterns. The process is illustrated in Figure 4. The two sub-CLBP patterns are treated as separate binary codes and combined during the feature vector generation.

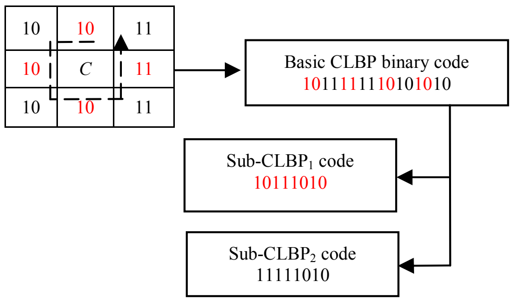


Figure 4. Generation of the two sub-CLBP patterns 10111010 and 11111010 from the original CLBP code 1011111110101010.

3.3. CLBP Feature Descriptor

After applying the CLBP operator on all the pixels of an image and splitting all the 16-bit CLBP patterns into the corresponding sub-CLBP patterns, we get two 8-bit binary codes for each pixel of the image. Thus, two encoded image representations are obtained for the two sub-CLBP patterns. Histograms generated from these two encoded images are then concatenated to form a spatially combined histogram, the CLBP histogram, which functions as a feature representation for the expression image. Figure 5 illustrates the CLBP histogram generation process from a sample expression image.

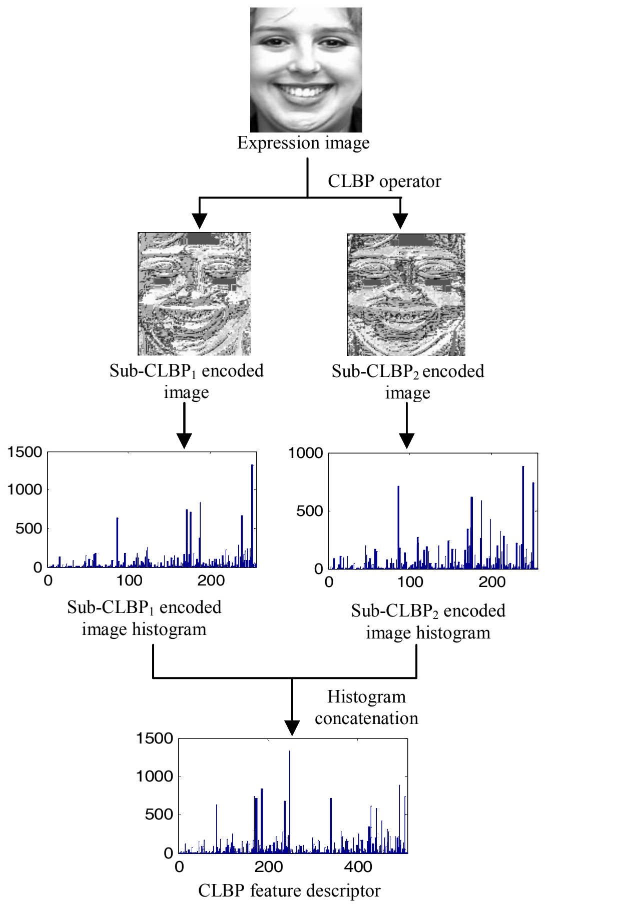


Figure 5. Illustration of the CLBP histogram (feature descriptor) generation process.

Histograms generated from the whole encoded image contain no location information of the micro-patterns, but merely their occurrences are expressed. However, presence of location information and spatial relationships provides a better facial feature representation and describes the image content more accurately [3, 10, 14]. Therefore, the CLBP histogram is modified to an extended histogram in order to incorporate some degree of location information. First, each image is partitioned into a number of regions and individual CLBP histograms are generated from each of those regions. Finally, the histograms of all the regions are concatenated to obtain the extended CLBP histogram. For the facial expression recognition process, this histogram collection is used as the facial feature vector. The extended histogram generation process is shown in Figure 6.

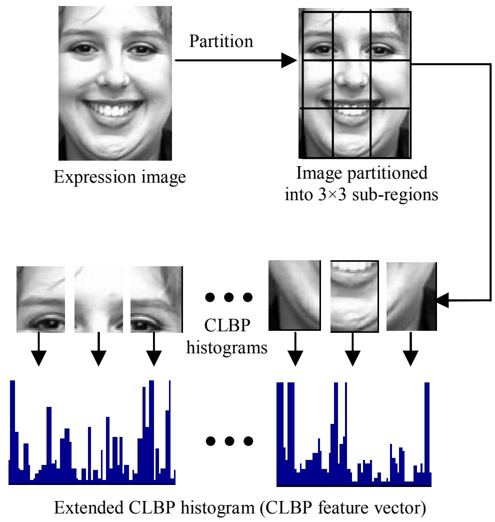


Figure 6. Each expression image is partitioned into a number of sub-regions, and the individual CLBP histograms generated from each of the sub-regions are concatenated to form the CLBP feature vector.

4. Classification Using Support Vector Machine

SVM is a state-of-the-art machine learning approach based on the modern statistical learning theory. It has been successfully applied in different classification problems. SVM performs the classification by constructing a hyper plane in such a way that the separating margin between positive and negative examples is optimal. This separating hyper plane then works as the decision surface. Given a set of labeled training samples $T=\{(x_i, l_i), i=1, 2, \dots, L\}$, where $x_i \in R^P$ and $l_i \in \{-1, 1\}$, a new test data x is classified by:

$$f(x) = sign\left(\sum_{i=1}^{L} \alpha_i l_i K(x_i, x) + b\right) \tag{4}$$

Here, α_i are Lagrange multipliers of dual optimization problem, b is a threshold parameter, and K is a kernel function. The hyper plane maximizes the separating margin with respect to the training samples with $\alpha_i>0$, which are called the support vectors.

SVM makes binary decisions. To achieve multi-class classification, the common approach is to adopt the one-against-rest or several two-class problems. In our study, we used the one-against-rest approach. Radial Basis Function (RBF) kernel was used for the classification problem. The radial basis function K can be defined as:

$$K(x_i, x) = exp(-\gamma \parallel x_i - x \parallel^2), \quad \gamma > 0 \tag{5}$$

$$\parallel x_i - x \parallel^2 = (x_i - x)^t (x_i - x) \tag{6}$$

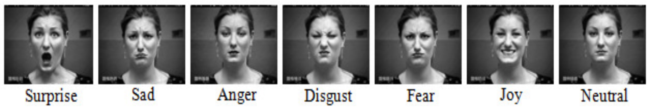
Here, γ is a kernel parameter. A grid-search was carried out for selecting appropriate parameter value, as suggested in [13].

5. Experimental Setup

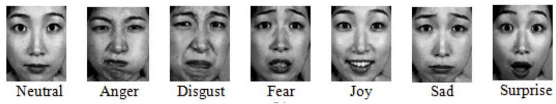
The recognition ability of the proposed method was evaluated based on a set of prototypic emotional expressions, which includes anger, disgust, fear, joy, sadness, and surprise. This 6-class expression can be further extended to a 7-class expression set by adding neutral face expression images. The performance evaluation was performed with two well-known image databases, namely the CK facial expression database [16] and the JAFFE database [17].

The CK database comprises 100 university students who were around 18 to 30 years old at the time of image acquisition. Among them, 65% were female, 15% were African-American, and 3% were Asian or Latino. A series of facial expression displays were performed by the subjects starting from neutral or near-neutral to one of the six prototypic emotional expressions stated before. The image sequences were digitized into 640×480 or 640×690 pixel resolution. In our setup, we first selected 1224 face image sequences from a total of 96 subjects, where each of the images was labeled as one of the six prototypic expressions. This 6-class expression dataset was then extended to a 7-class expression dataset by including additional 408 images of neutral expression face. Figure 7-a shows the sample prototypic expression images from the CK database.

The JAFFE database comprises facial expression images of 10 Japanese female subjects. All the images were digitized into a resolution of 256×256 pixels. The images were taken from a frontal pose, and the subjects' hair was tied back in order to facilitate the exposure of all the expressive zones of the face. In the image scene, an even illumination was created using tungsten lights. Instead of revealing the actual names, the subjects are referred with their initials, which are KA, KL, KM, KR, MK, NA, NM, TM, UY, and YM. In our setup, the 6-class expression dataset comprises a total of 283 images, while the 7-class expression set includes additional 50 neutral expression images. Figure 7-b shows the sample prototypic expression images from the JAFFE database.



a) CK database.



b) JAFFE database.

Figure 7. Sample images of each prototypic expressions.

The selected images were cropped from the original ones based on the positions of the two eyes and normalized to 150×110 pixels. The ground-truth of eye position data was provided for cropping. No alignment of facial features (such as alignment of mouth) was performed in our setup. Figure 8 shows a sample cropped facial image from CK database.

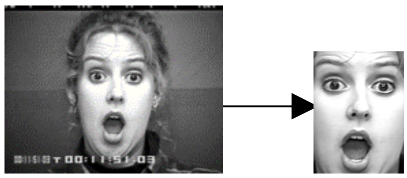


Figure 8. Cropping of a sample face image from the original one.

6. Results and Discussion

To evaluate the effectiveness of the proposed method, we carried out a ten-fold cross-validation scheme to measure the classification rate. In a ten-fold cross-validation, the whole dataset is randomly partitioned into ten subsets, where each subset comprises an equal number of instances. After that, one subset is used as the testing set and the classifier is trained on the remaining nine subsets. The average classification rate is calculated after repeating the above process for ten times.

The classification rate of the proposed method can be influenced by adjusting the number of regions into which the expression images are to be partitioned. We considered three cases in our experiments, where images were divided into 3×3, 5×5, and 7×6 regions. The performance of the CLBP feature descriptor is also compared with some widely-used facial feature representation approaches, namely LBP [20] and LTP [21].

6.1. Experiments with the CK Database

The CLBP feature descriptor achieves an excellent recognition accuracy of 94.4% for the 6-class expression dataset. On the other hand, for the 7-class dataset, the recognition accuracy is 90.4%, where inclusion of neutral expression results in a decrease in the accuracy. The reason is that, in this case more sample expressions are confused as neutral expression. For both the 6-class and the 7-class recognition problem, the highest classification rate is obtained for images partitioned into 5×5 images. Tables 1 and 2 shows the recognition rate of the 6-class and the 7-class expression datasets, respectively using different feature descriptors. It can be observed that, both LTP and CLBP achieves better recognition rate than LBP. However, the highest classification rate is achieved using CLBP for both the 6-class and the 7-class expression datasets.

Table 1. Recognition rate (%) for the CK 6-class expression dataset using different feature descriptors.

Operator	Classification rate (%) for different number of regions		
	3×3	5×5	7×6
LBP	79.3	89.7	90.1
LTP	87.3	92.3	93.6
CLBP	88.2	94.4	94.2

Table 2. Recognition rate (%) for the CK 7-class expression dataset using different feature descriptors.

Operator	Classification rate (%) for different number of regions		
	3×3	5×5	7×6
LBP	73.8	80.9	83.3
LTP	81.3	88.5	88.9
CLBP	82.1	90.4	89.2

Tables 3 and 4 show the confusion matrix of recognition using CLBP feature descriptor for the 6-class and the 7-class CK dataset, respectively in order to provide a better picture of the recognition accuracy of individual expression types for images partitioned into 5×5 regions. It can be observed that, one of the main reasons for the decrease of recognition rate in the 7-class problem is the misclassification of 20.3% sad expression images as neutral ones. We have also compared our method some other appearance-based feature descriptors, namely Gabor features [4], ICA [24], and enhanced ICA [24]. Table 5 shows the recognition rate of these methods for the 6-class expression dataset.

Table 3. Confusion matrix of CK 6-class recognition using CLBP for images partitioned into 5×5 regions.

	Anger (%)	Disgust (%)	Fear (%)	Joy (%)	Sad (%)	Surprise (%)
Anger	**97.1**	0	0	0	0	2.9
Disgust	0.6	**98.8**	0.6	0	0	0
Fear	0	2.5	**95.0**	0	0	2.5
Joy	1.7	0.6	0	**96.0**	1.7	0
Sad	0	1.3	0	0	**98.7**	0
Surprise	8.3	0	0	10.5	0	**81.2**

Table 4. Confusion matrix of CK 7-class recognition using CLBP for images partitioned into 5×5 regions.

	Anger (%)	Disgust (%)	Fear (%)	Joy (%)	Sad (%)	Surprise (%)	Neutral (%)
Anger	**95.7**	0	0	0	0	0.8	3.5
Disgust	0	**94.9**	1.9	0	0	3.2	0
Fear	0	4.4	**93.8**	0	0	0	1.8
Joy	1.8	0	0	**94.6**	0	1.8	1.8
Sad	0	1.0	0	0	**78.7**	0	20.3
Surprise	2.1	5.3	1.1	4.3	1.4	**85.8**	0
Neutral	5.3	0	0	0	5.3	0	**89.4**

Table 5. Comparison of the performance of CLBP and different appearance-based methods against the CK 6-class dataset.

Feature descriptor	Recognition rate (%)
Gabor	89.2
ICA	62.3
EICA	66.7
CLBP	94.2

6.2. Experiments with the JAFFE Database

For the JAFFE database, the CLBP feature descriptor achieves classification rates of 92.2% and 87.5% for the 6-class and the 7-class expression datasets, respectively. Table 6 and Table 7 show the recognition rates of different feature descriptors with the JAFFE 6-class and the 7-class expression datasets, respectively. It can be observed that, CLBP achieves the highest classification rate among the 3 methods. The recognition rate in the JAFFE database is relatively lower than the CK database. The main reason is the incorrect labeling of some of the facial expression images in the JAFFE database. Figure 9 shows examples of incorrect labeling of expression images in the JAFFE database.

Sad Surprise Joy

Figure 9. Incorrect labeling of expression images in the JAFFE database.

Table 6. Recognition rate (%) for the JAFFE 6-class expression dataset using different feature descriptors.

Operator	Classification rate (%) for different number of regions		
	3×3	5×5	7×6
LBP	84.1	87.6	90.5
LTP	84.3	87.9	90.9
CLBP	85.5	89.8	92.2

Table 7. Recognition rate (%) for the JAFFE 7-class expression dataset using different feature descriptors.

Operator	Classification rate (%) for different number of regions		
	3×3	5×5	7×6
LBP	81.5	82.3	85.3
LTP	84.6	85.0	86.7
CLBP	85.3	85.3	87.5

Tables 8 and 9 show the confusion matrix of recognition using CLBP feature descriptor for the 6-class and the 7-class JAFFE dataset, respectively. Here also, the inclusion of neutral expression images result in lower classification rate for the 7-class expression dataset.

Table 8. Confusion matrix of JAFFE 6-class recognition using CLBP for images partitioned into 7×6 regions.

	Anger (%)	Disgust (%)	Fear (%)	Joy (%)	Sad (%)	Surprise (%)
Anger	**97.5**	0	0	0	2.5	0
Disgust	8.7	**89.1**	0	0	2.2	0
Fear	0	1.3	**89.8**	3.7	3.4	1.8
Joy	0	0	0	**97.7**	0	2.3
Sad	4.4	1.9	4.6	3.7	**85.4**	0
Surprise	0	0	5.5	0	0	**94.5**

Table 9. Confusion matrix of JAFFE 7-class recognition using CLBP for images partitioned into 7×6 regions.

	Anger (%)	Disgust (%)	Fear (%)	Joy (%)	Sad (%)	Surprise (%)	Neutral (%)
Anger	**92.7**	4.9	0	0	2.4	0	0
Disgust	0	**95.1**	0	0	4.9	0	0
Fear	0	4.0	**80.0**	2.0	6.0	0	8.0
Joy	0	0	0	**97.6**	0	2.4	0
Sad	10.3	2.1	6.3	2.1	**77.1**	0	2.1
Surprise	0	0	7.3	0	0	**90.3**	2.4
Neutral	4.0	0	4.0	4.0	2.0	7.0	**79.0**

6.3. Experiments with Low-Resolution Images

Automated facial analysis is useful in smart meeting, surveillance and many other applications where often only low-resolution video data is available. Since geometric methods like detection of facial action units are difficult to accommodate in these scenarios, appearance-based methods seem to be a better solution. Therefore, the performance of the proposed method is also evaluated on low-resolution images. Experiments were conducted on images from the JAFFE 6-class and the JAFFE 7-class expression datasets. We considered 3 different image resolutions: 75×55, 48×36, and 37×27, as shown in Figure 10. The original images were down-sampled to obtain these low-resolution images. All the images were partitioned into 7×6 regions while forming the feature vector. Here also, the performance of the CLBP feature descriptor is compared with LBP and LTP. Table 10 and Table 11 show the classification rate of different feature descriptors against the JAFFE 6-class and the 7-class datasets, respectively.

75×55 48×36 37×27

Figure 10. Sample low-resolution images from the JAFFE database.

From the experimental results, it can be said that, facial feature representation based on the CLBP is more robust and provides higher classification rate than some existing feature representation methods, even with low resolution images. The superiority of the CLBP encoding is due to the utilization of the magnitude of the difference between the center and the neighbor gray values by integrating it with the basic LBP pattern to get a compound binary code, which preserves some important texture information discarded by the original LBP operator. Thus, this method provides an effective and efficient approach to person-independent facial expression recognition.

Table 10. Classification rate (%) of different feature descriptors for low-resolution images from the JAFFE 6-class dataset.

Operator	**Classification rate (%) for different resolutions**		
	75×55	**48×36**	**37×27**
LBP	83.9	81.7	77.9
LTP	86.4	84.2	81.1
CLBP	90.5	88.4	84.3

Table 11. Classification rate (%) of different feature descriptors for low-resolution images from the JAFFE 7-class dataset.

Operator	**Classification rate (%) for different resolutions**		
	75×55	**48×36**	**37×27**
LBP	83.1	80.2	75.3
LTP	85.8	83.5	80.6
CLBP	86.5	85.3	82.8

7. Conclusions

A new local texture pattern, the compound local binary pattern, and a feature descriptor constructed with the CLBP codes have been presented for facial expression recognition. The proposed method utilizes an encoding scheme that combines the magnitude information of the difference between two gray values with the original LBP pattern and thus provides increased robustness in many situations where LBP fails to generate consistent codes. Experimental results show that, the CLBP operator provides an effective and efficient approach for facial feature representation with high discriminative ability, which outperforms several existing feature representation methods. In future, we plan to incorporate temporal information with the CLBP method to recognize facial expressions in sequence images.

References

[1] Ahmed F., Hossain E., Bari A., and Shihavuddin A., "Compound Local Binary Pattern for Robust Facial Expression Recognition," *in Proceedings of the 12th IEEE International Symposium on Computational Intelligence and Informatics*, Hungary, pp. 391-395, 2011.

[2] Ahmed F. and Kabir M., "Directional Ternary Pattern for Facial Expression Recognition," *in Proceedings of the IEEE International Conference on Consumer Electronics*, Las Vegas, pp. 265-266, 2012.

[3] Ahonen T., Hadid A., and Pietkainen M., "Face Description with Local Binary Patterns: Application to Face Recognition," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037-2041, 2006.

[4] Bartlett M., Littlewort G., Frank M., Lainscsek C., Fasel I., and Movellan J., "Recognizing Facial Expression: Machine Learning and Application to Spontaneous Behavior," *in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, vol. 2, pp. 568-573, 2005.

[5] Bartlett M., Movellan J., and Sejnowski T., "Face Recognition by Independent Component Analysis," *IEEE Transaction on Neural Networks*, vol. 13, no. 6, pp. 1450-1464, 2002.

[6] Donato G., Bartlett M., Hager J., Ekman P., and Sejnowski T., "Classifying Facial Actions,"

IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 21, no. 10, pp. 974-989, 1999.

[7] Ekman P. and Friesen W., *Facial Action Coding System: A Technique for Measurement of Facial Movement*, Consulting Psychologists Press, USA, 1978.

[8] Fa C. and Shin F., "Recognizing Facial Action Units using Independent Component Analysis and Support Vector Machine," *Pattern Recognition*, vol. 39, no. 9, pp. 1795-1798, 2006.

[9] Fasel B. and Luettin J., "Automatic Facial Expression Analysis: A Survey," *Pattern Recognition*, vol. 36, no. 1, pp. 259-275, 2003.

[10] Gundimada S. and Asari V., "Facial Recognition using Multisensor Images Based on Localized Kernel Eigen Spaces," *IEEE Transaction on Image Processing*, vol. 18, no. 6, pp. 1314-1325, 2009.

[11] Guo G. and Dyer C., "Simultaneous Feature Selection and Classifier Training via Linear Programming: A Case Study for Face Expression Recognition," *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, USA, vol. 1, pp. 346-352, 2003.

[12] He D. and Cercone N., "Local Triplet Pattern for Content-based Image Retrieval," *in Proceedings of the 6th International Conference Image Analysis and Recognition Lecture Notes in Computer Science*, vol. 5627, pp. 229-238, 2009.

[13] Hsu C. and Lin C., "A Comparison on Methods for Multiclass Support Vector Machines," *IEEE Transaction on Neural Networks*, vol. 13, no. 2, pp. 415-425, 2002.

[14] Jabid T., Kabir M., and Chae O., "Robust Facial Expression Recognition Based on Local Directional Pattern," *Electronics and Telecommunications Research Institute Journal*, vol. 32, no. 5, pp. 784-794, 2010.

[15] Kabir H., Jabid T., and Chae O., "Local Directional Pattern Variance: A Robust Feature Descriptor for Facial Expression Recognition," *the International Arab Journal of Information Technology*, vol. 9, no. 4, pp. 382-391, 2012.

[16] Kanade T., Cohn J., and Tian Y., "Comprehensive Database for Facial Expression Analysis," *in Proceedings of the IEEE International Conference on Automated Face and Gesture Recognition*, Grenoble, pp. 46-53, 2000.

[17] Lyons M., Budynek J., and Akamatsu S., "Automatic Classification of Single Facial Images," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 21, no. 12, pp. 1357-1362, 1999.

[18] Ojala T., Pietikainen M., and Maenpaa T., "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971-987, 2002.

[19] Padgett C. and Cottrell G., "Representation Face Images for Emotion Classification," *in Proceedings of Advances in Neural Information Processing Systems*, USA, vol. 9, pp. 1-8, 1996.

[20] Shan C., Gong S., and McOwan P., "Facial Expression Recognition Based on Local Binary Patterns: A Comprehensive Study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803-816, 2009.

[21] Tan X. and Triggs B., "Enhanced Local Texture Feature Sets for Face Recognition under Difficult Lighting Conditions," *in Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, Brazil, vol. 4778, pp. 168-182, 2007.

[22] Tian Y., "Evaluation of Face Resolution for Expression Analysis," *in Proceedings of Computer Vision and Pattern Recognition Workshop*, Washington, pp. 1-7, 2004.

[23] Tian Y., Kanade T., and Cohn J., *Facial Expression Analysis*, Springer, New York, 2005.

[24] Uddin Z., Lee J., and Kim T., "An Enhanced Independent Component-Based Human Facial Expression Recognition from Video," *IEEE Transaction on Consumer Electronics*, vol. 55, no. 4, pp. 2216-2224, 2009.

[25] Valstar M. and Pantic M., "Fully Automatic Facial Action Unit Detection and Temporal Analysis," *in Proceedings of IEEE Computer Vision and Pattern Recognition Workshop*, New York, pp. 149, 2006.

[26] Valstar M., Patras I., and Pantic M., "Facial Action Unit Detection using Probabilistic Actively Learned Support Vector Machines on Tracked Facial Point Data," *in Proceedings of IEEE Computer Vision and Pattern Recognition-Workshops*, San Diego, vol. 3, pp. 76-84, 2005.

[27] Zhang Z., "Feature-Based Facial Expression Recognition: Sensitivity Analysis and Experiment with a Multi-Layer Perceptron," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 13, no. 6, pp. 893-911, 1999.

[28] Zhao W., Chellappa R., and Phillips P., "Face Recognition: A Literature Survey," *ACM Computing Survey*, vol. 35, no. 4, pp. 399-458, 2003.

[29] Zhou H., Wang R., and Wang C., "A Novel Extended Local Binary Pattern Operator for Texture Analysis," *Information Sciences*, vol. 178, no. 22, pp. 4314-4325, 2008.

Faisal Ahmed received his BSc degree in computer science and information technology from the Islamic University of Technology, Bangladesh in 2010. He is now working as a lecturer in the Department of CSE, Islamic University of Technology, Bangladesh. His research interests include texture analysis, computer vision, and pattern recognition.

Hossain Bari received his BSc degree in computer science and information technology from the Islamic University of Technology, Bangladesh in 2010. He is now working in the Samsung Bangladesh R & D Center Ltd, Bangladesh. His research interests include image processing and computer vision.

Emam Hossain received his BSc degree in computer science and information technology from the Islamic University of Technology, Bangladesh in 2010. He is now working as a lecturer in the Department of CSE, Ahsanullah University of Science and Technology, Bangladesh. His research interests include image processing, computer vision, artificial intelligence, and pattern recognition.