Compound Local Binary Pattern (CLBP) for Robust Facial Expression Recognition

Faisal Ahmed[1], Emam Hossain[2], A.S.M. Hossain Bari[3] and ASM Shihavuddin[4]
[1] Department of CIT, Islamic University of Technology, Gazipur, Bangladesh
[2] Department of CSE, Prime University, Dhaka, Bangladesh
[3] Samsung Bangladesh R & D Center Ltd, Dhaka, Bangladesh
[4] Universitat De Girona (UDG), Girona, Spain

***Abstract*—The local binary pattern (LBP) operator has been proved to be a simple and effective approach for facial feature representation. However, the LBP operator thresholds *P* neighbors at the value of the center pixel in a local neighborhood and encodes only the signs of the differences between the gray values. Thus, the LBP operator discards some important texture information. This paper presents a new local texture operator, the compound local binary pattern (CLBP), and a feature representation method based on CLBP codes for facial expression recognition. The CLBP operator combines extra *P* bits with the original LBP code, which are used to express the magnitude information of the differences between the center and the neighbor gray values. We empirically evaluate the effectiveness of the proposed feature representation for person-independent expression analysis. Extensive experiments show the superiority of the CLBP method against some other appearance-based feature representation methods.**

I. Introduction

Facial expression provides a non-verbal form of communication that facilitates the cognition of human emotions and intensions. Automated facial expression analysis has attracted much attention in the recent years due to its potential applicability in many areas such as human-computer interaction, data-driven animation, and customized applications for consumer products [1]. Although much work has been done, deriving an efficient and discriminative facial feature representation that can counteract variations in illumination, pose, and other changes is still a challenging task [2]. Commonly used facial feature extraction techniques include geometric feature-based methods and appearance-based methods [3].

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 [4] 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 [5] 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 [6], where they employed linear programming in order to perform simultaneous feature selection and classifier training. Recently, Valstar et al. [7], [8] 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 [1]. On the other hand, 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) [9] and independent component analysis (ICA) [10], [11] 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 [12], [13] and local feature analyses [14] are also explored in the literature.

Recently, facial expression analyses based on local binary pattern (LBP) [2], [15] and its variants have gained much popularity for their superior performances. The LBP operator was originally introduced for texture analyses [16] 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 [17]. To address this issue, local ternary pattern (LTP) [18] 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, Jabeed et al. [1] introduced local directional pattern (LDP). Instead of gray values, LDP employs the magnitude of the edge response values in different directions in order to encode the 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.

In this paper, we have introduced the compound local binary pattern (CLBP), an extension of the LBP operator and a feature representation method based on it for facial expression recognition. Unlike the original LBP operator that uses *P* bits to encode only the signs of the differences

978-1-4577-0045-3/11/$26.00 ©2011 IEEE

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 in a local neighborhood using a threshold. The motivation behind the proposed encoding scheme is to increase the robustness of the feature representation 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) facial expression database [19] 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.

II. LOCAL BINARY PATTERN (LBP)

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:

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

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

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 Fig. 1.

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 (ULBP), exploits certain LBP patterns, which appear more frequently in a significant area of the image. These patterns are known as the uniform patterns as they 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. [16] 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 [16].

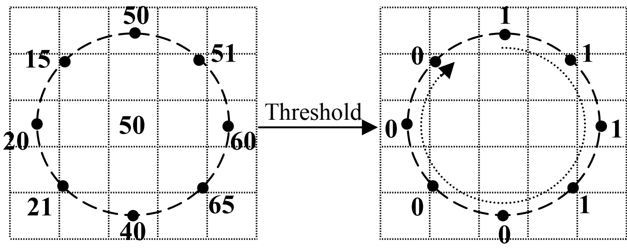

Figure 1. Illustration of the basic LBP operator. Here, the LBP code is 11110000 (decimal 240).

III. COMPOUND LOCAL BINARY PATTERN (CLBP)

The original LBP operator discards the magnitude information of the difference between the center and the neighbor gray values in a local neighborhood. As a result, this method tends to produce inconsistent codes. One example is shown in Fig. 2. Here, the 8-bit uniform LBP code (11111111) corresponds to a flat area or a dark spot at the center pixel [16], which is not correct in this case.

As LBP operator considers only the sign of the difference between two gray values, it often fails to generate appropriate binary code. 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 that employs one bit for each neighbor to express 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 represent 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 pattern and the other bit is used to encode the magnitude of the difference with respect to a threshold value, 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 (2) is replaced by the following function:

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

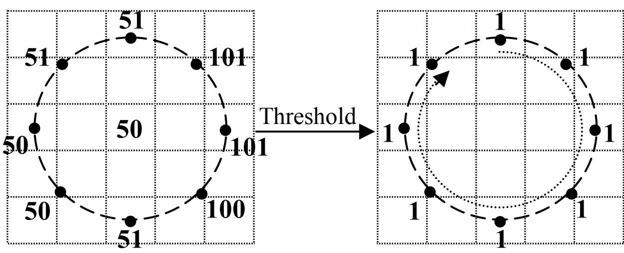

Figure 2. Generation of inconsistent binary pattern in the LBP encoding process.

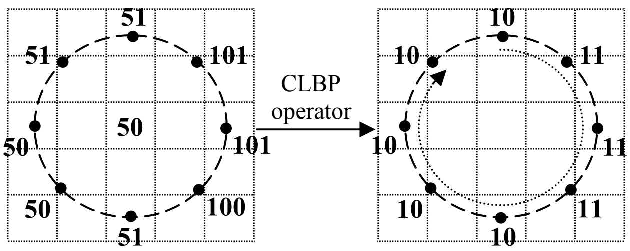

Figure 3. Illustration of the basic CLBP operator. Here, the binary pattern is 1011111110101010.

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 CLBP operator is illustrated in Fig. 3. It can be observed that, the proposed 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 consistent local pattern.

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 [20] have proposed to consider less number of neighbors while forming the binary patterns. Thus, this method discards some neighborhood information in order to reduce the length of the feature vector. Here, we have proposed 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 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 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 Fig. 4. The two sub-CLBP patterns are treated as separate binary codes and combined during the feature vector generation.

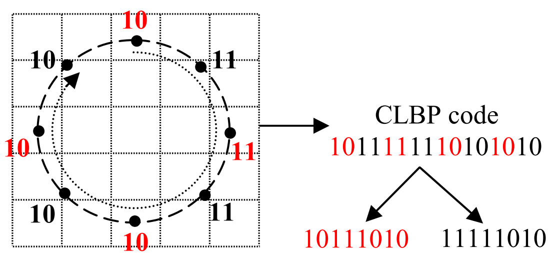

Figure 4. Generation of the two sub-CLBP patterns 10111010 and 11111010 from the original CLBP code 1011111110101010.

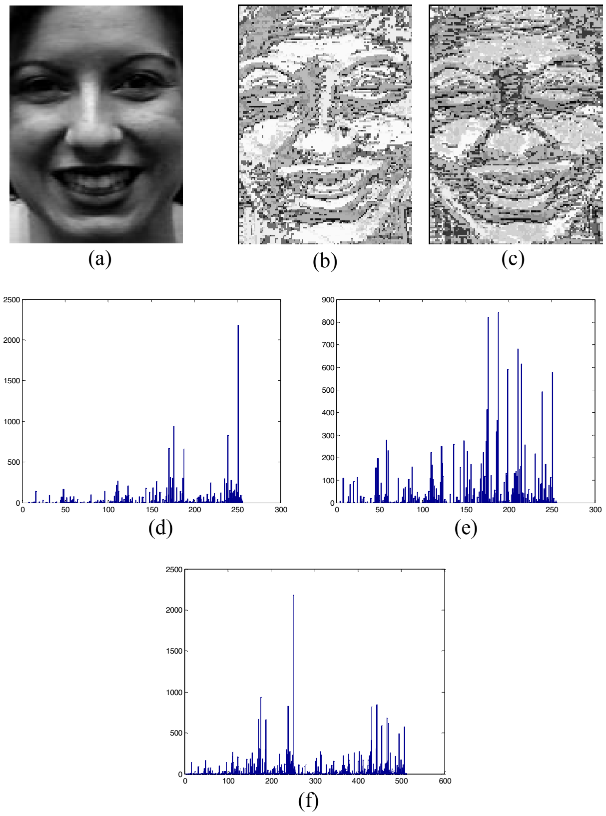

Figure 5. Illustration of the feature vector generation process, (a) is a sample expression image, (b) and (c) are the two sub-CLBP image representations, (d) and (e) are the histograms of (b) and (c), respectively, (f) is the feature vector generated by concatenating the histograms (d) and (e).

IV. FEATURE REPRESENTATION USING CLBP

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. Fig. 5 illustrates the CLBP histogram generation process from a sample expression image.

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 [1]. 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 histograms are generated from each of those regions. Finally, the histograms of all the regions are concatenated to obtain the extended 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 Fig.6.

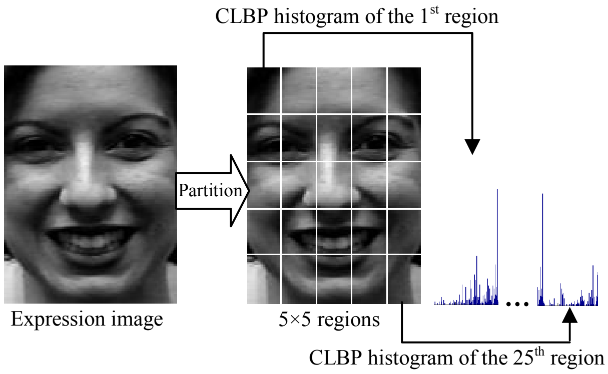

Figure 6. Each expression image is partitioned into a number of regions and individual CLBP histograms generated from each of the regions are concatenated to form the feature vector.

V. Classification Using Support Vector Machine (SVM)

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 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 $T = \{(\boldsymbol{x}_i, l_i), i = 1, 2, \ldots, L\}$, where $\boldsymbol{x}_i \in \mathbf{R}^P$ and $l_i \in \{-1, 1\}$, a new test data $\boldsymbol{x}$ is classified by

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

where 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 function K can be defined as

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

$$\| \boldsymbol{x}_i - \boldsymbol{x} \|^2 = (\boldsymbol{x}_i - \boldsymbol{x})^t (\boldsymbol{x}_i - \boldsymbol{x}) \tag{6}$$

Here, γ is a kernel parameter. A grid-search was carried out for selecting appropriate parameter value, as suggested in [21].

VI. Experimental Results

The performance of our proposed method is evaluated with a well-known image dataset, the Cohn-Kanade facial expression database [19]. The dataset comprises 1224 face images of a set of six prototypic emotional expressions, namely joy, surprise, sadness, anger, disgust, and fear. This six-expression dataset was extended to a 7-expression dataset by including additional 408 images of neutral expression face. Segmentation is an important part in facial expression analyses, as mixture of irrelevant background information may yield unreliable result. Thereofore, the selected images were cropped from the original ones using the ground truth of the two eye positions and then normalized to 150×110 pixels.

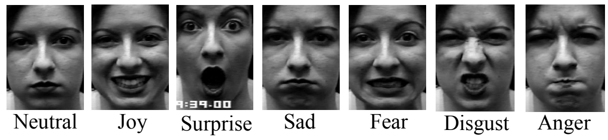

Figure 7. Sample images of the seven prototypic expressions.

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. 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 partitioned. We have considered three cases in our experiments, where images were divided into 3×3, 5×5, and 7×6 regions. The performance of the CLBP operator is compared with two widely-used local texture operators, namely the local binary pattern (LBP) operator [16] and the local ternary pattern (LTP) operator [18]. TABLE I and TABLE II show the 6-class and the 7-class expression dataset recognition rate, respectively.

TABLE I.
Recognition Rate of Different Methods Using the 6-Expression Dataset

Operator	Classification rate (%) for different number of regions		
	3×3	5×5	7×6
$LBP_{8,1}$	79.3	89.7	90.1
LTP	87.3	92.3	93.6
CLBP	88.2	94.4	94.2

TABLE II.
Recognition Rate of Different Methods Using the 7-Expression Dataset

Operator	Classification rate (%) for different number of regions		
	3×3	5×5	7×6
$LBP_{8,1}$	73.8	80.9	83.3
LTP	81.3	88.5	88.9
CLBP	82.1	90.4	89.2

It can be observed that, CLBP provides better recognition rate than LBP and LTP in all cases for both the 6-expression and the 7-expression datasets. For the 6-expression dataset, CLBP achieves a recognition rate of 94.4%. For the 7-expression dataset, the recognition rate is 90.4%. In addition, both LBP and LTP provide their best classification rate when the image is partitioned into 7×6 regions. On the other hand, the CLBP operator provides its best accuracy for images partitioned into 5×5 regions. Although both LTP and CLBP provide high recognition rates, CLBP requires a smaller feature vector than LTP in order to achieve its highest recognition rate. Therefore, computational complexity is reduced.

From the experimental results, it can be said that, facial feature representation based on the compound local binary pattern (CLBP) is more robust and provides higher classification rate than some existing feature representation methods. 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.

VII. Conclusion

This paper describes the compound local binary pattern, an extension of the original LBP texture operator and a feature representation method based on CLBP codes 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] T. Jabeed, M. H. Kabir, and O. Chae, "Robust Facial Expression Recognition Based on Local Directional Pattern," ETRI Journal, vol. 32, no. 5, pp. 784−794, October 2010.

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

[3] Y. Tian, T. Kanade, J. Cohn, "Facial Expression Analysis," in Handbook of Face Recognition, Springer, 2005.

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

[5] Z. Zhang, "Feature-Based Facial Expression Recognition: Sensitivity Analysis and Experiment with a Multi-Layer Perceptron," International Journal of Pattern Recognition and Artificial Intelligence, vol. 13, issue 6, pp. 893–911, 1999.

[6] G. D. Guo and C. R. Dyer, "Simultaneous Feature Selection and Classifier Training Via Linear Programming: A Case Study for Face Expression Recognition," IEEE Conference on Computer Vision and Pattern Recognition, pp. 346–352, 2003.

[7] M. Valstar, I. Patras, and M. Pantic, "Facial Action Unit Detection using Probabilistic Actively Learned Support Vector Machines on Tracked Facial Point Data," IEEE CVPR Workshop, vol. 3, pp. 76−84, 2005.

[8] M. Valstar and M. Pantic, "Fully Automatic Facial Action Unit Detection and Temporal Analysis," IEEE CVPR Workshop, p. 149, June 2006.

[9] C. Padgett and G. Cottrell, "Representation Face Images for Emotion Classification," Advances in Neural Information Processing Systems, vol. 9, 1997.

[10] M.S. Bartlett, J.R. Movellan, and T.J. Sejnowski, "Face Recognition by Independent Component Analysis," IEEE Transactoin on Neural Networks, vol. 13, no. 6, pp. 1450−1464, 2002.

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

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

[13] Y. Tian, "Evaluation of Face Resolution for Expression Analysis," IEEE Workshop on Face Processing in Video, 2004.

[14] G. Donato et al., "Classifying Facial Actions," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 21, no. 10, pp. 974−989, 1999.

[15] C. Shan, S. Gong, and P.W. McOwan, "Robust Facial Expression Recognition Using Local Binary Patterns," IEEE International Conference on Image Processing, pp. 370–373, 2005.

[16] T. Ojala, M. Pietikäinen, and T. Mäenpää, "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.

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

[18] X.Tan and B.Triggs, "Enhanced Local Texture Feature Sets for Face Recognition under Difficult Lighting Conditions," IEEE International Workshop on Analysis and Modeling of Faces and Gestures, LNCS 4778, pp.168−182, 2007.

[19] T. Kanade, J. Cohn, and Y. Tian, "Comprehensive Database for Facial Expression Analysis," IEEE International Conference on Automated Face and Gesture Recognition, pp. 46−53, March 2000.

[20] D. He and N. Cercone, "Local Triplet Pattern for Content-based Image Retrieval," Image Analysis and Recognition, pp. 229–238, 2009.

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