

Boosting Facial Expression Recognition Using LDGP - Local Distinctive Gradient Pattern

Mohammad Shahidul Islam
Assistant Professor, CSE dept.
Stamford University Bangladesh
Dhaka, Bangladesh
suva93@gmail.com

Surapong Auwatanamongkol
Associate Professor, CS dept.
National Institute of Development
Administration, Thailand
surapong@as.nida.ac.th

Md. Zahid Hasan
Lecturer, CSE dept.
Green University of Bangladesh
Dhaka, Bangladesh
hasan.ice@gmail.com

Abstract— Appearance based local feature methods are widely used for facial expression recognition because of their simplicity and high accuracy rates of recognition. However, the achieved accuracy rates and running time yet need to be improved. A new appearance based local feature method, called Local Distinctive Gradient Pattern (LDGP) is proposed in this paper. It derives two 4-bit local binary patterns from two different layers for a pixel by comparing the gray color intensity value of the pixel with its neighboring pixels in four distinct directions. Since each face image is divided into equal sized blocks, two histograms for the two 4-bit LDGP patterns of all pixels in each block can be constructed. The histograms of all blocks are then concatenated to build the feature vector for the given image. To evaluate the effectiveness of the proposed descriptor, experiments were conducted on the popular JAFFE dataset using Support Vector Machine (SVM) as the classifier. Extensive experimental results with seven prototype expressions show that proposed LDGP descriptor is superior to other appearance-based feature descriptors in terms of accuracy rates of recognition.

Keywords— LDGP, Pattern Recognition, Facial Feature Extraction, Feature Descriptor, JAFFE.

I. INTRODUCTION

A crucial part of human-human interaction is facial expression. To express emotion, it is considered to be the most important gesture [1]. Automatic recognition of facial expression is used in many areas for human-computer interaction e.g. emotion analysis, indexing and retrieval of video or image databases, creating animation etc., as human faces catch significant information about emotion and mind condition of every individual person [2], [3]. In human communication, only 7% is contributed by the verbal part, 38% by facial movement but 55% by facial expression [4]. This means that facial expression analysis is an important criterion in human-human as well as human-computer interaction.

In the past years, facial expression recognition concentrated on static images [5], [6]. Two approaches were investigated, feature-based and template-based. In recent years, image sequences and video data have been used to develop automated facial expression recognition system. Dynamic appearance and dynamic geometry contain various types of dynamic information, which are important for the

recognition of human expressions in the fields of computer vision [7] - [10] and psychology [11], [12].

Neutral, contempt, fear, sadness, disgust, anger, surprise and happiness are considered as the seven prototypes of facial expressions [13]. Most of the facial expression recognition systems built in the past are based on *Facial Action Coding System* (FACS), which involves very complex facial feature detection and extraction procedures [13], [14]. In FACS, the movement of muscle due to facial expression is coded with 44 different Action Units (AUs) [14]. More than 7000 different combinations can be formed using two or more AUs.

Combination of AUs with multi-stage face components and neural network as the classifier was used for facial expression recognition in [14]. The geometry-based feature extraction method proposed in the paper was time consuming and complex. In [15], Hidden Markov Models (HMMs) was used to model human facial expressions from video sequences and Naïve-Bayes classifier was used to recognize the expressions. In [16], Infrared (IR) illumination camera was used for facial feature detection, tracking and Dynamic Bayesian networks (DBNs) as classifier. In this system, 26 facial features around the regions of eyes, nose and mouth categorized the facial expressions. A real time system was constructed in [17] for facial expression recognition. A multichannel Gradient Model (MCGM) was initiated to capture facial movement signatures to identify facial expressions and the Support Vector Machine methodology was used as a classifier. Both space and time was considered in model [18].

A large range of human facial behaviors was presented in [19]. Facial features were tracked and rule-base reasoning was applied to recognize 20 AUs. In [20], some candidate grid nodes were manually placed on face for tracking facial landmarks to create facial wire frame model for facial expression recognition. All the above papers used geometry-based feature extraction methods.

The alternative methods are the appearance-based, where local feature based on color or texture or both are used for face representation. High accuracy rates are achieved by using these methods for Facial Expression Recognition (FER) than those of geometry-based methods e.g. AUs. Ahonen et al. [21] proposed a facial representation method for still images based on Local Binary Pattern (LBP). In this strategy, the LBP value

is calculated using the grey scale color intensities of the pixel and its neighboring pixels as follows:

$$LBP = \sum_{i=1}^p 2^{i-1} f(g(i) - c) \quad (1)$$

$$\text{Where, } f(x) = \begin{cases} 0 & \text{if } x \geq 0 \\ 1 & \text{if } x < 0 \end{cases} \quad (2)$$

‘c’ denotes the gray color intensity of the center pixel, g(i) is the gray color intensity of its neighbors and ‘p’ stands for the number of surrounding pixels. An extension to the original LBP operator named LBP_{RIU2} was proposed in [22], where length of the feature vector was reduced and implemented a rotated-invariant system. LBP is proven to be a powerful texture descriptor and adopted by many researchers for pattern recognition e.g. facial expression recognition [23], [24]. Though high accuracy rate is achieved using LBP descriptor for facial expression recognition but the feature extraction process is time consuming.

In 2008, Ojansivu et al. proposed a new texture descriptor LPQ (Local Phase Quantization) [25]. The LPQ method was blur insensitive and represented by a multiplication factor based on Phase invariant property and a Point Spread Function (PSF) of the original image in the frequency domain. Like LBP, LPQ also takes huge time for feature extraction. In [26], both local binary pattern and local phase quantization method were used to build a facial expression recognition system. The result was much better in this case than that of using LBP or LPQ separately but the system was too slow for high-resolution images.

Moment invariants are widely used in pattern recognition because of their discrimination power and robustness. They are invariant under shifting, scaling and rotation [27]. In [28], moment invariants were used for facial expression recognition. Moment invariants can reflect the deformation of facial features but cannot provide sufficient information of the displacement of facial features. Therefore, the authors added the feature displacement information and modified the general moment invariants feature vector to achieve high classification accuracy. The experiments were conducted on four prototypes of expressions only i.e. anger, disgust, happy and surprise.

Proposed feature representation method can capture more and distinct texture information from local 5x3 pixels area. The unique relations of the referenced pixel (‘c’) with four pixels at level one (‘a1’, ‘a2’, ‘a3’ and ‘a4’) and four pixels at level two (‘b1’, ‘b2’, ‘b3’, and ‘b4’) are not repeated while calculating LDGP pattern for any other pixels in that face (Fig. 1), whereas almost all the relations are repeated while calculating LBP pattern for the neighbors. In case of LPQ, each pixel is represented by eight-bit binary pattern derived using phase information, which results the feature vector length to 256 for a block, whereas in LDGP, a pixel can have only 32 different values. This makes LDGP quicker and more informative than both LBP and LPQ patterns (Table VI).

The rest of the paper is organized to explain the proposed local feature representation method in section II, data collection, system framework, experimental setup, result analysis in section III and conclusion in section IV.

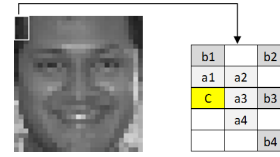


Fig 1 Pixels used to calculate LDGP code for the pixel ‘C’. To derive LDGP code for pixel ‘C’, the surrounding pixels a1, a2, a3, a4, b1, b2, b3, and b4 must exist.

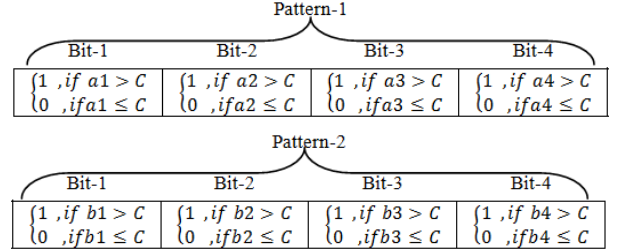


Fig. 2 Pattern P1 and P2 formation for the pixel ‘c’ shown in Fig. 1. Each of them is a 4 bit binary pattern.

II. PROPOSED FEATURE REPRESENTATION METHODOLOGY

At first, a 5x3 pixels local area as shown in Fig. 1 is used to calculate LDGP for a pixel. Where, ‘C’ is the referenced pixel. The gray color intensity values of the pixels C, a1, a2, a3, a4, b1, b2, b3 and b4 are used to formulate the LDGP binary pattern as shown in Fig. 2.

The LDGP pattern consists of two different level 4-bit binary patterns, say pattern-1 (P1) and pattern-2 (P2). P1 is computed using the pixel values of C, a1, a2, a3 and a4. P2 is calculated using the pixels value of C, b1, b2, b3 and b4 as shown in Fig. 2.

Each 4-bit binary pattern can have at most $2^4 = 16$ combinations. Therefore, P1 needs 16 bins to represent a single pixel and in the same way, P2 needs 16 bins too. Thus in LDGP, each pixel is represented by two values that range from 0 to 15. The two values represent the gradient directions of the gray color intensities of the neighboring pixels with respect to the referenced pixel in two different levels. The LDGP patterns can be derived for all the pixels where there exist eight neighboring pixels, a1, a2, a3, a4, b1, b2, b3, and b4 as like Fig 1 with the reference pixel ‘C’. A detailed example of obtaining LDGP patterns for a pixel is shown in Fig. 3 step by step.

Therefore to code a gray color facial image using proposed method, 16+16=32 bins are required. For example, for the first level pattern 0 to 15 (0000 to 1111) and for the second level pattern same as before 0 to 15 (0000 to 1111). Since a facial image is divided into equal sized blocks e.g. 5x5 or 7x7 or 9x9 or 9x9, 32 (16+16) bins for counting the numbers of occurrences of the two 4-bit patterns for a given block are needed. Finally, all histograms derived from all blocks are concatenated to form the feature vector for the given image.

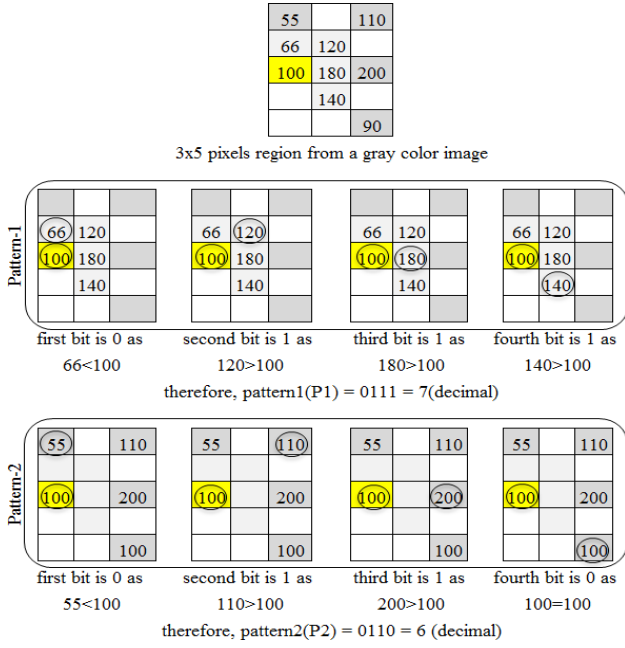


Fig. 3 Detailed example of calculating LDGP code for a pixel (here '100' in yellow color). Finally, the pixel is represented by two separate patterns '0111' and '0110'

A. Feature Vector

The gray scale facial image is divided into 81(9*9) equal sized square blocks and histogram of LDGP codes from each block is concatenated to form the feature vector as shown in Fig. 4. LDGP needs 32 bins only for a block of image. Therefore, histogram length for each block is 32 and the feature vector length for the whole image is 81*32= 2592.

B. Classification Using Support Vector Machine

Support vector machine, a well known linear classifier is successfully used in many research work for classification. It maps the feature data in to high dimensional feature space and draws a clear separation line between them. Let the set of training examples D be $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)\}$, where $\mathbf{x}_i = (x_1, x_2, \dots, x_r)$ is an input vector in a real-valued space $X \subseteq R^r$ and y_i is its class label (output value), $y_i \in \{1, -1\}$. 1: positive class and -1: negative class. SVM finds a linear function of the form (\mathbf{w} : weight vector)

$$f(\mathbf{x}) = \langle \mathbf{w} \cdot \mathbf{x} \rangle + b \quad (3)$$

So that an input vector \mathbf{x}_i is assigned to a positive class if $f(\mathbf{x}_i) \geq 0$, and to the negative class if $f(\mathbf{x}_i) < 0$.

$$y_i = \begin{cases} 1 & \text{if } \langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b \geq 0 \\ -1 & \text{if } \langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b < 0 \end{cases} \quad (4)$$

For linear data, it is easy to separate them but for non-linear data, SVM uses some sort of kernel function to create non-linear separation line. Some popular kernel functions are

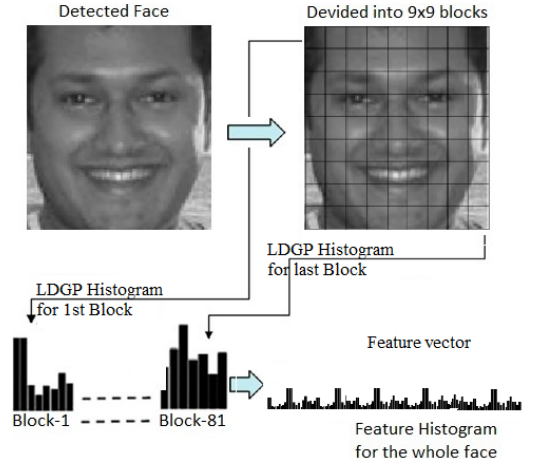


Fig. 4 Building feature vector for a single facial image. The feature histogram for the whole face is the feature vector or LDGP representator for this face.

polynomial, RBF (radial basis function) etc. Support vector machine is a binary classifier.

III. EXPERIMENTS AND RESULT ANALYSIS

The JAFFE [29] dataset was used for experiments to evaluate the effectiveness of the proposed method. The dataset contains 213 images of 7 facial expressions, six are basic facial expression and one is neutral expression, posed by 10 Japanese female models. The images were taken at the Psychology Department in Kyushu University. Every expression was taken more than once from each subject.

'fdlibmex' library from Matlab was used for face detection. This library consists of single 'mex' file with a single function that takes in image as input and returns the frontal face. The face was then masked using an elliptical shape to remove the unnecessary hair and neck-sides and divided into 81 equal sized blocks. Thus, LDGP features were extracted from each block. Concatenating feature histograms of all the blocks produces a unique feature vector of length $2 \times 16 \times 81 = 2592$ for a given image. Fig. 5 shows the proposed system framework.

Finally, A ten-fold none overlapping cross validation was performed in this paper. LIBSVM [30], a multiclass support vector machine was used for classification. From the dataset, 90% from each expression was taken for the training. The rest 10% images were used for testing. There was no overlap between any two folds and the experimental environment was user-dependent. Ten rounds of training and testing were conducted and the average confusion matrix for proposed method was reported and compared against the others. The kernel parameters for the classifier were set to: $s=0$ for SVM

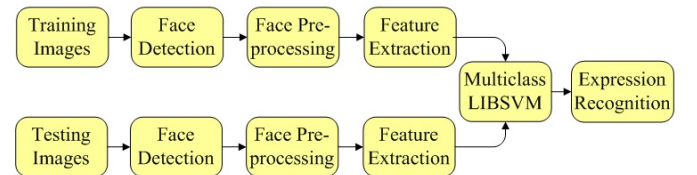


Fig. 5. Proposed System Framework

type C-Svc, $t=0/1/2$ for linear, polynomial and RBF kernel function respectively, $c=1$ is the cost of SVM, $g=1/$ (length of feature vector dimension), $b=1$ for probability estimation.

The accuracy achieved for different no of blocks is shown in Table I. Therefore keeping the block number fixed to 81, another set of experiments were conducted by differing the face dimension as shown in Table II. For JAFFE dataset, 99X99 pixels face dimension and 9X9=81 blocks gave the best performance in terms of accuracy. The experimental result for face expression recognition using the proposed feature representation method is shown using confusion matrix in Table III.

It can be seen from the confusion matrices that some particular expression classes, e.g., fear and sad, are consistently more difficult to classify than the others. Some instances of these expressions are difficult to distinguish even by a human and execution environment.

The achieved classification accuracy is shown in comparison with those of the other existing method in Table IV. It is clear that the LDGP method outperforms all the popular appearance-based methods. Moreover, the LDGP method yields shorter feature vector than other existing methods such as LBP and LPQ as shown in Table V.

The achieved classification accuracy, feature extraction time for a single facial image, learning time for a single fold and classification time of a single image are shown in comparison in Table VI.

TABLE I. CLASSIFICATION ACCURACY VS NO OF BLOCKS

Number of Blocks	Block Dimension (pixels)	Classification Accuracy (%)	Feature Vector Length	Face Dimension (Pixels)
3x3	33x33	88.28	9*32	99x99
9x9	11x11	94.6	81*32	
11x11	9x9	90.8	121*32	

TABLE II. RECOGNITION PERFORMANCE (%) IN DIFFERENT RESOLUTION OF IMAGES.





Resolution	81x81	99x99	117x117	135x135
Feature				
LDGP	93.8 ± 3.1	94.6 ± 3.0	94.8 ± 3.2	95 ± 3.1
LBP	89.6 ± 2.9	90.42 ± 3.1	90.52 ± 3.0	91.42 ± 3.1
LPQ	76.7 ± 1.7	79.56 ± 1.6	80.68 ± 1.7	81.12 ± 1.8

TABLE III. AVERAGE CONFUSION MATRIX USING LDGP ON JAFFE.

		Actual						
		Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
prediction	Angry	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	Disgust	0.0	93.1	6.9	0.0	0.0	0.0	0.0
	Fear	0.0	3.1	84.4	0.0	3.1	6.3	3.1
	Happy	0.0	0.0	0.0	96.8	0.0	3.2	0.0
	Neutral	0.0	0.0	0.0	0.0	100.0	0.0	0.0
	Sad	3.2	0.0	6.5	3.2	0.0	87.1	0.0
	Surprise	0.0	0.0	0.0	3.3	0.0	0.0	96.7

TABLE IV. RESULTS OF LDGP AND OTHER POPULAR METHODS IN SAME EXPERIMENTAL SETUP BUT DIFFERENT SVM KERNEL SETUP. (PERSON DEPENDENT)

person dependent	Classification Accuracy		
	Linear Kernel	Polynomial kernel	RBF kernel
LDGP	93.31%	94.60%	93.52%
LBP	89.72%	90.42%	89.31%
LBP _{U2}	91.43%	92.10%	91.06%
LPQ	77.56%	79.56%	78.14%

TABLE V. COMPARISON OF FEATURE VECTOR LENGTH AND CORRESPONDING CLASSIFICATION ACCURACY OF THE PROPOSED METHOD WITH SOME POPULAR METHODS

Method	Feature Vector Length (Dimension)	Classification Accuracy
LDGP	32 (2592)	94.60%
LBP	256 (20736)	90.42%
LBP _{U2}	59 (4779)	92.10%
LPQ	256 (20736)	79.56%

TABLE VI. CLASSIFICATION ACCURACY AND PROCESSING TIME COMPARISON FOR JAFFE DATASET. (PERSON DEPENDENT)

Method	Classification Accuracy (%)	Feature Extraction Time	Learning Time	Classification Time
LDGP	94.60%	0.010 sec	4.101 sec	0.005 sec
LBP	90.42%	0.026 sec	46.00 sec	0.070 sec
LBP _{U2}	92.10%	0.035 sec	25.00 sec	0.025 sec
LPQ	79.56%	0.095 sec	46.00 sec	0.070 sec
LBP _{U2} + LPQ	94.50%	0.130 sec	75.00 sec	0.100 sec

IV. CONCLUSION

A new appearance-based local feature descriptor is proposed in this paper. For each pixel in a gray scale image, the method extracts an LDGP code based on the differences between gray color intensity of the pixel and the gray color intensity of the surrounding pixels in four distinct directions as well as in two levels. The LDGP code for a pixel consists of two four bits binary patterns, p1 and p2, which yields the feature vector length much smaller than popular methods such as LBP and LPQ. The proposed method gives much better classification accuracy rates on static images, when compared to other previously proposed facial expression recognition methods. Future work may include extensive experiments of the proposed method on motion pictures or real-time videos and images having different viewpoints.

REFERENCES

- [1] J. A. Russell and J. M. Fernández-Dols, The Psychology of Facial Expression. Cambridge, U.K.: Cambridge Univ. Press, 1997.
- [2] I. S. Pandzic and R. Forchheimer, Eds., MPEG-4 Facial Animation. New York: Wiley, 2002.
- [3] P. Ekman and W. Friesen, Facial Action Coding System, Palo Alto, CA: Consulting Psychologists Press, 1978.
- [4] A. Mehrabian, Communication without words, psychology Today. 2(9) pp. 53-56, 1968.
- [5] M. Pantic and L. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 12, pp. 1424-1445, Dec. 2000.

- [6] J. Yu and B. Bhanu, "Evolutionary feature synthesis for facial expression recognition," *Pattern Recog. Lett.*, vol. 27, no. 11, pp. 1289–1298, Aug. 2006.
- [7] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 915–928, Jun. 2007.
- [8] T. Wu, M. Bartlett, and J. Movellan, "Facial expression recognition using Gabor motion energy filters," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recog. Workshop Human Commun. Behav. Anal.*, Jun. 2010, pp. 42–47.
- [9] Y. Tong, J. Chen, and Q. Ji, "A unified probabilistic framework for spontaneous facial action modeling and understanding," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 2, pp. 258–273, Feb. 2010.
- [10] P. Yang, Q. Liu, and D. Metaxas, "Boosting coded dynamic features for facial action units and facial expression recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2007, pp. 1–6.
- [11] Z. Ambadar, J. Schooler, and J. Cohn, "Deciphering the enigmatic face: The importance of facial dynamics to interpreting subtle facial expressions," *Psychol. Sci.*, vol. 16, no. 5, pp. 403–410, May 2005.
- [12] J. N. Bassili, "Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face," *Pers. Social Psychol.*, vol. 37, no. 11, pp. 2049–2058, Nov. 1979.
- [13] P. Ekman and W. Friesen, *Facial Action Coding System: A Technique for the Measurement of Facial Movement*, Consulting Psychologists Press, Palo Alto, CA., USA, 1978.
- [14] Y.L. Tian, T. Kanade and J.F. Cohn, "Recognizing action units for facial expressions analysis," *IEEE Trans. Pattern Analysis Machine Intell.*, Vol. 23, No. 2, pp. 97–115, March 2001.
- [15] I. Cohen, N. Sebe, A. Garg, L.S. Chen and T.S. Huang, "Facial expression recognition from video sequences: Temporal and static modeling," *Comput. Vision Image Understanding*, Vol. 91, pp. 160–187, August 2003.
- [16] Y. Zhang and Q. Ji, "Active and dynamic information fusion for facial expression understanding from image sequences," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 27, pp. 699–714, May 2005.
- [17] T. Anderson, A. Handid and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intel.*, Vol. 28, pp. 2037–2041, Dec. 2006.
- [18] Z. Yeasin, M. Pantic, G.I. Roisman and T.S. Huang, "A survey of affect recognition methods: Audio, visual and spontaneous expressions," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 31, pp. 39–58, Jan. 2009.
- [19] M. Pantic, M. I. Patras, "Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences," *IEEE Trans. Syst. Man Cybernet. Part B: Cybernet.*, Vol. 36: pp. 433–449, April 2006.
- [20] I. Kotsia and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines," *IEEE Trans. Image Process.*, Vol. 16, pp. 172–187, Jan. 2007.
- [21] T. Ahonen, A. Hadid and M. Pietikainen, "Face description with local binary patterns: Application to face recognition". *IEEE Trans. Pattern Anal. Mach. Intel.*, Vol. 28, pp. 2037–2041, Dec. 2006.
- [22] T. Ojala, M. Pietikainen and T. Maenpaa, "Multiresolution Gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Analysis Machine Intellig.*, Vol. 24, pp. 971–987, July 2002.
- [23] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 29: pp. 915–928, June 2007.
- [24] L. Ma and K. Khorasani, "Facial expression recognition using constructive feedforward neural networks," *IEEE Trans. Syst. Man Cybernet. Part B: Cybernet.*, Vol. 34, pp. 1588–1595, June 2004.
- [25] V. Ojansivu and J. Heikkila, "Blur insensitive texture classification using local phase quantization," *Proceedings of the 3rd International Conference on Image and Signal Processing*, July 1–3, 2008, Cherbourg-Octeville, France, pp. 236–243.
- [26] S. Yang and B. Bhanu, "Understanding discrete facial expressions in video using an emotion avatar image," *IEEE Trans. Syst. Man Cybern. B Cybern.*, Vol. 42, pp. 980–992, May 2012.
- [27] M. K. Hu, "Visual pattern recognition by moment invariants." *Information Theory, IRE Transactions on.* Vol. 8, no. 2, pp. 179–187, 1962
- [28] Y. Zhu, L. C. DE. Silva, C. C. Co, "Using Moment Invariant and HMM for Facial Expression Recognition," *Pattern Recognition Letters Elsevier*. Vol. 23, no. 1, pp. 83–91, 2002.
- [29] M.J. Lyons, M. Kamachi and J. Gyoba, (1997) The japanese female facial expression (JAFPE) database (Online). Available: http://www.kasrl.org/jaffe_info.html.
- [30] C.C. Chang and C. J. Lin, Libsvm: A Library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, Vol. 2, No.3, April 2011.