

Automatic Facial Expression Recognition Based on Hybrid Approach

Ali K. K. Bermani
College of Engineering,
Babylon University,
Babil, Iraq

Atef Z. Ghalwash
Computer Science Department,
Faculty of Computers and
Information, Helwan University,
Cairo, Egypt

Aliaa A. A. Youssif
Computer Science Department,
Faculty of Computers and
Information, Helwan University,
Cairo, Egypt

Abstract—The topic of automatic recognition of facial expressions deduce a lot of researchers in the late last century and has increased a great interest in the past few years. Several techniques have emerged in order to improve the efficiency of the recognition by addressing problems in face detection and extraction features in recognizing expressions. This paper has proposed automatic system for facial expression recognition which consists of hybrid approach in feature extraction phase which represent a combination between holistic and analytic approaches by extract 307 facial expression features (19 features by geometric, 288 feature by appearance). Expressions recognition is performed by using radial basis function (RBF) based on artificial neural network to recognize the six basic emotions (anger, fear, disgust, happiness, surprise, sadness) in addition to the natural. The system achieved recognition rate 97.08% when applying on person-dependent database and 93.98% when applying on person-independent.

Keywords- Facial expression recognition; geometric and appearance modeling; Expression recognition; Gabor wavelet; principal component analysis (PCA); neural network.

I. INTRODUCTION

Over the last years, face recognition and automatic analysis of facial expressions has one of the most challenging research areas in the field of computer vision and has received a special importance. The focus of the relatively recently initiated research area of facial expression has lied on sensing, detecting and interpreting human affective states and devising appropriate means for handling this affective information in order to enhance current human machine interface designs. The most expressive way humans display their emotional state is through facial expressions. A human can detect face and recognition on facial expressions without effort, but for a machine and in computer vision it is very difficult. Mehrabian [1] points out that 7% of human communication information is communicated by linguistic language (verbal part), 38% by paralanguage (vocal part) and 55% by facial expression. Therefore, facial expressions are the most important information for emotional perception in face to face communication. Automatic facial expression recognition is an interesting and challenging problem. Automatic recognition of facial expression has wide range of applications in areas such as human computer interaction (HCI), emotion analysis, Psychological area, virtual reality, video-conferencing, indexing and retrieval of image and video database, image understanding and synthetic face animation. In this paper and

most systems of recognition for facial expressions can be divided into three phases. (1) Face detection (2) Feature extraction (facial features). (3) Classification. Detect the face from the image is first important phase. Next and the most important phase is the mechanism for extracting the facial features from the facial image. The extracted facial features are either geometric features such as the shapes of the facial components (eyes, mouth, etc.) and the locations of facial characteristic points (corners of the eyes, mouth, etc.), or appearance features representing the texture of the facial skin in specific facial areas including wrinkles, bulges, and furrows. Appearance-based features include learned image filters from Principal Component Analysis (PCA), Gabor filters, features based on edge-orientation histograms, etc. or hybrid-based include both geometric and appearance features. The final phase is the classifier, which will classify the image into the set of defined expressions.

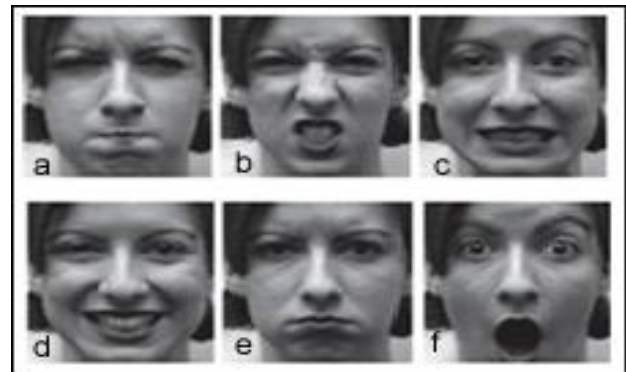


Figure 1. Basic emotion: (a) anger, (b) fear, (c) disgust, (d) joy, (e) sadness, (f) surprise.

There are six universally recognized expressions Angry, Disgust, Fear, Happy, Sad and Surprised (a facial expression into one of the basic facial expressions) as shown in Figure 1. The remainder of this paper is organized as follows: Section II is Related Work on facial expression recognition. Section III, describe Proposed System, In Section IV, experimental results are dataset used in the Proposed System and results. Finally, conclusions are given in Section V.

II. RELATED WORK

Study of facial expressions dates back to 1649, when John Bulwer has written a detailed note on the various expressions

and movement of head muscles in his book “Pathomyotomia“. In 19th century, one of the most important works on facial expression analysis that has a direct relationship to the modern day science of automatic facial expression recognition was the work done by Charles Darwin. In 1872, Darwin [2] wrote a treatise that established the general principles of expression and the means of expressions in both humans and animals. Since the mid of 1970s, different approaches are proposed for facial expression analysis from either static facial images or image sequences. In 1978, Ekman [3] defined a new scheme for describing facial movements. This was called Facial Action Coding Scheme (FACS). FACS combines 64 basic Action Units (AU) and a combination of AU's represents movement of facial muscles and tells information about face expressions. Before the mid of 1990s, facial motion analysis has been used by researchers to perform automatic facial data extraction. Generally existing approaches for facial expression recognition systems can be divided into three categories, based on how features are extracted from an image for classification. The categories are geometric-based, appearance-based, and hybrid-based.

A. Geometric-based approaches

The geometric feature measures the variations in shape, location, distance of facial components like mouth, eyes, eyebrows, nose, etc. In their earlier work Kobayashi and Hara [4], [5], proposed a geometric face model of 30 facial characteristic points FCPs. In their later work Kobayashi and Hara [6], use a CCD camera in monochrome mode to obtain a set of brightness distributions of 13 vertical lines crossing the FCPs. First, they normalize an input image by using an affine transformation so that the distance between irises becomes 20 pixels. From the distance between the irises, the length of the vertical lines is empirically determined. The range of the acquired brightness distributions is normalized to [0,1] and these data are given further to a trained neural network NN for expression emotional classification. Pantic et al.[7] proposed automated method for detecting facial actions by analyzing the contours of facial components, including the eyes and the mouth. A multi-detector technique is used to spatially sample the contours and detect all facial features. Arule-based classifier is then used to recognize the individual facial muscle action units(AUs). They achieve 86% recognition rate.

B. Appearance-based approaches

The appearance features present the appearance (skintexture) variations of the face, such as wrinkles, furrows, etc. It can be extracted on either the whole face image or specific regions in a facial image. Feng [8] uses Local Binary Patterns (LBP) to extract facial texture features and combines different local histograms to recover the shape of the face. Classification is performed through a coarse-to fine scheme, where seven templates were designed to represent the corresponding seven basic facial expressions. At first, two expression classes are selected based on the distance from the test image to the seven templates. The final classification is then done via a K-nearest neighbor classifier with the weighted Chi-square statistic. Deng et al.[9] proposed a facial expression recognition system based on the local Gabor filter bank. They selected some parts of the frequency and orientation parameters instead of using the entire global filter

bank to extract the features, and tested them on the JAFFE facial database which stores posed facial expressions. In order to select a few Gabor filters, they covered all the frequencies and orientations, but only selected one frequency for each orientation or each increase in the interval between neighboring frequencies with the same orientation. Lajevardi et al [10] they extract features from a picture of the whole face. The features are generated using the log-Gabor filters. The log-Gabor filters can be constructed with an arbitrary bandwidth that can be optimized to produce a filter with minimal spatial extent. Five scales and eight orientations are implemented to extract features from face images. This leads to 40 filter transfer functions $\{H_1, H_2, \dots, H_{40}\}$ representing different scales and orientations. The objective is to choose only the log-Gabor filters which minimize the absolute value of the spectral difference between the original image filter output and the noisy image filter output. Initially, the error is calculated for all of the training images, and for each image a set of 4 log-Gabor filters with the smallest value of the spectral difference is selected. The selected optimal subset of 4 log-Gabor filters allowed reducing the features dimension from 40 to only 4 arrays of size 60×60 for each image. The feature arrays are transferred to feature vectors of length 3600 for each image. Murthy and Jadon [11] propose a modified PCA (eigenspaces) for Eigen face reconstruction method for expression recognition. They have divided the training set of Cohn kanade and JAFFE databases into 6 different partitions and Eigen space is constructed for each class, then image is reconstructed. Mean square error is used as a similarity measure for comparing original and reconstructed image.

C. Hybrid-based approaches

Hybrid features uses both geometric and appearance based approaches to extract features. Zhang et al. [12] presented a Facial Expression Recognition system where they have compared the use of two types of features extracted from face images for recognizing facial expression. Geometric positions of set of fiducial points (34 facial points) and multiscale & multi orientation gabor wavelet coefficient extracted from the face image at the fiducial points are the two approaches used for feature extraction. These are given to neural network classifier separately or jointly and results are compared. Their system achieves a generalized recognition rate of 73.3% with geometric positions alone, 92.2% with Gabor wavelet coefficients alone, and 92.3% with the combined information.

Zhang and Ji [13] proposed a multi-feature technique that is based on the detection of facial points, nasolabial folds, and edges in the forehead area. In their method, facial features are extracted by associating each AU with a set of movements, and then classified using a Bayesian network model. Jun Ou et al.[14] presents the Automatic Facial Expression Recognition (AFER) using 28 facial feature points and Gabor wavelet Gabor wavelet filter provided with 5 frequencies, 8 orientations for facial feature localization, PCA for feature extraction and K nearest neighbor (KNN) for expression classification. Their system achieves a generalized recognition rate of 80%. Monwar & Rezaei [15] use location and shape features to represent the pain information. These features are used as inputs to the standard back-propagation in the form of a three-layer neural network with one hidden layer for

classification of painful and painless faces. They achieve 91.67% recognition rate using 10 hidden layer units.

III. PROPOSED SYSTEM

This paper proposes an automatic system for facial expression recognition, a system will address the problems of facial expression recognition about how to detect a human face in static images and how to represent and recognize facial expressions presented in those faces by using a hybrid approach for facial feature information extraction and then Classification of the facial expressions into the six basic emotions (anger, disgust, fear, happy, sad and surprise) in addition to the neutral one. The system operates after entering the image that contains the face which will begin The first phase; face detection process and segment the face image, Then extract feature vector (geometric feature and appearance feature) by apply the feature extraction process on face image, The final phase; uses this feature as an input into the radial basis function artificial neural network to recognize the facial expressions. Figure 2, shown the block diagram of the proposed system.

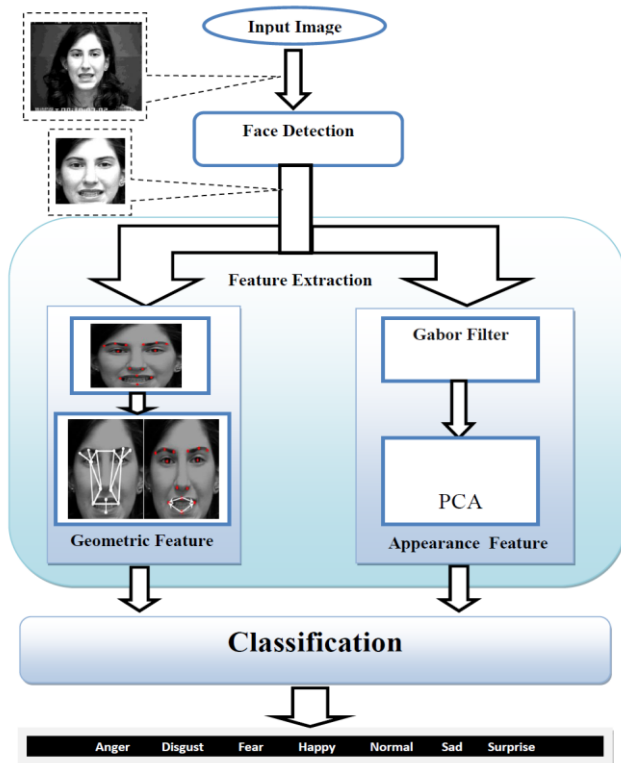


Figure 2. Block diagram of the facial expression recognition system.

A. Face Detection

In most systems of facial expression recognition the first phase is the face detection and segmentation from the image. The purpose of face detection is to determine the presence and the position of face in an image. Many algorithms have been proposed for face detection, e.g. the Viola-Jones face detector [16]. The open source code library (OpenCV) that employs algorithm based on Viola & Jones features Sreekar, Krishna [17],[16], used in this system. Sreekar [17] Works to determination face image, this system will segment the face

image the specific. 300x300 pixels is uniform size of face image that resulted from the Open CV face detector. Figure 3, shows example of the Open CV face detector.



Figure 3. An example of the OpenCV face detector.

B. Facial Features Extraction

The second phase and the most important in the systems of facial expression recognition is the facial feature extraction which is based on finding features that conveying the facial expression information. These problems can be viewed as a dimensionality reduction problem (transforming the input data into a reduced representation set of features which encode the relevant information from the input data). This system uses two approaches to extract the facial features: geometric and appearance features.

1) *Geometric Features Extraction*: In Geometric approach, 19 features are extracted from the face image. In this phase the same approach that Youssif et al. [18] will be used. They extracted 19 features from the face image by executing the segmentation process to divide the face image into three regions: mouth, nose and two eyes and two eyebrows and then located the facial characteristic points (FCPs) in each face component by using mouth, nose, eyes and eyebrows FCPs extraction techniques. Finally; certain lengths between FCPs are calculated using Euclidean distance D:

$$D = \sqrt{(x1 - x2)^2 + (y1 - y2)^2} \quad (1)$$

Where (x1, y1) and (x2,y2) are the coordinates of any two FCPs P1(x1,y1) and P2(x2,y2) respectively. Also two angles in the mouth area are computed to represent with geometric lengths the geometric features, typically results are in Figure 4 (a, b).

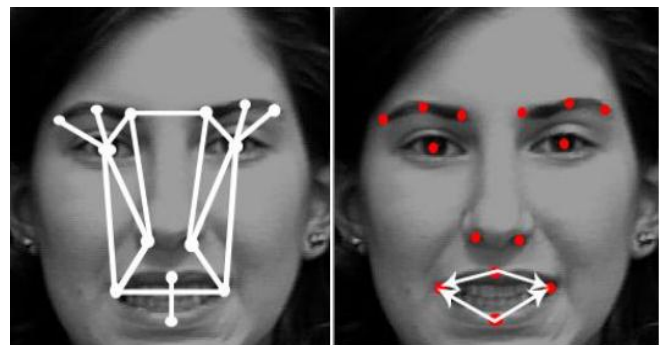


Figure 4. Geometric features: (a) geometric lengths, (b) mouth angles

2) *Appearance Features Extraction*: The second approach used to extract feature is appearance approach. Various feature extraction methods can be used to extract facial expression features such as, the active appearance models (AAM), eigenfaces, and Gabor wavelets that proposed in this paper. The Gabor wavelets reveal much about facial expressions as both transient and intransient facial features that often give rise to a change in contrast.

a) *pre-processing and Gabor filters*: After face detection process, 300x300 pixels is the size of face image, will be done by using cropping process on face image to be 200x300 pixels and then do resize to the cropped image to be 200x200. After pre-processing process of face image will use Gabor filter to extract features. Gabor wavelet represents the properties of spatial localization, orientation selectivity, and spatial frequency selectivity. The Gabor wavelet can be defined as

$$\Psi(\mathbf{Z}) = \frac{p_{u,r}^2}{\sigma^2} \exp\left(-\frac{p_{u,r}^2 \mathbf{Z}^2}{2\sigma^2}\right) \left[\exp(ip_{u,r}\mathbf{Z}) - \exp\left(-\frac{\sigma^2}{2}\right) \right] \quad (2)$$

Where $\mathbf{Z} = (x,y)$, and u and r define the orientations and scales of the Gabor wavelet, respectively [19]. $p_{u,r}$ is defined as

$$P_{u,r} = P_r e^{i\Phi_u} \quad (3)$$

Where $P_r = p_{\max}/f^r$ and $\Phi_u = \pi u/6$. P_{\max} is the maximum frequency, and f is the spacing factor between kernels in the frequency domain. In this experiment the Gabor wavelet with three scales ($r = 0,1,2$) and six orientations ($u = 0,1,2,3,4,5$), and parameters $\sigma = 2\pi$, $p_{\max} = \pi/2$, and $f = \sqrt{2}$ are applied. This reproduces 18 different Gabor kernels (3 scales * 6 orientations). The Gabor representation of an image, which is called a Gabor image, is produced from the convolution of the image with the Gabor kernels as defined by Equation 2. For each image pixel, it produces 18 Gabor images and each one consists of two Gabor parts: the real part and the imaginary part. Thereafter, these two parts are transformed into two kinds of Gabor feature: the magnitude and the phase. In this study only the magnitude features are used to represent the facial Gabor features. Changing the facial positions will cause only slight variations in the magnitude of the Gabor wavelet, while the phase of the Gabor wavelet output is very sensitive to the position of the face [20]. The 200x200 pixels in the Gabor image produce a large number of features is 40000 feature vectors. To produce a smaller number of features, the image size has been re-sized to 25x25 pixels, resulting in only 625 features.

b) *Dimension reduction with Principal component analysis (PCA)*: After re-sizing the Gabor image to 25x25 pixels the image is further compressed, resulting in 625x1 feature vectors for each image. PCA is a way of identifying patterns in the data and expressing the data in such a way as to highlight their similarities and differences [21]. Since the patterns in the data can be difficult to identify in data of high dimensions where the graphical representation is not available, PCA is a powerful tool for analyzing such data. The other

main advantage of PCA is its ability to reduce the number of dimensions without much loss of information. PCA will be used in this experiment to reduce the features of each image (its 18 Gabor images). To perform PCA, we reshape each Gabor image into one column producing a matrix have 625 rows and 18 columns, then the mean value of these Gabor images columns are subtracted from each Gabor image column. The subtracted means are the average across each dimension, and then the covariance matrix is calculated (which is 18*18 matrix). Then, based on the covariance matrix, eigenvectors and eigenvalues are calculated. Next, the 16 most significant eigenvalues are selected to obtain the eigenvectors (then will have a matrix of 16 columns and 18 rows). These eigenvectors are input into the radial basis function neural network classifier after reshaping it into 1 row and adding the geometry features to it (16*18=288 appearance features + 19 geometry features).

C. Facial Expression Classification

After extract feature from face image, the next and last phase in this system is the classification of the facial expressions. The facial feature vector of 307 feature elements (19 geometric features plus 288 appearance features) forms the input of radial basis function RBF neural network. As shown in Figure 5. The RBF neural network has one hidden layer uses neurons with RBF activation functions (ϕ_1, \dots, ϕ_m) for describing the receptors. The following equation is used to combine linearly the outputs of the hidden neurons:

$$y = \sum_{j=1}^m w_j \phi_j(\|x - t_j\|) \quad (4)$$

Where $\|x - t\|$ is the distance of the input vector $x = (x_1, \dots, x_{m1})$ from the vector $t = (t_1, \dots, t_{m1})$ which is called the center for m inputs (receptors). The proposed system use the Gaussian function described by the equation 5:

$$\phi(x) = e^{-\frac{x^2}{2\sigma^2}} \quad (5)$$

Where σ is the spread parameter of the Gaussian function. Number of neurons is 250 and RBF spread value is 250 used in this system with person-dependent datasets and Number of neurons is 80 and RBF spread value is 300 used in this system with person-independent datasets.

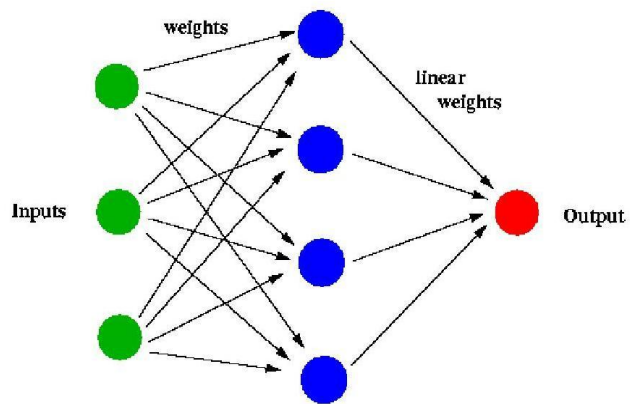


Figure 5. Radial basis function neural network architecture

IV. EXPERIMENTS AND RESULTS

The performance of proposed system is analysed on the Cohn-Kanade facial expression database Anitha et al. [22], which is commonly used in research on facial expression recognition. This database includes approximately 2000 image sequences with 640x480 pixels arrays and 8-bit precision for grayscale values from over 200 subjects. Subjects in the available portion of the database are 97 university students. They are ranged in age from 18 to 30 years. 65% are females, 15% are African-American, and 3% are Asian or Latino, Figure 6, Samples from original Cohn-Kanade Facial Expression Database. In order to compare the results of this system with that of Youssif et al [18], the database used in [18], has been adopted. They prepared two datasets from the available portion to examine the proposed system with different two manners. In the first manner the proposed system trained on each subject (person) different facial expression. So; all subjects with their facial expressions were exist in both training and testing datasets which called person-dependent dataset. In this dataset the last five images from each subject facial expression were taken and used the odds (60%) in training and the evens (40%) in testing. In the second manner the system trained in the facial expressions independent on the subjects. The person-independent dataset were prepared by selecting one image that convey facial expression information from each subject facial expressions images to form the seven facial expression classes. Then a 60% of the images were selected for training phase and the rest (40%) for testing. In both datasets the normal class was prepared by picking up the first image from each expression for subject.



Figure.6: Samples from original Cohn-Kanade Facial Expression Database.

The results in tables (1,2) show that the proposed system can classify the facial expressions correctly with recognition rates 97.08% using person-dependent dataset and recognition rate 93.98% using person-independent dataset. From these results we can conclude that the proposed system has improved recognition rate in both database used with the system used by Youssif et al [18], where the percentage was 96 using person-dependent dataset and 93 using person-independent dataset. The system is implemented in MATLAB 7.7.0 on 1.73GHz Microsoft Windows XP professional workstation.

%	Anger	Disgust	Fear	Happy	Normal	Sad	Surprise
Anger	100	0	0	0	0	0	0
Disgust	0	92.5	2.5	0	2.5	2.5	0
Fear	0	0	100	0	0	0	0
Happy	0	2.63	5.26	92.1	0	0	0
Normal	0	0	0	0	100	0	0
Sad	0	2.5	0	0	2.5	95	0
Surprise	0	0	0	0	0	0	100
Recognition rate	97.08						

%	Anger	Disgust	Fear	Happy	Normal	Sad	Surprise
Anger	93.33	0	0	0	6.66	0	0
Disgust	7.14	84.6	7.14	0	0	0	0
Fear	0	0	93.33	0	6.66	0	0
Happy	0	0	0	100	0	0	0
Normal	0	0	0	0	100	0	0
Sad	0	0	6.66	0	6.66	86.66	0
Surprise	0	0	0	0	0	0	100
Recognition rate	93.98						

V. CONCLUSIONS

This paper proposed a method for automatic facial expression recognition. A system is capable detect a human face in static images and extract features by a hybrid approach (geometric and appearance approach) and then classify expressions presented in those faces. A hybrid approach is used for facial features extraction as a combination between geometric and appearance facial features. Radial Basis Function RBF based neural network is used for facial expression recognition.

The proposed system can classify the facial expressions correctly with classification rates between 92% and 100% (recognition rate 97.08%) using person-dependent dataset and between 84.6% and 100% (recognition rate 93.98%) using person-independent dataset. The proposed system could classify anger, fear, normal, sad and surprise in maximum rates but the disgust and happy in minimum rate in person-dependent dataset. In person-independent dataset the anger, fear, happy, normal and surprise classes could be recognized at maximum rates but fear and sad classes registered the minimum recognition rate.

REFERENCES

- [1] A. Mehrabian, "Communication without Words," *Psychology Today*, Vo.1.2, No.4, pp 53-56, 1968.
- [2] C. Darwin, "The expression of the emotions in man and animal," J.Murray, London, 1872.
- [3] P. Ekman and W.V. Friesen, "Facial Action Coding System (FACS)," Manual. Palo Alto: Consulting Psychologists Press, 1978.
- [4] H. Kobayashi and F. Hara, "Recognition of Six Basic Facial Expressions and Their Strength by Neural Network," *Proc. Int'l Workshop Robot and Human Comm*, pp. 381-386, 1992.

- [5] H. Kobayashi and F. Hara, "Recognition of Mixed Facial Expressions by Neural Network," *Proc. Int'l Workshop Robot and Human Comm.*, pp. 387-391, 1992.
- [6] H. Kobayashi and F. Hara, "Facial Interaction between Animated 3D Face Robot and Human Beings," *Proc. Int'l Conf. Systems, an, Cybernetics*, pp. 3,732-3,737, 1997.
- [7] M. Pantic and L. J. M. Rothkrantz, "Facial action recognition for facial expression analysis from static face images," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 34(3):1449-1461, 2004.
- [8] X. Feng, "Facial expression recognition based on local binary patterns and coarse-to-fine classification," *In Proceedings of Fourth International Conference on Computer and Information Technology*, pages 178-183, 2004.
- [9] HB. Deng, LW. Jin, LX. Zhen and JC. Huang, "A new facial expression recognition method based on local Gabor filter bank and PCA plus LDA," *Int J Inf Technol 11(11)*:86-96. 2005.
- [10] S. Lajevardi and M. Lech, "Facial expression recognition from image sequences using optimized feature selection," *Image and Vision Computing New Zealand, IVCNZ 2008. 23rd International Conference*, pp. 1-6, 2008.
- [11] G. R. S. Murthy, R. S. Jadon, "Effectiveness of Eigenspaces for facial expression recognition," *International Journal of Computer Theory and Engineering*, Vol. 1, No. 5, 1793-8201, pp. 638-642. December 2009.
- [12] Z.Zhang, M. Lyons, M. Schuster and S. Akamatsu "Comparison between geometry-based and Gabor wavelets-based facial expression recognition using multi-layer perceptron," *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, pp. 454-459.1998.
- [13] Y. Zhang and Q. Ji. "Active and dynamic information fusion for facial expression understanding from image sequences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5):699-714, 2005.
- [14] Jun Ou, Xiao-Bo Bai, Yun Pei, Liang Ma, Wei Liu, "Automatic facial expression recognition using Gabor filter and expression analysis," *Second International Conference on Computer Modeling and Simulation*, IEEE, pp 215-218.2010.
- [15] M. Monwar, and S. Rezaei, "Pain recognition using artificial neural network," *Signal Processing and Information Technology, 2006 IEEE International Symposium on*. 2006.
- [16] P. Viola, and M. Jones. "Robust real-time object detection," *Second international workshop on statistical and computational theories of vision-modeling, learning, computing, and sampling*. 2001.
- [17] Sreekar Krishna. Open CV Viola-Jones Face Detection in Matlab. [Online] available :<http://www.mathworks.com/matlabcentral/fileexchange/19912> .June 6, 2010.
- [18] A. Youssif, and W. Asker, "Automatic Facial Expression Recognition System Based on Geometric and Appearance Feature," *Computer and Information Science*, Vol.4, No.2, March 2011.
- [19] TS. Lee "Image representation using 2D Gabor wavelets," *IEEE Trans Pattern Anal Mach Intell* 18:959-971, 1996.
- [20] I. Kotsia, I. Bucu and I. Pitas. "An analysis of facial expression recognition under partial facial image occlusion," *J Image Vision Computer*, 26:1052-1067, 2008.
- [21] LI. Smith, "A tutorial on principal component analysis," *Cornell University*, pp 2-22, 2002.
- [22] M. Anitha, K. Venkatesha and B. Adiga, "A survey on facial expression databases," *International Journal of Engineering Science and Technology*, Vol. 2(10), 5158-5174, 2010.