Feature Extraction for Facial Expression Recognition based on Hybrid Face Regions

Seyed M. LAJEVARDI, Zahir M. HUSSAIN
RMIT University, Australia
seyed.lajevardi@rmit.edu.au

***Abstract*—Facial expression recognition has numerous applications, including psychological research, improved human computer interaction, and sign language translation. A novel facial expression recognition system based on hybrid face regions (HFR) is investigated. The expression recognition system is fully automatic, and consists of the following modules: face detection, facial detection, feature extraction, optimal features selection, and classification. The features are extracted from both whole face image and face regions (eyes and mouth) using log Gabor filters. Then, the most discriminate features are selected based on mutual information criteria. The system can automatically recognize six expressions: anger, disgust, fear, happiness, sadness and surprise. The selected features are classified using the Naive Bayesian (NB) classifier. The proposed method has been extensively assessed using Cohn-Kanade database and JAFFE database. The experiments have highlighted the efficiency of the proposed HFR method in enhancing the classification rate.**

***Index Terms*—Facial expression recognition, Gabor filters, Face regions, Human computer interaction, Feature extraction**

I. INTRODUCTION

Since last decade, a growing interest in human computer interaction (HCI) systems has been developed. Automated Facial expression recognition is an important task in human computer interaction systems that include emotion processing. Humans are capable of producing thousands of facial actions during communication that vary in complexity, intensity, and meaning. Emotion or intention is often communicated by subtle changes in one or several discrete features. The addition or absence of one or more facial actions may alter its interpretation. In addition, some facial expressions may have a similar gross morphology but indicate varied meaning for different expression intensities.

Automatic facial expression analysis is a flourishing area of research in computer science. Problems that have been tackled with previously are the tracking of facial expression in static images and video sequences, the transfer of expressions to novel faces, the repurposing of a person's expression to a virtual model and recognition of facial expression. These recognition tasks have focused on the classification of emotional expressions [1], classification of complex mental states [2] or the automatic recognition of FACS action units [3].

A problem that is frequently encountered in each of these tasks is that of partial occlusions. Occlusions can introduce errors into the predicted expression or result in an incorrect expression being transferred to a virtual head. One type of partial occlusion is a temporary occlusion caused by a part of the face being obscured momentarily by an object or as a result of a person moving their head so that not all features of the face can be seen by a camera. Another type of occlusion is a systematic occlusion, which can be caused by a person wearing something such as a head-mounted display, which causes the features of the upper half of the face to be invisible. These types of occlusions are potentially more damaging since they result in whole features of relevance to judging facial expression being obscured.

Plenty of work has been done on facial expression recognition [15], [17], [19], [21]. In this study, we investigate the part of the face that contains the most discriminative information for facial expression recognition system and propose hybrid face region method for feature extraction. An automatic classification of facial expressions consists of two stages: feature extraction and feature classification. The feature extraction is extremely important to the whole classification process. If inadequate features are used, even the best classifier could fail to achieve accurate recognition. In most cases of facial expression classification, the process of feature extraction yields a prohibitively large number of features and subsequently a smaller sub-set of features needs to be selected according to some optimality criteria.

The Gabor wavelet feature representation showed high performance in the recognition of facial actions from image sequences. Although the Gabor wavelet facial feature representations have been widely adopted [6], [7], [20], it is computationally expensive to convolve the face images with the multi-level banks of the Gabor filters in order to extract the scale and the orientation coefficients. Furthermore, the Gabor wavelet analysis suffers from two major limitations: the maximum bandwidth of a Gabor filter is limited to approximately one octave and the Gabor filters are not optimal when the objective is to achieve broad spectral information with the maximum spatial localization. These drawbacks can be overcome when using the logarithmic form of the Gabor filters in the process of feature extraction.

The log Gabor filters are known to provide excellent simultaneous localization of spatial and frequency information, however the dimensionality of the resulting data is high. The dimensionality reduction can be achieved by selection of a small sub-set of the log Gabor features based on specified optimality criteria.

The dimensionality reduction can be achieved by selection of the more informative features based on feature selection and data reduction methods such as: principle component analysis (PCA), independent component analysis (ICA), mutual information (MI), etc. [10], [11], [12]. In this

Digital Object Identifier 10.4316/AECE.2009.03012

paper, the mutual information [25] is investigated to select the optimum features for classification. In contrast to the classical correlation-based feature selection methods, the mutual information can measure arbitrary relations between variables and it does not depend on transformations applied to different variables. It can be potentially useful in problems where methods based on linear relations between data are not performing well. Figure 1 illustrates a block diagram of the proposed facial expression recognition system.

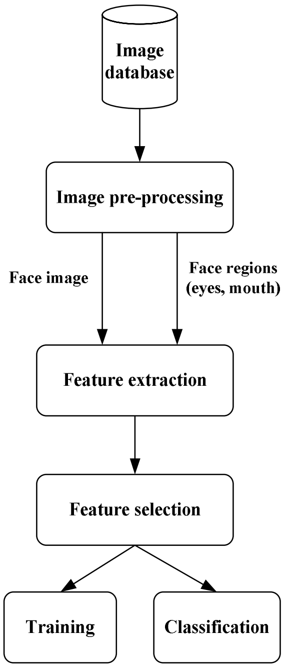

Figure 1. Block diagram of proposed system.

II. IMAGE DATASET

All the facial expression classification tests described here are performed using two popular databases: JAFFE database [6] and Cohn-Kanade database [1]. The Japanese Female Facial Expression (JAFFE) database contains 213 images of 6 basic facial expressions: happiness, anger, disgust, surprise, sadness and surprise, as well as the neutral expression. The images were taken from 10 Japanese female models. The emotions expressed by each picture were subjectively tested on 60 Japanese volunteers. A sample of images from JAFFE database is shown in Figure 2.

Cohn-Kanade database included 388 image sequences from 100 subjects. Each sequence contained 12-16 frames. The subject ages ranged from 18 to 30 years. Sixty five percent of subjects were female; and thirty five percent were male. Fifteen percent of subjects came from the African-American background, and three percent from the Asian or the Latino-American background. The image sequences represented 100 different subjects expressing different stages of an expression development, starting from a low arousal stage, reaching a peak of arousal and then declining. The facial expressions of each subject represented six basic emotions: anger, disgust, fear, happiness, sadness and surprise. Some subjects did not have image sequences corresponding to all of the six expressions, and in some cases, only one image sequence per expression was available. Figure 3 shows an example of sequence images from Cohn-Kanade database.

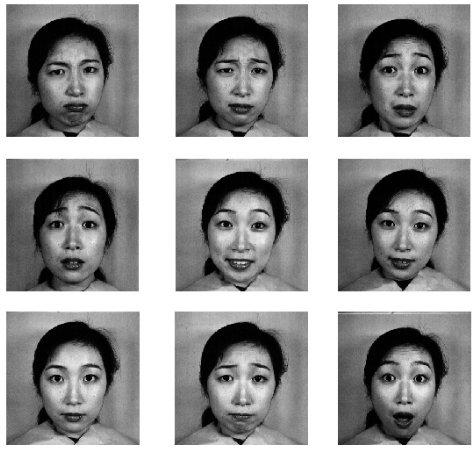

Figure 2. Sample static images from JAFFE database.

Figure 3. Image sequences from Cohn-Kanade database.

III. IMAGE PREPROCESSING

The image pre-processing procedure comes as a very important step in the facial expression recognition task. The aim of the pre-processing phase is to obtain images which have normalized intensity, uniform size and shape, and depict only a face expressing certain emotion. The pre-processing procedure should also eliminate the effects of illumination and lighting. The face area of an image is detected using the Viola-Jones method [4] based on the Haar-like features and AdaBoost learning algorithm. The Viola and Jones method is an object detection algorithm providing competitive object detection rates in real-time. It is primarily designed for the problem of face detection. The features used by Viola and Jones are derived from pixels selected from rectangular areas imposed over the picture and show high sensitivity to the vertical and horizontal lines. AdaBoost is an adaptive learning algorithm that can be used in conjunction with many other learning algorithms to improve their performance. AdaBoost is adaptive in the sense that subsequent classifiers built iteratively are made to fix instances misclassified by previous classifiers. At each iteration, a distribution of weights is updated such that, the weights of each incorrectly classified example, are increased (or alternatively, the weights of each correctly classified example are decreased), so that the new classifier focus more on those examples.

The final stage of the pre-processing for sequence images is detection of a face image which depicted certain emotion with the maximum level of emotion intensity. A new method based on mutual information (MI) which is called facial detection, is used [21], [26]. For each frame, the mutual information between the current frame and the initial frame is calculated, and the frame with the minimum mutual information is selected as the frame that represents an emotion with the maximum intensity. Finally, the images are scaled to the same size. Figure 4 shows the images after pre-processing step.

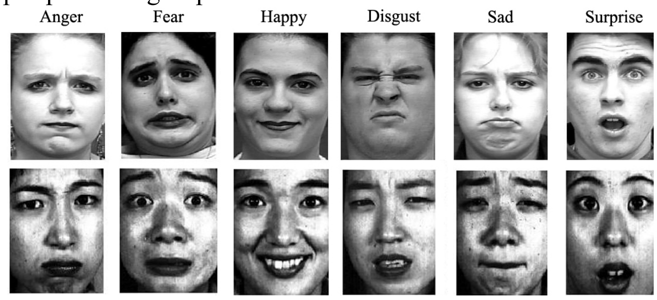

Figure 4. Images after pre-processing step.

For face region location, we use face model to extract the eyes and mouth from the face image. The location of eyes and mouth are shown in Figure 5. The sample of eyes and mouths which are used for testing and training set is illustrated in Figure 6 and Figure **7**.

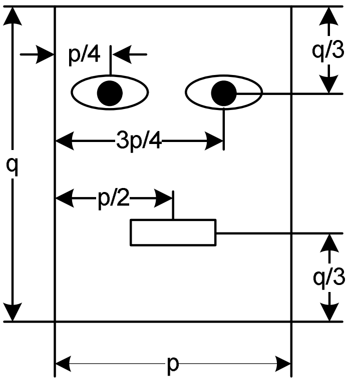

Figure 5. Face template used for eyes and mouth detection.

IV. Feature Extraction

Gabor filters are commonly recognized as one of the best choices for obtaining localized frequency information. However, they suffer from two major limitations. The maximum bandwidth of a Gabor filter is limited to approximately one octave and Gabor filters are not optimal if one is seeking broad spectral information with maximal spatial localization. Log Gabor filters, proposed by Field [13], circumvent this limitation. They always have a null DC component and can be constructed with arbitrary bandwidth which can be optimized to produce a filter with minimal spatial extent. Log Gabor filters in frequency domain can be defined in polar coordinates by $H(f, \theta) = H_f \times H_\theta$, where H_f is the radial component and H_θ, the angular one:

$$H(f,\theta) = \exp\left\{\frac{-[\ln(\frac{f}{f_0})]^2}{2[\ln(\frac{\sigma_f}{f_0})]^2}\right\}\exp\left\{\frac{-(\theta-\theta_0)^2}{2\sigma_\theta^2}\right\} \tag{1}$$

where f_0 is the filters centre frequency, θ_0, the filter direction. The constant σ_f, defines the radial bandwidth B in octaves:

$$B = 2\sqrt{2/\ln 2} \times \left|\ln\left(\frac{\sigma_f}{f_0}\right)\right| \tag{2}$$

the constant σ_θ, defines the angular bandwidth $\Delta\Omega$ in radians:

$$\Delta\Omega = 2\sigma_\theta\sqrt{2\ln 2} \tag{3}$$

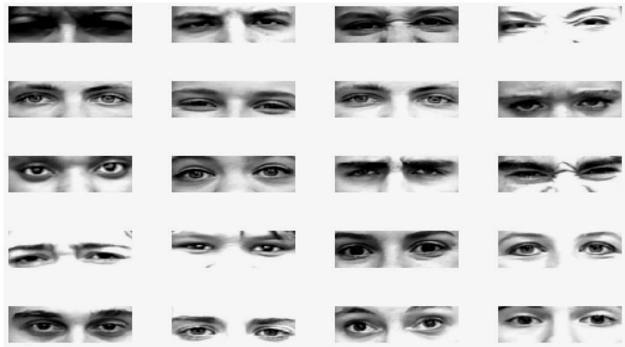

Figure 6. Eye images used for training and testing.

Figure 7. Mouth images used for training and testing.

In the study, the ratio σ_f/f_0 is kept constant for varying f_0, B is set to one octave and the angular bandwidth is set to $\Delta\Omega = \pi/4$ radians. There remains only σ_f, to be determined for a varying value of f_0.

Several scales and orientations are implemented to extract features from face images. This leads to different filter transfer functions representing different scales and orientations. The image filtering is performed in the frequency domain making the process faster as compared to the space domain convolution. After the 2D FFT transformation into the frequency domain, the image arrays I(x,y) are changed into the spectral vectors (I_f) and multiplied by the log Gabor transfer functions, producing different spectral representations for each image. The spectra are then transformed back to the spatial domain via the 2D inverse FFT. To create the feature vector based on HFR method, we extracted the features from eyes and mouth and concatenated them with whole face features (Equation 4).

$$\vec{\mathbf{f}}_{\text{HFR}} = \vec{\mathbf{f}}_{\text{eyes}} \bigcup \vec{\mathbf{f}}_{\text{mouth}} \bigcup \vec{\mathbf{f}}_{\text{face}} \quad (4)$$

This process results in prohibitively large number of feature arrays. For large training and testing sets, the computations are highly impractical. In order to improve the computational efficiency, it is critical to reduce the feature dimensions. This is achieved using feature selection process [5], [10], [25], [26].

V. Feature Selection

Optimal subset of features is selected on the basis of mutual information (MI) criterion [10], [25]. The mutual information represents a measure of information commonly found in two random variables, say X and Y, and it is given as:

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \quad (5)$$

where *p(x)* is the probability density function (pdf), defined as *p(x) = Pr{X=x}*, and *p(x,y)* is the joint pdf defined as *p(x,y) = Pr(X=x and Y=y)*. The MI can also be expressed in terms of the entropy:

$$I(X;Y) = H(X) - H(X \mid Y) \quad (6)$$

where, *H(X)* is the entropy of a random variable X, given as:

$$H(X) = -\sum_{x \in X} p(x) \log p(x) \quad (7)$$

H(X|Y) in Equation 6 is the conditional entropy given as:

$$H(X \mid Y) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log p(y \mid x) \quad (8)$$

The mutual information feature selection (MIFS) algorithm, described in [10] is applied to perform the feature selection. In this approach, starting from the empty set, the best available feature vectors are added, one by one to the selected feature set, until the size of the set reaches the desired value of N_S. The sub-set S of feature vectors are selected using a simultaneous maximization of the mutual information between the selected feature vectors in S and the class labels C, and a minimization of the mutual information between the selected feature vectors within S.

$$I(C;\mathbf{f}_i \mid S) = I(C;\mathbf{f}_i) - \beta \sum_{\mathbf{s}_k \in S} I(\mathbf{f}_i;\mathbf{s}_k) \quad (9)$$

where $I(C;f_i)$ is MI between feature and class label and $I(f_i;s_k)$ is the MI between selected feature and new feature.

As a result, an optimal sub-set $S \subset F$ of mutually independent and highly representative feature vectors is obtained.

VI. Classification

The Naive Bayesian (NB) classifier is a probabilistic method that has been shown to be effective in many classification problems [8], [14]. It assumes that the presence (or lack) of a particular feature of a class is unrelated to the presence (or lack) of any other feature. If using c to represent the value of the class variable, and $\{f_1, \ldots, f_k\}$ for the features, the classification decision is made using the following formula:

$$C = \arg\max_{c} \{p(c) \prod_{j=1}^{k} p(f_j \mid c)\} \quad (10)$$

where *p(c)*= Number of sample in class c / Total samples , (f_i|c) are conditional tables (or conditional density) learned in training by using examples, and k is the length of feature vector. Despite the independence assumption, NB has been shown to have very good classification performance for many real data sets, on par with many more sophisticated classifiers.

VII. Experiments and Results

In this study, we used JAFFE (JF) and Cohn-Kanade (C-K) databases to train and test the facial expression recognition system. Each test was performed 3 times using randomly selected testing and training sets and an average result was calculated. Training has been done for six expressions (C=6). The subjects represented in the training set were not included in the testing set of images, thus ensuring a person-independent classification of facial expressions. Automatic face detection, facial detection, and face region detection were used and the faces were also scaled. The tested images were classified using log Gabor filter for feature extraction and naïve Bayesian classifier. We extracted the features for different scales and orientations and tested them using naïve Bayesian classifier to choose the best scale and orientation for the log Gabor filters. Figure 8 illustrates the recognition rate for different five scales and eight orientations using C-K database. As a result, we have chosen the log Gabor filters with 3 scales and 6 orientations which have the maximum accuracy to do our experiments.

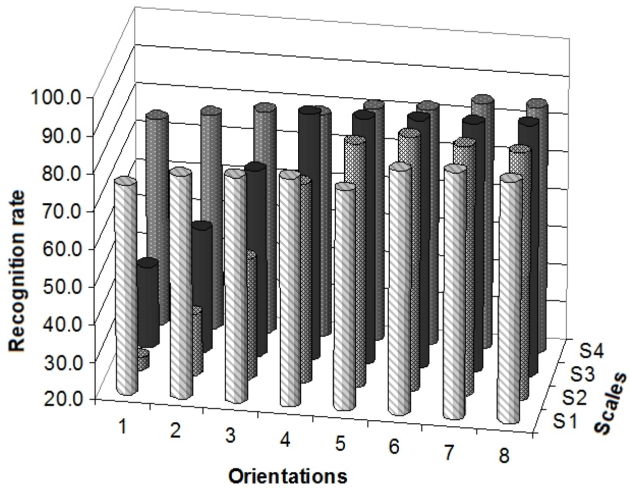

Figure 8. Percentage of recognition rate for 5 scales and 8 orientations based on log Gabor filters.

Furthermore, the expression recognition from different face regions is considered. The classification accuracy for different number of features is shown in Figure 9. The accuracy results are shown in Table I for both C-K and JF databases based on HFR method using log Gabor filter and mutual information. The recognition rate is increased from 87% and 93% on the basis of whole face image to 91.8% and 97.9% based on HFR method for C-K and JF databases respectively. Generally, the accuracy is improved with nearly 5% for both C-K and JF databases.

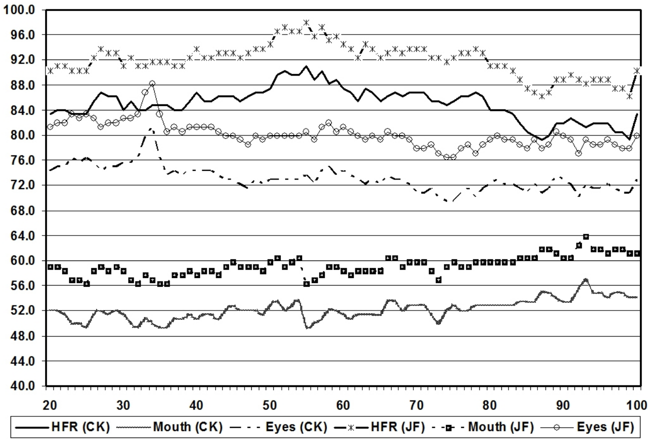

Figure 9. The correct classification rate for different number of features CK: Cohn-Kanade database, JF: JAFFE database.

VIII. CONCLUSION

Facial expression recognition based on hybrid face regions (HFR) has been investigated to determine the part of the face that contains most discriminative information for facial expression classification task. The feature vectors have been extracted from the original images and face regions by multiplying the images with log Gabor filters in frequency domain from face, eyes and mouth images. Then, the most informative features were selected on the basis of mutual information. The features are classified using Naïve Bayesian classifier. We compare the classification results when only eyes or mouth regions are used for classification. Overall, we found that the proposed HFR method is robust for recognizing different expressions.

TABLE I
PERCENTAGE OF CORRECT CLASSIFICATION BASED ON HFR METHOD.

(a) Cohn-Kanade database

	Anger	Disgust	Fear	Happy	Sad	Surprise
Anger	**81.2**	10.1	0.0	0.0	8.7	0.0
Disgust	8.7	**85.1**	6.3	0.0	0.0	0.0
Fear	0.0	0.0	**95.5**	4.5	0.0	0.0
Happy	0.0	0.0	3.0	**97.0**	0.0	0.0
Sad	8.0	0.0	0.0	0.0	**92.0**	0.0
Surprise	0.0	0.0	0.0	0.0	0.0	**100**
Average	**91.8**					

(b) JAFFE database

	Anger	Disgust	Fear	Happy	Sad	Surprise
Anger	**95.8**	0.0	0.0	0.0	4.2	0.0
Disgust	0.0	**96.5**	3.5	0.0	0.0	0.0
Fear	0.0	0.0	**97.9**	2.1	0.0	0.0
Happy	0.0	0.0	1.5	**98.5**	0.0	0.0
Sad	1.6	0.0	0.0	0.0	**98.4**	0.0
Surprise	0.0	0.0	0.0	0.0	0.0	**100**
Average	**97.9**					

REFERENCES

[1] Kanade, T., Cohn, J. F., and Tian, Y., "Comprehensive database for facial expression analysis", Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, Grenoble, France, pp. 46-53, 2000.
[2] Kaliouby, R. E., Robinson, P., "Real-time inference of complex mental states from facial expressions and head gestures", Conference on Computer Vision and Pattern Recognition Workshop, vol. 3, pp. 181-200, 2004.
[3] Tian, Y., Kanade, T., Cohn, J. F., "Recognizing action units for facial expression analysis", IEEE Tran. on Pattern Analysis and Machine Intelligence, vol. 23, no. 2, pp. 97-115, 2001.
[4] Viola, P., Jones, M., "Robust real-time object detection", International Journal of Computer Vision, 57(2), pp. 137-154, 2004.
[5] Guyon, I., Gunn, S., Nikravesh, M., Zadeh, A., "Feature Extraction Foundations and Applications", Springer, 2006.
[6] Lyons, M., Akamatsu, S., Kamachi, M., and Gyoba, J., "Coding facial expressions with Gabor wavelets", In FG'98: Proceedings of the 3rd International Conference on Face and Gesture Recognition, Washington, USA, 1998.
[7] Zheng, D., Zhao, Y., Wang, J., "Features extraction using a Gabor filter family", Proceedings of the Sixth IASTED International Conference Signal and Image Processing, Hawaii, USA, 2004.
[8] Rish, I., "An empirical study of the naive Bayes classifier", IJCAI Workshop on Empirical Methods in Artificial Intelligence, vol. 335, pp. 41-46, 2001.
[9] Claude, F. B., Chibelushi, C., "Facial Expression Recognition: A Brief Tutorial Overview", 2003.
[10] Battiti, R., "Using mutual information for selecting features in supervised neural net learning", IEEE Trans. on Neural Networks, vol. 5, no. 4, pp. 537-550, 1994.
[11] Liu, F., Wang, Z., Wang, L., Meng, X., "Facial expression recognition using HLAC features and WPCA", Lecture Notes in Computer Science, Springer, 2005.
[12] Buciu, I., Kotropoulos, C., and Pitas, I., "ICA and Gabor representation for facial expression recognition", International Conference on Image Processing, vol. 2, pp. 14-17, 2003.
[13] Field, D.J., "Relations between the images and the response properties of cortical cells", Jour. of the Optical Society of America, pp. 2379-2394, 1987.
[14] Duda, R. O., Hart, P. E., Stork, D. G., "Pattern Classification", Wiley, New York, 2001.
[15] Park, S., and Kim, D., "Subtle facial expression recognition using motion magnification", Pattern Recognition Letters, 30(7), pp. 708-716, 2009.
[16] Xie, X., and Lam, K.M., "Facial expression recognition based on shape and texture", Pattern Recognition, 42(5), pp. 1003-1011, 2009.
[17] Kotsia, I., Zafeiriou, S., and Pitas, I., "Novel multiclass classifiers based on the minimization of the within-class variance", IEEE Tran. on Neural Networks, 20(1), pp. 14-34, 2009.
[18] Geetha, A., Ramalingam, V., Palanivel, S., Palaniappan, B., "Facial expression recognition: a real time approach", Expert Systems with Applications, 36(1), pp. 303-308, 2009.
[19] Lajevardi, S. M., Lech, M., "Facial expression recognition using a bank of neural networks and logarithmic Gabor filters", Proceedings of Digital Image Computing: Techniques and Applications (DICTA'08), pp. 77-83, Australia, 2008.
[20] Lajevardi, S. M., Lech, M., "Averaged Gabor filter features for facial expression recognition", Proceedings of Digital Image Computing: Techniques and Applications (DICTA'08), pp. 71-76, Australia, 2008.
[21] Lajevardi, S. M., Lech, M., "Facial expression recognition from image sequences using optimised feature selection", 23rd International Conference on Image and Vision Computing (IVCNZ'08), pp. 1-6, New Zealand, 2008.
[22] Lajevardi, S. M., Hussain, Z. M., "Facial expression recognition: Gabor filters versus higher-order correlators", International Conference on Communication, Computer and Power (ICCCP'08), pp. 354-358, Oman, 2009.
[23] Lajevardi, S. M., Hussain, Z. M., "Facial expression recognition using log-Gabor filters and local binary pattern operators", International Conference on Communication, Computer and Power (ICCCP'08), pp. 349-353, Oman, 2009.
[24] Lajevardi, S. M., Hussain, Z. M., "Zernike moments for facial expression recognition", International Conference on Communication, Computer and Power (ICCCP'08), pp. 371-381, Oman, 2009.
[25] Lajevardi, S. M., Hussain, Z. M., "Feature selection for facial expression recognition based on mutual information", IEEE-GCC'09 Conference, Kuwait, 2009.
[26] Lajevardi, S. M., Hussain, Z. M., "Feature selection for facial expression recognition based on optimization algorithm", Second International Workshop on Nonlinear Dynamics and Synchronization (INDS'09), Klagenfurt, Austria, 2009.