AUT Journal of Modeling and Simulation

AUT J. Model. Simul., 51(1) (2019) 45-56
DOI: 10.22060/miscj.2019.15854.5144

Histogram Based Shape and Textural Characteristics for Facial Emotion Recognition

M. Imani*

Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran.

ABSTRACT: Emotion recognition has many applications in relation between human and machine. A facial emotion recognition framework for 6 basic emotions of happiness, sadness, disgust, surprise, anger and fear is proposed in this paper. The proposed framework utilizes the histogram estimate of shape and textural characteristics of face image. Instead of direct processing on the original gray levels of face image which may have not significant information about facial expression, the processing is done on transformed images containing informative features. The shape features are extracted by morphological operators by reconstruction and the texture ones are acquired by computing the gray-level co-occurrence matrix (GLCM), and applying Gabor filters. The use of whole face image may provide non-informative and redundant information. So, the proposed emotion recognition method just uses the most important components of face such as eyes, nose and mouth. After textural and shape feature extraction, the histogram function is applied to the shape and texture features containing emotional states of face. The simple and powerful nearest neighbor classifier is used for classification of fused histogram features. The experiments show the good performance of the proposed framework compared to some state-of-the-art facial expression methods such as local linear embedding (LLE), Isomap, Morphmap and local directional pattern (LDP).

Review History:

Received: 19 February 2019
Revised: 9 May 2019
Accepted: 14 May 2019
Available Online: 1 June 2019

Keywords:

facial emotion recognition
Gabor
morphology
GLCM
histogram

1. Introduction

Emotion recognition is an important task that has various applications in interaction between human and machine [1]-[3]. Automatic emotion recognition helps development of robots capable of understanding human and responding to people needs. For instance, the service robots which have the ability of emotion recognition are helpful for attendance and treatment of people with disabilities or elderly people. Automatic emotion recognition systems are also efficient in distance education environments, increasing drivers'safety, emotion-sensitive interfaces and management of online markets.

There are various ways for detection of emotional states of human such as direct asking from the user, tracking and investigation of implicit parameters, voice signal processing [4]-[5], vital signal processing [6]-[7], facial expression recognition [8] and gesture recognition [9]. Sometimes two or more techniques are used to form a hybrid method by fusing multi-modal emotional cues [10]-[12]. The most natural way for displaying human emotional states is facial expressions. The facial action coding system (FACS) is one of the main facial expression analysis methods in computer vision problems [13]. FACS defines 46 main action units with considering their intensity. This system can recognize six universally facial emotions: happiness, sadness, disgust, surprise, anger, fear and their combinations. Each facial expression is result of muscle contractions and the degree of changes in facial muscles can determine the intensity of the emotional state. Most of emotion recognition methods use the geometrical properties of face image for detection of emotional states. They determine some important points on the face images such as the middle point of upper part of brow, corner of eyes, the middle upper point of mouth and so on. Then, they compare the distance between the obtained points in the testing image with respect to those present in a natural image. Because different facial expressions generate different ratios about distances between the main face points, this comparison can recognize the emotional state. Several facial emotion recognition methods use the feature extraction methods to obtain an appropriate feature vector for representation of the face image and then fed it to a classifier. Some of the appearance features are histograms of oriented gradients (HOG) [14], principal component analysis (PCA) [15], local binary patterns (LBP) [16], Isomap [17] and local linear embedding (LLE) [18]. The Morphmap method has been used to construct a neighbourhood graph using intensity attenuation function where pixels in the layer that come under the angle of divergence can compose a part of group vectors and other vectors are deleted [19]. The multiclass support vector machine has been used to classify

*Corresponding author's email: maryam.imani@modares.ac.ir

various facial expressions. Emotion-related features of face contain prominent gradient magnitudes. So, methods such as local directional pattern (LDP) that extract edge based local features and histogram representation can achieve success in facial expression recognition problem [20].

Although many researchers have been worked on visible facial images, several works have used the thermal facial images. While visible images contain illumination information, the thermal images record temperature distribution which are robust to light changes [21]. Fusion of thermal and visible images can enhance the emotion recognition performance. Integration of both visible and thermal images by using support vector machine and deep models is suggested in [22].

Facial expressions are dynamically changed during the time. To dynamic study of emotions through decomposition of the facial expression event into muscle movements of different regions of face over different time instances, two issues can be considered: semantic based representation and temporal alignment. These issues are studied using manifold modelling of videos in [23]. Facial expression recognition is investigated in [24] where factors such as variations of head pose has been taken to account. Before recognition of expressions, an intermediate face representation is learned by using the global geometry features and the local appearance ones.

The proposed method in [25] uses the geometric features and the regional local binary patterns for emotion recognition. The extracted features are fused by an autoencoder and then classified by a self organization map based classifier. In [26], local directional histogram features are combined with the local directional strength features. Finally, the extracted features are used to train the convolutional neural network, which is a deep learning approach. Inaccurate alignment due to projection of 3D face on a 2D image plane causes major problems in the emotion recognition process. To deal with this difficulty, a multi angle pattern based deep learning method is proposed in [27]. The main benefits of this method are effective handling of illumination/pose variation and better facial alignment. Although deep learning method have great success in emotion recognition problems, but their efficiency is significantly affected by correct labels of target images. But, many image datasets have not be marked. To address this issue, a residential convolutional neural network has been proposed for keypoint detection in [28]. It does not consider the error in prediction of missing targets in the output layer.

Support vector regression is combined with a fuzzy model to recognize emotions in [29]. Three components of age, gender and province are used for data classification. The fuzzy model is used for intention generation using identification information. The use of fuzzy model is preferred than the single use of support vector regression. Fuzzy is also combined with neural models in many works such as [30].

The gray level co-occurrence matrix (GLCM) is one of the popular methods for texture feature extraction [31]. It acquires a parametric representation of spatial arrangement of image pixels with high level of discrimination between textural patterns. GLCM that is a complex matrix constructed in a local window consists of the frequency values of all co-occurring pixel intensities. A Gabor filter, consists of various directions and scales, with capturing potential physical structures of objects present in image is employed for obtaining the localization properties in both spatial and frequency domains [32]-[34]. Morphological profile (MP) by construction a multi-scale decomposition of a given image through applying a sequence of the opening and closing operators provides a good source of shape and geometrical structures of different objects of the image [35]-[38].

In this paper, a facial emotion recognition framework is proposed which extracts several useful appearance features from the face and then uses the simple nearest neighbor classifier for classification. It uses the shape features extracted by morphological filters or texture features extracted from GLCM or by Gabor filters. Then, the histogram function which provides a non-parametric statistical estimate of the probability distribution of face image is applied to the shape and texture features. Instead of direct applying the histogram function to whole face image, it is locally applied to some main components of face image that are effective for detection of emotional expressions. The proposed method is assessed from the recognition rate point of view compared to some state-of-the-art methods and the experiments show the superior performance of it.

II. PROPOSED FACIAL EMOTION RECOGNITION METHOD

A histogram based feature extraction framework by using shape and textural characteristics of face image is proposed for emotion recognition. The proposed framework consists of the following steps:

1- Detection of main components of the face image such as left eye, right eye, nose and mouth.

2- Extraction of shape or texture features from the main face components through applying morphological filters, Gabor filters or GLCM operators.

3- Calculation of local histogram of spatial (shape and texture) features extracted from each face component.

4- Fusion of histogram features of different face components through stacking.

5- Classification of the achieved feature vector by using th nearest neighbor classifier.

Each step of the proposed framework is described with more details in the following. At first the Viola-Jones algorithm is used for detection of face and its main components (eyes, nose and mouth) [39]. The Viola-Jones algorithm calculates the scalar product between the Haar like templates and image for Haar like feature extraction. Then, it uses Adaboost for feature selection and cascading classifiers for object detection through an efficient computational resource allocation.

A) GLCM

GLCM is an image processing technique for extraction of texture features. GLCM represents the occurance frequency of different combinations of gray level pixels in a given image. The contents of GLCM are used to provide a measure of variations in intensity at image pixels. GLCM extracts the second order statistical texture features. It is a matrix that has rows and columns with the same number of gray levels in the image. Let two pixels with gray level i and j are located within a given neighborhood with a pixel distance $(\Delta x,\Delta y)$. Each element of GLCM matrix denoted by $P(i,j|\Delta x,\Delta y)$ indicates the relative frequency of two mentioned pixels. In other representation, the GLCM matrix elements can be denoted by $P(i,j|d,\theta)$ where it indicates the second order statistical probability of occurance of two pixels with gray levels of i and j located in a distance d and an angle θ. By considering an input image with the size of $R\times C$ containing G gray levels (from 0 to G-1) and $I(m,n)$ as the intensity at sample $(r,c);r=1,\dots,R;c=1,\dots,C$, the GLCM element is calculated by [31], [41]:

$$P(i,j|\Delta x,\Delta y)=WL(i,j|\Delta x,\Delta y) \tag{1}$$

where

$$L(i,j|\Delta x,\Delta y)=\sum_{c=1}^{C-\Delta y}\sum_{r=1}^{R-\Delta x}A \tag{2}$$

$$A=\begin{cases}1 & ;if\ I(r,c)=i\ and\ I(r+\Delta x,c+\Delta y)=j\\ 0 & ;O.W.\end{cases} \tag{3}$$

and

$$W=\frac{1}{(R-\Delta x)(C-\Delta y)} \tag{4}$$

For each image pixel, a $G\times G$ GLCM matrix is calculated. Different kinds of texture features can be extracted from the GLCM matrices. Some of the main features are: variance, contrast, correlation, entropy, angular second moment and inverse different moment which are calculated by:

$$\text{Variance}=\sum_i\sum_j(i-\mu)^2P(i,j) \tag{5}$$

$$\text{Contrast}=\sum_{n=0}^{N_g-1}n^2\left\{\sum_{\substack{i=1\\|i-j|=n}}^{N_g}\sum_{j=1}^{N_g}P(i,j)\right\} \tag{6}$$

$$\text{Correlation}=\frac{\sum_i\sum_j(ij)P(i,j)-\mu_x\mu_y}{\sigma_x\sigma_y} \tag{7}$$

$$\text{Entropy}=-\sum_i\sum_j P(i,j)\log(P(i,j)) \tag{8}$$

$$\text{Angular second moment}=\sum_i\sum_j\{P(i,j)\}^2 \tag{9}$$

$$\text{Inverse different moment}=\sum_i\sum_j\frac{1}{1+(i-j)^2}P(i,j) \tag{10}$$

where $P(a,b)$ is the (a,b)th entry of the normalized GLCM:

$$P(i,j)=\frac{P(i,j)}{\sum_i\sum_j P(i,j)} \tag{11}$$

and μ_x, μ_y, σ_x, and σ_y are the means values and the standard deviations of $P_x(i)=\sum_{(j=1)}^{N_g}P(i,j)$ and $P_y(j)=\sum_{(i=1)}^{N_g}P(i,j)$.

B) Gabor

Gabor filters can analysis the textural characteristics of a given image through a multiresolution representation. A Gabor filter is a bandpass filter where its impulse response is generated by multiplying a complex oscillation with Gaussian envelope function. Extension of these functions to two-dimensional space allows creation of filters that are selective for orientation. To achieve localization characteristics in both frequency and spatial domains, a Gabor filter bank consisting of various directions and scales. The filters are convolved with the image to find the Gabor space. A set of Gabor filters with scale $s=1,\dots,N_s$ and direction $d=1,\dots,N_d$ is defined as follows where N_s and N_d are the number of scales and the number of directions, respectively [34], [41]:

$$h_{s,d}(x,y)=\alpha^{-s}\Phi(X,Y) \tag{12}$$

where $\alpha=(U_h/U_l)^{1/(Ns-1)})$ and $\Phi(X,Y)$ is the wavelet mother defined by:

$$\Phi(x,y)=\frac{1}{2\pi\sigma_x\sigma_y}exp\left\{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)+2\pi jU_hx\right\} \tag{13}$$

The definition of used parameters is as follows:

$$\begin{cases}X=\alpha^{-s}\left[(x-x_0)cos\left(\frac{\pi d}{N_d}\right)+(y-y_0)sin\left(\frac{\pi d}{N_d}\right)\right]\\ Y=\alpha^{-s}\left[-(x-x_0)sin\left(\frac{\pi d}{N_d}\right)+(y-y_0)cos\left(\frac{\pi d}{N_d}\right)\right]\end{cases} \tag{14}$$

$$\begin{cases}\sigma_x=\frac{(\alpha+1)\sqrt{ln4}}{2\pi(\alpha-1)U_h}\\ \sigma_y=\frac{\sqrt{ln4-\left(\frac{ln4}{2\pi U_h\sigma_x}\right)^2}}{2\pi\ tan\left(\frac{\pi}{2N_d}\right)\left[U_h-\frac{ln4}{U_h(2\pi\sigma_x)^2}\right]}\end{cases} \tag{15}$$

where U_l and U_h indicate the minimum and maximum centre frequency of the Gabor filters, respectively and (x_0,y_0) d denotes the filters centre coordination in the spatial domain.

C) Morphological profile (MP)

The morphological filters provide a multi-scale decomposition of a given image to extract spatial features such as shape and size of present objects in the image. The geometrical characteristics of the sliding window, called structural element (SE), determine the degree of processing. There are two basic operators in mathematical morphology: erosion and dilation. The erosion operator results in an output image that represents where SE fits the objects and the dilation operator provides an output image that represents where the SE hits the objects. Two widely used morphological filters are closing and opening filters. While the closing filter erodes the dilated image and removes the dark connected components, the opening filter dilates the eroded image and removes the connected components brighter than the intensity of neighboring pixels. The closing and opening profiles are obtained by applying a sequence of closing and opening filters with different SE sizes. The closing and opening operators by reconstruction are preferred for spatial feature extraction from the image than the conventional closing and opening operators. In the filtering by reconstruction, the shape of used SE is adaptive with respect to the present structures in the image. So, the operators by reconstruction can better maintain shapes than the conventional morphological operators. Let $\varphi_\lambda^*(x)$ be a morphological closing operator by reconstruction. The closing profile at pixel of the image indicated by $\Pi_\varphi(x)$ is obtained by using a SE with the size of λ [36]:

$$\Pi_\varphi(x) = \{\Pi_{\varphi\lambda}: \Pi_{\varphi\lambda} = \varphi_\lambda^*(x), \forall\lambda \in [0,n]\} \tag{16}$$

In duality, γ_λ^* is a morphological opening operator by reconstruction and the opening profile at the pixel x, $\Pi_\gamma(x)$ is defined by:

$$\Pi_\gamma(x) = \{\Pi_{\gamma\lambda}: \Pi_{\gamma\lambda} = \gamma_\lambda^*(x), \forall\lambda \in [0,n]\} \tag{17}$$

By applying n closing operators and n opening operators by reconstruction, a MP with $2n+1$ features containing spatial information is provided as:

$$MP_n(I) = \{\varphi_1^*(I), \dots, \varphi_n^*(I), I, \gamma_1^*(I), \dots, \gamma_n^*(I)\} \tag{18}$$

D) Histogram function

Histogram is an estimate of the probability distribution of a particular data such as a given image. Histogram of an image represents the tonal distribution of gray levels where it is contains the frequency of appearance of different intensities. An occurance histogram function as a non-parametric statistical estimate is applied to the texture features extracted by GLCM, morphological operators or Gabor filters from each face image component. For an image with gray levels in the range of [0,G-1], the histogram is defined as a discrete function:

$$hist(f_k, m) = n_k \tag{19}$$

where f_k is kth gray level value and n_k denotes the number of pixels with gray level of f_k and m denotes the number of bins. The number of bins determines the number of extracted features from each face component. The histogram features extracted from different face components are fused together by stacking to find the final feature vector of the face image.

E) Feature extraction

There are two main categories of information in a face image: 1-chromatic or intensity information and 2-shape and texture information. For detection of facial expression, the latter category has the most importance. If the histogram function is directly applied to the face image, i.e., to the original gray values of image, it provides only a representation of intensity values distribution which has not significant information about facial expression. But, when the histogram function is applied to the textural features such as morphological, GLCM and Gabor features, a representation of distribution of texture features is provided which is efficient for facial expression detection. Moreover, the whole face image does not contain the same amount of emotional information. So, by applying histogram function to entire face image globally, much redundant features are extracted which may decrease the correct recognition rate. To deal with this problem, the histogram function is individually applied to the face components (eyes, nose and mouth). Then, the histogram values of different face elements are stacked together to provide the final feature vector of the face image. The extracted feature vector is given to a classifier. The nearest neighbor classifier is chosen in this paper because of its efficiency and simplicity. The main advantages of the proposed facial emotion recognition framework are represented in the following:

1- The proposed framework uses the valuable spatial information extracted by GLCM, Gabor and morphological filters containing shape and textural characteristics of face which are efficient for facial expression detection.

2- Instead of using entire face, the proposed framework just uses the main components of face, i.e., eyes, nose and mouth. So, the redundant information which may increase classification error is decreased. Moreover, the computational burden is reduced.

3- The local histogram is used for extraction of contextual features from the GLCM, Gabor and morphological filtered face components. The use of histogram function with bins not only reduces the number of texture features to m features, but also, provides a nonparametric statistical estimate of the shape and textural characteristics of the main components.

III. EXPERIMENTS

The performance of the proposed facial emotion recognition method is evaluated in this section. The Japanese Female Facial Expression (JAFFE) dataset is used which contains 213 images acquired from 10 Japanese female models in 7 different facial expressions [40]. The expressions consist of happiness, sadness, surprise, anger, disgust, surprise, and neutrality. Some samples of this dataset are shown in Fig. 1.

The proposed framework can be individually implemented using MP, GLCM or Gabor approaches. The associated

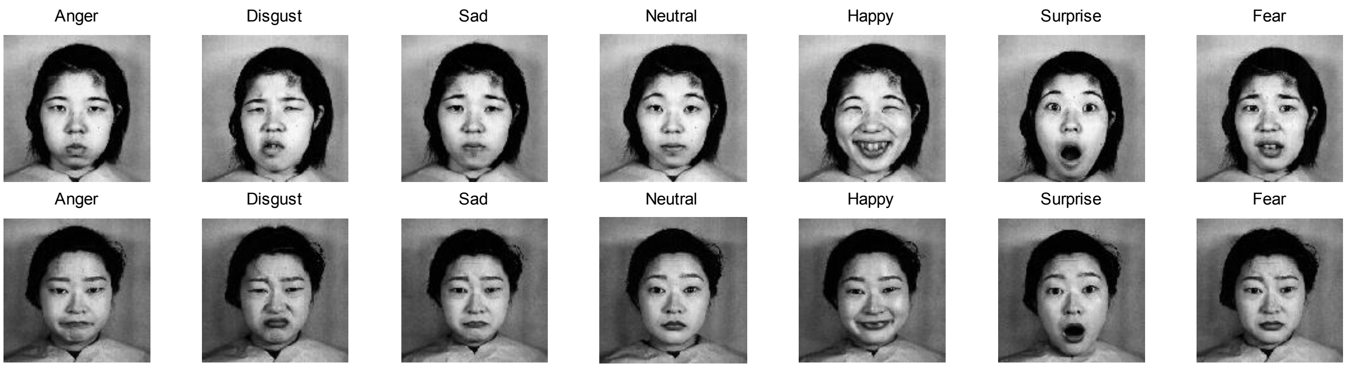

Fig. 1. Samples of JAFFE dataset.

proposed methods are called as Hist-MP, Hist-GLCM and Hist-Gabor. The first step in the proposed framework is detection of face components from the face image using the Viola-Jones algorithm. The results are shown in Fig. 2. After separation of facial components such as eyes, nose and mouth, the morphological filters are applied for shape characteristics extraction; GLCM operators are implemented for texture feature extraction; and the Gabor filters are applied to obtain texture features in different scales and directions. Some samples of morphological, GLCM and Gabor features of face components are shown in Figs. 3-5, respectively.

The latter step in the proposed framework is estimate of histogram from the texture features obtained from the previous step. In applying histogram function to the face image in the proposed framework, there are two main points, which are recalled in the following:

1- Histogram function, instead of direct applying to the intensity values of the face image, is applied to the MP, GLCM or Gabor filtered face components. Direct applying of histogram function to the original face image provides a distribution of pixels gray level which has not significant information about the facial expressions. The emotional states of face are generally expressed through variations in shape and texture of eyes, nose and mouth. So, at first, the shape features such as morphological ones or texture features, i.e., Gabor and GLCM ones are extracted. Then, the histogram function is applied to the filtered image.

2- Since the entire face has not the same contribution in composing the emotional expression, the use of whole face image for feature extraction generates non-informative and redundant features which may be mis-leading. So, instead of global applying the histogram function to entire face image, it is locally applied to the main face components.

The histogram curve, computed by (19), acquired from the morphological, GLCM and Gabor filtered face components are shown in Figs. 6-8, respectively. These histogram based features are finally given to a nearest neighbor classifier for classification.

The features extracted by three proposed methods, i.e., Hist-MP, Hist-GLCM and Hist-Gabor increase class discrimination and so, improve the classification accuracy. In other words, the extracted features reduce overlapping among different emotions in the new transformed feature space. As an example, the scatter plot of two emotions of happiness and anger are shown in Fig. 9 in a two dimensional feature space (using two extracted features). Beside the proposed methods, the scatter plot of emotions in the original feature space using the original grey levels of images is shown. As

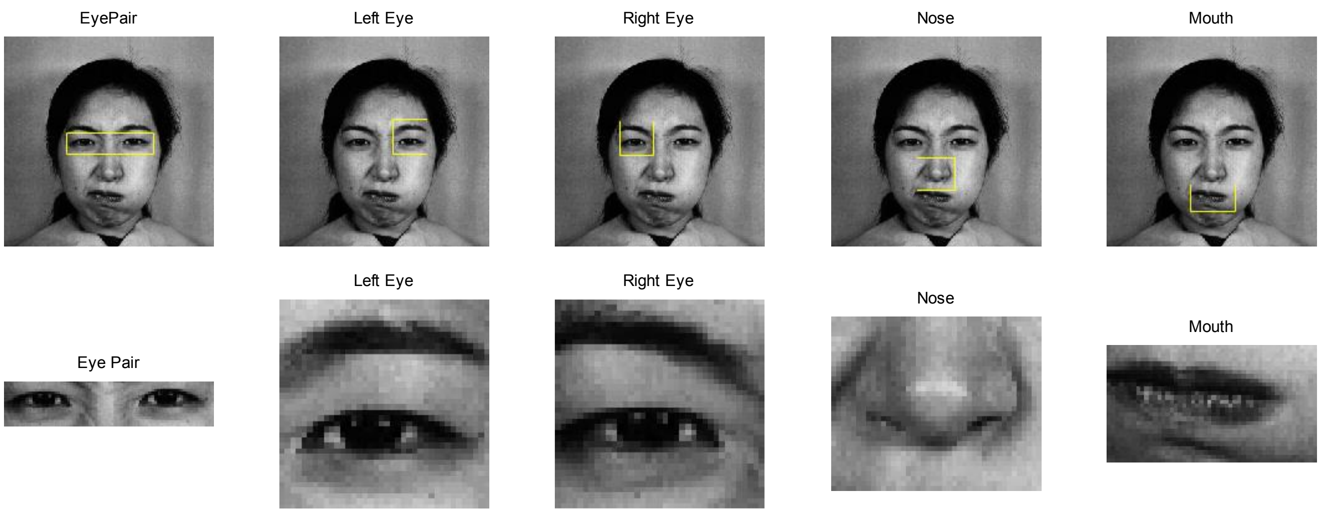

Fig. 2. Detection of face components using Viola-Jones algorithm.

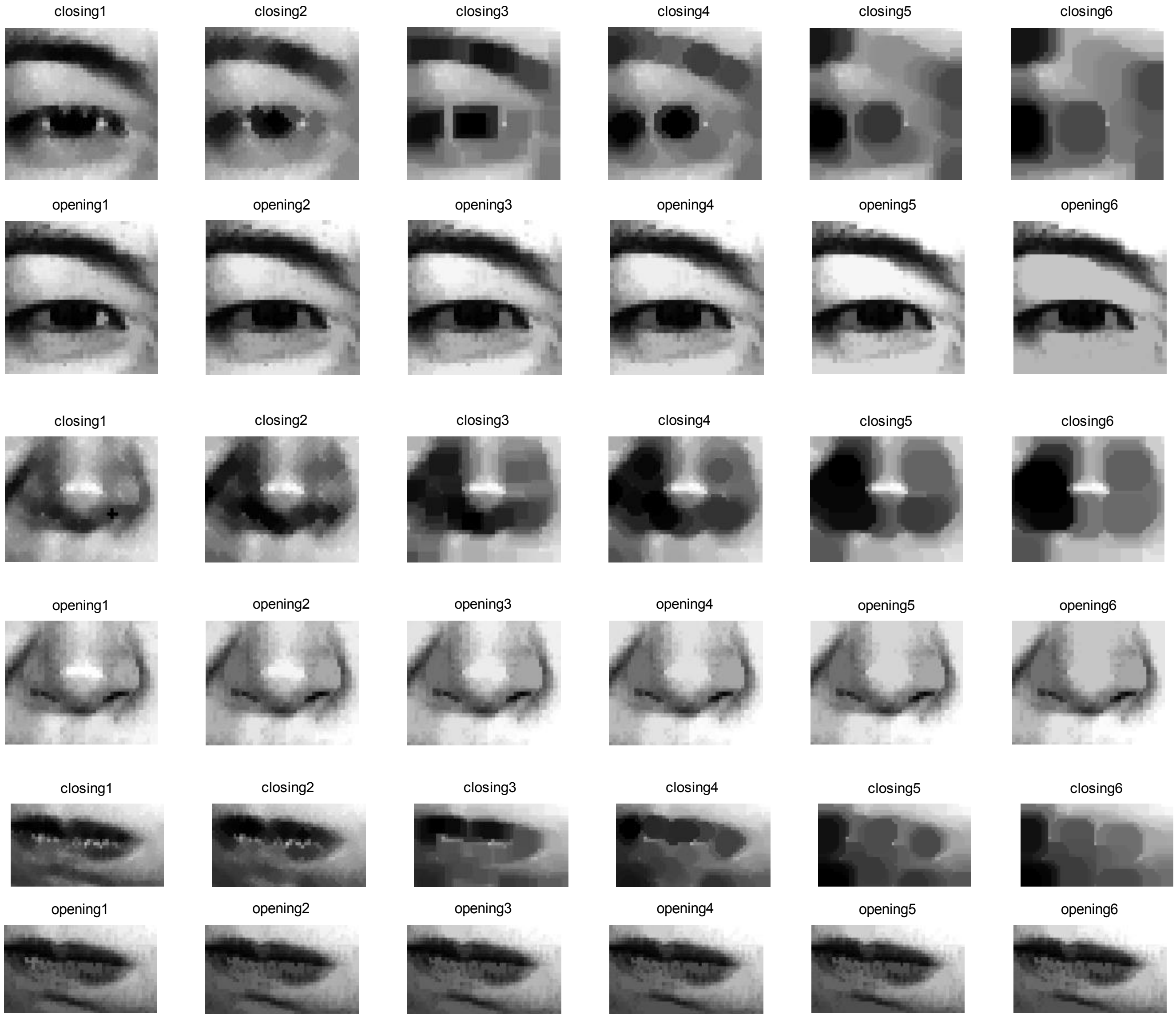

Fig. 3. Samples of morphological features of the face components (right eye, nose and mouth).

seen from this figure, two emotions are highly correlated in the original feature space. But, overlapping of two emotions is significantly decreased in the new feature space of Hist-GLCM and specially in the feature spaces of Hist-MP and Hist-Gabor.

The free parameters of each approach are set manually through experiments and according to previous studies. The fast GLCM method with d=1,θ=0 is used for calculation of GLCM matrices with a square neighbourhood window of size of $Winlen \times Winlen$ with $Winlen$=7 . According to [41], 16 features are calculated from the GLCM matrix. The Gabor filter is implemented with the following parameters: $Winlen$=8; U_h=1; U_l=0.01; N_s=6; N_d=2. The MP is also provided by applying a SE with the shape of 'disk' and radius

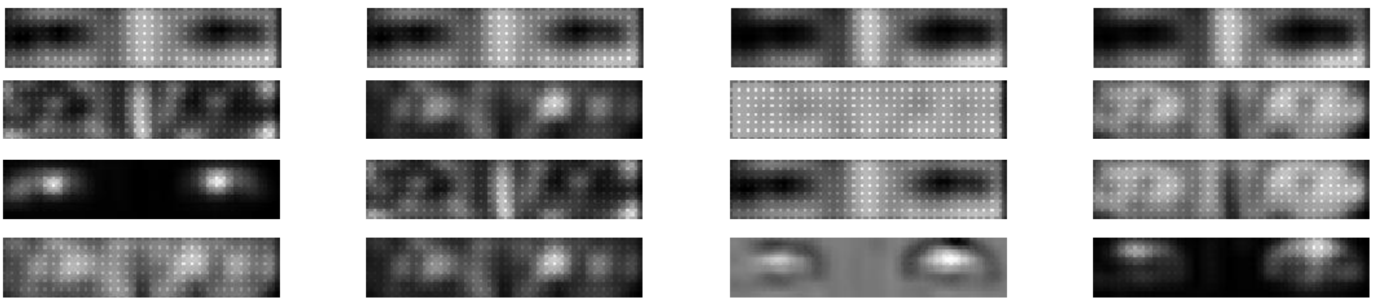

Fig. 4. Samples of GLCM features of the face components (eye pair).

Fig. 5. Samples of Gabor features of the face components (eye pair).

R□$\{1,2,,3,4,5\}$.

The experiments are done using two different number of histogram bins: m=10 and m=100. On each filtered component of face image such as eyes, nose and mouth, the histogram function with bins is applied. In other words, histogram features are extracted from the MP, GLCM or Gabor filtered face components. The confusion matrix results for Hist-MP, Hist-GLCM and Hist-Gabor with m=10 are shown in Tables 1-3 and the results associated with m=100 are reported in Tables 4-6, respectively. The following conclusions can be found from the obtained results:

1- Increasing the number of histogram bins can increase the classification accuracy. This is expected, because with increasing the number of histogram bins, a more accurate estimate of probability of face components is obtained.

2- Often, the nature state can be detected with the highest accuracy.

3- According to the confusion matrices, the emotions of sadness, fear and disgust are usually the most overlapped respect together and discrimination between them is a hard

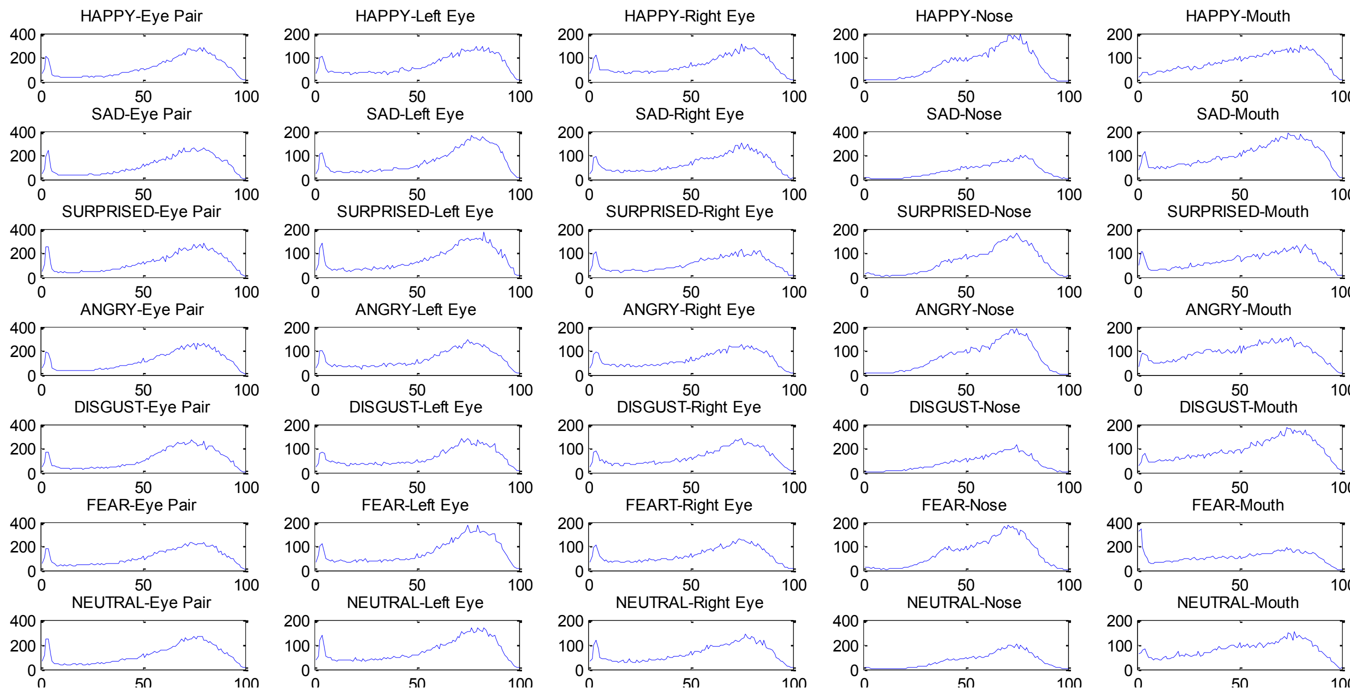

Fig. 6. Histogram of the morphological features of the face components.

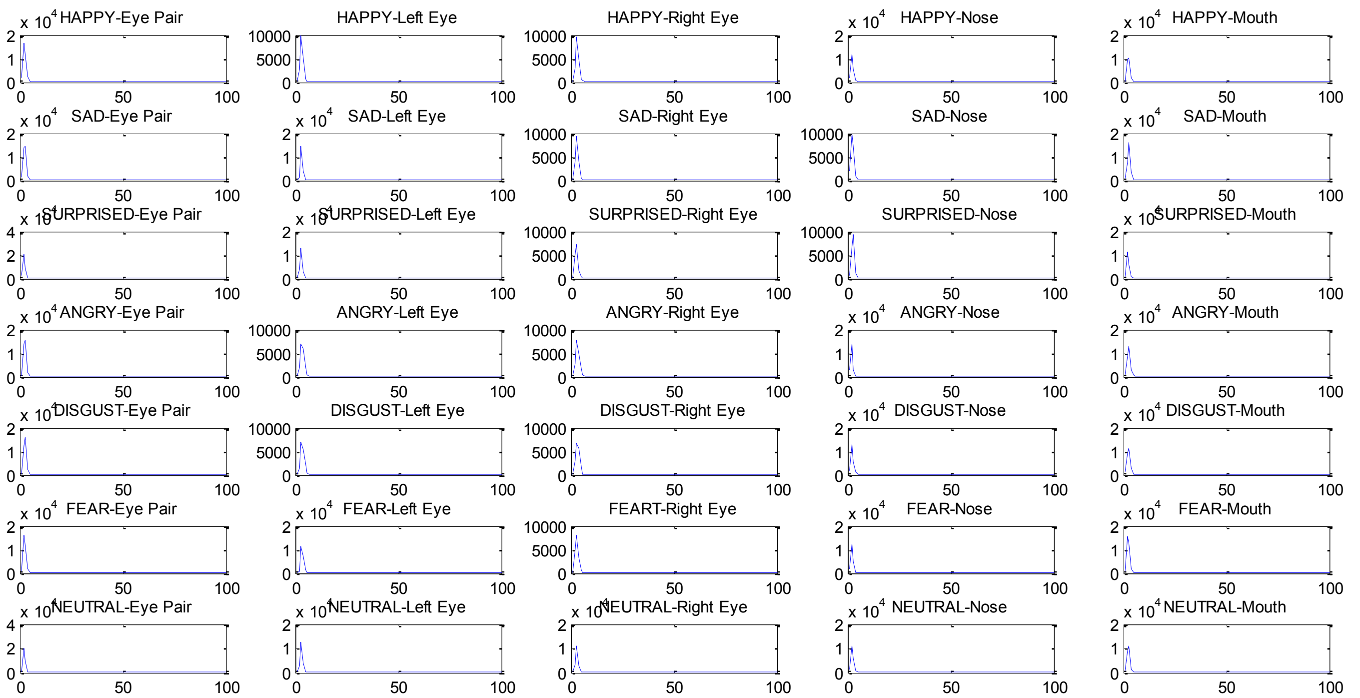

Fig. 7. Histogram of the GLCM features of the face components.

task.

4- Two emotions of happy and surprise usually have similar features. So, according to the obtained confusion matrices, these emotions are usually detected instead and wrongly confused.

5- With both m=10 and m=100 , the best facial recognition method is obtained by Hist-Gabor. After Hist-Gabor, the Hist-GLCM is better with m=10 and Hist-MP is better with m=100.

To provide a statistical assessment on the classification results, the McNemars test is used where the standardized normal test statistic is calculated by:

$$Z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}} \tag{20}$$

where f_{12} is the number of images correctly by method 1 and incorrectly by method 2. The sign of Z indicates whether

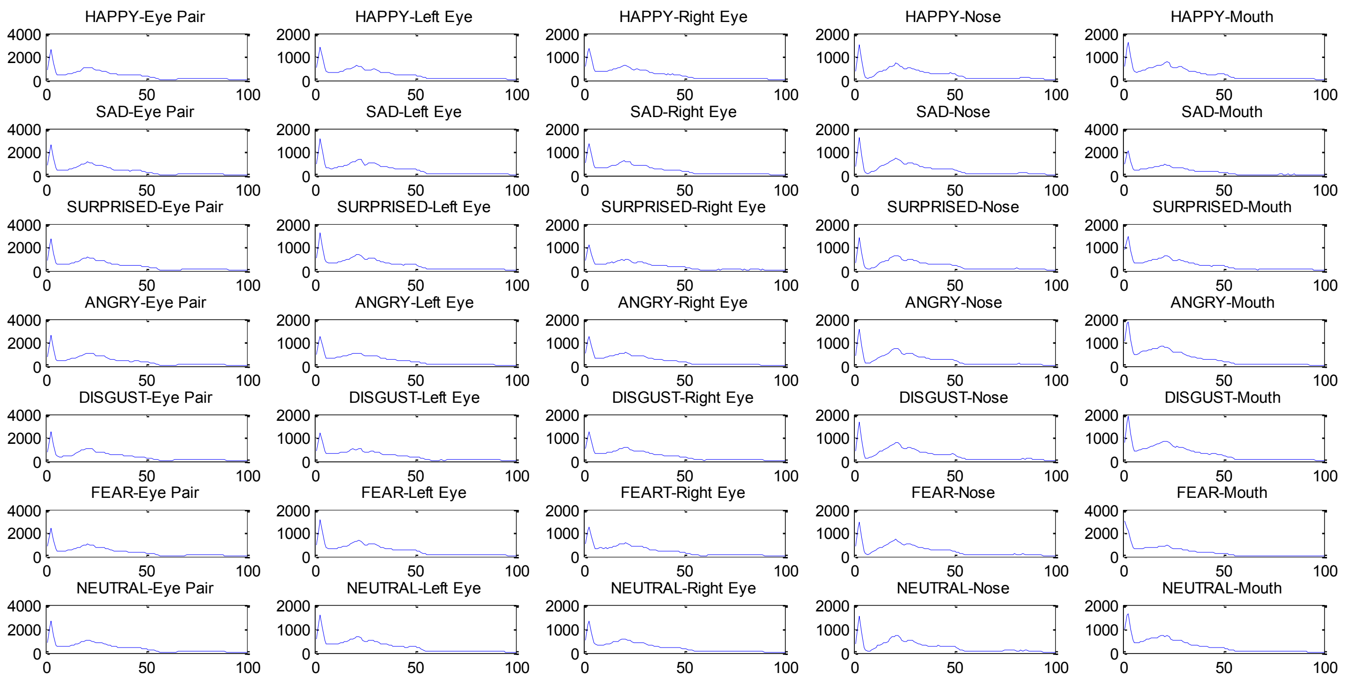

Fig. 8. Histogram of the Gabor features of the face components.

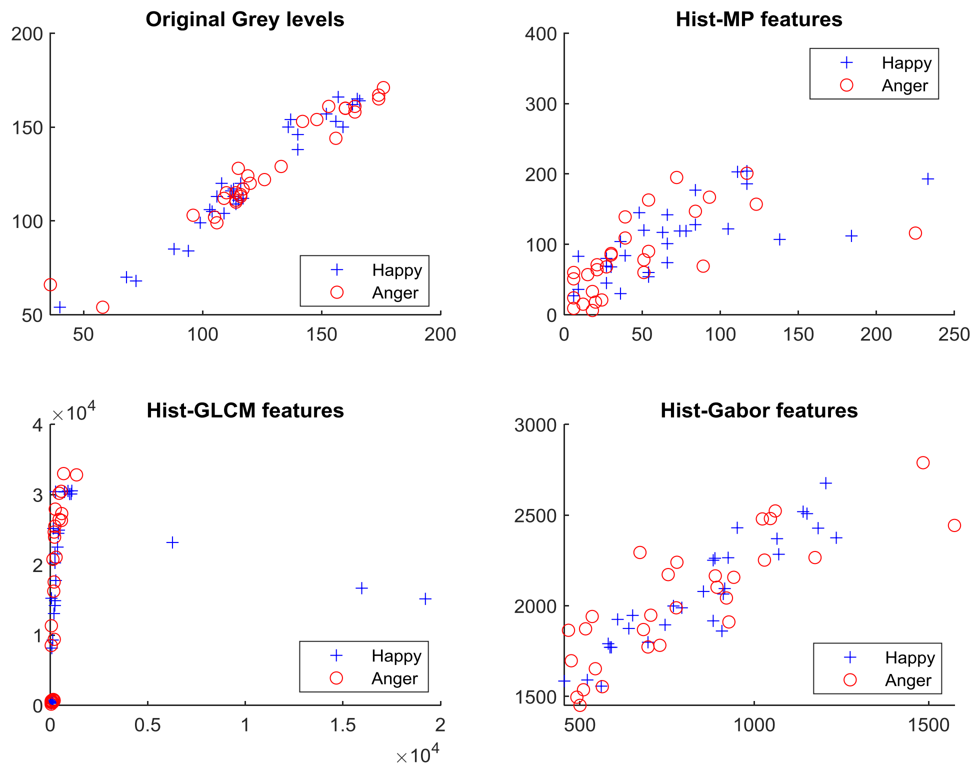

Fig. 9. Two dimensional scatter plot of two emotions of happiness and anger in the original feature space and in the transformed feature spaces of Hist-MP, Hist-GLCM and Hist-Gabor.

Table 1. Confusion matrix results for Hist-MP with *m*=10.

	Happy	Sadness	Surprise	Anger	Disgust	Fear	Neutral
Happy	67.74	9.68	6.45	9.68	3.23	0	3.23
Sadness	0	77.42	3.23	3.23	0	9.68	6.45
Surprise	3.33	0	73.33	3.33	6.67	0	13.33
Anger	6.67	6.67	0	80.00	3.33	3.33	0
Disgust	0	6.90	0	6.90	75.86	10.34	0
Fear	6.25	15.63	0	0	6.25	68.75	3.13
Neutral	3.33	6.67	3.33	0	0	3.33	83.33
Average accuracy= 75.21 Average reliability= 75.83							

Table 2. Confusion matrix results for Hist-GLCM with *m*=10.

	Happy	Sadness	Surprise	Anger	Disgust	Fear	Neutral
Happy	67.74	3.23	9.68	6.45	3.23	9.68	0
Sadness	3.23	77.42	3.23	6.45	3.23	3.23	3.23
Surprise	10.00	0	80.00	0	0	3.33	6.67
Anger	3.33	10.00	0	76.67	0	0	10.00
Disgust	0	0	0	10.34	79.31	6.90	3.45
Fear	3.13	6.25	9.38	0	3.13	68.75	9.38
Neutral	3.33	3.33	3.33	0	0	6.67	83.33
Average accuracy= 76.17 Average reliability= 76.45							

Table 3. Confusion matrix results for Hist- Gabor with *m*=10.

	Happy	Sadness	Surprise	Anger	Disgust	Fear	Neutral
Happy	74.19	3.23	6.45	6.45	0	6.45	3.23
Sadness	0	80.65	0	3.23	0	12.90	3.23
Surprise	10.00	3.33	76.67	0	3.33	0	6.67
Anger	6.67	13.33	0	73.33	0	0	6.67
Disgust	0	3.45	0	10.34	75.86	3.45	6.90
Fear	3.13	9.38	6.25	0	0	71.88	9.38
Neutral	3.33	0	3.33	0	0	0	93.33
Average accuracy= 78.00 Average reliability= 78.92							

Table 4. Confusion matrix results for Hist-MP with *m*=100.

	Happy	Sadness	Surprise	Anger	Disgust	Fear	Neutral
Happy	70.97	3.23	9.68	6.45	0	6.45	3.23
Sadness	0	80.65	3.23	3.23	3.23	6.45	3.23
Surprise	3.33	0	73.33	0	3.33	0	20.00
Anger	0	0	0	80.00	10.00	0	10.00
Disgust	0	0	0	6.90	82.76	10.34	0
Fear	3.13	0	0	0	15.63	81.25	0
Neutral	3.33	0	6.67	3.33	3.33	0	83.33
Average accuracy= 78.90 Average reliability= 80.05							

Table 5. Confusion matrix results for Hist-GLCM with *m*=100

	Happy	Sadness	Surprise	Anger	Disgust	Fear	Neutral
Happy	67.74	0	9.68	3.23	3.23	12.90	3.23
Sadness	3.23	83.87	3.23	6.45	0	0	3.23
Surprise	6.67	3.33	76.67	0	6.67	0	6.67
Anger	6.67	3.33	3.33	73.33	6.67	3.33	3.33
Disgust	3.45	0	0	0	89.66	6.90	0
Fear	3.13	6.25	6.25	9.38	6.25	68.75	0
Neutral	3.33	3.33	3.33	3.33	10.00	3.33	73.33
Average accuracy= 76.19 Average reliability= 76.31							

Table 6. Confusion matrix results for Hist- Gabor with *m*=100.

	Happy	Sadness	Surprise	Anger	Disgust	Fear	Neutral
Happy	70.97	6.45	9.68	0	6.45	3.23	3.23
Sadness	3.23	80.65	0	6.45	3.23	6.45	0
Surprise	16.67	0	73.33	0	3.33	0	6.67
Anger	3.33	10.00	0	80.00	0	3.33	3.33
Disgust	3.45	0	0	6.90	86.21	3.45	0
Fear	3.13	12.50	3.13	0	0	71.88	9.38
Neutral	6.67	0	3.33	0	0	0	90.00
Average accuracy= 79.00 Average reliability= 79.36							

Table 7. Comparison of McNemars test values.

	Hist-MP	Hist-GLCM	Hist-Gabor
Hist-MP	0	0.96	0.83
Hist-GLCM	-0.96	0	-0.55
Hist-Gabor	-0.83	0.68	0

Table 8. Comparison results between different facial emotion recognition methods.

	LLE	Isomap	Morphmap	Gabor	LDP	Hist-MP	Hist-GLCM	Hist-Gabor
Recognition Accuracy	71.23	72.46	75.85	76.15	78.04	78.90	76.19	79.00

method 1 is more accurate than method 2 ($Z > 0$) or vice versa ($Z < 0$). In addition, the difference in the accuracy between two methods is statistically significant if $|Z| > 1.96$. Three proposed methods, i.e., Hist-MP, Hist-GLCM and Hist-Gabor, are compared together from the McNemars test point of view. The obtained results are reported in Table 7. As seen from the obtained results, the Hist-MP method is preferred than other methods while Hist-GLCM is located in the last rank.

In addition, the performance of the proposed methods are evaluated compared to some widely used and some state-of-the-art facial emotion recognition methods such as LLE [18], Isomap [17], Morphmap [19], Gabor [32] and LDP [20]. The comparison results are reported in Table 8. The results obtained with m=100 are represented for the proposed methods in this table. As seen from the obtained results, among the competitors, i.e., LLE, Isomap, Morphmap, Gabor and LDP, the LDP method has the superior performance. Even, it can provide better results compared to Hist-GLCM. But, two other proposed methods (Hist-MP and Hist-Gabor) provide the best recognition results.

The proposed framework provides efficient features for facial emotion recognition. The extracted features contain shape and texture information of the main components of face, which are efficient in expression recognition. The extracted features increase discrimination between similar emotions, and so, recognize difference between various facial expressions with more accuracy.

IV. CONCLUSION

The shape and textural characteristics of face image extracted by GLCM, Gabor and morphological filters are used for emotion recognition in this work. GLCM provides frequencies of co-occurring intensities of pixels. MP results in a multi-scale decomposition of image containing shape features with different sizes corresponding to shape and size of the considered structure element. Gabor filters by providing image information in various scales and directions, efficiently extract textural features of face components. The proposed method contains five steps: 1- extraction of face components, 2-shape and texture feature extrication, 3- histogram computing of spatial features, 4- fusion of the computed histogram features, and 5- classification by using the nearest neighbor classifier. The histogram function is locally applied to the main component of filtered face image to obtain an efficient estimate of the distributed texture features. By considering the main components of face image such as eyes, nose and mouth instead of entire face image, the redundant information which may be mis-leading are removed. By locally applying the histogram function to the texture features extracted by GLCM, Gabor and morphological operators not only the features are reduced but also an appropriate estimate of texture features is obtained. According to the experimental results, the use of more number of histogram bins can provide better recognition accuracy. Generally, the proposed Hist-Gabor and Hist-MP methods provide higher classification accuracy compared to Hist-GLCM. Some group of emotions have high overlapping and discrimination of them is hard. The obtained results in the confusion matrices show confusing of sadness, fear and disgust emotions, and also confusing of happy and surprise emotions. Individual spatial feature extraction from the main components of face and applying the histogram operator to the shape and texture feature maps of the main components of the face image are among the main novelties of the proposed method, which improves the emotion recognition accuracy compared to other popular methods.

REFERENCES

[1] U. Thonse, R. V. Behere, S. K. Praharaj, P. S. V. N. Sharma, Facial emotion recognition, socio-occupational functioning and expressed emotions in schizophrenia versus bipolar disorder, Psychiatry Research, Available online 27 March 2018.

[2] D. K. Jain, P. Shamsolmoali, P. Sehdev, Extended deep neural network for facial emotion recognition, Pattern Recognition Letters, 120 (2019) 69-74.

[3] M. Z. Uddin, M. M. Hassan, A. Almogren, M. Zuair, G. Fortino, J. Torresen, A facial expression recognition system using robust face features from depth videos and deep learning, Computers & Electrical Engineering, 63 (2017) 114-125.

[4] D. Torres-Boza, M. C. Oveneke, F. Wang, D. Jiang, W. Verhelst, H. Sahli, Hierarchical sparse coding framework for speech emotion recognition, Speech Communication, 99 (2018) 80-89.

[5] A. Uribe, A. Gómez, M. Bastidas, O. L. Quintero, D. Campo, "A novel emotion recognition technique from voiced-speech," 2017 IEEE 3rd Colombian Conference on Automatic Control (CCAC), Cartagena (2017) 1-4.

[6] K. N. Minhad, S. H. M. Ali, M. B. I. Reaz, Happy-anger emotions classifications from electrocardiogram signal for automobile driving safety and awareness, Journal of Transport & Health, 7 (Part A) (2017) 75-89.

[7] S. Vaid, P. Singh, "De-noising of EEG signal for emotion recognition," 2015 International Conference on Signal Processing, Computing and Control (ISPCC), Waknaghat (2015) 159-162.

[8] M. Z. Uddin, M. M. Hassan, A. Almogren, A. Alamri, M. Alrubaian, G. Fortino, "Facial Expression Recognition Utilizing Local Direction-Based Robust Features and Deep Belief Network," IEEE Access, 5 (2017) 4525-4536.

[9] G. Caridakis, K. Karpouzis, S. Kollias, User and context adaptive neural networks for emotion recognition, Neurocomputing 71 (2008) 2553– 2562.

[10] X. Huang, J. Kortelainen, G. Zhao, X. Li, A. Moilanen, T. Seppänen, M. Pietikäinen, Multi-modal emotion analysis from facial expressions and electroencephalogram, Computer Vision and Image Understanding, 147 (2016) 114-124.

[11] M. A. Nicolaou, H. Gunes, M. Pantic, Output-associative RVM regression for dimensional and continuous emotion prediction, Image and Vision Computing, 30 (3) (2012) 186-196.

[12] M. Song, M. You, N., C. Li, Chen, A robust multimodal approach for emotion recognition. Neurocomputing, 71 (2008) 1913-1920.

[13] P. Ekman, W.V. Friesen, Facial action coding system: A technique for the measurement of facial movement, palo alto. Consulting Psychologists Press, CA (1988); P.C. Ellsworth, C.A. Smith, From appraisal to emotion: Differences among unpleasant feelings, Motivation and Emotion 12 (1978) 271–302.

[14] Z. Li, J.i. Imai, M. Kaneko, Facial-component-based bag of words and phog descriptor for facial expression recognition, IEEE International Conference on Systems, Man and Cybernetics (SMC 2009) (2009) 1353–1358.

[15] M. Turk, A. Pentland, Face recognition using eigen faces, IEEE Conference on Computer Vision and Pattern Recognition, Maui, HI, USA (1991).

[16] T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: Application to face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, 28 (2006) 2037–2041.

[17] J.B. Tenenbaum, V. de Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science, 290 (5500) (2000) 2319–2323.

[18] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding', Science, 290 (5500) (2000) 2323–2326.
[19] R. Kapoor, R. Gupta, Morphological mapping for non-linear dimensionality reduction, IET Computer Vision, 9 (2) (2015) 226-233.
[20] T. Jabid, M. H. Kabir, O. Chae, Robust facial expression recognition based on local directional pattern, ETRI journal, 32 (5) (2010) 784–794.
[21] Y. Yoshitomi, T. Asada, R. Kato, M. Tabuse, Facial expression recognition using facial expression intensity characteristics of thermal image, J. Robot. Netw. Artif. Life, 2 (1) (2015) 5–8.
[22] S. Wang, B. Pan, H. Chen, Q. Ji, Thermal Augmented Expression Recognition, IEEE Transactions on Cybernetics, 48 (7) (2018) 2203-2214.
[23] M. Liu, S. Shan, R. Wang, X. Chen, Learning Expressionlets via Universal Manifold Model for Dynamic Facial Expression Recognition, IEEE Transactions on Image Processing, 25 (12) (2016) 5920-5932.
[24] Q. Mao, Q. Rao, Y. Yu, M. Dong, Hierarchical Bayesian Theme Models for Multipose Facial Expression Recognition, IEEE Transactions on Multimedia, 19 (4) (2017) 861-873.
[25] A. Majumder, L. Behera, V. K. Subramanian, Automatic Facial Expression Recognition System Using Deep Network-Based Data Fusion, IEEE Transactions on Cybernetics, 48 (1) (2018) 103-114.
[26] M. Z. Uddin, W. Khaksar, J. Torresen, Facial Expression Recognition Using Salient Features and Convolutional Neural Network, IEEE Access, 5 (2017) 26146-26161.
[27] D. K. Jain, Z. Zhang, K. Huang, Multi angle optimal pattern-based deep learning for automatic facial expression recognition, Pattern Recognition Letters, In Press.
[28] S. Wu, J. Xu, S. Zhu, H. Guo, A Deep Residual convolutional neural network for facial keypoint detection with missing labels, Signal Processing,Volume 144, 2018, Pages 384-391.
[29] L. Chen, M. Wu, M. Zhou, Z. Liu, J. She and K. Hirota, Dynamic Emotion Understanding in Human-Robot Interaction Based on Two-Layer Fuzzy SVR-TS Model, IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2019, In Press.
[30] G. Acampora, A. Vitiello, Interoperable neuro-fuzzy services for emotion-aware ambient intelligence, Neurocomputing, 122 (2013) 3-12.
[31] M. Imani, G. A. Montazer, GLCM features and fuzzy nearest neighbor classifier for emotion recognition from face, 2017 7th International Conference on Computer and Knowledge Engineering (ICCKE), Mashhad (2017) 8-13.
[32] M. S. Bartlett, G. Littlewort, I. Fasel, and J. R. Movellan, Real time face detection and facial expression recognition: Development and applications to human computer interaction, Computer Vision and Pattern Recognition Workshop (CVPRW '03), 5 (2003) 53.
[33] E. Sariyanidi, H. Gunes, A. Cavallaro, Learning Bases of Activity for Facial Expression Recognition, IEEE Transactions on Image Processing, 26 (4) (2017) 1965-1978.
[34] M. Imani, H. Ghassemian, GLCM, Gabor, and Morphology Profiles Fusion for Hyperspectral Image Classification, IEEE proceedings of the 24th Iranian Conference on Electrical Engineering (ICEE 2016), Shiraz, Iran (2016) 460-465.
[35] S. Gupta, R. K. Singh, Mathematical morphology based face segmentation and facial feature extraction for facial expression recognition, 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), Noida (2015) 691-695.
[36] M. Imani, H. Ghassemian, Edge Patch Image-Based Morphological Profiles for Classification of Multispectral and Hyperspectral Data, IET Image Processing, 11 (3) (2017) 164-172.
[37] M. Imani, G. A. Montazer, Face Recognition Using Morphological Profile and Feature Space Discriminant Analysis, IEEE proceedings of the 25th Iranian Conference on Electrical Engineering (ICEE 2017), Tehran, Iran (2017) 1729-1734.
[38] R. A. Marklein, M. W. Klinker, K. A. Drake, H. G. Polikowsky, E. C. Lessey-Morillon, S. R. Bauer, Morphological profiling using machine learning reveals emergent subpopulations of interferon-γ–stimulated mesenchymal stromal cells that predict immunosuppression, Cytotherapy, 21 (1) (2019) 17-31.
[39] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, Computer Vision and Pattern Recognition, (2001) 511-518.
[40] M. J. Lyons, S. Akemastu, M. Kamachi, J. Gyoba, Coding Facial Expressions with Gabor Wavelets, 3rd IEEE International Conference on Automatic Face and Gesture Recognition (1998) 200-205.
[41] F. Mirzapour, H. Ghassemian, Using GLCM and Gabor filters for classification of PAN images, 21st Iranian Conference on Electrical Engineering (ICEE 2013), Mashhad, Iran (2013).

HOW TO CITE THIS ARTICLE

M. Imani, Histogram Based Shape and Textural Characteristics for Facial Emotion Recognition, AUT J. Model. Simul., 51(1) (2019): 45 -56.

DOI: 10.22060/miscj.2019.15854.5144