Automatic Facial Emotion Recognition Method Based on Eye Region Changes

Mina Navraan
Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran
m.navran@modares.ac.ir
Nasrollah Moghadam Charkari*
Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran
charkari@modares.ac.ir
Muharram Mansoorizadeh
Faculty of Electrical and Computer Engineering, Bu-Ali Sina University, Hamadan, Iran
mansoorm@basu.ac.ir

Received: 19/Apr/2015 Revised: 19/Mar/2016 Accepted: 19/Apr/2016

Abstract

Emotion is expressed via facial muscle movements, speech, body and hand gestures, and various biological signals like heart beating. However, the most natural way that humans display emotion is facial expression. Facial expression recognition is a great challenge in the area of computer vision for the last two decades. This paper focuses on facial expression to identify seven universal human emotions i.e. anger, disgust, fear, happiness, sadness, surprise, and neu7tral. Unlike the majority of other approaches which use the whole face or interested regions of face, we restrict our facial emotion recognition (FER) method to analyze human emotional states based on eye region changes. The reason of using this region is that eye region is one of the most informative regions to represent facial expression. Furthermore, it leads to lower feature dimension as well as lower computational complexity. The facial expressions are described by appearance features obtained from texture encoded with Gabor filter and geometric features. The Support Vector Machine with RBF and poly-kernel functions is used for proper classification of different types of emotions. The Facial Expressions and Emotion Database (FG-Net), which contains spontaneous emotions and Cohn-Kanade(CK) Database with posed emotions have been used in experiments. The proposed method was trained on two databases separately and achieved the accuracy rate of 96.63% for spontaneous emotions recognition and 96.6% for posed expression recognition, respectively.

Keywords: Facial Emotion Recognition; Gabor Filter; Support Vector Machine (SVM); Eye Region.

1. Introduction

Emotion recognition methods can be classified into different categories along a number of dimensions: speech emotion recognition vs. facial emotion recognition; machine learning methods vs. statistic methods. Furthermore, facial expression method can also be classified based on input data to a dynamic image sequences or static image.

Scientists and psychologists classify the emotions of people in seven different expressions: Anger, Disgust, Fear, Happiness, Neutral, Sadness, and Surprise. According to their studies, human emotions could be recognized via different ways like human expression, appearance, biological signals etc. Among them, analyzing speech and facial expressions to recognize human emotion are most popular approaches. Speech analysis is based on the vocal information whereas facial expression analysis deals with the changes and movement of the facial muscles.

According to [4], speech contributes to the emotions of the speaker by 50% for the spoken word and by 38% for the voice; whereas, the facial expression is affected by 55%. There are some interesting algorithms which have been introduced to analyze speech or facial expressions in order to recognize human emotion. According to the previous studies, both the aforementioned approaches succeed in classifying the emotions. However, the facial expression approaches have revealed more precise results than the speech ones. Furthermore, some studies used hybrid methods (i.e. speech and facial expressions) which yield more accurate results [5]. Other type of hybrid methods for emotion recognition used two different models for facial and hand gesture recognition by separate classifier in advance. Then, the results of both classifiers are combined using a third classifier to recognize the emotion [51].

In general, emotion recognition is not a difficult task for the majority of human beings. However, it is a very challenging issue for computer based methods. The method that is designed for automatic analysis of facial expression is usually called Facial Expression Recognition method (FER). Some studies made their effort to detect facial expression based on action units (AUs) activation according to Facial Action Coding System (FACS) [24, 42, and 44].

Building a standardized database for FER method is very crucial since expressions can be posed (on purpose

* Corresponding Author

as a voluntary action) or spontaneous (unconsciously). Experimental results show that posed and spontaneous expressions vary widely in terms of configuration, their characteristics, temporal dynamics and timings [26].

Volitional facial movements originate in the motor cortex, whereas the involuntary facial actions, originate in the sub-cortical areas of the brain [27]. Therefore, while some facial movements might be easier to make voluntarily, many actions are done spontaneously. This may make it difficult to collect posed expression data of all possible facial movement. In this regard, many of the posed expressions that researchers have used in evaluation of their FER methods are highly overstated in compare with spontaneous expressions datasets.

Fig. 1. Show exaggeration in expressing fear expression in posed Cohn-Kanade dataset and its comparison with fear expression in spontaneous FG-Net dataset. The research community shifts their focus from posed to spontaneous expression recognition.

The main characteristic of FER method is that it should be effortless and efficient. It is also preferred to FER methods to be real-time which is especially important in both: human-computer interaction (HCI) and human-robot interaction applications. Other characteristics of an efficient FER methods are, the capability to work with video as well as images, the ability to simultaneously recognize spontaneous expressions and posed expression, robustness against the changes of lighting conditions and view angles, properly working in the presence of occlusions, invariant to facial hair, glasses, makeup etc.

More importantly, good FER methods are person independent and work on people from different cultures, different skin colors, and different ages (in particular, recognize expressions of both infants, adults and the elderly). Finally, they must be able to work with videos and images of different resolution.

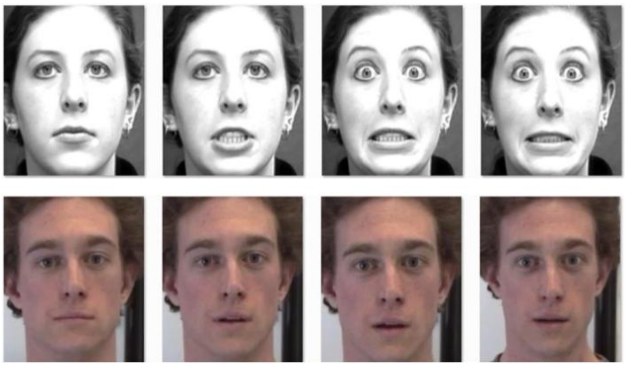


Fig. 1. Differences between the posed and spontaneous video frames in facial emotion recognition. First row: fear exaggeration in posed Cohn-Kanade dataset [55], Second row: fear expression in spontaneous FG-Net dataaset[56].

Feature extraction is a crucial step in facial expression recognition and largely defines the effectiveness of the performance. Therefore, selecting a suitable type of features for facial expression representation plays an essential role in modeling FER methods. Features can be classified into geometric and appearance (or texture) where each category has its own strength and weakness.

The final stage of any facial expression recognition method is the classification module based on the extracted features. Some works have been focused their works on proper classification by different classifiers. [15,16] studied static classifiers like the Naïve Bayes (NB), Tree Augmented Naïve Bayes (TAN), Stochastic Structure Search (SSS), and dynamic classifiers like Single Hidden Markov Models (HMM) and Multi-Level Hidden Markov Models (ML-HMM).

1.1 Motivation

According to the literature studies, the majority of facial expression recognition methods use Facial Action Coding System and action units (AUs) detection [52]. However, these methods generally suffer from time-intensive processes which result in high time complexity on video data. On the other hand, the importance of developing an approach to automatically differentiate between posed and spontaneous facial expression has been considered as one of the interesting issues within the last decade. In this regard, the main contributions of this work are summarized as follow;

It has been found that both eyes and mouth regions are parts of the face with the maximum amount of information in every facial expression [49]. So, we restrict the proposed method on extracting features from eye regions in facial expression recognition. This leads to a lower time complexity of each video frame with high accuracy rate. These regions are of great importance in differentiating between different facial expressions since the majority of facial activities particularly occur in upper face region.

The method is able to efficiently work on video and has the ability to recognize spontaneous and posed expression in low time complexity. To evaluate the method two different datasets (posed and spontaneous) have been employed in this paper. The Proposed FER method has been developed to process the video of facial emotion as well as to recognize the displayed actions in terms of seven basic emotions. In this regard, we have used both type of features (i.e. geometric and texture). We have employed spatial information of eyes for extracting geometric features whereas textural features are found by applying Gabor filter to eyes regions. Since features of eye region are extracted, feature space dimension will reduce dramatically. Furthermore, expression classifiers can be trained in much less time. Consequently, this leads to lower computational complexity where the emotional recognition rate is kept high in value. Finally, among different classification methods, we chose SVM because of its simplicity, flexibility and resistance to noise and outliers [42].

The rest of the paper is organized as follows: Section II reviews some related work. Our approach for FER method is presented in Section III and experimental results are reported in Section IV. Finally, Section V draws some conclusions.

2. Related Work

Numerous approaches have been presented to automatically analysis facial expressions from static image to dynamic image sequences. The early works have been summarized by [7, 8, 9, and 10]. Recent advances on automatic facial emotion recognition have been studied in [6]. Most of the existing FER methods employ various pattern recognition methods and emotions classification are based on 2D geometric and appearance facial features.

The typical procedure for almost all the emotion recognition method is depicted in Fig. 2. The first stage is face detection and tracking. It involves the process of locating face regions from the input data; align the face to a common coordinate method, and tracking the face region in every frames of video. The second stage is facial feature extraction and representation which is responsible for extracting and representing the facial variations caused by facial expressions. Finally, the last component is facial expression classification.

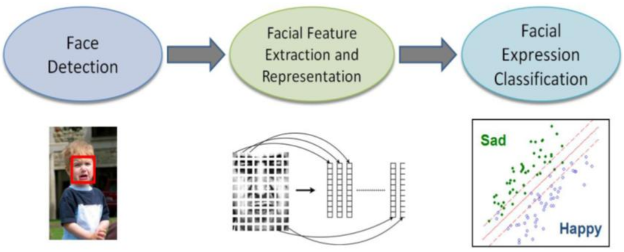


Fig. 2. Three main components of a typical facial expression recognition method [1].

2.1 Face Detection

Identifying the presence of faces on an image or frames of video and determining the locations and scales of them are the first step in facial emotion recognition. The accuracy of this stage is particularly significant in realistic condition.

A typical face detection algorithm performed the detection processes in the following steps. Given a set of training images acquired in a fixed pose (e.g. frontal, near-frontal, or profile view), histogram equalization or standardization is performed to dominate the effects of illumination. After this step, some face samples are extracted with knowledge based [28] or learning based methods [29, 30, and 31]. Here, the knowledge based methods model the face patterns by some definitive rules, such as facial components, face textures or skin color; the learning based methods model the face patterns by learning from a set of features with some discriminate functions [1]. With the extracted face patterns, the face detector method scans through the entire image to locate and detect the faces.

We have concentrated on the cascade based face detectors for its good performance. The AdaBoost based face detector by Viola and Jones is the most commonly used face detector in automatic face recognition and expression analysis [1].

The main distinctive idea of the Viola and Jones detector is to train a cascade classifier for haar-like rectangular features. The haar-like rectangular features can be efficiently computed with integral images [2], which facilitate the approach to gain real time detection approach. To further increase the detection speed while retaining the accuracy, AdaBoost was used to select the representative haar-like features [32]. In cascade based face detector, instead of training a single strong classifier, a number of weak classifiers are constructed and then combined into a cascade. The classifiers at the beginning of the cascade can efficiently reject the non-face regions, while the stronger classifiers later in the cascade simply need to classify the more face-like regions [1].

Several extensions have been made to Viola and Jones detector for detecting faces from different views [17, 18, and 32].

There are a few other face detection approaches in the recent literature, including the energy-based method that detects faces region and estimates the poses simultaneously [45], the face detectors using support vector machines (SVM) [46], the face detectors trained with only positive images [47], and the component-based face detector using Naive Bayes classifiers [48].

2.2 Feature Extraction

After face detection and tracking, the next step is to extract the representative and distinctive features of facial expression. Extracting effective features from the detected face image is crucial for successful facial expression recognition. Optimal feature extraction should be done in a way that minimizes within-class variations of expression while maximizing across-class variations. Features are generally divided into four groups:

Geometric features: Represent the shape and location of facial components or predefined facial feature points, which are extracted to form a feature vector to represent the face geometry [19, 20, 33, 34, 35]. Deformations between neutral state and current frame are parameters of facial expression [25].

The automatic and efficient detection and tracking of facial features point is still an open problem in many computer vision applications. This motivates the use of appearance (texture) based features for facial expression analysis.

Appearance features: Appearance based features measure the appearance changes which are mainly based on texture analysis. Typical appearance-based approaches use different image filters such as Gabor or linear filters, to extract feature vectors such as texture, correlation and gradient from whole face, specific regions, or regions of interest (ROI).

Gabor filters [50] have been found to be powerful in face expression analysis and widely used to extract the facial appearance changes as a set of multi scales and multi orientation coefficients for analyzing the texture. In problems with time and memory limitation another method called Local Binary Patterns (LBP) gains more popularity in facial texture analysis [38, 39, 40, and 41].

Fusion the features: Geometric feature capture macro-variation of facial structure while appearance-based approaches are capable of identifying local subtle changes [3]. By combining these two kinds of features the performance of facial emotion recognition method would be improved. Active Appearance Model (AAM) is a combination of appearance and texture information to construct a parameterized model of facial features. Relationships between AAM displacement and the image difference would be analyzed for facial expression detection. AAM has been widely used for facial feature extraction.

Temporal features: The main idea of extracting this kind of features is achieving the temporal information about the movements of facial components in video frames for facial expression [22].

2.3 Classification

The final stage of the FER method is based on machine learning theory; precisely it is the classification module. The input to the classifier is a set of features which has been retrieved from face region in feature extraction stage. The feature vector is formed to describe the facial expression. The first part of Classification module is training. The training set of classifier consists of labeled data. Once the classifier is trained, it can recognize input images by assigning them a particular expression class label.

All approaches for classification of facial expression can be divided into two groups: frame based recognition which only relies on a single frame; image sequence based approaches exploited the temporal behaviors of facial expressions.

In different researches, various classifiers have been applied such as , Neural Network, Bayesian Network (BN), Support Vector Machine (SVM), rule-based classifiers, and Hidden Markov Models (HMM) [21]. Some studies like [24, 42, and 44] in facial expression analysis, made their effort to classify action units (AU). Bartlett. et al. [42, 44] studied AU activations based on Gabor filter responses on whole face region. Other studies like [14, 23, 26, and 43] classified each emotional state based on the extracted features.

In [43] Gabor filters with different frequencies and orientations were applied on face region, and SVM classification method used for recognizing four basic emotion. Zhan. et al. [26] proposed a method to recognize seven basic spontaneous expressions using FG-Net dataset. For feature selection, Gabor filters with different frequencies and orientations were applied only to a set of facial landmark positions. [23] Used Adaboost to choose a subset of feature extracted from Gabor filters. The SVMs are then used for classification. Developed from statistical learning theory, is a widely used classifier for facial expression classification.

3. Proposed Method

In this section, we describe the details of our facial expression recognition method. The section is divided into three subsections. In Section A, we describe how face image are automatically processed for feature extraction. In section B, those features which have efficiently been employed to extract human emotions will be discussed. Finally, in section C, we adopt SVM for classifying anger, disgust, fear, happy sad, and surprise expressions. The main components of the proposed FER method are shown in Fig. 3.

3.1 Face Detection

In the first stage of FER method, after converting video to image sequences, facial regions are detected and tracked. For each frame of the video, an approximate region of a face is found by the Viola–Jones face detector. In order to perform facial emotion recognition, some processes are done in preprocessing step. Firstly, face region on each frame is extracted by means of Viola–Jones algorithm from the background and resized to 860*860 pixel. Second, key frame detector is performed to extract some video frames. Third, tracking process is conducted. Forth, Viola–Jones algorithm is employed for detecting the eyes region on key frames. Fig. 4. Show the extracted face and the detected eye regions, respectively.

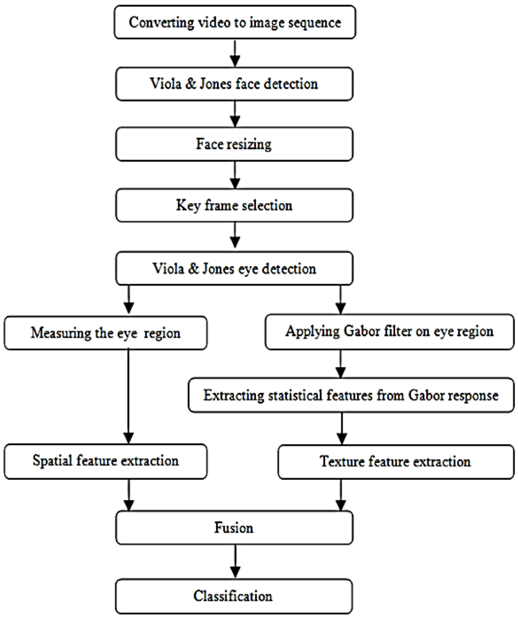


Fig. 3. overview of the proposed FER method

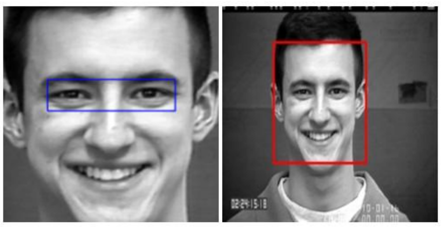


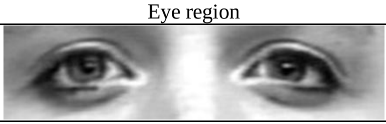
Fig. 4. output of Viola–Jones detector for detecting face and eyes regions in our work.

3.2 Feature Vector

The detected regions are used to extract two types of features: geometric and texture. Previous facial expression recognition methods typically used either geometric or appearance (texture) features [10]. We have employed both features since they provide complementary information. Two features are extracted separately and combined in the classification stage.

To define geometric features, after detecting eyes regions, their sizes are found. The changes in size of eyes regions form geometric feature and can be coped well with the variations in skin patterns or dermatoglyphics. One example of size changes in eye region during disgust expression in our work is shown in TABLE I.

Table 1. one example of size changes in eye region during disgust expression from Cohn dataset in our work.

#	Eye region	height	width
1		137	560
2		139	568
3		141	575
4		140	570
5		138	560
6		144	585
7		144	587
8		143	582
9		143	583

To extract texture features, Gabor filter has been applied. We convolve the extracted eye regions with Gabor filter banks with 1 spatial frequency that properly provides edges, Wrinkles and furrows, and 9 different orientations from 0 to 180 degrees with a 20degree steps. An example of Gabor response in one of the experiments is shown in Fig. 5.

Fig. 5. gabor response on eye region in our work

The Gabor responses from the whole image as features lead to huge dimensionality problem. To cope with these problem, different approaches like feature selection and employing Gabor filter on Regions-of-Interest (ROIs) have been suggested [11, 12, 36 and 37]. In this work, we have found the response of Gabor filter for every detected eye region.

To decrease dimensionality of feature space we use two strategies in Gabor feature extraction:

1) We restrict Gabor filter bank to one spatial frequency which properly provides edges, Wrinkles and furrows instead of using different spatial frequencies.

The reason is that, all the spatial frequencies don't provide useful information on image. The low spatial frequencies result in blurred edges in Gabor response where the high spatial frequencies lead to some broken and not continuous edges of Gabor response. Examples of Gabor filter responses with low and high spatial frequencies are shown in Fig. 6.

By applying Gabor filter banks with 9 different spatial frequencies from 1/2 to 1/32 in units, and 9 different orientations from 0 to 180 degrees with a 20 degree steps from the eye region with the size of 144 * 585 pixels, dimensionality becomes huge (9 * 9 * 144 * 585 = 6,823,440). By restricting the filter bank to 1 spatial frequency and 9 orientations, the dimensionality of the features reduce to 1 * 9 * 144 * 585 = 758,160 whilst better response with the most useful information would be obtained.

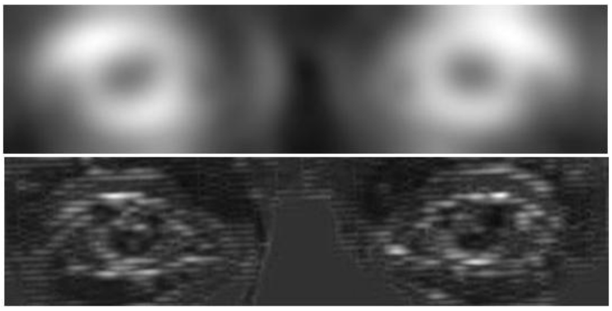


Fig. 6. gabor response with low spatial ferequncy in above picture and gabor response with high spatial ferequncy in below picture.

2) After getting the response of Gabor filter, we extract some statistical information such as histogram, mean, and standard deviation from every response of Gabor filter. This process reduces the dimensionality of the features dramatically.

Accordingly, histogram is defined as the number of pixels of each brightness level in the image. Mean is defined as the sum of the pixel values in Gabor response divided by the number of values. Finally, standard deviation is used to measure the amount of variation from the average value.

Horizontal & vertical profile: Data in a Gabor response matrix can be profiled. We get the vertical, horizontal or arbitrary profiles of the matrix data. The columns summation is defined as vertical profile. Similarly, horizontal profile is the rows summation.

3.3 Classification

In the final step, classification is employed. Facial emotion recognition requires classifier training with a set of images with particular emotions displayed. In the proposed FER method, the Support Vector Machine (SVM) is used as a classifier. We choose SVM because of its simplicity, flexibility and resistance to noise and outliers and having the advantage of solving non-linear, or high dimensional classification problem. The SVM is a classifier which receives labeled training data and transforms it into higher dimensional feature space. The SVM classification method computes separating hyper plane between classes. SVM determines the best separation between classes with respect to margin maximization. Regarding the kernels tested, we compare the most commonly used ones in the literature i.e. polynomial and rbf.

4. Experimental Results

There is still no standard method for evaluation of automatic FER methods. However, the majority of studies have reported the performance of their method on one or more available emotional face datasets. Many works have reported the results of their method on Cohn-kanade, FG-Net datasets, or both.

To investigate the efficiency of the proposed FER method, two facial expression datasets are used in this paper. The first dataset is the FG-Net Facial Expression which consists of MPEG video files with spontaneous emotions recorded. It contains some expression examples gathered from 18 different subjects (9 female and 9 male). Facial expressions of subjects are captured during watching videos. The dataset formed from FG-Net dataset consists of 1542 images of seven states, neutral and emotional (anger, disgust, fear, happy, sadness, and surprise). Accordingly, 222 frames for neutral, 221 frames for anger expression, 222 frames for disgust expression, 220 frames for fear expression , 218 frames for happiness, 220 frames for sadness, and 219 frames for surprise expression are used.

Second dataset is the Cohn-Kanade Facial Expression Dataset which contains image sequences displayed by 97 posers. The sequence displays the emotion from the start (neutral state) to the peak. The dataset contains 1532 images divided into seven classes: neutral, anger, disgust, fear, happiness, sadness, and surprise. Accordingly, 220 frames for neutral, 218 frames for anger, 218 frames for disgust, 218 frames for fear, 220 frames of happiness, 220 frames of sadness, and 218 frames for surprise expressions are used.

TABLE II. gives an overview of the existing methods where Gabor filter response is part of their extracted features, for expression analyzes.

Our work consists of two set of experiments: binary classification and multiclass classification. Evaluation criteria for classification problem are:

F-Measure: Combination of precision and recall by their harmonic mean. It varies between zero and one; when the value is closed to one it indicates a better performance of classification (1).

$$\mathrm{F-Measure} = \frac{2 \times \mathrm{precision} \times \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}} \quad (1)$$

False positive rate: It is also known as the false alarm ratio refers to the rate of occurrence of positive test results in samples known to be negative for which an individual is being tested (2). TN is the number of negative results which have detected correctly as negative results.

$$\text{false positive rate} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} \quad (2)$$

Table 2. Comparison of some facial expression analyses

Reference	Database	Feature type	Classification	Performance
2001 Tian.[13]	Cohn-Kanade	Permanent features: Gabor response on nasolabial region, nasal root, and eye corner transient features: canny edge detection	ANN	Recognition of upper face action units: 96.4% Recognition of lower face action units: 96.7% Average to independent databases:93.3%
2003 Bartlett.[23]	Cohn-Kanade	Gabor filter responses selected by Adaboost for each expression	SVM	Across 7 prototypic expression: AdaSVM+RBF: 90.7%
2008 Kotsia.[14]	Cohn-Kanade	Gabor filter output vectors on whole face have been concatenated to form a new long feature vector	SVM&MLP	Across 6 prototypic expression: Cohn : 91.6%,
2008 Zhan.[26]	FG-Net	Gabor filters with different frequencies and orientations are applied only to a set of facial landmark positions	SVM	Accuracy rate for 7 expressions : 82%
2010 Wu.[53]	Cohn-Kanade	Gabor motion energy filters (GME) output vectors on whole face	SVM	Average ROC over 6 expressions: on onset sequences classification: 78.56% on apex sequences classification: 97.81%
2014 L.Zhang.[54]	Cohn-Kanade	Gabor filters with different frequencies and orientations are applied on face region	SVM	Across 7 prototypic expression Cohn : 95.3%,
This work	Cohn-Kanade FG-Net	Gabor: statistical information from every response of Gabor filter Geometry: size of eye region	SVM	Average Measure for classifying 7 prototypic expression: Cohn: 96.6% FG-Net : 96.63%

4.1 Binary Classification

In the first experiment, we have conducted our method to detect emotional state from neutral face. So, we have used binary classification to determine the efficiency of the method.

In each video, the initial frames start from neutral state and the remaining frames rise to a complete emotional states. The results of the experiment are shown in TABLE III, IV, V, VI, VII, and VIII. In binary classification, the best result on posed Cohn-Kanade and spontaneous FG-Net datasets found for disgust expression with 98.64% and 99.22% correctly classified instances, respectively. The worst result on posed Cohn-Kanade dataset is for anger expression with 97.09% correctly classified instances. Similarly, on spontaneous FG-Net, the worst result is for surprise expression with 97.94% correctly classified instances.

Table 3. SVM classification of neutral & anger expression with poly & rbf kernels

	Kernel	FP Rate	%F-Measure	% average of Correctly Classified	dataset
neutral	Poly	0.05	96.67	96.68	Cohn-Kanade
anger		0.017	96.66		
average		0.033	96.67		
neutral	RBF	0.058	97.2	97.09	
anger		0	97		
average		0.029	97.1		
neutral	Poly	0.031	98.5	98.44	FG-Net
anger		0	98.4		
average		0.016	98.4		
neutral	RBF	0.023	98.9	98.83	
anger		0	98.8		
average		0.012	98.8		

Table 4. SVM classification of neutral & disgust expression with poly & rbf kernel

	Kernel	FP Rate	%F-Measure	% average of Correctly Classified	dataset
neutral	Poly	0.037	98.2	98.18	Cohn-Kanade
disgust		0	98.1		
average		0.019	98.2		
neutral	RBF	0.028	98.7	98.64	
disgust		0	98.6		
average		0.014	98.6		
neutral	Poly	0.031	98.5	98.45	FG-Net
disgust		0	98.4		
average		0.016	98.4		
neutral	RBF	0.016	99.2	99.22	
disgust		0	99.2		
average		0.008	99.2		

Table 5. SVM classification of neutral & fear expression with poly & rbf kernel

	Kernel	FP Rate	%F-Measure	% average of Correctly Classified	dataset
neutral	Poly	0.035	98.2	98.23	Cohn-Kanade
fear		0	98.2		
average		0.017	98.2		
neutral	RBF	0.044	97.4	97.34	
fear		0.009	97.3		
average		0.026	97.3		
neutral	Poly	0.036	97.9	97.84	FG-Net
fear		0.007	97.8		
average		0.021	97.8		
neutral	RBF	0.029	98.2	98.2	
fear		0.007	98.2		
average		0.018	98.2		

Table 6. SVM classification of neutral & happiness expression with poly & rbf kernel

	Kernel	FP Rate	%F-Measure	% average of Correctly Classified	dataset
neutral	Poly	0.035	98.2	98.23	Cohn-Kanade
happy		0	98.2		
average		0.017	998.2		
neutral	RBF	0.043	97.8	97.79	
happy		0	97.8		
average		0.021	97.8		
neutral	Poly	0.026	98.7	98.68	FG-Net
happy		0	98.7		
average		0.013	98.7		
neutral	RBF	0.033	98.4	98.35	
happy		0	98.3		
average		0.016	98.3		

Table 7. SVM classification of neutral & sadness expression with poly & rbf kernel

	Kernel	FP Rate	%F-Measure	% average of Correctly Classified	dataset
neutral	Poly	0.046	97.3	97.25	Cohn-K
sad		0.009	97.2		
average		0.028	97.2		
neutral	RBF	0.056	96.4	96.33	
sad		0.018	96.2		
average		0.037	96.3		
neutral	Poly	0.014	99	99.06	FG-Net
sad		0.005	99.1		
average		0.009	99.1		
neutral	RBF	0.009	98.8	98.83	
sad		0.015	98.9		
average		0.012	98.8		

Table 8. SVM classification of neutral & surprise expression with poly & rbf kernel

	Kernel	FP Rate	%F-Measure	% average of Correctly Classified	dataset
neutral	Poly	0.046	97.8	97.73	Cohn-Kanade
surprise		0	97.7		
average		0.023	97.7		
neutral	RBF	0.046	97.8	97.73	
surprise		0	97.7		
average		0.023	97.7		
neutral	Poly	0.049	97.2	97.12	FG-Net
surprise		0.008	97.1		
average		0.029	97.1		
neutral	RBF	0.033	98	97.94	
surprise		0.008	97.9		
average		0.02	97.9		

4.2 Multiclass Classification

Some studies like [24, 42, and 44] in facial expression analysis, made their effort to classify action units (AU). Other studies like [14, 23, 26, and 43] classified each emotional state based on the extracted features.

In this experiment, our objective is to detect and to distinguish anger, disgust, fear, happiness, sadness, surprise and neutral frames from each other. So, we face with 7- label classification problem. We have studied two popular strategies to solve such classifications problems; One-vs.-the-Rest and One-vs.-One.

One-Vs.-The-Rest: The strategy consists in fitting one classifier per class. For each classifier, the class is fitted against all the other classes. In addition to its computational efficiency, another advantage of this approach is its interpretability. In this strategy, for each class, we have one and only one classifier. So this feature makes it possible to gain the knowledge about the class by inspecting the corresponding classifier.

One-Vs.-One: It constructs one classifier per pair of classes. The class which receives the most votes at prediction step would be selected. This method is beneficial for algorithms such as kernel ones which don't scale well with n_samples. This is because each individual learning problem only involves a small subset of data whereas, with one-vs.-the-rest, the complete dataset is used n_classes times.

We have employed One-Vs-One scheme for solving our multi-classification problem. The results of the experiments are shown in TABLE IX, X.

Table 9. SVM classification of neutral, anger, disgust, fear, happiness, sadness & surprise expression with poly & rbf kernel in cohn dataset.

	Kernel	FP Rate	%F-Measure	% average of Correctly Classified	dataset
neutral	**Poly**	0.026	86.9	95.76	**Cohn-Kanade**
anger		0.002	98.9		
disgust		0.002	98.4		
fear		0.006	96.3		
happy		0.004	96.6		
sad		0.006	96.8		
surprise		0.004	96.8		
average		0.007	95.8		
neutral	**RBF**	0.02	88.9	96.6	
anger		0.001	99.5		
disgust		0.002	98.4		
fear		0.004	97.2		
happy		0.004	97.5		
sad		0.005	97.5		
surprise		0.005	97.2		
average		0.006	96.6		

Table 10. SVM classification of neutral, anger, disgust, fear, happiness, sadness & surprise expression with poly & rbf kernel in fg-net dataset.

	Kernel	FP Rate	%F-Measure	% average of Correctly Classified	dataset
neutral	**Poly**	0.015	92	95.91	FG-Net
anger		0.005	97.8		
disgust		0.002	97.7		
fear		0.011	94.1		
happy		0.002	98.4		
sad		0.004	97.5		
surprise		0.008	94		
average		0.007	95.9		
neutral	**RBF**	0.012	92.8	96.63	
anger		0.005	98		
disgust		0.002	98.2		
fear		0.008	95.7		
happy		0.004	98.2		
sad		0.002	97.9		
surprise		0.007	95.7		
average		0.006	96.6		

To validate the proposed FER method we use 10-fold cross validation method. In 10-fold cross validation, the original sample is randomly partitioned into 10 equal size subsamples. Of the 10 subsamples, a single subsample is retained as the validation data for testing the model, and the remaining 9 subsamples are used as training data. The cross-validation process is then repeated 10 times, with each of the 10 subsamples used exactly once as the validation data. The 10 results from the folds would then be averaged to produce a single estimation.

The confusion matrices in TABLE XI, XII indicates the performance of 10-fold cross validation on the posed and spontaneous datasets, respectively. The rows represent the emotion that was intended by the subject and the columns represent the emotion determined by the classifier. The total number of images that have been classified for each intended emotion is shown on the last column of the table.

For Cohn-kanade dataset, the overall percentages of correctly classified emotions were: 89.54% for intended neutral, 99.54% for anger, 97.71% for disgust, 96.79% for fear, 97.27% for happiness, 98.18% for sadness, and 97.25% for surprise.

Similarly, for FG-Net dataset, the overall percentages of correctly classified emotions were: 92.79% for intended neutral, 98.64% for anger, 97.75% for disgust, 96.36% for fear, 98.62% for happiness, 96.82% for sadness, and 95.43% for surprise.

Our proposed FER method has some drawbacks in spatial feature extraction. Despite the good performance of Viola Jones eye detection algorithm, this method doesn't provide the changes in the size of eye regions as expected. For example, during disgust expression the height of eye region decreased. However, as we have shown in TABLE I. viola Jones doesn't show height changes as expected.

5. Conclusion

The research community shifts their focus from posed to spontaneous expression recognition. Since spontaneous expression are more natural in compare with posed expression. Recognizing expressions which contain exaggeration in expressing facial emotion aren't useful in the real world applications. In future robust spontaneous expression recognizers will be developed and deployed in real-time methods and used in building many real-world applications especially in HCI applications.

The fully automated nature of the FER method allows us to perform facial expression analysis in many applications of HCI. By creating computer methods that can understand emotion, we enhance the communications that exists between humans and computers.

In this paper we introduced a fully automatic facial expression recognition method. Contrary to other approaches in emotional recognition which employ the whole face region or some interesting regions of faces, we restrict our FER method to analyze human emotional states based on eye region changes. We showed that how this region plays significant role for further analysis of facial expression process. Since the features of eye region are extracted, the size of feature dimension reduces dramatically. Therefore, expression classifiers can be trained in much less time. Consequently, this leads to lower computational complexity where the emotional recognition rate is kept high in value. We evaluate the proposed FER method on posed (Cohn-Kanade) and spontaneous (FG-Net) datasets. For feature extraction we use a fusion based approach. The classification rate on spontaneous dataset increased to 96.63%. On posed dataset the accuracy of FER method is 96.6%. The results are comparable to the existing facial expression recognition methods.

Proper recognition rate of the proposed FER method is because of extracting useful informative features during expressing emotions. We achieve this goal by restricting our method to analyze eye region changes during expressing facial emotion.

Table 11. The confusion matrices of 10-fold cross validation for multi-class expression classifier when tested on cohn-k dataset.

CLASSIFIED AS →	Neutral	Anger	disgust	Fear	Happy	Sad	Surprise	total
Intended Neutral	197	1	2	4	5	7	4	220
Intended Anger	1	217	0	0	0	0	0	218
Intended Disgust	5	0	213	0	0	0	0	218
Intended Fear	5	0	0	211	0	0	2	218
Intended Happy	6	0	0	0	214	0	0	220
Intended Sad	4	0	0	0	0	216	0	220
Intended Surprise	5	0	0	1	0	0	212	218

Table 12. The confusion matrices of 10-fold cross validation for multi-class expression classifier when tested on fg_net dataset.

CLASSIFIED AS →	Neutral	Anger	disgust	Fear	Happy	Sad	Surprise	total
Intended Neutral	206	4	1	1	5	1	4	222
Intended Anger	3	218	0	0	0	0	0	221
Intended Disgust	3	0	217	1	0	0	1	222
Intended Fear	3	0	1	212	0	0	4	220
Intended Happy	2	0	0	1	215	0	0	218
Intended Sad	3	1	1	2	0	213	0	220
Intended Surprise	2	1	0	6	0	1	209	219

References

[1] Kaimin Yu, “Toward Realistic Facial Expression Recognition“, Doctoral dissertation, School of Information Technologies at The University of Sydney, 2013.

[2] P.Viola, MJ. Jones, “Robust Real-Time Face Detection“, Interna-tional Journal of Computer Vision, 57(2):137{154, May 2004.

[3] Li,Peiyao, “Adaptive feature extraction and selection for robust facial expression“, Master of Engineering by Research thesis, University of Wollongong, School of Electrical, Computer and Telecommunications Engineering,UniversityofWollongong,2010.http://ro.uow.edu.au/theses/3268

[4] A. Mehrabian, “Communication without words“. Psychology today, vol.2, no.4, pp.53-56, 1968.

[5] M. Mansoorizadeh, N. Moghaddam Charkari, “Multimodal information fusion application to humanemotion recognition from face and speech“ Springer Science, Multimed Tools Appl 2010.

[6] Z. Zeng, M. Pantic, G.I. Rosman, and T.S. Huang. “A Survey of Af-fection Recognition Methods: Audio, Visual, and Spontaneous Expressions“.IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(1):39{58,January 2009.

[7] A.Samal, A. Lyengar. “Automatic recognition and analysis of human faces and facial expressions: a survey“. Pattern Recognition, 25(1):65{77,January 1992.

[8] M.Pantic, J.M. Rothkrantz. “ Automatic Analysis of Facial Ex pressions: The State of the Art“. IEEE Transaction on Pattern Analysis and Machine I Intelligence, 22(12):1424{1445, December 2000.

[9] B.Fasel, J.Luettin. “ Automatic facial expression analysis: a Survey“,Pattern Recognition, 36(1):259{275, January 2003.

[10] Y.Tian, Takeo Kanade,F. Cohn, “Facial Expression Analysis.In Handbook of Face Recognition“, Springer New York, 2005.

[11] Y.Tian, T. Kanade, and J. Cohn. “Eye-state action unit detection by Gabor wavelets“,In Proceedings of International Conference on Multi-modalInterfaces, pages 143{150, 2000.

[12] Y. Tian, T. Kanade, and J. Cohn. “Evaluation of gabor-wavelet-based facialaction unit recognition in image sequences of increasing complexity“, In IEEEInternational Conference on Automatic Face and Gesture Recognition, page 229,2002.

[13] Y. Tian, T. Kanade and J. Cohn, “Recognizing Action Units for Facial Expression Analysis,” IEEE Trans. Pattern Analysis and Machine Intelligence,vol. 23, no. 2, pp. 97–115, 2001.

[14] I. Kotsia, I. Buciu and I. Pitas, “An Analysis of Facial Expression Recognition under Partial Facial Image Occlusion,” Image and Vision Computing, vol. 26, no. 7, pp. 1052-1067, July 2008.

[15] I. Cohn, N. Sebe, F. Cozman, M. Cirelo, and T. Huang, “Learning Bayesian Network Classifiers for Facial Expression Recognition Using Both Labeled and Unlabeled Data,” Proc. IEEE Conf. Computer Vision and Pattern Recognition(CVPR),vol. 1, pp. I-595-I-604, 2003.

[16] I. Cohn, N. Sebe, A. Garg, L.S. Chen, and T.S. Huang, “Facial Expression Recognition From Video Sequences: Temporal and Static Modeling”, Computer Vision and Image Understanding,vol. 91, pp. 160-187, 2003.

[17] M. Jones, P. Viola. "Fast Multi-view Face Detection", Technical report, Mitsubishi Electric Research Laboratories, 2003.
[18] J. Wu, S.C. Brubaker, M. Mullin, and J. Rehg. "Fast asymmetric learning for cascade face detection", IEEE Transaction on Pattern Analysis and Machine Intelligence, 30(3):369{382, 2008.
[19] M. Pantic, M. Rothkrantz. "Facial action recognition for facial expressionanalysis from static face images", IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 34(3):1449{1461, June 2004.
[20] M. Pantic,I. Patras, "Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences",IEEETransactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 36(2):433{449, April 2006.
[21] G. Donato, M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski. "Classifying facial actions", IEEE Transactions on Pattern Analysis and Machine Intelligence,21(10):974{989, 1999.
[22] S.Koelstra, M.Pantic, "A Dynamic Texture-Based Approach to Recognition of Facial Actions and Their Temporal Models",IEEE TRANSACTIONS on Pattern Analysis and Machine Intelligence, VOL. 32, NO. 11, NOVEMBER 2010.
[23] M.S. Bartlett, G. Littlewort, I. Fasel, and R. Movellan, "Real Time Face Detection and Facial Expression Recognition: Development and Application to Human Computer Interaction," Proc. CVPR Workshop on Computer Vision and Pattern Recognition for Human-Computer Interaction,vol. 5, 2003.
[24] A.Savran, B. Sankur, M. Taha Bilge, "Regression-based intensity estimation of facial action units", Image and Vision Computing 2012.
[25] M. Pantic and L. Rothkrantz, "Expert System for Automatic Analysis of Facial Expression", Image and Vision Computing J., Vol. 18, No. 11, p. 881-905,2000.
[26] C. Zhan, W. Li, P. Ogunbona,F. Safaei, A Real-Time Facial Expression Recognition Systemfor Online Games, International Journal of Computer Games Technology, Volume 2008.
[27] Ekman, P. and Friesen, W. V, 'Detecting deception from body or face.'Journal of Nonverb Behav,(1974).
[28] D.Reisfeld and Y.Yeshurun,"Robust detection of facial features by generalized symmetry", in Proceeding of 11th IAPR International Conference on Pattern Recognition, PP.117-120, 1992.
[29] G.A. Ramirez and O.Fuentes,"Face detection using combination of classifiers", in Proceeding of The Canadian Conference on Computer and Robot Vision, PP.610-615,2005.
[30] P.Zhang,"A video-based face detection and recognition system using cascade face verification modules",in Proceeding of 37th IEEE Applied Imagery Pattern Recognition Workshop, PP.1-8, 2008.
[31] B.Yipp,"Face and eye rectification in video conference using artificial neutral network", in IEEE International Conference on Multimedia and Expo,PP.690-693,2005.
[32] [32] P.Viola, M.J.Jones, and D.Snow,"Dtecting pedestrians using patterns of motion and appearance", in Proceeding of Ninth IEEE International Conrefence on Computer Vision, Vol.2,PP.734-741, 2002.
[33] Z. Zhang, M.J. Lyons, M. Schuster, and S. Akamatus. "Comparison betweengeometry-based and Gabor-wavelets-based facial expression recognition usingmulti-layer perceptron". In IEEE International Conference on Automatic Face& Gesture Recognition, 1998.
[34] R.E. Kaliouby and P. Robinson. "Real-time inference of complex mental states from facial expressions and head gestures". In IEEE CVPR Workshop on Real-time Vision for HumanCComputer Interaction, 2004.
[35] Michel Valstar, I. Patras, and M. Pantic. "Facial Action Unit Detection using Probabilistic Actively Learned Support Vector Machines on Tracked Facial Point". In IEEE Conference on Computer Vision and Pattern Recognition, page 76, 2005.
[36] G. Ford. "Tutorial on gabor filters". Technical report, Machine Perception Lab, Institute of Neural Computation, 2002.
[37] Ying-Li Tian, T. Kanade, and J. Cohn. "Eye-state action unit detection by Gabor wavelets". In Proceedings of International Conference on Multi-modal Interfaces, pages 143{150, 2000.
[38] T. Ojala, M. Pietikainen, and D. Harwood. "A comparative study of texture measures with classification based on featured distribution". Pattern Recognition, 29(1):51{59, 1996.
[39] T. Ahonen, A. Hadid, and M. Pietikainen. "Face recognition with local binary Patterns". In European Conference on Computer Vision, 2004.
[40] X. Feng, A. Hadid, and M. Pietikainen. "A coarse-to-fine classification scheme for facial expression recognition". In International Conference on Image Analysis and Recognition, 2004.
[41] Caifeng Shan, Shaogang Gong, and Peter W. McOwan. "Facial expression recognition based on Local Binary Patterns: A comprehensive study". Image andVision Computing, 27(6):803{816, May 2009.
[42] M.S.Barlett, G.Littlewort, M.Frank, C.Lainscsek, I.Fasel, and J.Movellan. "Recognizing Facial Expression: Machine Learning and Application to Spontaneous Behavior". In Computer Vision and Pattern Recognition, 2005.
[43] P. Lekshmi.V1. , Dr.M.Sasikumar, "Analysis of Facial Expression using Gabor and SVM" , International Journal of Recent Trends in Engineering, Vol. 1, No. 2, May 2009.
[44] M.S. Bartlett, G. Littlewort, M.G. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Fully Automatic Facial Action Recognition in Spontaneous Behavior," Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition (AFGR '06), pp. 223-230, 2006.
[45] M. Osadchy, Y. LeCun, and M. Miller. "Synergistic face detection and pose estimation with energy-based models". Journal of Machine Learning Research, 8:1197{1215, May 2007.
[46] B. Heisele, T. Serre, and T. Poggio. "A component-based framework for face detection and identification. ",International Journal of Computer Vision, 74(2):167{181, 2007.
[47] D. Keren, M. Osadchy, and C. Gotsman. "Antifaces: A novel fast method for image detection". IEEE Transaction on Pattern Analysis and Machine Intelligence,23(7):747{761, 2001.
[48] H. Schneiderman and T. Kanade." Object detecion using the statistics of parts". International Journal of Computer Vision, 56(3):151{177, 2004.
[49] M. Pardas and A. Bonafonte, "Facial animation parameters extraction and expression recognition using Hidden Markov Models," Signal Processing: Image Communication,vol. 17, pp.675–688, 2002.
[50] M.J. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba. "Coding facial expression with Gabor wavelets". In IEEE International Conference on Automatic Face &Gesture Recognition, pages 200{205, 1998.

[51] Priya Metri, Jayshree Ghorpade, Ayesha Butalia, " Facial Emotion Recognition Using Context Based Multimodal Approach", International Journal of Emerging Sciences, 2(1), 171-182, March 2012.
[52] Ekman, Paul, and Wallace V. Friesen. “Manual for the facial action coding system”. Consulting Psychologists Press, 1978.
[53] Wu, Tingfan, Marian S. Bartlett, and Javier R. Movellan. "Facial expression recognition using gabor motion energy filters." In Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, pp. 42-47. IEEE, 2010.
[54] Zhang, Ligang, Dian Tjondronegoro, and Vinod Chandran. "Random Gabor based templates for facial expression recognition in images with facial occlusion." Neurocomputing 145 (2014): 451-464.
[55] Kanade, Takeo, Jeffrey F. Cohn, and Yingli Tian. "Comprehensive database for facial expression analysis." In Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on, pp. 46-53. IEEE, 2000.
[56] Wallhoff, Frank. "Facial expressions and emotion database." Technische Universität München (2006).

Mina Navran received her B.Sc. from Islamic Azad University of Qazvin in 2008 and M.Sc. in Computer Engineering from Tarbiat Modares University of Tehran in 2014, respectively. Her main interests are Affect Computing, image analysis, and data mining.

Nasrollah Moghaddam Charkari received his B.Sc. in Computer Engineering from Shaheed Beheshti University of Tehran in 1987, and his M.Sc. and Ph.D. in Computer Science and Information System Engineering from Yamanashi University of Japan in 1992 and 1995, respectively. He is currently an Associate Professor at School of Electrical and Computer Engineering at Tarbiat Modares University of Tehran, Iran. His main research interests are image analysis and understanding, complex networks, Parallel Algorithms and Processing, and bioinformatics.

Muharram Mansoorizadeh received his B.Sc. from University of Isfahan in 2000, and his M.Sc. and Ph.D. from Tarbiat Modares University of Tehran in 2002, and 2008 in Computer Science, respectively. He is currently an Assistant Professor at Bu-Ali Sina university of Hamadan, Iran.His main research interests are Affective Computing, Machine Learning, and Machine Vision.