Facial expressions recognition using an ensemble of feature sets based on key-point descriptors

M. Sultan Zia[1,2] and M. Arfan Jaffar*[1,3]

The authors in this study proposed facial expression recognition system in order to improve the expression recognition performance over the recently proposed systems. Feature sets for all training samples are constructed based on speed up robust feature descriptors. An ensemble of feature sets is then created incrementally. To achieve high diversity of ensemble, the dissimilarities between the training samples for each class are computed. This high diversity led to a high recognition rate. Experimentation on two publicly available datasets is performed. The system achieved 98·6% accuracy on JAFFE dataset and 96·3% accuracy on Multimedia Understanding Group dataset. The results of proposed system are compared with recently proposed work in this area and proved the soundness of the proposed method.

Keywords: Facial expressions recognition, Ensemble learning, SURF descriptors, Template matching, Ensemble diversity

Introduction

Communication is fundamental for a successful human life on this planet. There are two fundamental forms of communication verbal and non-verbal. Facial expressions are the most proficient form of non-verbal communication.[1] Even during a verbal communication of humans, significant information is transferred through facial expressions. Facial expressions propose substantial information about the emotional state, mindset, and intentions of the individual. Sometimes facial expressional communication becomes even more significant than verbal communication. Recognition of facial expressions is important not only for humans but also for the computers for a natural human computer interaction. The applications of automatic facial expression recognition are not limited to human computer interaction; there are many other application areas where facial expressions have an essential role like interactive games, psychology, humanoid robots, virtual reality, medicine, entertainment, security, computer assisted learning and deceiving/lie detection, etc.[2,3]

Six expressions are considered as basic expressions and they are 'Angry', 'Disgust', 'Fear', 'Happy', 'Sad', 'Surprise'. Neutral state is also considered as one of the basic expressions (Fig. 1).

Implementation of facial expression recognition system requires two important and fundamental things: one is detection of the facial feature points in input facial images as well as designing a suitable representation of these feature points and the other is a suitable discriminative method which can classify these feature points. In this study a novel automatic facial expression recognition system is proposed in order to improve the expression recognition performance over the recently proposed systems. We used speed up robust feature (SURF)[4] descriptors to transform the expression domain into feature domain. Speed up robust feature descriptors have high discriminative properties when used to represent the expression domain.[5] A novel and simple descriptor sets based classifier is proposed. Then a novel ensemble construction mechanism also defined. The classification accuracy of our proposed method is higher than the recently proposed facial expression recognition systems. The performance measure and comparison with recently proposed methods is established by using two publicly available benchmark datasets. The comparison of results with recent papers in this field shows that our proposed method has higher recognition rate.

The rest of the paper is organised as follows: the section on 'Previous work' contains previous work; detailed description of the proposed system is described in the section on 'Methodology'; the section on 'Experimentations and results' contains the details of experimentation results and discussions; conclusion is presented in the section on 'Conclusion and future work'; Acknowledgments and references are at the end.

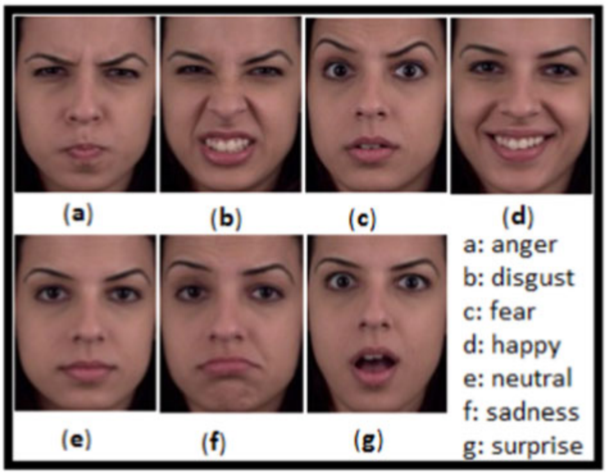
Previous work

Facial expression recognition techniques reported in literature can be divided into two broad categories according to the mechanism they adopted to extract facial expression information. The one is model template

[1] National University of Computer and Emerging Sciences, Islamabad, Pakistan
[2] COMSATS, Institute of Information Technology, Sahiwal, Pakistan
[3] College of Computer and Information Sciences, Al Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia

*Corresponding author, email arfan.jaffar@gmail.com

Received 14 September 2012; accepted 26 November 2014
DOI 10.1179/1743131X14Y.0000000097



1 Seven basic emotions (pictures courtesy MUG database)

based and the other is feature based.[6] The model template based methods use two-dimensional (2D) or three-dimensional (3D) facial models as templates to extract facial expression information while in feature based methods textured or geometrical information of the face is extracted known as features. Our proposed method falls in the second category.

Model template based methods

A point based model of face is proposed by Pantic and Rothkrantz.[7] Their model is composed of two 2D face views namely profile and frontal. Both of these views are used to interact with deformation of facial features. Then a correspondence with Fast Action Units (FAUs) is achieved. Then a set of decision rules to classify different facial expressions is established.

A 3D facial model is proposed by Braathen *et al.*[8] In an image sequence, first of all they launch the head pose. Then by using a canonical geometry the faces are warped into the facial model. Subsequently, a rotation to frontal view has been made and the model is projected back to image plane. After linearly resolving the brightness of pixels a set of Gabor filters is convolved. Finally, a bank of Support Vector Machines (SVMs) is used for facial expression recognition. A 3D shape model of face is also presented by Berretti *et al.*[9] Scale-invariant feature transform (SIFT) descriptor features are computed from the depth images of face around the facial key-points. Then they selected a subset of features with maximum relevance. Finally, support vector machine classifier is used in one versus all fashion for facial expression classification.

Recently, an interesting work is proposed by Fang *et al.* in 3D and four-dimensional (4D) domains.[10] They proposed an advanced annotated face model technique. They proposed an improvement in the fitting of atomic force microscopy and hence achieved a dance point correspondence by the combination of thin plate splines and atomic force microscopy. Then the facial expression recognition algorithm is used based on component based point distribution model. They also proposed a dance registration of 4D data. They established its effectiveness for 4D FER temporal coherence of 4D data. For further readings on 3D and 2D model based approaches a survey has been conducted by Fang *et al.* [11]

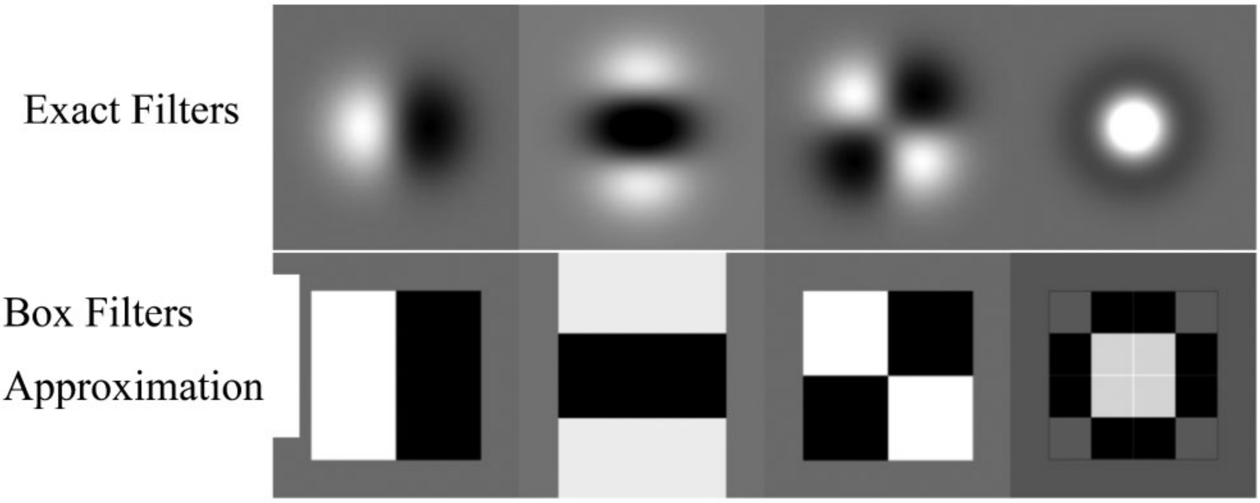
Feature based methods

A facial feature detection and tracking system is proposed by Zhang *et al.*[12] The approach provides visual information which is robust against varying lighting conditions and head motion. Dynamic Bayesian Network is used for expression recognition. Two approaches for segmentation and classification of facial expression are proposed by Cohen *et al.*,[13] one is static and the other is dynamic. Dynamic Bayesian Network with a tree structure organisation is used in static approach while a classifier based on multilevel HMM is used in dynamic approach. A real time facial expression recognition system is proposed by Bartlett *et al.*[14] On an input video stream a frontal face detector is employed and then they represent the face image in Gabor domain. Finally, the Gabor domain representation is given to a bank of SVMs for classification of seven basic expressions. Guo and Dyer[15] also used Gabor filters for feature extraction. They used many classification mechanisms like SVM, Bays, AdaBoost and linear programming.

Many authors proposed classification of facial expressions based on Neural Networks (NNs). Fasel[16] proposed a convolutional NN based solution. Groups of neurons are used for feature extraction. The system they developed is robust for changes in location of face as well as variations in scale. Matsugu *et al.*[17] proposed an FER system which uses convolutional NN for face detection. For expression classification a rule based algorithm is employed. A feed forward NN with one hidden layer is employed for expression classification by Ma and Khorasani.[18] They used Discrete Cosine Transform (DTC) for facial feature extraction.

Principle Component Analysis (PCA) is used for feature selection by Dubuisson *et al.*[19] After applying PCA a decision tree based classifier is employed for expression classification. PCA is also used by Chen and Huang[20] along with linear discriminant analysis (LDA) for dimensionality reduction of feature space. A feature set constructed in lexicographic order and finally, a nearest neighbour based rule employed for classification.

Gao *et al.*[21] extracted geometrical and structured features from the user sketched expression models. Then these features are subjected to linear edge mapping based classifier. An active appearance model based face modelling is used by Abboud *et al.*[22] Three or one PCA is used to construct the face model. The classification is carried out in active appearance model space. Bashyal and Venayagamoorthy[23] proposed an FER system based on Gabor wavelets for feature extraction and an unsupervised clustering algorithm learning vector quantisation (LVQ-1) for classification. Zhao and Zhang[24] used a kernel based manifold learning method, which nonlinearly extract the discriminant information. Local Binary Patterns (LBP) facial features are extracted and Euclidean distance based nearest neighbour classifier is used. Euclidean metric based nearest neighbour classifier is also used by Yan *et al.*[25] They proposed an adaptive discriminative metric learning for feature extraction. Owusu *et al.*[26] used AdaBoost with a neural network as base classifier. The Gabor features are extracted from the face images after reducing the feature space by Bessel transform. Fusion of Gabor features and local binary patterns is used by Zavaschi *et al.*[27] A pool of base classifiers (SVMs) is created and then a multi-objective



2 Comparison of original filters and corresponding approximations using box filters (left to right) Col.1: $D_x(x,\sigma)$, Col.2: $D_{yy}(x,\sigma)$, Col.3:$D_{xy}(x,\sigma)$, Col.4: $DoH(x,\sigma)$

genetic algorithm is used to find the best ensemble. The objectives of genetic search are accuracy and ensemble size.

Methodology

In general, a facial expression recognition system consists of a sequential configuration of processing units, which perceive to a classic pattern recognition system. These building blocks are: image/video acquisition, pre-processing, feature extraction, classification, and post-processing.[28]

In this paper a novel facial expression classification system is proposed in order to improve the facial expression classification performance over the recently proposed systems. The detail of constituting parts of the system is presented in the following subsections.

Image preprocessing

In order to implement our proposed methodology, the first step we performed is the preprocessing of the facial images. Expression representation can be sensitive to noise, lighting conditions as well as translation, scaling, and rotation of the head in an image. To combat the effect of these unwanted transformations, the facial image may be preprocessed and geometrically standardised before classification. We did not apply any geometrical standardisation like eye-alignment, etc. all the images are resized to a fix size and then we applied histogram equalisation and smoothing (3 × 3 mean filter).

Feature extraction

In order to recognise facial expressions from static images of frontal face, a set of key parameters that describes a particular facial expression is required to be extracted from the image, so that this parameter set can be used to discriminate between different expressions. If the feature set of a face image belonging to an expression class matches with that of another face belonging to some other expression class, no feature based classification technique will be able to correctly classify both of the faces. This situation is called feature overlap, and it should never occur in an ideal feature extraction technique.

Let $X=\{x_1,x_2,x_3,\cdots,x_N\}$ be the set of face images, where N stands for number of images. Let $X_{tr}=\{x_1,x_2,x_3,\cdots,x_K\}$ and $X_{te}=\{x_1,x_2,x_3,\cdots,x_M\}$ represent the training and testing samples respectively drawn from dataset X, where $N=K+M$. A label $l_i\in\{1,2,\cdots,\Omega\}$, is assigned to each face image x_i for supervised learning algorithms. Where Ω stands for number of expression classes, $\Omega=7$ in our case, as shown in Fig. 1. We used K_j to denote the number of samples belonging to the jth class. Therefore

$$K=\sum_{j=1}^{\Omega}K_j \tag{1}$$

During the training phase each image in the training set X_{tr} is transformed to the feature domain representation. Let $F_{tr}=\{f_1,f_2,f_3,\cdots,f_K\}$, $f_i\in R^{\alpha}$ be the feature domain representation of facial images of training dataset, where K is the size of training dataset and α is the feature dimension of each sample. Descriptors of SURFs[4] are used to represent the facial images in feature domain. i.e.

$$f_i=\text{SURF\_Discrp}(x_i),\quad i=1,2,3,\cdots,N \tag{2}$$

In the next few subsections we shell briefly describe the computation of SURF descriptors.

Speed up robust feature outperforms or is comparable to existing schemes in terms of repeatability, distinctiveness, and robustness, with much faster performance. The SURF descriptors are scale and in-plane rotation invariant. The SURF descriptor builds on the strengths of the leading existing detectors and descriptors. Speed up robust features have been successfully used in a broad range of applications including face authentication and 2D as well as 3D face recognition.[29] Therefore we decided to investigate their performance in facial expression recognition. The initial results of descriptor based template matching (dissimilarity measure) using SURF descriptors were quite auspicious and interesting. Therefore, we proposed a descriptor based template matching algorithm using SURF descriptors.

The SURF algorithm can be described in terms of the interest point detector and descriptor. The detector locates the key points in the image, and the descriptor describes the features of the key points and constructs the descriptors of the key points.

Interest point detection

The determinant of Hessian matrix is used to detect the interest points. The location of interest points assumed to be where the determinant has maximum value. An approximation of Hessian matrix is computed by using the integral images for computational efficiency. Let $p=(x,y)$ be a point in a facial image x, then the Hessian matrix $\mathbf{H}(p,\sigma)$ can be computed with a given scale σ

$$\mathbf{H}(p,\sigma)=\begin{bmatrix} L_{\mathrm{xx}}(p,\sigma) & L_{\mathrm{xy}}(p,\sigma) \\ L_{\mathrm{xy}}(p,\sigma) & L_{\mathrm{yy}}(p,\sigma) \end{bmatrix} \tag{3}$$

where $L_{\mathrm{xx}}(p,\sigma)$ is the convolution of the Gaussian second order derivative with the facial image x in point p, and similarly for $L_{\mathrm{xy}}(p,\sigma)$ and $L_{\mathrm{yy}}(p,\sigma)$. The box filters are used to approximate the Gaussian convolutions. The computation of box filters is much faster. The corresponding approximations using box filters are denoted by $D_{\mathrm{xx}}(p,\sigma)$, $D_{\mathrm{xy}}(p,\sigma)$ and $D_{\mathrm{yy}}(p,\sigma)$ respectively. A comparison of original filters and corresponding approximations using box filters is shown in Fig. 2 and detail can be found in Ref. 4. Therefore, the determinant of Hessian matrix is approximated as follows

$$|\mathbf{H}(p,\sigma)|=D_{\mathrm{xx}}(p,\sigma)D_{\mathrm{yy}}(p,\sigma)-wD_{\mathrm{xy}}(p,\sigma) \tag{4}$$

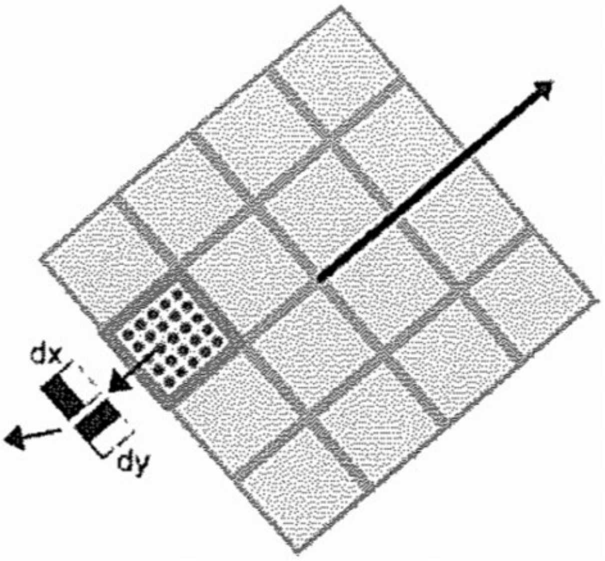
where $w(\simeq 0{\cdot}9)$is the relative weight of filter response. A pyramid of scale space is constructed to achieve scale space invariant. This can be achieved in an efficient way by changing the size of box filters.

We detected feature point candidates at locations where this determinant has maximum value. These candidates are then validated, if the response is above a given threshold. In the detection step, the local maxima of the Hessian determinant operator applied to the scale-space are computed to select interest point candidates. These candidates are then validated, if the response is above a given threshold. In order to save computation time, interest points are discarded which are near borders and corners of the face image.

Interest point description

The SURF used the sum of the Haar wavelet responses to describe the feature of a key point. Haar wavelets are used for the integral images to increase robustness and decrease computation time. For the extraction of the descriptor, the first step consists of constructing a square region centred at the key point and oriented along the orientation decided by the orientation selection method. The region is split up equally into smaller 4 × 4 square sub-regions. For each sub-region, the Haar wavelet responses are computed at 5 × 5 regularlyspaced sample points (as shown in Fig. 3). We call the dx Haar wavelet response in horizontal direction and dy the Haar wavelet response in vertical direction.

Then, the wavelet responses dx and dy are summed up over each sub-region and form a first set of entries in the feature vector. In order to bring information about the polarity of the intensity changes, the sum of the absolute values of the responses, |dx| and |dy| were also extract. Hence, each sub-region has a four-dimensional descriptor vector $\mathbf{v}$ for its underlying intensity structure $\mathbf{v}=\left(\sum \mathrm{d_x}, \sum \mathrm{d_y}, \sum |\mathrm{d_x}|, \sum |\mathrm{d_y}|\right)$. Each sub-region contributes four values to the descriptor vector leading to an overall vector of length 4 × 4 × 4=64.



3 Presentation of descriptor building

Classification

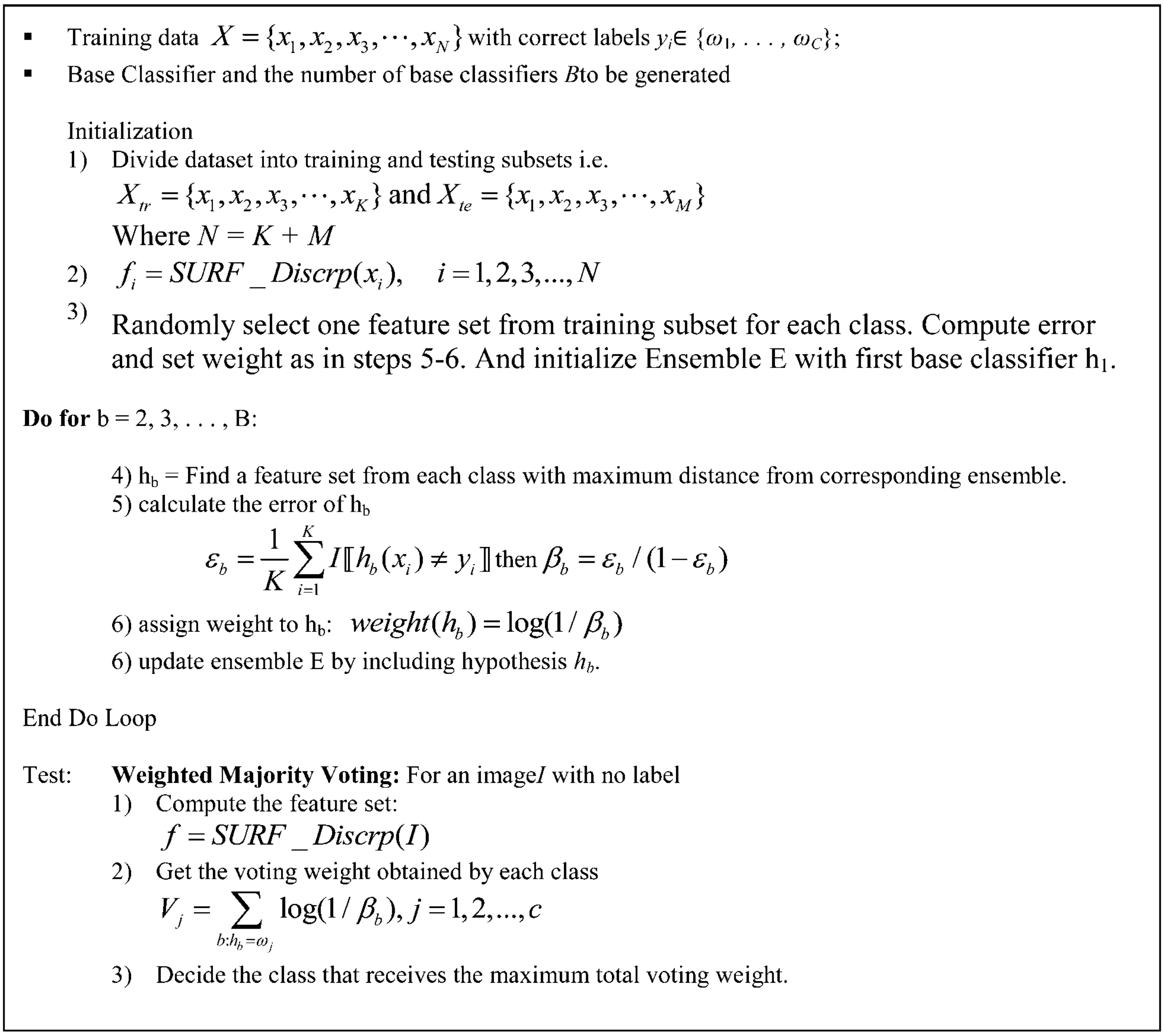
After extraction of a suitable feature set from a face image the final and most important task is classification of facial expressions on the bases of extracted features. A proper classifier selection which is fast and robust enough to any particular problem is important. In this section, we performed facial expression recognition on the bases of descriptor vectors extracted in the previous section. For this purpose we designed an ensemble of base classifiers in an incremental fashion. By taking a combination of decisions of a board of several members may provide a superior solution as compared to any single decision made by any member. It has been proved that a classifier (strong classifier) of high accuracy can be constructed by combining the outputs of several member classifiers (weak classifiers) which can barely outperform than random guessing. Another benefit of such combination is the reduction of variance of decisions and raises the confidence level of decision.

We constructed the ensemble incrementally. The algorithm to construct the ensemble from feature sets is shown in Fig. 4. Each image x_{i} is represented by a feature set f_{i} as described in equation (2). Each feature set f_{i} is consists of a set of descriptors. We initialise the ensemble classifier by selecting a representative feature set from each expression class. Each base classifier is consisted of seven feature sets selected one from each expression class. Then the classification can be performed by nearest neighbour method. Euclidian distance is used to find the similarity between the two templates

$$\mathrm{d}(D,T^{\mathrm{c}})=\sum_{\mathrm{i}=1}^{\mathrm{m}} \min_{\mathrm{k}=1}^{\mathrm{n}} \sum_{\mathrm{j}=1}^{\mathrm{M}} \left(D_{\mathrm{i,j}}-T^{\mathrm{c}}_{\mathrm{k,j}}\right)^2 \tag{5}$$

where M is number of descriptor bins, D_{i} is the descriptor of a face image and $T^{\mathrm{c}}_{\mathrm{k}}$ is kth descriptor of reference template of class C.

Now the big question is that how the representative descriptors can be selected. To find the answer of this question, we get idea from the theory of support vector machines, where the most significant feature vectors (support vectors) are those which are located near the



```
▪ Training data $X = \{x_1, x_2, x_3, \cdots, x_N\}$ with correct labels $y_i \in \{\omega_1, \ldots, \omega_C\}$;
▪ Base Classifier and the number of base classifiers $B$to be generated

  Initialization
  1) Divide dataset into training and testing subsets i.e.
     $X_{tr} = \{x_1, x_2, x_3, \cdots, x_K\}$ and $X_{te} = \{x_1, x_2, x_3, \cdots, x_M\}$
     Where $N = K + M$
  2) $f_i = SURF\_Discrp(x_i), \quad i = 1,2,3,...,N$
  3) Randomly select one feature set from training subset for each class. Compute error
     and set weight as in steps 5-6. And initialize Ensemble E with first base classifier h1.

Do for b = 2, 3, . . . , B:

     4) $h_b$ = Find a feature set from each class with maximum distance from corresponding ensemble.
     5) calculate the error of $h_b$
        $\varepsilon_b = \frac{1}{K}\sum_{i=1}^{K} I[\![h_b(x_i) \neq y_i]\!]$ then $\beta_b = \varepsilon_b / (1 - \varepsilon_b)$
     6) assign weight to $h_b$: $weight(h_b) = \log(1/\beta_b)$
     6) update ensemble E by including hypothesis $h_b$.

End Do Loop

Test:  Weighted Majority Voting: For an image$I$ with no label
       1) Compute the feature set:
          $f = SURF\_Discrp(I)$
       2) Get the voting weight obtained by each class
          $V_j = \sum_{b:h_b=\omega_j} \log(1/\beta_b), j = 1,2,...,c$
       3) Decide the class that receives the maximum total voting weight.
```

4 Incrementally learning algorithm (Learn++)

boundary area of classes.[30] These critical feature points are called support vectors and have key role in designing the SVM. Moreover, the theory on ensemble systems states that the key concept in ensemble systems is diversity.[31] Obviously, we cannot expect much benefit from a combination of similar classifiers. Therefore, we selected representative feature sets for subsequent hypothesis (base classifiers) which have maximum distance from previously constructed ensemble. The first hypothesis is constructed on randomly selected feature sets. Therefore, first of all seven feature sets are selected randomly one for each class. This set of seven feature sets is called our initial hypothesis h_1. Then the performance of h_1 is tested and its error rate is computed

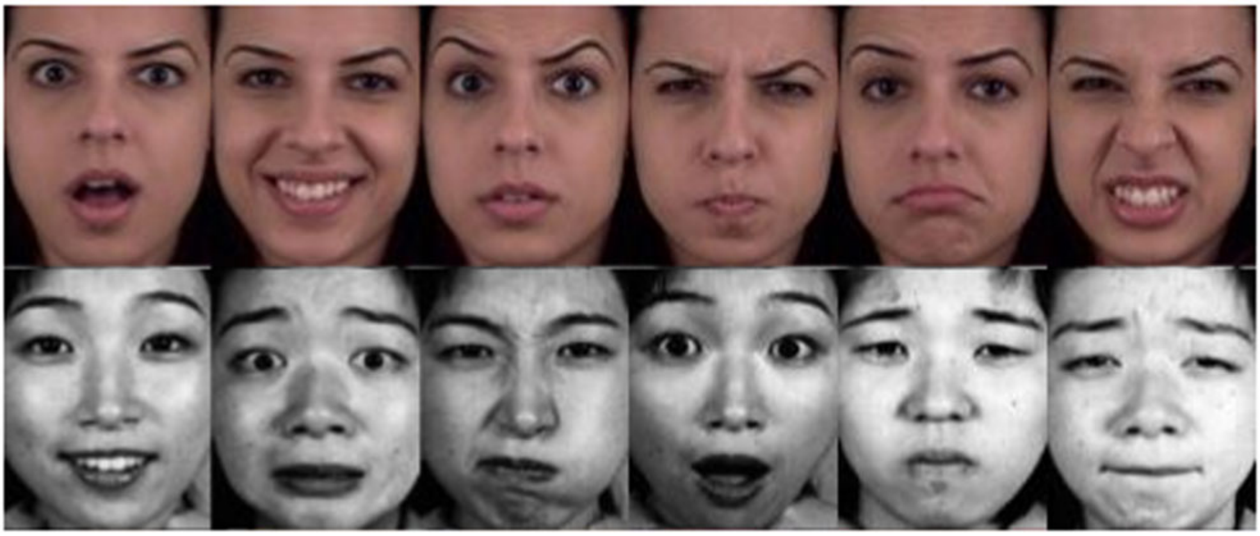
$$\varepsilon_b = \frac{1}{K}\sum_{i=1}^{K} I[h_b(x_i) \neq y_i] \tag{6}$$

After normalising this error, it is used to assign the weight of hypothesis h_1

$$\beta_b = \varepsilon_b/(1-\varepsilon_b) \tag{7}$$

$$\text{weight}(h_b) = \log(1/\beta_b) \tag{8}$$

The ensemble is initialised with this hypothesis (base classifier) and we called this ensemble E_1. The subsequent (B–1) classifiers are obtained by using the concepts of support vector machines and diversity as discussed above. We selected those feature sets (images) which are difficult to classify by the previously constructed ensemble. For this we computed Euclidian distance between the descriptors for each feature set in the training subset. Therefore, for each class we computed the distances of all training samples and find the ones with the highest distance. For example, we have created an ensemble E_b so far, and now we want to extend it to compute E_{b+1}. It means the current ensemble constituted of b base classifiers $\{h_1, h_2, h_3, \cdots, h_b\}$, hence, there are 'b' number of feature sets for each class. Let f_i^c represent ith feature set belonging to class 'c'. Therefore, we computed the total distance of each f_i^c from the 'b' feature sets belonging to class 'c' in the ensemble E_b. The feature set with maximum total distance is selected as representative of class 'c' for subsequent base classifier h_{b+1}. Similarly,



5 Examples of facial expression images from MUG (Row 1) and JAFFE (Row 2) databases

the process is repeated for each class and a representative feature set for each class is included in h_{b+1}. Then the performance of this base classifier is tested and error is calculated. This error is normalised and used to assign the weight to base classifier h_{b+1}. The algorithm to construct the ensemble incrementally is shown in Fig. 4.

To classify an image using the ensemble classifier, first of all the image is transformed into feature domain using the equation (2). The feature set is then given as input to all the base classifiers. Each base classifier h_b predicts the class of input image along with a voting weight assigned to that classifier during the training phase. Then the total voting weight is computed obtained by each class

$$V_j = \sum_{b:h_b=\omega_j} \log(1/\beta_b), j = 1,2,\cdots,c \tag{9}$$

Then, the final decision is made by weighted majority voting, i.e. the class with the largest total voting weight is declared as the final decision.

Experimentations and results

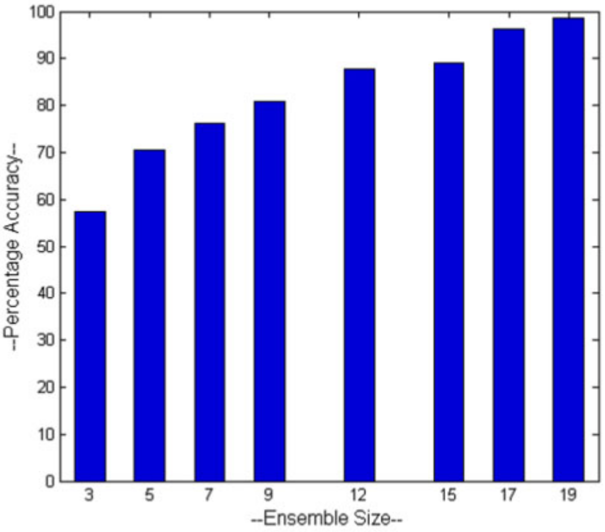
The proposed facial expression recognition system is evaluated for accuracy using two publicly available datasets namely JAFFE[32] and Multimedia Understanding Group (MUG).[33] The first experiment is carried out on JAFFE dataset that consists of 213 images of 10 Japanese female subjects. The dataset consists of grey scale images of seven basic expressions (Angry, Disgust, Fear, Happy, Neutral, Sad and Surprise). The images were captured in multiple sessions. Resolution of images is 256 × 256. The number of images of each subject corresponding to each expression is almost the same. Some sample images from JAFFE dataset are shown in the second row of Fig. 5.

The experiments on JAFFE dataset are performed in two sessions. In the first session, we constructed the ensemble of base classifiers incrementally. The dataset is divided into two parts: the training subset and the testing subset. Since each subject has at least one image for each expression and there are seven expressions including neutral. We constructed a testing subset by taking one image per subject per expression. Therefore, total 70 images in testing subset and remaining 143 are considered as training subset. This training and testing scheme is used by many authors.[23,34]

The face images are resized to 120 × 160 and a 3 × 3 mean filter is applied.

Figure 6 shows the classification performance with different ensemble sizes. X-axis shows the ensemble size that means the number of base classifiers. On the Y-axis percentage accuracy of classification of seven facial expressions is represented. From the Fig. 6 it is clear that the system learned incrementally. Our proposed system achieved high accuracy with a relatively small ensemble size. The system achieved 98·6% accuracy with an ensemble size 19. This ensemble size is obviously small as compared to SVM based ensemble[27] and Learn++.[5] This leads to a small training as well as testing time. Another advantage of our proposed system is that it works fine with small sized datasets provided that the training set should be a well representative of the data. On the other hand, the scheme proposed in Ref. 5 required a large training dataset.

For further elaboration of results, the confusion matrices corresponding to ensemble sizes five, twelve



6 Performance of ensemble system (incremental learning)

Table 1 Confusion matrix of seven-class expression recognition on JAFFE dataset obtained by ensemble size 5

	ANG	DIS	FEA	HAP	SAD	SUR	NEU
ANG	80·0%	10·0%	0%	3·3%	6·7%	0%	0%
DIS	17·3%	69·9%	10·4%	0%	3·4%	0%	0%
FEA	0%	9·4%	62·5%	9·4%	18·7%	0%	0%
HAP	0%	0%	6·4%	80·7%	9·7%	0%	3·2%
SAD	0%	3·3%	0%	12·9%	83·8%	0%	0%
SUR	0%	0%	16·7%	10·0%	0%	56·6%	16·7%
NEU	0%	0%	16·7%	10·0%	6·7%	0%	66·6%

Table 3 Confusion matrix of seven-class expression recognition on JAFFE dataset obtained by ensemble size 19

	ANG	DIS	FEA	HAP	SAD	SUR	NEU
ANG	100%	0%	0%	0%	0%	0%	0%
DIS	6·9%	89·7%	3·4%	0%	0%	0%	0%
FEA	0%	0%	100%	0%	0%	0%	0%
HAP	0%	0%	0%	100%	0%	0%	0%
SAD	0%	0%	0%	0%	100%	0%	0%
SUR	0%	0%	0%	0%	0%	100%	0%
NEU	0%	0%	0%	0%	0%	0%	100%

and nineteen are shown in Tables 1, 2 and 3 respectively. From confusion matrices it is clear that some expressions are easy to classify like angry, happy while some expressions are difficult to classify like disgust and fear. Table 3 shows that ensemble size 19 successfully classified all the seven basic expressions but disgust. Only three images out of 29 were misclassified, two as angry and one as fear.

In the second session of this experimentation on JAFFE dataset, we compared the results of our proposed approach with the recently proposed approaches. Only three approaches proposed in 2012, 2013 and 2014 were discussed here. Further comparisons are presented in Table 4. In 2012, Yan *et al.*[25] proposed an adaptive discriminative metric learning method. They achieved 96·0% accuracy by using a method of large and small penalties on between-class samples and those samples with large differences respectively to get more discriminative information. Nearest neighbour classifier used. In 2013, Zavaschi *et al.*[27] obtained 96·2% classification accuracy on JAFFE dataset by using an ensemble of 73 SVM classifiers in three groups: three LBP, 30 Gabor scale-based, and 40 Gabor orientation-based. Owusu *et al.*[26] in 2014, claimed 96·8% classification accuracy on a three-layered feed forward neural network by reducing the feature space by Bessel transform. They extracted Gabor features from reduced feature space. Our proposed method outperformed all these methods with 98·6% classification accuracy as shown in Table 4.

Our next experimentation is performed on facial expression dataset of MUG. Six hundred images of 12 subjects (both male and female) are selected for this experiment. Image sequences were captured under a controlled laboratory environment. Each person is was asked to express six basic expressions and a neutral facial expression. The images were captured at rate 19 fps with a resolution 896 × 896, in a well equipped photographic studio and in uniform lighting conditions. The captured image sequences started and ended at neutral expression; therefore, the peak of represented expression is somewhere in the middle of the sequence. Row 1 in Fig. 5 shows some sample expression images taken from MUG dataset.

Table 2 Confusion matrix of seven-class expression recognition on JAFFE dataset obtained by ensemble size 12

	ANG	DIS	FEA	HAP	SAD	SUR	NEU
ANG	100·0%	0%	0%	0%	0%	0%	0%
DIS	24·1%	72·4%	3·4%	0%	0%	0%	0%
FEA	9·7%	0%	74·2%	3·2%	9·7%	3·2%	3·2%
HAP	0%	0%	0%	86·7%	0%	0%	13·3%
SAD	3·2%	0%	0%	3·2%	93·7%	0%	0%
SUR	0%	0%	0%	0%	0%	87·7%	13·3%
NEU	0%	0%	0%	0%	3·3%	0%	96·7%

For this experiment, we selected three types of images from three different locations of the sequences: 1 with peak expressions, 2 with moderate expressions and 3 with weak expressions. The training and testing subsets are selected randomly. Fifty per cent randomly selected images are used for training and the rest of the 50% for testing purposes. The experimentation results are reported 96·3%. The confusion matrix of classification accuracy on MUG dataset is shown in Table 5.

It is interesting to note that the proposed system achieved 100% accuracy on MUG dataset when the system is trained, tested and evaluated on images with peak expressions only. These experiments on MUG dataset shows that the system successfully classified all peak expressions while it has reasonable recognition accuracy (96·33%) for moderate and weak expression representations.

Conclusion and future work

In this study, we have proposed a facial expression recognition system that has high recognition accuracy. After preprocessing the images are transformed to feature sets. The feature sets are computed by SURF descriptors. An ensemble of feature sets is constructed by maximising the diversity. This realised diversity in ensemble classifiers leads to a high recognition rate. The proposed system is evaluated on two well known facial expression datasets. The obtained results support the said claim of high classification accuracy. The results

Table 4 Comparison of classification accuracy with different approaches on JAFFE database

Reference approach	Accuracy
Liu and Wang[36] (2006), Gabor filters	92·5%
Bashyal and Venayagamoorthy[23] (2008), Gabor and LVQ features	90·2%
Zhi and Ruan[35] (2008), 2D locality preserving projections	95·9%
Cheng *et al.*[37] (2010), Gaussian process	95·2%
Oliveira *et al.*[38] (2011), 2DPCA with feature selection and SVM	94·0%
Yan *et al.*[25] (2012), adaptive discriminative metric learning	96·0%
Zavaschi *et al.*[27] (2013), Ensemble based on Gabor and LBP	96·2%
Owusu *et al.*[26] (2014), Gabor filters, Bessel Transform, AdaBoost	96. 8%
Proposed approach	98·6%

Table 5 Confusion matrix of seven-class expression recognition on MUG dataset obtained by ensemble size 22*

	ANG	DIS	FEA	HAP	SAD	SUR	NEU
ANG	95·35%	0%	0%	0%	4·65%	0%	0%
DIS	0%	98·84%	0%	1·16%	0%	0%	0%
FEA	0%	0%	96·51%	0%	3·49%	0%	0%
HAP	0%	0%	0%	100%	0%	0%	0%
SAD	0%	0%	0%	0%	97·67%	0%	2·33%
SUR	0%	0%	6·98%	2·32%	0%	90·70%	0%
NEU	3·57%	0%	0%	0%	1·19%	0%	95·24%

*Average: 96·33%.

are compared and hence, outperformed the existing methodologies.

In future, we are planning to elaborate the concept of diversity for other base classifiers. Moreover, how to modify and apply this approach for coloured features remains another interesting direction for future work.

Acknowledgements

The authors would like to thank Higher Education Commission (HEC), Govt. of Pakistan and NU-FAST for providing funds and required resources to complete this work.

References

1. Khan, R. A., Meyer, A., Konik, H. and Bouakaz, S. Framework for reliable, real-time facial expression recognition for low resolution images. *Pattern Recogn. Lett*., 2013, **34**, (10), 1159–1168.
2. Rudovic, O., Pantic, M. and Patras, I. Coupled Gaussian processes for pose-invariant facial expression recognition. *IEEE Trans. Pattern Anal. Mach. Intell*., 2013, **35**, (6), 1357–1369.
3. Vinciarelli, A., Pantic, M. and Bourlard, H. Social signal processing: Survey of an emerging domain. *Image Vis. Comput*., 2009, **27**, (12), 1743–1759.
4. Bay, H., Ess, A., Tinne Tuytelaars, T., Van Gool, L. Speeded-up robust features (SURF). *Comput. Vis. Image Understand*., 2008, **110**, (3), 346–359.
5. Zia, M. S. and Jaffar, M. A. An adaptive training based on classification system for patterns in facial expressions using SURF descriptor templates. *Multimedia Tools Appl*., 2013, 1–19.
6. Kotsia, I. and Pitas, I. Facial expression recognition in image sequences using geometric deformation features and support vector machines. *IEEE Trans. Image Process*., 2007, **16**, (1), 172–187.
7. Pantic, M. and Rothkrantz, LJM. Expert system for automatic analysis of facial expressions. *Image Vis. Comput*., 2000, **18**, (11), 881–905.
8. Braathen, B., Bartlett, S., Littlewort, G., Smith, E., Movellan, J. R. An approach to automatic recognition of spontaneous facial actions. Proc. 5th IEEE Int. Conf. on *Automatic face and gesture recognition*, Washington, DC, USA, May 2002, IEEE.
9. Berretti, S., Boulbaba, B. A., Daoudi, M., del Bimbo, A. 3D facial expression recognition using SIFT descriptors of automatically detected keypoints. *Visual Comput*., 2011, **27**, (11), 1021–1036.
10. Fang, T., Zhao, X., Ocegueda, O., Shah, S.K., Kakadiaris, I.A. 3D/4D facial expression analysis: an advanced annotated face model approach. *Image Vis. Comput*., 2012, **30**, (10), 738–749.
11. Fang, T., Zhao, X., Ocegueda, O., Shishir, S. K., Kakadiaris, I. A. 3D facial expression recognition: A perspective on promises and challenges. Proc. IEEE Int. Conf. on *Automatic face & gesture recognition and workshops (FG 2011)*, Santa Barbara, CA, USA, March 2011, IEEE.
12. Zhang, Y. M. and Ji, Q. Active and dynamic information fusion for facial expression understanding from image sequences. *IEEE Trans. Pattern Anal. Mach. Intell*., 2005, **27**, (5), 699–714.
13. Cohena, I., Sebeb, N., Gargc, A., Chend, L. S., Huang, T. S. Facial expression recognition from video sequences: temporal and static modeling. *Comput. Vis. Image Understand*., 2003, **91**, (1), 160–187.
14. Bartlett, M. S., Littlewort, G., Fasel, I., Movellan, J. R. Real time face detection and facial expression recognition: development and applications to human computer interaction. Computer Vision and Pattern Recognition Workshop, Madison, WI, USA, June 2003, IEEE, Vol. 5.
15. Guo, G. D. and Dyer, C. R. Learning from examples in the small sample case: face expression recognition. *IEEE Trans. Syst. Man Cybern. Part B: Cybern*., 2005, **35**, (3), 477–488.
16. Fasel, B. Multiscale facial expression recognition using convolutional neural networks. Proc. 3rd Indian Conf. on *Computer vision, graphics image processing: ICVGIP 2002*, Ahmadabad, India, December 2002, IISc.
17. Matsugu, M., Mori, K., Mitari, Y. Subject independent facial expression recognition with robust face detection using a convolutional neural network. *Neural Netw*., 2003, **16**, (5), 555–559.
18. Ma, L. Y. and Khorasani, K. Facial expression recognition using constructive feedforward neural networks. *IEEE Trans. Syst. Man Cybernetics, Part B: Cybern*., 2004, **34**, (3), 1588–1595.
19. Dubuisson, S., Davoine, F. and Masson, M. A solution for facial expression representation and recognition. *Signal Process. Image Commun*., 2002, **17**, (9), 657–673.
20. Chen, X.-W. and Huang, T. Facial expression recognition: a clustering-based approach. *Pattern Recog. Lett*., 2003, **24**, (9), 1295–1302.
21. Gao, Y., Leung, M. K. H., Hui, S. C., Tananda, M. W. Facial expression recognition from line-based caricatures. *IEEE Trans. Syst. Man Cybern., Part A: Syst. Humans*, 2003, **33**, (3), 407–412.
22. Abboud, B., Davoine, F. and Dang, M. Facial expression recognition and synthesis based on an appearance model. *Signal Proces.: Image Commun*., 2004, **19**, (8), 723–740.
23. Bashyal, S. and Venayagamoorthy, G. K. Recognition of facial expressions using Gabor wavelets and learning vector quantization. *Eng. Appl. Artif. Intell*., 2008, **21**, (7), 1056–1064.
24. Zhao, X. M. and Zhang, S. Q. Facial expression recognition based on local binary patterns and kernel discriminant isomap. *Sensors*, 2011, **11**, (10), 9573–9588.
25. Yan, H., Ang, M. H. and Poo, A. N. Adaptive discriminative metric learning for facial expression recognition. *IET Biom*., 2012, **1**, (3), 160–167.
26. Owusu, E., Zhan, Y. Z. and Mao, Q. R. A neural-AdaBoost based facial expression recognition system. *Expert Syst. Appl*., 2014, **41**, (7), 3383–3390.
27. Zavaschi, T. H. H., Britto Jr, A. S., Oliveira, L. E. S., Koerich, A. L. Fusion of feature sets and classifiers for facial expression recognition. *Expert Syst. Appl*., 2013, **40**, (2), 646–655.
28. Jain, A. K., Duin, R. P. W. and Mao, J. C. Statistical pattern recognition: A review. *IEEE Trans. Pattern Anal. Mach. Intell*., 2000, **22**, (1), 4–37.
29. Dreuw, P., Steingrube, P., Hanselmann, H., Ney, H. SURF-Face: face recognition under viewpoint consistency constraints. *BMVC 2009*, London, UK, September 2009, British Machine Vision Association.
30. Hearst, M. A., Dumais, S. T., Osman, E., Platt, J., Scholkopf, B. Support vector machines. *IEEE Intell. Syst. Their Appl*. 1998, **13**, (4), 18–28.
31. Polikar, R. Bootstrap-inspired techniques in computation intelligence. *IEEE Signal Process. Mag*., 2007, **24**, (4), 59–72.
32. Lyons, M., Akamatsu, S., Kamachi, M., Gyoba, J. Coding facial expressions with Gabor wavelets. Proc. 3rd IEEE Int. Conf. on *Automatic face and gesture recognition*, Nara, Japan, April 1998, IEEE.
33. Aifanti, N., Papachristou, C. and Delopoulos, A. The MUG facial expression database. Proc. 11th Int. Workshop on *Image analysis for multimedia interactive services*, Desenzano del garda, Italy, April 2010, IEEE.
34. Saha, A. and Wu, Q. M. J. Curvelet Entropy for Facial Expression Recognition. Advances in multimedia information processing-PCM 2010, 2011, pp. 617–628 (Berlin/Heidelberg, Springer).
35. Zhi, R. C. and Ruan, Q. Q. Facial expression recognition based on two-dimensional discriminant locality preserving projections. *Neurocomputing*, 2008, **71**, (7), 1730–1734.
36. Liu, W. F. and Wang, Z. F. Facial expression recognition based on fusion of multiple Gabor features. Proc. 18th Int. Conf. on Pattern recognition: *ICPR 2006*, Hong Kong, China, August 2006, IEEE, Vol. 3.
37. Cheng, F., Yu, J. S. and Xiong, H. L. Facial expression recognition in JAFFE dataset based on Gaussian process classification. *IEEE Trans. Neural Netw*., 2010, **21**, (10), 1685–1690.
38. Oliveira, L. S., Koerich, A. L., Mansano, M., Britto Jr, A. S. 2D principal component analysis for face and facial-expression recognition. *Comput. Sci. Eng*., 2011, **13**, (3), 9–13.