Proficient Automatic Facial Expression Recognition Techniques Based on PCA

Bushra Naz[1], Faraz Ahsan[2], Sajjad Mohsin[1]

[1]Department of Computer Science, COMSATS Institute of Information Technology, Islamabad, Pakistan
Correspondence: bushra.naz@comsats.edu.pk
[2] Department of Computer Science, HITEC University, Taxila, Pakistan

Abstract: Facial expression recognition has potential applications in different aspects of day to day life not yet realized due to absence of effective expression recognition techniques. This paper aims to present computationally efficient hybrid models based on automatic facial feature extraction and artificial neural network (ANN). For promising accuracy up to 7% as compared to existing models, a cascade of geometric features and appearance based features, namely: FLIP TRIANGLE and HOURGLASS. Numerical results markedly demonstrate average accuracies ranging from 83% to 87% which provides 5.63% increase on overall average accuracy as compared with existing technique.
[Bushra NAZ, Faraz AHSAN, Sajjad MOHSIN. **Proficient Automatic Facial Expression Recognition Techniques Based on PCA.** *Researcher* 2016;8(8):52-59]. ISSN 1553-9865 (print); ISSN 2163-8950 (online). http://www.sciencepub.net/researcher. 10. doi:10.7537/marsrsj080816.10.

Keywords: Facial expression recognition, Principal Component Analysis, automatic face representation, artificial neural network, JAFFE

1. Introduction

Human's cognitive process and emotions in their social interaction is shown by facial expressions. Human Behavior Interpretation (HBI) and Human Computer Interactive (HCI) interfaces minimize the gap between human thinking and machine understanding. These interfaces are used in many application areas such as intelligent tutoring software, control monitoring systems, virtual reality etc. and hence desired an embedded system for automated expression recognition with little computations. It is a challenging task and lots of methods have been proposed to solve this problem under different variations. However the human expressions are highly adapted and elusive. Hence researchers define seven distinct classes of expressions namely anger, disgust, fear, happy, sad, surprise and neutral.

The roots of Facial Expression Recognition (FER) systems can be traced down in 1971, when Ekman and Friesen defined six primary emotions each is expressed with a distinctive unique expression. Different methods have been used to recognize these expressions. An automatic FER has been developed in three phases: acquisition, representation and expression classification, where acquisition is a pre-processing part to detect face from input image and align these images independent of varying illumination and orientation conditions. In addressing FER, face region can be extracted by using either manual or automatic cropping [1-4]. Viola and Jones proposed an automatic and fast face detection algorithm which provides robust results [6].

The next two phases are main processing part which aims to extract salient features from the image and recognize facial expression based on these features [4,5]. Facial feature representation involves two steps: feature extraction and feature selection. Feature extraction describes the input face in the form of appropriate features that are used to find the correct expression. A set of pixels, often known as crucial parameters, needs to be extracted from the image that can best represent each of the expression class. This set of pixels is represented in the form of vector called feature vector and plays significant role in classification which is next step of FER.

The first step is selecting the face proportion for processing, called Region of Interest (ROI). ROI for FER can be either face region or cropped face sub-regions. A feature extraction technique is applied on this ROI. Based on scheme of feature extraction, there are two types of features: appearance based scheme and geometric based scheme. A method that analyzes whole ROI is categorized as appearance based. In geometrical based methods, the expression is analyzed by dividing the ROI into smaller components. They form features according to the shape and location of each pixel in the image. In 1978, Ekman and Friesen proposed a coding system, known as the Facial Action Coding System (FACS) for the manual labeling of facial actions. To identify emotions from facial clues, FACS enables facial expression analysis through standardized coding of changes in facial motion in terms of atomic facial actions called Action Units (AUs). This work inspired many researchers to analyze facial expressions. Another face model for FER is Active Appearance Model (AAM). An AAM based real-time technique with neural network was presented by [7]. The accurate convergence of AAM is achieved by second order minimization. Adaboost algorithm for eye tracking is used for positioning AAM to extract local shape features [8]. These geometric models provide robust results than appearance based but there are some disadvantages regarding these techniques [9]: for initial frame, the model is positioned manually on the contours of required features and components for each individual subject. Identifying facial AUs in varying pose and illumination conditions is difficult and hence targets the robustness of this method.

Kazmi et al. use 3-level two dimensional DWT for appearance based feature extraction [3]. Dongcheng builds a model by combining DWT with PCA/ FLD for finding distinct features of different expressions [10]. Gabor wavelet based feature extraction methods are largely used in the literature for robust results. Oshidari et al. proposed an adaptive Gabor wavelet transform method [11]. Tsai combine three facial feature extraction techniques that are Angular Radial Transform (ART), DCT and Gabor filter [12]. Hosseini proposes wavelet-based salient point's technique for facial components extraction [2]. Bashyal et al. build a model by applying 18 Gabor filters, with three special frequencies and six distinct orientations, on 34 manually selected facial points [4]. Mahesh also proposed a hybrid method that manually selects fiducial points and applies Gabor wavelets and PCA for feature selection [13].

Typical FER methods need to acquire whole face to process and infer about the expression. However, the more image region, facial data and action points are covered, the more processing time and memory is consumed. Hence, in this study we investigate two spatial models for facial feature representations, by introducing a cascade of geometric features and appearance based features, namely: FLIP TRIANGLE and HOURGLASS. Our results are found better from earlier work in expression recognition. As well, the proposed feature extraction methods are automatic rather than manual. Moreover, the proposed techniques minimize 44% to 51% computation time and 28 % (Hourglass) to 41 % (Flip Triangle) features as compared to whole face. The proposed system is evaluated on a publicly available JAFFE database containing seven expressions of different subjects.

2. Preliminaries

In this section, our objective is to describe the necessary preparations in the model. This study focuses on the problem of image acquisition for facial expression recognition of an individual from static images. In order to solve this problem our ROI is the face of that individual.

For automatic face detection, we used a method that Paul Viola and Michael Jones published in 2001. Their approach to detect objects in images uses Adaboost classifier in combination with Viola & Jones features [3]. It is a strong, **binary classifier** built of several *weak detectors*. Each weak classifier is simply a binary classifier. During the learning stage, each classifier works as a feature selection method. A class of local features as simple rectangles is computed as difference of sums of pixel intensities on the input image. The Viola Jones features are rapidly and efficiently calculated by using an intermediate image representation called *integral image*. This approach for face detection proves high detection accuracy while minimizing computation time for face detection.

Figure 1. Face Detection by Viola Jones Method

3. Proposed Method

For quickly solving the recognition problem, this section describes the framework of our proposed feature selection methodology. We design two hybrid facial depictions based on geometric and appearance based features for automatic FER system. The geometry of a face tells the reliable location of facial points that are used for further facial analysis. Preliminary systems employ state of the art trackers, that track the motion of these facial points but accurate and robust tracking is still ongoing research. Appearance based features depicts pixels information of whole face or specific region, represented in the form of feature vector. The block diagram is shown in Figure 2. The main phases of the proposed method(s) are explained in detail in the following sections.

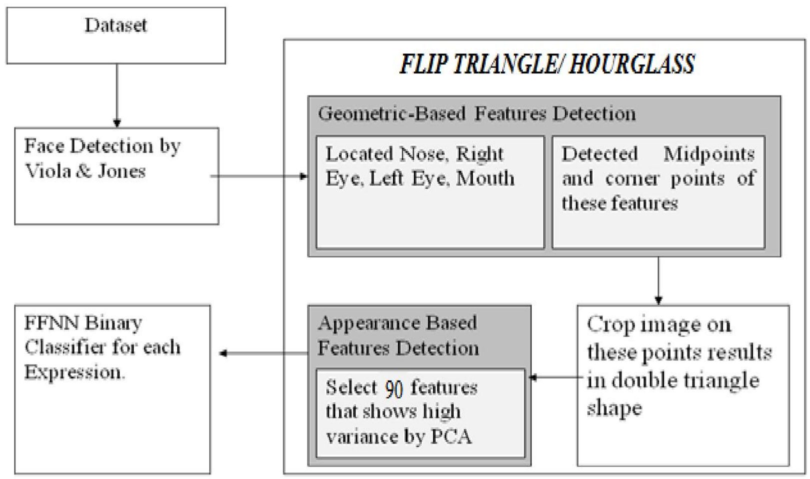


Figure 2: Block Diagram of Proposed Technique

a. Face Representation

After detecting ROI, the selection of significant points, known as features, is performed. This phase, also known as face representation, highly affects the overall performance of the system as the accuracy of expression classifier depends on these features. Preliminary approaches show that eyes and mouth regions contain most discriminative features to represent an expression [6]. The set of 'm' feature vectors, for 'm' facial images, is extracted in such a way that the feature vector belonging to one class of expression must be different from other classes of expressions so that a feature based classifier can correctly classify the expressions.

In this research, we present two mathematical geometric depictions namely: FLIP TRIANGLE and HOURGLASS. Feature extraction is automatically performed by using geometric based seven feature points. For selection of most discriminative features, Principal Component Analysis (PCA) is applied.

I. Flip Triangle Technique

From the input face image, we first create a geometrical face representation by automatic detection of seven feature points called "Flip Triangle". We have noticed that when an eye changes its position e.g. stretched or widely opened, it has same impact on whole eye muscles. This point led ground for expression recognition with half eye features instead of storing full eye features. Similarly nose midpoint and lips corner point incorporate the information required for each expression. Hence we created Flip Triangle by automatically detecting significant feature points and cropped the face based on these points. Figure 5 shows the process of creating Flip Triangle geometrical method.

A frontal face view has mouth, nose and eyes as shown in Figure 3. Here El and Er represent left and right eyes respectively, while n and m represents nose and mouth features. Let the ROI has h height and w width.

To utilize the structural relationship between aforementioned frontal face features, we have created two congruent triangles by taking nose as facial midpoint. As discussed the facial eye and mouth regions conveys maximum expression information, system will only store upper and lower triangles and ignores left and right.

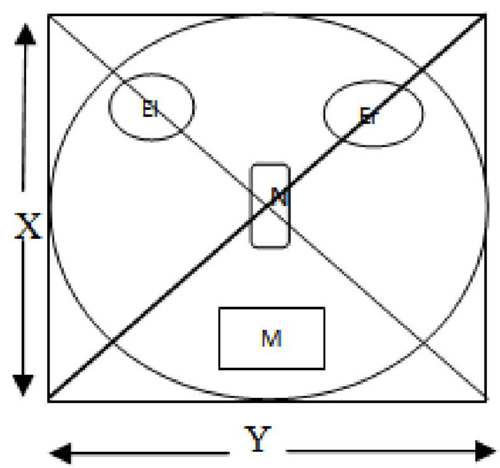


Figure 3: Simulation by Triangles of face rectangle

Consider an image of X × Y i-e 'X' number of pixels in each column and 'Y' number of pixels in each row; area of a rectangle having 'b' breadth and 'h' height can be obtained by eq 1.

$$Area = \frac{1}{2} b \times h \qquad \text{...............eq (1)}$$

Or

$$Area = \frac{1}{2} X \times Y$$

The congruent triangle (pair of upper and lower triangles) covers same width as the actual triangle but each triangle has half height as shown in Figure 4.

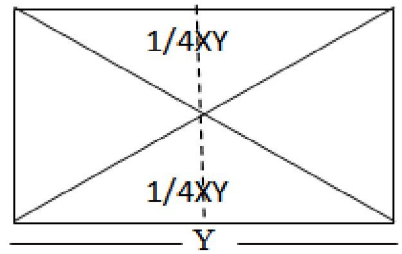


Figure 4: Two Congruent Triangles paired with other triangles that make up a rectangle.

Hence above equation (eq 1) becomes

$$Area = \frac{1}{2}(X \times \frac{1}{2} Y)$$

$$= \frac{1}{4} X \times Y \qquad \text{....eq (2)}$$

By eq 2, this method stores only 1/2 features as compared to whole face image.

After calculation it is found that 41% features are reduced from original image. The resultant image is shown in Figure 5.

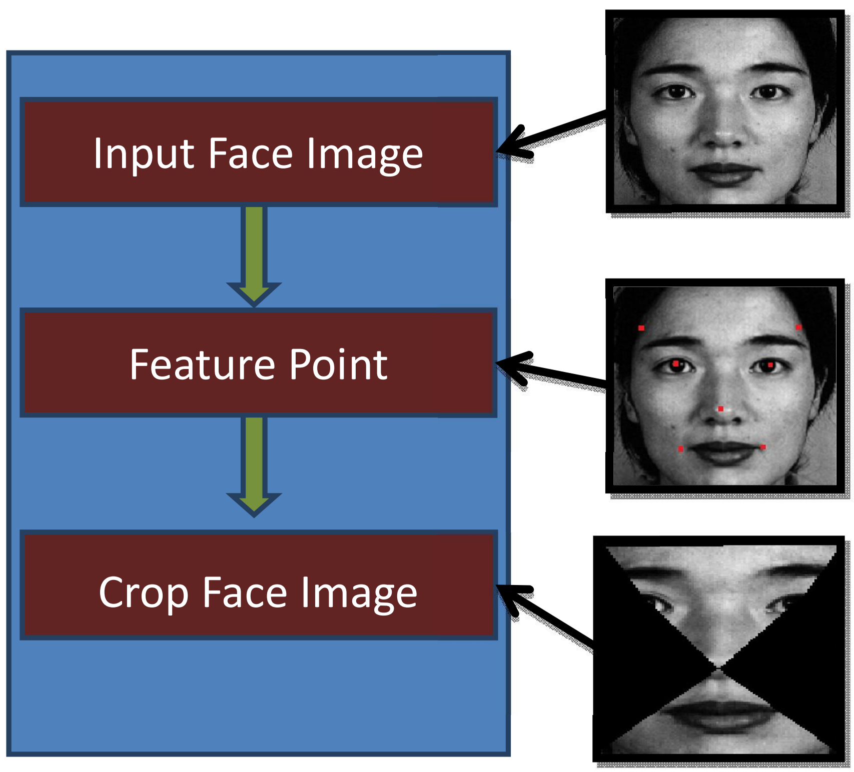


Figure 5: Flip Triangle Represe

II. Hourglass Technique

The next geometric representation created is called "Hourglass" and is more efficient than the discussed flip triangle method. It also automatically detects seven salient facial points. It incorporates facial wrinkles information and hence proves importance of facial wrinkles in each expression. These wrinkles are present around mouth and eyes. For instance, wrinkles around eyes and mouth region when a person opens mouth for smile expression or forehead wrinkles in disgust or sad expression etc. The resultant image from Hourglass representation is presented in Figure 7.

The amount of features stored in memory for hourglass representation can be viewed by eq 3 and Figure 6.

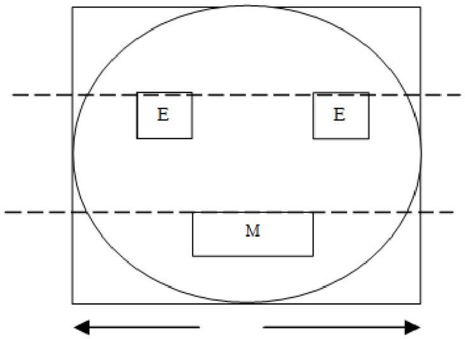


Figure 6: Simulation of Facial Features for Hourglass approach

By eq 1

$$\begin{aligned} Area &= \tfrac{1}{2} b \times h \\ &= \tfrac{1}{2} X \times Y \\ &= \tfrac{1}{2} X \times (\tfrac{1}{8} Y + \tfrac{1}{2} Y) \\ &= \tfrac{1}{2} X \times (\tfrac{5}{8} Y) \\ &= \tfrac{5}{16} X \times Y ....eq(3) \end{aligned}$$

By eq(3), this method reduces approximately 31% features of original facial image. This theoretical calculation is near to our practical approach which reduces 28% features from original image.

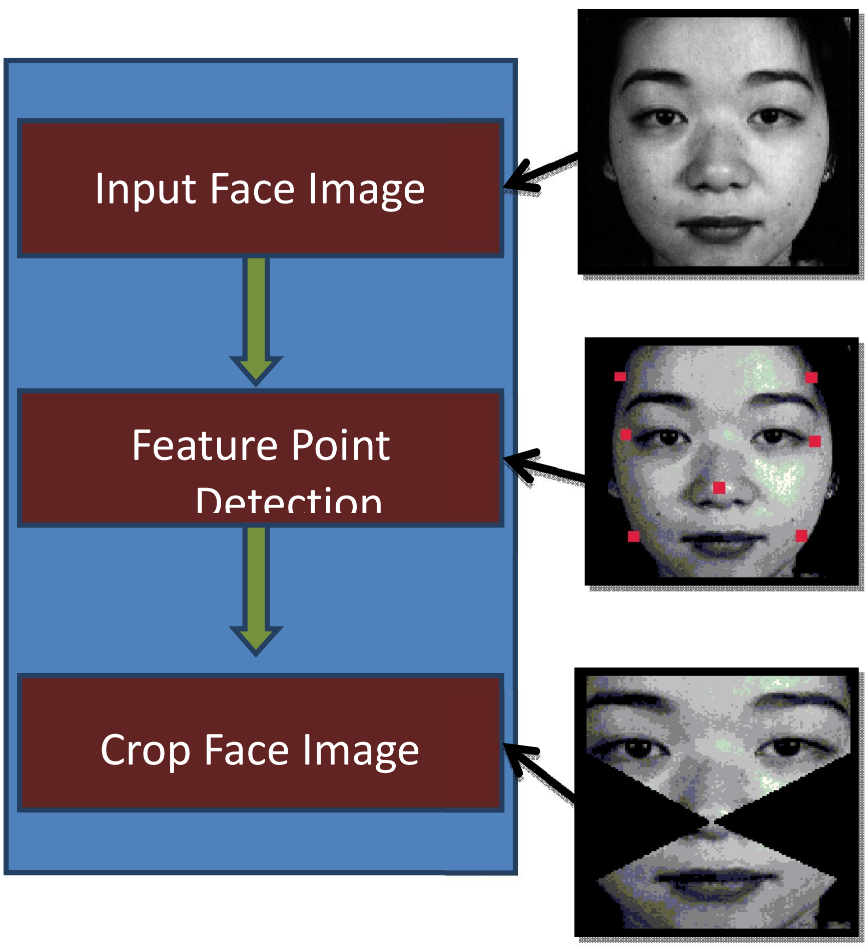
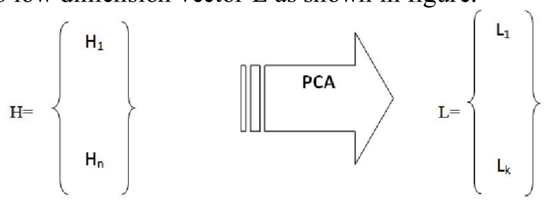


Figure 7: Hourglass Represe

III. Principal Component Analysis

In our experiment, we applied PCA on the cropped face region to select optimal and distinctive features which minimize inter class variations of an expression while maximize intra class variations. PCA is a statistical technique used for data mining. It reduces data dimensions from high variables to a few variables that can represent original data effectively as shown in figure.

Let a high dimension vector H of the image contain n number of rows; use PCA that converts data to low dimension vector L as shown in figure.



where k<n.

Let $H = \{h_1, h_2, ..., h_n\}$ be the set of n vectors, the method runs in six steps

1) Computing mean of sets h.

$$\bar{h} = \frac{1}{m}\sum_{i=1}^{m} h_i$$

2) Subtract the mean

$$\Psi_i = h_i - \bar{h}$$

3) Set the matrix $A = [\Psi_1, \Psi_2, ...\Psi_n]$ and compute

$$C = AA^t = \frac{1}{m}\sum_{i=1}^{m} \Psi_i \Psi_i^T$$

4) Compute eigenvalues of C

$$\lambda_1 > \lambda_2 > .... > \lambda_n$$

5) Compute eigenvectors of C, $u_1, u_2, ...u_n$.

$$h - \bar{h} = L_1 u_1 + L_2 u_2 + ...L_n u_n = \sum_{i=1}^{n} L_i u_i$$

6) Dimentionality Reduction

$$\hat{h} - \bar{h} = \sum_{i=1}^{k} L_i u_i$$

Where k<<n, u_i are the eigenvectors corresponding to the k largest eigenvalues. The representation of $\hat{h} - \bar{h}$ into the basis $u_1, u_2, ...u_k$

$$L = \begin{Bmatrix} L_1 \\ \\ L_k \end{Bmatrix}$$

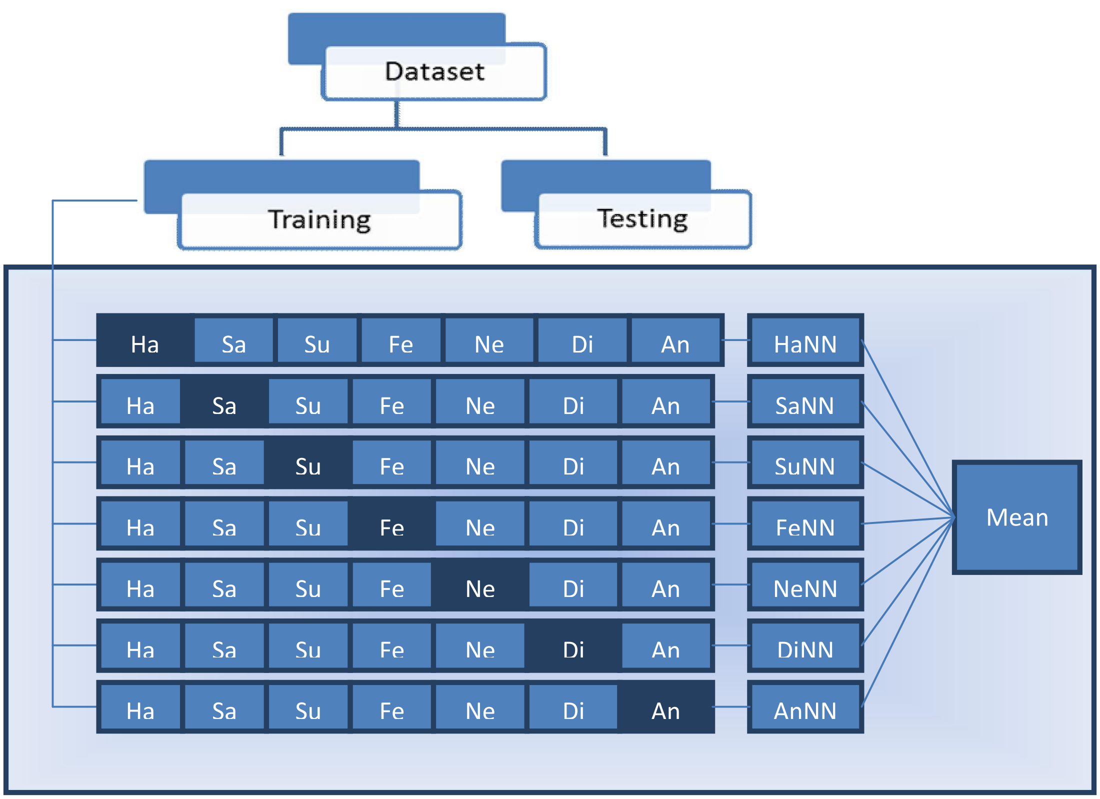


Figure 8: Architecture of Proposed Classifier

This reduced set of variables is called principal components. These are linear combination of the underlying original variables and makes high dimensional data comprehend and easy to visualize. PCA is an unsupervised method, in the sense that no information regarding groups is used in the dimension reduction. This means that PCA shows a visual representation of the most dominant patterns in a data set.

IV. Recognition

Lastly, a classification model is created to classify accurate expression, by using features detected at previous step. The features are provided as input; the classifier performs classification and results in one of seven output classes. Number of output classes depends on number of expressions to classify.

By literature, we deduced that FFNN is much simple and gives most promising results for recognition. We combined multiple binary classifiers, one against all approach, to improve generalization performance as compared to single classifier [3]. We designed seven FFNN for seven different expressions. Each one is able to recognize one of the seven expressions. We combined the results of this pool of classifiers by mean rule. The architecture of our proposed classifier is shown in Figure 8.

4. Implementation Details

We experiment our proposed technique using Japanese standard expression library JAFFE. It contains seven concrete classes of expressions namely happy, sad, neutral, anger, disgust, surprise and fear. There are 213 grey-scale images of female expressions and each image is of 256×256 size in tiff format. The heads of the subjects are in frontal pose.

The proposed technique is implemented on a Corei3 processor. An engineering and mathematical programming language MATLAB 2012 is used for programming environment. Image processing toolbox and statistics toolbox are used for code development.

5. Results and Discussion

We perform our experiment on JAFFE database with static images. 213 images of 10 persons were tested. Seven expressions were recognized namely: happy, sad, anger, disgust, surprise, fear and neutral. The facial images are represented as "Flip Triangle" geometric representation and PCA is used for selecting 90 high variance features. These features are then used for training and testing a bank of seven binary FFNN classifiers for expression recognition. Ten-fold cross validation is used for network training and testing in a way that the images used in training were not included in the testing phase. During training phase, network parameters are adjusted: number of nodes at input layer are 90 (90 features are selected per image), two hidden layers with 45 and 20 nodes and an output layer with 2 nodes (binary classifier). The learning rate of the network is varied to find the best learning rate and hence 0.5 works the best.

The results are shown in Table-1. It is shown that our proposed flip triangle method outperforms previous technique in all expressions with 2% to 3% on average except for 'happy' expression. Hence we concluded that in recognizing facial expressions, wrinkles around mouth and eyes plays significant role.

Table 1. Comparison of Appearance Based Technique [3] And Proposed Hybrid

Expression	DWT [3]	Proposed	
		FlipTriangle	Hourglass
Anger	81.1	83.9	84.48
Disgust	82.4	89.01	89.67
Fear	81.2	90.12	92.47
Happy	81.3	81.09	87.51
Sad	80.05	83.8	85.95
Surprise	81.5	82.9	89.61
Neutral	81	84.9	87.49

For incorporating such useful information, we enhanced the proposed method to Hourglass representation. The Hourglass representation improves recognition rate from 4% to 7%.

Our automatic system outperforms existing techniques. In terms of feature extraction and selection, significantly minimizes computational cost and memory demand. We achieved accurate expression recognition up to 87% with 44% to 51% less computation time. The memory demand is 28% to 41% less as compared to whole face. We developed a memory and CPU time efficient technique without degrading recognition accuracy. Hence we are able to conclude that the proposed technique can efficiently and effectively recognize facial expressions.

Other parameter to compute efficacy of an automatic FER system is the computation time for feature vector calculation per image. Table-2 shows excellence of our method as compared to whole face region for recognition of seven expressions on JAFFE.

Table 2: Comparison in Terms of Computation Time

Experiment	Computation Time (ms)
Full Face+PCA	75.6 ms
Flip Triangle+PCA	37.4 ms
Hourglass + PCA	42.2 ms

6. Conclusion

Two new hybrid feature extraction techniques, namely Flip Triangle and Hourglass, are proposed which provides automatic detection of best features for expression recognition. The benchmark of our proposed methods is to find approximate geometric points automatically. We reduced 28% to 41% features as compared to whole face. Optimal expression details are detected by applying appearance based PCA on cropped sub-region. The eyes and mouth regions can be used to recognize an expression efficiently. We developed a memory and CPU time efficient technique without degrading recognition accuracy.

The proposed hybrid feature extraction is tested on 2-D grayscale images. In future, we intend to extend the proposed method to 3-D and colored images. The proposed method can also be improved to be used in real time. Our method may be enhanced with varying classifiers for other image databases, as well.

References

1. Buciu I., Kotropoulos C., Pitas I., ICA and Gabor representation for facial expression recognition. In: IEEE 2003 International Conference of Image Processing Volume 2, 14-17 Sep 2003, pp: II-855-8.
2. Aabidin Z., Harjoko A., A neural network based facial expression recognition using fisherface. Journal of Computer Applications 2012; 59(3): 30-34.
3. Kazmi S., Ain Q., Jaffer M., Wavelets based facial expression recognition using a bank of neural networks, In: IEEE 2010 Future Information Technology (FutureTech); 21-23 May 2010; Busan, South Korea: IEEE. pp. 1-6
4. Bashyal S., Venayagamoorthy G., Recognition of facial expressions using Gabor wavelets and learning vector quantization. Engineering Applications of Artificial Intelligence 2008; 21(7): 1056-1064.
5. Zhang Z., Feature-based facial expression recognition. INT J PATTERN RECOGN 1999; 13(6): 893-911.
6. Gupta S., Agrwal S., Meena Y., Nain N, A Hybrid Method of Feature Extraction for Facial Expression Recognition. In: IEEE 2011 Signal-Image Technology & Internet-Based Systems (SITIS); Nov 28- Dec 1 2011; Dijon, France : IEEE. pp. 422-425.
7. Choi H., Oh S., Real Time Facial Expression Recognition using Active Appearance Model and Multilayer Perceptron, SICE- ICASE International Joint Conference; 18-21 Oct 2006; Busan South Korea: IEEE. pp. 5924-5927
8. Tang F., Deng B., Facial Expression Recognition using AAM and Local Facial Features, Third International Conference on Natural Computation ICNC 2007; 24-27 Aug 2007; Haikou, China: IEEE. pp. 632-635.
9. Zor C. Facial expression recognition, Master of Science in signal processing and machine intelligence, University of Surrey: Guildford, Surrey, UK, 2008.
10. Dongcheng S., Jieqing J. The method of facial expression recognition based on DWT-PCA/LDA, Third International Congress on Image and Signal Processing (CISP) 2010; 16-18 Oct 2010; Yantai, China: IEEE. pp. 1970-1974.
11. Oshidari B., Araabi B. An effective feature extraction method for facial expression recognition using adaptive Gabor wavelet, IEEE International Conference on Progress in Informatics and Computing (PIC) Volume 2; 10-12 Dec 2010; Shangai, China: IEEE pp:776-780.
12. Tsai H., Lai Y., Zhang Y. Using SVM to design facial expression recognition for shape and texture features, International Conference on Machine Learning and Cybernetics (ICMLC); 11-14 July 2010; Qingdao, China: IEEE pp. 776-780.
13. Kumbhar M., Jadhav A., Patil M. Facial expression recognition based on image feature; International Journal of Computer and Communication Engineering 2012; 1(2): 117-119.

8/24/2016