TTIC, 2020, Vol. 4, 1-9

Face Expression Recognition using Shape Factors Induced by Landmark Triangulation

Avishek Nandi
Department of computer & System Sciences
Visva-Bharati University
Santiniketan, India.
avisheknandi10@gmail.com

Paramartha Dutta
Department of computer & System Sciences
Visva-Bharati University
Santiniketan, India.
paramartha.dutta@gmail.com

Md Nasir
Department of computer & System Sciences
Visva-Bharati University
Santiniketan, India.
nasir.vb@gmail.com

***Abstract*—Facial Expression Recognition (FER) has many real life applications in Human-Computer Interaction (HCI). The detection of basic facial expressions such as Anger, Disgust, Fear, Happiness, Sadness, and Surprise in a given face image is a challenging problem. The authors propose a novel method where 1) The Active Appearance Model (AAM) is used to generate *sixty-eight* facial landmark points. 2) Then *twenty* salient landmark points out of sixty-eight points are identified and used to form triangulation on the face. 3) Then *seven* different geometric shape factors are calculated for each triangle in the triangulation set. 4) Each of their shape factors is trained with Multi- Layer Perceptron (MLP) for the classification of expressions. 5) Then the best performing shape factor is selected as the final feature. The proposed method is well tested on benchmark databases viz. CK+, JAFFE, MMI, and MUG. The effective and efficient learning of the shape factor with MLP shows extremely encouraging results.**

***Index Terms*—Facial Expression Recognition, Active Appearance Model (AAM), Shape Factors**

I. INTRODUCTION

Facial expression recognition a focal research interest of the affective computing community for about the last two decades applications encompassing Medical, Business, Education, Security and, Surveillance. Some studies reveal that among manydifferent forms of human emotions the one emanating from the face has the highest ability to differentiate one emotion from another. Thus, expression reflected in a face contributes a significant amount of characteristic information for automated recognition of human emotions. To classify the different types of emotions effective grouping of emotions in distinct classes is necessary. For this purpose, six classes of atomic emotions with a substantial difference in appearance and emotional quality are grouped into six classes of expressions viz. Anger (ANG), Disgust (DIS), Fear (FER), Happiness (HAP), Sadness (SAD) and, Surprise (SUR). According to Ekman et al., these six expressions are indispensable constituents among different people despite their caste, creed, and origin [13]. These six basic expressions are uniquely identifiable because a single class of emotion emulates a unique combination of muscular movements yielding classifiable feature cues corresponding to that particular class of emotion. Another important consideration towards automatic facial expression recognition is that the quality of the detected feature, which plays an important role in the segregation of facial expression in different classes. Facial Action Coding System (FACS) is one such system that describes the movement of facial muscles to classify each facial expression and effectively associate them to its respective affect classes [8], Although FACS is a powerful feature descriptor Action Unit (AU) detectors of FACS suffer from misclassification in the case of occultation and pose variation. In contrast to the FACS system, the geometric-based methods directly use location and shape information of relevant facial components viz. eyes, eyebrows, nose, and mouth to extract the emotion-related features from the face induced by location points. These, in turn, are fed into a classifier without any intermediate stage of Action Unit detection. Design of an effective Facial Expression Recognition (FER) System involves the following steps of computations i) Face Identification ii) Feature Extraction and Reduction and iii) Classification learning. For the present, the viola Jones cascade face detection method [17] is used for face identification. The most popular geometric-based feature extraction method is the Active Appearance Model (AAM) [9]. Most popular texture- based methods are Histogram oriented Gradients (HoG) [11], Gabor Filters [12], and Wavelet Filters [32], etc. The classifiers used in designing an effective FER system are Support Vector Machine (SVM), Radial Basis Function (RBF), Multi-Layer Perceptron (MLP), Convolutional Neural Network (CNN), etc. The main challenges of designing a Facial Expression Recognizer (FER) system pervade changes in head orientation, pose of the subject, the lighting condition, interpersonal variance, occlusion on the face, resolution of the captured image, noise conditions, etc. To overcome these issues the authors prefer the Active Appearance Model (AAM) which is a statistical object recognition model and capable of representing a shape-changing object using a set of landmark points [9]. The AAM represents the shape of a face using 68 landmark points even though all of those points don't have the same significance for recognizing an expression. So, only *twenty one* important out of *sixty eight* landmark points is considered for further processing. To extract the geometric shape properties of the face triangulation sets are formed constituting all possible triangles that can be generated using that *twenty ones* landmark points. These landmark points are selected from the eyes, eyebrows, nose, and mouth region of the face leaving the points on the outer perimeter of the face. The authors identify *six* landmark points on both the eyebrows, *eight* landmark points on both the eyes, *three* landmark points on the nose and *four* landmark points on the mouth region. The shape properties of the triangle are extracted using *seven* different shape factors formulae of a triangle. These shape factors are individually learned with a Multi-Layer Perceptron (MLP) Network to find out the best shape descriptor in the context of a relevant database.

A. Gap Analysis

In the section Related Works we have discussed some state of the works relevant to our works. Most geometric-based works are applied on sequence databases and with a single geometric-based feature. Some studies also include the application of geometric-based features on static databases. But the application of various types of geometric shape-based features for efficient and effective recognition of geometric facial features still needs to be explored in more detail.

B. Motivation

The geometric shape of the face is the key attribute of facial expression recognition. The geometric-based methods have advantages over other methods that it is generally immune to head rotation, lighting conditions, resolution, subject pose. The proposed method is different from the method in [4] in that authors have only used a single shape property of a triangle whereas our paper includes *seven* different shape factors for each triangle. Our paper also includes geometrically more powerful features for facial expression recognition.

C. Contribution

The proposed method has many contributions which are listed below. (1) The selection of salient landmark points sensitive to basic facial expression. (2) *Seven* different shape-factors describing the shape properties of each triangle. (3) Individual learning of shape factors helps to find out the best features. (4) Application of simplistic learning machine MLP for robust and real-time facial expression recognition. (5) Validation of the proposed method with four benchmark databases to establish the performance stability of the algorithm. (6) Comparison of MLP with Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), K-Nearest Neighbour (KNN) to prove the performance superiority of the proposed method over other states of the art classifier.

D. Organization

The paper is organized as follows apart from the introduction section; section II discusses the state of the works relevant to our studies. Section III gives a brief introduction about the Active Appearance Model (AAM) and Multi-Layer Perceptron (MLP) which are the foundation of the proposed work. Section IV illustrates the methodology of the proposed algorithm in detail. Section V contains the results and description of CK+, JAFFE, MMI, and MUG databases in detail. Section VI describes discussion and comparisons. Finally, section VII concludes our study.

II. RELATED WORKS

The identification of proper kinds of facial features is an indispensable requirement of a Facial Expression Recognition (FER) system. There are many works on facial expression recognition systems here the authors only reviewed recent geometric-based works that are relevant to this paper. The authors also compared the results with all the state of the art recent works which are discussed here. The works of Kotsia and Pitas include geometric features on the grid nodes with facial landmarks on the image sequences along with a Support Vector Machine (SVM) classifier for classification of six basic expressions [21]. Another geometric feature-based by Saeed et al. achieved the state-of-the-art performance with only *eight* facial landmark points [28] in the sequence database. The pioneering work of Kamarol et al. used spatiotemporal features extracted by the Harris corner algorithm which are learned with Support Vector Machine (SVM) classifier [20]. Barman and Dutta used geometric-based distance and shape signature along with some statistical measures which are learned with MLP classifier [2] for effective recognition of facial expressions. In contrast to that Happy and Routray used salient facial patches connected with emotion elicitation with some modification to Active Appearance Model (AAM) landmark localization method [15]. Cheon et al. used the distance between neutral and peak expression class as discriminatory features which are learned with K-Nearest Neighbors (K-NN) based classifier [7]. Another work by Barman and Dutta used Distance and Texture based features with a bag of the classifier to enhance the expression recognition performance [3]. The work by Xei et al. focuses on Intraclass Variation Reduced Feature (IVRF) to remove intraclass negative influence which is seen during the training KNN classifier [31].

III. PREREQUISITE

In this section, the authors briefly described the (1) Active Appearance Model (AAM) and the (2) Multi-Layer Perceptron (MLP) which are essential for understanding the methodology [section IV].

A. Active Appearance Model (AAM)

AAM is a model of shape and appearance of any deformable object which is developed by Cootes et al. and they have shown its application in the detection of the shape of a face [9]. The model is trained with the shape and appearance patterns of given examples images. The model tries to fit a basic shape pattern to an object shape by predicting the changes in the model parameters. An AAM shape instance of a face can be represented as

$$S = [\, x_1, y_1, \ldots., x_L, y_L]^T \quad (1)$$

S is a $2L \times 1$ vector consisting of L landmark points. The AAM model is trained using a set of N images $I_1, I_2, \ldots.., I_N$ that is annotated with a set of L landmarks. Then the Principal Component Analysis (PCA) is used to find the frame-wise normalized landmark points that are approximated using formula 2.

$$x = \bar{x} - P_S W_S Q_S c, g = \bar{g} - P_g Q_g c \quad (2)$$

Where,

$$Q = (Q_S, Q_g) \quad (3)$$

Where $\bar{x}$ is mean shape, P_S are modes of orthogonal variations and c is the shape parameter. W_S is the shape parameter matrix. Q_S and Q_g are eigenvectors controlling shape and Gray level parameters c are the appearance parameters.

B. Multi Layer Perceptron (MLP)

The MLP is a learnable machine of the Neural Network (NN) class that has at least three layers of neurons and it can be used for function approximation, classification, etc [16]. The MLP has three stages of computation 1) Learning 2) Testing 3) Prediction. The back-propagation algorithm is used to train an MLP and it has two phases of computation 1) the forward pass 2) the backward pass. In the forward pass network weight is fixed and the input signal is propagated from the input layer to the hidden layer towards the output layer and network response is recorded. In the backward pass at first, the supervision error is computed by differentiating network response from the actual response, and this error signal is propagated from the output layer to the hidden layer towards the input layer. As the error signal is back-propagated the network weights are adjusted or in other words, the network is being trained. This process of training is repeated until the error values are very small in other words the network is producing the desired response. In the second stage of learning the trained MLP model is tested with a set of input values that are new to the trained model. This is to ensure that the network is learning the features from the input dataset instead of memorizing. At the final stage, the trained MLP model is used for the prediction of unknown inputs.

Network outputs of MLP can be mathematically defined with equation 4.

$$Y_j(n) = \emptyset(y_i(n)) \quad (4)$$

where,

$$y_j(n) = \sum_0^m W_{ij}(n) y_i(n) \quad (5)$$

In the backward pass at first, the supervision error is computed by differentiating network output with an actual response and the error signal are propagated from the output layer to hidden layer to the input layer $e_j(n) = d_j(n) - y_j(n)$, As the error signal is back-propagated the network weights are adjusted with the formula 6.

$$\Delta W_{ij}(n) = \rho \delta_j(n) y_i(n) \quad (6)$$

Where, $\Delta W_{ij}(n)$ is weight correction, ρ is learning rate and $y_i(n)$ is the input signal of the j^{th} neuron, n iteration index. If j is an output neuron $\delta_j(n)$ is calculated using equation 7.

$$\delta_j(n) = e_j(n) \varphi'_j \left(\vartheta_j(n)\right) \quad (7)$$

If j is a hidden neuron then$\delta_j(n)$ is calculated using equation 8.

$$\delta_j(n) = \varphi'_j \left(\vartheta_j(n)\right) \sum_k \delta_k(n) W_{kj}(n) \quad (8)$$

IV. METHODOLOGY

The Proposed Facial Expression Recognition (FER) system has two phases of computation i) The Training Phase and ii) The evaluation Phase. The training phase has the following steps of computation. (1) Image input. (2) Facial Landmark Localization with AAM. (3) Salient Landmark Selection. (4) Triangulation Formation. (5) Shape-Factors calculation using three angles of triangle (6) Classification learning with MLP. The second phase (evaluation phase) is the as same as the training phase but only differs in the last stage which is "Predicted expression level" instead of ''Classification Learning with MLP". The design of the proposed system is given in figure 1.

The authors have also presented the proposed method in algorithmic format. The algorithm 1 takes expression images and associated levels as input and outputs a trained model with associated shape factor and accuracy.

A. Facial Landmark Plotting with AAM

To determine the landmark points on the face from a given input image firstly the face region needs to be detected properly. The Viola-Jones cascade face detection algorithm is used for detecting the rectangular bounding region of the face [17]. Next, the Active Appearance Model (AAM) fitting model is used to localize landmark points on eyes, eyebrows, nose, lips, and perimeter region of the face the implementation of which can be found on [29]. Two example expression images with landmark points plotted on the face displayed in figure 2.

B. Salient Landmark Selection

The AAM generates a total of *sixty eight* landmark points covering all the components of the face but all those points are doesn't carry the same level of information for recognizing an expression. So the authors have selected only the salient land-mark points which are highly sensitive to facial expression. Figure 3 shows the landmark points plotted on two examples face image. It can be observed from figure 3 that *three* points are selected from both the eyebrows $[3 \times 2]$, *four* points are selected from both the eyes $[4 \times 2]$, *tree* points on the nose

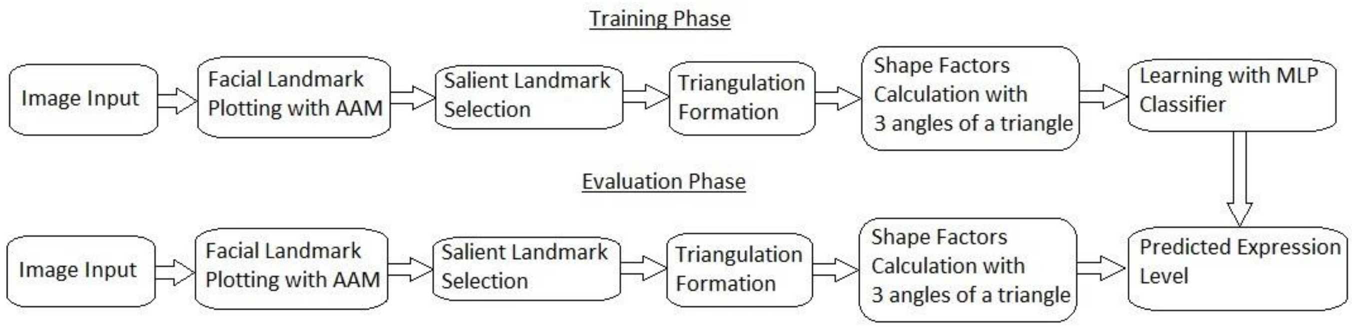

Fig. 1. Flow Chart of the proposed Facial Expression Recognition (FER) system.

Algorithm 1: Proposed Training Algorithm for Facial Expression Recognition

INPUT: A set of N expression images of size $n \times m$ and Expression Levels E

STEP1: Extract the Landmark points using AAM fitting algorithm.

STEP2: Select the 21 salient landmark points $L = [(x_1, y_1), (x_2, y_2), \dots, (x_{21}, y_{21})]$.

STEP3: Generate the set of Triangulation T using Landmark points, where $T = [(x_i, y_i), (x_j, y_j), (x_k, y_k)]$ and $i, j, k \in [1, L]$.

STEP4: for each t_i in T Compute distance triplet $D_{ijk} = [a, b, c]$, $a = dist\big((p_1, p_2), (p_3, p_4)\big)$, $b = dist\big((p_3, p_4), (p_5, p_6)\big)$, $c = \big((p_1, p_2), (p_5, p_6)\big)$ and using D_{ijk} compute angle A, B and C of t_i

STEP5: Compute shape factors sf_i where $i \in [1, 7]$ using angle A, B and C of t_i

STEP6: Construct an MLP machine M_i with shape factors sf_i as input and expression level E as output.

STEP7: Train M_i with (sf_i, E) for $i \in [1, 7]$.

STEP8: Perform Cross-Validation on $v_i = CrossVlaiate(M_i)$ for $i \in [1, 7]$.

STEP9: Select sf_i with $max\big(Accuracy(M_i)\big)$ where $i \in [1, 7]$.

OUTPUT: Shape Factor sf_i, Trained Model M_i, validation accuracy v_i.

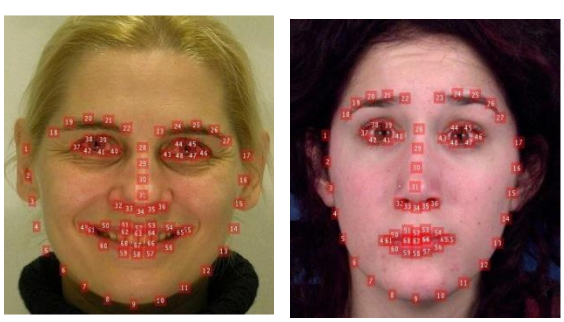

Fig. 2. 68 landmark points generated by AAM plotted on face image.

[3] and *four* landmark points on the mouth region [4]. So total $3 \times 2 + 4 \times 2 + 3 + 4 = 21$ landmarks are selected as salient landmark points. The landmark points are selected based on Facial Action Unit (FAU) shown on table I.

Fig. 3. The 21 salient landmark points are plotted on the face image.

C. Triangulation Formation

To mathematically model the geometric shape of the face tingles are formed on the face using that twenty-one landmark points. The maximum number possible of unique triangles which can be drawn using *twenty one* landmark points are ${}^{21}_{3}C = 1330$. Those triangles altogether form triangulation set T_r. We can mathematically define T_r as follows. If the set of salient landmark points are

$$Sl = [(x_1, y_1), (x_2, y_2), \dots, (x_{21}, y_{21})]$$

then the triangulation set (T_r) can be represented as equation 9.

$$T = \left\{ \begin{array}{c} \{(x_1, y_1), (x_2, y_2), (x_3, y_3)\}, \\ \{(x_1, y_1), (x_2, y_2), (x_4, y_4)\}, \dots\dots, \\ \{(x_{19}, y_{19}), (x_{20}, y_{20}), (x_{21}, y_{21})\} \end{array} \right\} \quad (9)$$

The formation of triangulation set T is displayed on figure 4.

D. Shape Factor Calculation

The authors have used *seven* different shape factors to describe the shape properties of a triangle. The shape factors of a triangle are calculated as follows.

TABLE I
SELECTION OF FACIAL LANDMARK POINT BASED ON FACIAL ACTION UNIT (FAU)

Emotions	Action Units	Description	Selected Points
Happiness	6 + 12	Cheek Raiser, Lip Corner Puller	7,10,11,14,18,19
Sadness	1+4+15	Inner Brow Raiser, Brow Lowerer, Lip Corner Dipressor	1,2,3,4,5,6,18, 20,21,19
Surprise	1+2+5+26	Inner Brow Raiser, Outer Brow Raiser, Upper Lid Raiser, Jaw Drop	1,2,,3,4,5,6,7,8,10, 11,12,14
Fear	1+2+4+5+ 7+20+26	Inner Brow Raiser, Outer Brow Raiser, Brow Lowerer, Upper Lid Raiser, Lid Tightener, Lip Stretcher, Jaw Drop	1,2,3,4,5,6,7,8,9,10,11, 12,13,14,18,19,20,21
Anger	4+5+7+23	Brow Lowerer, Upper Lid Raiser, Lid Tightener, Lip Tightener	1,2,3,4,5,6,7,8,10,11,12,14, 18,19
Disgust	9+15+16	Nose Wrinkler, Lip Corner Depressor, Lower Lip Depressor	15,16,17,18,19

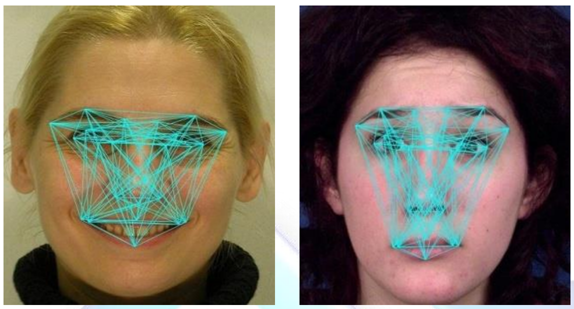

Fig. 4. Formation of triangulation using the *twenty one* salient landmark points.

Firstly, for a given triangle $t = \{(x_1, y_1), (x_2, y_2), (x_3, y_3)\}$ three side lengths a, b and c are calculated using the formulas 10 to 12.

$$a = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{10}$$

$$b = \sqrt{(x_2 - x_3)^2 + (y_2 - y_3)^2} \tag{11}$$

$$c = \sqrt{(x_1 - x_3)^2 + (y_1 - y_3)^2} \tag{12}$$

Next, the three angles A, B *and* C of the triangle t are calculated using the formulae 13 to 15.

$$\mathrm{A} = \cos^{-1}\left(\frac{\mathrm{b}^2+\mathrm{c}^2-\mathrm{a}^2}{2\mathrm{bc}}\right) \tag{13}$$

$$\mathrm{B} = \cos^{-1}\left(\frac{\mathrm{a}^2+\mathrm{c}^2-\mathrm{b}^2}{2\mathrm{ac}}\right) \tag{14}$$

$$\mathrm{C} = \cos^{-1}\left(\frac{\mathrm{a}^2+\mathrm{b}^2-\mathrm{c}^2}{2\mathrm{ab}}\right) \tag{15}$$

Next, the seven different shape factors $sf_1 to\ sf_7$ are calculated using the formulae 16 to 22.

$$sf_1 = \frac{max(A,B,C)}{\pi} \tag{16}$$

$$sf_2 = \frac{min(A,B,C)}{\pi} \tag{17}$$

$$sf_3 = \frac{max(A,B,C)-min(A,B,C)}{\pi} \tag{18}$$

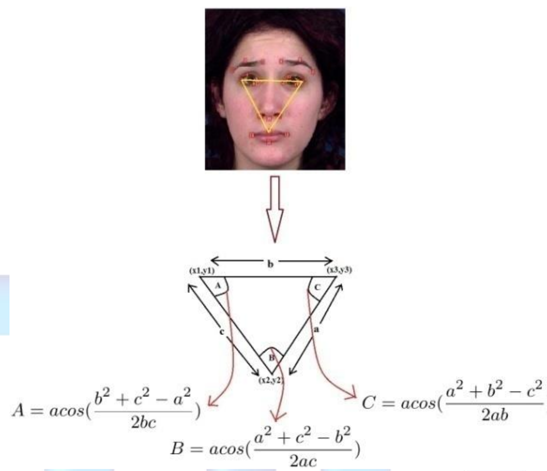

Fig. 5. Formation of a single triangle on a sample face image and calculation of thee angles of this triangle is demonstrated in the above figure.

$$sf_4 = \frac{max(A,B,C)+min(A,B,C)}{\pi} \tag{19}$$

$$sf_5 = \frac{max(A,B,C)-min(A,B,C)}{max(A,B,C)} \tag{20}$$

$$sf_6 = \frac{max(A,B,C)+min(A,B,C)}{\max(A,B,C)} \tag{21}$$

$$sf_7 = \frac{\pi-max(A,B,C)}{\pi} \tag{22}$$

These shape factors are the main features that are learned with MLP for the classification of expressions. The formation of a triangle and the calculation of the angles of the triangles are pictorially explained in figure 5.

E. Learning with MLP classifier

A brief introduction to the working principle of the MLP classifier is given in section III-B. The authors have used the Scaled Conjugate Gradient (SCD) back-propagation learning algorithm from [25]. The MLP classifier architecture consists of 4 layers: An input layer having input neurons equals the number of extracted features (2660), a hidden layer having

20 hidden neurons, an output layer having output neurons equals the number of output classes (6/7), and a SoftMax layer at the end to distribute the classification probability among the output classes. The input dataset is subdivided into a 70:15:15 ratios for training, testing, and validation respectively. Firstly, the MLP classifier is trained with each shape factor $[sf_i : i \in (1,7)]$ and the performance is measured. Then the top-performing two shape factors are merged and trained jointly. It is evident from section V that the effective selection of shape factors learning enhances the system accuracy.

F. Experimental Setup

We make use of AAM by Sagonas et. al. for generating face description by identifying 68 landmark points on the face [29]. Python-based Dlib implementation of [29] is used for computation of facial landmarks available on “https://github.com/davisking/dlib-models”. The network keeps training until there is no significant improvement in the validation accuracy to manage over-fitting. Sigmoid is used for activation. Network errors are computed as Root Mean Squared Error (RMSE). Architecture of the MLP is displayed in figure 6.

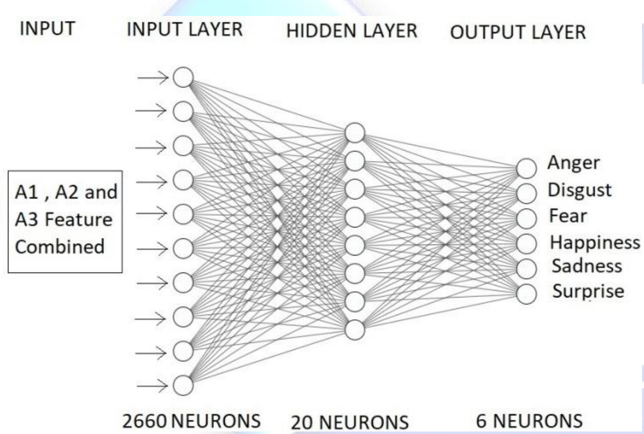

Fig.6. Architecture of the proposed neural network

V. EXPERIMENTAL RESULTS

The authors have tested the proposed methods in 4 well known benchmark facial expression databases which are CK+ [22], JAFFE [23], MMI [30] and MUG [1]. The number of images in each expression class for CK+, JAFFE, MMI, and MUG database is given on table V. It can be observed from the table that a total of *six* expressions are considered viz. Anger (ANG), Disgust (DIS), Fear (FER), Happiness (HAP), Sadness (SAD), and Surprise (SUR).

The details of the database with the obtained results are described in subsection V-A-V-D. Accuracy of the proposed method for *seven* different shape-factors on CK+, JAFFE, MMI,and MUG databases are shown on table V.

It can be observed from table V that following shape factors sf_2, sf_4, sf_5, sf_6 achieved highest accuracy on CK+, JAFFE, MMI and MUG databases with 99.68%, 99.03%, 96.85%, and 99.01% accuracy respectively.

TABLE II
THE NUMBER OF IMAGES IN EACH EXPRESSION CLASS FOR CK+, JAFFE, MMI AND MUG DATABASE

Expressions	CK+	JAFFE	MMI	MUG
Anger (ANG)	45	30	62	57
Disgust (DIS)	59	29	24	71
Fear (FER)	25	32	24	47
Happiness (HAP)	69	31	45	87
Sadness (SAD)	28	31	28	48
Surprise (SUR)	83	30	39	66
Total:	309	183	222	376

TABLE III
ACCURACY FOR DIFFERENT SHAPE FACTOR IN CK+, JAFFE, MMI, AND MUG DATABASES.

	sf_1	sf_2	sf_3	sf_4	sf_5	sf_6	sf_7
CK+	99.1	**99.6**	98.8	99.0	98.5	98.9	98.4
JAF	98.7	98.5	97.6	**99.0**	98.3	97.8	98.1
MMI	94.3	95.2	96.3	94.6	95.4	**96.3**	96.6
MUG	98.4	97.9	98.0	98.3	**99.0**	98.0	98.3

A. The Extended Chon-Kanade (CK+) Database

The CK+ database consists of 593 video sequences of posed and non-posed expressions from 123 subjects where each sequence starts from neutral expression to the peak expression [22]. The database has seven different expressions which are Anger (ANG), Contempt (CON), Disgust (DIS), Fear (FER), Happiness (HAP), Sadness (SAD), and Surprise (SUR). The models are of age in between 18 to 50 where 69% female, 81%, Euro-American, 13% Afro-American, and 6% of other groups. The 327 of the 593 sequences of the database have a proper expression tag. The authors have selected 309 peak images from each sequence of six basic expressions for training the model. The proposed system achieved 99.68% accuracy with (sf_2) in the overall dataset. The confusion matrix of for (sf_2) is given in table IV.

TABLE IV
CONFUSION MATRIX IN THE CK+ DATABASE.

	AN	DI	FE	HA	SA	SU
AN	100	0	0	0	0	0
DI	0	100	0	0	0	0
FE	0	0	100	0	0	1.19
HA	0	0	0	100	0	0
SA	0	0	0	0	100	0
SU	0	0	0	0	0	98.80

It can be observed from the table IV that Anger, Disgust, Fear, Happiness, and Sadness, and Surprise expression of the CK+ database are recognized with 100% accuracy, and Surprise expression is recognized with 98.80% accuracy. 19% fear expression is confused with Surprise.

B. The Japanese Female Facial Expression Database (JAFFE)

The JAFFE dataset contains 213 images from 10 subjects where each subject posed *tree* or *four* variations of *six* basic expressions and *one* neutral expression [23]. This is posed database and expression ratings are created by taking opinions from 97 students. The proposed system achieved the highest 99.03% accuracy with sf_5 feature in the overall dataset. The confusion matrix of the JAFFE dataset with sf_5 feature is given on table V.

TABLE V
CONFUSION MATRIX OF THE JAFFE DATABASE.

	AN	DI	FE	HA	SA	SU
AN	100	0	0	0	0	0
DI	0	100	0	0	0	0
FE	0	0	100	0	0	0
HA	0	0	0	100	0	0
SA	0	0	0	0	100	0
SU	0	0	3.22	0	0	96.77

It can be observed from the table V that Anger, Disgust, Fear, Happiness, Sadness expressions of the JAFFE database are recognized with 100% accuracy, Surprise expressions are recognized with 96.77% accuracy. We can also observe from table V that 3.22% of surprise expression is confused with Fear.

C. The Multimedia Imaging database (MMI)

The MMI dataset is a spontaneous facial expression database the dataset contains video, as well as still images [30]. The spontaneous expressions are recorded by simulating expression triggering videos and sounds in an audio- visual system. The authors have taken 222 well-expressed images from 484 images expressed by 5 subjects from PART- IV of the database for the experiment. This dataset contains only six basic expressions. The proposed system achieved 96.39% accuracy with (sf_6) feature in the overall dataset. The confusion matrix of MMI database with (sf_6) feature is given on table V-C.

TABLE VI
CONFUSION MATRIX OF THE MMI DATABASE.

	AN	DI	FE	HA	SA	SU
AN	98.3	0	0	0	0	1.6
DI	0	100	0	0	0	0
FE	0	0	100	0	0	0
HA	4.16	0	0	95.8	0	0
SA	2.2	0	0	2.2	95.5	0
SUR	7.1	0	0	0	0	92.5

Table V-C shows that Anger, Disgust, Fear, Happiness, Sadness, and Surprise expressions of the MMI database is recognized with 98.36%, 100%, 100%, 95.83%, 95.55%, and 92.58% respectively. The 1.63% of Anger expression is confused with Surprise, 4.16% of Happiness expression is confused with Anger, 2.22%, and 2.22% of sadness expression are confused with Anger and Happiness, and 7.14% of Surprise expression is confused with Anger.

D. The Multimedia Understanding Group Database (MUG)

The mug database is a posed facial expression video sequence database from 86 participants [1]. The subjects are 35 men and 51 women of Caucasian race and age in between20 to 35. The publicly available part of the dataset contains 401 images from 26 subjects. This dataset contains six basic expressions. The proposed system achieved 99.01% accuracy in the overall dataset with the (sf_5) feature. The confusion matrix is given on table V-D.

TABLE VII
CONFUSION MATRIX OF THE MUG DATABASE.

	AN	DI	FE	HA	SA	SU
AN	93.4	0	0	0	0	0
DI	0	98.6	0	0	0	0
FE	3.2	0	100	0	0	5.7
HA	0	1.3	0	100	0	0
SA	3.2	0	0	0	100	0
SUR	0	0	0	0	0	94.2

We can observe from table V-D that Anger, Disgust, Fear, Happiness, Sadness, Surprise of the MUG database are recognized with 93.44%, 98.61%, 100%, 100%, 100%, and 994.28% of accuracy respectively. Also, we can observe that 3.27% Sadness and Fear expression is confused with Anger. 5.71% of Fear expression is confused Surprise, 1.38% Happiness expression is confused with Disgust expression.

VI. DISCUSSION AND COMPARISONS

Expression-specific accuracy in CK+, JAFFE, MMI, and MUG database is compared in table. It can be observed from table that Anger expression achieved 100% accuracy in CK+ and JAFFE database. Disgust expression got 100% accuracy in CK+, JAFFE, and MMI database. Fear expression achieved 100% accuracy in all four databases. The Happiness and Sadness expression got 100% accuracy in CK+, JAFFE, and MUG database.

TABLE VIII
ACCURACY IN ANGER, DISGUST, FEAR, HAPPINESS, SADNESS ANDSURPRISE EXPRESSION CLASS IN CK+, JAFFE, MMI,AND MUG DATABASE.

	AN	DI	FE	HA	SA	SU
CK+	100	100	100	100	100	98.8
JAFFE	100	100	100	100	100	96.7
MMI	98.3	100	100	95.8	95.5	92.5
MUG	93.4	98.6	100	100	100	94.2

Type, No. of Images, No. of subjects, No. of Expressions, 10 Fold Cross-Validation and Overall Accuracy in CK+, JAFFE, MMI, and MUG database are given on table VI. We can observe from table VI that the CK+, MMI, and MUG database contains Posed and spontaneous expressions whereas the JAFFE database only contains posed expressions. All the databases have many sample images varying between 222 to 401. The CK+ database has 182 subjects which are the highest among all followed by 26 subjects on the MUG database, 10 subjects on the JAFFE database, and *five* subjects on the MMI database. All databases have *six* basic expressions. The CK+ database has Contempt expression as extra and JAFFE and MUG database have Neutral expression as an extra. The representative shape factor selected for CK+, JAFFE, MMI and MUG databases are sf_2, sf_4, sf_6 and sf_5 respectively.

TABLE IX
TYPE, NO. OF IMAGES, NO.OF SUBJECTS, NO.OF EXPRESSIONS AND OVERALL ACCURACY IN CK+, JAFFE, MMI AND MUG DATABASE.

DB Name	Type of Database	No. Image	No. Sub	No. Exp	SF	Cross Validation	Accuracy
CK+	Posed + Spontaneous	327	182	6 basic	sf_2	97.02%	99.68%
JAFFE	Posed	212	10	6 basic	sf_5	96.44%	99.03%
MMI	Posed + Spontaneous	222	5	6 basic	sf_6	93.47%	96.39%
MUG	Posed + Spontaneous	401	26	6 basic	sf_6	96.21%	99.01%

TABLE X
COMPARISON OF PERFORMANCE OF MLP WITH LDA, SVM, KKN AND RBF CLASSIFIER IN CK+, JAFFE, MMI, AND MUG DATABASES.

	CK+	JAFFE	MMI	MUG
LDA [14]	91.40	79.30	77.00	83.30
SVM [26]	87.80	76.10	81.50	84.80
KNN [10]	82.00	78.90	82.00	75.30
RBF [6]	96.4	97.0	97.0	97.7
MLP [25]	**99.68**	**99.03**	**96.39**	**99.01**

TABLE XI
CROSS-DATABASE VALIDATION AMONG CK+, JAFFE, MMI, AND MUG DATABASE

	CK+	JAFFE	MMI	MUG
CK+	98.80	63.82	77.71	48.67
JAFFE	76.44	97.85	56.40	39.88
MMI	83.28	69.86	95.25	33.66
MUG	50.44	39.51	15.22	97.63

TABLE XII
COMPARISON OF THE TRIANGLE SHAPE-FACTOR BASED METHOD WITH OTHER EXISTING RELEVANT WORKS.

Recent Works	**Accuracy CK+**	**Accuracy JAFFE**	**Accuracy MMI**	**Accuracy MUG**
Mollahosseini et al. [24]	93.20%		77.60%	
Jung et al. [18]	97.25%		70.24%	
Rahulamathavan et al. [27]		94.37%	95.24%	
Kotsia and Pitas [21]	86.70			
Saeed et al. [28]	83.15			
Kamarol et al. [19]	97.70			
Happy and Routray [15]	94.07			
Cheon et al. [7]	86.49			
Barman and Dutta [5]	98.70	97.60	94.30	**99.30**
Our Proposed Method	**99.68%**	**99.03%**	**96.39%**	**99.01%**

The highest accuracy achieved is 99.68% on the CK+ database followed by 99.03% on the JAFFE database, 99.01% on the MUG database, and 96.39% on the MMI database. We have also computed the 10-fold cross-validation accuracy on CK+, JAFFE, MMI, and MUG databases the proposed model achieved 97.02%, 96.44%, 93.47%, and 96.21% accuracy respectively.

The authors also have compared the performance of Multi-Layer Perceptron (MLP) with Support Vector Machine (SVM), K-Nearest Neighbours (KNN), Linear Discriminant Analysis (LDA), and Radial Basis Function (RBF) classifier. The results of that are shown on CK+, JAFFE, MMI, and MUG database on table VI.

It is evident from table VI that the proposed system achieved the highest accuracy with MLP classifier alone in all four databases.

The author has also performed Cross-Database validation in CK+, JAFFE, MMI, and MUG databases. The results of which are displayed on table. The training databases are presented row-wise and validation databases are provided column-wise in table.

The authors have also compared the results obtained with many other existing states of the artworks in the facial expression recognition domain which are discussed in section II. It can be observed from table VI that the proposed method achieved superior performance in the CK+, JAFFE, and MMI database with 99.68%, 99.03%, and 96.39% accuracy respectively. In the MUG database the proposed method performed with 99.01% accuracy where the state of the art accuracy is 99.30%.

It can be observed from table VI that MLP outperformed SVM, KNN, and LDA in all four databases. It is also evident from table VI, VI and VI that the proposed system achieved very good accuracy in all four databases and it achieved the highest accuracy in the MMI database.

VII. CONCLUSIONS

It can be concluded from this study firstly, the selection of salient landmarks helps to identify more prominent features and reduces the search space. Secondly, the introduction of seven novel shape-factors helps to identify the best feature descriptor. Thirdly learning in multiple state-of-the-art databases with MLP, KNN and SVM prove the robustness of the system. Lastly, The shape factors introduced in our study shows quite an effective extraction of facial features and the training with 7 different shape factors helps to identify best feature descriptor which is robust against different illumination condition, head and occultation in the face.

ACKNOWLEDGEMENTS

The authors want to take this opportunity to thank Dr. A. Delopoulos for providing access to the MUG database andalso to Prof. M Pantic for giving access to the MMI database. The authors also like to UGC for providing the NET-JRF fellowship (UGC-Ref. No.: 34347/(OBC)(NET-JAN 2017)) to the first author of this paper.

REFERENCES

[1] Niki Aifanti, Christos Papachristou, and Anastasios Delopoulos. The mug facial expression database. In *11th International Workshop on Image Analysis for Multimedia Interactive Services WIAMIS 10*, pages 1–4. IEEE, 2010.

[2] Asit Barman and Paramartha Dutta. Facial expression recognition using distance and shape signature features. *Pattern Recognition Letters*, 2017.

[3] Asit Barman and Paramartha Dutta. Facial expression recognition using distance and texture signature relevant features. *Applied Soft Computing*,77:88–105, 2019.

[4] Asit Barman and Paramartha Dutta. Influence of shape and texture features in facial expression recognition. *IET Image Processing*, 04 2019.

[5] ASIT BARMAN and Paramartha Dutta. Influence of shape and texture features in facial expression recognition. *IET Image Processing*, 2019.

[6] Sheng Chen, Colin FN Cowan, and Peter M Grant. Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Transactions on neural networks*, 2(2):302–309, 1991.

[7] Yeongjae Cheon and Daijin Kim. Natural facial expression recognition using differential-aam and manifold learning. *Pattern Recognition*, 42(7):1340–1350, 2009.

[8] Jeffrey F Cohn, Zara Ambadar, and Paul Ekman. Observer-based measurement of facial expression with the facial action coding system. *The handbook of emotion elicitation and assessment*, pages 203–221, 2007.

[9] Timothy F Cootes, Gareth J Edwards, and Christopher J Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (6):681–685, 2001.

[10] Scott Cost and Steven Salzberg. A weighted nearest neighbor algorithm for learning with symbolic features. *Machine learning*, 10(1):57–78, 1993.

[11] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 886–893. IEEE, 2005.

[12] Hong-Bo Deng, Lian-Wen Jin, Li-Xin Zhen, Jian-Cheng Huang, et al. A new facial expression recognition method based on local gabor filter bank and pca plus lda. *International Journal of Information Technology*,11(11):86–96, 2005.

[13] Paul Ekman and Wallace V Friesen. Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17(2):124, 1971.

[14] Yaqian Guo, Trevor Hastie, and Robert Tibshirani. Regularized linear discriminant analysis and its application in microarrays. *Biostatistics*, 8(1):86–100, 2007.

[15] SL Happy and Aurobinda Routray. Automatic facial expression recog- nition using features of salient facial patches. *IEEE transactions on Affective Computing*, 6(1):1–12, 2015.

[16] Simon Haykin. *Neural Networks and Learning Machines, 3/E*. Pearson Education India, 2010.

[17] Michael Jones and Paul Viola. Fast multi-view face detection. *MitsubishiElectric Research Lab TR-20003-96*, 3(14):2, 2003.

[18] Heechul Jung, Sihaeng Lee, Junho Yim, Sunjeong Park, and Junmo Kim. Joint fine-tuning in deep neural networks for facial expression recognition. In *Proceedings of the IEEE international conference on computer vision*, pages 2983–2991, 2015.

[19] Siti Khairuni Amalina Kamarol, Mohamed Hisham Jaward, Heikki Kälviäinen, Jussi Parkkinen, and Rajendran Parthiban. Joint facial expression recognition and intensity estimation based on weighted votes of image sequences. *Pattern Recognition Letters*, 92:25–32, 2017.

[20] Siti Khairuni Amalina Kamarol, Mohamed Hisham Jaward, Jussi Parkki- nen, and Rajendran Parthiban. Spatiotemporal feature extraction for facial expression recognition. *IET Image Processing*, 10(7):534–541, 2016.

[21] Irene Kotsia and Ioannis Pitas. Facial expression recognition in image sequences using geometric deformation features and support vector machines. *IEEE transactions on image processing*, 16(1):172–187, 2006.

[22] Patrick Lucey, Jeffrey F Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, pages 94–101. IEEE, 2010.

[23] Michael Lyons, Shigeru Akamatsu, Miyuki Kamachi, and Jiro Gyoba. Coding facial expressions with gabor wavelets. In *Proceedings Third IEEE international conference on automatic face and gesture recogni- tion*, pages 200–205. IEEE, 1998.

[24] Ali Mollahosseini, David Chan, and Mohammad H Mahoor. Going deeper in facial expression recognition using deep neural networks. In *2016 IEEE Winter conference on applications of computer vision (WACV)*, pages 1–10. IEEE, 2016.

[25] Martin Fodslette Møller. A scaled conjugate gradient algorithm for fast supervised learning. *Neural networks*, 6(4):525–533, 1993.

[26] Kristiaan Pelckmans, Johan AK Suykens, T Van Gestel, J De Brabanter, L Lukas, B Hamers, B De Moor, and J Vandewalle. Ls-svmlab: a matlab/c toolbox for least squares support vector machines. *Tutorial. KULeuven-ESAT. Leuven, Belgium*, 142:1–2, 2002.

[27] Yogachandran Rahulamathavan, Raphael C-W Phan, Jonathon A Cham- bers, and David J Parish. Facial expression recognition in the encrypted domain based on local fisher discriminant analysis. *IEEE Transactions on Affective Computing*, 4(1):83–92, 2012.

[28] Anwar Saeed, Ayoub Al-Hamadi, Robert Niese, and Moftah Elzobi. Frame-based facial expression recognition using geometrical features. *Advances in Human-Computer Interaction*, 2014:4, 2014.

[29] Christos Sagonas, Epameinondas Antonakos, Georgios Tzimiropoulos, Stefanos Zafeiriou, and Maja Pantic. 300 faces in-the-wild challenge: Database and results. *Image and vision computing*, 47:3–18, 2016.

[30] Michel Valstar and Maja Pantic. Induced disgust, happiness and surprise: an addition to the mmi facial expression database. In *Proc. 3rd Intern. Workshop on EMOTION (satellite of LREC): Corpora for Research on Emotion and Affect*, page 65, 2010.

[31] Siyue Xie, Haifeng Hu, and Ziyu Yin. Facial expression recognition using intra-class variation reduced features and manifold regularisation dictionary pair learning. *IET Computer Vision*, 12(4):458–465, 2017.

[32] Zhengyou Zhang, Michael Lyons, Michael Schuster, and Shigeru Aka- matsu. Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron. In *Proceed- ings Third IEEE International Conference on Automatic face and gesture recognition*, pages 454–459. IEEE, 1998.