Landmark triangulation-induced Altitude Signature for change detection of human emotion from face image sequence

Md Nasir[1] · Paramartha Dutta[1] · Avishek Nandi[1]

Abstract

Video-based facial expression recognition is a potential alternative for detecting the transitional processes of human emotion. In this paper, we present the landmark triangulation-induced altitude signatures to depict gradual changes in human emotion from face video frames. In our proposed approach, we consider the geometry-based triangulation technique which results from the triangles formed by important landmark points on face images. In this work, we measure three altitudes from all triangles and generate three different variations of altitude signature viz. AS1, AS2 and AS3 for the classification of six basic emotions viz. anger (AN), disgust (DI), fear (FE), happiness (HA), sadness (SA) and surprise (SU) in several ways. The performance capability of our proposed recognition system is tested on three benchmark face video datasets: Extended Cohn–Kanade (CK+), M&M Initiative (MMI), and Multimedia Understanding Group (MUG) through the application of Multilayer Perceptron (MLP) classifier and also validated by tenfold cross-validation method. Experimental results, 98.77%, 93.06%, and 98.75% on CK+, MMI, and MUG, respectively, are found very impressive. System efficacy is vindicated by showing the superiority of the proposed technique over the other existing techniques.

1 Introduction

Analysis of face expression is becoming an effective alternative for human behavior understanding because of which it has been able to attract enormous attention to relevant researches in the recent era. Accordingly, spreading the impact of affective computations and its many application fields such as computer technology, security, and psychology could result in facial expression which provides enormous scope to the people for interaction with each other [1]. According to [2], in human interaction, fifty-five percentage of total information is contributed via facial expression of the persons concerned, thirty-eight percentage of information is contributed via vocal part and remaining is contributed verbally. These facts create a research interest in automatic facial expression recognition from face images. In our daily life, human facial expressions are observed due to expression shrinkage of facial muscles, resulting in changes in facial components like eyebrows, nose, eyes, lips, and jaws. Such changes of facial components may involve an impact on both of geometrical structure and surface of the face [3]. Classification of emotion is an identification of movement of face components and facio-muscular motion [4]. These deformations of facial components are reflected in different facial expression classes in terms of emotions. Ekman and Friesen [5] did identify six universal human emotions. These emotions, however, may vary according to the human cultures, ethnicity, etc. Identification of appropriate facial features may ensure the robustness of the facial expression recognition system (FERS). The optimal feature extraction technique should provide sufficient features to minimize the dissimilarity within the same class of expressions and maximize the dissimilarity between different classes of expression [6]. The finest recognizer may be a failure to recognize the emotion with higher accuracy due to insufficient features of high discrimination capability. Capturing the displacement of face

✉ Md Nasir
  nasir.vb@gmail.com

  Paramartha Dutta
  paramartha.dutta@gmail.com

  Avishek Nandi
  avisheknandi10@gmail.com

1   Visva-Bharati University, Santiniketan, WestBengal, India

points can recognize the movement of facial features. Most of the recognition techniques based on geometric features use an active appearance model (AAM) to find out the geometric location of facial points. Many techniques [7–9] used static face images for extracting geometric features of every basic facial expression individually. It is not adequate to provide sufficient information in recognizing the changes in emotion. In [10], the authors reported that there are two approaches for developing the system of emotion recognition. One is a static image-based approach, while the other is a motion image-based approach. The best representation of emotion through geometric features may not always ensure higher recognition accuracy. The fluctuation, on the contrary, is observed occasionally according to the performance capability of different classifiers. Various existing techniques such as [11–13] are found to have used different type of recognizers viz. support vector machine (SVM), Naïve Bayesian classifier (NB), and artificial neural Network (ANN), respectively, to distinguish one emotion from others.

Motivation: Facial feature extraction either from static images or image sequences can be estimated in two several procedures with one being a geometry-based procedure for finding shape features, while the other being an appearance-based procedure for extracting texture features [14,15]. In [16,17] and [18], authors have applied their feature extraction techniques on sequenced video frames to observe dynamic variations of emotion. For the present, we, however, focus on a spatio-geometric approach to deal with different kinds of basic expressions for recognition of their time-varying behavior. In this method, landmark points are identified to represent geometric structures of various face components [19]. Facial expression can be represented by showing the geometric relations among these different components. It is identified that static images are unable to capture the variations of those relationships due to the persistent gap of geometric information. On the contrary in the case of image sequences, several image frames from neutral to basic expression are available, and changes in locations of every landmark points over time may be observed via all these consecutive image frames. These concepts help us to motivate for working with video sequences so that we can identify the dynamic behavior of every individual expression framewise. We use the triangulation altitude signature as a geometric feature to establish the geometric relationships among facial components, and after that, we have tried to detect the variations in geometric relations over the emotional sequence.

Our Contribution: 1) Formation of triangles from every image frame of sequence: Facial feature points are taken into account covering the major facial components such as nose, eyes, eyebrows, and lips. Triangles are formed by connecting those feature points identified through the application of the active appearance model (AAM) on every image frame.

2) Altitude signature is computed by using the triangulation mechanism. Here, triangle altitude is calculated for every triangle generated from the image frame. This altitude signature is used to measure the dynamic relationships among various facial components over time, and they are considered as facial features. 3) After getting a dynamic representation of expression via normalized facial features, multilayer perceptron is introduced for identification of the human emotion changes.

The remaining sections of the article are organized as follows. Various relevant existing works are given in Sect. 2. Our proposed method is presented by Sect. 3. Results and discussion on our experimentation are described in Sect. 4. Comparison with other state-of-the-art methods is included in Sect. 5, In the end, the conclusion is drawn in Sect. 6.

2 Related works

Since it is very much important to understand the facial behavior of a human in their non-verbal communication, we need to focus on different activities of face areas when one expresses his emotion in person-to-person communication. So, there can be one among the most valuable options for interpreting the expressional meaning is the determination of suitable facial features that distinguish distinct emotional signals. There are a lot of works done on the identification of emotional signals through introducing several features. The work of Yaddaden et al. [20] has found geometry-based facial features are more efficient for the identification of emotions. That's why they include facial points tracking step into their work to generate two distinctive types of features (distance and variance) and applied those features into two separate recognizers such as K-Nearest Neighbors (KNN) and support vector machines (SVM). In [21], authors have found only eight facial points among sixty-eight points as suitable for geometric feature generation and performed their recognition task. Moreover, they tried to investigate the influence of feature point detection error on emotional signals and enhanced their system performance accordingly. Authors in [22] have proposed a new landmark-based geometric approach for the expression classification. For the enhancement of accuracy level, they used three feature selection methods (sequential backward feature selection, sequential forward feature selection, and principal component analysis). It reduces the dimension of feature sets which are extracted by taking Euclidean distances between pairs of landmarks. They implemented the recognition task through the application of the support vector machines classifier on both reduced sizes of feature and non-reduced size of the feature. In [23], authors reported a novel work for emotional expression recognition through several types of geometrical representation of feature set in triangular form. Important

face marker points are first identified by them. Later, triangles shapes are formed by those face markers to generate three different types of geometric features (area of the triangle, inscribed circle circumference, and inscribed circle area of a triangle). With the usage of several machine learning algorithms (K nearest neighbor, decision tree, probabilistic neural network, random forest, extreme learning machine, and support vector machine), authors examine the consequence of various extracted features in recognizing different emotional face sequences. After doing the cross-validation of all features, they found the efficient performance of the random forest classifier with the usage of the inscribed circle area of a triangle in classification. The study [24] stated that the traditional emotion recognition systems obtained bounding box in detecting the face regions which contains some noise information within it. As result, it affects the recognition accuracy. Therefore, authors used landmarks to extract the face regions. Afterward, texture information on extracted regions is computed by using local binary pattern (LBP). Here, pixel information is considered for pattern identification of landmark movement, and the corresponding histogram is calculated to be stored in the feature vector. The learning of this feature set is done by using a convolution neural network to get better classification results. Many traditional deep learning-based human emotion recognition approaches (convolutional neural network, deep belief network) focus only on the generation of deep feature from face pixel rather than focusing on the correlation between geometric locations of face landmarks. The authors in [25] presented a new deep learning-based framework to extract hybrid features that combine both pixel-level features and deep geometric features. They introduced spatial attention CNN for the invention of pixel information and performed series of long short-term memory network learning tasks to establish the geometric correlation between landmarks. The usage of a hybrid network manifests the holistic local feature extraction method to improve the discrimination power of features in recognition of emotional expression. The study [26] used geometric points in sequence database for measuring the displacement of points, lines, and the shape of the triangle over the emotional signals. They tried to enhance the utility of face points in recognizing the emotion by showing the accuracy for each of the three feature sets. Authors [27] have presented a framework where salient landmarks are used for shape and texture feature invention. The performance of those features is tested on an MLP classifier by measuring different statistical parameters on both individual and combined features. Similarly, they reported their other work on facial expression in the other article [28] where distance and texture features are enhanced with relevant landmarks on major face portions, thereafter features are applied into several classifiers for significant recognition performance. In paper [29], authors introduced an integrated framework for

emotion recognition from multiple face frames. Here, both dynamic and static texture features are reflected through integrating the local binary feature with Gabor multi-orientation fusion histogram. Ghimire et al. used histogram-oriented gradient feature-based classification technique where histogram features are extracted by dividing images into smaller sub-images [30]. Such features are applied to the extreme learning machine (ELM) for the recognition of images. Authors in [31] have chosen landmarks for localization of important patches on key portions of the face and generated hybrid features by considering shape and appearance features extracted from salient patches.

3 Proposed method

Our main aim is to identify the dynamic movements of every facial component viz. nose, eyes, eyebrows, and mouth. The possibility of capturing the dynamic changes of emotion is found in the video sequence. We considered only ten frames from every image sequence according to their perfect projection of landmark points on those face image frames. Active appearance model is employed on frames to detect the region of each facial component so that we can measure the geometric relations among them. Further, we can find out the time-varying behavior of each basic expression by capturing geometric profiles of expressions over the image frames. These geometric profiles are computed by applying a geometric-based approach which includes a triangulation mechanism to generate a triangle from landmark points. Here, we extracted all triangles computed from those important feature points located at major facial components. Such triangles are used to generate altitude signature as geometric features are fed into multilayer perceptron (MLP) to map the image sequences into several basic emotions. The several steps of our procedure are shown in Fig. 1.

3.1 Face components detection and feature points identification

In the recognition system, components of the face play an imperative responsibility in characterizing individual basic expression. Changing geometric shapes of those components on the face may cause reflection on the formation of facial expressions. Geometric points describe the shapes and location of those facial components. The structural changes of components depend on the movement of those points over the facial region. One of the most frequently used statistical approaches is active shape model (ASM) which is used in [32] to detect the component's shapes by localizing geometric points on face images. Another statistical approach is the active appearance model (AAM) [33]. The model is developed by considering statistical measurement between
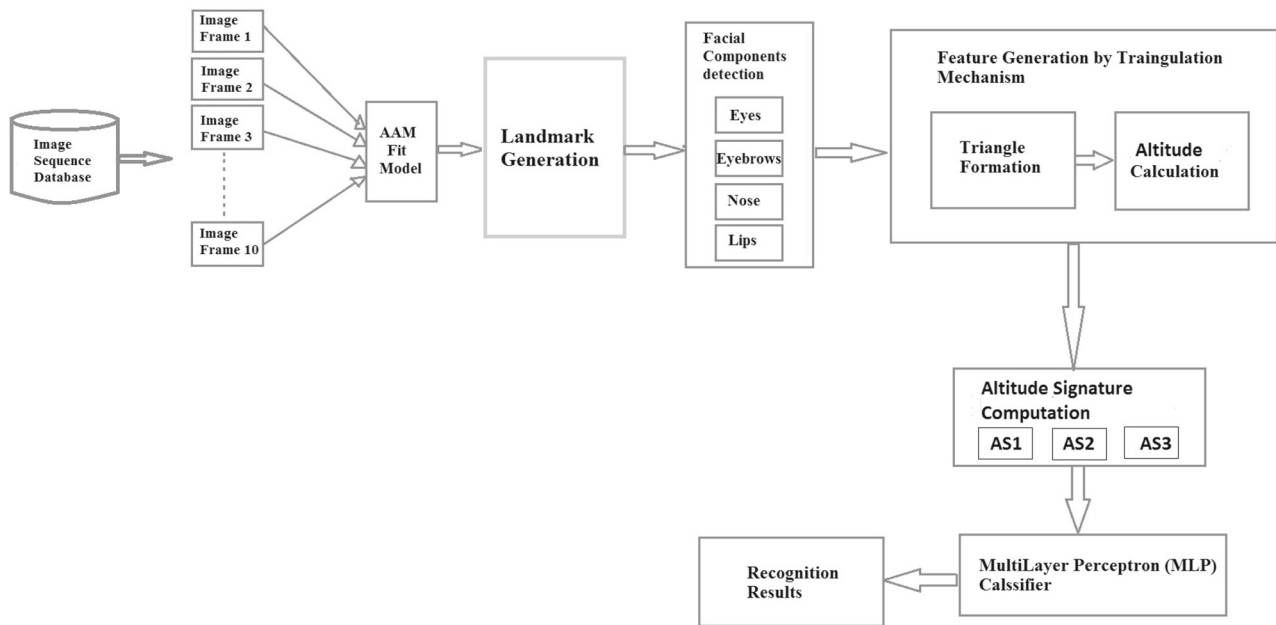
Fig. 1 Workflow of our proposed procedure

the shape and appearance features of deformable objects. This model follows the optimization approach which is initiated by taking a set of sample images labeled with landmark points. Each image is considered as a vector v_s of landmark points describing shape features. All images in the training set are repeatedly trained and approximated through the application of principal component analysis (PCA) to normalized data for making a shape model. The model is defined by Eq. 1

$$v_s = \bar{v}_s + O_s P_s \tag{1}$$

where $\bar{v}_s$ is mean shape, O_s and P_s are orthogonal and shape parameter. Similarly, statistical model of appearance feature is generated by considering normalized gray-level vector v_g for each training image. Here, mean $\bar{v}_g$ of gray-level images is computed by going through a recursive process. After that PCA is applied on normalized gray-level vectors to obtain the gray-level model. The model is defined Eq. 2

$$v_g = \bar{v}_g + O_g P_g \tag{2}$$

where O_g and P_g are orthogonal and gray-level parameter. AAM is better than ASM [34]. Here, we have applied a Fast-Sic [33] AAM fitting model to generate feature points on face images. A total of 68 points is found for every single image. Only 23 points we chose and these are taken as follows: eyebrows(6), left eye(4), right eye(4), nose(3), upper lip(3), and lower lip(3). It is found that these 23 points are very sensitive concerning facial expression as they are only the feature points to represent the region of major facial components. We have simply discarded other feature points from 68 points as

their movement of geometric position is not observed during the changes of facial expression. Figure 2 shows the detection of the facial components captured by crucial landmark points from image frames.

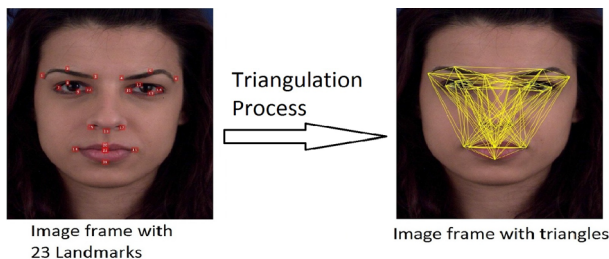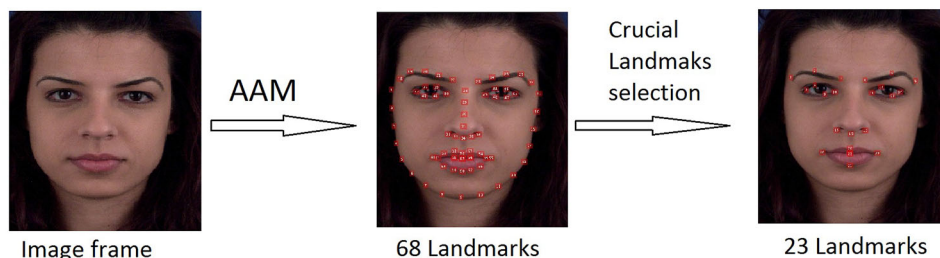### 3.2 Feature generation from landmark points

An effective feature extraction technique play a vital role in the emotion recognition system to develop high accuracy, reduce complexity, and increase the efficiency label of the system [35]. Our proposed method used a triangulation mechanism on image sequence to get triangle-altitude information-based geometric features from every image frame.

Altitude Signature:
The altitude of a triangle depends on the side which is called the base. In a triangle, first, select two of the landmark points and call the line segment joining these points the base. The perpendicular distance from the third landmark point of the triangle to the line containing the first two landmark points is called the altitude. In such a way, we got three altitudes of a particular triangle. These altitudes are relative to the bases of the triangle. Altitudes also depend on the area of the triangle, and this area is varied with the emotional changes in the transition from neutral to basic expression. Such changes of triangle shape area under consecutive sequential frames are observed by computing altitude signature.

Altitude Signature($AS1$): It first takes the difference between the maximum and minimum altitude of the triangle, and after that, it finds a summation of three altitudes.

Fig. 2 Facial Components detection with crucial Landmarks from single image frame

Image frame   AAM   68 Landmarks   Crucial Landmaks selection   23 Landmarks

Image frame with 23 Landmarks   Triangulation Process   Image frame with triangles

Fig. 3 Triangle generation from landmark points collected from facial components

Finally, altitude signature $AS1$ is formed by taking the ratio of altitude difference and altitude summation.

Altitude Signature($AS2$): The ratio between maximum altitude and total summation of three altitudes.

Altitude Signature($AS3$): The ratio between minimum altitude and total summation of three altitudes.

Figure 3 shows the triangle generation from landmark points collected from facial components. We have taken into account ten image frames for every sequence having 23 important landmark points of each. For each frame, $n = 1771$ triangles are constructed by connecting geometric locations for every possible combination of three landmark points out of 23 landmark points, and besides this, all three altitudes of those triangles are computed.

Let, $(a_l, b_l, c_l, A_l, B_l, C_l, h1_l, h2_l, h3_l)_{i,j,k}^m$ denotes three sides, three angles and three altitudes of triangle formed by i, j, k, three landmark points (as corner points) in l^{th} frame in m^{th} sequence. Sides are computed by using Euclidean distance formula described in Eqs. 3,4 and 5, angles are calculated by Eqs. 6,7 and 8, and altitudes are computed by Eqs. 9,10 and 11.

$$a_l = \sqrt{(y_k - y_j)^2 + (x_k - x_j)^2} \tag{3}$$

$$b_l = \sqrt{(y_k - y_i)^2 + (x_k - x_i)^2} \tag{4}$$

$$c_l = \sqrt{(y_j - y_i)^2 + (x_j - x_i)^2} \tag{5}$$

$$A_l = \cos^{-1}(\frac{b_l^2 + c_l^2 - a_l^2}{2 \times b_l \times c_l}) \tag{6}$$

$$B_l = \cos^{-1}(\frac{a_l^2 + c_l^2 - b_l^2}{2 \times a_l \times c_l}) \tag{7}$$

$$C_l = \cos^{-1}(\frac{a_l^2 + b_l^2 - c_l^2}{2 \times a_l \times b_l}) \tag{8}$$

$$h1_l = b_l \times \sin^{-1} A_l \tag{9}$$

$$h2_l = b_l \times \sin^{-1} C_l \tag{10}$$

$$h3_l = c_l \times \sin^{-1} A_l \tag{11}$$

Now, three different variations of altitude signatures are computed by Eqs. 12,13 and 14.

$$(AS1_l)_{i,j,k}^m = \frac{max(h1_l, h2_l, h3_l) - min(h1_l, h2_l, h3_l)}{h1_l + h2_l + h3_l} \tag{12}$$

$$(AS2_l)_{i,j,k}^m = \frac{max(h1_l, h2_l, h3_l)}{h1_l + h2_l + h3_l} \tag{13}$$

$$(AS3_l)_{i,j,k}^m = \frac{min(h1_l, h2_l, h3_l)}{h1_l + h2_l + h3_l} \tag{14}$$

Finally, Eqs. 15, 16 and 17 define feature vectors in three different ways to represent each single video face sequence.

$$(AS1)_{i,j,k}^m = [(AS1_0)_{i,j,k}^m, (AS1_1)_{i,j,k}^m, (AS1_2)_{i,j,k}^m, \ldots, (AS1_n)_{i,j,k}^m] \tag{15}$$

$$(AS2)_{i,j,k}^m = [(AS2_0)_{i,j,k}^m, (AS2_1)_{i,j,k}^m, (AS2_2)_{i,j,k}^m, \ldots, (AS2_n)_{i,j,k}^m] \tag{16}$$

$$(AS3)_{i,j,k}^m = [(AS3_0)_{i,j,k}^m, (AS3_1)_{i,j,k}^m, (AS3_2)_{i,j,k}^m, \ldots, (AS3_n)_{i,j,k}^m] \tag{17}$$

Here, n denotes the frame index in each sequence. For our work, n=10. So size of each feature vector is $10 \times 1771 = 17710$. The computation of three altitudes from a triangle on a single frame is shown in Fig. 4.

3.3 Facial expression classification

In our system, multilayer perceptron (MLP) is employed as a classifier for discrimination of facial expression into different basic expression classes. The network stability depends on the error estimation on the neuron. So the selection of the number of neurons in each network layer has imperative responsibility in the improvement of performance accuracy. In particular, the hidden neurons have more influence in

Fig. 4 Illustration for computation of three altitudes from triangle on single frame
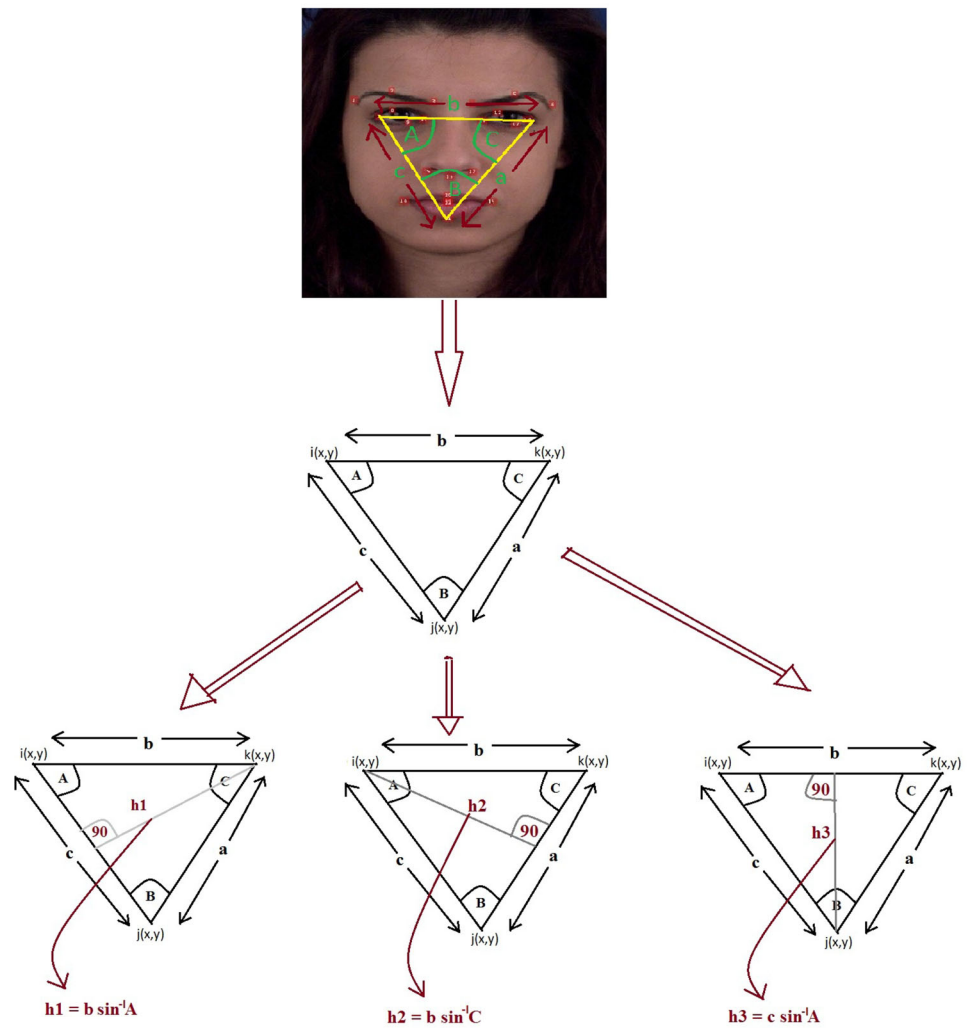
controlling the network error than the other neurons in the network. A very higher number or a very lower number of hidden neurons may cause the overfitting or underfitting problem respectively. In view of this, we used the trial rule for the fixing of the number of hidden neurons. We repeatedly examined the network with the selection of several numbers of hidden neurons to get minimal error. Based on this trial rule, we found the best selection of ten hidden neurons which obtain the higher accuracy and find the network stability. On the contrary, we selected 17710 numbers of input neurons because this number is equivalent to the size of the input feature vector, and six output neurons are used because we want to identify six basic emotions separately. Extracted Altitude Signature on an image sequence is used as input of this classifier containing $10 \times 1771 = 17710$ number of input neurons in the input layer. We have organized the signature matrix by combining ten rows into a single row containing 17710 height-length information and put it into MLP. As MLP is a feedforward artificial neural network, input signals are forwarded to the next hidden layer of ten neurons, and finally, the signals coming from hidden layers are fed into the output layer of six neurons to classify the extracted features into six basic expressions. This process is continued until there is no such signature matrix is to be organized in the training dataset. Here, network is trained with the scaled conjugate gradient backpropagation process. The training process is stopped when the network achieves minimum error computed by the mean square error formula, and it also adjusts weight and bias values during the training period. Algorithm 1 shows every step for the classification with MLP.

Algorithm 1 MLP classification algorithm using Altitude Signature

---

Require: Image video sequence
Ensure: Recognition of transforming emotion
1: **for** each sequence $s \in S$ **do**
2:    **for** each frame $f \in s$ **do**
3:       AAM fitting Model is applied on f to extract landmark points
4:       Generate all possible triangles T
5:       **for** each triangle $t \in T$ **do**
6:          compute components of triangle $(a_f, b_f, c_f, A_f, B_f, C_f, h1_f, h2_f, h3_f)_t$, where $(a_f, b_f, c_f) \rightarrow$ three sides, $(A_f, B_f, C_f) \rightarrow$ three angles and $(h1_f, h2_f, h3_f) \rightarrow$ three altitudes of triangle t.
7:          Compute Altitude Signatures such as $(AS1_f)_t^s$, $(AS2_f)_t^s$, $(AS3_f)_t^s$ and store it into vector $AS1_{f,t}^s$, $AS2_{f,t}^s$, $AS3_{f,t}^s$, respectively.
8:       **end for**
9:       Generate final sequence vectors in frame wise such as: $AS1^s = AS1_f^s \cup AS1_{f+1}^s$, $AS2^s = AS2_f^s \cup AS2_{f+1}^s$, $AS3^s = AS3_f^s \cup AS3_{f+1}^s$
10:    **end for**
11:    Each sequence vector is applied on MLP classifier.
12:    Get the recognition result.
13: **end for**

Algorithm 1 involves three steps: scanning of frame one by one, generation of the triangle using crucial landmarks, and computation of triangle components followed by extraction of altitude signature. Here, the computation of the triangle's component has the complexity of $O(1)$. There is a total t number of triangles in each frame, so the total computation complexity for the components is $O(t)$. Similarly, computation of the altitude information has the same complexity of $O(t)$. Every t number of the triangle is generated in every single scan of frame in a particular sequence. Since $|s|$ number of frames is available in each sequence, hence computational time complexity for extracting altitude signature becomes $O(|s|) \times (O(t) + O(t)) = O(|s|t)$. As $|s| = 10$, $t = 1771$, and $t >> |s|$, the total complexity tends to $O(t)$. In [20], authors have extracted two geometric features (distance and variance) for recognition of emotion separately. In their study, number of frames in each sequence is $|s| = 18$, and in each frame number of distance value is $d = 2278$ and number of variance value is $v = 2278$. So time complexity for extracting both features becomes $O(d)$ and $O(v)$. Hence, the time complexity in our case is $O(t) < O(d)$ or $O(t) < O(v)$. On the other side, the study [26], used total $T = 22100$ number of triangle's information for the recognition task. So the time complexity for this triangle based feature extraction is $O(T)$, which is higher than our proposed feature extraction complexity , $i.e.$, $O(T) > O(t)$

4 Experimental result and discussion

We have experimented our proposed feature extraction technique on popular benchmark video data: CK+ [36], MUG [37] and MMI [38] to discriminate human emotion from each other. Figure 5 visualizes the sequence of expressions is available in above mentioned three video datasets. Before applying the feature into MLP, we have arranged all databases by dividing as follows: 70% sequences are used for network training, 15% sequences are taken to validate the network performance, and the remaining 15% sequences are reserved for testing purposes. To measure the efficiency of our proposed technique, we have derived three other performance parameters from confusion matrix for all databases, and these are calculated by following Recall $= \frac{TP}{TP+FN}$, Precision $= \frac{TP}{TP+FP}$ and $F-\text{score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$. Here, TP(True Positive) = number of image sequences are correctly identified, FN(False Negative) = number of image sequences is incorrectly identified with other classes, and FP(False Positive) = number of image sequences are incorrectly identified as an actual class in the database. Details analysis on the results of all three databases is shown in sections 3.1,3.2 and 3.3. Several datasets have a different number of sequences for every emotion, and it is given in Table 1.

4.1 CK+ Database

The dataset consists of the mature face (posed and non-posed) from 210 persons. The ages of them are 18 to 50 years and most of them are female, others are from Euro-American, Afro-American, and other groups. In our experiment, 327 image sequences out of 593 sequences are considered for emotion recognition as those sequences are found to meet the criteria for seven different emotions including six basic and one contempt (CON). Confusion matrix computed from CK+ [36] dataset shown in Table 2. Here, altitude signature AS1 identifies anger, fear, happiness, sadness with 100% accuracy. The lowest accuracy 83.33% is achieved in the case of contempt expression but two sequences of contempt class are misclassified with sadness and one sequence of contempt is mismatched with anger. Altitude signature AS2 can detect the transition of contempt and fear emotion with 100% accuracy but lower accuracy of 92.85% is detected in case of sadness emotion. Signature AS2 perfectly recognizes 44 sequences of anger, 57 sequences of disgust, and 68 sequences of happiness but two sequences of disgust and two sequences of sadness are mismatched with anger. Signature AS3 identifies 100% accuracy on anger, contempt, and surprise emotion,

Fig. 5 Visualization of three datasets with sample image sequences
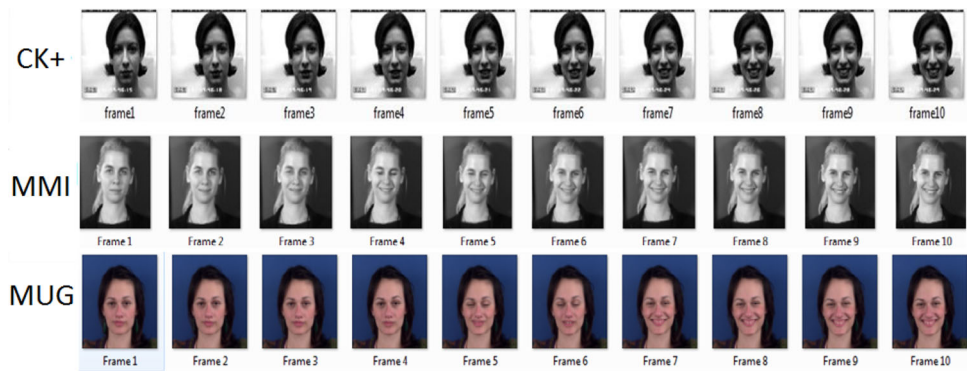
Table 1 Sequence number/emotion in three datasets

Dataset	Anger	Contempt	Disgust	Fear	Happiness	Sadness	Surprise	Total
CK+	45	18	59	25	69	28	83	327
MMI	31	–	32	28	42	28	41	202
MUG	149	–	117	150	107	133	145	801

Table 2 Confusion matrix on CK+

	AN	CON	DI	FE	HA	SA	SU
Altitude Signature AS1							
AN	45	0	0	0	0	0	0
CON	1	15	0	0	0	2	0
DI	1	0	58	0	0	0	0
FE	0	0	25	0	0	0	0
HA	0	0	0	0	69	0	0
SA	0	0	0	0	0	28	0
SU	0	0	0	1	0	0	82
Altitude Signature AS2							
AN	44	0	1	0	0	0	0
CON	0	18	0	0	0	0	0
DI	2	0	57	0	0	0	0
FE	0	0	0	25	0	0	0
HA	0	0	0	0	68	0	1
SA	2	0	0	0	0	26	0
SU	0	0	0	1	0	0	82
Altitude Signature AS3							
AN	45	0	0	0	0	0	0
CON	0	18	0	0	0	0	0
DI	1	0	58	0	0	0	0
FE	0	0	0	24	1	0	0
HA	0	1	0	0	68	0	0
SA	1	0	0	0	0	27	0
SU	0	0	0	0	0	0	83

and lower accuracy of 96% is found on recognition of fear emotion. Signatures AS1, AS2, and AS3 are able to get the overall accuracy of 98.47%, 97.85%, and 98.77% on the CK+ dataset, respectively. Here, signature AS3 shows the higher

performance ability on this dataset. Table 6 shows another measurement of performance accuracy.

4.2 MUG Database

In the MUG [37] database, all images are collected from 52 several subjects, and they are labeled with six basic emotions. We have utilized 801 image sequences from the MUG dataset to categorize them into basic expression labels. Among them, 561, 120, and 120 sequences are considered for training, validation, and testing purposes. Table 3 shows the confusion matrix on MUG, respectively. Signature AS1 is found to achieve 100% accuracy on sadness emotion and lower accuracy 97.43% is found on disgust expression. A total of 146 expression profiles of anger, 149 expression profiles of fear, 105 of happiness, and 144 of surprise are correctly classified by the signature AS1. Among all anger expressions, only one is misclassified with disgust and two are confused with happiness. Two sequences of happiness, one sequence of fear, three sequences of disgust, and one sequence of surprise are confused with anger, disgust, surprise, and fear, respectively. Signature AS2 is found to compute 100% accuracy on recognition of fear emotion and lower accuracy of 97.19% is obtained on happiness. Here, four sequences of anger, two sequences of disgust, three sequences of happiness are misclassified with happiness, surprise, and anger, respectively. Two expression profiles of sadness are confused with anger and fear. Two profiles of surprise are mismatched with disgust and sadness. Signature AS3 identifies fear emotion with 100% recognition accuracy but here also it gets lower recognition results of 97.19% on happiness. More than 99% accuracy is observed on disgust and surprise emotion by the signature AS3. The overall accuracy of 98.75%, 98.37%,

Table 3 Confusion matrix on MUG dataset

	AN	DI	FE	HA	SA	SU
Altitude Signature AS1						
AN	146	1	0	2	0	0
DI	0	114	0	0	0	3
FE	0	1	149	0	0	0
HA	2	0	0	105	0	0
SA	0	0	0	0	133	0
SU	0	0	1	0	0	144
Altitude Signature AS2						
AN	145	0	0	4	0	0
DI	0	115	0	0	0	2
FE	0	0	150	0	0	0
HA	3	0	0	104	0	0
SA	1	0	1	0	131	0
SU	0	1	0	0	1	143
Altitude Signature AS3						
AN	146	1	0	2	0	0
DI	0	116	0	0	0	1
FE	0	0	150	0	0	0
HA	3	0	0	104	0	0
SA	2	0	0	0	131	0
SU	0	0	0	0	1	144

Table 4 Confusion matrix on MMI

	AN	DI	FE	HA	SA	SU
Altitude Signature AS1						
AN	29	1	0	0	1	0
DI	2	29	0	0	1	0
FE	4	0	22	1	0	1
HA	0	0	0	42	0	0
SA	2	1	0	0	25	0
SU	1	0	0	0	0	40
Altitude Signature AS2						
AN	30	0	0	0	0	1
DI	3	27	0	1	1	0
FE	0	0	25	1	0	2
HA	0	0	1	41	0	0
SA	2	1	0	0	25	0
SU	0	0	0	1	0	40
Altitude Signature AS3						
AN	28	0	0	1	2	0
DI	0	32	0	0	0	0
FE	0	0	23	0	3	2
HA	1	0	0	41	0	0
SA	3	0	0	0	25	0
SU	2	0	0	0	0	39

98.75% are obtained on the MUG dataset by the signatures AS1, AS2, and AS3, respectively. From these results, it is found that AS1 and AS3 performed well on this dataset. Table 6 shows another measurement of performance accuracy.

4.3 MMI database

Our proposed signatures are also derived from the MMI [38] sequence database to perform the recognition task in our system. It is a very challenging task to deal with this database as it involves both frontal and side-view images in a single frame. We have used only frontal images for our consideration and discarded all those image sequences which have no emotion label. After preprocessing the database, we got 202 videos with six emotion labels from 236 videos are available in the MMI database. Such sequence contains expressional profile including following phases: neutral, onset, apex, and offset. The detailed recognition result of this database is shown in Table 4. Altitude signature AS1 recognizes happiness emotion with 100% accuracy, whereas lower accuracy of 78.57% is found in recognition of fear emotion. Two anger emotions are misclassified with disgust and sadness but a total of 29 out of 31 anger sequences are perfectly recognized by the signature AS1. Two sequences of sadness, two sequences of disgust, and one sequence of surprise are confused with anger class. Twenty-five sequences of sadness, 29 sequences

of disgust, and 40 sequences of surprise are identified accurately. Altitude Signature AS2 achieves higher accuracy of 97.61% in recognition of happiness, and lower accuracy of 84.37% is obtained in disgust emotion. Here, this signature recognizes 30 expression profiles of anger, 25 of fear, 25 of sadness, and 40 of surprise perfectly. Altitude Signature AS3 obtains 100% accuracy in disgust. Lower accuracy of 82.14% is achieved in fear. It identifies 28 anger, 41 happiness, 25 sadness, and 39 surprises accurately. Altitude Signature AS1, AS2, and AS3 achieve overall accuracy of 92.57%, 93.06%, and 93.06% on the MMI dataset, respectively. Here, we found Altitude Signature AS1 and AS2 as better signatures to perform recognition tasks on this dataset. We also computed another measurement of performance accuracy from the MMI confusion matrix shown in Table 6.

Table 5 evaluates the comparison task for the proposed three signatures over three datasets.

4.4 Cross-validation

The most popular procedure N-fold cross-validation is used on all three video datasets for measuring the performance skill of our system. In this process, the dataset is randomly divided into N number of different groups. In our system, we used $N = 10$ and that is why it is becoming a tenfold cross-validation procedure. For each fold, the dataset is divided into

Table 5 Comparison of recognition results (%) of three signatures over three datasets

Dataset	Feature	AN	CON	DI	FE	HA	SA	SU	Avg. Acc.
CK+	AS1	100	83.33	98.3	100	100	100	98.79	98.47
	AS2	97.77	100	96.61	100	98.55	92.85	98.79	97.85
	AS3	100	100	98.3	96	98.55	96.42	100	98.77
MUG	AS1	97.98	–	97.43	99.33	98.13	100	99.31	98.75
	AS2	97.31	–	98.29	100	97.19	98.49	98.62	98.37
	AS3	97.98	–	99.14	100	97.19	98.49	99.31	98.75
MMI	AS1	93.54	–	90.62	78.57	100	89.28	97.56	92.57
	AS2	96.77	–	84.37	89.28	97.61	89.28	97.56	93.06
	AS3	90.32	–	100	82.14	97.61	89.28	95.12	93.06

Table 6 Recall, Precision and F-score on CK+,MUG and MMI dataset

			AN	CON	DI	FE	HA	SA	SU
CK+	AS1	Recall	1	0.83	0.98	1	1	1	0.99
		Precision	0.96	1	1	0.96	1	0.93	1
		F-score	0.98	0.91	0.99	0.98	1	0.96	0.99
	AS2	Recall	0.97	1	0.96	1	0.98	0.92	0.98
		Precision	0.91	1	0.98	0.96	1	1	0.98
		F-score	0.93	1	0.96	0.97	0.98	0.95	0.98
	AS3	Recall	1	1	0.98	0.96	0.98	0.96	1
		Precision	0.95	0.94	1	1	0.98	1	1
		F-score	0.97	0.96	0.98	0.97	0.98	0.97	1
MUG	AS1	Recall	0.98	–	0.97	0.99	0.98	1	0.99
		Precision	0.99	–	0.98	0.99	0.98	1	0.98
		F-score	0.98	–	0.97	0.99	0.98	1	0.98
	AS2	Recall	0.97	–	0.98	1	0.97	0.98	0.98
		Precision	0.97	–	0.99	0.99	0.96	0.99	0.98
		F-score	0.97	–	0.98	0.99	0.96	0.98	0.98
	AS3	Recall	0.97	–	0.99	1	0.97	0.98	0.99
		Precision	0.96	–	0.99	1	0.98	0.99	0.99
		F-score	0.96	–	0.99	1	0.97	0.98	0.99
MMI	AS1	Recall	0.93	–	0.9	0.78	1	0.89	0.98
		Precision	0.76	–	0.94	1	0.98	0.93	0.98
		F-score	0.84	–	0.92	0.88	0.99	0.91	0.98
	AS2	Recall	0.96	–	0.84	0.89	0.97	0.89	0.97
		Precision	0.85	–	0.96	0.96	0.93	0.96	0.93
		F-score	0.9	–	0.89	0.92	0.94	0.92	0.94
	AS3	Recall	0.9	–	1	0.82	0.97	0.89	0.95
		Precision	0.82	–	1	1	0.97	0.83	0.95
		F-score	0.85	–	1	0.9	0.97	0.85	0.95

ten numbers of different groups. Here, one group is used as a testing sample, and the remaining groups are used as a training sample. In the same tactic, this process is performed for the remaining ninefolds so that each group can be tested once. After performing on this set up, we got ten different recognition accuracy but we have calculated average accuracy across the tenfolds to reduce the biases of the recognition results. Table 7 shows tenfold cross-validation accuracy over CK+ [36], MUG [37] and MMI [38] datasets and corresponding comparison graph is shown in Fig. 6.

Fig. 6 Corresponding comparison graph of Table 7
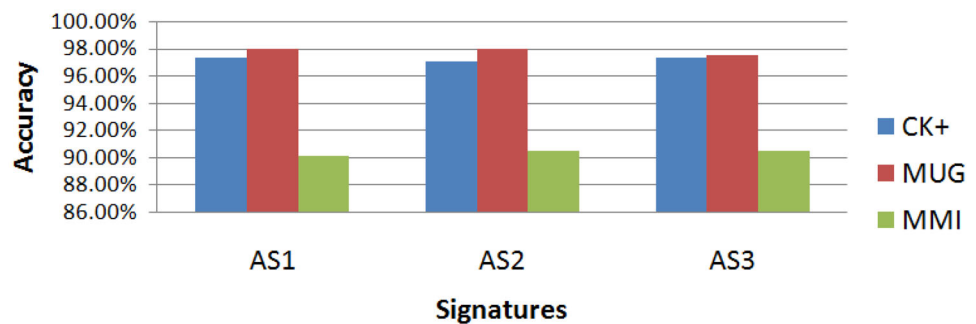
Table 7 Tenfold cross-validation accuracy of three Altitude Signatures on CK+, MUG and MMI

Dataset	Signature AS1	Signature AS2	Signature AS3
CK+	97.32%	97.04%	97.33%
MUG	97.92%	97.93%	97.49%
MMI	90.10%	90.41%	90.46%

5 Comparison with other state-of-the-art methods

The performance superiority of our proposed recognition system is evaluated by showing a comparison of the accuracy between our method and some existing methods. Triangle Altitude information is used as a geometric feature for MLP classification. Our proposed signature AS1, AS2, AS3 gives superior recognition performance of 98.47%,97.85%, 98.77% in ck+ dataset, 92.57% , 93.06%, 93.06% in MMI dataset and 98.75% in MUG dataset, respectively. All comparisons are given in Table 8.

CK+: In the system, [20] has used a distance-based feature for performing recognition tasks on the CK+ image sequence database and has achieved 90% of overall accuracy using an SVM classifier. They have used a total of 309 image sequences with six basic expressions but in our proposed system total of 327 image sequences are taken with six basic expressions and one contempt expression. Here, all three signatures show higher accuracy than the system [20]. In [39], even though the system has used the same number of expressions to perform their classification task as our system used but overall accuracy of 83.90% in the CK+ dataset is achieved by the study [39] which is lower than the accuracy obtained by our proposed system. Authors in [26] used 52 landmarks points and extracted triangle's components through the utilization of those facial points. Maximum changes in emotional sequence, they found through the application of extracted triangle feature into their SVM classification-based system. The system in [26] obtained an average recognition rate of 97.80% on CK+, while it shows higher performance than other existing systems reported in Table 8 but the accuracy of the system [26] still is lower than the accuracy of our

all three signatures. Hence, the system [26] could not beat our system performance. On the other hand, the approach in [29] has found 100% accuracy on disgust and happiness with an overall accuracy of 95.80%. Though they beat our system for only disgust emotion, overall accuracy is still lower than our three signatures. Similarly, our signatures also outperforms two other methods [21,40] and it is described in Table 8. Authors in [21] extracted geometric feature from eight facial points but they found approximately 6% lower recognition accuracy by using it than the accuracy level (83.01%) obtained by using 68 facial points into feature extraction process. The study [40] has not utilized the feature information directly from feature points for the recognition of emotion. They tried to infer expressional states from the corresponding action units of 26 feature points. Authors in [40] claim that detection accuracy of action units significantly affects expression recognition, and they achieved 87.43% average accuracy on the CK+ dataset which is lower than the accuracy levels of our proposed three signatures.

MUG: In the case of the MUG dataset, it is observed from Table 8 that the signatures AS1 and AS2 are able to show better performance than the system [41] through achieving higher overall accuracy as well as obtaining higher results in recognition of individual six emotions. On the other hand, the approach in [26] has achieved 100% accuracy on anger, disgust, and happiness with an average classification rate of 95.50%. Here, though we are not able to reach 100% recognition accuracy for those three emotions, our proposed method presents higher overall accuracy than the method in [26].

MMI: Authors in [42] implemented their recognition approach on the MMI dataset and achieved 100% accuracy on sadness and surprise transition which is higher than our all signatures but the overall accuracy of 71.43% achieved by the system [42] is quite lower than the accuracy obtained in our approach. From Table 8, it is seen that signature AS1, AS2, and AS3 could outperform the method in [26] by showing better accuracies in both individual emotions as well as overall recognition. Authors in [43] have shown two approaches (KSOM-based classification and MLP-based classification) and achieved an overall accuracy of 93.53% and 72.15% respectively. Our all three signatures could not achieve the

Table 8 Results (%) comparison with other state-of-the-art system

Dataset	Methods	AN	CON	DI	FE	HA	SA	SU	Overall ACC.
CK+	[21]	80	–	83	72	100	64.2	98.7	83.01
	[29]	93.3	–	100	88	100	85.7	96.4	95.80
	[40]	66.67	–	91.02	80	97.08	80	96.88	87.43
	[26]	97.5	–	96.67	96	100	96.67	100	97.80
	[20]	91	–	97	87	99	90	97	90
	[39]	81	44	90	58	97	50	98	83.90
	AS1	100	83.33	98.3	100	100	100	98.79	98.47
	AS2	97.77	100	96.61	100	98.55	92.85	98.79	97.85
	AS3	100	100	98.3	96	98.55	96.42	100	98.77
MUG	[26]	100	–	100	85	100	90	98	95.50
	[41]	96.3	–	93.83	88.89	95.06	80	78	95.24
	AS1	97.31	–	98.29	100	97.19	98.49	98.62	98.75
	AS2	97.98	–	99.14	100	97.19	98.49	99.31	98.37
	AS3	93.54	–	90.62	78.57	100	89.28	97.56	98.75
MMI	[26]	70	–	80	70	85	73.33	85	77.22
	[42]	50	–	66.67	75	66.67	100	100	71.43
	[43]	98.3	–	88.7	90.2	90.5	88.7	98.1	93.53
	[43]	79.3	–	54.6	75.1	97.3	55.2	71.4	72.15
	AS1	93.54	–	90.62	78.57	100	89.28	97.56	92.57
	AS2	96.77	–	84.37	89.28	97.61	89.28	97.56	93.06
	AS3	90.32	–	100	82.14	97.61	89.28	95.12	93.06

accuracy level of 93.53% but they are found to outperform the MLP-based classification approach used in the study [43].

6 Conclusion

In this paper, we tried to generate spatiotemporal information from video sequences through the application of the triangulation mechanism. Analysis of landmark distortion on face plane is becoming one of the most potential alternatives for detecting gradual changes in human emotion, while evolving of facial expression starts from neutral to extreme expressional face. Only the single location information is not sufficient to analyze the deformation of face components because the components are geometrically connected with each other. For this reason, we used triangle shape collecting multiple landmark location information, and generated altitude signatures from triangles for the recognition of transformative emotion. Geometric relations among the face components during the deformation of expression over the video frames are well-identified by employing our proposed varieties of altitude signature into MLP recognizer. Experimental results on CK+, MUG, and MMI accomplished the superior ascertainment of our proposed features. Further, the overall performance of our proposed signatures is validated by showing impressive tenfold cross-validation accuracy as well as it is justified through the comparison of results with other existing methods.

References

1. Zabih R, Woodfill J (1994) Non-parametric local transforms for computing visual correspondence. In: European conference on computer vision, pp 151–158. Springer
2. Mehrabian A (2007) Nonverbal communication. New Brunswick
3. Tao J, Tan T (2005) Affective computing: a review. In: International conference on affective computing and intelligent interaction. Springer, pp 981–995
4. Beat F, Juergen L (2003) Automatic facial expression analysis: a survey. Pattern Recognit 36(1):259–275
5. Ekman P, Friesen Wallace V (1971) Constants across cultures in the face and emotion. J Personal Soc Psychol 17(2):124
6. Caifeng S, Shaogang G, McOwan Peter W (2009) Facial expression recognition based on local binary patterns: a comprehensive study. Image Vis Comput 27(6):803–816
7. Pantic M, Rothkrantz LJM (2004) Facial action recognition for facial expression analysis from static face images. IEEE Trans Syst Man Cybern Part B (Cybernetics) 34(3):1449–1461

8. Naveen Kumar HN, Jagadeesha S, Jain Amith K (2016) Human facial expression recognition from static images using shape and appearance feature. In: 2016 2nd international conference on applied and theoretical computing and communication technology (iCATccT), pp 598–603. IEEE

9. Akkoca BS, Gökmen M (2014) Facial expression recognition from static images. In: 2014 22nd signal processing and communications applications conference (SIU), pp 1291–1294. IEEE

10. Ghimire D, Jeong S, Lee J, Park SH (2017) Facial expression recognition based on local region specific features and support vector machines. Multimed Tools Appl 76(6):7803–7821

11. Irene K, Ioannis P (2006) Facial expression recognition in image sequences using geometric deformation features and support vector machines. IEEE Trans Image Process 16(1):172–187

12. Lajevardi SM, Lech M (2008) Facial expression recognition from image sequences using optimized feature selection. In: 2008 23rd international conference image and vision computing New Zealand, pp 1–6. IEEE

13. Haris RA (2018) Face expression recognition using artificial neural network (ANN) model back propagation. In: Proceedings of the joint workshop KO2PI and the 1st international conference on advance & scientific innovation, pp 126–134. ICST (Institute for Computer Sciences, Social-Informatics and ...)

14. Yang S, Bhanu B (2012) Understanding discrete facial expressions in video using an emotion avatar image. IEEE Trans Syst Man Cybern Part B (Cybernetics) 42(4):980–992

15. Zavaschi Thiago HH, Britto Jr Alceu S, Oliveira Luiz ES, Koerich Alessandro L (2013) Fusion of feature sets and classifiers for facial expression recognition. Expert Syst Appl 40(2):646–655

16. Wan C, Tian Y, Liu S (2012) Facial expression recognition in video sequences. In: Proceedings of the 10th world congress on intelligent control and automation, pp 4766–4770. IEEE

17. Liu X, Cheng T (2003) Video-based face recognition using adaptive hidden Markov models. In: 2003 IEEE computer society conference on computer vision and pattern recognition, 2003. Proceedings, vol 1, p I. IEEE

18. Pantic M, Patras I (2006) Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences. IEEE Trans Syst Man Cybern Part B (Cybernetics) 36(2):433–449

19. Khairuni AKS, Hisham JM, Jussi P, Rajendran P (2016) Spatiotemporal feature extraction for facial expression recognition. IET Image Process 10(7):534–541

20. Yaddaden Y, Adda M, Bouzouane A, Gaboury S, Bouchard B (2017) Facial expression recognition from video using geometric features

21. Anwar S, Ayoub A-H, Robert N, Moftah E (2014) Frame-based facial expression recognition using geometrical features. Adv Human-Comput Interact 2014:4

22. Nuri Ö, Mehmet BG (2021) New geometric based features for facial expression recognition. Technical report, EasyChair

23. Murugappan M, Mutawa A (2021) Facial geometric feature extraction based emotional expression classification using machine learning algorithms. Plos One 16(2):e0247131

24. Singh A, Khan MA, Baghel N (2020) Face emotion identification by fusing neural network and texture features: facial expression. In: 2020 international conference on contemporary computing and applications (IC3A), pp 187–190. IEEE

25. Chang L, Kaoru H, Junjie M, Zhiyang J, Yaping D (2021) Facial expression recognition using hybrid features of pixel and geometry. IEEE Access 9:18876–18889

26. Deepak G, Joonwhoan L, Ze-Nian L, Sunghwan J (2017) Recognition of facial expressions based on salient geometric features and support vector machines. Multimed Tools Appl 76(6):7921–7946

27. Asit B, Paramartha D (2019) Influence of shape and texture features on facial expression recognition. IET Image Process 13(8):1349–1363

28. Asit B, Paramartha D (2019) Facial expression recognition using distance and texture signature relevant features. Appl Soft Comput 77:88–105

29. Lei Z, Zengcai W, Zhang G (2017) Facial expression recognition from video sequences based on spatial-temporal motion local binary pattern and gabor multiorientation fusion histogram. Mathematical Problems in Engineering

30. Deepak G, Joonwhoan L (2014) Extreme learning machine ensemble using bagging for facial expression recognition. J Inf Process Syst 10(3):443–458

31. Happy SL, Routray A (2015) Robust facial expression classification using shape and appearance features. In: 2015 eighth international conference on advances in pattern recognition (ICAPR), pp 1–5. IEEE

32. Cootes Timothy F, Taylor Christopher J, Cooper David H, Jim G (1995) Active shape models-their training and application. Comput Vis Image Underst 61(1):38–59

33. Tzimiropoulos G, Pantic M (2013) Optimization problems for fast AAM fitting in-the-wild. In: Proceedings of the IEEE international conference on computer vision, pp 593–600

34. Barman A, Dutta P (2021) Facial expression recognition using distance and shape signature features. Pattern Recognit Lett 145:254–261

35. Ying-Ming W, Wang H-W, Lu Y-L, Yen S, Hsiao Y-T (2012) Facial feature extraction and applications: a review. In: Asian conference on intelligent information and database systems. Springer, pp 228–238

36. Lucey P, Cohn JF, Kanade T, Saragih J, Ambadar Z, Matthews I (2010) 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, pp 94–101. IEEE

37. Aifanti N, Papachristou C, Delopoulos A (2010) The mug facial expression database. In: 11th international workshop on image analysis for multimedia interactive services WIAMIS 10, pp 1–4. IEEE

38. Valstar M, Pantic M (2010) Induced disgust, happiness and surprise: an addition to the mmi facial expression database. In: Proceedings of 3rd international workshop on EMOTION (satellite of LREC): Corpora for research on emotion and affect, p 65. Paris, France

39. Amin M, Hassan A, Farzad T (2015) Video-based facial expression recognition by removing the style variations. IET Image Process 9(7):596–603

40. Yongqiang L, Shangfei W, Yongping Z, Qiang J (2013) Simultaneous facial feature tracking and facial expression recognition. IEEE Trans Image Process 22(7):2559–2573

41. Yogachandran R, Phan Raphael C-W, Chambers Jonathon A, Parish David J (2012) Facial expression recognition in the encrypted domain based on local fisher discriminant analysis. IEEE Trans Affect Comput 4(1):83–92

42. Shiqing Z, Xianzhang P, Yueli C, Xiaoming Z, Limei L (2019) Learning affective video features for facial expression recognition via hybrid deep learning. IEEE Access 7:32297–32304

43. Anima M, Laxmidhar B, Subramanian Venkatesh K (2014) Emotion recognition from geometric facial features using self-organizing map. Pattern Recognit 47(3):1282–1293