



# MQSMER: a mixed quadratic shape model with optimal fuzzy membership functions for emotion recognition

R. Vishnu Priya<sup>1</sup> · V. Vijayakumar<sup>2</sup> · João Manuel R. S. Tavares<sup>3</sup> 

Received: 24 November 2017 / Accepted: 18 December 2018  
© Springer-Verlag London Ltd., part of Springer Nature 2019

## Abstract

The traditional geometrical-based approaches used in facial emotion recognition fail to capture the uncertainty present in the quadrilateral shape of emotions under analysis, which reduces the recognition accuracy rate. Furthermore, these approaches require extensive computational time to extract the facial features and to train the models. This article proposes a novel geometrical fuzzy-based approach to accurately recognize facial emotions in images in less time. The four corner vertices of the mouth are the most important features to recognize facial emotions and can be extracted without the need of a reference face. These extracted features can then be used to define the quadrilateral shape, and the associated degree of impreciseness in the shape can be accessed using the proposed geometric fuzzy membership functions. Hence, four fuzzy features are derived from the membership functions and given to classifiers for emotion evaluations. In our tests, the fuzzy features achieved an accuracy rate of 96.17% in the Japanese Female Facial Expression database, and 98.32% in the Cohn-Kanade Facial Expression database, which are higher than the ones achieved by other common up-to-date methods. In terms of computational time, the proposed method required an average of 0.375 s to build the used model in a common PC.

**Keywords** Emotion recognition · Fuzzy shape · Geometric features · Facial expression

## 1 Introduction

Emotions play an important role in our daily lives. A study on communications through emotions conducted by a psychologist [1] found that 55% of our usual messages are transmitted through facial expressions or emotions, vocal

cues convey 38% and the remainder 7% is expressed using verbal cues. This suggests that facial expressions play a major role in human social interactions. Typically, facial expressions are created through shrinking of one or more facial muscles, which temporarily deform facial components. Ekman and Friesen [2] developed a well-accepted study on facial expressions and suggested that expressions are universal across human ethnicities and cultures. Their research also stated that there are six basic emotions: anger, disgust, fear, happiness, sadness and surprise, which can be evaluated based on facial muscle movements generated by 44 anatomical action units (AUs) defined in the facial action coding system (FACS). In recent years, several authors managed to recognize facial emotions using AUs [3–6]. However, it is a very laborious task to determine emotions using the FACS; consequently, attention has been given to automatic recognition of emotions. The recent progress in automation has seen a fast growth in facial expression analysis with applications in computer vision, pattern recognition and human–computer interaction. Several other applications, such as Emo chat [7], intelligent tutoring system [8], facial animation, and virtual reality of

---

✉ João Manuel R. S. Tavares  
tavares@fe.up.pt

R. Vishnu Priya  
vishnupriya.r@vit.ac.in

V. Vijayakumar  
vijayakumar.v@vit.ac.in

<sup>1</sup> Department of Computer Applications, Maulana Azad National Institute of Technology, Bhopal, Madhya Pradesh, India  
<sup>2</sup> School of Computing Science and Engineering, VIT University, Chennai Campus, Chennai, Tamilnadu, India  
<sup>3</sup> Instituto de Ciência e Inovação em Engenharia Mecânica e Engenharia Industrial, Departamento de Engenharia Mecânica, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal

facial emotions, have also been developed for the recognition of emotions.

Systems for automatic detection of facial expressions can extract relevant facial features from either static images or image sequences that are input to computational classifiers to recognize the respective emotions. Usually, there are two ways to recognize facial expressions, namely geometric-based and appearance-based approaches. The geometric-based approach uses the shape and position of the face under analysis, while the appearance-based features approach uses wrinkles, bulges, furrows, and other facial peculiarities and obtain essential information about facial expressions through micro-patterns. Several appearance-based algorithms have been proposed Happy and Routray [9], Poursaberi et al. [10], Zhong et al. [11], Zhang and Tjondronegoro [77], Song et al. [12] and Uddin et al. [13]. However, the major challenge of appearance-based features is its inability to generalize appearance-based features across different human races. Although geometric features also have their drawbacks, for example, they are very difficult to track and can easily be affected by noise, they can generate all the necessary information to recognize facial expressions [14]. In fact, humans have an extraordinary ability to recognize expressions, and, for example, even when a cartoon image has only facial contours, they can easily recognize the associated expression [15]. Therefore, geometric-based features seem to be the best option for the development of computational systems to recognize human expressions.

Most of the algorithms in the literature to detect facial expressions accurately can be classified as holistic or local. Eigenfaces and Fisherfaces [16, 17] are holistic methods that extract facial features from the complete face under analysis. On the other hand, local methods separate a face image into a few small blocks and apply certain feature extraction algorithms. Heisele et al. [18], Zou et al. [19] reported that the performance of facial expression recognition is significantly increased when local features are used compared to the whole face. Those local descriptors are identified through deformations of eyebrows, eyes, nose, mouth and 42 muscles. Among the local regions, Zhang et al. [74] stated that expression recognition based on the mouth is more rewarding than one based on the upper part of the face, that is the eyes. This statement can be justified since: First, the extraction of feature points from the mouth is easier than from the eye because the feature points in the mouth are much more clearly distinguished from each other. Majumder et al. [20] reported that the feature detection in the eyes is a challenging task due to the presence of eyelashes, shadows between the eyes and eyebrows, and the very small gap between eyes and eyebrows. Moreover, the eye vertices are located in the skin region without a distinctive grey scale. Second, the main

deformations in the face due to emotions are in the mouth region. Third, the main discerning features associated to facial expressions are distributed in the lower part of the face [21]. Fourth, although it is well known that the eye is highly sensitive to emotions, the stimulus response to the emotions is very small. Furthermore, each longitudinal section of the face seems to be a mirror of the other one, but the symmetrical view does not resemble the same, as can be observed in Fig. 1. These findings clearly suggest that the full mouth region with its geometrical nature can generate promising results.

This work introduces a fully automatic method for facial expression recognition using geometric features. A set of four corner vertices is extracted from the mouth region of the static image under analysis. The extracted features are used to define the quadrilateral shape and are then processed by the proposed fuzzy membership functions. The fuzzy features are derived from the membership functions with the ability to deal with uncertainty and are processed by a classifier that recognizes the presence of any basic expression. The experimental results show that the proposed approach achieves high recognition rates. The flow chart of the proposed approach is shown in Fig. 2.

This article is organized as follows: related works are reviewed in Sect. 2, the proposed approach is presented in Sect. 3; the results of the proposed approach are compared to the ones obtained by other up-to-date approaches in Sect. 4, and finally, Sect. 5 brings the conclusions and suggests future works.

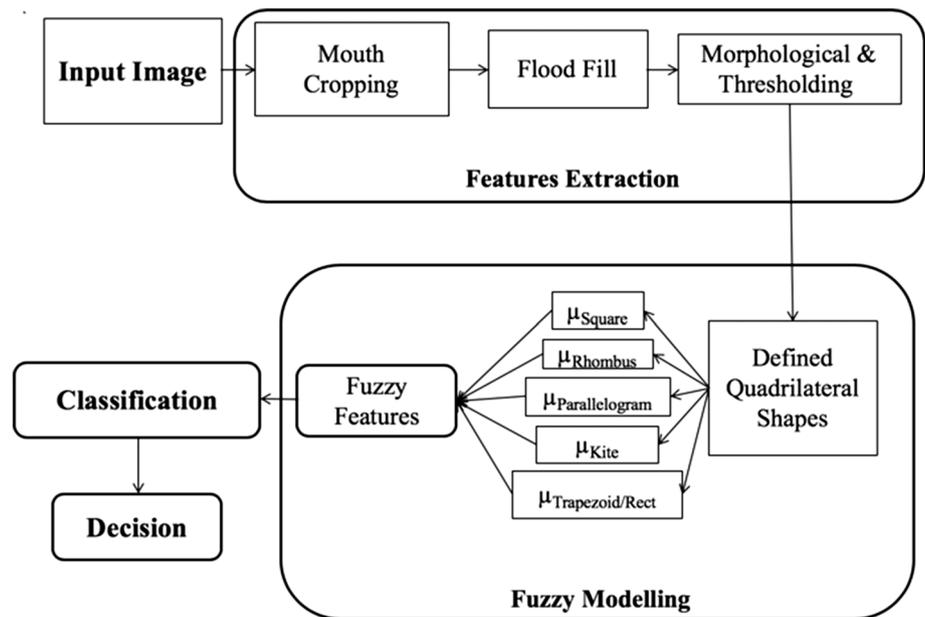
## 2 Related work

Researchers have worked on human facial emotion recognition for several decades, and various techniques and approaches to recognize emotions have been proposed. Some of these techniques and approaches are reviewed in the following subsections.



**Fig. 1** Two sample examples of the symmetrical view of the face: the face on the left is the original face and the face in the centre is the left symmetry and on the right is the right symmetry

**Fig. 2** Flow chart of the proposed approach



## 2.1 Emotion recognition from whole faces

To recognize emotions from whole faces, researchers have exploited pixel-based information [22, 23], wavelet transform [24, 25], Gabor filtering [26, 27], edges and skin detection [28], discrete cosine transform [29, 30], optical flow analysis [31], thermal analysis [32], local binary pattern [6, 34–39] and level set [40]-based methods. These methods extract features from whole faces of different persons, which increases the dimensionality of the recognition problem, and the required computational time and complexity grows.

## 2.2 Appearance-based approaches

The major disadvantage of active-based model methods, like the active appearance model (AAM) [41, 42] and the active shape Model (ASM), is the need for prior information concerning the expected shape features. During the training phase, the shape features of these models have to be identified, usually manually [43], and the recognition rate also strongly depends on the sample set used for training. A recent study to recognize facial expressions addressed the problem through the selection of the region near salient facial components: the extraction and matching of salient patch-based Gabor features were suggested in [77]. However, the proposed appearance-based method achieved low recognition rates due to the inefficiency in selecting suitable patches for matching. Gu et al. [15] used a radial encoding strategy based on Gabor filters to recognize facial expressions. The self-organizing map was applied to check the homogeneity of the encoded contours.

The experimental results obtained using faces without occlusion, i.e. whole faces, and with local occlusions, showed interesting results. Xie and Lam [44] introduced the shape and texture-based method for facial expression recognition. Zavaschi et al. [45] used a multi-objective genetic algorithm to select the best features from a pool built using Gabor filtering and local binary patterns. However, the selection of the more suitable features from the salient regions increased the required processing time.

## 2.3 Geometric-based approaches

In these approaches, the geometric features are extracted from areas of facial components, e.g. eyes, mouth and nose, and then the geometric relations among the extracted features are processed. Kobayashi and Hara [46] developed a local facial features model using geometric facial points. Zhang et al. [47] suggested the use of the position of fiducial points of the face under analysis, the multi-scale and multi-orientation Gabor wavelet coefficients at the same points or their combination to address the problem of facial expression recognition. Several recent geometric-based approaches are based on geometric feature tracking [12, 13, 48, 49], discriminant non-negative matrix factorization [50], graph-based feature point tracking [51] and facial contours [15]. In a common approach, the deformation of facial components is assessed by tracking the variation of feature points from the expressive image under study to the related neutral image. Usually, humans have the ability to recognize facial expressions without any reference face. Hence, the development of solutions for facial expression recognition using reference faces reduces

their success, as they are very different from the way humans perceive emotions, and also it increases the pre-processing time. Moreover, emotion analysis based on geometrical shapes always contains a certain level of ambiguity, which was not been taken into account in the previously mentioned approaches.

## 2.4 Recognition modules

Various classifiers have been used to build recognition modules for facial expressions. The well-known recognition modules are based on support vector machines (SVMs), hidden Markov models (HMM), Random Forest, Boosting, Bagging, Gaussian mixture models (GMM), dynamic Bayesian networks (DBN), and Multilayer Perceptron (MLP). For example Asthana et al. [52], Ghimire and Lee [53], Kotisa and Pitas [48], Moore and Bowden [54], Rudovic et al. [55], Saeed et al. [56], Zhang et al. [37], Bartlett et al. [57] and Bargal et al. [58], MLP-based networks in Mayor Torres and Stepanov [59], Tarnowski et al. [60] and Zhang [61], Deep Neural networks in Ding et al. [62], Huang and Lu [63] and Barros et al. [64], and Radial Basis Function Networks (RBFN) in Rosenblum et al. [65] to classify facial emotions directly, but always without taking into account the vagueness presented in the model, which can reduce the recognition rates.

The above review shows that the recent approaches have failed to capture the ambiguity presented in the geometric shape under analysis. Also the deformations associated with the expression need to be found by relating them to a corresponding neutral facial image. This reduces the efficiency and increases the required computational time and complexity. In our approach, the reference image is not needed, and a reduced number of features are extracted from the mouth to be analysed. The extracted features are then used to define the quadrilateral shape for each emotion, and the fuzzy membership functions are derived from the shape. The proposed fuzzy membership functions are a square, rhombus, kite and an isosceles trapezoid. These four fuzzy functions produce the fuzzy features to capture the impreciseness and vagueness, i.e. the uncertainty, present in the shape. Then, SVM and Random Forest-based classifiers are used for recognition. The results show that the recognition rate of the proposed method is higher than the ones from other recent approaches found in the literature.

## 3 Mixed quadratic shape model

Facial expression analysis is generally divided into three main phases: feature extraction, geometric transformation and expression classification. Here, the first phase

concerns the detection and extraction of feature points. The challenging issue in this phase is to find the optimal number of feature points to be used. The maximum number of extracted feature points found in the literature was 185 [77]. However, the number of extracted features should be as low as possible in order to reduce computational times. In the other more common related works, a facial reference image is needed, i.e. a face in a neutral state. Then reference features are extracted from the image for analysis. This causes an additional delay in the pre-processing stage and is also very different from the way humans perceive objects. Most of the recent works fail to discriminate emotions using traditional classification methods because impreciseness and vagueness present in the geometrical shapes are not captured. In this work, the aforesaid disadvantages are overcome by extracting a minimum number of feature points from the mouth and using the geometric fuzzy membership functions. The fuzzy features derived from fuzzy membership functions are used to classify the six basic emotions. The adopted fuzziness has the ability to deal with the uncertainty in shape that helps to effectively discriminate the emotions. The idea of the proposed mixed quadratic shape model (MQSM) developed to identify the emotions is described in the following subsections.

### 3.1 Background

A geometric-based approach can be used to describe the shape associated to a face. Some of the facial geometrical features commonly used in the literature are: point, line, triangle, circle, oval, ellipse and quadrilateral. However, to initialize and track facial shapes is challenging. Vadivel et al. [66] tracked the oval shape of the mouth using 13 feature points, but tracking all the points along the border of a shape is a difficult and time-consuming task. They also interconnected the centre point with the vertex points to measure the deformation involved, which requires extra computational time. Ghimire and Lee [67] tracked 52 facial key features modelled on points and lines to recognize facial expressions. Saeed et al. [56] used eight facial key-points to model the geometric structure of the face. Recently, Ghimire et al. [68, 69] extracted 52 facial key-points to develop their facial geometric model based on lines and triangles. They proved that the triangle-based representation outperforms both line- and point-based representations. The triangle is half of a quadrilateral.

The proposed approach defines the quadrilateral shape from four vertices of the mouth. The defined shape failed to match the quadrilateral shapes in geometry due to the ambiguity involved in the defined shape; however, this is overcome by using the proposed fuzzy membership functions.

### 3.2 Region of interest

As per discussion in the introduction, the mouth region has the highest deformation levels in faces due to emotions; therefore, it is considered as the region of interest (ROI) in this work. Moreover, psychologically, the left half of the entire body is controlled by the right part of the brain and the right half is controlled by the left part of the brain. As per Nielsen et al. [70] stated, the emotions are more expressive in the left half of the face of the people with right brain activity and vice versa. Therefore, the emotions extracted from the full mouth region are more truthful. The poses of the mouth can be used to find the associated deformations as listed in Table 1 for different type of facial emotions [71]. Based on Table 1, one can conclude that the left, upper, right and lower mouth vertices are the highlights for each emotion. Therefore, these four feature points of the mouth are employed in the current work. This low number of points reduces the required processing time, resource and storage space substantially, which facilitates, for example, the implementation in micro- and nano-electronic devices. The proposed approach is explained in the following subsections on a step-by-step basis.

### 3.3 Feature points extraction

The face to be analysed is localized in the input image, and to reduce the computational time, only three quarters of the lower part of the face is considered here as the ROI. Then, the mouth is cropped manually from the previous defined ROI; this results in the image  $C(x, y)$ , Fig. 3a. Then the flood fill algorithm is applied to obtain the intensity values of dark regions that are enclosed by lighter regions to the same intensity level, and the enhanced image  $C_{\text{ffalgo}}(x, y)$  is obtained, Fig. 3b. The latter image is further processed through thresholding and a morphological opening operation to obtain the contour boundaries:

$$T_h(x, y) = \sum_{i=0}^x \sum_{j=0}^y \begin{cases} 1, & \text{if } C_{\text{ffalgo}}(x, y) \geq T, \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

$$g(x, y) = (T_h \ominus S) \oplus S \quad (2)$$

where  $T$  is the global threshold and  $S$  is the  $3 \times 3$  structuring element. The contour boundary is used to find the four vertices based on the min and max values of the ‘ $x$ ’ and ‘ $y$ ’ coordinates of its points, respectively. These four vertices are denoted as  $A, B, C$  and  $D$ , which represent the left, right, top and bottom of the mouth, respectively:

$$A = g_A(x, y); g_A(x) = \text{Min}, \forall y \text{ and } g_A(y) = \text{Max}(\text{pixel value}) \quad (3)$$

$$B = g_B(x, y); g_B(x) = \text{Max}, \forall y \text{ and } g_B(y) = \text{Max}(\text{pixel value}) \quad (4)$$

$$C = g_C(x, y); g_C(x) = \text{Max}(\text{pixel value}) \text{ and } g_C(y) = \text{Max}, \forall x \quad (5)$$

$$D = g_D(x, y); g_D(x) = \text{Max}(\text{pixel value}) \text{ and } g_D(y) = \text{Min}, \forall x \quad (6)$$

Using the four points  $A, B, C$  and  $D$ , the quadratic shape is defined, as shown in Fig. 2. Using this shape, different human emotions can be recognized.

### 3.4 Quadrilateral shape definition

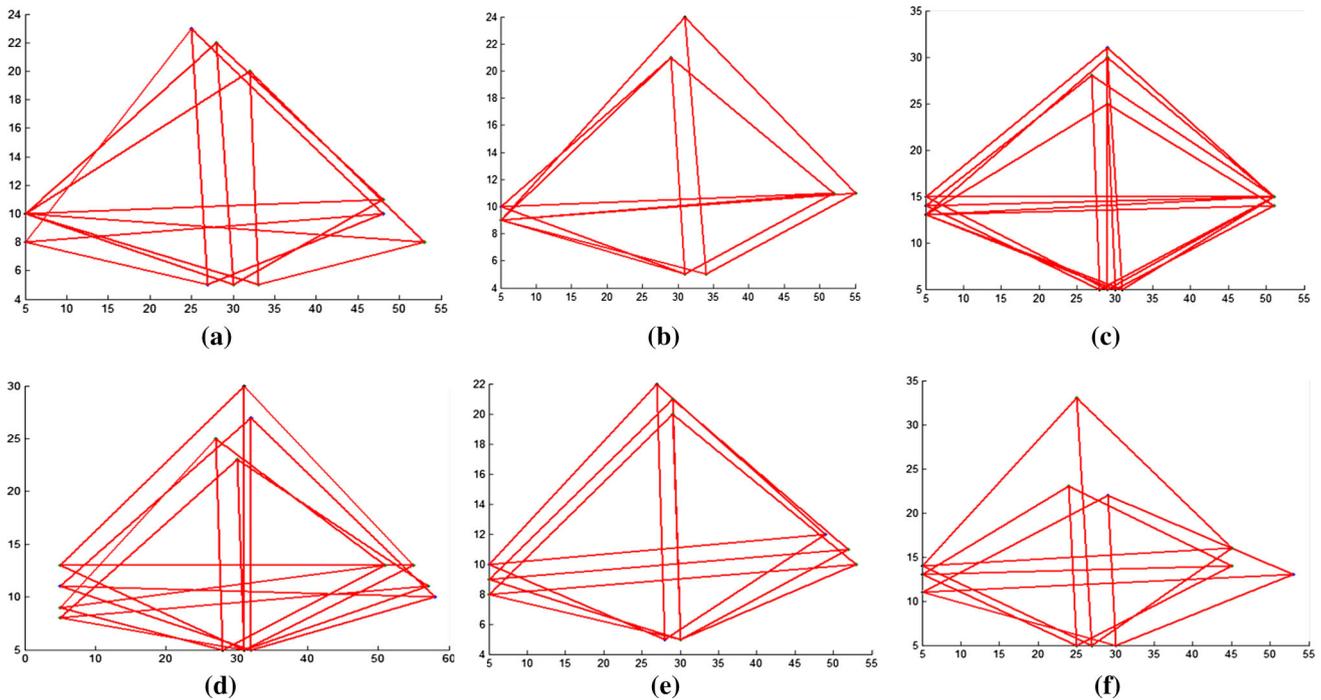
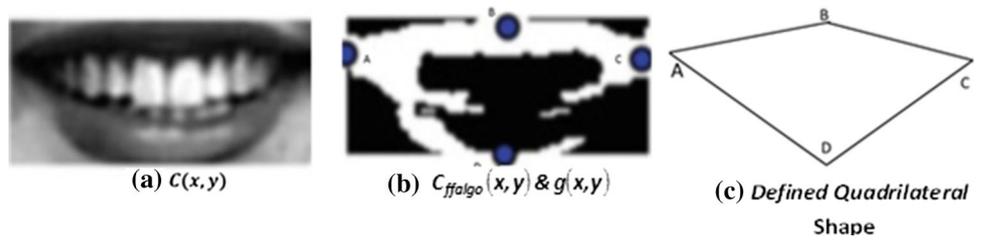
Figure 4 shows the defined quadratic shapes of the mouth in the images indexed with ‘KA.’ from the Japanese Female Facial Expression (JAFPE) benchmark dataset, which contains 6 emotions:  $E = \{\text{angry, disgust, fear, happy, sad and surprise}\}$ . The group of defined quadrilateral shapes for the  $e$ th emotion in Fig. 4 is denoted as  $GpDQS_e^L$ , where  $e \in \{1, 2, \dots, E\}$  represents the emotion index and represents the group index, and a single quadrilateral in a group is denoted as  $DQS_e^L$ .

Furthermore, the quadrilateral shapes in geometry: square, rhombus, parallelogram, kite and isosceles trapezoid, are denoted as  $GeoQSs$ . Figure 4 shows that each  $DQS_e^L$  in  $GpDQS_e^L$  failed to match up with  $GeoQSs$ . However, the impreciseness of  $DQS_e^L$  can be calculated using fuzzy membership functions. This can be done by calculating the contribution of the fuzzy functions of square, rhombus, parallelogram, kite and trapezoid in  $DQS_e^L$ . The contribution of  $DQS_2^2$  in  $GeoQSs$  is shown in Fig. 5, where

**Table 1** Emotions and respective mouth poses

Emotion	Mouth poses
<i>Fear</i>	Lip corners pulled sideways, tighten and elongating the mouth
<i>Happy</i>	Lips corners pulled up
<i>Anger</i>	Lips tighten and pressed together
<i>Surprise</i>	Mouth opened as jaw drops
<i>Disgust</i>	Mouth opened with upper lip raised, and tongue stuck out
<i>Sadness</i>	Lips corner pulled straight

**Fig. 3** Example of the low-level feature extraction from a mouth in an input image: **a** mouth segmented region; **b** four mouth corner vertices; **c** defined quadrilateral shape built for the mouth



**Fig. 4** Defined quadrilateral shapes for different emotions from the JAFFE dataset: **a** angry (KA.AN1–KA.AN3), **b** disgust (KA.DI1–KA.DI3), **c** fear (KA.FE1–KA.FE4), **d** happy (KA.HA1–KA.HA4), **e** sad (KA.SA1–KA.SA3), and **f** surprise (KA.SU1–KASU3)

the shapes in dotted lines represent the *GeoQSs* and the shapes drawn in continuous lines represent the defined shape. The proposed fuzzy membership functions for square, rhombus, parallelogram, kite and rectangle/isosceles trapezoid are presented in the following subsection, and the degrees of the functions are defined as:

$$\mu_c(X) = \begin{cases} \text{if } \mu_c(X) = 0, & \text{Not a member of fuzzy set} \\ 0 \leq \mu_c \leq 1, & \text{Membership degree varies} \\ \mu_c(X) = 1, & \text{Member of fuzzy set} \end{cases} \quad (7)$$

### 3.5 Fuzzy membership functions for the defined quadratic shape

As aforementioned, the proposed approach calculates the contribution of each *GeoQSs* in  $DQS_e^L$ . The variables used in the proposed mixed quadratic shape model are shown in Fig. 6 and in Table 2. In Fig. 6, lines ‘A’ and ‘B’ indicate

the widths of the MQSM and the vertical lines the lengths, and hence A as four points ‘ $a_1$ ’, ‘ $a_2$ ’, ‘ $a_3$ ’ and ‘ $a_4$ ’, in the same way, B as five points named as ‘ $b_1$ ’, ‘ $b_2$ ’, ‘ $b_3$ ’, ‘ $b_4$ ’ and ‘ $b_5$ ’. The point ‘ $c_1$ ’ is the starting point of the kite. The dotted lines between two endpoints are the diagonals of the respective shape. The proposed fuzzy membership functions are built as explained in the following subsections.

#### 3.5.1 Square

In geometry, a four-sided regular quadrilateral with all sides equal is called a square. A logical representation of a square is depicted in Fig. 6 with  $a_1, a_2, b_2$  and  $b_1$ . Based on the properties of this shape, the four equal lengths of the square are given by the distance between any two adjacent points, i.e.  $a_1a_2 = a_2b_2 = b_2b_1 = b_1a_1$ . The length of the diagonals is the distance between opposite vertices:  $a_1b_2 = b_1a_2$ . Only one of the diagonals is shown in Fig. 6 to preserve the clarity of the diagram. It can be noted that the

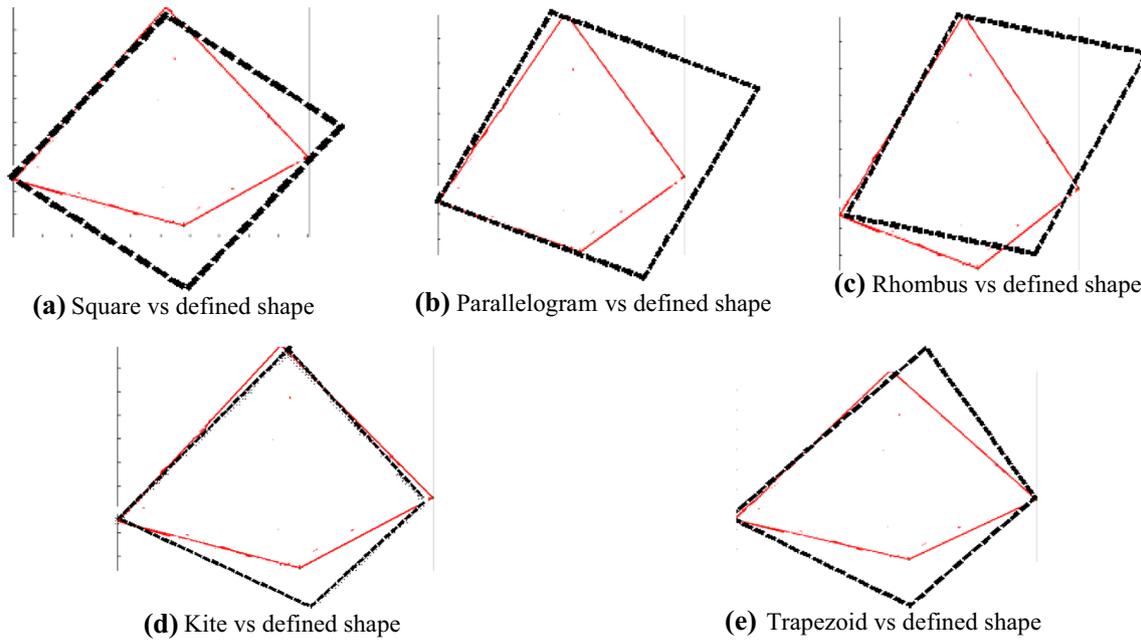


Fig. 5 Geometrical shapes (dotted shapes) versus defined shape (continuous shapes)

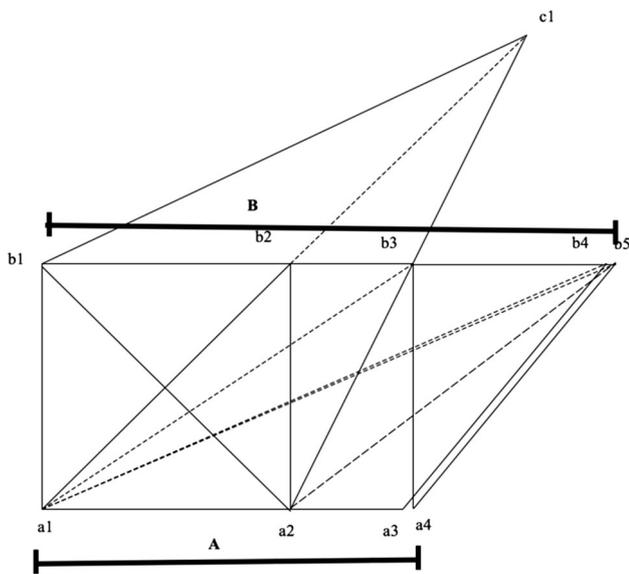


Fig. 6 Variables used in the mixed quadratic shape model

diagonal of any square is always greater than its side by  $\sqrt{2}$  times; therefore, the fuzzy membership function for a perfect square is defined as:

$$\mu_{Sq}(MQS) = e^{-[MQS_{Sq}-1]^2} \tag{8}$$

where  $MQS_{Sq} = [(2 * (a_1b_1 + a_1a_2))/(a_1b_2 + a_2b_1)]$ , i.e.  $MQS_{Sq}$  is defined as twice the ratio between the sum of two adjacent sides and the sum of the two diagonals.  $MQS_{Sq}$  is assigned to each  $CQS$  of  $GCQS_e$  to check the perfectness of the square. Hence in Eq. (8) the values range from:

$$\mu_{Sq}(MQS) = \begin{cases} 1 \leq \mu \leq 1.4, & \text{Perfect square} \\ 0 \leq \mu < 1, & \forall \text{ others} \end{cases} \tag{9}$$

### 3.5.2 Parallelogram

The well-known property of a parallelogram (PP) is that both lengths and widths are different, as depicted in Fig. 6

Table 2 Six primitive shapes inferred from the MQS model

Shape	Sides	Length 1	Width 1	Length 2	Width 2	Diagonal 1	Diagonal 2
Square	$a_1a_2b_1b_2$	$a_1b_1$	$a_1a_2$	$a_2b_2$	$b_1b_2$	$a_1b_2$	$a_2b_1$
Rhombus	$a_1a_3b_2b_4$	$a_1b_2$	$a_1a_3$	$a_3b_4$	$b_2b_4$	$a_1b_4$	$a_3b_2$
Parallelogram	$a_1a_4b_2b_5$	$a_4b_5$	$a_1a_4$	$a_1b_2$	$b_2b_5$	$a_1b_5$	$a_4b_2$
Kite	$a_1a_2b_1c_1$	$a_2c_1$ (L)	$a_1a_2$ (S)	$a_1b_1$ (S)	$b_1c_1$ (L)	$a_1c_1$	$a_2b_1$
Rectangle	$a_2a_4b_3b_2$	$a_2b_2$	$a_2a_4$	$a_4b_3$	$b_2b_3$	$a_2b_3$	$a_4b_2$
Isosceles trapezoid	$a_2a_4b_5b_1$	$a_2b_1$	$a_2a_4$	$a_4b_5$	$b_1b_5$	$a_2b_5$	$a_4b_1$

by  $a_1, a_4, b_2$  and  $b_5$ . The lengths  $b_2b_5$  and  $a_1a_4$  are the upper and lower sides of parallelogram, and  $a_1b_2$  and  $a_4b_5$  are the widths. However, the opposite sides of the PP are equal, i.e.  $b_2b_5 = a_1a_4$  and  $a_1b_2 = a_4b_5$ . The PP has a long diagonal and a short one:  $a_1b_5 \neq a_4b_2$ . Using these properties, the fuzzy membership function for a perfect parallelogram is derived as:

$$\mu_{Pll}(MQS) = e^{\{-[MQS_{Pll}-1]^2\}} \tag{10}$$

where  $MQS_{Pll} = [((a_4b_5 - a_1a_4) * (a_1b_5 - a_4b_2))/((b_2b_5 - a_1b_2) * (a_4b_2 - a_1b_5))]$ , which is a ratio of two products where the difference of two adjacent sides and the difference of the two diagonals (with a change in sign) in both the numerator and the denominator and the values always range between [0, 1].

### 3.5.3 Rhombus

Rhombus is a special type of parallelogram with two diagonals: one long and one short. It differs from a parallelogram since all of its four sides have the same length. A logical representation of a rhombus with its diagonals is shown in Fig. 6 by  $a_1, a_3, b_2$  and  $b_4$ , with the lengths indicated by  $a_1b_2 = a_1a_3 = a_3b_4 = b_2b_4$  and the diagonals by  $a_1b_4 \neq a_3b_2$ .

The membership value of a perfect rhombus is the ratio between two adjacent sides with the differences in the diagonals which is equals to 1 (one):

$$\mu_{Rh}(MQS) = e^{\{-[MQS_{Rh}-1]^2\}} \tag{11}$$

where  $MQS_{Rh} = [(a_1b_2 * (a_1b_4 - a_3b_2))/(b_2b_4 * (a_1b_4 - a_3b_2))]$ .

### 3.5.4 Kite

A kite is a quadrilateral with four sides grouped into two sets of equal length sides which are adjacent to one another. Interestingly, a parallelogram also has two sets of equivalent length sides; however, they are opposite to one another. A pictorial representation of a kite is shown in Fig. 6 by  $a_1, a_2, b_1$  and  $c_1$ . The fuzzy membership function using the variables of a kite is given by Eq. 12, where  $b_1c_1 = a_2c_1$  and  $b_1a_1 = a_1a_2$  according to the property of the two segments joining opposite points of tangency of equal length; additionally, the diagonals connecting opposite ends have different lengths:

$$\mu_{Kt} = e^{\left\{-\left[\frac{(b_1c_1 - a_1a_2) * (a_1c_1 - a_2b_1)}{(a_1b_1 - a_2c_1) * (a_2b_1 - a_1c_1)} - 1\right]^2\right\}} \tag{12}$$

### 3.5.5 Rectangle and isosceles trapezoid

A rectangle is shown in Fig. 6 by  $a_2, b_2, b_3$  and  $a_4$ , where the diagonals are of equal length and the adjacent sides are not, i.e.  $a_2b_3 = b_2a_4$  and  $a_2a_4 \neq b_2b_3$ . Figure 6 represents the rectangle pictorially that appears within the isosceles trapezoid. The isosceles trapezoid is typically considered a special type of rectangle as shown in Fig. 6 by  $a_2, b_1, b_5$  and  $a_4$ , where two opposite sides are parallel and the other two sides are of equal length, i.e.  $b_1b_5 \parallel^{el} a_2a_4$  and  $a_2b_1 = a_4b_5$ , respectively, which means that adjacent sides do not have equal lengths. The diagonals split each other into similar regions with lengths that are pairwise equal. As pictured in Fig. 6, the diagonals  $b_1a_4 = b_5a_2$  have the same length. Hence, the fuzzy membership function is derived for both rectangular and isosceles trapezoid as:

$$\mu_{R/It} = e^{\left\{-\left[\frac{(a_2b_3) * (a_2b_2 - a_2a_4)}{(a_4b_2) * (a_4b_3 - b_2b_3)} - 1\right]^2\right\}} \tag{13a}$$

or

$$\mu_{R/It} = e^{\left\{-\left[\frac{(a_2b_5) * (a_2b_1 - a_2a_4)}{(a_4b_1) * (a_4b_5 - b_1b_5)} - 1\right]^2\right\}} \tag{13b}$$

### 3.6 Fuzzy set

A set ( $\mu_m$ ) of membership degrees obtained from Eqs. (8)–(13) for each  $DQS_e^l$  of the different  $e \in \{1, 2, \dots, E\}$  generates the fuzzy set  $E_{FS}$ , where  $m$  represents the total number of fuzzy membership functions. Each degree in  $E_{FS}$  varies in the real unit interval of [0, 1], say  $\mu_m : E_{FS} \rightarrow [0, 1]$ . Using the matrix of the fuzzy relation  $E_{FR} : E_{FS} \rightarrow E_s$ , which relates the membership degrees with its associated emotion, the classification decision can be achieved.  $E_{FS}$  reveals that the minimum number of fuzzy values is used for predicting facial expression; hence, the required processing time, resources and storage are substantially reduced. Then,  $E_{FR}$  is the input for a machine learning algorithm, which plays an important role to improve the classification accuracy. The proposed approach was tested using recent machine learning algorithms as described and discussed in the next section.

## 4 Experimental results

In this section, the recognition rates of the proposed approach are assessed and compared against other common approaches.

### 4.1 Dataset

The JAFFE and Cohn-Kanade Facial Expression (CK++) databases for facial expression analysis were used to assess the proposed and the other state-of-the-art approaches. JAFFE has 213 grayscale facial expression images (neutral—30, angry—30, disgust—29, fear—33, happy—30, sad—31, and surprise—30) of ten subjects. This controlled database was taken under similar lighting conditions and without occlusion. All the images have a resolution of  $256 \times 256$  pixels. As previously described, the three quarters of the lower part of each input image were considered the ROI, from which the mouth region is extracted and used in the posterior processing steps. In addition, the CK++ dataset available at CMU, in Pittsburgh, USA (Kanade et al. 2000), was used, which consists of 593 image sequences from 123 subjects. The facial expressions in each grayscale sequence begin with a neutral face and increase to the height of the emotion given in the last frame. These peak expression frames from each sequence were used to validate the performance of the proposed approach. Experiments were carried out for the six emotions: anger, disgust, fear, happy, sad and surprise; however, the neutral images were discarded. Therefore, all images used in this study are publicly available and were acquired according to the Ethics Commissions of the related Institutions.

### 4.2 Feature extraction from mouth

The accuracy rate decreases substantially when the entire set of facial features is considered [72]. Recently, researchers recognized emotions using important facial regions, particularly the eyes and mouth [73]. As discussed previously, the mouth provides more promising results than eyes; hence, the four vertices of the mouth in each input image of the JAFFE dataset were extracted, Fig. 7.

Using the four vertices  $A, B, C$  and  $D$ , the  $DQS_e^L$  for all facial expression images were defined. Then, Eqs. (8)–(13) were applied on each  $DQS_e^L$  to generate the  $E_{FS}$ . Each

collection of membership degrees in  $E_{FS}$  was mapped to the respective emotion to obtain the  $E_{FR}$ , and the recognition rate associated to each emotion was computed.

### 4.3 Mouth against eyes

The highest recognition rates obtained by Gu et al. [21] using the full faces on the JAFFE and CK++ databases were 89.67% and 91.57%, respectively. In this approach, the accuracy of the results was enhanced once again with the eyes/mouth occlusion. Three mask sizes, namely small, medium and large, were overlapped on the eyes and mouth, and the results were evaluated. Based on the evaluation perform, it was concluded that the expressions with a masked mouth were more difficult to recognize than those with masked eyes. The results for the CK++ database with large masks are given in Table 3, where there was a 12% improvement in the recognition rate for masked eyes.

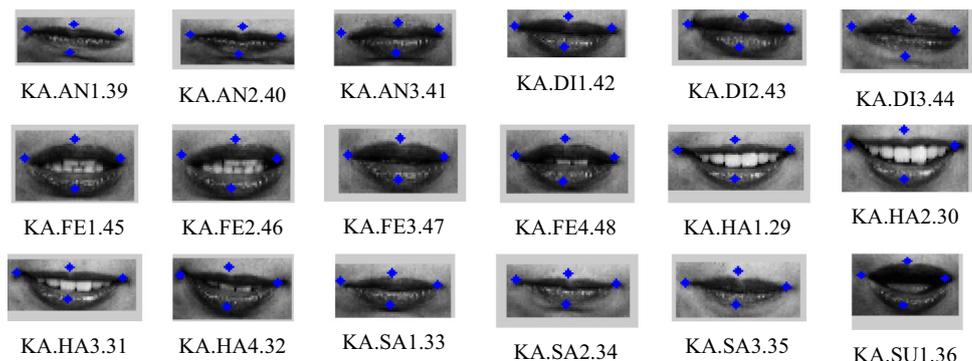
Zhang et al. [74] obtained a recognition rate of 78.6% for lower AUs and of 71.3% for the upper ones. Moreover, in [48, 50] the recognition accuracy obtained was 96.3% with eyes occlusion, and 93.7% with mouth occlusion. This concludes that the recognition accuracy for the mouth provides more promising outcomes than the eyes.

Table 4 presents the confusion matrices obtained by the proposed work and by the work presented in [21] for the CK++ dataset. The data in Table 4 show that the recognition rate of the proposed approach was 96.5% and was 85.47% for the approach under comparison. This suggests that the proposed approach gives encouraging recognition rates compared to other similar works for the mouth and that the mouth leads to more promising results than the eyes.

### 4.4 Recognition rate

The local features extracted from the JAFFE images were analysed using the J48 decision tree to prune unwanted features from the dataset. Table 5 presents the confusion matrix obtained for the JAFFE dataset using Fuzzy

Fig. 7 Cropped mouths from images indexed with “KA.” in the JAFFE dataset with the four vertices highlighted



**Table 3** Results obtained by the method proposed by Gu et al. [21] with mouth and eyes occluded with large masks

Gu et al. [21]	Happy	Sad	Surprise	Disgust	Angry	Scared	Recognition rate (%)
Mouth Masked	80.2	50.97	93.01	85.53	62.12	69.55	73.56
Eyes masked	89.8	86.43	96.97	91.86	63.00	84.78	85.47

**Table 4** Results with the CK++ images by the proposed method and the one proposed by Gu et al. [21] only based on mouth features

Approach (mouth)	Angry	Disgust	Fear/scared	Happy	Sad	Surprise	Recognition rate (%)
Proposed work	93.1	95.3	93.36	99.67	97.7	100	96.52
Gu et al. [21]	63.00	91.86	84.78	89.8	86.43	96.97	85.47

**Table 5** Confusion matrix from the JAFFE dataset using the J48 algorithm

	Angry	Disgust	Fear	Happy	Sad	Surprise	Percentage (%)
Angry	<b>26</b>	1	1	0	2	0	<b>86.66</b>
Disgust	2	<b>20</b>	2	1	4	0	<b>68.99</b>
Fear	0	2	<b>29</b>	0	1	0	<b>90.65</b>
Happy	0	0	0	<b>31</b>	0	0	<b>100</b>
Sad	2	1	1	0	<b>27</b>	<b>0</b>	<b>87.09</b>
Surprise	3	0	2	0	0	25	<b>83.33</b>

Average of Correctly Classified Instances: 86.12%

Bold values indicate the best results

Membership Functions (FMF) with the J48 decision tree algorithm. The table includes the individual recognition percentage of each emotion along with the overall correct recognition percentage. The J48 decision tree algorithm is used to identify the essential features and reduces the outliers in each class. It greatly reduces both the input dimension and the required computational time. On using the J48 algorithm, the values of  $\mu_{PII}$  were found to be similar, which may not help in the discrimination of the emotions. Thus, the other four fuzzy features, namely  $\mu_{Sq}$ ,  $\mu_{Rh}$ ,  $\mu_{Kt}$  and  $\mu_{R/It}$ , were selected for further use in the machine learning algorithm to improve the accuracy rate.

The machine learning methods used in the facial emotion recognition were: SVM, Multilayer Perceptron (MLP) and Ensemble learning, which have all been successfully used in other similar works. The parameters were selected separately for each classifier in order to find the best values that led to the highest classification. The LibSVM toolbox of the SVM was used with the Linear, Polynomial, Radial Basis Function (RBF) and Sigmoid kernels. Among these, the best result was obtained with the RBF kernel, and with the parameters  $C = 20$  and  $\gamma = 0.1$ , see Table 6. The MLP classifier presented good results with a topology of 4 neurons in the input layer, 20 neurons in the hidden layer and 7 in the output layer, and the learning rate = 0.3, momentum = 0.2 and training time = 500. Finally, a recent machine learning technique, mainly ensemble learning,

was used to assess the performance of the proposed approach. In the first model, an Adaboost meta-algorithm with Random Forest was used, and in the second model, bagging with Hidden Markov Model was combined to build the model for prediction. In the Random Forest tree, resampling was applied as a pre-processing step, which is a supervised filter to produce a random subset of the input dataset. Each classifier was trained using tenfold cross-validation. Using these configurations, the recognition accuracies of the JAFFE dataset for the different classifiers were computed and the confusion matrices were built, Tables 6, 7, 8 and 9.

Tables 6, 7, 8 and 9 show that all classifiers provided more or less the same outcome and efficiency: with an accuracy from 95.1 to 96.2%. Beside the overall efficiency, the recognition rate of all classifiers was equal to 100% for the emotions angry and sad. The emotions that follow the highest recognition rates were happy and disgust. It seems that the recognition rates for the surprise class tend to be lower compared to the other emotions, as it was confused with the other expressions. The probable reason reported in Zhang et al. [74] for this finding is that the surprise images on the JAFFE database have a closed or only slightly open mouth. However, the proposed approach showed an interesting result for the surprise emotion using the Random Forest-based ensemble learning classifier.

**Table 6** Confusion matrix obtained with the JAFFE dataset using SVM with RBF

	Angry	Disgust	Fear	Happy	Sad	Surprise	Percentage (%)
Angry	<b>30</b>	0	0	0	0	0	<b>100</b>
Disgust	1	<b>28</b>	0	0	0	0	<b>96.6</b>
Fear	0	0	<b>30</b>	1	1	0	<b>93.8</b>
Happy	0	0	0	<b>30</b>	1	0	<b>96.8</b>
Sad	0	0	0	0	<b>31</b>	0	<b>100</b>
Surprise	1	2	0	2	0	<b>25</b>	<b>83.33</b>

Average of Correctly Classified Instances: 95.08%

Bold values indicate the best results

**Table 7** Confusion matrix obtained with the JAFFE dataset using MLP

	Angry	Disgust	Fear	Happy	Sad	Surprise	Percentage (%)
Angry	<b>30</b>	0	0	0	0	0	<b>100</b>
Disgust	1	<b>28</b>	0	0	0	0	<b>96.6</b>
Fear	0	0	<b>31</b>	0	1	0	<b>96.9</b>
Happy	0	0	0	<b>30</b>	1	0	<b>96.8</b>
Sad	0	0	0	0	<b>31</b>	0	<b>100</b>
Surprise	1	2	0	2	0	<b>25</b>	<b>83.3</b>

Average of Correctly Classified Instances: 95.63%

Bold values indicate the best results

**Table 8** Confusion matrix obtained with the JAFFE dataset using the Bagging + HMM model

	Angry	Disgust	Fear	Happy	Sad	Surprise	Percentage (%)
Angry	<b>30</b>	0	0	0	0	0	<b>100</b>
Disgust	1	<b>27</b>	0	0	1	0	<b>93.1</b>
Fear	0	0	<b>31</b>	0	1	0	<b>96.9</b>
Happy	0	0	0	<b>30</b>	1	0	<b>96.8</b>
Sad	0	0	0	0	<b>31</b>	0	<b>100</b>
Surprise	1	2	0	2	0	<b>25</b>	<b>83.33</b>

Average of Correctly Classified Instances: 95.08%

Bold values indicate the best results

**Table 9** Confusion matrix obtained with the JAFFE dataset using the Adaboost + Random Forest model

	Angry	Disgust	Fear	Happy	Sad	Surprise	Percentage (%)
Angry	<b>30</b>	0	0	0	0	0	<b>100</b>
Disgust	1	<b>27</b>	0	0	1	0	<b>93.1</b>
Fear	0	0	<b>31</b>	0	1	0	<b>96.9</b>
Happy	0	0	0	<b>30</b>	1	0	<b>96.8</b>
Sad	0	0	0	0	<b>31</b>	0	<b>100</b>
Surprise	1	1	0	1	0	<b>27</b>	<b>90.0</b>

Average Correctly Classified Instances: 96.17%

Bold values indicate the best results

The consolidated results for the different classifiers in terms of average recognition rate using tenfold cross-validated and percentage split are shown in Table 10.

In terms of percentage split, the proposed approach was evaluated using the default values of the classifiers with 66% for training purposes and 34% for evaluation. Table 10 shows that the average accuracy rates obtained by

the classifiers were always higher than 95% reaching a peak at 99.8%. The proposed approach achieved the best accuracy rate (99.8%) with percentage split, which is better than the ones obtained by the state-of-the-art approaches. However, the tenfold cross-validator was the best estimator and the Adaboost + Random Forest model obtained the best results among all the classifiers. Hence, the

**Table 10** Average recognition rate of the proposed approach using different classifiers

Classifier	Tenfold cross-validated (%)	Training samples (66%) Testing samples (34%) (%)
SVM	95.08	96.8
BAGGING + HMM	95.08	97.1
MLP	95.63	<b>99.8</b>
ADABOOST + RANDOM FOREST	<b>96.17</b>	98.4

Bold values indicate the best results

Adaboost + Random Forest was chosen as the proposed model classifier.

Apart from the recognition rate, statistical tests are required to prove the performance of a new classification method. Thus, the statistical measures, namely sensitivity (SEN), specificity (SPEC), positive predictive value (PPV), *F*-measures, were calculated for our method applying all the classifiers used in this work, Table 11.

Sensitivity and specificity measure the percentage of positive and negative samples that are correctly recognized, respectively. PPV defines the proportion of positive outcomes in a statistical test. The data in Table 11 confirm that the overall averages were 95.5% for sensitivity, 99% for specificity, 95.5% for PPV and 95.5% for the *F*-measure. These results show that the overall performance of the proposed approach was good. The Adaboost + Random Forest classifier achieved the best results followed by MLP, then the SVM classifier with the RBF kernel and the HMM-based classifier which had similar performances. Furthermore, the proposed approach took only 0.18–0.64 s (TT) to train the model; this high computational speed was due to fact that the model was trained using only four fuzzy features. Also, the proposed fuzzy membership functions used only the elementary arithmetical operations and operated over: the length, width and diagonals of the quadrilateral shapes, which takes a fixed computational time.

The Adaboost + Random Forest model was found to outperform the other classifiers in both recognition rate and the statistical metrics. The performance of this classification model was assessed using the receiving operator

characteristics (ROC) curve, Fig. 9. The *X*-axis of the ROC curve represents false positive (1-Specificity) and the *Y*-axis the true positive (Sensitivity). The best compromise is found when both sensitivity and specificity are highest at the same time. Further, the area under the ROC curve, called AUC, is 1 (one) for a perfect predictive power. Figure 8 shows that the best predictive power for all emotions was found using the Adaboost + Random Forest classifier model.

However, the ROC curves might be mislead when handling highly unbalanced datasets. Therefore, graphs were drawn for precision versus recall (PR) to interpret the performance of the proposed method in a more objective manner. Hence, the PR curves in Fig. 9 indicate the number of true positives that are likely to be obtained in a competent predictive system.

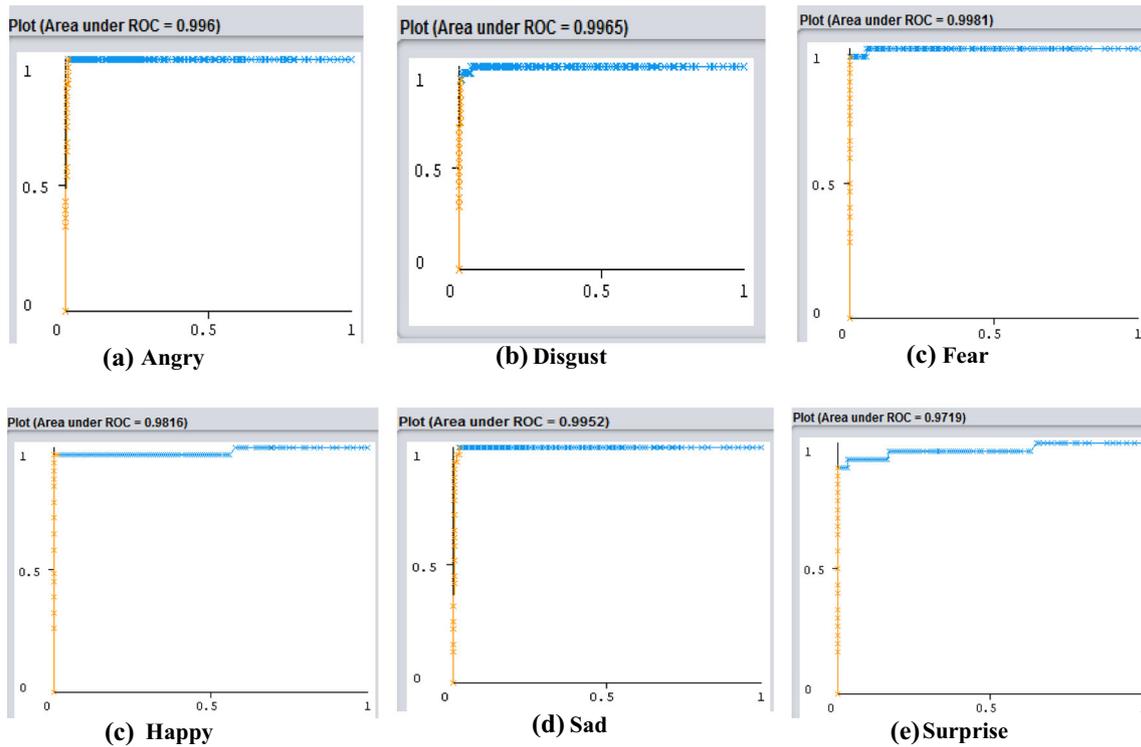
Table 12 presents the confusion matrix for the methods under comparison, and the proposed approach outperformed all the other approaches, and without the need of a neutral image as reference. The recognition rate obtained by Wu et al. [5] was lower than 80% and poorer compared to all the other approaches. The results obtained by Zhang et al. [37] and Gu et al. [21] were greater than 90%, and the performances as to happy emotion obtained by Happy and Routray [9], Zhang and Tjondronegoro [77] and Rahulathavan et al. [23] reached 90%. Gaidhane et al. [75] more recently presented an approach that reached 94%. In short, the proposed approach gave encouraging results compared to the other approaches. These encouraging results are because the impreciseness and vagueness in the shapes used to classify each emotion were built using fuzzy membership functions.

When the proposed method was evaluated with the CK++ dataset, the default parameter values were used for the classifiers and the SVM with the RBF kernel classifier gave the best results. Table 13 gives the confusion matrices of the proposed method as well as the other methods under comparison for the CK++ database. The recognition rates for the surprise and happy emotions obtained by the proposed method were higher than those of the other methods. Recently, Ghimire et al. [68, 69] and Gu et al. [21] suggested that the results for anger, fear and sadness emotions in the CK++ dataset were more similar than the ones obtained for happy and surprise emotions. The best results

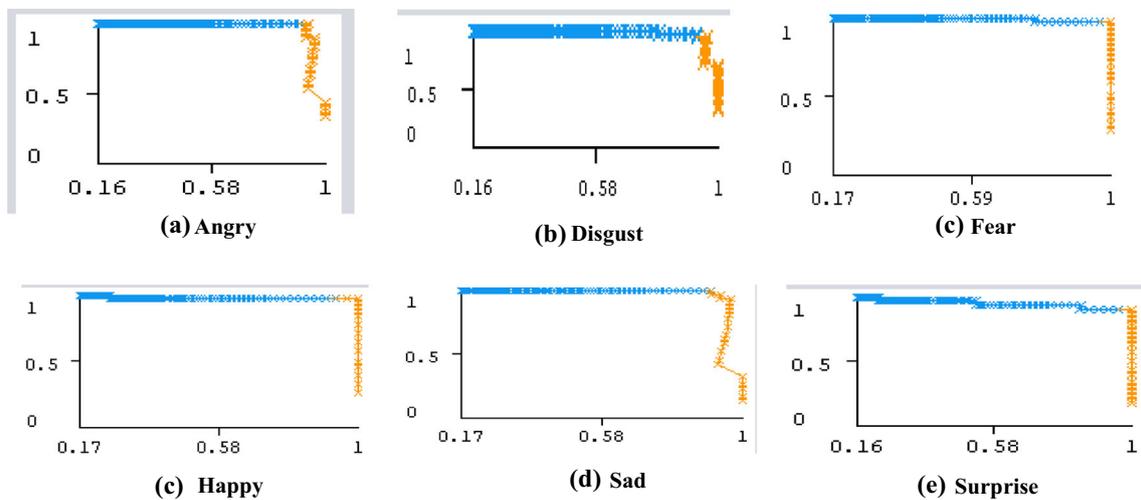
**Table 11** Overall performance of the proposed approach using different classifiers

CLASSIFIER	SEN	SPEC	PPV	<i>F</i> -measure	TT (s)
SVM	0.951	0.99	0.954	0.95	<b>0.18</b>
MLP	0.956	0.991	0.958	0.956	0.64
BAGGING + HMM	0.951	0.99	0.953	0.950	0.33
ADABOOST + RANDOM FOREST	<b>0.962</b>	<b>0.992</b>	<b>0.964</b>	<b>0.962</b>	0.35

Bold values indicate the best results



**Fig. 8** ROC curves for all emotions using the Adaboost + Random Forest classifier model



**Fig. 9** PR curves for all emotions when classified using the Adaboost + Random Forest classifier model

were obtained when they took two peak expression frames for anger, fear and sadness and one for each of the happy and surprise emotions. Finally, they achieved a 97.25% recognition rate. The results of our proposed approach were similar with a value of 98.32%; however, we only used the last frame of each case in the input database. Zhang et al. [74] reported that the surprise images in the CK++ database have an exaggerated open mouth and are easily distinguished from the other emotions. Also, the results for the surprise and anger emotions contrast with those obtained

with the JAFFE database. Table 13 shows that the average results obtained (98.32%) are better than those obtained by the other methods. The highest accuracy obtained by the proposed approach was for the emotion of surprise, and the lowest was for the emotion of anger.

Tables 14 and 15 indicate the performance obtained previously by similar works on facial emotion recognition using the JAFFE and CK++ databases, respectively. In general, the literature considers that a performance comparison with other approaches may not be analysed directly

**Table 12** Results obtained by different state-of-the-art approaches when classifying the different emotions on the JAFFE database

Literature	Angry	Disgust	Fear	Happy	Sad	Surprise	Average (%)
Proposed work	100	93.1	96.9	96.8	100	90.0	96.17
Gaidhane et al. [75]	92.6	94.4	91.0	96.0	94.2	96.5	94.11
Happy and Routray [9]	100	86.2	93.8	96.8	77.4	96.7	91.81
Wang et al. [88]	84.2	87.2	78.1	96.3	92.4	96.1	89.0
Rahulamathavan et al. [23]	96.7	93.1	93.8	93.5	90.3	93.3	93.45
Zhang et al. [37]	92.4	90.8	87.5	96.2	84.2	88.3	89.88
Gu et al. [21]	93.3	86.2	75.0	100	93.3	96.7	90.75
Zhang and Tjondronegoro [77]	96.7	90.0	93.8	93.6	93.6	90.0	92.92
Wu et al. [5]	83	68	67	88	78	88	78.66

**Table 13** Results of various approaches in the classification of the different emotions from the CK++ database

Approach	Angry	Disgust	Fear	Happy	Sad	Surprise	Average
Proposed work	95.3	100	95.8	100	98.8	100	98.32%
Ghimire et al. [68, 69]	97.5	96.7	96	100	96.7	100	97.81%
Agarwal and Mukherjee [89]	97.5	94.6	82.6	98.2	92.3	100	94.2%
Gaidhane et al. [75]	94	98.7	93.1	99.6	94.5	99.7	96.47
Hsieh et al. [90]	93.3	93.8	90.5	94.5	–	96.1	93.6
Zhang et al. [3]	85	95	85	97	90	98	90.38%
Pu et al. [41]	75	94.6	68.6	97.7	88.9	92.5	89.37%
Happy and Routray [9]	87.8	94.3	93.3	94.2	96.4	98.5	94.09%
Wang et al. [88]	70.4	94.3	80	94.4	87	98	87.4%
Hsu et al. [91]	86.7	96.6	68.0	97.1	75.0	97.6	86.8%
Zhang et al. [74]	90	83	65	80	60	77	75.83%
Gu et al. [21]	63	91.9	84.8	89.8	86.4	97	86.63%
Poursaberi et al. [10]	87.0	91.9	91	96.9	84.6	91.2	90.38
Zhong et al. [11]	71.4	95.3	81.1	95.4	88.0	98.3	88.26%
Zhang and Tjondronegoro [77]	87.1	90.2	92	98.1	91.5	100	93.14%
Zhao and Zhang [36]	97.6	94.2	99.6	95.5	89.8	97.2	95.66%
Jain et al. [4]	76.7	81.5	94.4	98.6	77.2	99.1	87.90%
Song et al. [12, 13]	90.6	86.0	84.6	93.6	90.2	92.3	89.56%
Wu et al. [5]	82.9	67.7	66.7	87.7	78.4	87.9	78.55%
Shan et al. [6]	85.1	97.5	79.9	97.5	74.7	97.3	88.83%
Uddin et al. [13]	82.5	97.5	95	100	92.5	92.5	93.33%

because of differences in method, subjects, number of features, classifier, number of classes, number of images used as well as differences in partitioning the datasets. However, the results of each method can be analysed by taking the recognition result from their respective articles and tabulated as in Table 14. Among all the state-of-the-art methods, the proposed approach obtained very good recognition rates using the JAFFE database. Table 14 shows that the proposed approach reached a recognition rate of 96.17%, which is significantly higher than the rates obtained by the other methods. The number of features used to recognize emotions in the literature varies from 11 to 185. However, the proposed approach considers only four fuzzy features and, even so, attained a recognition accuracy of 96.17%.

The proposed approach partitioned the used input database through tenfold cross-validation and obtained a recognition accuracy of 96.17% using the JAFFE dataset. Tsai and Chang [76] achieved a recognition accuracy of 95.71%, which is close to the one obtained by the proposed approach. Other approaches, like those proposed by Happy and Routray [9], Rahulamathavan et al. [23], Zhang and Tjondronegoro [77], Gupta et al. [30] and Zhang et al. [37], used leave-one-out and conventional approaches for the input dataset partition and achieved more than 90% of recognition accuracy. Zhao and Zhang [36] and Gu et al. [15] suggested other approaches based on cross-validation and obtained performances below 90%. Based on all these findings, the recognition rate of the proposed approach, which used minimal feature points and the best cross-

**Table 14** Recognition rates obtained by different approaches using the JAFFE database

Approach	Method	No. of features	Images	Accuracy (%)
Proposed work	Fuzzy geometry	4	184 (without neutral)	96.17 (Cross-validation)
Tsai and Chang [76]	HOG + U-LTP	N/A		95.71
Happy and Routray [9]	Appearance feature	18 facial patches	183	91.7
Rahulamathavan et al. [23]	LFDA (in the encrypted domain)	40	213	94.37 (leave-one-out)
Zhang et al. [37]	Local binary pattern + LFDA	11	213	90.70 (leave-one-out)
Gupta et al. [30]	Hybrid (discrete cosine transform + Gabor filter + Wavelet transform + Gaussian distribution)	Unknown	213	93.40 (Conventional)
Zhao and Zhang [36]	Local binary pattern + KDIsoMap	20	213	81.59 (Cross-validation)
Zhang and Tjondronegoro [77]	Patch-based Gabor	185	203	93.48 (leave-one-out)
Gu et al. [15]	Radial encoded Gabor jets	49	213	89.67 (Cross-validation)

**Table 15** Results of recognition rates of the different approaches using the CK++ database

Approach	Method	No. of feature	Accuracy (%)
Proposed work	Fuzzy geometry	4 features	98.32% (Cross-validation)
Ghimire et al. [68]	LBP + NCM features	29 local features	97.25%
Ghimire et al. [69]	Salient geometric features	52 facial patches	98.30%
Wu and Lin [84]	GM + W-CR-AFM	–	89.84%
Vo and Le [85]	AlexNet + SVM	–	86.83%
Mollahosseini et al. [78]	CNN	–	93.2%
Elaiwat et al. [79]	Spatio-temporal	–	95.66%
Jung et al. [86]	CNN	–	80.6%
Li and Lam [80]	CNN	–	96.8%
Siddiqi et al. [81]	Stepwise linear discriminant analysis	–	96.83%
Zhang et al. [3]	Action Unit	56 features	90.38% (Cross-validation)
Pu et al. [41]	Active appearance model	–	89.37% (Cross-validation)
Happy and Routray [9]	Appearance feature	18 facial patches	94.1%
Ghimire and Lee [67]	HOG feature, ELM Ensemble	–	97.30%
Liu et al. [82]	Spatio-temporal	–	94.19%
Saeed et al. [56]	Geometric features, SVM classifier	8 facial key points	83.01%
Cruz et al. [87]	Temporal features	–	71.83%
Aifanti and Delopoulos [83]	Facial key point tracking	–	94.31%
Zhao and Zhang [36]	LBP + KDIsoMap	30 features	94.88 (Cross-validation)
Zhang and Tjondronegoro [77]	Patch-based Gabor	180 features	94.48 (Leave-one-out)
Gu et al. [15]	Radial encoded Gabor jets	49 features	91.51 (Cross-validation)

validation estimator, is encouraging relative to the rates of the other current methods.

Table 15 compares the recognition rates obtained by the various approaches using the CK++ database. The number of features used in the CK++ database to recognize the

expressions varies from 30 to 180. Meanwhile, the four fuzzy features used in the proposed approach through partitioning the input dataset using cross-validation achieved the highest accuracy of 98.32%. Ghimire et al. [68, 69] achieved an accuracy close to our result using 29

and 52 features, respectively. Zhao and Zhang [36], Zhang and Tjondronegoro [77] and Gu et al. [15] used 30 to 180 features, obtaining 90–95% of accuracy. Saeed et al. [56] used a low number of facial key points, only 8, but they only reached an accuracy of 83%. The other action units and appearance-based methods proposed by Zhang et al. [3], Pu et al. [41] and Happy and Routray [9], achieved 89–94% of accuracy. The recent classifiers used in Molahosseini et al. [78], Elaiwat et al. [79], Li and Lam [80], Siddiqi et al. [81], Ghimire and Lee [67], Liu et al. [82] and Aifanti and Delopoulos [83] obtained accuracies of 93–97%, and in Wu and Lin [84], Vo and Le [85], Jung et al. [86] and Cruz et al. [87] of 70–89.9%.

## 5 Conclusion

In this article, a new approach based on a minimum number of features extracted from the mouth region and without using a reference face was proposed to recognize human emotions. The extracted features were used to draw a quadrilateral for the face under analysis and the associated degree of impreciseness was addressed using the proposed mixed quadratic shape model through fuzzy membership functions. The proposed fuzzy membership functions were square, rhombus, kite and isosceles trapezoid. To validate the proposed approach, common learning methods were used to classify the human emotions and their recognition rates compared. The best recognition rates of the proposed approach were 96.17% and 98.32% for the JAFFE and CK++ datasets, respectively, and which were comparatively higher than the ones obtained by other recently proposed approaches.

The major advantages of the proposed approach are: only four facial features, fuzzy membership functions and fuzzy features are used to accurately identify the human emotions under evaluation here. The development of the proposed model based only on four fuzzy features reduced the computation time and space. The detection of the emotions without the need of a reference face brings this computational approach closer to the human system of perception. Finally, the evaluation with a competent cross-validator and with statistical tests confirmed the efficiency of our approach.

A future work will continue the proposed method but will also consider other factors like, age and gender, which also play vital roles in emotion recognition.

**Acknowledgements** João Manuel R.S. Tavares gratefully acknowledges the funding of Project NORTE-01-0145-FEDER-000022—SciTech—Science and Technology for Competitive and Sustainable Industries, co-financed by “Programa Operacional Regional do Norte” (NORTE2020), through “Fundo Europeu de Desenvolvimento Regional” (FEDER).

## Compliance with ethical standards

**Conflict of interest** The authors report no conflict of interest.

## References

- Mehrabian A (1968) Communication without words. *Psychol Today* 2(4):53–56
- Ekman P, Friesen W (1978) Action coding system: a technique for the measurement of facial movement. Consulting Psychologists Press, Palo Alto
- Zhang L, Mistry K, Jiang M, Neoh SC, Hossain MA (2015) Adaptive facial point detection and emotion recognition for a humanoid robot. *Comput Vis Image Underst* 140:93–114
- Jain S, Hu C, Aggarwal CK (2011) Facial expression recognition with temporal modeling of shapes. In: *IEEE international conference on computer vision workshops (ICCVWorkshops)*, Barcelona, pp 1642–1649
- Wu T, Bartlett M, Movellan JR (2010) Facial expression recognition using Gabor motion energy filters. In: *Proceedings of IEEE computer society conference on computer vision and pattern recognition workshops (CVPRW'10)*, pp 42–47
- Shan C, Gong S, McOwan P (2009) Facial expression recognition based on local binary patterns: a comprehensive study. *Image Vis Comput* 27(6):803–816
- Anderson K, McOwan PW (2004) Robust real-time face tracker for use in cluttered environments. *Comput Vis Image Underst* 95(2):184–200
- Whitehill J, Bartlett M, Movellan J (2008) Automatic facial expression recognition for intelligent tutoring systems. In: *Proceedings of computer vision and pattern recognition workshops*, Anchorage, AK, USA
- Happy SL, Routray A (2015) Automatic facial expression recognition using features of salient facial patches. *IEEE Trans Affect Comput* 6(1):1–12
- Poursaberi A et al (2012) Gauss-Laguerre wavelet textural feature fusion with geometrical information for facial expression identification. *EURASIP J Image Video Process*. <https://doi.org/10.1186/1687-5281-2012-17>
- Zhong L, Liu Q, Yang P, Liu B, Huang J, Metaxas DN (2012) Learning active facial patches for expression analysis. In: *IEEE Conference on computer vision and pattern recognition (CVPR)*, Providence, RI, USA
- Song M, Tao D, Liu Z, Li X, Zhou M (2010) Image ratio features for facial expression recognition application. *IEEE Trans Syst Man Cybern Part B Cybern* 40(3):779–788
- Uddin M, Lee J, Kim T (2009) An enhanced independent component-based human facial expression recognition from video. *IEEE Trans Consum Electron* 55:2216–2224
- Valstar M, Patras I, Pantic M (2005) Facial action unit detection using probabilistic actively learned support vector machines on tracked facial point data. In: *IEEE computer society conference on computer vision and pattern recognition, CVPR workshops*, p 76
- Gu W, Venkatesh Y, Xiang C (2010) A novel application of self-organizing network for facial expression recognition from radial encoded contours. *SoftComput Fusion Found Methodol Appl* 14(2):113–122
- Belhumeur P, Hespanha J, Kriegman D (1997) Eigenfaces vs fisher face: recognition using class specific linear projection. *IEEE Trans Pattern Anal Mach Intell* 19(7):711–720
- Turk M, Pentland A (1991) Eigenfaces for recognition. *J Cogn Neurosci* 3(1):71–86

18. Heisele B, Ho P, Wu J, Poggio T (2003) Face recognition: component-based versus global approaches. *Comput Vis Image Underst* 91:6–21
19. Zou J, Ji Q, Nagy G (2007) A comparative study of local matching approach for face recognition. *IEEE Trans Image Process* 16(10):2617–2628
20. Majumder A, Behera L, Subramanian VK (2014) Emotion recognition from geometric facial features using self-organizing map. *Pattern Recogn* 47:1282–1293
21. Gu W et al (2012) Facial expression recognition using radial encoding of local Gabor features and classifier synthesis. *Int J Pattern Recogn* 45:80–91
22. Wang Z, Ruan Q (2010) Facial expression based orthogonal local fisher discriminant analysis. *Proc ICSP 2010*:1358–1361
23. Rahulmathavan Y, Phan RC-W, Chambers JA, Parish DJ (2013) Facial expression recognition in the encrypted domain based on local fisher discriminant analysis. *IEEE Trans Affect Comput* 4(1):83–92
24. Shih FY, Chuang CF, Wang PSP (2008) Performance comparisons of facial expression recognition in JAFFE database. *Int J Pattern Recognit* 22(3):445–459
25. Kazmi SB, Qurat-ul-Ain JMA (2012) Wavelet-based facial expression recognition using a bank of support vector machines. *Soft Comput* 16(3):369–379
26. Donato G, Bartlett MS, Hager JC, Ekman P, Sejnowski TJ (1999) Classifying facial actions. *IEEE Trans Pattern Anal Mach Intell* 21(10):974–989
27. Deng HB, Jin LW, Zhen LX, Huang JC (2005) A new facial expression recognition method based on local Gabor filter bank and PCA plus LDA. *Int J Inf Technol* 11(11):86–96
28. Ilbeygi M, Hosseini HS (2012) A novel fuzzy facial expression recognition system based on facial feature extraction from color face images. *Eng Appl Artif Intell* 25:130–146
29. Kharat GU, Dudul SV (2009) Emotion recognition from facial expression using neural networks. In: *Human–computer systems interaction, AISC 60*, pp 207–219
30. Gupta SK, Agrwal S, Meena YK, Nain N (2011) A hybrid method of feature extraction for facial expression recognition. In: *7th international conference on signal image technology & internet-based systems*, pp 422–425
31. Anderson K, McOwan PW (2006) A real-time automated system for the recognition of human facial expressions. *IEEE Trans Syst Man Cybern Part B Cybern* 36(1):96–105
32. Jarlier S, Grandjean D, Delplanque S, N'Diaye K, Cayeux I, Velazco MI, Sander D, Vuilleumier P, Scherer KR (2011) Thermal analysis of facial muscles contractions. *IEEE Trans Affect Comput* 2(1):2–9
33. Feng XY, Hadid A, Pietikainen M (2004) A coarse-to-fine classification scheme for facial expression recognition. In: *The 1st international conference on image analysis and recognition*, pp 668–675
34. Liu WF, Yi SJ, Wang YJ (2009) Automatic facial expression recognition based on local binary patterns of local areas. In: *WASE international conference on information engineering*, pp 197–200
35. Moore S, Bowden R (2009) The effects of pose on facial expression recognition. In *Cavallaro A, Prince S, Alexander D (eds) Proceedings of the British machine conference*. BMVA Press, pp 79.1–79.11
36. Zhao X, Zhang S (2011) Facial expression recognition based on local binary patterns and kernel discriminant isomap. *Sensors* 11:9573–9588
37. Zhang S, Zhao X, Lei B (2012) Robust facial expression recognition via compressive sensing. *Sensors* 12:3747–3761
38. Luo Y, Wu C-M, Zhang Y (2013) Facial expression recognition based on fusion feature of PCA and LBP with SVM. *Int J Light Electron Opt* 124(17):2767–2770
39. Khana RA, Meyer A, Konik H, Bouaka S (2013) Framework for reliable, real time facial expression recognition for low resolution images. *Pattern Recogn Lett* 34:1159–1168
40. Sohail ASM, Bhattacharya P (2011) Classifying facial expressions using level set method based lip contour detection and multi-class support vector machines. *Int J Pattern Recognit Artif Intell* 25(06):835–862
41. Pu X, Fan K, Chen X, Ji L, Zhou Z (2015) Facial expression recognition from image sequences using twofold random forest classifier. *Neurocomputing* 168:1173–1180
42. Luo R, Huang C, Lin P (2011) Alignment and tracking of facial features with component-based active appearance models and optical flow. In: *International conference on advanced intelligent mechatronics (AIM)*. IEEE, pp 1058–1063
43. Lanitis A, Taylor C, Cootes T (1997) Automatic interpretation and coding of face images using flexible models. *IEEE Trans Pattern Anal Mach Intell* 19:743–756
44. Xie X, Lam K-M (2009) Facial expression recognition based on shape and texture. *Pattern Recogn* 42:1003–1011
45. Zavaschi THH, Britto AS Jr., Oliveira LES, Koerich AL (2013) Fusion of feature sets and classifiers for facial expression recognition. *Expert Syst Appl* 40:646–655
46. Kobayashi H, Hara F (1997) Facial interaction between animated 3d face robot and human beings. In: *IEEE international conference on systems, man, and cybernetics. Computational cybernetics and simulation, vol 4*. IEEE, pp 3732–3737
47. Zhang Z, Lyons M, Schuster M, Akamatsu S (1998) Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron. In: *Proceedings of 3rd international conference on automatic face and gesture recognition*, pp 454–459
48. Kotsia I, Pitas I (2007) Facial expression recognition in image sequences using geometric deformation features and support vector machines. *IEEE Trans Image Process* 16(1):172–187
49. Valstar MF, Pantic M (2012) Fully automatic recognition of the temporal phases of facial actions. *IEEE Trans Syst Man Cybern Part B Cybern* 42(1):28–43
50. Kotsia I, Buciu I, Pitas I (2008) An analysis of facial expression recognition under partial facial image occlusion. *Image Vis Comput* 26:1052–1067
51. Zafeiriou S, Pita I (2008) Discriminant graph structures for facial expression recognition. *IEEE Trans Multimed* 10(8):1528–1540
52. Asthana A, Saragih J, Wagner M, Goecke R (2009) Evaluating AAM fitting methods for facial expression recognition. In: *Proceeding of the international conference on affective computing and intelligent interaction*, pp 1–8
53. Ghimire D, Lee J (2013) Geometric feature-based facial expression recognition in image sequences using multi-class AdaBoost and support vector machines. *Sensors* 13(77):14–7734
54. Moore S, Bowden R (2011) Local binary patterns for multi-view facial expression recognition. *Comput Vis Image Underst* 115:541–558
55. Rudovic O, Pantic M, Patras I (2012) Coupled Gaussian processes for pose-invariant facial expression recognition. *IEEE Trans Pattern Anal Mach Intell* 25:1357–1369
56. Saeed A, Al-Hamadi A, Niese R, Elzobi M (2014) Frame-based facial expression recognition using geometric features. *Adv Hum Comput Interact* 2014:1–13
57. Bartlett M, Littlewort G, Frank M, Lainscsek C, Fasel I, Movellan J (2005) Recognizing facial expression : machine learning and application to spontaneous behavior. *IEEE Comput Soc Conf Comput Vis Pattern Recognit* 2:568–573

58. Bargal SA, Barsoum E, Ferrer CC, Zhang C (2016) Emotion recognition in the wild from videos using images. In: ICM: Proceedings of the 18th ACM international conference on multimodal interaction. ACM
59. Mayor Torres JM, Stepanov EA (2017) Enhanced face/audio emotion recognition: video and instance level classification using ConvNets and restricted Boltzmann Machines. In: Proceedings of the international conference on web intelligence. ACM
60. Tarnowski P et al (2017) Emotion recognition using facial expressions. In: International conference on computational science—ICCS 2017, 12–14 June 2017, Zurich, Switzerland
61. Zhang Z (1999) Feature-based facial expression recognition: sensitivity analysis and experiments with a multi layer perceptron. *Int J Pattern Recognit Artif Intell* 13:893–911
62. Ding W, Xu M, Huang D, Lin W, Dong M, Yu X, Li H (2016) Audio and face video emotion recognition in the wild using deep neural networks and small datasets. In: ICMI 2016: proceedings of the 18th ACM international conference on multimodal interaction. ACM
63. Huang Y, Lu H (2016) Deep learning driven hypergraph representation for image-based emotion recognition. In: Proceedings of the 18th ACM international conference on multimodal interaction. ACM
64. Barros P et al (2017) Emotion-modulated attention improves expression recognition: a deep learning model. *Neurocomputing* 253:104–114
65. Rosenblum M, Yacoob Y, Davis L (1996) Human expression recognition from motion using a radial basis function network architecture. *IEEE Trans Neural Netw* 7(5):1121–1138
66. Vadivel A, Shanthi P, Shaila SG (2015) Estimating emotions using geometric features from facial expressions, 3rd edn. *Encyclopedia of Information Science and Technology*, p 8
67. Ghimire D, Lee J (2014) Extreme learning machine ensemble using bagging for facial expression recognition. *J Inf Process Syst* 10(3):443–458
68. Ghimire D, Lee J, Li Z-N, Jeong S (2017) Recognition of facial expressions based on salient geometric features and support vector machines. *Int J Multimed Tools Appl* 76:7921–7946
69. 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* 76:7803–7821
70. Nielsen JA, Zielinski BA, Ferguson MA, Lainhart JE, Anderson JS (2013) An evaluation of the left-brain vs. right brain hypothesis with resting state functional connectivity magnetic resonance imaging. *PLOS ONE. Cogn Neurosci-Connect*. <https://doi.org/10.1371/journal.pone.0071275> Featured in PLOS Collections
71. Barthomeuf L, Droit-Volet S, Rousset S (2012) How emotions expressed by adults' faces affect the desire to eat liked and disliked foods in children compared to adults. *Br J Dev Psychol* 30:253–266
72. García HF, Álvarez MA, Orozco AA (2017) Dynamic facial landmarking selection for emotion recognition using Gaussian processes. *J Multimodal User Interfaces* 11:327–340
73. Ithaya Rani P, Muneeswaran K (2017) Recognize the facial emotion in video sequences using eye and mouth temporal Gabor features. *J Multimed Tools Appl* 76:10017–10040
74. Zhang L, Jiang M, Farid D, Hossain MA (2013) Intelligent facial emotion recognition and semantic-based topic detection for a humanoid robot. *Expert Syst Appl* 40:5160–5168
75. Gaidhane VH, Hote YV, Singh V (2016) Emotion recognition using eigenvalues and Levenberg–Marquardt algorithm-based classifier. *Indian Acad Sci* 41(4):415–423
76. Tsai H-H, Chang Y-C (2017) Facial expression recognition using a combination of multiple facial features and support vector machine. *Int J Soft Comput*. <https://doi.org/10.1007/s00500-017-2634-3>
77. Zhang L, Tjondronegoro D (2011) Facial expression recognition using facial movement features. *IEEE Trans Affect Comput* 2(4):219–229
78. Mollahosseini A, Chan D, Mahoor MH (2016) Going deeper in facial expression recognition using deep neural networks. In: Proceedings of IEEE winter conference on applications of computer vision (WACV), Lake Placid, NY, USA, pp 1–10
79. Elaiwat S, Bennamoun M, Boussaid F (2016) A spatio-temporal RBM-based model for facial expression recognition. *Int J Pattern Recognit* 49:152–161
80. Li J, Lam EY (2015) Facial expression recognition using deep neural networks. In: Proceedings of IEEE international conference on imaging systems and techniques (IST), pp 1–6
81. Siddiqi MH, Ali R, Khan AM, Park Y-T, Lee S (2015) Human facial expression recognition using stepwise linear discriminant analysis and hidden conditional random fields. *IEEE Trans Image Process* 24(4):1386–1398
82. Liu M, Shan S, Wang R, Chen X (2014) Learning expression lets on spatio-temporal manifold for dynamic facial expression recognition. In: Proceedings of IEEE conference on computer vision and pattern recognition (CVPR), Columbus, OH, USA, 2014, pp 1749–1756
83. Aifanti N, Delopoulos A (2014) Linear subspace for facial expression recognition. *Signal Process Image Commun* 29:177–188
84. Wu B-F, Lin C-H (2017) Adaptive feature mapping for customizing deep learning based facial expression recognition model. *IEEE access* 2017
85. Vo DM, Le TH (2016) Deep generic features and SVM for facial expression recognition. In: 3rd National Foundation for Science and Technology Development Conference on Information and Computer Science (NICS), pp 80–84
86. Jung H, Lee S, Park S, Kim B, Kim J, Lee I, Ahn C (2015) Development of deep learning-based facial expression recognition system. In: 21st Korea-Japan joint workshop on frontiers of computer vision (FCV), pp 1–4
87. Cruz AC, Bhanu B, Thakoor NS (2014) Vision and attention theory based sampling for continuous facial motion recognition. *IEEE Trans Affect Comput* 5:418–431
88. Wang H, Huang H, Makedon F (2014) Emotion detection via discriminant laplacian embedding. *Univ Access Inf Soc* 13:23–31
89. Agarwal S, Mukherjee DP (2017) Facial expression recognition through adaptive learning of local motion descriptor. *Int J Multimed Tools Appl* 76:1073–1099
90. Hsieh CC, Hsieh MH, Jiang MK, Cheng YM, Liang EH (2015) Effective semantic features for facial expressions recognition using SVM. *Int J Multimed Tools Appl* 75(11):6663–6682
91. Hsu FS, Lin WY, Tsai TW (2014) Facial expression recognition using bag of distances. *Int J Multimed Tools Appl* 73(1):309–326

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.