Pattern Anal Applic (2013) 16:41–54
DOI 10.1007/s10044-012-0286-6

THEORETICAL ADVANCES

Facial emotional classification: from a discrete perspective to a continuous emotional space

Isabelle Hupont · Sandra Baldassarri · Eva Cerezo

Received: 27 March 2010 / Accepted: 10 July 2012 / Published online: 27 July 2012

Abstract User emotion detection is a very useful input to develop affective computing strategies in modern human computer interaction. In this paper, an effective system for facial emotional classification is described. The main distinguishing feature of our work is that the system does not simply provide a classification in terms of a set of discrete emotional labels, but that it operates in a continuous 2D emotional space enabling a wide range of intermediary emotional states to be obtained. As output, an expressional face is represented as a point in a 2D space characterized by evaluation and activation factors. The classification method is based on a novel combination of five classifiers and takes into consideration human assessment for the evaluation of the results. The system has been tested with an extensive universal database so that it is capable of analyzing any subject, male or female of any age and ethnicity. The results are very encouraging and show that our classification strategy is consistent with human brain emotional classification mechanisms.

Keywords Affective computing · Algorithms · Facial expression analysis · Intelligent user interfaces

I. Hupont (✉)
Multimedia Technologies Division,
Aragon Institute of Technology, Zaragoza, Spain
e-mail: ihupont@ita.es

S. Baldassarri · E. Cerezo
Computer Science and Systems Engineering Department,
Engineering Research Institute of Aragon (I3A),
University of Zaragoza, Zaragoza, Spain
e-mail: sandra@unizar.es

E. Cerezo
e-mail: ecerezo@unizar.es

1 Introduction

Human computer intelligent interaction is an emerging field aimed at providing natural ways for humans to use computers as aids. It is argued that for a computer to be able to interact with humans it needs to have the communication skills of humans. One of these skills is the affective aspect of communication [3]. The most expressive way humans display emotions is through facial expressions. Facial expression is the most powerful, natural and direct way used by humans to communicate and understand each other's affective state and intentions [19]. Thus, the interpretation of facial expressions is the most common method used for emotional detection and forms an indispensable part of affective human computer interface (HCI) designs. However, developing a system that correctly interprets facial expressions is a difficult task. It initially involves extracting certain information from the face that is subsequently used to feed the classification system that will judge the affectivity of the facial expression. Therefore, to design such a system three main decisions must be made, relating to:

1. *The facial model* to feed the classification system, it is necessary to determine whether it is enough to select a small set of facial characteristics or whether the face as a whole is to be taken into account to achieve an accurate analysis.
2. *The description level* a set of categories must be defined (level of emotional description) to be used for the classification of facial expressions.
3. *Classification mechanisms* a method must be established to categorise the facial posture shown, in terms of the defined description level and based on the facial information taken from the face model.

Regarding facial modelization, a careful choice of the facial information to be used as input will affect the quality

and speed of the system. Some studies suggest that all the necessary information for the recognition of expressions is contained in the deformation of a set of carefully selected characteristics of the eyes, mouth and eyebrows [9]. This idea, known as the "feature-based approach", has given rise to several works [16, 20, 27]. Its main advantage is that the computational time required to process the facial information is short. Nevertheless, other studies state that the face must be considered as a whole ("appearance-based approach"), and this involves the consideration of more facial information than studying the displacement of a set of points [1, 38]. This supposed improvement in the quality of facial information is generally detrimental to the speed of the system, critical in an HCI system operating in real time. As a compromise between the advantages and disadvantages of the two approaches, other studies have developed hybrid systems for facial expression recognition [25, 26, 43].

In the second place, classification requires a clear definition of the level of emotional description to work with. Two main streams in current research on automatic analysis of facial expressions consider either discrete categories of facial affect (emotion labels) or location in a continuous 2D emotional space:

- *Discrete categories of facial affect* Perhaps the most long-standing way that facial affect has been described by psychologists is in terms of discrete categories, an approach that is rooted in the language of daily life. The most commonly used emotional categories in expression recognition systems are the six universal emotions proposed by Ekman [9] which include "happiness", "sadness", "fear", "anger", "disgust" and "surprise". Examples of studies using this categorization are [16, 23, 40]. The labeling scheme based on categories is very intuitive and thus matches peoples' experience. However, discrete lists of emotions fail to describe the wide range of emotions that occur in daily communication settings. There are a few tentative efforts to detect non-basic affective states from deliberately displayed facial expressions, including "fatigue" [17], and mental states such as "agreeing", "concentrating", "interested", "thinking", "confused", and "frustrated" [18, 41]. However, a set of emotions is a mere list of labels with no real link between them. It does not represent a dimensional space and has no algebra: every emotion must be studied and recognized independently.
- *Location in continuous 2D emotional space* To overcome the problems cited above, some researchers, such as Whissell [36] and Plutchik [29], prefer to view affective states related to one another in a systematic manner. They consider emotions as a continuous 2D space which dimensions are evaluation and activation. The evaluation dimension measures how a human feels, from positive to negative. The activation dimension measures whether humans are more or less likely to take some action under the emotional state, from active to passive. Dimensional representations are attractive because they provide a way of describing a wide range of emotional states. In real life scenarios, emotional states do not jump from one universal emotion label to another. They rather vary over time crossing many intermediate emotions. Dimensional approaches are much more able to deal with non-discrete emotions and variations in emotional states over time [14]. However, very few works have chosen a dimensional description level, and the few that do are more related to the design of synthetic faces [33], data processing [8] or psychological studies [13] than to emotion recognition. Moreover, in existing affective recognition works the problem is simplified to a two-class (positive vs. negative and active vs. passive) [11], a four class (quadrants of 2D space) [4], or other multiclass (where classes represent different affective space sub-areas) classification [24], thereby the descriptive potential of 2D space is lost.

Finally, the third main decision is related to the mechanism used for classifying emotions. Independently of the facial model and the level of description chosen, in the literature most facial expression analyzers obtain better emotional classification performances using neural networks, rule-based expert systems, support vector machines and Bayesian nets based classifiers. In [42], an excellent state-of-the art summary is given of the various methods recently used in facial expression emotional recognition. However, it is difficult to compare the effectiveness of existing emotional classification mechanisms because in most cases different data sets and different assessment criteria are used. An additional problem is that the majority of studies in the literature select only one type of classifier for emotional detection, or at the most compare different classifiers and then use that which provides the best results [22].

In this paper, an effective system for facial emotional classification is described. The face modeling selected as input for the system follows a feature-based approach: the inputs of the classifiers are a set of facial distances and angles chosen on the basis of a feature selection technique, so that the face is modelled in a computationally simple way without losing relevant information about the facial expression. This is especially important when dealing with HCI systems working in real-time. For the description level, we start with a classification method in discrete categories that is subsequently expanded to be able to work in a continuous emotional 2D space and thus to consider a wide range of intermediate emotional states. With regard to

the classification itself, the system combines the outputs of different classifiers simultaneously using a weighted majority voting strategy. In this way, the overall risk of making a poor selection with a given classifier for a given input is reduced.

The main outstanding feature of our work is that the system is the first in the current state of the art that recognizes the location (coordinates) of the user's facial expression in the emotional 2D space. Other works using the evaluation-activation space for emotional classification confine themselves to providing information about the polarity of facial expression (positive/negative or active/passive) or the quadrant of space to which the image belongs. Another noteworthy feature of the work is that the system is tested with an extensive universal database showing individuals of different races and gender. Furthermore, human assessment is taken into consideration in the evaluation of the classification system. This type of study has not been adopted in other works, and it provides substantial added value to a system that deals with human-computer intelligent interaction.

The structure of the paper is the following: Sect. 2 describes the classification method into discrete categories. In Sect. 3 the step from the discrete perspective to the continuous emotional space is explained in detail and Sect. 4 comprises concluding remarks and a description of future work.

2 Discrete emotional classification: a novel combination of facial classifiers

In this section, an effective method is presented for the automatic classification of facial expressions into discrete emotional categories. The method is able to classify the user's emotion in terms of the six Ekman's universal emotions (plus "neutral"), giving a confidence value to each emotional category. This output will be the base for the further expansion to a 2D continuous emotional recognition.

Figure 1 illustrates the general process followed by the proposed method that is detailed in the subsections below: firstly, Sect. 2.1 explains the acquisition and extraction process of the input features necessary to the system; then, Sect. 2.2 describes the implemented emotional classification mechanism itself; finally, the results and their validation by human assessment are given in Sect. 2.3.

2.1 Data acquisition and facial features extraction

The first step of the method involves capturing the user's facial expressions with a camera (webcam, camcorder, etc.)

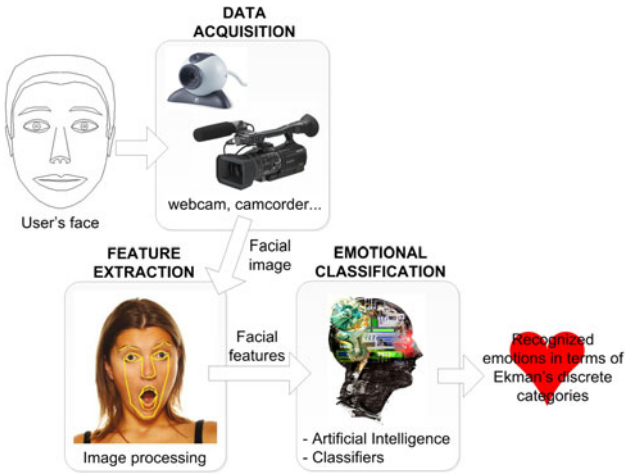

Fig. 1 Discrete facial emotional classification general process

and extracting certain information from the recorded facial image by means of image processing techniques. Facial action coding system (FACS) [10] was developed by Ekman and Friesen to code facial expressions in which the individual muscular movements in the face are described by action units (AUs). This work inspired many researchers to analyze facial expressions by means of image and video processing, where by tracking of facial features and measuring a set of facial distances and angles, they attempt to classify different facial expressions. This approach involves the development of a precise facial feature tracker and the careful selection of the most appropriate facial distances and angles to achieve classification.

Regarding facial trackers, although some of the detectors presented in the literature seem to perform quite well when localizing a small number of facial feature points such as the corners of the eyes and the mouth, none of them detects more than 20 facial feature points [5, 27, 32] and, more importantly, none performs the detection with high accuracy. Most of them are limited in terms of occlusions, fast movements, large head rotations, lighting, facial deformations, skin color, beards, glasses, etc. Moreover, most facial trackers work in 2D and very few are able to provide 3D facial features coordinates [32].

With regard to facial measures, existing works demonstrate that a high emotional classification accuracy can be obtained by analyzing a small set of facial distances and angles. Examples are the system of Soyel and Demirel [32] that studies six 3D facial distances; the work by Tang and Huang [34] that empirically demonstrates that 10–30 facial distances are sufficient to yield good facial emotional classification results; the method proposed by Hammal et al. [16] that analyzes a set of five 2D facial distances; or the approach of Chang et al. [5] that measures 12 feature distances.

Following that methodology, the initial inputs of our classifiers were established in a set of distances and angles obtained from 20 characteristic facial points. In fact, the inputs are the variations of these measures with respect to the "neutral" face. The chosen set of initial inputs compiles the distances and angles that have been proved to provide the best classification performance in existing works of the literature, such as the aforementioned. The points are obtained thanks to faceAPI [30], a commercial real-time facial feature tracking program that provides cartesian facial 3D coordinates. It is able to track up to ±90° of head rotation and is robust to occlusions, lighting conditions, presence of beard, glasses, etc. Figure 2a shows the correspondence of these points with those defined by the MPEG4 standard. The initial set of parameters obtained from faceAPI's 3D points information is shown in Fig. 2b. In order to make the distance values consistent (independently of the scale of the image, the distance to the camera, etc.) and independent of the expression, all the distances are normalized with respect to the distance between the eyes (MPEG4 Facial Animation Parameter Unit -FAPU- called "ESo"). The choice of angles provides a size invariant classification and saves the effort of normalization.

In order to determine the goodness and usefulness of the parameters, a study of the correlation between them was carried out using the data (distance and angle values) obtained from a set of training images. For this purpose, two different facial emotion databases were used: the FGNET database [35] that provides spontaneous (non-acted) video sequences of 19 different young Caucasian people, and the MMI Facial Expression Database [28] that holds 1,280 acted videos of 43 different subjects from different races (Caucasian, Asian, South American and Arabic) and ages ranging from 19 to 62. Both databases show Ekman's six universal emotions plus the "neutral" one and provide expert annotations about the emotional apex frame of the video sequences. A new database has been built for this work with a total of 1,500 static frames selected from the apex of the video sequences from the FG-NET and MMI databases. It has been used as a training set in the correlation study and in the tuning of the classifiers. A correlation-based feature selection technique [15] was carried out in order to identify the most influential parameters in the variable to predict (emotion) as well as to detect redundant and/or irrelevant features. Subsets of parameters that are highly correlated with the class while having low intercorrelation are preferred. A set of important conclusions were extracted from the results: (a) symmetrical distances (e.g. LD5 and RD5) are highly correlated and thus redundant; (b) distance D3 and angle A2 also present a high correlation value; (c) angles LA3 and RA3 are not influential for achieving the emotional classification. Therefore, from the initial set of parameters, only the most significant ones were selected to work with: RD1, RD2, RD5, D3, D4, D6 and A1 (marked in bold in Fig. 2b). This kind of feature selection process is not carried out in other existing works even though it is important since it reduces the number of irrelevant, redundant and noisy inputs in the model and thus computational time, without losing relevant facial information.

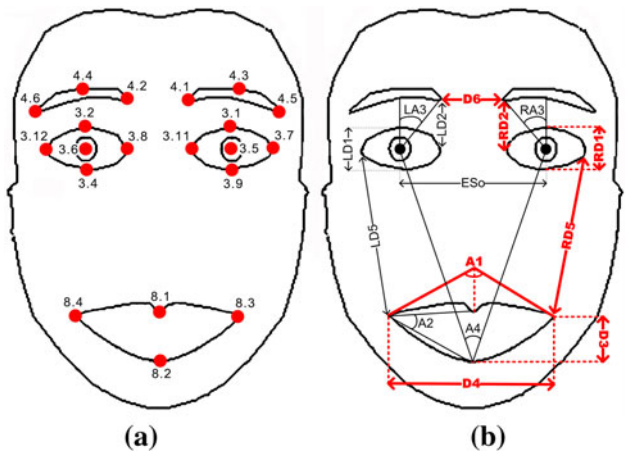

Fig. 2 **a** Tracked facial feature points according to MPEG4 standard and **b** corresponding facial parameters tested (in *bold*, the final selected parameters)

2.2 Discrete facial emotional classification

Once the facial input features have been extracted, the second step of the method involves the implementation of a classification system that will judge the affectivity of the facial expression in terms of the six Ekman's universal emotions (plus "neutral"). The implemented classification mechanism intelligently combines the outputs of different well-known Artificial Intelligence classifiers. Section 2.2.1 describes the criteria taken into account when selecting the various classifiers which are then combined in the way explained in Sect. 2.2.2.

2.2.1 Selection of classifiers

In order to select the best classifiers, the Waikado Environment for Knowledge Analysis (Weka) tool was used [39]. It provides a collection of machine learning algorithms for data mining tasks. From this collection, five classifiers were selected after tuning and benchmarking: RIPPER, Multilayer Perceptron, SVM, Naive Bayes and C4.5. The selection was based on their widespread use as well as on the individual performance of their Weka implementation. A tenfold cross-validation test over the 1,500 training images has been performed for each selected classifier. The success rates obtained for each classifier and

Table 1 Success rates obtained with a tenfold cross-validation test over the 1,500 training images for each individual classifier and each emotion (first five rows) and when combining the five classifiers (sixth row)

	Disgust (%)	Joy (%)	Anger (%)	Fear (%)	Sadness (%)	Neutral (%)	Surprise (%)
RIPPER	50.00	85.70	66.70	48.10	26.70	80.00	80.00
SVM	76.50	92.90	55.60	59.30	40.00	84.00	82.20
C4.5	58.80	92.90	66.70	59.30	30.00	70.00	73.30
Naive Bayes	76.50	85.70	63.00	85.20	33.00	86.00	71.10
Multilayer Perceptron	64.70	92.90	70.40	63.00	43.30	86.00	77.80
Combination of classifiers	94.12	97.62	81.48	85.19	66.67	94.00	95.56

each emotion are shown in the first five rows of Table 1. As can be observed, in general, the "correct" percentages obtained individually by the classifiers are not very favorable, especially for "disgust", "sadness", "anger" and "fear". However, each classifier is very reliable for detecting certain specific emotions but not so much for others. For example, the C4.5 is excellent at identifying "joy" (92.90 % correct) but is only able to correctly detect "fear" on 59.30 % of occasions, whereas Naive Bayes is way above the other classifiers for "fear" (85.20 %), but is below the others in detecting "joy" (85.70 %) or "surprise" (71.10 %). It would, therefore, appear that an intelligent combination of the five classifiers in such a way that the strong and weak points of each are taken into account could be a good solution for developing a method with a high success rate.

2.2.2 *Combination of classifiers*

When dealing with matters of great importance, people often seek a second opinion before making a decision, sometimes a third and sometimes many more. In doing so, the individual opinions are weighed up and combined through some sort of thought process before a final decision that is presumably the most informed one is reached. Following this idea, the combination of the outputs of several classifiers by averaging may reduce the risk of an unfortunate selection of a poorly performing classifier. The averaging may or may not surpass the performance of the best classifier in the ensemble, but it certainly reduces the overall risk of making a particularly poor selection.

The classifier combination chosen follows a weighted majority voting strategy. The voted weights are assigned depending on the performance of each classifier for each emotion. From each classifier, a confusion matrix formed by elements $P_{jk}(E_i)$, corresponding to the probability of having emotion i knowing that classifier j has detected emotion k, is obtained. The probability assigned to each emotion $P(E_i)$ is calculated as:

$$P(E_i) = \frac{P_{1k'}(E_i) + P_{2k''}(E_i) + \cdots + P_{5k^v}(E_i)}{5}, \quad (1)$$

where $k', k'' \ldots k^v$ are the emotions detected by classifiers $1, 2 \ldots 5$, respectively.

The assignment of the final output confidence value corresponding to each basic emotion is done following two steps:

1. Firstly, the confidence value $\mathrm{CV}(E_i)$ is obtained by normalizing each $P(E_i)$ to a 0 through 1 scale:

$$\mathrm{CV}(E_i) = \frac{P(E_i) - \min\{P(E_i)\}}{\max\{P(E_i)\} - \min\{P(E_i)\}}, \quad (2)$$

where

- $\min\{P(E_i)\}$ is the greatest $P(E_i)$ that can be obtained by combining the different $P_{jk}(E_i)$ verifying that $k \neq i$ for every classifier j. In other words, it is the highest probability that a given emotion can reach without ever being selected by any classifier.
- $\max\{P(E_i)\}$ is that obtained when combining the $P_{jk}(E_i)$ verifying that $k = i$ for every classifier j. In other words, it is the probability that obtains a given emotion when selected by all the classifiers unanimously.

2. Secondly, a rule is established over the obtained confidence values in order to detect and eliminate emotional incompatibilities. The rule is based on the work of Plutchik [29], who assigned "emotional orientation" values to a series of affect words. For example, two similar terms (like "joyful" and "cheerful") have very close emotional orientation values while two antonymous words (like "joyful" and "sad") have very distant values, in which case Plutchik speaks of "emotional incompatibility". The rule to apply is the following: if emotional incompatibility is detected, i.e. two non-null incompatible emotions exist simultaneously, that chosen will be the one with the closer emotional orientation to the rest of the non-null detected

emotions. For example, if "joy", "sadness" and "disgust" coexist, "joy" is assigned zero since "disgust" and "sadness" are emotionally closer according to Plutchik.

2.3 Discrete facial emotional classification results

This section presents the classification results obtained when applying the previous facial affect recognition method to the 1,500 images of the database and the validation strategies used for validation purposes. Section 2.3.1 reports the obtained success rates and presents how users' assessment is then taken into account to analyze classification performance. Section 2.3.2 compares the results with the ones obtained in other existing representative works.

2.3.1 Results: success rates and human assessment

The initial results obtained when applying the strategy explained in the previous section to combine the scores of the five classifiers with a tenfold cross-validation test are shown in the sixth row of Table 1. As can be observed, the success rates for the "neutral", "joy", "disgust" and "surprise" emotions are very high (94.00–97.62 %). However, the system tends to confuse "fear" with "surprise" and "anger" with "sadness" or "disgust"; therefore, the performances for those emotions are slightly worse. Confusion between these pairs of emotions occurs frequently in the literature and for this reason many classification works do not consider them. The lowest result of our classification is for "sadness": it is confused with the "neutral" emotion on 20 % of occasions, due to the similarity of their facial expressions. Nevertheless, the results can be considered positive as emotions with distant "emotional orientation" values (such as "disgust" and "joy" or "neutral" and "surprise") are confused on less than 2.5 % of occasions and incompatible emotions (such as "sadness" and "joy" or "fear" and "anger") are never confused. Table 2 shows the confusion matrix obtained after the combination of the five classifiers.

Another relevant aspect to be taken into account when evaluating the results is human opinion. The labels provided in the database for training classifiers correspond to the real emotions felt by users although they do not necessarily have to coincide with the perceptions other human beings may have about the facial expressions shown. Undertaking this kind of study is very important when dealing with human-computer interaction, since the system is proved to work in a similar way to the human brain. However, such studies are not performed in other classification works.

In order to take into account the human factor in the evaluation of the results, 60 persons were told to classify the 1,500 images of the database in terms of emotions. As a result, each one of the frames was classified by ten different persons in five sessions of 50 images. With this information, the evaluation of the results was repeated: the recognition was marked as "good" if the decision was consistent with that reached by the majority of the human assessors. It is important to realize that, according to Bassili [2], a trained observer can correctly classify facial emotions with an average of 87 %.

For example, in the image shown in Fig. 3, the FG-NET database classifies it exclusively as "disgust" while the assessors recognized it as "anger" and "sadness" as often as "disgust". The users' results are similar to those of our method which obtains a confidence value of 0.83 for "anger", 0.51 for "disgust" and 0.35 for "sadness".

The results of considering users' assessment are shown in the second row of Table 3. As can be seen, the success ratios have considerably increased. Therefore, it can be concluded that the confusions of the algorithms go in the same direction as those of the users': our classification strategy is consistent with human classification.

2.3.2 Comparison with other representative methods

In order to demonstrate the outstanding position of the presented discrete emotional classification method within the current state of the art, Table 4 compares it with other representative existing approaches. Those works have been chosen for comparison due to the following reasons:

Table 2 Confusion matrix obtained combining the five classifiers

Emotion ⟶ is classified as ↓	Disgust (%)	Joy (%)	Anger (%)	Fear (%)	Sadness (%)	Neutral (%)	Surprise (%)
Disgust	**94.12**	0.00	2.94	2.94	0.00	0.00	0.00
Joy	2.38	**97.62**	0.00	0.00	0.00	0.00	0.00
Anger	7.41	0.00	**81.48**	0.00	7.41	3.70	0.00
Fear	3.70	0.00	0.00	**85.19**	3.70	0.00	7.41
Sadness	6.67	0.00	6.67	0.00	**66.67**	20.00	0.00
Neutral	0.00	0.00	2.00	2.00	2.00	**94.00**	0.00
Surprise	0.00	0.00	0.00	2.22	0.00	2.22	**95.56**

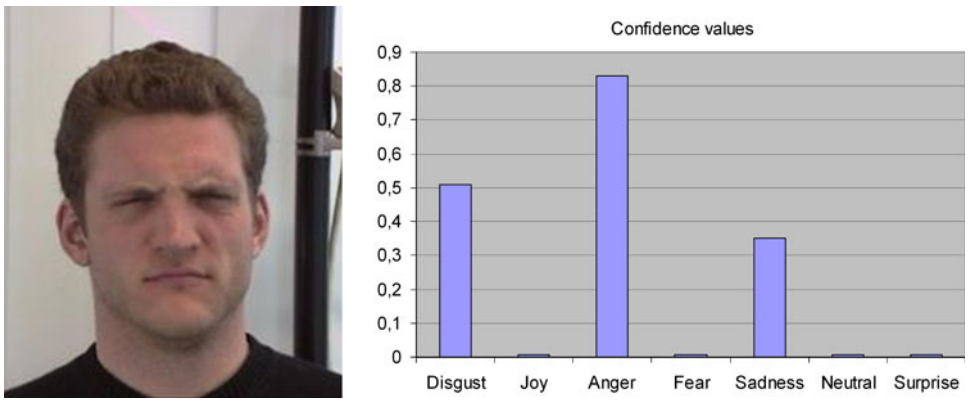

Fig. 3 Frame classified as "disgust" by the FG-NET database [35]. Our method classifies it as a mixture of "anger", "disgust" and "sadness"

Table 3 Comparison of results obtained when combining the five classifiers and after considering human assessment

	Disgust (%)	Joy (%)	Anger (%)	Fear (%)	Sadness (%)	Neutral (%)	Surprise (%)
Combination of classifiers	94.12	97.62	81.48	85.19	66.67	94.00	95.56
Combination of classifiers + human assessment	97.06	97.62	88.89	96.30	86.67	98.00	97.78
Growth rate	3.12	0.00	9.09	13.04	30.00	4.26	2.32

Table 4 Success rates of the presented facial images discrete classification method (in gray), both with and without taking into account human assessment, and comparison with other representative works

	Proposed method		Method of Hammal et al. [16]	Method of Datcu and Rothkrantz [7]	Method of Cohen et al. [6]	Method of Zhang et al. [44]
Type of classifier	Combination		rule-based	SVM	Bayesian net	Neural Network
Database	1,500 frames, 62 subjects		630 frames, 8 subjects	474 frames	40 subjects	213 frames, 9 japanese females
Validation strategy	Tenfold cross-validation		hold-out method	Twofold cross-validation	leave-one-out cross-validation	Tenfold cross-validation
User assessment	No	Yes	No	No	No	Yes
Sucess rates						
Joy	97.62%	97.62%	87.26%	72.64%	97.00%	
Surprise	95.56%	97.78%	84.44%	83.80%	85.00%	90.10%
Disgust	94.12%	97.06%	51.72%	80.35%	88.00%	The only available data is the overall recognition rate of the 6 + "neutral" universal emotions.
Anger	81.48%	88.89%	Not recognized	75.86%	80.00%	
Sadness	66.67%	86.67%	Not recognized	82.79%	85.00%	
Fear	85.19%	96.30%	Not recognized	84.70%	93.00%	
Neutral	94.00%	98.00%	88.00%	Not recognized	96.00%	

- They have the same experimental proposal, i.e. classifying facial images in terms of Ekman's universal emotions.
- Each work has the best performance using a specific type of classifier from the ones that have proved to obtain better results in the literature. The selected works achieve emotional classification by means of a Neural Network [44], a rule-based expert system [16], a support vector machine (SVM) [7] and a Bayesian net [6], respectively.
- They detail the evaluation strategy followed to obtain classification results. Many works do not detail whether they have used or not cross-validation, so the direct comparison of results is not always possible.

As it can be observed, the success rates of our method are better even when a more strict validation strategy is used. Besides, other methods' confusion matrices uniformly distribute classification errors among all emotions. For example, in [16] and [7] the probability of confusing two incompatible emotions, such as "joy" and "sadness", is similar to the one of confusing other pairs of emotionally distant affective states, such as "joy" and "surprise". The other methods [6, 44] do not even provide information about confusions among emotions. It is also important to realize that the database used in this work is bigger than the used in the other ones and, therefore, more universal.

3 Expansion to a dimensional description of affect: a 2D emotional space for the extraction of continuous emotional information

As discussed in the "Sect. 1", the use of a discrete set of emotions (labels) for emotional classification has important limitations. To avoid these limitations and expand the emotional output information from the system in terms of intermediate emotions, use has been made of one of the most influential evaluation-activation 2D models in the field of psychology: that proposed by Whissell [36]. Section 3.1 describes this model and the methodology followed to obtain as final output of the system the (x, y) coordinates in the evaluation-activation space of the analyzed facial expression. The general results of emotional classification obtained in the 2D space are presented in Sect. 3.2 and analyzed in greater detail in Sect. 3.3 taking human assessment into account.

3.1 From a discrete perspective of emotions to a continuous emotional space through Whissell's model

The "Dictionary of Affect in Language" (DAL) [36] is a tool developed by the psychologist Cynthia Whissell to quantify the emotional meanings of words. Words included in the DAL were selected in an ecologically valid manner from an initial corpus of more than one million words following the process detailed in [37]. Whissell assigns a pair of values $\langle evaluation, activation \rangle$ to each word of the DAL. These emotional coordinates were calculated on the basis of volunteers' responses to the words along the two emotional dimensions of activation and evaluation. As a result, the most current form of the DAL contains close to 9,000 emotional words which volunteers rated along each of these two dimensions in roughly a quarter of a million separate judgements. Figure 4 graphically shows as an example the position of some of the DAL words in the evaluation-activation space, where both affective dimensions are bounded

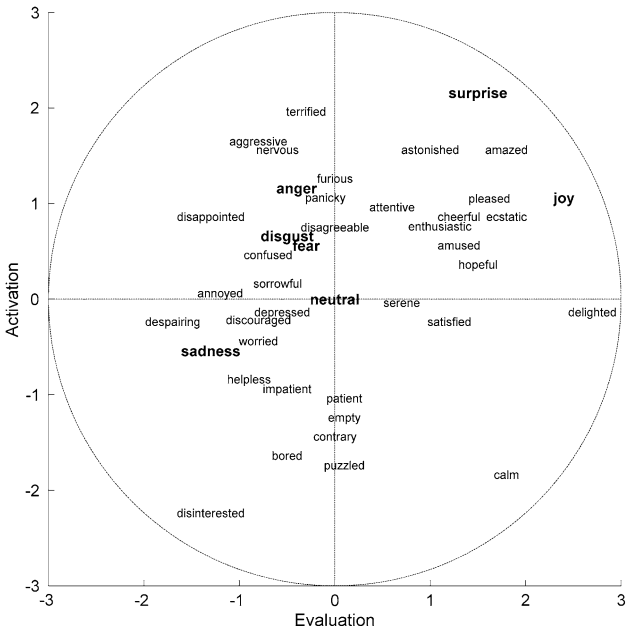

Fig. 4 Simplified Whissell's evaluation-activation space: position of some of the DAL words in the 2D emotional space are shown

within a range of $[-3, +3]$ and the "neutral" state is placed at original point $(0, 0)$.

The next step is to build an emotional mapping so that an expressional face image can be represented as a point on this plane whose coordinates (x, y) characterize the emotion property of that face. It can be seen that the emotion-related words corresponding to each one of six Ekman's emotions have a specific location (x_i, y_i) in the Whissell space (in bold in Fig. 4). Thanks to this, the output information of the classifiers (confidence value of the facial expression to each emotional category) can be mapped in the space. This emotional mapping is carried out considering each of six Ekman's basic emotions plus "neutral" as 2D weighted points in the evaluation-activation space. The weights are assigned depending on the confidence value $CV(E_i)$ obtained for each emotion. The final (x, y) coordinates of a given image are calculated as the centre of mass of the seven weighted points in the Whissell space following (3) (see Fig. 5).

$$x = \frac{\sum_{i=1}^{7} x_i \mathrm{CV}(E_i)}{\sum_{i=1}^{7} \mathrm{CV}(E_i)} \quad \text{and} \quad y = \frac{\sum_{i=1}^{7} y_i \mathrm{CV}(E_i)}{\sum_{i=1}^{7} \mathrm{CV}(E_i)} \tag{3}$$

In this way, the output of the system is enriched with a larger number of intermediate emotional states.

3.2 General results

The method described in the previous section has been put into practice with the outputs of the classification system when applied to the database facial expressions images.

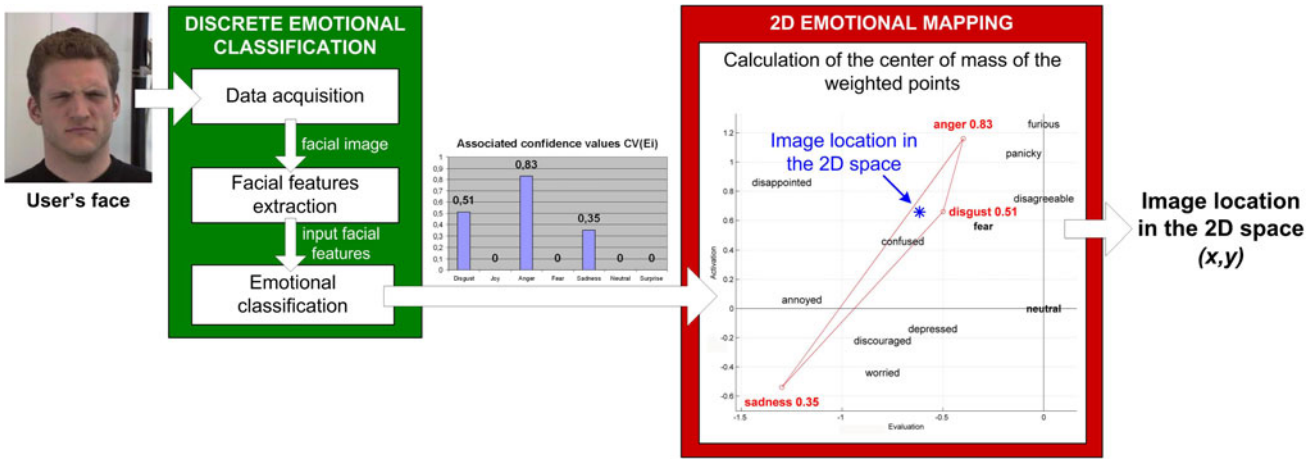

Fig. 5 Overall block diagram for obtaining the location of a facial image in the 2D emotional space. A graphic illustration of the 2D emotional mapping process is included as an example

Figure 6 shows several images with their nearest label obtained in the Whissell space with our system.

In Fig. 7, the location of all the classified images is plotted in the Whissell space (marker size is proportional to the percentage of images situated at the same location). If the locations of the facial expressions in the Whissell space are analyzed, a set of interesting conclusions can be extracted.

1. A great number of images are located along the straight lines that link, respectively, "surprise"/"fear" and "sadness"/"neutral" (Fig. 8). Firstly, this result highlights the capability of the method to find a large amount of intermediate states between those pairs of emotions, which is extremely important since it enriches the output of the system. Secondly, the difficulty is made clear of establishing a threshold that separates certain facial expressions of those pairs of emotions. Ekman [9] already noticed this difficulty for humans when trying to distinguish "surprise" and "fear" one from another, although both are distinguishable from the rest of the emotions. This fact is corroborated by the results obtained by the system, where facial expressions are located along a straight line that shows the gradual passing from one emotion to the other. Something similar happens between "neutral"/"sadness", emotions that have already caused difficulties in a great number of works when attempts have been made to distinguish them, probably due to the fact that some commonalities of facial movements are shared between them.
2. Another interesting point is the isolation of the emotion "joy" in the Whissell space. The system hardly detects intermediate states between "joy" and the rest of the emotions, classifying the emotion in the great majority of cases purely as "joy" (see Fig. 7).

Fig. 6 Example of images from the database with their nearest label in the Whissell space, according to the method described in Sect. 3.1

According to Fredrickson [12], positive emotions are difficult to study since they are comparatively few and barely distinguishable from each other. For instance

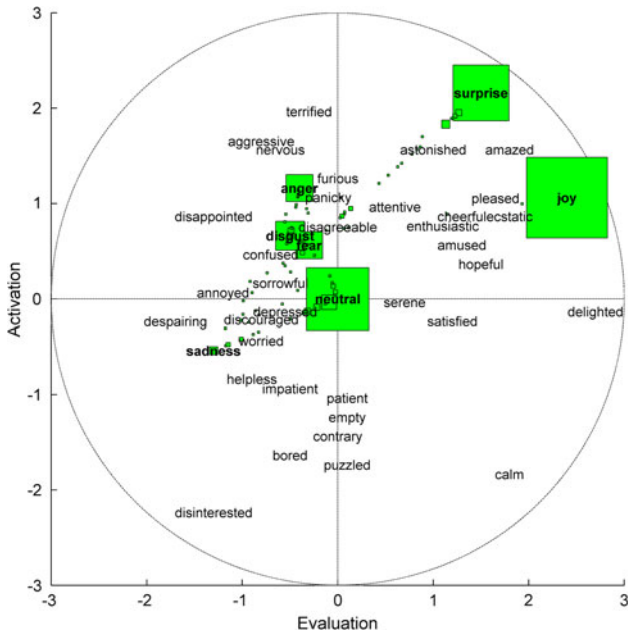

Fig. 7 Location of the different images of the database in the Whissell space, according to the method explained in Sect. 3.1 (markers size is proportional to the percentage of images situated at the same location)

"joy", "cheerful" and "pleased" are not easily distinguishable. In fact, these emotions do not have a special characteristic signal. They all share the "Duchenne smile", where lip corners rise and muscles around the eyes contract, raising the cheeks.

3. In contrast to positive emotions, "anger", "fear", "disgust" and "sadness" are clearly different experiences, although they are very close within the Whissell space. Taxonomically, basic emotions identify three or four negative emotions for each positive one, and this imbalance is also reflected in the number of emotional words in the English language. From a physical point of view, negative emotions have facial configurations that entail universally recognized signals. Faces expressing "sadness", "anger", "disgust" or "fear" can be easily identified. This fact explains the confinement in the negative central zone of a great number of images from the database (Fig. 7), corresponding to intermediate states of negative emotions.

3.3 Evaluation of results taken human assessment into account

The database used in this work provides images labelled with one of the six Ekman universal emotions plus "neutral", but there is no a priori known information about their location in the Whissell 2D space. In order to evaluate the system results, there is a need to establish the region in the Whissell space where each image can be considered to be correctly located. For this purpose, a total of 43 persons participated in one or more evaluation sessions (50 images per session). In the sessions, they were told to locate a set of images of the database in the Whissell space (Fig. 9).

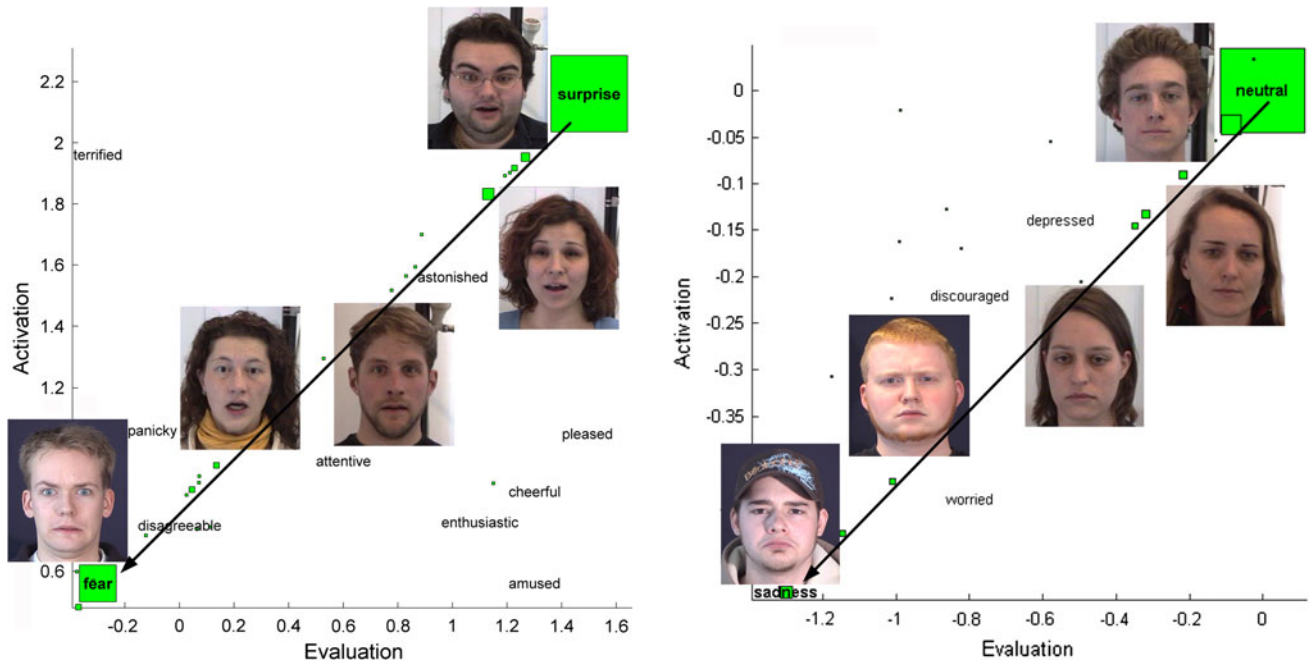

Fig. 8 Zoom on different zones of the Whissell space. Selected images are shown near their location, calculated according to the method explained in Sect. 3.1

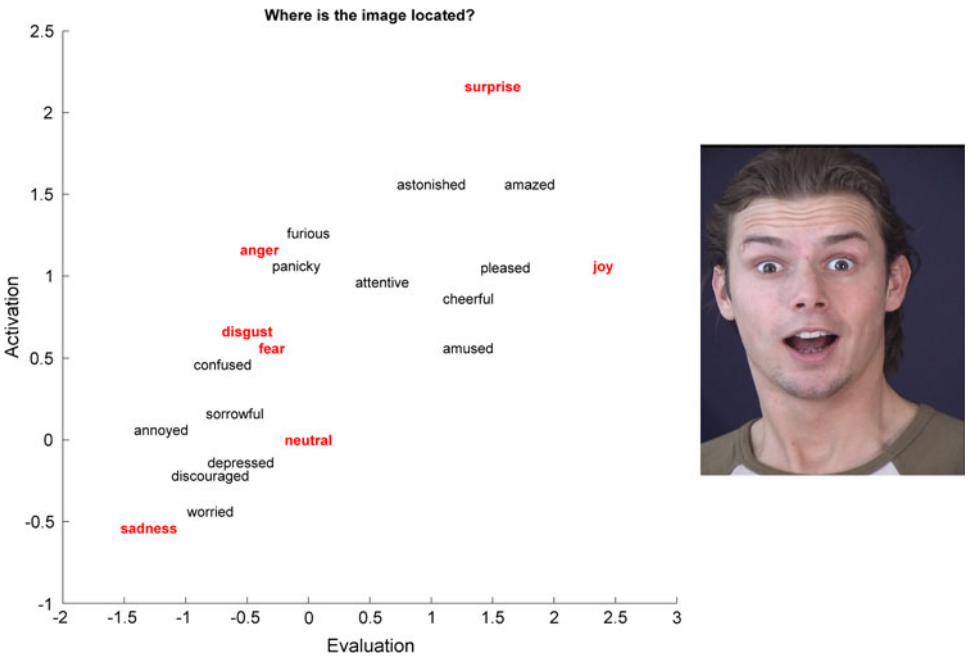

Fig. 9 Example of evaluation session. The user is told to locate the image shown in the Whissell space

As result, each one of the frames was located in terms of evaluation-activation by 16 different persons.

The collected evaluation data have been used to define an ellipsoidal region where each image is considered to be correctly located. The algorithm used to compute the shape of the region is based on Minimum Volume Ellipsoids (MVE). MVE looks for the ellipsoid with the smallest volume that covers a set of data points. Although there are several ways to compute the shape of a set of data points (e.g. using a convex hull, rectangle, etc.), we chose the MVE because of the fact that real-world data often exhibits a mixture of Gaussian distributions, which have equi-density contours in the shape of ellipsoids. First, the collected data are filtered in order to remove outliers: a point is considered an outlier if its coordinate values (in both dimensions) are greater than the mean plus three times the standard deviation. Then, the MVE is calculated following the algorithm described by Kumar and Yildrim [21]. The MVEs obtained are used for evaluating results at four different levels:

1. Ellipse criteria. If the point detected by the system (2D coordinates in the Whissell space) is inside the defined ellipse, it is considered a success; otherwise it is a failure. These criteria are illustrated in Fig. 10.
2. Quadrant criteria. The output is considered to be correctly located if it is in the same quadrant of the Whissell space as the ellipse centre.
3. Evaluation axis criteria. The system output is a success if situated in the same semi-axis (positive or negative) of the evaluation axis as the ellipse centre. This information is especially useful for extracting the positive or negative polarity of the shown facial expression.
4. Activation axis criteria. The same criteria projected to the activation axis. This information is relevant for measuring whether the user is more or less likely to take an action under the emotional state.

The results obtained following the different evaluation strategies are presented in Table 5. As can be seen, the success rate is 73.73 % in the most restrictive case, i.e. when the output of the system is considered to be correctly located when inside the ellipse. It rises to 94.12 % when considering the evaluation axis criteria.

Objectively speaking, these results are very good, especially when it is remembered that a trained observer can correctly classify facial emotions with an average of 87 %. Moreover, certain affective states are so close in the 2D space or share so many facial features that it turns out really difficult even for a human being to clearly distinguish among them (see Fig. 6). In the performed evaluation sessions, when the human evaluators found a given image hard to locate in the 2D space this fact was reflected in a greater size of the calculated MVE. Therefore, the use of ellipsoids allows to analyse if the classification mechanism is coherent with human classification (positioning the image inside the corresponding ellipsoid), which is, in fact, the aim of our work.

However, the obtained results are difficult to compare with other emotional classification studies that can be found in literature, given that most of such studies do not recognize emotions in evaluation-activation terms. Moreover the few that do have not been tested under

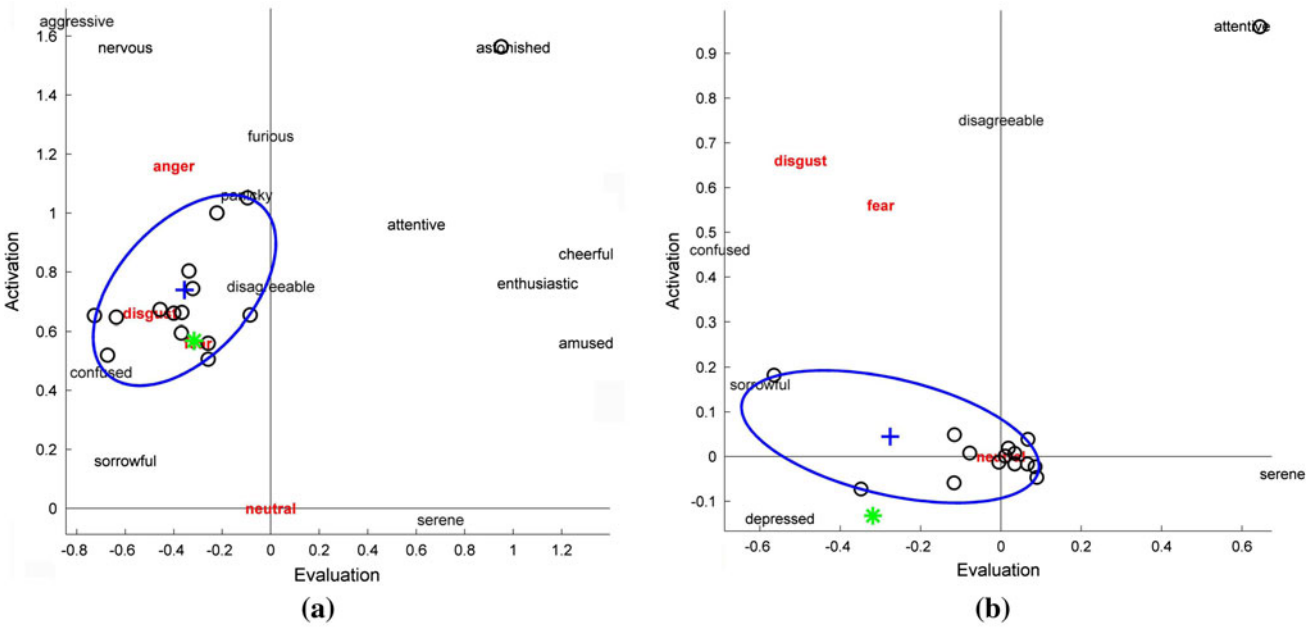

Fig. 10 Example of calculated MVEs. *Black dots* are the data points obtained from users' assessment; *blue cross* is the ellipse center; and *green asterisk* is the 2D location detected by the system. **a** Success case following ellipse criteria. **b** Failure case following ellipse criteria

Table 5 Results obtained according to different evaluation criteria

	Ellipse criteria (%)	Quadrant criteria (%)	Evaluation axis criteria (%)	Activation axis criteria (%)
Success rate	73.73	87.45	94.12	92.94

common experimental conditions (e.g. different databases, ground-truth data or evaluation strategies are used) and do not provide as output the coordinates of the studied facial image in the 2D space. The few existing works that manage dimensional concepts reduce the descriptive potential of the 2D space to a discrete approach where the output classification labels correspond to a set of zones of the evaluation-activation space. As a reference, the work of Fragopanados and Taylor [11] achieves success rates of 88 and 64 % when classifying facial expressions in terms of evaluation and activation polarity (positive vs. negative and active vs. passive), respectively; the study of Shin [31] obtains recognition results of 90.9 % in evaluation dimension and 56.1 % in activation dimension; and Caridakis et al. [4] achieve a success ratio of 67 % in four class (quadrants of 2D space) classification. The presented system's success rates largely surpass those performances, even though our system's output is much more rich and our validation strategy much more strict.

4 Conclusions and future work

This paper describes an effective method for facial emotional classification. The inputs to the system are a set of facial parameters (angles and distances between characteristic points of the face) chosen by means of a correlation-based feature selection technique so that the face is modelled in a simple way without losing relevant facial expression information. In order to make the emotional classification mechanism more robust, the system intelligently combines in a novel manner the five most commonly used classifiers in the literature, obtaining at the output a confidence value of the facial expression to each of the six Ekman's universal emotions (plus "neutral"). This information is then emotionally mapped on to Whissell's 2D evaluation-activation space with the aim of obtaining the location (coordinates) of the input facial expression in the space. The final output of the system does not, therefore, simply provide a classification in terms of a set of emotionally discrete labels, but goes further by extending the emotional information over an infinite range of intermediate emotions. This work is able to recognize the location of the user facial expression in a 2D evaluation-activation space, and not only its polarity (positive/negative or active/passive) as in existing studies working with a dimensional description level of affect. This kind of output is especially useful in HCI scenarios, such as human-robot or human-virtual agent interaction contexts, since it emulates the way humans detect

emotions from their interlocutor in real human-human communication contexts. Another noteworthy feature of the work is that it has been tested with an extensive database of 1500 images showing individuals of different races and gender, giving universal results with very promising levels of correctness. Human assessment has been taken into account in the evaluation of the system, that has been proved to work in a similar way to the human brain, leading to similar confusions.

The recent focus on the research area of Affective Computing lies on sensing emotions from multiple modalities, since natural human-human interaction is multimodal: people communicate through speech and use body language (posture, facial expressions, gaze) to express emotion, mood, attitude, and attention. A main question related to multimodality that still remains unsolved is how to fuse the information coming from different channels (audio, video, etc.). The multimodal fusion problem reinforces the limitations of categorical descriptions of affect: discrete emotional labels have no real link between them and, at the fusion stage, every studied emotion must be recognized independently. Therefore, all available multimodal recognizers have designed and/or used ad-hoc solutions for fusing information coming from multiple modalities but cannot accept new modalities without re-defining and re-training the whole system. The use of a continuous emotional space in the way described in this paper opens the door to the fusion of different modules coming from different channels in a simple and scalable fashion. The dimensional approach provides an algebra and allows the different emotional inputs coming from different modalities, with different levels of description of affect, to be related mathematically. In fact, authors are currently considering the integration of new multimodal emotional recognition input modules to the system (user's speech, gestures, gaze, mouse-clicks, keyboard use) making use of the Whissell space.

Another future work line is to pass from static facial expression recognition to the analysis of the dynamic evolution in time of user's facial expressions in video sequences. Every time with more force, the psychological investigation argues that the timing of the facial expressions is a critical factor in the recognition of emotions since humans inherently display facial emotions following a continuous temporal pattern. With this postulate and thanks to the use of the 2-dimensional description of affect, an emotional facial video sequence could be viewed as a point (corresponding to the location of a particular affective state in time t) moving through this space over time. In this way, by introducing dynamic information in the system, emotion recognition could be improved and become more useful in real-time interaction contexts.

Acknowledgments This work has been partly financed by the Spanish Government through the DGICYT contract TIN2011-24660, by the project FEDER ATIC, and the SISTRONIC Group of the Aragon Institute of Technology (Ref. T84).

References

1. Anderson K, McOwan P (2006) A real-time automated system for the recognition of human facial expressions. IEEE Trans Syst Man Cybern Part B 36:96–105
2. Bassili J (1979) Emotion recognition: the role of facial movement and the relative importance of upper and lower areas of the face. J Pers Soc Psychol 37:2049–2058
3. Boukricha H, Becker C, Wachsmuth I (2007) Simulating empathy for the virtual human max. In: Proceedings 2nd international workshop on emotion and computing in conjunction with the German conference on artificial intelligence (KI2007), pp 22–27
4. Caridakis G, Malatesta L, Kessous L, Amir N, Paouzaiou A, Karpouzis K (2006) Modeling naturalistic affective states via facial and vocal expression recognition. In: International conference on multimodal interfaces, pp 146–154
5. Chang C, Tsai J, Wang C, Chung P (2009) Emotion recognition with consideration of facial expression and physiological signals. In: Proceedings of the 6th annual IEEE conference on computational intelligence in bioinformatics and computational biology, pp 278–283
6. Cohen I, Sebe N, Garg A, Chen L, Huang T (2003) Facial expression recognition from video sequences: temporal and static modeling. Comput Vis Image Underst 11(1–2):160–187
7. Datcu D, Rothkrantz L (2007) Facial expression recognition in still pictures and videos using active appearance models: a comparison approach. In: Proceedings of the international conference on computer systems and technologies, pp 1–6
8. Du Y, Bi W, Wang T, Zhang Y, Ai H (2007) Distributing expressional faces in 2-d emotional space. In: Proceedings of the 6th ACM international conference on image and video retrieval, pp 395–400
9. Ekman P, Dalgleish T, Power M (1999) Handbook of cognition and emotion. Wiley, Chihester
10. Ekman P, Friesen W, Hager J (2002) Facial action coding system. Research Nexus eBook
11. Fragopanagos N, Taylor J (2005) Emotion recognition in human–computer interaction. Neural Netw 18:389–405
12. Fredrickson B (2003) The value of positive emotions. Am Sci 91:330–335
13. Gosselin F, Schyns P (2001) Bubbles: A technique to reveal the use of information in recognition tasks. Vis Res 41:2261–2271
14. Gunes H, Schuller B, Pantic M, Cowie R (2011) Emotion representation, analysis and synthesis in continuous space: a survey. In: 2011 IEEE international conference on automatic face gesture recognition and workshops (FG 2011), pp 827–834
15. Hall M (1998) Correlation-based feature selection for machine learning. PhD thesis, Hamilton, New Zealand
16. Hammal Z, Couvreur L, Caplier A, Rombaut M (2007) Facial expression classification: An approach based on the fusion of facial deformations using the transferable belief model. Int J Approx Reason 46:542–567
17. Ji Q, Lan P, Looney C (2006) A probabilistic framework for modeling and real-time monitoring human fatigue. IEEE Trans Syst Man Cybern Part A 36:862–875
18. Kapoor A, Burleson W, Picard R (2007) Automatic prediction of frustration. Int J Hum Comput Stud 65:724–736
19. Keltner D, Ekman P (2000) Facial expression of emotion. Handb Emot 2:236–249

20. Kotsia I, Pitas I (2007) Facial expression recognition in image sequences using geometric deformation features and support vector machines. IEEE Trans Image Process 16: 172–187
21. Kumar P, Yildirim E (2005) Minimum-volume enclosing ellipsoids and core sets. J Optim Theory Appl 126:1–21
22. Littlewort G, Bartlett M, Fasel I, Chenu J, Movellan J (2004) Analysis of machine learning methods for real-time recognition of facial expressions from video. In: Computer vision and pattern recognition, face processing workshop
23. Littlewort G, Bartlett M, Fasel I, Susskind J, Movellan J (2006) Dynamics of facial expression extracted automatically from video. Image Vis Comput 24:615–625
24. Liu W, Lu J, Wang Z, Song H (2008) An expression space model for facial expression analysis. In: Congress on image and signal processing (CISP 08), vol 2, pp 680–684
25. Lucey S, Ashraf A, Cohn J (2007) Investigating spontaneous facial action recognition through aam representations of the face. In: Handbook on Face Recognition, pp 275–286
26. Maalej A, Amor BB, Daoudi M, Srivastava A, Berretti S (2011) Shape analysis of local facial patches for 3d facial expression recognition. Pattern Recogn 44(8):1581–1589
27. Pantic M, Bartlett M (2007) Machine analysis of facial expressions. Face recognition pp 377–416
28. Pantic M, Valstar M, Rademaker R, Maat L (2005) Web-based database for facial expression analysis. In: IEEE international conference on multimedia and expo, pp 317–321
29. Plutchik R (1980) Emotion: a psychoevolutionary synthesis. Harper & Row, New York
30. SeeingMachines (2008) Face api technical specifications brochure. http://www.seeingmachines.com/pdfs/brochures/faceAPI-Brochure.pdf. Accessed 1 Mar 2010
31. Shin Y (2007) Facial expression recognition based on emotion dimensions on manifold learning. In: International conference on computational science. Springer, Berlin, pp 81–88
32. Soyel H, Demirel H (2007) Facial expression recognition using 3d facial feature distances. Lect Notes Comput Sci 4633: 831–838
33. Stoiber N, Seguier R, Breton G (2009) Automatic design of a control interface for a synthetic face. In: Proceedings of the 13th international conference on intelligent user interfaces, pp 207–216
34. Tang H, Huang T (2008) 3D facial expression recognition based on automatically selected features. In: IEEE international conference on computer vision and pattern recognition, pp 1–8
35. Wallhoff F (2006) Facial expressions and emotion database. http://www.mmk.ei.tum.de/waf/fgnet/feedtum.html, Technische Universität München
36. Whissell C (1989) The dictionary of affect in language. In: Emotion: theory, research and experience. The measurement of emotions, vol 4. Academic, New York
37. Whissell C (2000) Whissell's dictionary of affect in language technical manual and user's guide. http://www.hdcus.com/manuals/wdalman.pdf. Accessed 18 Dec 2010
38. Whitehill J, Bartlett M, Movellan J (2008) Automated teacher feedback using facial expression recognition. In: Workshop on CVPR for human communicative behavior analysis, IEEE conference on computer vision and pattern recognition
39. Witten I, Frank E (2005) Data Mining: practical machine learning tools and techniques, 2nd edn. Morgan Kaufmann, San Francisco
40. Wu Y, Liu H, Zha H (2005) Modeling facial expression space for recognition. In: 2005 IEEE/RSJ International conference on intelligent robots and systems (IROS 2005), pp 1968–1973
41. Yeasin M, Bullot B, Sharma R (2006) Recognition of facial expressions and measurement of levels of interest from video. IEEE Trans Multimedia 8:500–508
42. Zeng Z, Pantic M, Roisman G, Huang T (2009) A survey of affect recognition methods: audio, visual, and spontaneous expressions. IEEE Trans Pattern Anal Mach Intell 39–58
43. Zhang Y, Ji Q (2005) Active and dynamic information fusion for facial expression understanding from image sequences. IEEE Trans Pattern Anal Mach Intell 699–714
44. 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 the 3rd international conference on face and gesture recognition, pp 454–459