Emotional facial expression classification for multimodal user interfaces

Abstract. We present a simple and computationally feasible method to perform automatic emotional classification of facial expressions. We propose the use of 10 characteristic points (that are part of the MPEG4 feature points) to extract relevant emotional information (basically five distances, presence of wrinkles and mouth shape). The method defines and detects the six basic emotions (plus the neutral one) in terms of this information and has been fine- tuned with a data-base of 399 images. For the moment, the method is applied to static images. Application to sequences is being now developed. The extraction of such information about the user is of great interest for the development of new multimodal user interfaces.

1 Introduction

Facial expression is the most powerful, natural and direct way between humans to communicate emotions, valuations and intentions. As pointed out by Bruce [1], human face-to-face communication is an ideal model for designing a multimodal human-computer interface (HCI).

A system capable of extracting emotional information from user's facial expressions would be of great interest for developing new interfaces which follow the human face-to-face communication model in the most realistic way. In particular, the creation of virtual environments populated by 3D virtual characters capable of understanding users' expressions and reacting accordingly represents, nowadays, a challenging but affordable task.

Nevertheless, to develop a system that interprets facial expressions is difficult. Three kinds of problems have to be solved: face detection in a facial image or image sequence, facial expression data extraction and facial expression classification (e.g. into emotional categories). In this paper we are going to deal with the third problem: classification. This implies the definition of the set of categories we want to deal with, and the implementation of the categorization mechanisms.

The structure of the paper is as follows: in Section 2 works and problems in facial classification are presented. Section 3 explains our method. Results are presented in Section 4, whereas conclusions and comments about future work are discussed in Section 5.

2 The Problems of Facial Expressions Classification

Facial expression analyzers make use of three different methods of classification: patterns, neuronal networks or rules. If a pattern-based method is used [2,3,4], the face expression found is compared with the patterns defined for each expression category. The best matching decides the classification of the expression. Most of these methods first apply PCA and LDA algorithms to reduce dimensionality. In the systems based on neuronal networks [5,6], the face expression is classified according to a categorization process "learned" by the neuronal network during the training phase. In general, the entrance to this type of systems is a set of characteristics extracted from the face (points or distances between points). The rule-based methods [7] classify the face expression into basic categories of emotions, according to a set of face actions previously codified. In [8] an excellent state-of-the-art on the subject can be found.

In any case, the development of automatic facial classification systems presents several problems. Most of the studies on automated expression analysis perform an emotional classification. The emotional classification of Ekman [9] is the most followed one. It describes six universal basic emotions: joy, sadness, surprise, fear, disgust and anger. Nevertheless, the use of Ekman's categories for developing automating facial expression emotional classification is difficult. First, his description of the six prototypic facial expressions of emotions is linguistic and, thus, ambiguous. There is no uniquely defined description either in terms of facial actions or in terms of some other universally defined facial codes. Second, classification of facial expressions into multiple emotion categories should be possible (e.g. raised eyebrows and smiling mouth is a blend of surprise and happiness) but, still, there is no psychological scrutiny on this topic. Another important issue to be considered is individualization. The system should be capable of analyzing any subject, male or female of any age and ethnicity and of any expressivity.

3 A Simple Method for the Automatic Analysis of Face Expressions

Our method is based on the work of Hammal et al [10]. They have implemented a facial classification method for static images. The originality of their work consists, on the one hand, in the supposition that all the necessary information for the recognition of expressions is contained in the deformation of certain characteristics of the eyes, mouth and eyebrows and, on the other hand, in the use of the Belief Theory to make the classification. Nevertheless, their method has important restrictions. The greater restriction comes from the fact that it is only able to discern 3 of the 6 basic emotions (without including the neutral one). This is basically due to the little information they handle (only 5 distances). It would not be viable, from a probabilistic point of view, to work with many more data, because the explosion of possible combinations would remarkably increase the computational cost of the algorithm.

3.1 General Description of the Method

Our method studies the variation of a certain number of face parameters (distances and angles between some feature points of the face) with respect to the neutral expression. The objective of our method is to adjudge a score to each emotion, according to the state acquired by each one of the parameters in the image. The emotion (or emotions in case of draw) chosen will be the one that obtains a greater score.

For example, let's imagine that we study two face parameters (P_1 and P_2) and that each one of them can take three different states (C^+, C^- and S, following the nomenclature of Hammal). State C^+ means that the value of the parameters has increased with respect to the neutral one; state C^- that its value has diminished with respect to the neutral one; and the state S that its value has not varied with respect to the neutral one. First, we build a descriptive table of emotions, according to the state of the parameters, like the one of the Table 1. From this table, a set of logical tables can be build for each parameter (Table 2). That way, two vectors of emotions are defined, according to the state taken by each one of the parameters (C^+, C^- or S) in a specific frame. Once the tables are defined, the implementation of the identification algorithm is simple. When a parameter takes a specific state, it is enough to select the vector of emotions (formed by 1's and 0's) corresponding to this state. If we repeat the procedure for each parameter, we will obtain a matrix of so many rows as parameters we study and 7 columns, corresponding to the 7 emotions. The sum of 1's present in each column of the matrix gives the score obtained by each emotion.

	P1	**P2**
Joy	C-	S/C-
Surprise	C+	C+
Disgust	C-	C-
Anger	C+	C-
Sadness	C-	C+
Fear	S/C+	S/C+
Neutral	S	S

Table 1. Theoretical table of parameters' states for each emotion.

Compared to the method of Hammal, ours is computationally simple. The combinatory explosion and the number of calculations to make are reduced considerably, allows us to work with more information (more parameters) of the face and to evaluate the seven universal emotions, and not only 4 of them, as Hammal does.

		E1 joy	E2 surprise	E3 disgust	E4 anger	E5 sadness	E6 fear	E7 neutral
C1	C+	**0**	**1**	**0**	**1**	**0**	**1**	**0**
	C-	1	0	1	0	1	0	0
	N	0	0	0	0	0	1	1
		E1 joy	**E2 surprise**	**E3 disgust**	**E4 anger**	**E5 sadness**	**E6 fear**	**E7 neutral**
C2	C+	0	1	0	0	1	1	0
	C-	**1**	**0**	**1**	**1**	**0**	**0**	**0**
	N	1	0	0	0	0	1	1

Table 2. Logical rules table for each parameter.

3.2 Feature Selection

The first step of our method consists of extracting the 10 feature points of the face that will later allow us to analyze the evolution of the face parameters (distances and angles) that we wish to study. Figure 1 shows the correspondence of these points with the ones defined by the MPEG-4 standard. For the moment, the extraction of the points is made manually, by means of a landmarking program made in Matlab. We are now developing an automatic features extraction, which will allow as well to analyze a greater number of images and to even study the evolution of the parameters in video sequences, and not only in static images.

The characteristics points are used to calculate five distances shown in Figure 2. These five distances can be translated in terms of MPEG-4 standard, putting them in relation to the feature points shown in Figure 1 and with some FAPs defined by the norm. All the distances are normalized with respect to the distance between the eyes (MPEG FAPU "ESo"), which is a distance independent of the expression. This way, the values will be consistent, independently of the scale of the image, the distance to the camera, etc.

3.3 Database

In order to define the emotions in terms of the parameters states, as well as to find the thresholds that determine if parameter is in a state or another, it is necessary to work with a wide database. In this work we have used the facial expressions and emotions database FG-NET of the University of Munich [11] that provides video sequences of 19 different people showing the 7 universal emotions from Ekman (Fig.3).

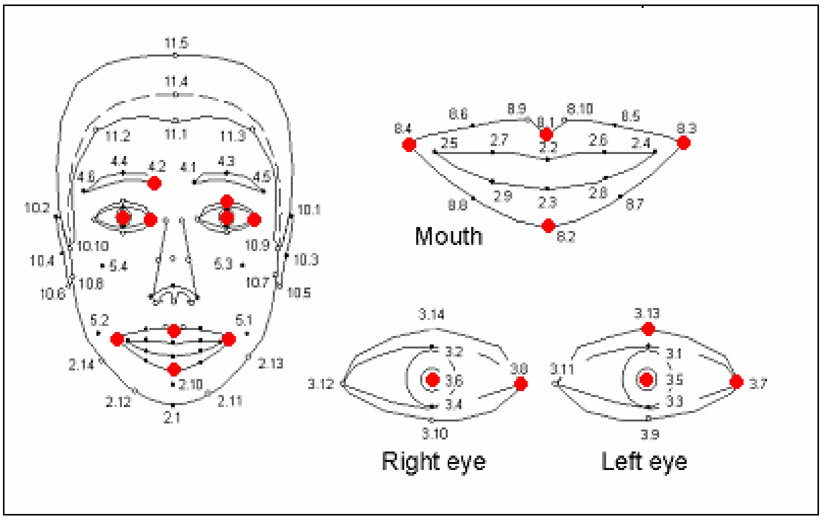
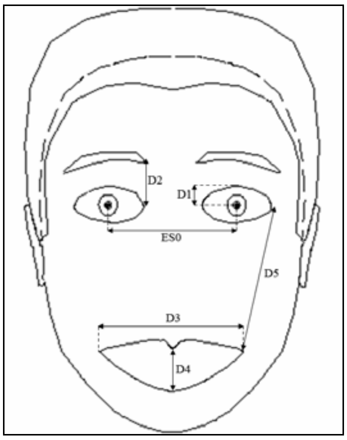


Fig. 1. Facial feature points used for the later definition of the parameters to analyze, according to MPEG-4 standard.



MPEG-4 FAPs NAME	FEATURE POINTS USED FOR DISTANCES
close_upper_l_eyelid close_lower_l_eyelid	D1=d(3.5, 3.1)
raise_r_i_eyebrow	D2=d(4.2, 3.8)
stretch_l_cornerlip stretch_r_cornerlip	D3=d(8.4, 8.3)
open_jaw	D4=d(8.1 , 8.2)
raise_r_cornerlip	D5=d(8.3, 3.7)

Fig. 2. Characteristic distances used in our method (left). On the right, relationship between the five characteristic distances of and the MPEG-4 FAPs and feature points.

Fig. 3. Example of selected frames of the FG-NET database [11].

4. Results

4.1 Initial results

First we considered to work with the same parameters as the Hammal method, ie, with the 5 characteristic distances shown in Figure 2. In order to build a descriptive table of each emotion in terms of states of distances, we must determine the value of the states of distances that define each emotion (C^+, C^- or S), as well as evaluate the thresholds that separate a state from another, for each distance. To do this, we studied the variation of each distance with respect to the neutral one, for each person of the database and for each emotion. An example of the results obtained for distance D_4 is shown in Figure 4. From these data, we can make a descriptive table of the emotions according to the value of the states (Table 3).

The last step to complete our algorithm is to define the values of the thresholds that separate a state of another one, for each studied distance. Two types of thresholds exist: the upper threshold (marks the limit between neutral state S and state C^+) and the lower threshold (the one that marks the limit between neutral state S and state C^-). The thresholds' values are determined by means of several tests and statistics. Figure 4 shows an example of thresholds estimation for the distance D_4.

Once the states that characterize each emotion and the value of the thresholds are established, the algorithm has been proved on the 399 images of the database. In the evaluation of results, the recognition is marked as "good" if the decision is coherent with the one taken by a human being. To do this, we have made surveys to several people to classify the expressions shown in the most ambiguous images.

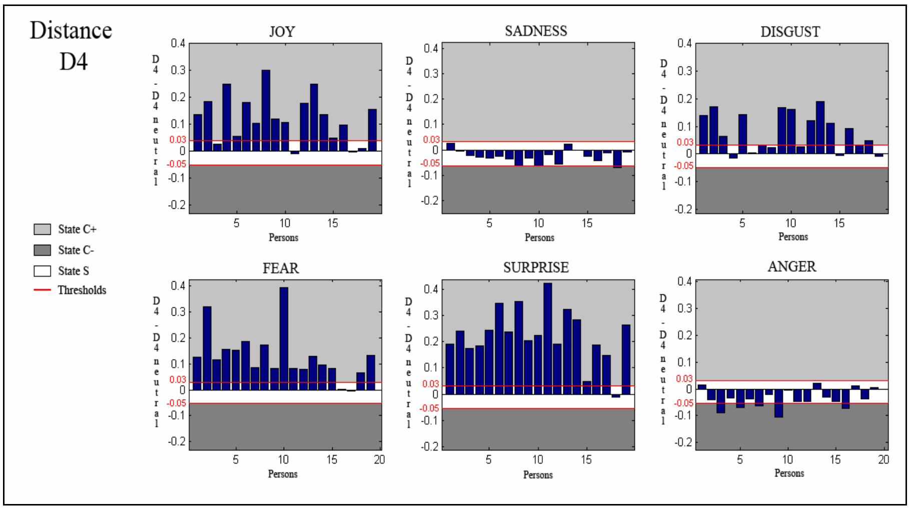
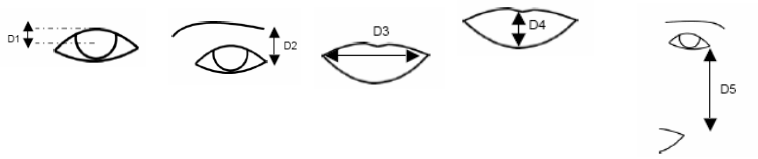


Fig. 4. Statistics results obtained for distance D_4. Thresholds estimations are also shown.



Joy	C-	S/C-	C+	C+	C-
Surprise	S/C+	S/C+	S/C-	C+	S/C+
Disgust	C-	C-	S/C+/C-	S/C+	S/C-
Anger	C-	C-	S/C-	S/C-	S/C+/C-
Sadness	C-	S	S/C-	S	S/C+
Fear	S/C+	S/C+/C-	C-	C+	S/C+
Neutral	S	S	S	S	S

Table 3. Theoretical table of the states taken by D_i for each emotion, according to the results of the statistics obtained from the FG-NET database. Some distances do not provide any information of interest for emotions (squares in gray)

For example, in the image shown in Figure 5, the surveyed people recognized as much "disgust" as "anger", although the FG-NET database classifies it like "disgust" exclusively. Our method obtains a draw.

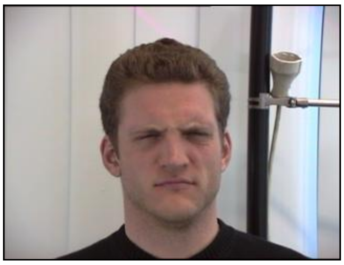


Fig. 5. Frame classified like “disgust” by the FG-NET database[15].

The obtained results are shown in the third column Table 4. As it can be observed, the percentage of success obtained for the emotions “disgust”, “anger”, “sadness”, “fear” and “neutral” are acceptable and similar to the obtained by Hammal (second column). Nevertheless, for “joy” and “surprise” the results are not very favorable. In fact, the algorithm tends to confuse “joy” with “disgust” and “surprise” with “fear”, which comes justified looking at Table 3, where it can be seen that a same combination of states of distances can be given for the mentioned pairs of emotions. Related to classification success, it is interesting to realize that human mechanisms for face detection are very robust, but this is not the case of those for face expressions interpretation. According to Bassili [12], a trained observer can correctly classify faces showing emotions with an average of 87%.

4.2 Addition of characteristics: information about the wrinkles in the nasal root

In order to improve the results obtained in “joy”, we introduce a new face parameter: the presence or absence of wrinkles in the nasal root, typical of the emotions “disgust” and “anger”. This way, we will mark a difference between “joy” and “disgust”. The obtained success rates are shown in the forth column in Table 4. We observe, as it was expected, a considerable increase in the rate of successes, especially for “joy” and “disgust”. However, the rates still continue being low for “sadness” and “surprise”, which makes us think about the necessity to add more characteristics to the method.

4.3 Addition of characteristics: information about the mouth shape

A key factor to analyze in the recognition of emotions is the mouth shape. For each one of the 7 basic emotions, its contour changes in many different ways. In our method, we have extracted 4 feature points of the mouth that are shown in Figure 6. Results are shown in the fifth column in Table 4. As it can be seen, the new information has introduced a great improvement in our results. The importance of the mouth shape in the expression of emotions is thus confirmed.

EMOTION	% SUCCESS HAMMAL METHOD	% SUCCESS OUR METHOD	% SUCCES WRINKLES NASAL ROOT	% SUCCES MOUTH SHAPE
Joy	87,26	36,84	100	100
Surprise	84,44	57,89	63,16	63,16
Disgust	51,20	84,21	94,74	100
Anger	not recognized	73,68	94,74	89,47
Sadness	not recognized	68,42	57,89	94,74
Fear	not recognized	78,95	84,21	89,47
Neutral	88%	100	100	100

Table 4. Classification rates of Hammal [13] (second column), of our method with the 5 distances (third column), plus wrinkles in the nasal root (fourth column) plus mouth shape information (fifth column).

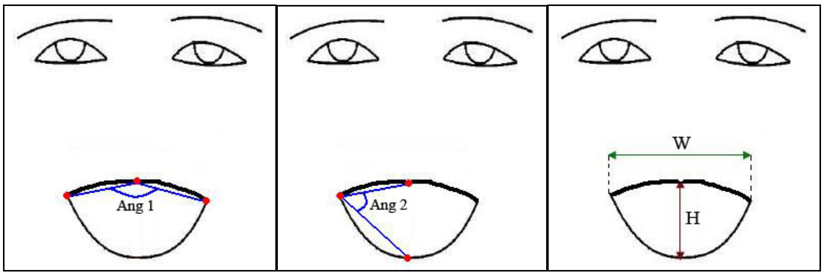


Fig. 6. Extra information added about the mouth shape.

5. Conclusions and future work

We have presented a simple and effective method for the automatic classification of facial expressions. The introduction of several additional parameters barely increases the computational cost of the algorithm, given its simplicity, and produces very significant rates of improvement. In a future it is hoped to introduce new characteristics, in the form of face distances or angles (for example the angle formed by the eyebrows). Another noticeable objective in the short term is to make the tracking of the landmarks in an automatic way. Thanks to it, we will be able to introduce dynamic information in our method, that is to say, to study the evolution in the time of the evaluated parameters. Every time with more force, the psychological investigation argues that the timing of the facial expressions is a critical factor in the interpretation of expressions. In the midterm, the objective is to add the system to the ambient intelligent applications that the group is developing, to enrich user interaction.

Acknowledments

This work has been partially financed by the Spanish "Dirección General de Investigación", contract number Nº TIN2004-07926 and by the Aragon Government through the WALQA agreement (ref. 2004/04/86).

References

1. V. Bruce: What the Human Face Tells the Human Mind: some Challenges for the Robot-Human Interface. Proc. Int'l workshop Robot and Human Comm.A. (1992) 44-51.
2. G.J. Edwards, T.F. Cootes y C.J. Taylor, "Face Recognition Using Active Appearance Models", Proc. European Conf. Computer Vision, Vol. 2, pp. 581-695, 1998.
   H. Hong, H. Neven y C. von der Malsburg, "Online Facial Expression Recognition Based on Personalized Galleries", Proc. Int'l Conf. Automatic Face and Gesture Recognition, pp. 354-359, 1998.
3. H. Hong, H. Neven y C. von der Malsburg, "Online Facial Expression Recognition Based on Personalized Galleries", Proc. Int'l Conf. Automatic Face and Gesture Recognition, pp. 354-359, 1998.
4. M.J. Lyons, J. Budynek y S. Akamatsu, "Automatic Classification of Single Facial Images", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 21, nº12, pp. 1357-1362, 1999.
5. Z. Zhang, M. Lyons, M. Schuster y S. Akamatsu, "Comparison between Geometry-Based and Gabor Wavelets-Based Facial Expression Recognition Using Multi-Layer Perceptron", Proc. Int'l Conf. Automatic Face and Gesture Recognition, pp. 454-459, 1998.
6. M. Wallace, A. Raouzaiou, N. Tsapatsoulis y S. Kollias, "Facial Expression Classification Based on MPEG-4 FAPs: The Use of Evidence and Prior Knowledge for Uncetainty Removal", IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)**,** Vol. 1, pp. 51-54, Budapest (Hungría), 25-29 de Julio de 2004.
7. M. Pantic y L.J.M. Rothkrantz, "Expert System for Automatic Analysis of Facial Expression", Image and Vision Computing J., Vol. 18, nº. 11, pp. 881-905, 2000.
8. Pantic, M. and Rothkrantz, L.J.M.: Automatic Analysis of Facial Expressions: The State of the Art. Pattern Analysis and Machine Intelligence, IEEE Transactions on Volume 22, Issue 12 (December 2000) 1424–1445.
9. P. Ekman: Facial Expression, The Handbook of Cognition and Emotion. John Wiley et Sons (1999).
10. Hammal Z., Couvreur L., Caplier A. and Rombaut M.: Facial Expressions Recognition Based on The Belief Theory: Comparison with Diferent Classifiers. Proc. 13th International Conference on Image Analysis and Processing (Italy 2005).
11. http://www.mmk.ei.tum.de/~waf/fgnet/feedtum.html (Reviewed in February 2006).
12. Bassili, J.N.: Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. Journal of Personality and Social Psychology nº 37 (1997) 2049-2059.