Effective Emotional Classification Combining Facial Classifiers and User Assessment

Isabelle Hupont[1], Sandra Baldassarri[2], Rafael Del Hoyo[1], and Eva Cerezo[2]

[1] Instituto Tecnológico de Aragón, Zaragoza (Spain)
{ihupont,rdelhoyo}@ita.es
[2] Departamento de Informática e Ingeniería de Sistemas,
Instituto de Investigación en Ingeniería de Aragón, Universidad de Zaragoza (Spain)
{sandra,ecerezo}@unizar.es

Abstract. An effective method for the automatic classification of facial expressions into emotional categories is presented. The system is able to classify the user facial expression in terms of the six Ekman's universal emotions (plus the neutral one), giving a membership confidence value to each emotional category. The method is capable of analysing any subject, male or female of any age and ethnicity. The classification strategy is based on a combination (weighted majority voting) of the five most used classifiers. Another significant difference with other works is that human assessment is taken into account in the evaluation of the results. The information obtained from the users classification makes it possible to verify the validity of our results and to increase the performance of our method.

Keywords: Facial Expressions, Emotional Classifiers, Multimodal Interfaces, Affective Computing

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. Recent researches have focused on the development of virtual environments populated by 3D virtual characters capable of understanding the emotional state of the user and reacting accordingly. Addressing user's emotions in human-computer interaction significantly enhances the believability and lifelikeness of virtual humans [1].

The most expressive way humans display emotions is through facial expressions. Thus, the interpretation of facial expressions is the most followed method for achieving user's emotions detection. The process implies the extraction of facial expression information from the user's image or image sequence and the classification of that information into emotional categories. This paper focuses in the problem of classification, which involves the definition of the set of categories we want to deal with and the implementation of the categorization mechanisms.

F.J. Perales and R.B. Fisher (Eds.): AMDO 2008, LNCS 5098, pp. 431–440, 2008.

In the literature, facial expression analysers consider a set of characteristics extracted from the face (points or distances/angles between points) and use different methods of classification [2] in order to determine emotional categories. Most of the systems that work in real-time use a feature based representation of facial data for classification. The real-time methods that obtain better results are neural networks, rule-based expert systems, support vector machines (SVM) and Bayesian nets.

In the systems based on neuronal networks [3,4], the face expression is classified according to a categorization process “learned” by the neuronal network during the training phase.

Rule-based expert systems are also known as knowledge-based systems since they establish a set of rules based on the knowledge of an expert or on objective statistical observations. These systems have the advantage of being easily understood and implemented. They have been used in several works [5,6,7].

Support Vector Machines (SVM), instead, are a set of supervised learning methods used for classification that simultaneously minimize the empirical classification error and maximize the geometric margin; hence, they are also known as maximum margin classifiers. It has been only recently that this technique has been used for the classification of facial expressions [8,9].

Finally, Bayesian networks are directed acyclic graphs whose nodes represent variables, and whose arcs encode conditional dependences between the variables. Most of the algorithms that employ Bayesian networks use Hidden Markov Models [10,11].

There are other approaches for facial expression classification as the ones that use holistic face models; they often rely on Gabor filters [12] or Linear Discriminant Analysis [13], but they are very time consuming and they are not suited for real-time classification.

Regarding emotional categories, the classification proposed by Ekman [14] is the most followed one. It describes six universal basic emotions: joy, sadness, surprise, fear, disgust and anger. However, most existing approaches only consider a subset of these emotions and don’t allow to represent a blend of them.

In this paper we present an effective method for the automatic classification of facial expressions into emotional categories based on a novel combination of existing classifiers. The system is able to classify the user emotion in terms of the six Ekman’s universal emotions (plus the neutral one), giving a membership confidence value to each emotional category. The method is capable of analysing any subject, male or female of any age and ethnicity and takes into account emotional classification performed by humans to analyse the results.

The structure of the paper is the following: Section 2 discusses the selection of the facial features that are the inputs of the facial classifiers. In Section 3 the five different classifiers used are presented and the way these classifiers are combined is explained. Section 4 presents success rates and shows how users’ assessment is used to analyse and improve results. Section 5 compares our results with the ones obtained with other methods. Finally, Section 6 is devoted to present the conclusions and future work.

2 Setting the Classifiers' Input: Extraction and Selection of Facial Information

It has been stated 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 [14]. Making use of this property, and taken also other relevant works into account [5,6,15], we established the initial inputs of our classifiers to a set of distances and angles obtained from twenty characteristic facial points. In fact, the inputs will be the variations of these angles and distances with respect to the neutral face. The points are obtained thanks to a real-time facial feature tracking program presented elsewhere [16]. Figure 1 shows the correspondence of these points with those defined by the MPEG4 standard. The initial set of parameters obtained from these points is shown in Figure 2. In order to make the distances' values consistent (independently of the scale of the image, the distance to the camera, etc.), all the distances are normalized with respect to the distance between the eyes (MPEG4 Facial Animation Parameter Unit -FAPU- called "ESo"), which is a distance independent of the expression. The choice of angles for classification provides a size invariant classification and saves the effort of normalization.

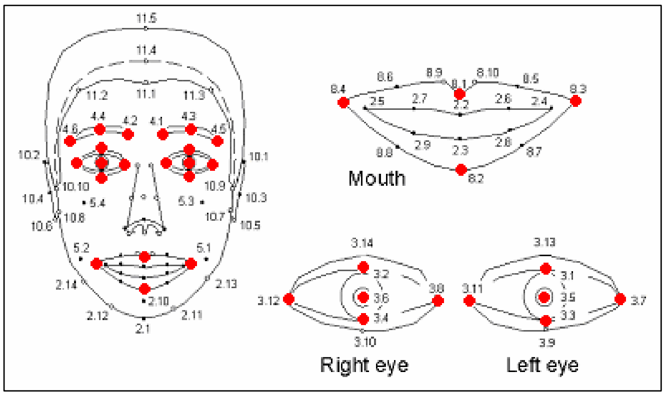
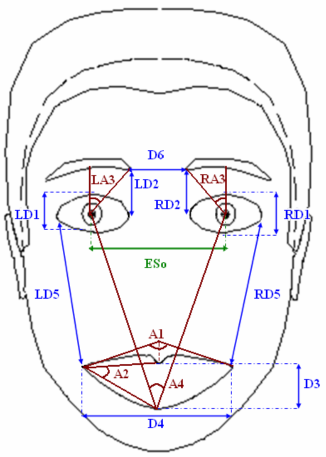


Fig. 1. Facial feature points according to MPEG4 standard

In other to determine the goodness and usefulness of the parameters, a study of the correlation between parameters was carried out using the data (distances and angles values) obtained from a training set of images. For this purpose, two different facial emotion databases have been used: the FG-NET database [17] that provides video sequences of 19 different Caucasian people; and the MMI Facial Expression Database [18] that holds 1280 videos of 43 different subjects from different races (Caucasian, Asian and Arabic). Both databases show the seven universal emotions of Ekman. A total of 1500 static frames were carefully selected from those databases to be use as training sets in the correlation study and in the tuning of the classifiers.

A study of the correlation matrix and dispersion diagrams between parameters was done. The idea was to determine the parameters most influential to the variable to predict (emotion) as well as to detect redundant parameters. From the obtained results, a set of important conclusions were extracted: 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, we decided to work only with the most significant ones, marked in grey in Figure 2.



PARAMETER	MPEG4 FEATURE POINTS
LD1	distance(3.14,3.10)
RD1	distance(3.13,3.9)
LD2	distance(4.2,3.8)
RD2	distance(4.1,3.11)
D3	distance(8.1,8.2)
D4	distance(8.4,8.3)
LD5	distance(8.4,3.12)
RD5	distance(8.3,3.7)
D6	distance(4.2,4.1)
A1	angle(8.4,8.1,8.3)
A2	angle(8.1,8.4,8.2)
LA3	angle(4.4,3.6,4.2)
RA3	angle(4.1,3.5,4.3)
A4	angle(3.6,8.2,3.5)

Fig. 2. Facial parameters tested (left). On the right, relationship between the parameters and the MPEG4 feature points

3 A Novel Combination of Classifiers

3.1 Classifiers Selection

In order to select the classifiers used in this work, the Waikado Environment for Knowledge Analysis (Weka) tool was used. It provides a collection of machine learning algorithms for data mining tasks [19]. From this collection, five classifiers were selected after suitable tuning them: 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:

- RIPPER is a propositional “if…then…” rule-learner algorithm named “Repeated Incremental Pruning to Produce Error Reduction”. The reason for having chosen this algorithm is the simplicity of its rules and its good performance, especially on large noisy datasets.
- Multilayer Perceptron is the most widely used neural network for classification tasks. The multilayer perceptron's power comes from its similarity to certain biological neural networks in the human brain, which is very useful for our working domain.
- The selection of the SVM classifier is due to the growing interest it arised in the literature in the last years.

- Naive Bayes is a simple probabilistic classifier based on applying Bayes' theorem with strong (naive) independence assumptions. Naive Bayes classifiers can be trained very efficiently in a supervised learning setting.
- C4.5 is also a rule-based classifier, but it is used to generate a decision tree. Its good performance is due to the use of the concept of Information Entropy to establish the classification mechanisms.

The classification results obtained for each classifier and emotion are shown in Table 1. These results are obtained from a 10-fold cross-validation test over the 1500 training images. Cross-validation measures the ability of the classifier to adapt itself to new data; this is why more realistic measurements are obtained. Therefore, values obtained with k-fold cross-validation are usually poorer that the ones obtained without cross-validation, as it can be observed in Table 2. This is an important issue when comparing classifiers' success rates.

Table 1. Success rates obtained for each classifier and each emotion with a 10-fold cross-validation test over the 1500 training images

	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%

Table 2. Success rates obtained for each classifier and each emotion without cross-validation test over the 1500 training images

	Disgust	Joy	Anger	Fear	Sadness	Neutral	Surprise
RIPPER	94.10%	100.00%	88.90%	85.20%	46.70%	86.00%	84.40%
SVM	94.10%	100.00%	77.80%	77.80%	70.00%	88.00%	93.30%
C4.5	91.20%	95.20%	92.60%	85.20%	93.30%	92.00%	91.10%
Naive Bayes	85.30%	90.50%	70.40%	85.20%	50.00%	92.00%	80.00%
Multilayer Perceptron	100.00%	100.00%	88.90%	100.00%	76.70%	100.00%	97.80%

3.2 Classifiers Combination

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 weighted and combined through some thought process to reach a final decision that is presumably the most informed one. 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 beat the performance of the best classifier in the ensemble, but it certainly reduces the overall risk of making a particularly poor selection [20].

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. To illustrate the combination strategy, let's imagine a simple case where only 2 classifiers are used and where the confusion matrix of each one of them is the given in Table 3.

Table 3. Confusion matrix of two classifiers (example)

Classifier 1

Emotion --> is classified as	Disgust	Joy	Anger	Fear	Sadness	Neutral	Surprise
Disgust	68.75%	6.25%	9.38%	3.13%	9.38%	3.13%	0.00%
Joy	6.98%	90.70%	0.00%	0.00%	0.00%	0.00%	2.33%
Anger	7.41%	0.00%	70.37%	0.00%	14.81%	7.41%	0.00%
Fear	10.34%	0.00%	0.00%	58.62%	3.45%	0.00%	27.59%
Sadness	17.39%	4.35%	4.35%	4.35%	56.52%	13.04%	0.00%
Neutral	0.00%	0.00%	7.02%	0.00%	15.79%	75.44%	1.75%
Surprise	0.00%	0.00%	0.00%	18.18%	0.00%	2.27%	79.55%

Classifier 2

Emotion --> is classified as	Disgust	Joy	Anger	Fear	Sadness	Neutral	Surprise
Disgust	57.78%	4.44%	11.11%	6.67%	11.11%	2.22%	6.67%
Joy	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Anger	9.52%	0.00%	71.43%	0.00%	14.29%	4.76%	0.00%
Fear	11.54%	0.00%	0.00%	61.54%	7.69%	3.85%	15.38%
Sadness	12.50%	0.00%	16.67%	0.00%	50.00%	20.83%	0.00%
Neutral	0.00%	1.85%	5.56%	0.00%	12.96%	77.78%	1.85%
Surprise	0.00%	0.00%	0.00%	17.39%	2.17%	0.00%	80.43%

When a classifier outputs a specific emotion, the row of its confusion matrix table corresponding to this emotion is selected. If the procedure is repeated for each classifier, a matrix of as many rows as classifiers studied and seven columns, corresponding to the seven emotions is created. The sum of the values present in each column of the matrix gives the score obtained by each emotion ("global output"). If the score result is divided by the number of classifiers, the average global output of the classifier is obtained.

For example, if classifier 1 outputs "disgust" and classifier 2 outputs "anger", the rows in grey in Table 3 is selected. From these data, the matrix shown in Table 4 is created. The sum of the columns allows to obtain the average global output for each emotion. As it can be seen, in the example the system detects emotions "anger" and "disgust" approximately with the same confidence value. This strategy has been applied to combine the five classifiers. Results obtained are presented in next section.

Table 4. Integrating the results of two classifiers (example)

C1 output	68.75%	6.25%	9.38%	3.13%	9.38%	3.13%	0.00%
C2 output	9.52%	0.00%	71.43%	0.00%	14.29%	4.76%	0.00%
Average global output	39.14%	3.13%	40.40%	1.56%	11.83%	3.94%	0.00%

4 Results: Success Rates and Human Assessment

The results obtained when applying the strategy explained in the previous section to combine the results of the five classifiers are shown in first row of Table 5. As it can be observed, the success rates for the neutral, joy and surprise emotions are very high (84.44%-95.23%). However, the system tends to confuse disgust with fear and fear with surprise; therefore, the performances for those emotions are slightly smaller. This is a problem that usually arises with these three emotions; this is why many classification works do not consider them. The showiest result of our classification is for sadness: it is confused with the neutral emotion in the 68% of the occasions, owing to the similarity of their facial expressions. Nevertheless, results can be considered positive, as the confusion matrices of two incompatible emotions (such as sadness and joy or joy and disgust) intermingle in less than the 0.2% of the occasions.

Table 5. Initial results obtained combining the five classifiers (first row) and after considering human assessment (second row)

Emotion	Joy	Surprise	Disgust	Anger	Sadness	Fear	Neutral
Initial results	95.23%	84.44%	79.41%	74.07%	30%	62.96%	92.00%
After human assessment	95.23%	91.11%	84.24%	77.78%	66.67%	92.59%	98.00%

In order to take into account the human factor in the evaluation of the results, 60 persons were told to classify the 1500 images of the database in terms of emotions. As a result, each one of the frames was classified by 10 different persons that were surveyed in 5 sessions of 50 images. With this information, the evaluation of results was repeated: the recognition was marked as "good" if the decision was coherent with the one taken by a human being. It is important to realize that, according to Bassili [21], a trained observer can correctly classify faces showing emotions with an average of 87%.

This kind of evaluation revision is really interesting and, as we will see, useful but it is not performed in other classification works. For example, in the image shown in Figure 3, the FG-NET database classifies it like "disgust" exclusively while the surveyed people recognized it as much "disgust" as "anger" and "sadness". Users' results are similar to our method, which obtains: 54.59% anger, 33.36% disgust, 12.15% sadness.

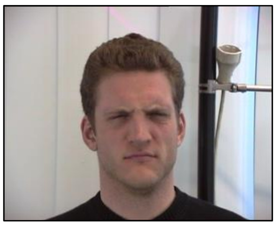


Fig. 3. Frame classified as "disgust" by the FG-NET database [17]

The results of considering users' assessment are shown in second row of Table 5. As it 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 than the users' ones: our classification strategy is appropriate and coherent with human classification.

5 Comparison with Other Methods

Table 6 compares the presented system, in the grey column, with other recent approaches with the same experimental proposal [5,8,11,3]. Many works do not detail if they have used or not cross-validation, so the direct comparison of results is not always possible. As it can be observed, our success rates are generally better. Besides, their confusion matrices uniformly distribute classification errors among all emotions: the probability of confusing joy with sadness, two incompatible emotions, is the same that the one of confusing joy with surprise. This is not the case in our method, as it has been explained in the previous section. It is also important to realize that the database used in this work is bigger than the used in the other ones (1500 images of 62 individuals of all races and genders), and therefore more universal.

Table 6. Classification rates of our method plus user assessment (in grey); and comparisons with the rates obtained by other recent approaches

		Combination of 5 classifiers + user assessment	**Method of Hammal et al. [5]**	**Method of Datcu & Rothkrantz [8]**	**Method of Cohen et al. [11]**	**Method of Zhang et al. [3]**
Type of classifier		combination	rule-based	SVM	Bayesian net (HMM)	neural network
Database		1500 frames, 62 subjects	630 frames, 8 subjects	474 frames	>40 subjects	213 frames, 9 japanese females
Validation strategy		10-fold cross-validation	hold-out method	2-fold cross-validation	leave-one-out cross-validation	10-fold cross-validation
User assesment		yes	no	no	no	yes
Success rates	**Joy**	95.23%	87.26%	72.64%	97.00%	90.10% The only available data is the overall recognition rate of the 6 + neutral universal emotions.
	Surprise	91.11%	84.44%	83.8%	85.00%	
	Disgust	88.24%	51.20%	80.35%	88.00%	
	Anger	77.78%	not recognized	75.86%	80.00%	
	Sadness	66.67%	not recognized	82.79%	85.00%	
	Fear	92.59%	not recognized	84.70%	93.00%	
	Neutral	98.00%	88.00%	not recognized	96.00%	

6 Conclusions and Future Work

This paper describes an effective method for the automatic classification of facial expressions. The method is based on the combination of the 5 classifiers most widely used in the literature. The combination strategy has been weighted majority voting. The classification results are obtained from a 10-fold cross-validation test over 1500 training images, and results are promising.

A comparison with other four recent works has been presented. The distinguishing features of our work are:

- the presented method is able to consider the six Ekman's emotions, plus the neutral one,
- it gives a confidence level for all classified emotions,
- the success rates are generally better,
- classification errors are not uniformly distributed: the confusion matrices of incompatible emotions intermingle in less than the 0.2% of the occasions,
- it has been tuned with a large database of individuals of all races and genders.

Another significant difference is the use of human assessment in the evaluation of the classification results. 60 persons were told to classify the 1500 images of the database in terms of emotions: our classification strategy has been proved to work in a similar way as human brain, leading to similar confusions.

Our emotional classifier is being used as a new multimodal input to Maxine [22], an engine developed by the group for managing 3D virtual scenarios and characters. Maxine focuses in the use of 3D characters to enrich user interaction in different application domains. User's emotion detection is a very useful input to develop affective computing strategies: the general vision is that if a user's emotion could be recognized by a computer, human-computer interaction would become more natural, enjoyable and productive. The computer could offer help and assistance to a confused user or try to cheer up a frustrated user, and hence react in more appropriate ways.

In the near future, we are considering new inputs to the system like adding information about the user's speech (frequency, volume, speed, etc.), introducing dynamic information (i.e. the evolution in the time of the evaluated facial parameters) or making a fuzzification of the input variables.

Acknowledgments. The authors wish to thank the Computer Graphics, Vision and Artificial Intelligence Group of the University of the Balearic Islands for providing us the real-time facial tracking module to test our classifier.

This work has been partially financed by the Spanish "Dirección General de Investigación", Nº TIN2007-63025 and by the Aragon Government through the WALQA agreement (ref. 2004/04/86) and the CTPP02/2006 project.

References

1. Boukricha, H., Becker, C., Wachsmuth, I.: Simulating Empathy for the Virtual Human Max. In: 2nd International Workshop on Emotion and Computing in conj. with the German Conference on Artificial Intelligence, Osnabrück, Germany, pp. 22–27 (2007)
2. Pantic, M., Rothkrantz, L.J.M.: Automatic Analysis of Facial Expressions: The State of the Art. Pattern Analysis and Machine Intelligence, IEEE Transactions 22(12), 1424–1445 (2000)
3. Zhang, Z., Lyons, M., Schuster, M., Akamatsu, S.: Comparison between Geometry-Based and Gabor Wavelets-Based Facial Expression Recognition Using Multi-Layer Perceptron. In: Proc. Int'l Conf. Automatic Face and Gesture Recognition, pp. 454–459 (1998)
4. Wallace, M., Raouzaiou, A., Tsapatsoulis, N., Kollias, S.: Facial Expression Classification Based on MPEG-4 FAPs: The Use of Evidence and Prior Knowledge for Uncetainty Removal. In: IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), vol. 1, pp. 51–54 (2004)

5. Hammal, Z., Couvreur, L., Caplier, A., Rombaut, M.: Facial Expressions Recognition Based on the Belief Theory: Comparison with Diferent Classifiers. In: Proc. 13th International Conference on Image Analysis and Processing (2005)
6. Pantic, M., Rothkrantz, L.J.M.: Expert System for Automatic Analysis of Facial Expression. Image and Vision Computing J. 18(11), 881–905 (2000)
7. Cerezo, E., Hupont, I.: Emotional Facial Expression Classification for Multimodal Interfaces. In: Perales, F.J., Fisher, R.B. (eds.) AMDO 2006. LNCS, vol. 4069, pp. 405–413. Springer, Heidelberg (2006)
8. Datcu, D., Rothkrantz, L.J.M.: Facial Expression Recognition in still pictures and videos using Active Appearance Models. A comparison approach. In: Proceedings of the 2007 International Conference on Computer Systems and Technologies, vol. 285(112) (2007)
9. Qinzhen, X., Pinzheng, Z., Wenjiang, P., Luxi, Y., Zhenya, H.: A Facial Expression Recognition Approach Based on Confusion-Crossed Support Vector Machine Tree. In: International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 309–312 (2006)
10. Limin, M., Chelberg, D., Celenk, M.: Spatio-temporal modeling of facial expressions using Gabor-wavelets and hierarchical hidden Markov models. In: IEEE International Conference on Image Processing, vol. 2, pp. 57–60 (2005)
11. Cohen, I., Sebe, N., Garg, A., Chen, L.S., Huang, T.S.: Facial expression recognition from video sequences: temporal and static modeling. Computer Vision and Image Understanding 11(1-2), 160–187 (2003)
12. Duang-Duang, Y., Lian-Wen, J., Jun-Xun, Y., Li-Xin, Z., Jian-Cheng, H.: Facial expression recognition with Pyramid Gabor Features and Complete Kernel Fisher Linear Discriminant Analysis. International Journal of Information Technology 11(9), 91–100 (2005)
13. Edwards, G.J., Cootes, T.F., Taylor, C.J.: Face Recognition Using Active Appearance Models. In: Proc. European Conf. Computer Vision, vol. 2, pp. 581–695 (1998)
14. Ekman, P.: Facial Expression, the Handbook of Cognition and Emotion. John Wiley et Sons, Chichester (1999)
15. Esau, N., Wetzel, E., Kleinjohann, L., Keinjohann, B.: Real-time facial expression recognition using a fuzzy emotion model. In: IEEE International Fuzzy Systems Conference FUZZ-IEEE 2007, pp. 1–6 (2007)
16. Manresa-Yee, C., Varona, J., Perales, F.J.: Towards hands-free interfaces based on real-time robust facial gesture recognition. In: Perales, F.J., Fisher, R.B. (eds.) AMDO 2006. LNCS, vol. 4069, pp. 504–513. Springer, Heidelberg (2006)
17. FG-Net Database, http://www.mmk.ei.tum.de/~waf/fgnet/feedtum.html
18. Pantic, M., Valstar, M.F., Rademaker, R., Maat, L.: Web-based Database for Facial Expression Analysis. In: Proc. IEEE Int'l Conf. Multmedia and Expo. (2005)
19. Witten, I., Frank, E.: Data Mining: Practical machine learning tools and techniques, 2nd edn. Morgan Kaufmann, San Francisco (2005)
20. Polikar, R.: Ensemble Based Systems in Decision Making. IEEE Circuits and Systems Magazine 6(3), 21–45 (2006)
21. 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 37, 2049–2059 (1997)
22. Cerezo, E., Baldassarri, S., Seron, F.: Interactive agents for multimodal emotional user interaction. In: IADIS Multi Conferences on Computer Science and Information Systems 2007, pp. 35–42 (2007)