An Expert System for Multiple Emotional Classification of Facial Expressions

M. Pantic and L.J.M. Rothkrantz

Delft University of Technology
Faculty of Information Technology and Systems (ITS)
Department of Knowledge Based Systems
P.O. Box 356, 2600 AJ Delft, the Netherlands
{M.Pantic,L.J.M.Rothkrantz}@cs.tudelft.nl.

Abstract

This paper discusses Integrated System for Facial Expression Recognition (ISFER), which performs facial expression analysis from a still dual facial view image. The system consists of three major parts: facial data generator, facial data evaluator and facial data analyser. While the facial data generator applies fairly conventional techniques for facial features extraction, the rest of the system represents a novel way of performing a reliable identification of 30 different face actions and a multiple classification of expressions into the six basic emotion categories. An expert system has been utilised to convert low level face geometry into high level face actions, and then this into highest level weighted emotion labels. The system evaluation results demonstrated rather high concurrent validity with human coding of facial expressions using FACS [4] and formal instructions in emotion signals [5].

1. Introduction

The user interface for computer systems is currently evolving to an intelligent multi-modal tool. Processing, understanding and emulating auditory and visual human communicative signals by a computer will facilitate a revolutionary human-like man-machine interface.

The existing expression recognition systems (e.g. Hong et al. [9], Otsuka and Ohya [13], Kobayashi and Hara [12], Thalmann et al. [16]) mostly deal with the analysis and singular classification of the six prototypic facial expressions as defined by Ekman [5] (happiness, anger, disgust, fear, surprise, sadness). However, it is not certain at all that any facial expression able to be displayed on the face can be singularly classified under the six basic emotion categories. A psychological discussion on the topic can be found in Russell [15] and Ekman [6]. Experimental proofs can be found in the studies of Asian researchers such as Zhang et al. [19], which reported that the Asian subjects have difficulties to express some of the basic expressions such as disgust and fear. To achieve a realistic analysis of human facial displays, automated discrimination of subtle changes in facial expression and a multiple classification of these into expression categories are needed.

The Facial Action Coding System (FACS) [4] is a system designed for human observers to detect subtle changes in facial appearance. It is a system that linguistically describes all possible visually detectable facial changes in terms of 44 so-called Action Units (AUs). So far, several studies on vision-based facial gesture analysis suggested that FACS AUs could be detected from digitised face images.

Kearney and McKenzie [11] reported on a self-adaptive expert system that converts facial data into a set of face actions and then this into a set of emotion labels. The system recognises 36 different face actions but uses hand-measured manually supplied face image data that is difficult to track automatically.

Essa and Pentland [7] proposed a method for recognition of facial expressions based on differential patterns of optical flow. They used spatio-temporal templates to recognise 2 face actions and 3 prototypic emotional expressions. However templates are unsuitable for face action recognition since for each individual face action and each combination of various face actions a separate template should be defined.

Black and Yacoob [2] also used an optical flow model of image motion for facial expression analysis. Although their system utilises some mid-level predicates that describe the facial change, the specificity of optical flow to action unit discrimination was not described. The discrimination of facial expressions remained at the level of the basic emotion prototypes rather than on a finer level of face actions.

Cohn et al. [3] proposed an optical-flow-based method for discriminating between AUs in the eyebrow, eye and mouth regions. The method can identify 8 individual AUs and 7 AUs combinations.

The method is still tentative. It requires a manual labelling of some facial landmarks in the first frame of the examined image sequence. Also, it cannot deal with arbitrary image sequences – the examined image sequence should start with a neutral facial expression and may not contain more than one face action in a row. The method does not deal at all with expression classification.

We developed a system, referred to as ISFER, which can robustly perform both, recognition of a complex facial expression composed of several individual face actions and multiple classification of the expression into the six basic emotion categories. An advanced human-computer interface could employ our expression-recognition tool as pre-processing for interpretation of the encountered expression and for animation of that expression by a virtual actor.

ISFER forms a part of the ongoing research on intelligent anthropomorphic multi-modal human-machine interface, conducted at the Knowledge Based Systems department of TU Delft. This paper discusses the aspects of the conducted research and the resulting implementation of ISFER.

The overall structure and the main characteristics of ISFER are presented in section 2. The framework for hybrid facial feature tracking is briefly described in section 3. The facial data evaluation is discussed in section 4. The dual-view face model and facial data analysis are explained in section 5. Section 6 provides concluding remarks and a short overview of the future work.

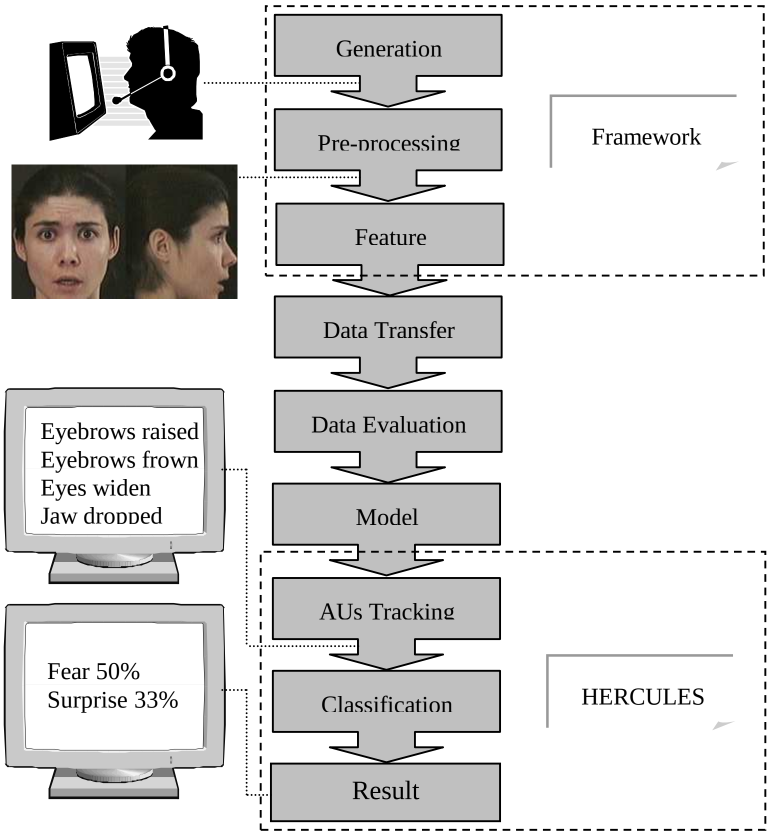

Figure 1. ISFER Structure

2. ISFER

Our system consists of three integral parts (Figure 1): data generator, data evaluator and data analyser. The Facial Data Generator is in fact a framework for hybrid facial feature tracking, which for each facial feature executes multiple feature detectors on the examined dual-view face image. The Facial Data Evaluator makes the best possible selection from the redundantly tracked facial features and substitutes the missing data by setting and checking the hypothesis on the overall facial appearance. The Human Emotion Recognition Clips Utilised Expert System (HERCULES), which converts the evaluated face geometry into face actions and classifies the encountered facial expression into the six basic emotion categories as proposed by Ekman [5], forms the Facial Data Analyser of the system.

Dual view face images are acquired using two digitised cameras mounted on holders attached to a headphone-like device. One camera holder is placed in front of the face at approximately 15 centimetres from the tip of the nose (frontal view). The other camera is placed on the right side of the face at approximately 15 centimetres from the centre of the right cheek (side view). This camera setting is not prone to the problems of rigid head motions. The cameras are moving together with the head and insure the scale and orientation invariance of the acquired images.

ISFER deals with static face actions. This means that only the end-state of the facial movement is measured in comparison to an expressionless face of the same subject. The movement itself is not measured. In other words, the system deals with still face images, not with image sequences.

Since the system detects the examined facial expression from the difference between that expression and the neutral facial expression, the accuracy of the analysis of the expressionless face is crucial. To ensure correct extraction of the facial features from someone's neutral facial expression, it is highly recommended that the results of the automatic feature tracking are visually inspected and, if necessary, that the choice of facial feature detectors is further manually made. Analysis of each next expression of the observed person is performed in a completely automatic way.

The reasoning of the system is person-independent. This means that the process of automatic recognition of facial gestures does not depend on physiognomic variability of the subjects. The generic face model and the rules for recognition of face actions that are based on the person-independent FACS rules facilitate this.

ISFER has three major imperfections. The system does not deal with the intensities of face actions. A face action either underlies the observed facial expression (face action intensity is set to 100%) or not (face action intensity is set to 0%). Further, the system does not deal with minor inaccuracies of the face geometry delimited by the Facial Data Generator. These effect the face action recognition accuracy and, in turn, the emotional classification accuracy. Finally, ISFER performs emotional classification into the six basic emotion categories. Still, it is not at all certain that all facial expressions able to be displayed on the face can be classified under the six basic emotion categories. Allowing the user to define his/her own classification categories would probably yield more realistic interpretation of the encountered expressions.

3. Facial Data Extraction

The existing automated face analysers usually utilise only one kind of facial feature detectors. In contrast, we are proposing a hybrid approach to facial feature tracking. The Facial Data Generator represents in fact a framework that per facial feature concurrently applies multiple feature detectors of different kinds. For instance, a neural network-based approach originally proposed by Vincent et al. [17] that finds the micro-features of the eyes or an active contour method proposed by Kass et al. [10] with a greedy algorithm for minimising the snake's energy function [18] can currently perform automatic tracking of the eyes. But, any other detector picked up "off the shelves" that performs tracking of the eye contour can be use instead. We are combining known techniques rather than fine-tuning the existing facial feature detectors or inventing new ones.

There are two motivations for combining detectors. First, no time will be spent to invent and implement a new facial feature detector. The existing detectors could be just picked up "off the shelves" and integrated into the framework we are proposing here. Second, it is expected that a combined detector will have increased quality. Each algorithm has circumstances under which it performs extremely well and the facial features that it can track better. This implies that a combined detector will have less weak properties and perform better than the best single detector. Finally, introducing the redundancy by applying multiple detectors per facial feature and then choosing the best of the acquired results will finally yield in a more complete set of the detected facial features.

The framework for hybrid facial feature tracking is a Java-implemented tool that has been developed according to the multi-detector paradigm. The overall design of the framework and its GUI is explained in Rothkrantz et al. [14]. Here, we are providing merely a short overview of the framework structure.

The modules of the framework are classified into three groups. The modules for generating digital dual-view images and for filtering the image data belong to the pre-processing group. The modules that perform tracking of the facial regions (i.e. head contour, profile contour, eyebrows region, eyes region and mouth region) belong to the detection group. The modules that track the contours of the facial features (i.e. eyebrows, eyes, nostrils, mouth and chin) belong to the extraction group. For each facial feature, several detectors have been already integrated into the framework. Still, if adding another feature detector will increase the quality of the current facial feature extraction, the detector can be easily integrated into the framework at any point.

After invoking all of the feature detectors that belong to the extraction group of modules, the result of each one is stored in a separate file. Those files form the input to the Facial Data Evaluator.

4. Facial Data Evaluation

The files that form the input to the facial data evaluation part of ISFER contain redundant data. In the case that none of the detectors of a certain facial feature performs a successful tracking, the output files of the Facial Data Generator will contain missing data about that facial feature. The files can also contain highly inaccurate data. The main function of the Facial Data Evaluator is to make the best possible selection from the redundantly tracked facial features and to deal with the encountered ambiguities in the selected facial data.

The process of dealing with ambiguous facial data is, in fact, the process of checking, reducing and adjusting the set of files that form the output of the framework for hybrid facial feature tracking. The whole process is based on two kinds of knowledge, namely, the evaluation of a specific feature detector and the facial anatomy.

Based on the evaluation results obtained for a given detector, we assigned a certain priority to each facial feature detector integrated into the framework. The facial feature detectors and their priority levels are given in Table 1. Since the evaluation of a particular detector determines our confidence in the accuracy of that detector, the priorities assigned to the detectors are used to select the most confident results of the performed feature tracking.

Table 1. The priority levels of the extraction modules integrated into the framework

Module	Priority
Find Profile Contour	2
Fuzzy Mouth	2
Snake Mouth	1
Curve fitting of the Mouth	2
Snake Eye	1
Eye NN	2
Chain Code Eyebrow	1
Curve fitting of the Eyebrow	2
Find Nose/Chin	1

The knowledge about facial anatomy concerns the facts like “the inner corners of the eyes are stable facial points”, “the face is symmetric”, “a movement of the eyes can be unilateral but most often it’s bilateral”, etc. This knowledge is used to check the correctness of the performed facial feature tracking as well as to substitute missing data.

4.1. Checking the Facial Data

The set of the framework output files is evaluated first in terms of missing data. If a single point represents a tracked facial feature, the file containing that feature is labelled as *missing*. In the case of the pair features (eyes and eyebrows), only if a single point represents each feature, the file is labelled as *missing*. If only one of the features is tracked as a single point then the file is labelled as *missing one*.

The output files that haven’t been labelled as *missing* are evaluated further in terms of highly inaccurate data. The evaluation process consists of the following steps.

1. To conclude that the profile contour is badly tracked the tip of the nose and the top of the forehead should deviate for at least ten pixels from these points tracked in the neutral facial expression. The file containing the tracked profile contour will rarely (if ever) be labelled as *highly inaccurate* considering the overall performance of the algorithm with an average localisation error of 2 pixels.
2. To conclude that the eyes are badly tracked one of the following two requirements should be fulfilled. First, the points representing the inner corners of the eyes are immovable points considering the camera setting. If the position of these points deviates for at least five pixels from the neutral-expression-position of these points, one or both eyes will be flagged as badly tracked. A slight deviation in the position of the inner corners of the eyes uncovers an inaccurate- but not a highly inaccurate tracking. Although the narrowing and the widening of the eyes can be unilateral, it is almost always bilateral [4]. So, the proportion of one eye comparing to the other should be the same in the examined expression as in the neutral expression. If this is not the case, one or both eyes will be flagged as badly tracked. If both eyes are flagged as badly tracked, the file containing the tracked eyes will be labelled as *highly inaccurate*. If only one eye is flagged as badly tracked, the file will be labelled as *highly inaccurate one*. This procedure is applied to each file containing the result of an eye detector.
3. In the case of the eyebrows, the important fact is that no muscle contraction can elongate or de-elongate the eyebrow [4]. This and the camera setting, ensure that the area size of each eyebrow remains the same in each examined frontal-view of the observed person. If the size of the eyebrow area deviates for at least ten pixels from the size of that area measured in the neutral facial expression, the eyebrow will be flagged as badly tracked. If both eyebrows are flagged as badly tracked, the file containing this information will be labelled as *highly inaccurate*. If only one eyebrow is flagged as badly tracked, the file will be labelled as *highly inaccurate one*. This procedure is applied to each file containing the result of an eyebrow detector.
4. The points representing the centres of the nostrils are immovable points considering the camera setting. If the tracked location of the nostrils deviates for more than five pixels from the neutral expression position of the nostrils, the file containing the output of the module Find Nose /Chin will be labelled as *highly inaccurate*.
5. Checking the accuracy of a mouth-tracking algorithm is a pretty difficult task considering the diversity of the possible mouth movements. The mouth can be elongated or de-elongated, wide open or tightened, puckered or sucked in, laughing or crying. The check that we are performing consists of two steps. First, the opening of the mouth calculated from the tracked mouth contour is compared to the distance between the lips calculated from the profile contour. If the compared distances deviate for more than five pixels, the file containing the tracked mouth contour will be labelled as *highly inaccurate*. The second step utilises the mouth-detector checking facility implemented as the Fuzzy Mouth module. The output of the module is a classification of mouth expression into one of the smile, neutral and sad categories. The rules such as “if smile then the mouth corners are up” extend the fuzzy classifier and facilitate a comparison of the fuzzy classifier output with the output of another mouth detector. If the tracked mouth contour doesn’t pass this test, the file containing it will be labelled as *highly inaccurate*. This procedure is applied to each file containing the result of a mouth detector.

At this point, the files that haven’t been labelled as *missing* or *highly inaccurate* are labelled as *good*.

4.2. Reduction / Adjustment of the Facial Data

After the framework output files that are labelled as *missing* are discarded, the reduction and the adjustment of the files proceed as follows.

1. Each output file, which contains the result of an eye detector and has been labelled as *highly inaccurate*, is discarded. If there is no eye-detector file left, the missing data is substituted with the eyes tracked in the neutral facial expression. Otherwise, the non-discarded result of the eye detector with a highest priority (Table 1) will be used in system's further processing. If the eye-detector file with a highest priority is labelled as *missing one* or *highly inaccurate one*, the result of an eye detector with a lower priority will be used to substitute the data about the badly tracked eye. If there is no detector with a lower priority, the successfully tracked eye substitutes the badly tracked eye.
2. In the case of the eyebrows the processing is the same as in the case of the eyes.
3. If the file containing the result of the module Find Nose/Chin is labelled as *highly inaccurate*, the nostrils are set to the neutral-expression-position of the nostrils.
4. Each file, which contains the result of a mouth detector and has been labelled as *highly inaccurate*, is discarded. If there is no mouth-detector file left, the missing data is substituted with the mouth tracked in the neutral facial expression. Otherwise, the non-discarded result of the mouth detector with a highest priority will be used in system's further processing.

The Facial Data Evaluator has three shortcomings. The currently implemented data evaluation process will not discover a mouth contour that greatly extends the horizontal length of the actual mouth. Second, all data labelled as *highly inaccurate* will be discarded and, if no data has been labelled as *good*, the relevant facial feature tracked in the neutral facial expression will substitute the missing feature. By doing so, the accurate information about the examined facial expression gets lost. Finally, ISFER is not able to deal with minor inaccuracies encountered in the framework output.

To enhance the system we should implement both, dealing with face image sequences (the features tracked in a previous frame could be used to substitute missing data) and fuzzy reasoning on face image data.

5. Facial Data Analysis

The Facial Data Evaluator results in unambiguously defined face geometry determined as a set of files. The features defined by our face model can be extracted straightforwardly from these files. The extraction is performed in the Model Data Acquiring step of the system's processing (see Figure 1). The obtained face model-based face geometry forms further the input to the reasoning mechanism of the system, HERCULES.

5.1. Face model

We utilise a point-based face model composed of two 2D facial views, namely the frontal and the side view. There are several motivations for this choice. First, the rules of the FACS can be converted in a straightforward manner into the rules for deforming a point-based face model. The validity of the model can be inspected visually by comparing the changes in the model and the changes in the modelled expression. Finally, combining a dual facial view into a single model yields a more realistic representation of 3D face and avoids manual initialisation of a 3D face model (e.g. Thalmann [16]).

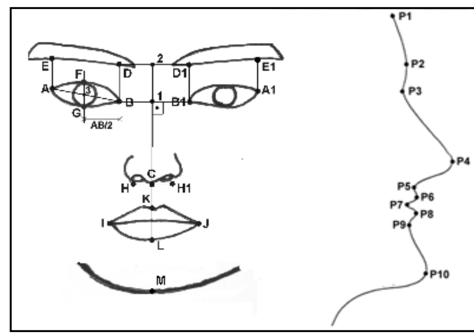

Figure 2. Face model

The frontal-view face model is composed of 19 facial points illustrated in Figure 2. The utilised side-view face model is similar to the profile model proposed by Harmon [8]. It consists of 10 profile points, which correspond with the peaks and valleys of the curvature of the profile contour function (Figure 2).

5.2. Automatic Face Action Tracking

We achieved an automatic face action tracking in two steps. First we perform the automatic tracking of the facial features in the examined face image by utilising the multi-detector processing of the Facial Data Generator. Since the images are scale- and orientation invariant, extraction of the model features from the tracked contours of the facial features is straightforward. Then the obtained face geometry is automatically converted into a set of activated AUs. From a total of 44 AUs defined in FACS, 28 AUs (i.e. 30 face actions) can be uniquely described using our face model. FACS description of AUs and the representation of AUs-codes in terms of our model are given in Table 2 in an informal reader-oriented pseudo-code. These are the rules employed by the system's inference engine, HERCULES.

Table 2. The rules for the recognition of the face actions based on our face model. Points P1 - P10 belong to the side-view face model while the rest of the points belong to the frontal-view face model (Figure 2). Point 3 is the centre of the distance AB and point 4 is the centre of the distance A1B1.

AU	Description	Recognition	AU	Description	Recognition
1	Raised inner brows	increased ∠BAD and ∠B1A1D1	19	Tongue showed	curvature between P6 and P8 contains two valleys and a peak
2	Raised outer brow	increased ∠BAD or ∠B1A1D1	20	Mouth stretched	increased f16, non-increased f12, non-increased f13
4	Lowered / frowned brows	P2 downwards, not increased curvature between P2 and P3	23	Lips tightened but not pressed	non-activated AU28b, non-activated AU28t, non-activated AU8, decreased KL, KL > 0, non-decreased IJ, non-increased IB, non-increased JB1
5	Raised upper lid	increased 3F or increased 4F1			
6	Raised cheek	activated AU12			
7	Raised lower lid	non-activated AU12, non-activated AU9, FG > 0, F1G1 > 0, 3F > 0, 4F1 >0, decreased 3G or decreased 4G1	24	Lips pressed together	non-activated AU28b, non-activated AU28t, non-activated AU8, decreased KL, KL > 0, decreased IJ < t1
8	Lips towards each other (teeth visible, lips tensed and less visible)	increased P5P6, P6 outwards, P8 outwards, curvature between P6 and P8 angular ([ ), increased P8P10	25	Lips parted	increased P6P8, P4P10 < t2
			26	Jaw dropped	t2 < P4P10 < t3
			27	Mouth stretched	P4P10 > t3
9	Wrinkled nose	increased curvature between P2 and P3	28	Lips sucked in	Points P6 and P8 are absent
10	Raised upper lip	P6 upwards, P6 outwards, decreased P5P6, not increased curvature between P2 and P3	28b	Bottom lip sucked	Point P8 is absent
			28t	Top lip sucked in	Point P6 is absent
12	Mouth corners pulled up	decreased IB, decreased JB1, increased CI, increased CJ	36t	Bulge above the upper lip produced by the tongue	increased curvature between P5 and P6
13	Mouth corners pulled sharply up	decreased IB, decreased JB1, decreased CI, decreased CJ	36b	Bulge under the lower lip produced by the tongue	Point P9 is absent
15	Mouth corner pulled downwards	increased IB or increased JB1	38	Nostrils widened	absent AUs: 8, 9, 10, 12, 13, 14, 15, 18, 20, 24, 28 increased HH1
16	Depressed lower lip	P8 downwards, P8 outwards, decreased P8P10	39	nostrils compressed	decreased HH1
17	Raised chin	P10 inwards	41	Lid dropped	(non-decreased 3G, decreased FG, decreased 3F) or (decreased F1G1, decreased 4F1, non-decreased 4G1)
18	Lips puckered	decreased IJ > t1			

These rules have been validated twice. First, we asked three certified FACS coders to produce the facial expressions of separate AU activation, according to the rules given in Table 2. Only the changes described in the table have been produced, the appearance of other facial features is left unchanged. 90 recorded dual views were given for evaluation to other two certified FACS coders. In 100% of the cases, the image representing the activation of a certain AU, produced according to our rules, has been labelled with the same AU-code by the FACS coders.

The second validation test of the rules for AU recognition concerns the automatically performed face action tracking from 496 dual views. The images are 31 expressions of separate face actions shown by eight certified FACS coders twice (2x8x31). The images have been made strictly according to the rules given in Table 2. Dual views have been recorded under constant illumination using fixed light sources and none of the subjects had a moustache, a beard or wear glasses. Subjects were of both sexes and ranged in age (22-33) and ethnicity (European, South American and Asian). The average recognition rate was 92% for the upper face AUs (AU1-AU7) and 86% for the lower face AUs. For 2% of the images, the tracking failed completely.

5.3. Emotional Classification of Expressions

The set of HERCULES' production rules given in Table 3 performs the multiple classification of the face actions into the emotion categories. These rules have been acquired in a straightforward manner from the linguistic descriptions of the prototypic expressions given by Ekman [5]. Five certified FACS coders have validated the rules using a set of 129 dual view images representing the relevant combinations of AUs. In 85% of the cases, the human observer labelled the observed expression as given in Table 3.

An AU-coded description of the shown expression and its classification given in terms of weighted emotion labels conclude the facial expression analysis performed by ISFER. A weight is assigned to an emotion label according to the assumption that each AU, forming a part of a certain prototypic expression (Table 4), has the same influence on intensity of that expression. For instance, an expression formed by activating AU6, AU12 and AU25 will be classified as 75% of happiness.

Table 4. AUs-coded description of the basic emotions

Expression	Description in terms of AUs
Happiness	6 + (12 with or not 16 + (25 or 26))
Sadness	(1 with or not 4) + (6 or 7) + 15 + 17 + (25 or 26)
Anger	4 + 7 + (((23 or 24) with or not 17) or (16 + (25 or 26)) or (10 + 16 + (25 or 26))) with or not 2
Disgust	((10 with or not 17) or (9 with or not 17)) + (25 or 26)
Fear	(1 + 4) + (5 + 7) + 20 + (25 or 26)
Surprise	(1 + 2) + (5 without 7) + 26

To evaluate the semantic correctness of the rules of Table 4, we asked three certified FACS coders to produce facial expressions according to these rules. The acquired 54 images (3 times 6 expressions shown by 3 subjects) were given for evaluation to other five certified FACS coders. The achieved average of the correct recognition ratio of 86% validated the rules [1].

The overall performance of the automatic emotional classification of facial expressions performed by the system has been tested on a set of 265 face images. The images are: 129 images used to validate the rules of Table 3, 56 images representing "pure" basic emotional expressions (including neutral expression) and 80 images of various blended emotional expressions shown by 8 certified FACS coders. Image acquisition has been performed under constant illumination using fixed light sources and none of the subjects had a moustache, a beard or wear glasses. Subjects were of both sexes and ranged in age (22-33) and ethnicity

Table 3. The rules of HERCULES for multiple classification of facial expression into the six basic emotion categories

AUs	Emotion	AUs	Emotion	AUs	Emotion	AUs	Emotion
1+2	surprise	4	anger	23+17	anger	10+17	disgust
2	anger	5	surprise	23+26	anger	10+(25/26)	disgust
1	sadness	6	happiness	23	anger	10	disgust
1+4+5+7	fear	7	anger	24+17+26	anger	9+(25/26)	disgust
1+4+5	fear			24+17	anger	9+17	disgust
1+4+7	sadness	27	surprise	24+26	anger	9	disgust
1+5+7	fear	20+(25/26)	fear	24	anger	12+(25/26)	happiness
1+4	sadness	20	fear	10+16+(25/26)	anger	12	happiness
1+5	fear	15+(25/26)	sadness	10+17+(25/26)	disgust	16+(25/26)	anger
1+7	sadness	15	sadness	9+17+(25/26)	disgust	17	sadness
5+7	fear	23+17+26	anger	12+16+(25/26)	happiness	26	surprise

(European, South American and Asian). First, the images were manually classified according to the rules of Table 3. The performance of the automatic classification is then evaluated by counting the images that have been correctly classified and weighted by the system. In only 2% of the images (6 images) the tracking failed completely. The average correct recognition ratio was 91% (Table 5).

Table 5. Distribution of the correct recognition ratio and the misrecognition ratio of 265 emotional expressions; "B" stands for blended emotional expression

Exp.	Recognised expression						
	Sur	Fear	Dis	Ang	Hap	Sad	B
surprise	97	1	0	0	0	0	2
fear	0	84	0	0	0	9	7
disgust	0	0	82	14	0	0	3
anger	0	1	12	84	0	0	2
happy	1	0	0	0	98	0	1
sad	0	2	0	0	0	96	2
B	3	1	0	0	2	1	93

Average: 90.57%

6. Conclusion

This paper presents a prototype of the person- and situation independent system for vision-based facial gesture analysis, which utilises a framework for hybrid facial feature tracking and an Expert System for face action tracking and multiple emotional classification of facial expressions. By a number of experiments, we demonstrated the validity of the rules that have been employed. The evaluation of the overall performance of the fully automated system indicates that the facial feature tracking, the face action tracking and the face action emotional classification are performed rather accurately by the system.

Our current work is focused on a threefold. Modelling the facial motion and its intensity (i.e. dealing with face image sequences and AU intensity) will increase the overall performance of the system. Developing a Fuzzy Expert System for face action tracking and face action emotional classification will increase the quality of the system by allowing it to reason about the involved face actions according to the accuracy of the performed facial feature tracking. Designing and developing a learning facility, which will allow the user to define his/her own interpretation categories, will yield a broader and more realistic classification of the encountered expressions.

References

[1] J.N. Bassili, "Facial motion in the perception of faces and of emotional expression", *Journal of Experimental Psychology* 4, pp. 373-379, 1978.

[2] M.J. Black and Y. Yacoob, "Recognising Facial Expressions in Image Sequences Using Local Parameterised Models of Image Motion", *Journal of Computer Vision* 25 (1), pp. 23-48, 1997.

[3] J.F. Cohn, A.J. Zlochower, J.J. Lien and T. Kanade, "Feature-Point Tracking by Optical Flow Discriminates Subtle Differences in Facial Expression", *Proc. 3rd IEEE FG*, pp. 396-401, 1998.

[4] P. Ekman, *Facial Action Coding System (FACS)*. Palo Alto: Consulting Psychologists Press, 1978.

[5] P. Ekman, *Emotion in the Human Face*. Cambridge: Cambridge University Press, 1982.

[6] P. Ekman, "Strong Evidence for Universals in Facial Expressions: A Reply to Russell's Mistaken Critique", *Psychological Bulletin* 115 (2), pp. 268-287, 1994.

[7] I. Essa and A. Pentland, "Coding, Analysis, Interpretation and Recognition of Facial Expressions", *IEEE Trans. PAMI* 19 (7), pp. 757-763, 1997.

[8] L.D. Harmon, M.K. Khan, R. Lash and P.F. Raming, "Machine identification of Human Faces", *Pattern Recognition* 13, pp. 97-110, 1981.

[9] H. Hong, H. Neven and C. von der Malsburg, "Online Facial Expression Recognition based on Personalised Galleries", *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, pp. 354-359, 1998.

[10] M. Kass, A. Witkin and D. Terzopoulos, "Snake: Active Contour Model", *Proc. 1st ICCV*, pp. 259-269, 1987.

[11] G.D. Kearney, S. McKenzie, "Machine Interpretation of Emotion", *Cognitive Science* 17 (4), pp. 589-622, 1993.

[12] H. Kobayashi and F. Hara, "Facial Interaction between Animated 3D Face Robot and Human Beings", *Proc. IEEE Conf. System, Man and Cybernetics*, pp. 3732-3737, 1997.

[13] T. Otsuka and J. Ohya, "Spotting Segments Displaying Facial Expression from Image Sequences Using HMM", *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, pp. 442-447, 1998.

[14] L. J. M. Rothkrantz, M. van Schouwen, F. Ververs and J. Vollering, "A multimedial workbench for facial expression analysis", *Proc. EUROMEDIA*, pp. 94-101, 1998.

[15] J.A. Russell, "Is there Universal Recognition of Emotion From Facial Expression? A Review of the cross-cultural studies", *Psychological Bulletin* 115 (1), pp. 102-141, 1994.

[16] N.M. Thalmann, P. Kalra and M. Escher, "Face to Virtual Face", *Proc. IEEE* 86 (5), pp. 870-883, 1998.

[17] J.M. Vincent, D.J. Myers and R.A. Hutchinson, "Image feature location in multi-resolution images using a hierarchy of multi-layer perceptrons", *Neural Networks for Speech, Vision and Natural Language*, pp. 13-29, 1992.

[18] D.J. Williams and M. Shah, "A fast algorithm for active contours estimation", *Computer Vision and Image Processing: Image Understanding* 55 (1), pp. 14-26, 1992.

[19] Z. Zhang, M. Lyons, M. Schuster and 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.