An Expert System for Recognition of Facial Actions and Their Intensity

M. Pantic and L.J.M. Rothkrantz

Delft University of Technology
Faculty of 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

The Facial Action Coding System (FACS) is an objective method for quantifying facial movement in terms of 44 component actions, i.e. Action Units (AUs). This system is widely used in behavioral investigations of emotion, cognitive process and social interaction. Highly trained human experts (FACS coders) presently perform the coding. This paper presents a system that can automatically recognize 30 AUs, their combinations and their intensity. The system employs a framework for hybrid facial feature detection and an expert system for facial action coding in static dual-view facial images. Per facial feature, multiple feature detection techniques are applied and the resulting redundant data is reduced so that an unequivocal facial expression geometry ensues. Reasoning with uncertainty is used to encode and quantify the encountered facial actions based on the determined expression geometry and the certainty of that data. Eight certified FACS coders tested the system. The recognition results demonstrated rather high concurrent validity with human coding.

Introduction

Facial expressions play the main role in the non-verbal aspect of human communication [11]. Besides, facial movements that comprise facial expressions provide information about affective state, personality, cognitive activity and psycho-pathology. The Facial Action Coding System (FACS) [4] is the leading method for measuring facial movement in behavioural science. FACS is currently executed manually by highly trained human experts (i.e. FACS coders). Recent advances in computer technology open up the possibility for automatic measurement of facial signals. An automated system would make classification and quantification of facial expressions widely accessible as a tool for research and assessment in behavioural science and medicine. Such a system could also form the front-end of an advanced human-computer interface that performs interpreting (e.g. [10]) communicative facial expressions.

This paper presents a system that performs facial expression recognition as applied to automated FACS encoding. From 44 facial actions defined by FACS, our system automatically recognizes 30 facial actions, their combinations and their intensity by applying different AI techniques and non-AI techniques integrated into a single system. We use a hybrid approach, i.e. a combination of different image processing techniques, to extract facial expression information from a static dual-view image. Then we employ a rule-based expert system to encode and quantify the encountered facial actions from the extracted facial expression information and the certainty of that data. Finally another expert system is applied to adjust this result (if necessary), based on an emotional classification of the encountered facial expression. Validation studies on the prototype demonstrated that the recognition results achieved are in 90% consistent with those of eight FACS coders. In addition it has been shown that the quantification of the facial action codes achieved by the system deviates in average for 8% from that done by the FACS coders.

Facial Action Coding System

The Facial Action Coding System (FACS) [4] has been developed to facilitate objective measurement of facial activity for behavioural science investigations of the face. It is a system designed for human observers to visually detect independent subtle changes in facial appearance caused by contractions of the facial muscles. In a form of rules, FACS provides a linguistic description of all possible visually detectable facial changes in terms of 44 so-called Action Units (AUs). Using these rules, a trained human FACS coder decomposes an observed expression into the specific AUs that produced the expression.

Although FACS is the most prominent method for measuring facial expressions in behavioral science, a major impediment to its widespread use is that its manual application is time consuming in addition to the time required to train human experts. Each minute of videotape takes approximately one hour to score and it takes 100 hours of training to achieve minimal competency on FACS. Automating FACS would not only make it widely accessible as a research tool, it would also increase the speed of coding and improve the precision and reliability of facial measurement.

In addition to providing a tool for behavioral science research, a system that outputs facial action codes would provide an important basis for man-machine interaction systems. In natural interaction only 7% of the meaning of a communicative message is transferred vocally while 55% is transferred by facial expressions [11]. FACS provides a description of the basic elements of any facial expression. Integration of automated systems for facial action coding, speech recognition and interpretation of those communicative signals would make human-computer interaction more natural, more efficient and more effective.

Automatic Recognition of Facial Actions

Recent advances in computer vision and pattern analysis facilitated automatic analysis of facial expressions from images. Different approaches have been taken in tackling the problem: analysis of facial motion [6], [1], [12], grey-level pattern analysis [20], analysis of facial features and their spatial arrangements [2], [8], [13], [10], holistic spatial pattern analysis [7], [17]. The image analysis techniques in these systems are relevant to the goal of automatic facial expression data extraction, but the systems themselves are of limited use for behavioural science investigations of the face. In many of these systems the discrimination of expressions remained at the level of few emotion categories, such as happy, sad or surprised, rather than on a finer level of facial actions. Yet, for investigations of facial behaviour itself, such as studying of the difference between genuine and simulated affective state, an objective and detailed measure of facial activity such as FACS is needed.

Explicit attempts to automate facial action coding in images are few [3]. Black et al. [1] use local parameterised models of image motion and few mid-level predicates that are derived from the estimated motion parameters and describe the encountered facial change. Here the specificity of optical flow to action unit discrimination has not been described. Essa et al. [6] use spatio-temporal templates to recognise two facial actions and four prototypic emotional expressions. Cohn et al. [2] achieved some success in automating facial action coding by feature point tracking of a set of points manually located in the first frame of an examined facial image sequence. Their method can identify 8 individual AUs and 7 AUs combinations. Here, each image sequence should start with a neutral facial expression and may not contain more than one face action in a row.

In fact, it is not known whether any of the methods reported up-to-date is sufficient for describing the full range of facial behaviour. None of the systems presented in the literature deals with both, facial action coding and quantification of the codes.

A New Approach

This paper presents a system capable of interpreting static dual-view facial images in terms of facial actions and their intensities involved in the shown facial expression. The system was developed to achieve both:

1. person independent, robust, fully automatic extraction of facial expression information from a dual-view
2. robust, fully automatic quantified facial action coding.

The study of feasibility demonstrated that a rule-based expert system, combined with image analysis techniques for facial expression information extraction, is appropriate paradigm for expression recognition as applied to automated FACS encoding. Here, the rule-based character of FACS and the overall characteristics of the task (i.e. it is a cognitive task that involves reasoning rather than numerical computation on a stable and narrow knowledge domain defined by FACS) decided the issue.

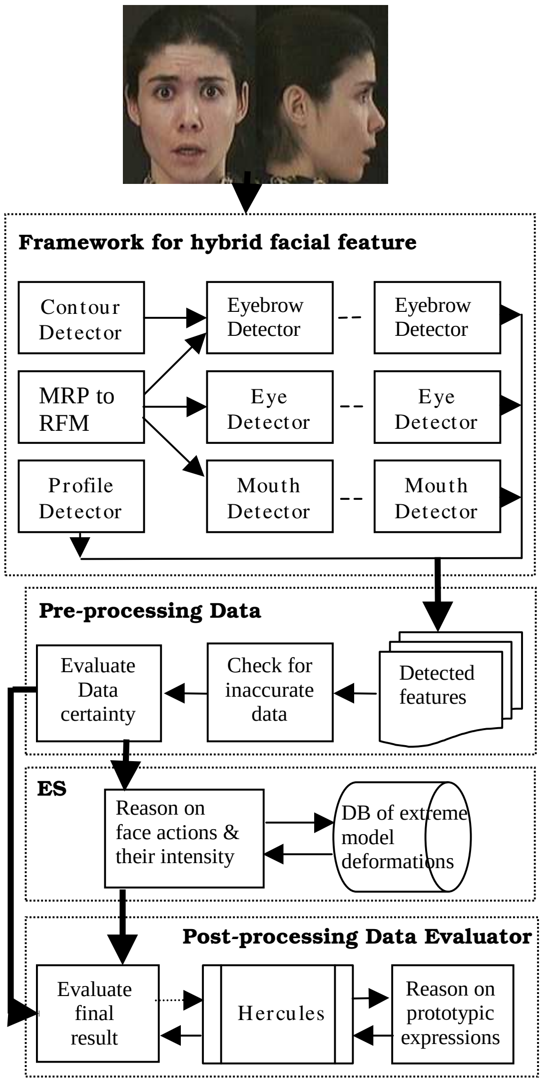

Figure 1. System architecture

Our system consists of four integral parts (Figure 1): data generator, data evaluator, data analyzer and post-processor. The Facial Data Generator is a framework for "hybrid" facial expression information extraction from a dual-view facial image where for each prominent facial feature (eyes, eyebrows, nose and mouth) multiple feature detectors are applied. This part of the system is presented first. Then the Data Evaluator is explained. The Facial Data Evaluator selects per facial feature the best from the results of the applied detectors, substitutes missing data by setting and checking hypotheses about the overall facial appearance and assigns certainty measures (i.e. our confidence in data) to the evaluated data. The Facial Data Analyzer, presented next in this paper, has been implemented as a rule-based expert system that converts the evaluated facial expression data into quantified facial action codes. Finally the Post-Processor is presented. It is a CLIPS implemented rule-based expert system, which classifies the current expression into one of the six basic emotions [5] and based on the result adjusts (if necessary) the result obtained in the previous processing stages. The paper provides technical data on system development, software environment, testing procedures and results. A discussion about the strengths and limitations of the system concludes the paper.

Facial Data Generator

FACS was primarily developed for human observers to perform facial action encoding from full-face photographs of an observed person. Efforts have recently turned to measuring facial actions by image processing of video sequences [2], [6], [1]. This became a trend since there is a growing psychological research that argues that facial expression dynamics are critical in expression analysis. Nevertheless, our work is more in line with the original purpose of FACS – measuring of static facial actions. In our system 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.

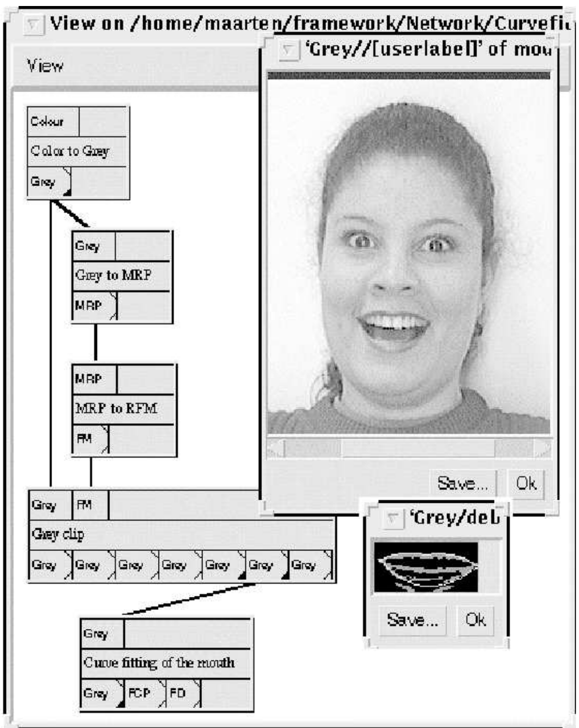

Figure 2. Screen shot of stand-alone mode of the Facial Data Generator

The system deals with static dual-view facial images. Two digitized cameras mounted on the head of the user acquire the images. The cameras are fastened to two holders attached to a headphone-like device. One camera holder is placed in front of the face at approximately 15 centimeters from the tip of the nose (obtains the frontal view). The other camera is placed on the right side of the face at approximately 15 centimeters from the center of the right cheek (obtains the side view). The camera setting ensures the presence of the face in the scene and some out-of-plane head motions cannot be encountered together with the non-rigid facial motion (i.e. the images are scale and orientation invariant).

The existing systems for facial image analysis usually utilize a single kind of feature detectors [3]. In contrast, we are proposing a hybrid approach to facial expression data extraction. To localize the contours of the prominent facial feature (eyebrows, eyes, nose and mouth), for each feature the Facial Data Generator concurrently applies multiple detectors of different kinds. For instance, a neural network-based approach originally proposed by Vincent et al. [16] that finds the micro-features of the eyes and an active contour method proposed by Kass et al. [9] with a greedy algorithm for minimizing the snake's energy function [19] perform currently automatic detection of the eyes. But, any other detector picked up "off the shelves" that achieves localization of the eye contour can be used instead. For profile detection, a spatial approach to sampling the profile contour from a thresholded side-view image is applied [18]. Instead of fine-tuning the existing feature detectors or inventing new ones, known techniques are combined.

The motivation for integrating multiple detectors is the increase in quality of a "hybrid detector". Each typical feature detector has circumstances under which it performs better than another detector. Hence, the chances for successful detection of a given feature increase with the number of integrated detectors. Therefore, by integrating different detectors per facial feature into a single framework, the percentage of missing data is reduced.

The requirement posed on the development of the Facial Data Generator was the integration of the existing detectors in an easy-to-enlarge interactive user-friendly platform that can operate stand-alone as well as a part of a larger system. The stand-alone mode, illustrated in Figure 2, is used for testing of different detectors. Availability of JDK and JNI made Java perfectly suitable for the development of such a software platform. More details about the design of the Facial Data Generator and the integrated feature detectors can be found in [14].

After invoking all integrated detectors, each localized facial feature contour is stored in a separate file. The files form the input to the Facial Data Evaluator (Figure 1).

Facial Data Evaluator

The Facial Data Evaluator operates in two stages. First it delimits the geometry of the encountered expression by choosing the "best" of the redundantly detected facial features stored in the files, which form the output of the Facial Data Generator. In the second stage, the defined facial expression geometry is represented in terms of our face model. The set of the face-model points, together with the assigned *certainty factors* (CFs), forms the input to the Facial Data Analyzer.

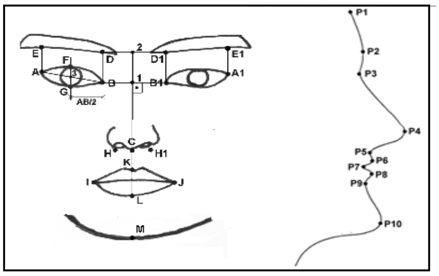

Figure 3. Face model

Selection among the Detected Facial Features

The reasoning of the first stage applies the knowledge about the facial anatomy (e.g. the inner corners of the eyes are immovable points) to check the correctness of the results achieved by the facial feature detectors. Based on this check, each file forming the output of the Facial Data Generator is flagged with one of the labels: *good*, *missing*, *missing one*, *highly inaccurate* and *highly inaccurate one*. If a single point represents the localized contour of a facial feature, the file containing that result is labeled as *missing*. In the case of the pair features (eyes and brows), the file may be labeled as *missing one*. A file is labeled as *highly inaccurate* if there is a lack of consistency in the extracted facial expression geometry. For example, a file containing the result of an eye detector is labeled as *highly inaccurate one* if the localized inner corner of an eye deviates for more than 5 pixels from the inner corner of the pertinent eye localized in the expressionless face of the same subject. The files that pass this check are labeled as *good*. Finally an inter-file consistency check is performed. If the contour stored in the tested file deviates for more than 10 pixels in any direction from relevant contours stored in the other files, the tested file is discarded.

When all of the files are evaluated in terms of missing and highly inaccurate data, the files labeled as *missing* or *highly inaccurate* are discarded and the facial expression geometry is determined by the results stored in the left over files. To make the best choice between the results of different detectors, which detect the same feature, the priorities $n \in \mathbb{N}$ are used. These have been off-line manually assigned to the integrated detectors based on their overall evaluation results. Each facial feature is delimited by the content of a not discarded file that comprises that feature detected by the detector of the highest priority. The priority of the selected detector n (where $n = N$ is the highest priority a detector can have) determines the CF assigned to the feature as given in formula (1).

In the case of the eyes and eyebrows it may happen that the remained files are labeled as *missing one* or *highly inaccurate one*. The eye/eyebrow that has been successfully localized by a detector with the highest priority is used to substitute its pair feature that has been badly localized. The CF of the successfully detected feature is set according to formula (1), while the CF of the feature being replaced is calculated as given in formula (2).

$$CF = (1 / N) * n \qquad (1)$$

$$CF = (1 / (N + 1)) * n \qquad (2)$$

If detection of a certain feature fails (i.e. all of the relative files are discarded), the pertinent feature detected in the expressionless face of the same subject is used to substitute the missing feature. The CF assigned to the feature being substituted in this way is set to *1/2N*.

Representation by the Face Model

We utilize a point-based face model composed of two 2D facial views, namely the frontal and the side view (Figure 3). There are two main motivations for this choice. First, the rules of FACS can be converted straightforwardly into the rules for deforming a point-based face model. Second, the validity of the model can be inspected visually by comparing the changes in the model and the changes in the modeled expression.

The frontal-view face model is composed of 19 facial points. The utilized side-view face model consists of 10 profile points, which correspond with the peaks and valleys of the curvature of the profile contour function [18].

Since all of the detectors integrated into the Facial Data Generator extract contours of the facial features and since the images are scale- and orientation invariant, localizing the model points from the extracted contours of the facial features is straightforward. For instance, point A and point B are localized as the outermost left, respectively, the outermost right point of the contour of the left eye. Point F and point G are localized as the upper, respectively, the lower intersection point of the eye contour with a line going parallel to the vertical face axis through the middle of the line AB (as illustrated in Figure 3).

To each of the model points a CF is assigned that is equal to the CF assigned to the facial feature to which the point belongs. For example, the CF assigned to the points of the side-view model is equal to the CF that has been assigned to the sampled profile contour.

Facial Data Analyzer

The Facial Data Analyzer is the kernel of our system. It performs reasoning with uncertainty about facial actions and their intensity. Table 1 provides the mapping between 30 FACS rules and 30 rules of our expert system.

Each rule of the knowledge base given in Table 1 recognizes activation of a single AU based on the facial change caused by that AU. This means that each rule encodes a certain facial action based on discrepancy of the spatial arrangement of the model points between the current and the neutral expression of the same person.

Table 1. User-oriented pseudo-code of the rules for facial action coding from the face model deformation (Figure 3)

AU	FACS rule	ES rule
1	Raised inner brows	increased ∠BAD and ∠B1A1D1
2	Raised outer brow	increased ∠BAD or ∠B1A1D1
4	Lowered / frowned brows	P2 downwards, not increased curvature P2-P3
5	Raised upper lid	increased 3F or increased 4F1
6	Raised cheek	activated AU12
7	Raised lower lid	no AU12 & AU9 FG>0 F1G1>0, 3F>0, 4F1>0, decreased 3G or decreased 4G1
8	Lips towards each other (teeth visible, lips tensed & less visible)	increased P5P6, P6 outwards, P8 outwards, curvature P6-P8 [increased P8P10
9	Wrinkled nose	increased curvature P2-P3
10	Raised upper lip	P6 upwards, P6 outwards, decreased P5P6, not increased curvature P2-P3
12	Mouth corners pulled up	decreased IB, decreased JB1, increased CI, increased CJ
13	Mouth corners pulled sharply up	decreased IB, decreased JB1, decreased CI, decreased CJ
15	Mouth corner downwards	increased IB or increased JB1
16	Depressed lower lip	P8 downwards, P8 outwards, decreased P8P10
17	Raised chin	P10 inwards
18	Lips puckered	decreased IJ>t1
19	Tongue showed	curvature P6-P8 contains 2 valleys and a peak
20	Mouth stretched	increased f16, not increased f12 not increased f13
23	Lips tightened but not pressed	no AU28b, no AU28t, no AU8, decreased KL, KL>0, not decreased IJ, not increased IB, not increased JB1
24	Lips pressed together	no AU28b, no AU28t, no AU8, decreased KL, KL>0, decreased IJ<t1
25	Lips parted	increased P6P8, P4P10<t2
26	Jaw dropped	t2<P4P10<t3
27	Mouth stretched	P4P10>t3
28	Lips sucked in	Points P6 and P8 are absent
28b	Bottom lip sucked in	Point P8 is absent
28t	Top lip sucked in	Point P6 is absent
36t	Bulge above the upper lip caused by tongue	increased curvature P5-P6
36b	Bulge under the lower lip	Point P9 is absent
38	Nostrils widened	absent AUs: 8, 9, 10, 12, 13, 14, 15, 18, 20, 24, 28 increased HH1
39	nostrils compressed	decreased HH1
41	Lid dropped	not decreased 3G decreased FG, decreased 3F or decreased F1G1, decreased 4F1, not decreased 4G1

The rules have been uniquely defined. In other words, each model deformation corresponds to unique set of AU-codes.

We utilized a *relational list* (R-list) to represent the relations between the rules of the knowledge base. The used R-list is a four-tuple list where the first two columns identify the conclusion clause of a certain rule that forms the premise clause of another rule, identified in the next two columns of the R-list. Each premise clause of each rule given in Table 1 is associated with an S-function, as defined in (3), which influences a so-called cumulative *membership grade* (MG) of the premise of the rule.

$$\begin{aligned}
&S(x;\alpha,\beta,\gamma)=0 && \text{for } x\le\alpha\\
&S(x;\alpha,\beta,\gamma)=2[(x-\alpha)/(\gamma-\alpha)]^2 && \text{for } \alpha<x<\beta\\
&S(x;\alpha,\beta,\gamma)=1-2[(x-\gamma)/(\gamma-\alpha)]^2 && \text{for } \beta<x<\gamma\\
&S(x;\alpha,\beta,\gamma)=1 && \text{for } x\ge\gamma
\end{aligned} \tag{3}$$

where α and γ are function's end points and $\beta=(\alpha+\gamma)/2$ is so-called crossover point

The parameters of S-function are on-line defined by the contents of the database (DB) containing the maximal encountered deformations of the face model. For instance, the S-functions associated with the premises of the rule for recognition of AU5 are defined as *S5_1(x; 0, ½max_3F, max_3F)* and *S5_2(x; 0, ½max_4F1, max_4F1)* where *x* is the actual deformation of the distance 3F, respectively 4F1, and the *max_3F* and *max_4F1* are retrieved from the DB.

The database of extreme model deformations is on-line altered. For each facial-distance/ profile-contour (defined in Table 1), the difference is calculated between that feature detected in the expressionless face and the pertinent feature detected in the current expression. If the determined difference is higher than the related value stored in the DB, the content of the DB is adjusted. The initial values of the extreme model deformations are set off-line, prior the system execution, based on a representative set of facial expressions of the currently observed person. This representative set of facial expressions, i.e. observed persons' *individual extreme-displays (IED) set*, consists of the 6 basic emotional expressions, neutral expression and 4 maximal displays of AU8, AU18, AU39 and AU41. This set of 11 expressions has been experimentally proved to be sufficient for initialisation of the values stored in the DB of extreme model deformations (see rules for facial expression emotional classification in [13]).

Fast direct chaining as defined by Schneider et al. [15] has been applied as the inference procedure. It is a breadth-first search algorithm that starts with the first rule of the knowledge base and then searches the R-list to find if the conclusion of the fired rule forms a premise of another rule that will be fired in the next loop. Otherwise, the process will try to fire the rule that in the knowledge base comes after the rule last fired.

The model points delimited by the Facial Data Evaluator (Figure 3) determine the facial-distances/ profile-contours employed by the rules (Table 1). The CFs associated with the model points define the CF of the related distance/ contour as given in formula (4).

$$CF\_feature = min\ (CF\_point1,\ldots,CF\_pointN) \quad (4)$$

The overall certainty of the premise of a fired rule is calculated as defined by Schneider et al. [15]:

1. For the portion of the premise that contains clauses *c1* and *c2* related as *c1 AND c2*, *CF = min (CF_c1, CF_c2)*.
2. For the portion of the clause that contains clauses *c1* and *c2* related as *c1 OR c2*, *CF = max (CF_c1, CF_c2)*.
3. If the premise contains only clause *c*, *CF = CF_c*.

Further, the cumulative membership grade *MG_p* of the premise of a rule is calculated and multiplied by 100% to obtain the quantification of the AU code encrypted by that rule. *MG_p* of a rule's premise *p* is calculated from the membership grades *MG_c* associated with the clauses *c* of the premise *p*.

1. For a clause *c* of a kind "certain AU (not) activated", *MG_c = 1*. For the portion of the premise that contains *c AND c1* or *c OR c1*, where the clause *c1* is of another kind, *MG_p = MG_c1*.
2. For a clause *c* of a kind "certain point absent /present", *MG_c = 1*. For the portion of the premise that contains *c AND c1* or *c OR c1*, where the clause *c1* is of another kind, *MG_p = MG_c1*.
3. For a clause *c* where two values are compared, $MG\_c = S(x;\ \alpha,\ \beta,\ \gamma)$, where *S* is the S-function associated with *c*. For a portion of the premise that contains *c AND c1*, where *c* and *c1* are of the same kind, *MG_p = avg (MG_c, MG_c1)*. For a portion of the premise that contains *c OR c1*, *MG_p = max (MG_c, MG_c1)*.
4. If the premise contains only clause *c*, *MG_p = MG_c*.

A processing loop of the inference engine ends with updating the DB of the extreme model deformations, updating a *list of fired rules* (LFR) and searching the R-list for a rule that the process will try to fire in the next loop. LFR prevents the inference engine from firing a rule twice. If a rule has fired, its number is added to this list.

Post Processor

In the case a certain facial feature fails to be detected by the Facial Data Generator, the Facial Data Evaluator utilises the pertinent feature detected in the expressionless face to substitute missing data. Hence, exact information about the examined expression is lost. To diminish this loss, we exploit a higher level "emotional grammar" of facial expressions defined by Ekman [5]. The main idea is that there is a higher possibility that a smile is coupled with "smiling" eyes than with expressionless eyes.

The system's post-processor utilizes an existing CLIPS-implemented expert system, HERCULES, to classify the observed facial expression into the six basic emotion categories. Since HERCULES has been presented elsewhere [13], just a short description of its processing is provided here. The attention is paid on integration and actual employment of HERCULES within the system for automated facial action encoding.

HERCULES accepts an AU-coded description of the encountered expression and converts this into a set of emotion labels. The rules for emotional classification of the facial actions are straightforwardly acquired from the linguistic descriptions of the prototypic facial expressions given by Ekman [5]. Five certified FACS coders have validated these rules using a set of 129 dual view images representing the relevant combinations of AUs. In 85% of the cases, the human observer and the system evenly labeled the observed expression [13].

HERCULES returns a set of quantified emotion labels. An emotion label is quantified according to the assumption that each AU, forming a part of a certain basic expression, has an equal influence on that expression's intensity.

Input to the Post-Processor consists of the expression geometry delimited by the Facial Data Evaluator and the quantified AU-codes determined by the Facial Data Analyzer. The geometry of the current expression is checked for presence of an expressionless facial feature. A simple control of the assigned CFs performs this check. A CF equal to *1/2N* is assigned to a facial feature only if the pertinent feature detected in the expressionless face has substituted the feature. If there is a feature having CF equal to *1/2N*, HERCULES is invoked. Otherwise, the system's processing terminates and displays the result – the quantified AU-codes and the certainty of these conclusions.

If HERCULES is invoked, this result is adjusted upon the acquired emotional classification of the analyzed expression. The returned list of emotion labels is searched and a kind of backward reasoning of HERCULES' inference engine is performed for the emotion label with the highest weight and the facial feature marked as expressionless. The rules given in Table 2 are used to reason about the possible deformation of the marked facial feature whereupon the system's final result is then adjusted.

Table 2. The rules for determining the appearance of the missing facial feature (i.e. the appropriate AU code) based on emotional classification of the encountered expression

	Eyes	Eyebrows	Mouth
Sadness	7 if 1	1	15
Fear	5+7	1 if 5	20
Happiness	6	-	12
Surprise	5	1+2	26
Disgust	9	9	9
Anger	7	4	24

In order to quantify appropriately the newly added AU, the AU-codes comprising the analyzed expression are compared to the AU-codes comprising the prototypic expression, which characterizes the emotion category to which the analyzed expression has been classified. The AU-codes that belong to both are marked and their average intensity is assigned to the newly added AU. The CF assigned to this AU is obtained as given in formula (5).

$$CF = \frac{1}{2} * min\{CFs\_marked\_AU\text{-}codes\} \qquad (5)$$

System Development and Evaluation

The system is developed according to the Incremental Development model. This model is characterized by integrated prototyping where the design phases - coding, integration and implementation - are split in successive increments of functionality. The successive increments, covering the full breadth of the system in an easy-to-integrate way, were selected according to the main parts of the system: Facial Data Generator, Data Evaluator, Data Analyzer and Post-Processor. Each part has been developed independently and then integrated into the operational and tested prototype presented in this paper. Chronologically, the Facial Data Generator and the Post-Processor have been developed in parallel and before the other parts of the system.

Since the system is to be used on different software platforms for purposes of behavioral science research as well as a part of human-computer interface, robustness, user-friendliness and portability were the requirements posed on the development. Integrating multiple detectors into a single workbench for facial expression information extraction and applying the reasoning with uncertainty on the extracted data insure robustness and precision of the system. JDK and JNI made Java a proper tool for fulfilling all other constraints posed on the development.

The operational prototype presented here has not been deployed in a real-world environment. The aim is to develop a robust, fully operational, intelligent multi-modal/media human-computer interface which will perform encoding and interpreting of all human communicative signals, namely, speech, facial expressions, body movements, vocal and physiological reactions. Still, if regarded merely in the scope of human-behavior-interpretation application domain, the prototype has been evaluated by the end-users since eight certified FACS coders have performed the validation studies on the prototype. Validation studies addressed the question whether the interpretations acquired by the system are acceptable to human experts judging the same images.

Testing Images and Testing Subjects

The overall performance of the system's prototype has been evaluated on a database containing 1040 dual views (see Figure 1 for a testing image example). Eight certified FACS coders participated in building of this database. Subjects were of both sexes and ranged in age (22-34) and ethnicity (European, South American and Asian).

The database of testing images contains the dual views of each subject displaying 2x30 expressions of separate AU activation, 4 maximal displays of AU8, AU18, AU39 and AU41, 2x6 basic emotional expressions, a neutral expression and 53 expressions representing combinations of AU activation. The images have been recorded under constant illumination using fixed light sources attached next to the mounted cameras and none of the subjects had a moustache, a beard or wear glasses.

Facial Action Encoding Performance

Two certified FACS coders validated the rules for AU coding by evaluating 90 expressions of separate AU activation displayed by other three coders. In 100% of the cases the image representing the activation of a certain AU, produced according to our rules (Table 1), has been labeled with the same AU-code by the coders. This result has been expected, however, since all of the rules have been acquired from FACS in a straightforward manner.

The facial action coding achieved by the system was 89.6% (i.e. 90% for the upper face AUs, 85% for the lower face AUs and 94% for the AUs combinations) when compared to human coding of all images in the database.

Facial Action Codes Quantification Performance

In order to compare quantification of the AU-codes done by our system with that done by humans, we collected the data from a questionnaire. For each image from the database shown by a certain subject, we asked the other seven subjects to assign an *individual index of intensity impression* to each of the activated AU(s) displayed in the image. While determining the indexes for the images of an observed subject, the coders used that persons' individual extreme-displays (IED) set. Finally, for each image in the database, an average index of intensity impression has been calculated.

For each of the eight subjects, his/her IED-set was also used to set the initial values in the database of extreme model deformations. The rest of his/her dual views have been used to evaluate the performance of the system by comparing the system's result and the average index of intensity impression related with a relevant image. Then the results for a total of 952 testing images have been averaged. The average disagreement between the AU intensity assigned by the system and the relevant average index of intensity impression was 0.08 (i.e. 8%), respectively 0.16 (i.e. 16%), in the case of the correctly recognized AU with a CF >= 0.3, respectively CF < 0.3. Disagreements were mostly caused by "inaccuracy" of the human eye when comparing the currently observed facial deviation with a relevant deviation shown in the images of the observed subject's IED-set.

Conclusion

The system presented in this paper brings together three fundamentally diverse technologies: psychologically and anatomically founded FACS [4], image analysis and AI. The system encodes and quantifies 30 different facial actions from static dual-view facial images.

By a large number of experiments, a confident system performance measurement is obtained that indicates rather robust and accurate facial action coding that the system accomplishes. When tested on 1040 dual-view images, facial action correct recognition rate achieved by the system was 89.6%. Average disagreement between the facial action intensity calculated by the system and that assigned by human experts was 0.08 (i.e. 8%), respectively 0.16 (i.e. 16%), in the case of the correctly recognized facial action with a CF >= 0.3, respectively CF < 0.3.

In comparison to the existing explicit attempts to automate facial action coding [2], [1], [6], the system presented in this paper is new and fundamentally different by the use of AI technology. Also it deals with automatic facial action coding in a more effective way. The best of the existing similar systems [2] performs recognition of 15 different facial actions. None of the existing similar systems quantifies the facial action codes. Our system performs accurate fully automatic coding and quantification of 30 different facial actions in static facial dual-views.

There are a number of ways in which the presented system could be improved. First of all, the system cannot encode the full range of facial behavior. From a total of 44 AUs defined in FACS, the presented prototype can encode 30 AUs from a dual-view image of encountered facial expression. The facial feature detectors integrated into the system are far from perfect and have not been proved capable of detecting all facial changes underlying a full range of facial behavior. The facial motions should be modeled and real-time spatio-temporal detectors of facial movement should be integrated into the system to allow tracking of fast facial actions such as wink, blink and wiping of the lips. Modeling the facial motion will also allow analysis of facial expression dynamics, which seems to be crucial in expression analysis.

Another limitation of the presented prototype is evident in a time-consuming performance. While the execution of the reasoning process takes some 3-4 seconds, complete processing of a single image takes 3 minutes in average due to the time-consuming image processing. Real-time image analysis would need to be achieved if the system is to be used as a part of a realistic man-machine interface.

We are not aware of any system, including our own, which perfects automatic facial action coding either in photographs or in video sequences. We still seek and investigate the possibilities.

References

[1] Black, M.J.; Yacoob, Y. 1998. Recognizing facial expressions in image sequences using local parameterized models of image motion. *International Journal on Computer Vision* 25(1): 23-48.

[2] Cohn, J.F.; Zlochower, A.J; Lien, J.J.; Kanade, T. 1998. Feature-point tracking by optical flow discriminates subtle differences in facial expression. *Proc. IEEE FG*, 396-401.

[3] Donato, G.; Bartlett, M.S.; Hager, J.C.; Ekman, P.; Sejnowski, T.J. 1999. Classifying Facial Actions. *TPAMI* 21(10): 974-989.

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

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

[6] Essa, I.; Pentland, A. 1997. Coding, analysis, interpretation and recognition of facial expressions. *TPAMI* 19(7): 757-763.

[7] Hong, H.; Neven, H.; von der Malsburg, C. 1998. Online facial expression recognition based on personalized galleries. *Proc. IEEE FG*, 354-359.

[8] Huang, C.L.; Huang, Y.M. 1997. Facial expression recognition using model-based feature extraction and action parameters classification. *Journal of Visual Communication and Image Representation* 8(3): 278-290.

[9] Kass, M.; Witkin A.; Terzopoulos, D. 1987. Snake: Active Contour Model. *Proc. IEEE ICCV*, 259-269.

[10] Kearney G.D.; McKenzie, S. 1993. Machine interpretation of emotion: design of JANUS. *Cognitive Science* 17(4): 589-622.

[11] Mehrabian, A. 1968. Communication without words. *Psychology Today* 2(4):53-56.

[12] Otsuka, T.; Ohya, J. 1998. Spotting segments displaying facial expression from image sequences using HMM. *Proc. IEEE FG*, 442-447.

[13] Pantic, M.; Rothkrantz, L.J.M. 1999. An expert system for multiple emotional classification of facial expressions. *Proc. IEEE ICTAI*, 113-120.

[14] Rothkrantz, L.J.M.; van Schouwen, M.R.; Ververs, F.; Vollering, J.C.M. 1998. A multimedia workbench for facial expression analysis. *Proc. Euromedia*, 94-101. Ghent: SCS Press.

[15] Schneider, M.; Kandel, A.; Langholz, G.; Chew, G. 1996. *Fuzzy Expert System Tools*. Chichester: John Wiley & Sons Ltd.

[16] Vincent, J.M.; Myers, D.J.; Hutchinson, R.A. 1992. Image feature location in multi-resolution images. *Neural Networks for Speech, Vision and Natural Language*, 13-29. Chapman & Hall.

[17] Wang, M.; Iwai, Y.; Yachida, M. 1998. Expression recognition from time-sequential facial images by use of expression change model. *Proc. IEEE FG*, 324-329.

[18] Wojdel, J.C.; Wojdel, A.; Rothkrantz, L.J.M. 1999. Analysis of facial expressions based on silhouettes. *Proc. of ASCI*, 199-206. Delft, NL: ASCI Press.

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

[20] Zhang, Z.; Lyons, M.; Schuster, M.; Akamatsu, S. 1998. Comparison between geometry-based and Gabor wavelets-based facial expression recognition using multi-layer perceptron. *Proc. IEEE FG*, 454-459.