Automated Facial Expression Analysis

M. Pantic & L.J.M. Rothkrantz

Delft University of Technology
Faculty of Information Technology and Systems (ITS),
Zuidplantsoen 4, 2628 BZ Delft, the Netherlands
{M.Pantic,L.J.M.Rothkrantz}@cs.tudelft.nl.

Keywords: Intelligent Analysis, Interpretation of MM data, Motion Capture, Java

Abstract

Human Emotion Recognition Clips Utilised Expert System was designed to recognise and interpret facial expressions of the observed person in an automatic way [1]. Still, input to HERCULES has been manually supplied. Through integrating HERCULES into the *Integrated System for Facial Expression Recognition* (*ISFER)* a complete process of automatic analysis of facial expressions has been achieved [2]. ISFER forms a part of the project *Automated System for Non-verbal Communication* [3] which is an ongoing project at the Knowledge Based Systems department of Delft University of Technology. The theoretical formulation of the implemented facial expression recognition has been acquired from FACS [4]. The interpretation of the recognised facial expression is currently based [1][7] on the recognition of so-called six basic emotions defined by Ekman [5][6]: happiness, sadness, fear, surprise, disgust and anger. Validation of the implemented knowledge and performing of the facial expression analysis in a completely automated way, are the main topics of this paper.

1. Introduction

Human communication has two main aspects: verbal and non-verbal. As the words are the atomic information units of the verbal communication, so the phenomena like facial expressions, body movements and physiological reactions are the atomic units of the non-verbal communication. Non-verbal communication forms guidance, amplification or even a replacement of the verbal communication. This fact triggered our interest in automation of non-verbal communication.

The goal of the project Automated System for Non-verbal Communication is the development of an automated system that has to provide qualitative and quantitative information for assessment of mood and intentions of the observed communicator. The system has to cope with the registration, processing and interpretation of the non-verbal communication signals. The input to the system consists of sound-, image-, and sensor data. This multimodal input is further analysed in a parallel way through performing of voice-, facial expression-, body movements-, and physiological appearances analysis [3]. Each analysis has to be implemented as an independent module that can operate as a separate unit as well as an integral part of the system. The final output of the system would be a data-fusion of the results of each separate module, representing a hypothesis about the mood and the intentions of the observed person.

ISFER has to perform automatic facial expression analysis of the system described above. In other words, it has to perform:

1. automatic facial features extraction,
2. reasoning about shown facial expression and eventual interpretation of it in terms of emotions,
3. appropriate presentation of the acquired results.

ISFER has and will have quite a broad area of application [7].

- Training in enlarging of the communicating skills, important for desk-staff in any branch.
- Training in emotion and stress control, important for any job where stressful situations are daily present (e.g. pilots, policemen, nuclear power-plant operator).
- Monitoring hazardous work places and “on-line” assessing the stress of the “key” employee (e.g. pilot).
- Enhancing automated information systems and multi-modal communication systems with a feedback information about the mood and intentions of the user.

2. Knowledge acquisition

To achieve automated facial expression analysis it is necessary to implement the knowledge on facial expression recognition and interpretation [1][7]. It is quite difficult to acquire this kind of knowledge from a living expert that has only a slight idea about the ways of information technology. On the other hand, it was possible to find a lot of information about this topic in widely acknowledged psychological studies. Therefore we chose to acquire the necessary knowledge from the written sources. We chose the most prominent and the most commonly used study on facial expressions in the last fifteen years written by Paul Ekman.

Facial expressions represent a visible consequence of the facial muscles’ contractions. Therefore, if we can detect which muscles have been contracted we can recognise the shown facial expression.

Detection of the facial muscles' contractions can be performed with Facial Action Coding System [4]. The activation of the muscles is described by the system as the activation of 44 Action Units. Each AU corresponds to the contraction produced by one or a group of related muscles. Activation of an AU is described in terms of changes of the facial appearance (i.e. changes of the facial features such as eyes, eyebrows and mouth) caused by the contraction of the underlying muscle(s). By tracing the contours of the facial features, changes in the appearance of the features will be traced. This will reveal the activation of the facial muscles and therefore the shown facial expression. This knowledge forms the basis of the facial expression recognition engine of ISFER [1][7].

In the rest of the text we use the term "contours of the facial features" to indicate the approximated curves of the actual contours of the facial features. The modules of the ISFER workbench do tracing of the curves and their approximation. An approximated curve represents a deformable template cast around the extracted snake or a connection of the extracted features (e.g. micro-features) [2].

Currently performed process of the emotion recognition is based on the definition of six basic emotions [5][6]. Ekman describes each basic emotion as a facial expression characteristic for that emotion. Together with the knowledge on facial expression recognition, the knowledge about the expression characteristic for each basic emotion forms a complete reasoning mechanism of ISFER.

3. Knowledge Modelling

The goal posed on a further development of the system for automated analysis of facial expressions was a suitable modelling of the acquired knowledge so that the posed task can be performed. We built a model of facial points that fulfils a threefold of requirements.

1. The points defined by the model must reveal changes in the appearance of the facial features such as eyes, eyebrows and mouth.
2. It must be possible to establish a simple and unique relation between the changes in the points' position and the activation of separate AUs.
3. It must be possible to extract the points from the automatically traced contours of the facial features such as eyes, eyebrows and mouth.

The model is illustrated in *Fig. 1* and the facial points are described in *Table 1*. A detail description of the model can be found in [1] and [7].

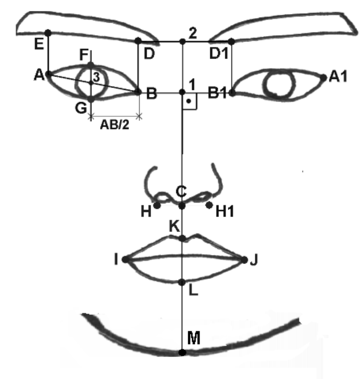

Fig. 1: Model

4. Knowledge Implementation

Though our model can't describe all of 44 AUs defined in FACS, it describes 32 AUs. By observing more than 500 photos of the people photographed during their anticipation in some discussion we came to the conclusion that 12 AUs, which cannot be

Point	Facial position	Stability & Methods of Automatic Tracing
B	right eye inner corner	Stable point; tracing using the micro-features of the eye traced with a NN [3]
B1	left eye inner corner	Stable point; tracing using the micro-features of the eye traced with a NN [3]
A	right eye outer corner	Stable point; tracing using the micro-features of the eye traced with a NN [3]
A1	left eye outer corner	Stable point; tracing using the micro-features of the eye traced with a NN [3]
H	right nostril inner corner	Stable point; tracing by using colour analysis and template matching[3]
H1	left nostril inner corner	Stable point; tracing by using colour analysis and template matching [3]
C	middle point between nostrils	calculated as ½d(HH1) from the traced contour of the nostrils[3]
D	on the right brow above B	Non-stable point; tracing by using template matching [3]
D1	on the left brow above B1	Non-stable point; tracing by using template matching [3]
E	on the right brow above A	Non-stable point; tracing by using template matching [3]
E1	on the left brow above A1	Non-stable point; tracing by using template matching [3]
F	top of the right eye ellipse	Non-stable point; tracing using the micro-features of the eye traced with a NN [3]
F1	top of the left eye ellipse	Non-stable point; tracing using the micro-features of the eye traced with a NN [3]
G	bottom of the right eye ellipse	Non-stable point; tracing using the micro-features of the eye traced with a NN [3]
G1	bottom of the left eye ellipse	Non-stable point; tracing using the micro-features of the eye traced with a NN [3]
K	top of the upper lip	Non-stable point; tracing by applying colour analysis or template matching [3]
L	bottom of the lower lip	Non-stable point; tracing by applying colour analysis or template matching [3]
I	right corner of the mouth	Non-stable point; tracing by applying colour analysis or template matching [3]
J	left corner of the mouth	Non-stable point; tracing by applying colour analysis or template matching [3]
M	bottom of the chin parabola	Non-stable point; tracing by using chin parabola template matching [3]

Table 1: Facial points defined by our model

recognised by the system, seldom occur. Such conclusion is also logical when one knows that those unrecognisable AUs are "Tongue Show", "Jaw Sideways", "Bulge on the cheeks produced by the pushed tongue", etc.

As an example, we can consider the description of AU6, AU12 and AU25 with our model. *Fig. 2* illustrates the facial expression caused by the activation of AU6 + AU12 + AU25, and *Table 2* describes the rules of the system's facial expression recognition engine. The facial expression interpretation engine will conclude that the shown emotion is happiness by comparing the definition of happiness given by Ekman and the facial expression illustrated below.

Fig.2: Default expression & AU6+AU12+AU25

5. Software Platform

Since 1992, a research regarding automated facial expression analysis is an ongoing project at the Knowledge Based Systems department. Different software solutions for different parts of facial analysis process were developed in the past few years. Therefore, the requirement posed on the development of ISFER platform was the integration of all the existing modules. The result of such integration should be an interactive user-friendly platform for an easy performing of automated facial expression analysis [3]. Availability of JDK and JNI made Java perfectly suitable for development of such a kind of software platform [8].

ISFER has been organised in three major layers according to three major phases of facial expression analysis – image data pre-processing, analysis and interpretation. Each layer contains a cluster of modules, which perform different parts of the appropriate phase of automatic facial expression analysis.

The modules for generating digital full-face images, for feeding the system with this information and for filtering the image data are placed in the pre-processing layer of ISFER. The modules that extract a particular facial feature such as eye or mouth from a full-face image and the modules that trace the contours of those facial features, belong to the data analysis layer of ISFER. HERCULES performs the facial expression recognition and interpretation and belongs to the interpretation layer of the workbench. The structure of ISFER can be illustrated as shown in *Fig. 3*.

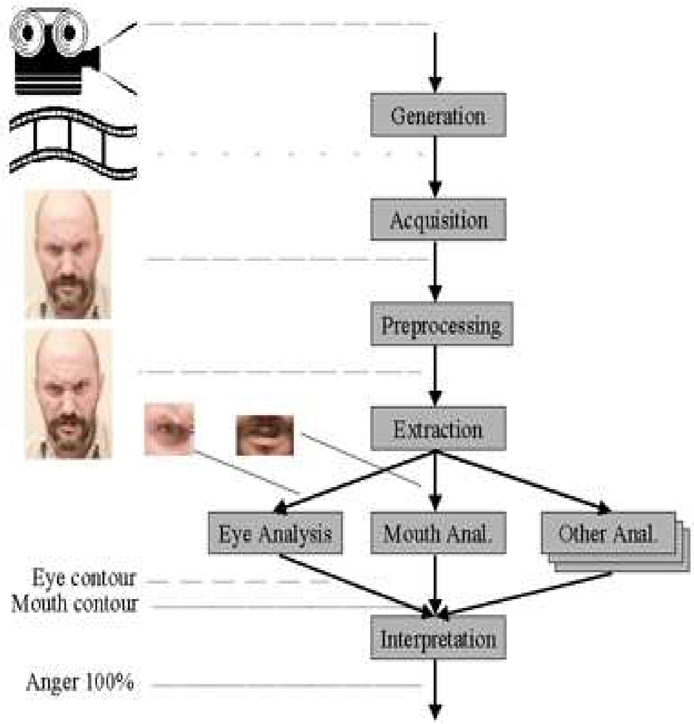

Fig.3: Algorithmic representation of ISFER structure

Each module of the system has been integrated as an independent unit into the platform. Still, GUI of the system has been built so that an arbitrary number of modules belonging to the arbitrary layers of the system can be selected and then connected to form a so-called network of modules that will perform the desired analysing task. The modules are related through data on which they operate. Two modules can be connected in a network when the output of one module forms the input to the other module. Each time when a connection is made, it will be checked if the data types of the modules match. Only if the data types match, the connection between the modules will be allowed.

At any moment the current network is displayed to the user in a form of a directed graph where the nodes of the graph depict the modules and the branches of the

AU	Ekman's description	Mapped on HERCULES' Model	Rule Name
AU6	cheeks raised, wrinkled outer corners of the eyes	the number of black pixels in the outer half (right from the line AE) of the circle with r(½d(AB)), C(A) becomes greater; the number of black pixels in the outer half (left from the line A1E1) of the circle with r(½d(A1B1)), C(A1) becomes greater.	rAU6
AU12	mouth corners pulled up	d(IB) becomes smaller; d(JB1) becomes smaller; d(CI) becomes greater; d(CJ) becomes greater.	rAU12
AU25	lips parted	d(KL) becomes greater; d(CM) does not become greater.	rAU25

Table 2: Implementation of FACS - HERCULES' facial expression recognition

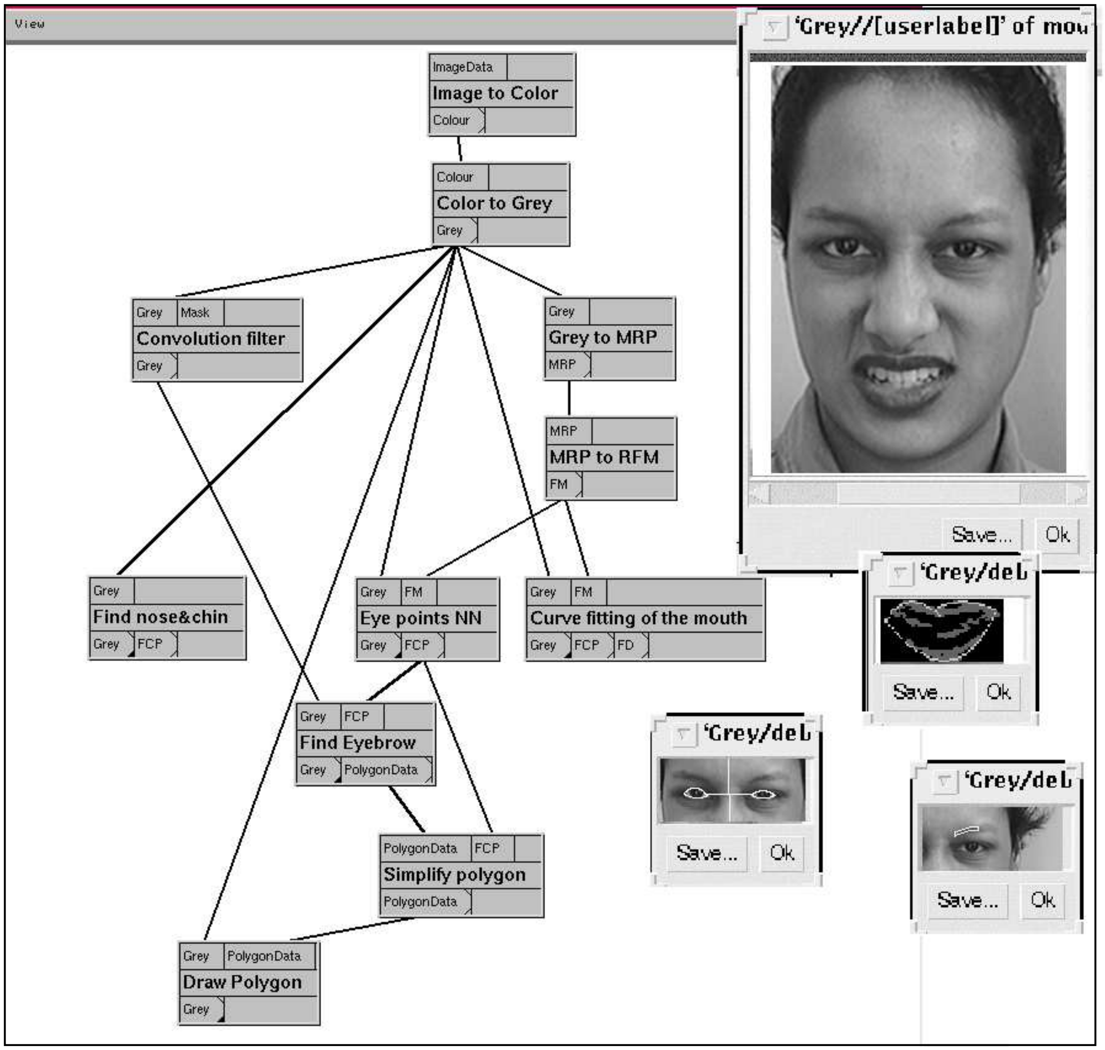

Fig.4: Screen shot of ISFER workbench

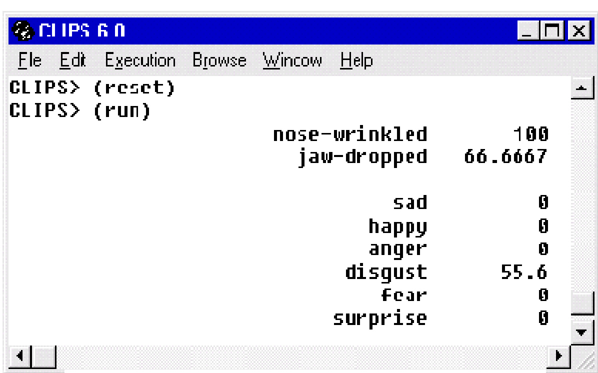

Fig. 5: Screen shot of HERCULES

graph represent the connections between the modules. An example of a network that results in displayed traced contours of the facial features is given in *Fig.4*.

The points of the traced contours will be saved in a file that is handed to the recognition and interpretation part of ISFER embodied in HERCULES. Based on the passed data about the traced contours, the points of the model will be extracted. These points form the input to the inference engine of HERCULES, who's processing results in a description of the shown facial expression and its emotional interpretation [1][7]. In the case of the facial expression illustrated in *Fig. 4* the output of HERCULES will be the description of that expression, given as the description of AU9 + AU27 activation, and the interpretation of that expression given as a percentage of disgust (see *Fig. 5*). One certainly do understand that the output of the system given in *Fig. 5* results from the reasoning strictly based on the knowledge extracted from Ekman [4][5]. So, one can interpret the facial expression illustrated in *Fig. 4* differently than the system does. To facilitate the system with the possibility of labelling the expressions with the user defined labels a learning facility will be built. However, this issue isn't within the scope of this paper.

6. Automated Facial Expression Analysis

Fig. 4 and *Fig. 5* illustrate the whole trajectory of automatic facial expression analysis. It should be stressed that for each facial feature there are several modules integrated into the ISFER platform that can be used to trace that feature in the examined image. However we made the illustrated choice deliberately because the chosen modules had the best testing results.

The complete trajectory of automated facial expression analysis consists of a tenfold of program [2].

1. Image to Color – conversion of Java *Image* data to a flat array of pixels (integrated in the pre-processing workbench layer).
2. Color to Grey – conversion of the colour picture to a grey picture (integrated in the pre-processing workbench layer).
3. Convolution Filter – improving the picture quality using the convolution filter (C module integrated in the pre-processing workbench layer).
4. Grey to MRP – converting the grey-scale picture to the Multi Resolution Pyramid format [2] (C module integrated in pre-processing workbench layer).
5. MRP to RFM – generating the Raw Feature Map, which represents a global approximation of head, left and right eye, mouth and nose (C module integrated in the data analysis workbench layer).
6. Find Nose & Chin – tracing the contours of the nostrils and approximating the chin parabola (C module integrated in the data analysis workbench layer).
7. Eye Points NN – locating the contours of the eyes using a neural network (C module integrated in the data analysis workbench layer).
8. Curve fitting of the mouth – fitting a single second-degree curve around the lower lip and two second-degree curves around the upper lip (C module integrated in the data analysis workbench layer).
9. Find Eyebrow, Simplify Polygon, Draw Polygon – based on the estimated points of the eye, the eyebrow is located and then approximated with two simplified second-degree curves (C modules integrated in data analysis workbench layer).
10. HERCULES – a rule based expert system for recognition and interpretation of facial expression [1] [7]. HERCULES is still not physically integrated into the workbench. The connection with the workbench modules is accomplished through data transfer in form of a file containing the traced contours of the facial features. The facial features defined by the model are then extracted from the contours and the reasoning is performed. The result of the system is a description of the shown facial expression (a description of the activated AUs) and an emotional interpretation represented in the form of the calculated intensities of six basic emotions (see *Fig 5*).

7. Knowledge Verification and Validation

During the process of verification and validation of the model, 20 human observers have been asked to detect manually the position of the points defined by model in the full-face images. The aim of this manual measurement of the points was determination of the error that a human observer will make when measuring. The reasoning that has been followed concerns the extent of the found error. If the error made by manual measuring of a particular point extends the extreme possible deviation in the position of that point then the model isn't reliable neither for the human observers nor for an automatic extraction.

Each point of the model has a certain degree of freedom that ranges from 0 to 2 [1]. For example point I, representing the right corner of the mouth, has the degree of freedom 2 because it can change its facial position, under the facial muscles contraction, in both directions, x and y. Point K, representing the top of the lips, has, on the other hand, the degree of freedom 1 because its x co-ordinate is set to be equal to the x co-ordinate of point C. The standard deviation of the x- and/or y co-ordinate (measured in pixels) of each of the points defined by the model having a degree of freedom greater than 0 is given in *Table 3*. In opposition to all expectations, the found standard deviation is a close neighbourhood of one pixel in the case of all points defined by the model except for the points representing the inner corners of the eyes. In other words, the points defined by HERCULES' model are rather accurately traceable.

Point	σ	**Point**	σ
x_B	0.7399	y_E	0.5099
y_B	0.7071	y_{E1}	0.5568
x_{B1}	0.9734	y_F	0.7746
y_{B1}	0.6000	y_{F1}	0.7921
x_A	1.2359	y_G	0.6982
y_A	0.4899	y_{G1}	0.6000
x_{A1}	1.9046	y_K	0.3162
y_{A1}	1.3000	y_L	0.7483
x_H	0.8292	x_I	0.5831
y_H	0.4000	y_I	0.4975
x_{H1}	0.9695	x_J	0.7263
y_{H1}	0.9206	y_J	0.5362
y_D	0.9734	y_M	0.7399
y_{D1}	0.7921		

Table 3: Standard deviation of the manual measurement

For two different photos of the same person the stable points B and B1 remain stable and will be measured by different human observers with a divergence of the average error given for the particular point in *Table 3*. Further, for each non-stable point the standard deviation of the measurement of that point is much smaller than the extreme possible deviation in the position of that point. Considering these facts it could be concluded that the model is reliable in the case of a manual measuring of the points defined by the model.

As already explained, the points defined by the model aren't extracted directly from a digital image. First the contours of the facial features such as eyes, eyebrows and mouth are automatically extracted and from those traced contours the points defined by the model are further extracted. Supposing that all of the modules for automated tracing of the facial features are accurate enough the reliability of the model, in the case of the automatically performed measuring, primarily depends on the possibility of extracting the points defined by the model from the traced contours of the facial features. All of the points defined by the model can be automatically extracted from already traced contours of the facial features. As an illustration of the extraction process of the points defined by the model, extraction of the points of the mouth from an automatically traced mouth contour (as illustrated in *Fig. 6*) is described in *Table 4*.

Fig. 6: Automated tracing of the mouth contour

Point	Part of the traced contour
I	Most right point of the contour
J	Most left point of the contour
K	Upper intersection point of the vertical axis of the face and the traced contour
L	Lower intersection point of the vertical axis of the face and the traced contour

Table 4: Extraction of the mouth points

Unfortunately, the modules for automated tracing of the facial features are not accurate enough to say that the reliability of the model depends only on the possibility of extracting the points from the traced contours. A module for dealing with highly inaccurate and missing data must be built and integrated between the ISFER workbench and the system's reasoning mechanism. Some issues hereabout are given in the next paragraph.

The reasoning of the system has been tested on the manually supplied input to the system' reasoning mechanism (HERCULES). The input contained the sets of the measurements manually made on:

- 125 photographs representing the facial expressions caused by the activation of separate AUs,
- 30 photographs representing the facial expressions of *pure* emotions ([5] [6]), and
- 125 photographs representing *blended* emotions ([6]).

The results of the tests confirm that the changes in the position of the points defined by the model uniquely reveal the activation of Action Units, that is, the shown facial expression. In fact, for the same input (i.e. the same set of changes in the position of the points defined by the model) the same result has been always generated, that is the same description of the shown facial expression. Further, the reasoning mechanism of the system is completely based on, and acquired from FACS [4]. When the system is applied on the same set of pictures, the generated results always match the results acquired by applying the FACS' rules. All of these facts led to the conclusion that the model is valid in the case of the manual measuring of the points defined by the model.

The same inference engine has been used in the manual input version as well as in the automatic input version of the system [1][7]. This means that the reasoning of the system does not depend on the way in which the input is supplied to the system. Therefore, the validity of the knowledge implemented in the recognition and interpretation part of ISFER is ensured in the case of both, the manual and the automatic input version of the system.

8. Future Developments

Unfortunately, all of the workbench modules for automatic tracing of the contours of the facial features have not been made robust enough yet. For example, all of the modules are prone to changes of the light conditions. Therefore it cannot be expected that the contours of the facial features will always be accurately traced as illustrated in *Fig. 6*. The inaccurate or missing data must be expected (see *Fig. 7*).

Fig. 7: Example of inaccurately traced right eye and failed tracing of the mouth contour

Nevertheless, the current version of the system's reasoning mechanism, embodied in HERCULES, assumes that the provided input to the inference engine will contain no inaccurate, missing or contradictory data. This means that the system will generate a correct result about shown facial expression only for the accurately traced contours of the facial features. Therefore, dealing with the inaccurate or missing data must be implemented in a future version of the system. This can be accomplished through implementation of the following.

- Knowledge about the facial proportions can be built into the system. This will reduce the effects of the inaccurately traced contours of the facial features. Further, this is a method of dealing with the missing data.

- The reasoning mechanism of the system embodied in HERCULES can be made as a Fuzzy Expert System. This will introduce the certainty of the generated results, which will depend on the certainty of the input data [9]. This will take the inaccuracy of the traced contours into account.

In order to build the knowledge about the facial proportions into the system, a new requirement must be posed on the observation process. Considering the fact that a 3D face is captured on a 2D image, the facial proportions are not retained if the head is rotated. Therefore, the camera must be set so that no head rotation can be encountered. If the head rotations can be encountered then an inaccurately traced contour of a facial feature can be interpreted as a consequence of a head rotation around y-axis (as illustrated in *Fig. 7*, for example). In other words, if the head rotations can be encountered there is no use of building the knowledge about the facial proportions into the system. This means that the camera must move together with the head of the observed person. In order to achieve this, the camera must be fixed on the head of the observed person in front of his/her face. Having a mounted camera above the microphone on a headphone-like holder, usually used by a pilot or a telephonist, will serve the purpose. Posing this requirement on the observation process will not reduce the applicability of the system in any case [7]. To the contrary, it will enhance the system performance because it will be possible to trace the face of the observed person at each moment, without a chance of having the face out of the camera view.

An assessment of the errors made when the facial points defined by the model are automatically extracted must be obtained. The assessment can be made in the same way the assessment of the errors made when the points of the model are manually measured has been obtained. This error assessment can be used further to make HERCULES a fuzzy expert system where the imprecision of the extracted facial features will be taken into account when the system's knowledge is processed.

A real-time processing of the system is very important for some of the above named areas of application. Also, processing of a sequence of images, in contrast to the processing of a single image currently implemented, must still be developed. But, as the mechanics of the system's reasoning have been set, a fully automated process of facial expression analysis has been achieved.

References

[1] Pantic, M. & Rothkrantz, L.J.M. (1997). *Automatic Recognition of Facial Expressions and Human Emotions*. Proceedings of ASCI'97 conference, ASCI, Delft, NL, ISBN 90-803086-2-5, pp. 196-202.

[2] Rothkrantz, L.J.M., van Schouwen, M., Ververs, F. & Vollering (1998). *A multimedial workbench for facial expression analysis*. Proceedings of EUROMEDIA'98, SCS International, Ghent, Belgium, ISBN 1-56555-140-0, pp. 94-101.

[3] van Vark, R.J., Rothkrantz, L.J.M. & Kerckhoffs, E.J.H. (1996). *Prototypes of Multimedial Stress Assessment*. Proceedings of MediaComm95, SCS International, Ghent, Belgium, ISBN 1-56555-046-3, pp. 108-112.

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

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

[6] Ekman, P. & Friesen, W.V. (1975). *Unmasking the Face.* Prentice Hall, New Jersey, USA.

[7] Pantic, M., Rothkrantz, L.J.M. & Koppelaar, H. (1998). *Automation of non-verbal communication of facial expressions*. Proceedings of EUROMEDIA'98, SCS International, Ghent, Belgium, ISBN 1-56555-140-0, pp. 86-93.

[8] Flanagan, D. (1996). *Java in a Nutshell*. O'Reilly & Associates Inc., Sebastopol, USA.

[9] Schneider, M., Kandel, A., Langholz, G. & Chew, G. (1996). *Fuzzy Expert System Tools*. John Wiley & Sons Ltd., Chichester, West Sussex, UK.