Automatic Recognition of Facial Expressions and Human Emotions

Maja Pantic & Leon Rothkrantz

Knowledge Based Systems Section, Faculty of Technical Mathematics and Informatics,
Delft University of Technology,
Julianalaan 132, 2628 BL Delft, The Netherlands

e-mail : maja@kgs.twi.tudelft.nl

Keywords: Expert System, Automatic tracing, Reasoning, Real-time, Parallelism.

Abstract

Human Emotion Recognition Clips Utilised Expert System -HERCULES [10] forms a part of an Automated System for Non-verbal Communication. It was designed to analyse a shown facial expression, based on psychological study called FACS [3], and to interpret it in terms of six basic emotions [2]. HERCULES now accepts manually made measurements on a full-face photograph and returns a description of the shown facial expression, quantitative measurement of it, its interpretation in terms of the six basic emotions and quantitative measurement of that emotional interpretation. Although the current version of HERCULES operates with a manual input, its kernel is built to accept facial measurements made automatically and processed in parallel. However, a fully automated real-time version of the system as well as dealing with inaccurate, redundant and faulty input is still under development.

1 Introduction

Human communication has two main aspects : verbal and non-verbal. In a dialogue the exchange of information does not take place only through words but also through facial expressions, the total body posture and the physiological reactions such as colouring of the face or sweating.

The goal of the project ***Automated Systems for Non-verbal Communication***, carried out by the Knowledge Based Systems group, is development of an automated system for non-verbal communication. The input to the system, consists of sound-, image-, and sensor data processed in parallel. Based on that input, voice-, facial expression-, body movements- and physiological appearances analysis have to be performed in parallel so that the output of the system represents a real-time qualitative and quantitative information about mood and intentions of the observed person [13].

The system will have quite a broad area of application. For example, it can be used for training in emotion and stress control, for training in enlarging the communicating skills, for enhancing multi-modal communication systems (e.g. videoconferencing), etc.

Human Emotion Recognition Clips Utilised Expert System (HERCULES) has to perform automatic facial expression analysis of the above described system. In other words, it has to perform:

1. automatic facial features extraction,
2. reasoning about shown facial expression in terms of emotions, and
3. displaying of the acquired results.

2 Theoretical Basis

One hundred years ago, Darwin wrote that facial expressions of emotion are universal, not learned differently in each culture, but biologically determined as the product of man's evolution. Since Darwin's time many writers have disagreed this statement. However, Ekman's psychological investigations [2] show that the facial appearances of at least six emotions - happiness, sadness, surprise, fear, anger and disgust - are indeed universal. Ekman calls those emotions the **six basic emotions** and he describes each of them [2] in terms of the facial expression characteristic for that emotion.

On the other side, in FACS [3] Ekman discusses ways of recognising facial expressions by human observer. Facial Action Coding System (FACS) is a system that can measure all possible visually distinguishable facial movements. The measurements units of FACS, so called **Action Units (AUs)**, are described in terms of facial muscle actions which are responsible for each and every facial movement. Therefore, with FACS, each facial expression shown by the observed person can be described in terms of activated AUs.

Considering the fact that each of the six basic emotions is described in terms of the facial expression characteristic for it, and that a facial expression is caused by facial muscular action, and that AUs are described in terms of facial muscle actions, it is quite obvious that if AUs could be automatically recognised, the facial expressions related to the emotions, which can be described in terms of activated AUs, can also be recognised.

Nevertheless, FACS isn't made for computers but for human observers - recognition of AUs' activation cannot be performed directly because recognition of the facial muscles contraction isn't detectable automatically. We need an indirect way of performing automatic recognition of AUs' activation. Such a way is automatic facial features extraction. By tracing the deviations in the facial features position, shape or intensity found between some default facial expression and the examined facial expression, which are caused by facial muscles contraction that in turn underlies the AUs' activation, the activation of AUs can be traced.

Once the activation of AUs have been revealed, the shown facial expression is also revealed (in terms of activated AUs) as well as the shown emotion (by classifying revealed facial expression against facial expressions characteristic for six basic emotions). This forms the main principle of HERCULES' reasoning.

3 The Model

There are several existing models describing facial features (e.g. FCP model [8] or Facial Landmarks model [7]) but, those are not suitable for an automatic facial features extraction. So, we built a new model based on the following two constraints :

1. Defined facial points from which the facial features will be derived must be traceable automatically.

2. It must be possible to establish a simple and unique relation between the deviations in facial features position, shape or intensity and the activation of separate AUs.

Our model is illustrated in *Figure 1* and the facial points are described in *Table 1*.

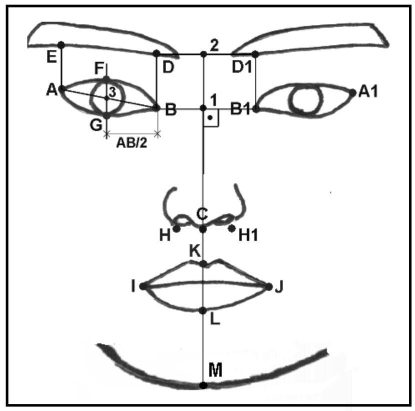


Fig. 1. : Facial Points Model

Point	Facial position	Method(s) used for Automatic Tracing
B	right eye inner corner	tracing: by tracing the **micro-features** of the eye with a NN [14] [5]
B1	left eye inner corner	tracing: by tracing the **micro-features** of the eye with a NN [14] [5]
A	right eye outer corner	tracing: by tracing the **micro-features** of the eye with a NN [14] [5]
A1	left eye outer corner	tracing: by tracing the **micro-features** of the eye with a NN [14] [5]
H	right nostril inner corner	tracing: by tracing the nostrils **template** [15] [12] [1]
H1	left nostril inner corner	tracing: by tracing the nostrils **template** [15] [12] [1]
C	middle point between nostrils	calculated as ½d(HH1) from the traced nostrils **template** [15] [12] [1]
D	on the right brow above B	tracing is still under development
D1	on the left brow above B1	tracing is still under development
E	on the right brow above A	tracing is still under development
E1	on the left brow above A1	tracing is still under development
F	top of the right eye ellipse	tracing: by tracing the **micro-features** of the eye with a NN [14] [5]
F1	top of the left eye ellipse	tracing: by tracing the **micro-features** of the eye with a NN [14] [5]
G	bottom of the right eye ellipse	tracing: by tracing the **micro-features** of the eye with a NN [14] [5]
G1	bottom of the left eye ellipse	tracing: by tracing the **micro-features** of the eye with a NN [14] [5]
K	top of the upper lip	tracing: by performing the **colour analysis** of the mouth [11] [9] or by applying the **active contours (snakes)** method [15] [12] [6]
L	bottom of the lower lip	tracing: by performing the **colour analysis** of the mouth [11] [9] or by applying the **active contours (snakes)** method [15] [12] [6]
I	right corner of the mouth	tracing: by performing the **colour analysis** of the mouth [11] [9] or by applying the **active contours (snakes)** method [15] [12] [6]
J	left corner of the mouth	tracing: by performing the **colour analysis** of the mouth [11] [9] or by applying the **active contours (snakes)** method [15] [12] [6]
M	bottom of the chin parabola	tracing: by tracing the chin parabola **template** [15] [12] [1]

Table 1. : Description of the Facial Points belonging to the Model

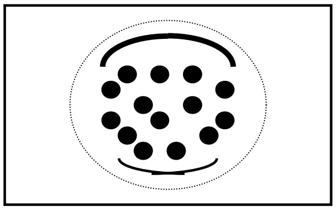
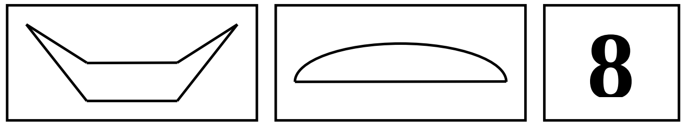


Fig. 2. : Chin pattern



(a) (b) (c)

Fig. 3. : Shapes of the mouth

AU	Ekman's description	Mapped on our Model	Rule
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. : An Example of Implementation of FACS - A part of HERCULES' facial expression recognition

Used facial features, derived from the above described facial points, are :

- Two angles ∠BAD and ∠B1A1D1, and sixteen distances (see *Figure 1*): d(AE), d(A1E1), d(3F) where point 3 is the midpoint of the line AB, d(4F1) where point 4 is the midpoint of the line A1B1, d(3G), d(4G1), d(FG), d(F1G1), d(CK), d(IB), d(JB1), d(CI), d(CJ), d(IJ), d(KL), and d(CM) where M is the tip of the chin.
- The intensity of black pixels in the seven facial areas (see *Figure 1*): in the upper half (above the line DD1) of the circle with r(½d(BB1)), C(2); in the lower half (below the line DD1) of the circle with r(½d(BB1)), C(2); in the outer half (right from the line AE) of the circle with r(½d(AB)), C(A); in the outer half (left from the line A1E1) of the circle with r(½d(A1B1)), C(A1); in the outer half (right from the line through point I, normal to the line BB1) of the circle with r(¼d(BB1)), C(I); in the outer half (left from the line through point J, normal to the line BB1) of the circle with r(¼d(BB1)), C(J); and the number (or existence) of white pixels on the line KL.
- Four facial shapes: the pattern of black pixels on the chin that becomes more visible if the chin is raised (see *Figure 2*); the shape of the lower lip which is present if the lower lip is depressed (see *Figure 3 (a)*); the shape of the mouth which is present if the lower lip is sucked into the mouth (see *Figure 3 (b)*); the 8-shape of the mouth which is present if the cheeks are sucked into the mouth (see *Figure 3 (c)*).

Neither the intensity of black pixels in the particular facial areas nor the existence of the above illustrated facial shapes can be traced automatically so far. Tracing of those features using pattern recognition, colour analysis and deformable templates is now under development.

4 Knowledge Implementation

In FACS, Ekman describes activation of each of 44 AUs as a specific change in facial appearance.

For 32 AUs, each of the facial appearance changes that characterise those AUs can be uniquely mapped on the set of deviations in position, shape or intensity of the facial features defined in our model. That mapping forms implementation of Ekman's knowledge and represents reasoning mechanism of facial expressions recognition part of HERCULES. The mapping between the six basic emotions and the set of activated AUs forms implemented Ekman's knowledge that represents inference engine of emotions recognition part of HERCULES.

As an example of the mapping between activation of an AU and the set of deviations in position, shape or intensity of the in our model defined facial features, we can consider AU6, AU12 and AU25 (see *Figure 4* for facial expression caused by the activation of AU6 + AU12 + AU25, and *Table 2* describing HERCULES' recognition of that facial expression - i.e. describing the rules that represent implemented Ekman's knowledge about AU6, AU12 and AU25).

The same example can be used for illustrating emotional interpreting of the shown facial expression. By classifying recognised facial expression against facial expressions characteristic for the six basic emotions (see *Figure 5* illustrating the mapping between happiness and the set of activated AUs that participate in forming the facial expression characteristic for that basic emotion), the facial expression illustrated in *Figure 4* will be interpreted as happiness (see also *Figure 7* illustrating HERCULES processing of example used here).

Fig. 4. : Default expression & "Happy" expression

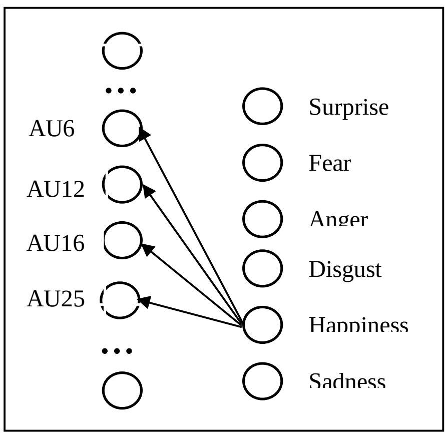


Fig. 5. : Mapping of happiness on AUs

5 System Structure

The input of the system is the set of deviations :

1. between facial distances measured in examined and in some default facial expression,

2. between number of black pixels traced in examined and in default facial expression, and

3. between existence of facial shapes in examined and in default facial expression.

The facial features have to be traced on the full-face image in an automatic way with the support of image-processing techniques and Neural Networking (currently the facial features are manually traced and measured). Based upon the registered deviations each facial expression is recognised and classified by the system in terms of activated AUs. The output of the first part of the system is a description of the shown facial expression (i.e. a description of each of the activated AUs) present on the examined intensity image, together with a quantitative measurement of that facial expression (i.e. together with the intensity of each of the activated AUs; currently either 0% or 100%).

The second part of HERCULES matches the set of activated AUs with the set of AUs characteristic for the six basic emotions. The output of this part of the system is an interpretation of the examined facial expression in terms of intensity of each of six basic emotions (based on the intensity of the activated AUs and on the mapping illustrated in *Figure 5*).

The first two parts of the system (i.e. HERCULES' kernel) consider processing of one intensity image - they deal with time dependent reasoning about the shown facial expression. The last part of the system (which isn't implemented yet) has to deal with relating the separate intensity images to each other as a sequence of intensity images in time - it has to deal with time independent reasoning about the shown facial expressions. The output of this part of the system has to be a graphical representation of the registered sequence of facial expressions as a function of time.

The structural representation of HERCULES as a whole can be represented as illustrated in *Figure 6*.

In **pre-processing** the facial features have to be extracted; the result of the extraction has to be controlled in order to avoid inaccuracy, incompleteness and/or redundancy; normalisation of extracted facial distances has to be performed; and the set of deviations (found by comparing the examined facial expression with some default facial expression) has to be calculated.

The set of deviations will form the input to HERCULES' **kernel** which, based on it, will give the interpretation of the currently examined facial expression in terms of :

- (time-dependent reasoning) description of the facial expression together with quantitative measurement of it, and
- (time-dependent reasoning) recognised (shown) emotion in terms of quantitative measurement (intensities) of the six basic emotions.

Post-processing of facial expressions involves time-independent reasoning aspects. That is :

- graphical representation of intensities of the six basic emotions as a function of time, and
- (optional) reasoning about emotional state (based on some "grammar" of emotions) of the observed person during the performed session.

6 System Implementation

HERCULES is an Expert System. The main reason for implementing facial expression- and human emotions recognition as an Expert System lies in the characteristics of that task.

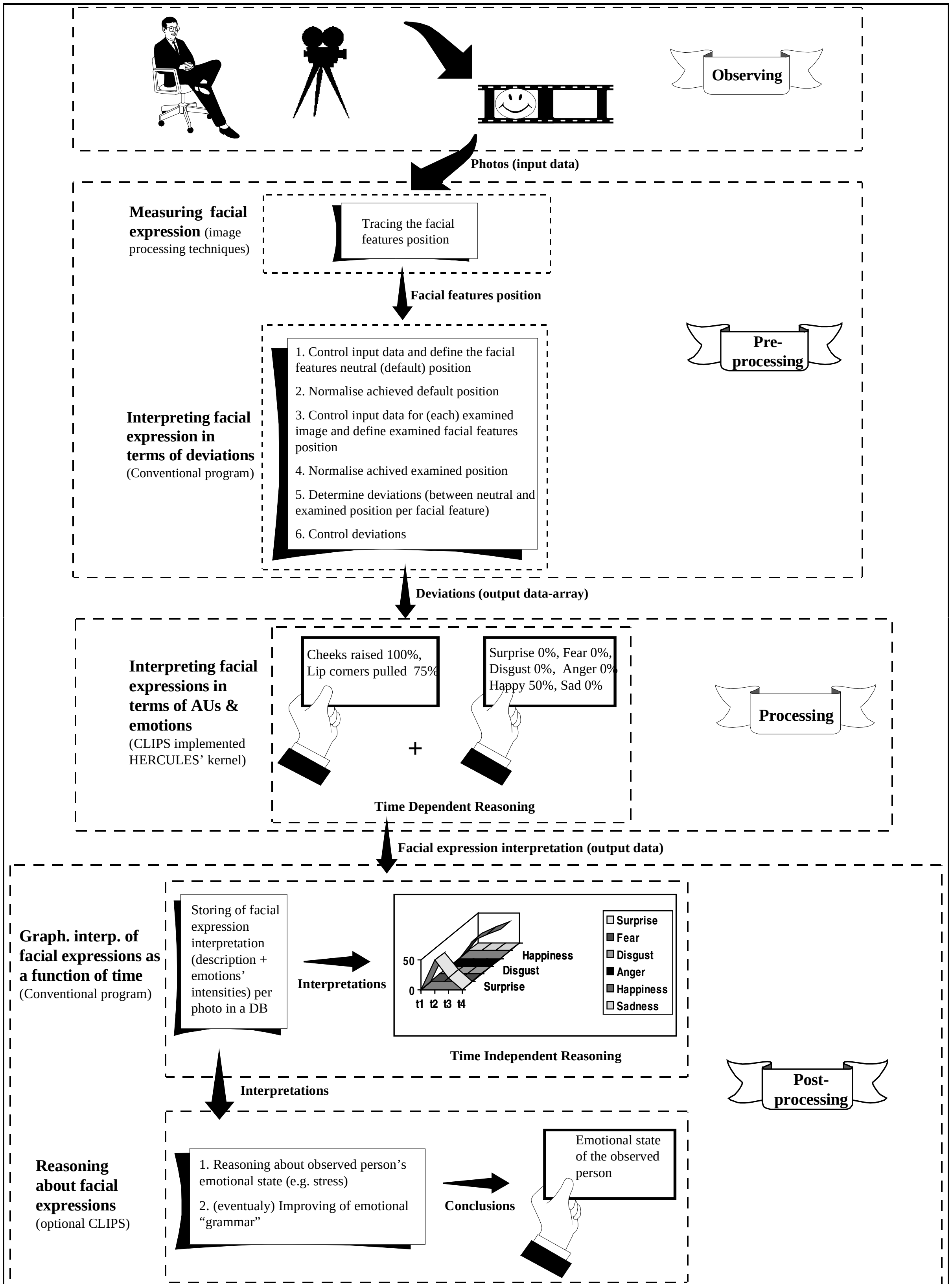


Fig. 6. : HERCULES

There is no algorithmic solution to it, it involves reasoning rather than numerical computation, the domain of the knowledge that has to be built in the system does not involve common sense or general-world knowledge.

Considering the global characteristics of a task which can be best implemented using an Expert System such as complexity of the task, requiring of reasoning about well defined problem domain rather than numerical computation and impossibility of solving it using a conventional computing method [4], it can be concluded that facial expressions- and human emotions recognition task is a sort of task which can be best implemented using an Expert System.

HERCULES is built as a Rule-Based Expert System. There are several reasons for this. The recognition of shown facial expression is done through recognition of the set of activated AUs. The activation of each particular AU is described by a set of condition clauses (see *Table 2*) - if the condition clauses are true, that AU is activated. The best way of implementing condition clauses is to implement them as rules. The LHS of a rule will contain condition clause(s), describing the activation conditions of a particular AU, while the RHS of the rule will state that the activation of that particular AU is present. Secondly, the rule-based expert systems have a so-called modular nature which makes possible creating of independent units that can in later stages still be integrated in the final system and that, in turn, makes expanding of the expert system fairly easy.

HERCULES' kernel (facial expressions- and emotions recognition) is implemented in CLIPS. CLIPS is an acronym for C Language Integrated Production System. CLIPS is a forward reasoning rule-based language, designed at NASA with the specific purpose of providing high portability, low cost and easy integration with external systems [4].

7 Testing

The system has been tested using the set of manually made measurements on :

- more than 50 photographs representing facial expressions caused by the activation of separate AUs,
- 10 photographs representing facial expressions of *pure* emotions (i.e. emotions expressed at maximal intensity [3]), and

(a) default fac. expression (b) examined fac. expression

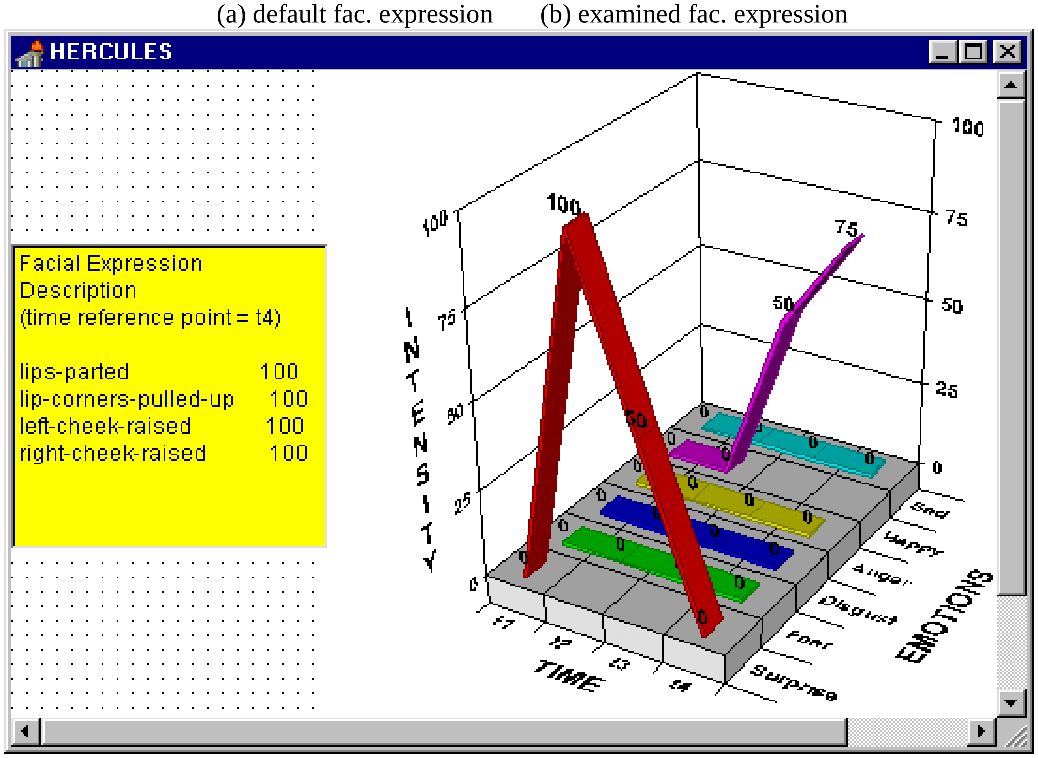


Fig. 7. : HERCULES' processing for the manually supplied input measured on photo (b) in comparison to photo (a)

- 50 photographs representing *blended* emotions (i.e. combinations of six basic emotions or, pure emotions not expressed at maximal intensity [3]).

An example of HERCULES' processing is given on *Figure 7*.

8 Current State-of-the-art

As already mentioned, goal placed on the development of HERCULES is threefold:

1. automatic facial features extraction,
2. reasoning about shown facial expression in terms of emotions, and
3. displaying of the acquired results.

Except for automatic facial features extraction, which is still under development, the goal has been achieved. For the same set of deviations in facial features position, shape or intensity the system will always give the same result - the same facial expression description and its emotional interpretation.

Although the system must still be improved (e.g. implementation of the pre-processing is still to be completed; making the system robust i.e. dealing with inaccurate, incomplete and redundant input is still to be done; the number of intensity levels that can be associated with an activated AU could be increased; the validity of the mapping illustrated partially in *Figure 5* has yet to be ensured; emotional "grammar" is still to be researched and the whole post-processing must be implemented; parallel execution of different algorithms for automatic facial features extraction as well as parallel processing of sequence of images has yet to be performed in order to come closer to the real-time performance of the system; parallel processing of sequence of images, in contrast to the processing of a single image currently implemented, as well as reasoning about such sequence of images must still be developed), the mechanics of the system's reasoning have been set.

Moreover, the kernel of the system is built so that the module for manual input handling is implemented separately from the reasoning modules. By this, fully automated system, although not a real-time one, is only a matter of time - as soon as the automatic facial features extraction and pre-processing of the system are fully implemented and robustness of the system ensured, different parts of the system can be integrated in a final product that will perform automatic recognition of shown emotion through automatic recognition of shown facial expression.

9 References

[1] Burianek, J., Zyka, V., Rothkrantz, L.J.M. & Hlavacz, V. (1996). *Nostril and Chin Segmentation*. KBS Group, TU Delft, NL.

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

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

[4] Giarratano, J. & Riley, G. (1990). *Expert Systems - Principles and Programming*. PWS-KENT Publishing Company, Boston, USA.

[5] Jonge, D. de (1995). *Toward Automatic Facial Features Extraction and Recognition of Facial Expressions*. Master thesis, KBS Group, TU Delft, NL.

[6] Juran, P., Krsek, P., Rothkrantz, L.J.M. & Hlavacz, V. (1996). *Mouth Segmentation Using Active Contours*. KBS Group, TU Delft, NL.

[7] Kearney, G.D. & McKenzie, S. (1993). Design of a Memory-Based Expert System for Interpreting Facial Expressions in Terms of Signaled Emotions (JANUS). *Cognitive Science* 17, pp. 589-622. University of Greenwich, London, UK.

[8] Kobayashi, H. & Hara, F. (1992). Study on Face Robot for Active Human Interface. *IEEE International Workshop on Robot and Human Communication*, pp. 276-281.

[9] Libal, V., Rothkrantz, L.J.M. & Vlcek, M. (1996). *Mouth Segmentation Based on Colour Analysis*. KBS Group, TU Delft, NL.

[10] Pantic, M. (1996). *Human Emotion Recognition Clips Utilised Expert System : HERCULES*. Master thesis, KBS Group, TU Delft, NL.

[11] Pitas, I. (1993). *Digital Image Processing Algorithms*. Prentice-Hall International Ltd, UK.

[12] Reinders, M.J.T. (1995). *Model Adaptation for Image Coding*. PhD thesis, Faculty of Electrical Engineering, TU Delft, NL.

[13] Vark, R.J. van, Rothkrantz, L.J.M. & Kerckhoffs, E.J.H. (1996). *Prototypes of Multimedial Stress Assessment subject to Laboratory Experiments*. KBS Group, TU Delft, NL.

[14] Vincent, J.M., Myers, D.J. & Waite J.B. (1992). *Neural Networks for Vision, Speech & Natural Languages*. British Telecommunications. Chapman & Hall, London, UK.

[15] Yuille, A.L., Hallinan, P.W. & Cohen, D.S. (1992). Feature Extraction from face Using Deformable Templates. Proceedings *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 104-109, San Diego, USA.