# EXPLORING THE TIME COURSE OF FACIAL EXPRESSIONS WITH A FUZZY SYSTEM

*F. Piat and N. Tsapatsoulis*
*Dept of Electrical and Computer Engineering,*
*National Technical University of Athens, Zographou 15773, Greece*
piat@image.ntua.gr , ntsap@image.ntua.gr

## ABSTRACT

Recognizing facial expressions is one of the current challenges of research in HCI. Previous research show the limits of recognition based on a single static image, and analyzing video sequences seems more promising. We present here a fuzzy system for the classification of basic facial expressions and use it to examine the time course dynamics of facial expressions. The system's inputs are the variations of distances defined by salient facial points from one frame to the next. For maximum compatibility, those facial points (eyebrows, eyes, mouth) were chosen from the set of points defined in the standard MPEG-4 specifications, and so that their automatic extraction is tractable. The first results suggest that some expressions can be recognized early after the onset. For other expressions, it is in general possible to reduce significantly the number of possibilities. Forming early hypotheses regarding the expression could be necessary for a system to work in real-time, since other steps may have to follow: prediction of user's action, or choice of computer's action, etc... This also has implications for the recognition of milder expressions.

## 1. INTRODUCTION

One of the most important features of multimedia systems is the interactivity they offer to the user. We increasingly want systems able to adapt themselves to the user and to customize their behaviour accordingly. Computers may start to be truly intelligent when they understand the user's goals and expectations. A non-intrusive way to get some of this information is for the computer to decode the user's face expression, as this in general continuously provides a lot of information relating to his satisfaction. This information can be used by a system to get feed-back regarding the result or appropriateness of the last action, or to anticipate the user's next action, for instance. Some research [1] suggests that the meaning of an emotion may reside in the predispositions for possible actions it entails.

One of the numerous problems to recognize facial expressions is that they are ambiguous and we often need additional information in order to interpret them, for instance the speech accompanying the expression, or some contextual information. In particular, knowing the temporal position of a frame relative to the whole sequence is absolutely necessary. Without this information, we cannot distinguish a mild expression at its maximum of intensity (called apex) from the onset of a strong one: we would confuze a light smile with the beginning of a hearty laughter. This problem imposes severe restrictions on the possibility to recognize static expressions. Therefore it is necessary to explore the dynamics of the expressions, the way facial changes unfold in time over the course of the whole sequence. The speed of some movements and their relative temporal order may be a valuable information to improve recognition, even for off-line systems. Further, this should give us clues as to how early before the apex can an expression be recognized, and how to achieve early recognition, which is important for on-line systems. Recognizing expressions as early as possible could be crucial for many real-time applications, since recognition will most likely not be the ultimate goal of the application, but should be followed by some action, which choice and execution also require some time. Even when early recognition cannot be achieved, it is likely that restricting the number of possible responses to just a few will be helpful, as this will play a role similar to 'orienting the attentional focus' of the system: fewer hypotheses can be more thoroughly tested for improved accuracy. Finally, understanding the development of expressions over time can have implications for the recognition of mild expressions, as mentioned above.

In the next section we propose a fuzzy system built to recognize expressions from salient facial points. For now, it is limited to the recognition of the 6 basic 'universal' expressions, for tractability reasons (Smile, Anger, Sadness, Disgust, Fear and Surprise). Section 3 details the results of the analysis of whole sequences with the fuzzy system.

## 2. PROPOSED FUZZY SYSTEM

It is important for the generality of such a system to be compatible with the standards defined for video sequences, so we adopt the framework outlined in [2]. It defines sets of parameters that allow the description and animation of human bodies and faces for the MPEG-4 standard. Over 50 feature facial points (Facial Definition Parameters or FDP) are used to define a given face, and as many basic actions (Facial Action Parameter or FAP) that a face can perform are used to describe its movements and lead to the rendering of any expression. The FAPs express relations between FDPs, for instance the distance between two feature points. They are normalized according to some distances independent of the expression, in order to get consistent values regardless of the scale of the picture, distance from the camera, etc… Specifically, those FAP units are the distance between the eyes (ESo) for horizontal distances and the distance from the middle of the eyes to the tip of the nose (ENSo) for vertical distances.

## 2.1 Inputs

In order to reduce the amount of data to be dealt with and computation time, we choose to focus on the most expressive parts of the face: eyes, eyebrows and mouth. This should still contain enough information for categorization, and those are salient points that can be detected automatically in the future. At this stage, the localization of the points is still manual. From these 19 facial points, we compute 14 of the FAPs defined by their distances. After inspection of emotional sequences, it became apparent that some attitudes (e.g., position of the head) are very revealing of the emotional state but not captured by any of these 14 FAPs, so 6 FAPs were added (numbered 15-20), some being combinations of MPEG FAPs, some totally new. They will be used for an improved version of the system. Figure 1 shows the facial points and the distances computed, including those used as units for normalization. The original numbers of the FDP points in [2] are retained, so they are not consecutive anymore.

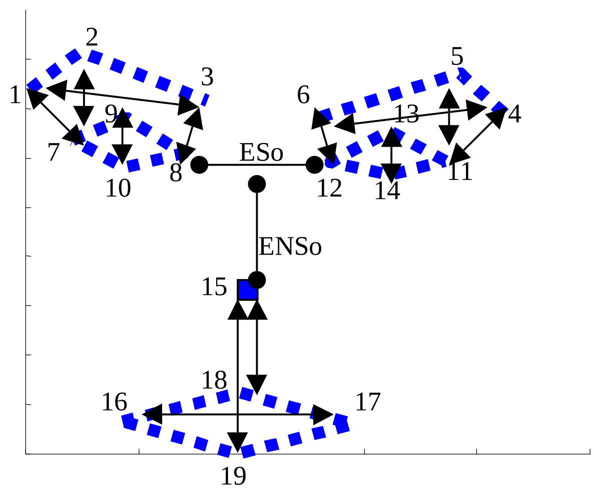


**Figure 1**. The 19 FDP points and the FAP distances they define, which time derivatives constitute the input to the fuzzy system. Eso and ENSo are the distances used for normalization.

Table 1 lists the FAPs computed from these points. Their values depend on the expression of the face, but also on the particular structural configuration of a given face. To feed the fuzzy system information that relates only to the facial movements, the FAP changes from one frame to the next are computed to constitute the input: $I^{i}_{t+1} = FAP^{i}_{t+1} - FAP^{i}_{t}$

## 2.2 Fuzzy Inference System

The continuity of the emotion space as well as the uncertainty involved in the feature estimation process, whether automatic or manual, make the use of fuzzy logic appropriate for the feature-to-emotion mapping. The structure of the proposed fuzzy inference system is described in Figure 2. The input is actually a 14-tuple, one value for each parameter of the FAP subset shown in Table 1, and describes the increment (or decrement) of the corresponding FAP. The output is a 6-tuple, one value for the degree of the belief that the expression is anger, sadness, joy, disgust, fear and surprise. On the universe of discourse of each input (or output) parameter, a fuzzy linguistic partition is defined: the fuzzification stage associates one of the 3 values low, medium or high to each FAP. The linguistic terms of the fuzzy partitions (for example medium open_jaw) are connected with the aid of the IF-THEN rules of the Rule Base. These IF-THEN rules are heuristically constructed and express the a priori knowledge of the system. The activation of the antecedents of a rule causes the activation of the consequences, i.e. the degree of belief that the emotion is X concluded from the degree of the increment (or decrement) of the FAPs. This is done for all 6 expressions and the expression with the highest degree of belief is considered the winner. The FAPs used in the inference rules for each expression are shown in Table 2.

# 3. RESULTS

The predictions of the fuzzy system were recorded for all frames after the first one for each of the 20 video sequences, totalling 192. The number of frames varied with each sequence. The average recognition rate was 75%, mostly because of the confusion between anger and disgust. This can probably be improved with a fine tuning of the parameters of the membership functions.

The average number of frames needed to give a reliable prediction was recorded for each type of expression. Half way through the sequence, the system can generally recognize a smile, or guess that the expression belongs to the group fear/surprise or anger/disgust.

Full results and discussion in the full, camera-ready paper.

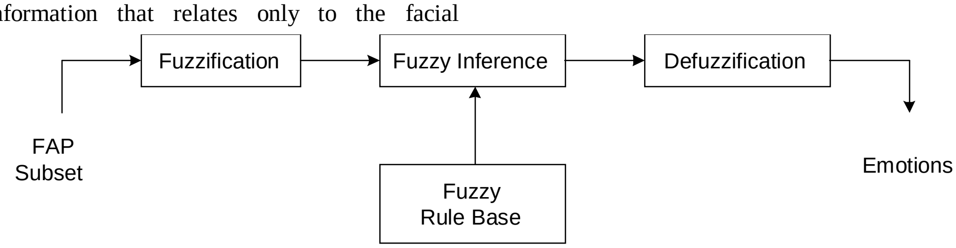


**Figure 2:** Structure of the fuzzy system.

| FAP name | Features used for the description | Positive Intensity |
|---|---|---|
| Squeeze_l_eyebrow | $f_1 = \frac{s(1,3)}{ESo}$ , $\frac{df_1}{dt}$ | $\frac{df_1}{dt} < 0$ |
| Squeeze_r_eyebrow | $f_2 = \frac{s(4,6)}{ESo}$ , $\frac{df_2}{dt}$ | $\frac{df_2}{dt} < 0$ |
| raise_u_midlip | $f_3 = \frac{s(16,30)}{ENSo}$ , $\frac{df_3}{dt}$ | $\frac{df_3}{dt} < 0$ |
| raise_l_midlip | $f_4 = \frac{s(16,33)}{ENSo}$ , $\frac{df_4}{dt}$ | $\frac{df_4}{dt} < 0$ |
| raise_l_i_eyebrow | $f_5 = \frac{s(3,8)}{ENSo}$ , $\frac{df_5}{dt}$ | $\frac{df_5}{dt} > 0$ |
| raise_r_i_eyebrow | $f_6 = \frac{s(6,12)}{ENSo}$ , $\frac{df_6}{dt}$ | $\frac{df_6}{dt} > 0$ |
| raise_l_o_eyebrow | $f_7 = \frac{s(1,7)}{ENSo}$ , $\frac{df_7}{dt}$ | $\frac{df_7}{dt} > 0$ |
| raise_r_o_eyebrow | $f_8 = \frac{s(4,11)}{ENSo}$ , $\frac{df_8}{dt}$ | $\frac{df_8}{dt} > 0$ |
| raise_l_m_eyebrow | $f_9 = \frac{s(2,7)}{ENSo}$ , $\frac{df_9}{dt}$ | $\frac{df_9}{dt} > 0$ |
| raise_r_m_eyebrow | $f_{10} = \frac{s(5,11)}{ENSo}$ , $\frac{df_{10}}{dt}$ | $\frac{df_{10}}{dt} > 0$ |
| open_jaw | $f_{11} = \frac{s(16,33)}{ENSo}$ , $\frac{df_{11}}{dt}$ | $\frac{df_{11}}{dt} > 0$ |
| close_upper_l_eyelid – close_lower_l_eyelid | $f_{12} = \frac{s(9,10)}{ENSo}$ , $\frac{df_{12}}{dt}$ | $\frac{df_{12}}{dt} < 0$ |
| close_upper_r_eyelid – close_lower_r_eyelid | $f_{13} = \frac{s(13,14)}{ENSo}$ , $\frac{df_{13}}{dt}$ | $\frac{df_{13}}{dt} < 0$ |
| stretch_l_cornerlip – stretch_r_cornerlip | $f_{14} = \frac{s(28,29)}{ESo}$ , $\frac{df_{14}}{dt}$ | $\frac{df_{14}}{dt} > 0$ |
| Mouth_corners_down | $f_{15} = \frac{sy([16,17],15)}{ENSo}$ , $\frac{df_{15}}{dt}$ | $\frac{df_{15}}{dt} > 0$ |
| Mouth_vertical_ asymetry | $f_{16} = Slope(16,17) - f_{20}$ | $f_{16} \neq 0$ |
| Mouth_occlusion (detection failure) | $f_{17} = boolean\ value$ | |
| Head_up/down | $f_{18} = \frac{ENSo}{ESo}$ , $\frac{df_{18}}{dt}$ | $\frac{df_{18}}{dt} < 0$ |
| Head_ rotation | $f_{19} = \frac{ESo}{ENSo}$ , $\frac{df_{19}}{dt}$ | $\frac{df_{19}}{dt} < 0$ |
| Head _tilt | $f_{20} = Slope([7,8],[11,12])$ | $f_{20} \neq 0$ |

**Table 1**: Description of FAP set using a subset of the FDP points. *Note:* s(i,j)=Euclidean distance between FDP points i and j, sy(i,j)=their vertical distance, [i,j] their middle point, ESo, ENSo=Horizontal and vertical distances used for normalization.

| | |
|---|---|
| Anger | squeeze_l_eyebrow, squeeze_r_eyebrow, raise_u_midlip, raise_l_midlip |
| Sadness | raise_l_i_eyebrow, raise_r_i_eyebrow, close_upper_l_eyelid, close_upper_r_eyelid, close_lower_l_eyelid, close_lower_r_eyelid |
| Joy | close_upper_l_eyelid, close_upper_r_eyelid, close_lower_l_eyelid, close_lower_r_eyelid, stretch_l_cornerlip, stretch_r_cornerlip |
| Disgust | close_upper_l_eyelid, close_upper_r_eyelid, close_lower_l_eyelid, close_lower_r_eyelid, raise_u_midlip |
| Fear | raise_l_o_eyebrow, raise_r_o_eyebrow, raise_l_m_eyebrow, raise_r_m_eyebrow_m, raise_l_i_eyebrow, raise_r_i_eyebrow, squeeze_l_eyebrow, squeeze_r_eyebrow, open_jaw |
| Surprise | raise_l_o_eyebrow, raise_r_o_eyebrow, raise_l_m_eyebrow, raise_r_m_eyebrow_m, raise_l_i_eyebrow, raise_r_i_eyebrow, open_jaw |

**Table 2**: Description of the six primary expressions using the FAPs.

## 4. REFERENCES


[1] Review of existing techniques for human emotion understanding and their applications in human-computer interaction, Technical Report, Research contract FMRX-CT97-0098 (DG12-BDNC), October 1998.

[2] ISO/IEC JTC1/SC29/WG11 MPEG96/N1365, "MPEG4 SNHC : Face and body definition and animation parameters", 1996.