A FUZZY SYSTEM FOR EMOTION CLASSIFICATION BASED ON THE MPEG-4 FACIAL DEFINITION PARAMETER SET

Nicolas Tsapatsoulis, Kostas Karpouzis, George Stamou, Frederic Piat and Stefanos Kollias

Department of Electrical and Computer Engineering, National Technical University of Athens
Heroon Polytechniou 9, 157 73 Zographou, GREECE
e-mail: {ntsap,kkarpou}@image.ntua.gr

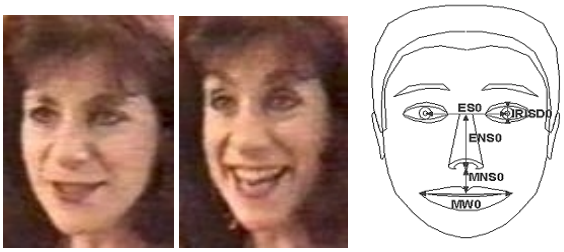
ABSTRACT

The human face is, in essence, an advanced expression apparatus; despite its adverse complexity and variety of distinct expressions, researchers has concluded that at least six emotions, conveyed by human faces, are universally associated with distinct expressions. In particular, sadness, anger, joy, fear, disgust and surprise form categories of facial expressions that are recognizable across different cultures. In this work we form a description of the six universal facial expressions, using the MPEG-4 Facial Definition Parameter Set (FDP) [1]. According to the MPEG-4 Standard, this is a set of tokens that describe minimal perceptible actions in the facial area. Groups of such actions in different magnitudes produce the perception of expression [2]. A systematic approach towards the recognition and classification of such an expression is based on characteristic points in the facial area that can be automatically detected and tracked. Metrics obtained from these points feed a fuzzy inference system whose output is a vector of parameters that depicts the systems' degree of belief with respect to the observed emotion. Apart from modeling the archetypal expressions we go a step further: by modifying the membership functions of the involved features according to the *activation* parameter [3] we provide an efficient way for recognizing a broader range of emotions than that related with the archetypal expressions.

1 INTRODUCTION

Research in emotion analysis has mainly concentrated on primary or archetypal emotions, which are universally associated to distinct expressions [4]. Very few studies [5] that explore non-archetypal emotions, have appeared in the computer science literature. In contrary, psychological researchers have extensively investigated [3][6] a broader variety of emotions. Although exploitation of the results obtained by psychologists is far from being straightforward, computer scientists can use some hints to their research. Whissel [3] suggests that emotions are points in a space with a relatively small number of dimensions, which at a first approximation, seem to be *activation* and *evaluation*. From the practical point of view, *evaluation* seems to express internal feelings of the subject and its estimation through face formations is intractable. On the other hand, *activation* is related to the facial muscles' movement and can be more easily estimated based on facial characteristics. Figure 1(a) and Table 1 illustrate the relation between face formation, expressed through the magnitude of the movement of some FDPs and the *activation* dimension due to Whissel (the *activation* for the term "delighted" is 4.2 while for "Eager" is 5).

The establishment of the MPEG standards, and especially MPEG-4, indicate an alternative way of analyzing and modeling facial expressions and related emotions [1]. Facial Animation Parameters (FAPs) and Facial Definition Parameter Set (FDP) are utilized in the framework of MPEG-4 for facial animation purposes. Automatic detection of particular FDPs in a video sequence is an active research area [8], which can be employed within the MPEG-4 standard for analyzing and encoding facial expressions.



(a) (b)

Figure 1: (a) Expressions labeled as "delighted" and "eager" (b) The Facial Animation Parameter Units (ES = ESo/1000; NS = ENSo/1000; MNS = MNSo/1000; MW = MWo/1000) [1]

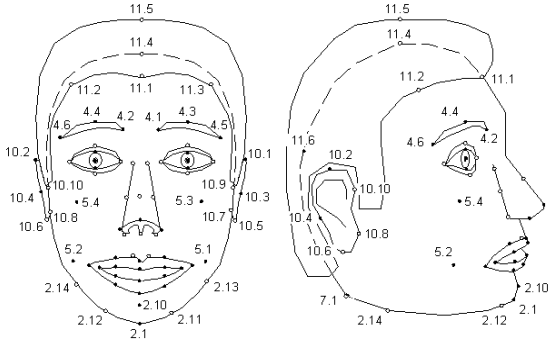


Figure 2: The 3D feature points of the FDP set [1]

	Activ	Eval		Activ	Eval
Afraid	4.9	3.4	Angry	4.2	2.7
Bashful	2	2.7	Delighted	4.2	6.4
Disgusted	5	3.2	Eager	5	5.1
Guilty	4	1.1	Joyful	5.4	6.1
Patient	3.3	3.8	Sad	3.8	2.4
Surprised	6.5	5.2			

Table 1: Selected emotion words from [3] and [6]

The continuity of the emotion space as well as the uncertainty involved in the detection of FDP points, which influences the feature estimation process, make the use of fuzzy logic appropriate for the feature-to-emotion mapping. Furthermore, gained experience from psychological researchers, as it is expressed through the *activation* parameter, can be incorporated into the system allowing the analysis of a larger number of emotions [9].

2 PARAMETER SETS FOR DEFINITION AND ANIMATION OF FACES

The Facial Definition Parameter set (FDP) and the Facial Animation Parameter set (FAP) were designed in the MPEG-4 framework to allow the definition of a facial shape and texture, as well as animation of faces reproducing expressions, emotions and speech pronunciation. The FAPs are based on the study of minimal facial actions and are closely related to muscle activation, in the sense that they represent a complete set of atomic facial actions; therefore they allow the representation of even the most detailed natural facial expressions, even those that cannot be categorized as particular ones. All the parameters involving translational movement are expressed in terms of the Facial Animation Parameter Units (FAPU). These units are defined with respect to specific distances in a neutral pose in order to allow interpretation of the FAPs [1] on any facial model in a consistent way. As a result, description schemes that utilize FAPs produce reasonable results in terms of expression and speech related postures (visemes) irrespectively. The FAPUs are illustrated in Figure 1(b) and correspond to fractions of distances between some key facial features.

In general, facial expressions and emotions can be described as a set of measurements (FDPs and derived features) and transformations (FAPs) that can be considered atomic with respect to the MPEG-4 standard; in this way, one can describe both the anatomy of a human face, as well as any animation parameters with groups of distinct tokens, the FDPs and the FAPs, thus eliminating the need to explicitly specify the topology of the underlying geometry. These tokens can then be mapped to automatically detected measurements and indications of motion on a video sequence and thus help recognize and recreate the emotion or expression conveyed by the subject.

3 RELATING THE FDP POINTS OF MPEG-4 WITH THE ARCHETYPAL EXPRESSIONS

Although muscle actions [4] are of high importance, with respect to facial animation, one is unable to track them analytically without resorting to explicit electromagnetic sensors. However, a subset of them can be deduced from their visual results, that is, the deformation of the facial tissue and the movement of some facial surface points. This reasoning resembles the way that humans visually perceive emotions, by noticing specific features in the most expressive areas of the face, the regions around the eyes and the mouth. The six archetypal emotions, as well as intermediate ones, employ facial deformations strongly related with the movement of some prominent facial points that can be automatically detected. These points can be mapped to a subset of the MPEG-4 FDP feature point set.

Table 2 illustrates our proposition [9] for the description of the archetypal expressions and some variations of them, using the MPEG-4 FAPs [1] terminology. Hints for the modeling were obtained from psychological studies [2][7], which refer to face formation during expressions, as well as from experimental data provided from classic databases like Ekman's (static) and MediaLab's (dynamic).

Table 3 shows the FDPs involved in the modeling FAPs as well as the actual features used for the description. The correlation between FAP and FDP subsets is mainly achieved through distances between the FDP points. Time derivatives of the computed distances are also used and serve two different purposes: first, they define the positive intensities for the FAP set and second, they characterize the development of the expressions and mark the expressions "apex". The $f_{i\text{-}NEUTRAL}$ refers to the particular distance when the face is in the neutral condition.

Anger	squeeze_l_eyebrow (+) lower_t_midlip (-) raise_l_i_eyebrow (+) close_t_r_eyelid (-) close_b_r_eyelid (-)	squeeze_r_eyebrow(+) raise_b_midlip (+) raise_r_i_eyebrow (+) close_t_l_eyelid (-) close_b_l_eyelid (-)
Sadness	raise_l_i_eyebrow (+) close_t_l_eyelid (+) raise_l_m_eyebrow (-) raise_l_o_eyebrow (-) close_b_l_eyelid (+)	raise_r_i_eyebrow (+) close_t_r_eyelid (+) raise_r_m_eyebrow (-) raise_r_o_eyebrow (-) close_b_r_eyelid (+)
Surprise	raise_l_o_eyebrow (+) raise_l_i_eyebrow (+) raise_l_m_eyebrow (+) squeeze_l_eyebrow (-) open_jaw (+)	raise_r_o_eyebrow (+) raise_r_i_eyebrow (+) raise_r_m_eyebrow(+) squeeze_r_eyebrow (-)
Joy	close_t_l_eyelid (+) close_b_l_eyelid (+) stretch_l_cornerlip (+) raise_l_m_eyebrow (+) lift_r_cheek (+) lower_t_midlip (-) **OR** open_jaw (+)	close_t_r_eyelid (+) close_b_r_eyelid (+) stretch_r_cornerlip (+) raise_r_m_eyebrow(+) lift_l_cheek (+) raise_b_midlip (-)
Disgust	close_t_l_eyelid (+) close_t_r_eyelid (+) lower_t_midlip (-) squeeze_l_cornerlip (+) **AND / OR** squeeze_r_cornerlip (+)	close_b_l_eyelid (+) close_b_r_eyelid (+) open_jaw (+)
Fear	raise_l_o_eyebrow (+) raise_l_m_eyebrow (+) raise_l_i_eyebrow (+) squeeze_l_eyebrow (+) open_jaw (+) **OR** close_t_l_eyelid (-) lower_t_midlip (-) **OR** lower_t_midlip (+)	raise_r_o_eyebrow (+) raise_r_m_eyebrow (+) raise_r_I_eyebrow (+) squeeze_r_eyebrow(+) close_t_r_eyelid (-)

Table 2: FAPs involved in the six archetypal expressions

3.1 Automatic Detection of Facial Protuberant Points

The detection of the FDP subset used to describe the involved FAPs was based on the work presented in [8]. However, for accurate detection in many cases human assistance was necessary. The authors are working towards a fully automatic implementation of the point detection procedure.

FAP name	Features for description / Utilized feature	Positive Intensity
squeeze_l_eyebrow	$f_1 = s(1,3)$ $F_1 = f_1 - f_{1\text{-NEUTRAL}}$	$F_1 < 0$
squeeze_r_eyebrow	$f_1 = s(4,6)$ $F_2 = f_2 - f_{2\text{-NEUTRAL}}$	$F_2 < 0$
lower_t_midlip	$f_1 = s(16,30)$ $F_3 = f_3 - f_{3\text{-NEUTRAL}}$	$F_3 < 0$
raise_b_midlip	$f_1 = s(16,33)$ $F_4 = f_4 - f_{4\text{-NEUTRAL}}$	$F_4 < 0$
raise_l_i_eyebrow	$f_1 = s(3,8)$ $F_5 = f_5 - f_{5\text{-NEUTRAL}}$	$F_5 > 0$
raise_r_i_eyebrow	$f_1 = s(6,12)$ $F_6 = f_6 - f_{6\text{-NEUTRAL}}$	$F_6 > 0$
raise_l_o_eyebrow	$f_1 = s(1,7)$ $F_7 = f_7 - f_{7\text{-NEUTRAL}}$	$F_7 > 0$
raise_r_o_eyebrow	$f_1 = s(4,11)$ $F_8 = f_8 - f_{8\text{-NEUTRAL}}$	$F_8 > 0$
raise_l_m_eyebrow	$f_1 = s(2,7)$ $F_9 = f_9 - f_{9\text{-NEUTRAL}}$	$F_9 > 0$
raise_r_m_eyebrow	$f_1 = s(5,11)$ $F_{10} = f_{10} - f_{10\text{-NEUTRAL}}$	$F_{10} > 0$
open_jaw	$f_1 = s(30,33)$ $F_{11} = f_{11} - f_{11\text{-NEUTRAL}}$	$F_{11} > 0$
close_t_l_eyelid – close_b_l_eyelid	$f_1 = s(9,10)$ $F_{12} = f_{12} - f_{12\text{-NEUTRAL}}$	$F_{12} < 0$
close_t_r_eyelid – close_b_r_eyelid	$f_1 = s(13,14)$ $F_{13} = f_{13} - f_{13\text{-NEUTRAL}}$	$F_{13} < 0$
stretch_l_cornerlip – stretch_r_cornerlip	$f_1 = s(28,29)$ $F_{14} = f_{14} - f_{14\text{-NEUTRAL}}$	$F_{14} > 0$
squeeze_l_eyebrow & squeeze_r_eyebrow	$f_1 = s(3,6)$ $F_{15} = f_{15} - f_{15\text{-NEUTRAL}}$	$F_{15} < 0$

Table 3: Description of FAP set using a subset of the MPEG-4 FDP set. *Note*: s(i,j)=Euclidean distance between FDPs i and j

4 FUZZY INFERENCE SYSTEM

The structure of the proposed system is shown in Figure 3. For each picture / frame that illustrates a face in an emotional state, a *15*-tuble feature vector, corresponding to the FAPs depicted in Table 3, is computed and feeds the fuzzy inference system. The input vector is fuzzified according to the membership functions of the particular elements. Details about the fuzzification procedure are given in the following section.

The output is an *n*-tuple, where *n* refers to the number of modeled emotions; for the archetypal emotions each particular output value expresses the degree of the belief that the emotion is anger, sadness, joy, disgust, fear or/and surprise. On the universe of discourse of each input (or output) parameter, a fuzzy linguistic partition is defined. 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 based on Tables 2 and 4 and express the a priori knowledge of the system.

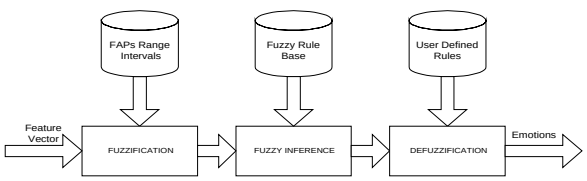


Figure 3: The structure of the fuzzy system

		A	Sa	J	D	F	Su
F_1 (ES)	Mean	-57	*	*	*	*	*
	StD	28	*	*	*	*	*
F_2 (ES)	Mean	-58	*	*	*	*	*
	StD	31	*	*	*	*	*
F_3 (MNS)	Mean	-73	*	-271	-234	*	*
	StD	51	*	110	109	*	*
F_4 (MNS)	Mean	*	*	*	-177	218	543
	StD	*	*	*	108	135	203
F_5 (ENS)	Mean	-83	85	*	*	104	224
	StD	48	55	*	*	69	103
F_6 (ENS)	Mean	-85	80	*	*	111	211
	StD	51	54	*	*	72	97
F_7 (ENS)	Mean	-66	*	*	*	*	54
	StD	35	*	*	*	*	31
F_8 (ENS)	Mean	-70	*	*	*	*	55
	StD	38	*	*	*	*	31
F_9 (ENS)	Mean	-149	*	24	-80	72	144
	StD	40	*	22	53	58	64
F_{10} (ENS)	Mean	-144	*	25	-82	75	142
	StD	39	*	22	54	60	62
F_{11} (MNS)	Mean	*	*	*	*	291	885
	StD	*	*	*	*	189	316
F_{12} (IrisD)	Mean	*	-153	-254	-203	244	254
	StD	*	112	133	148	126	83
F_{13} (IrisD)	Mean	*	-161	-242	-211	249	252
	StD	*	109	122	145	128	81
F_{14} (MW)	Mean	*	*	234	*	*	-82
	StD	*	*	98	*	*	39
F_{15} (ES)	Mean	-69	-56	*	-52	*	86
	StD	51	35	*	34	*	60

Table 4: Experimentally verified FAPs involved in archetypal expressions (**A**nger, **Sa**dness, **J**oy, **D**isgust, **F**ear, **Su**rprise)

4.1 Fuzzification of the input vector

Table 4 is our basis for constructing the membership functions for the feature vector elements. Using some data sets that represent archetypal expressions like Ekman's images and Media Lab's video sequences we computed the parameters of Table 4 for the FAPs employed in the archetypal expressions. In this way we estimate the universe of discourse for the particular features. For example a reasonable range of variance for F_5 is $[m_{A5} - 3\cdot\sigma_{A5} \quad m_{Su5} + 3\cdot\sigma_{Su5}]$ where m_{A5}, σ_{A5} and m_{Su5}, σ_{Su5} are the mean values and standard deviations of feature F_5 corresponding to expressions *anger* and *surprised* respectively. For unidirectional features like F_{11} either the lower or upper limit is fixed to zero. Table 4 can be also used to determine how many and which linguistic terms should be assigned to a particular feature; for example the linguistic terms *medium* and *high* are sufficient for the description of feature F_{11}.

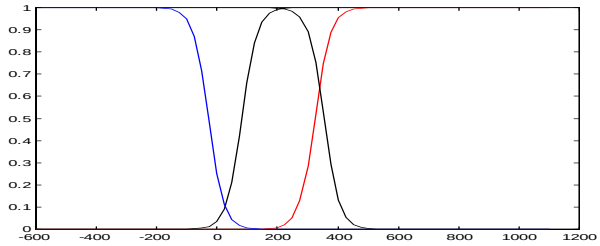


Figure 4: Membership functions for feature F_4

The membership functions for the particular features have been also derived based on the statistics provided in Table 4. Figure 4 illustrates the membership functions for linguistic terms *low*, *medium* and *high* corresponding to feature F_4.

4.2 Recognition of a broader variety of emotions

The system described in Section 4 can be modified to analyze more than the archetypal emotions. In order to do that we need to: (a) estimate which of features participate to the emotions and, (b) modify, with a reasonable manner, the membership functions of the features to correspond to the new emotions.

As a general rule, one can define six general categories, each one characterized by a fundamental archetypal emotion; within each of these categories intermediate expressions are described by different emotional and optical intensities, as well as minor variation in expression details. For example, the emotion group "fear" also contains "worry" and "terror"; these two emotions can be modeled by translating appropriately the positions of the linguistic terms, associated with the particular features, in the universe of discourse axis. The same rationale can also be applied in the group of "disgust" that also contains "disdain" and "repulsion".

Keeping the above in mind the difference in *activation* values a_Y and a_X corresponding to expressions Y and X, which belong to the same category, is split in the membership functions based on the following rules:

Rule 1: Emotions of the same category involve the same features F_i.

Rule 2: Let μ_{XZi} and μ_{YZi} be the membership functions for the linguistic term Z corresponding to F_i and associated with emotions X and Y respectively. If the μ_{XZi} is centered at value m_{XZi} of the universe of discourse then μ_{YZi} should be centered at $m_{YZi} = \frac{a_Y}{a_X} m_{XZi}$

Rule 3: a_Y and a_X are known values obtained from Whissel's study [3].

5 EXPERIMENTAL RESULTS

In order to evaluate our algorithm we have performed experiments on two different datasets; the first consist of static images showing only archetypal emotions and the other contains video sequences showing a variety of expressions. The results are summarized in Table 5 and 6.

	Fear	Disgust	Joy	Sadness	Surprise	Anger
Static Set	67%	73%	92%	76%	94%	85%
PHYSTA	58%	64%	87%	61%	85%	68%

Table 5: Experimental Results on archetypal emotions

Material: The first dataset consists of: (a) 80 pictures of CMU database showing the emotions *neutral*, *joy*, *sadness* and *anger*, (b) 60 pictures of Yale database corresponding to emotions *normal*, *joy*, *surprise* and *sadness*, (c) 100 selected frames from MediaLab's database corresponding to *neutral*, *joy*, *sadness, surprise, disgust,* and *anger,* and (d) 30 pictures from various sources showing the emotions *neutral* and *fear*. Pictures corresponding to neutral condition were used as the first frame for all other emotions. The second dataset is a pilot database created in the framework of project PHYSTA of the Training Mobility and Research Program of the European Community [5]. The PHYSTA pilot database contains both video and audio signals, showing humans in various emotional states –not only archetypal ones- and has been recorded from BBC's broadcasted program. In our simulations only the video signal has been considered.

Discussion: Table 5 shows that the classification rates of PHYSTA dataset are lower than the ones of Static Set. This fact emanates from the content of pictures; the video sequences of PHYSTA dataset show emotional states recorded from real life and not extreme cases contained in the databases of the Static Set. It is also shown in Table 5 that emotions corresponding to larger muscle movements –higher *activation* parameter- are more easily recognizable. Table 6 shows some preliminary results on classifying variations of archetypal emotions. Although the classification rates are low, they still be above chance level; this fact implies that is not intractable to discriminate related emotions based on the scheme proposed in Section 4.2.

	Disdain	Disgust	Repulsion	Delighted	Eager	Joy
Rec. Rate	48%	60%	52%	61%	65%	75%

Table 6: Results on variations of archetypal emotions

6 CONCLUSION

In this study we have shown that the FAP and FDP sets of MPEG-4 standard, accompanied with a fuzzy inference system, provide an efficient means for the recognition of emotions. The fuuzy system accounts for the continuity of the emotion space as well as for the uncertainty of feature estimation process. Moreover, experts knowledge is included in the system by the use of a rule base; the latter has been constructed from psychological studies and verified experimentally. Finally we have introduced the use of the *activation* parameter as the basis of extending the system to recognize a broader range of emotions.

7 REFERENCES

[1] M. Tekalp, "Face and 2-D Mesh Animation in MPEG-4," Tutorial Issue On The MPEG-4 Standard, Image Communication Journal, Elsevier, 1999.

[2] G. Faigin, "The Artist's Complete Guide to Facial Expressions," Watson-Guptill, New York, 1990.

[3] C. M. Whissel, "The dictionary of affect in language," R. Plutchnik and H. Kellerman (Eds) Emotion: Theory, research and experience: vol 4, The measurement of emotions. Academic Press, New York, 1989.

[4] P. Ekman and W. Friesen, "The Facial Action Coding System," Consulting Psychologist Press, San Francisco, CA, 1978.

[5] EC TMR Project PHYSTA Report, "Development of Feature Representation from Facial Signals and Speech," January 1999.

[6] R. Plutchik, "Emotion: A Psychoevolutionary Synthesis," Harper and Row, New York, 1980.

[7] F. Parke and K. Waters, "Computer Facial Animation," A K Peters, 1996

[8] Kin-Man Lam and Hong Yan, "An Analytic-to-Holistic Approach for Face Recognition Based on a Single Frontal View," IEEE Trans. on PAMI, vol. 20, no. 7, July 1998.

[9] K. Karpouzis, N. Tsapatsoulis and S. Kollias, "Moving to Continuous Facial Expression Space using the MPEG-4 Facial Definition Parameter (FDP) Set," in Proc. of the Electronic Imaging 2000, San Jose, CA, USA, Jan. 2000.