

# Fast facial expression recognition for emotion awareness disposal

Martina Eckert, Almudena Gil, Diego Zapatero, Juan Meneses, José F. Martínez Ortega  
 Research Center on Software Technologies and Multimedia Systems for Sustainability (CITSEM)  
 Technical University of Madrid, Spain  
 Martina.eckert@upm.es

**Abstract**—This paper presents a simple and fast expression recognition algorithm aimed at running in a secondary plane to provide emotion awareness for primary applications as e.g. exergames, in real time. The algorithm is based on the extraction of 19 facial landmarks which are used to detect some of the Action Units (AUs) defined in the Facial Action Coding System (FACS) and a newly created one. In addition, the new concept of Combined Action Units (CAUs) is presented. Those are grouped AUs which are detected as a unit. The applied emotion classification is based on logical rules, no learning is involved. First implementations have been made on a mobile platform.

**Keywords**—*facial emotion recognition; action units; rule based classification; facial landmarks; FACS; emotion awareness*

## I. INTRODUCTION

Facial emotion recognition has a very long history and much effort is still being done to achieve a better performance year by year. Actually, very good and robust systems are already available, which not only recognize basic emotions, but also subtle expressions called “in the wild”. Most of the systems are very sophisticated and based on highly complicated algorithms such as facial characteristics extraction via Active Appearance Meshes (AAM) or Gabor features and classification algorithms that imply learning mechanics like Neural Networks or Support Vector Machines (SVM). Those techniques work well, but are slow, need training and require a high computational amount.

There are also very promising fast techniques, as e.g. [1] which is a technique based on Local Binary Patterns (LBP) and achieves a real time detection. Also the technique of [2], based on edge detection and rules is fast and presents quite good results. Nevertheless, no recent work could be found that extracts facial landmarks and uses them to detect AUs, which afterwards are used to detect the emotion.

This work has been elaborated to find a suitable algorithm that quickly provides information about basic emotional states, i.e. emotion awareness that could be used by other applications. E.g. an intelligent rehabilitation gaming software, could use this knowledge to react to the user’s feelings by adapting the level of difficulty, the type of exercise etc., which would increase the efficiency of a rehabilitation process. Other examples could be to work in the background of a commercial application to detect if advertisements are working, or of a political debate shows, to estimate the public’s opinion about politicians.

Above all, smartphone applications are addressed, which already occupy a fixed place in our lives. The presented approach could be very useful to adapt any application to the user’s preferences and necessities, e.g. change common functionalities for elderly people: when the system detects e.g. a difficulty while using a mobile phone, it could react by augmenting the letter size or similar.

Therefore, the aim is to read the user’s face in the background of another application, detecting only a useful subset of the six basic emotions or the global mood. E.g. for exergames, it would be essential to detect Happiness as an index for fun, and Disgust and Angriiness for high effort or even suffering, whereas Fear, Sadness and Surprise are not so important.

Basic requirements of the implemented detection method are swift responses and robustness, with low processing times and low memory occupation. Our investigation and comparisons lead to the conclusion to implement a rule-based system which classifies easily distinguishable indicators, in this case, facial AUs (Action Units). Here, the aim is not to learn from a given data set, but to find out ever suitable rules. Learning algorithms are generally very expensive in computational effort and have to be trained for each different situation and environment.

The main implementations and tests were performed in Matlab, currently, performance is also tested on a mobile phone. In the following, the method is shortly introduced, a detailed description would be provided in the full paper.

## II. METHOD

Our method is based on the standard approach consisting in preprocessing, feature extraction and machine learning. Here, preprocessing is nearly inexistent, only a histogram equalization is performed to enhance contrast in difficult lightning conditions. Feature extraction is based on 19 characteristic facial points, which is the minimum needed to describe basic facial movements. To classify those movements, we adopt the Facial Action Coding System (FACS) proposed by Ekman and Friesen in 1978 [3], with some modifications to fit for 19 points and realize the decision with a fixed decision tree, based on the analysis of multiple publications and own observations.

### A. Facial landmark extraction

The first step is the extraction of 19 characteristic facial points which later are used to recognize selected AUs. The Viola-Jones algorithm [4] is applied for face and eye detection. Based on the eye's positions, basic rules for face composition are applied to roughly extract the regions of eyebrows, nose and mouth (Fig. 1, below, left). Furthermore, those regions undergo a simple thresholding process in the grey scale range (color is transformed to YCbCr using Y) which is followed by different morphological operations to extract well-defined areas of eyebrows, eyes and mouth, see example in Fig. 1, above. Their extreme points provide four landmarks on each eye, three on each brow, one on the nose tip, and four around the mouth (Fig. 1, below, center). The rightmost image below shows the result of a real-time test implementation in a mobile phone.

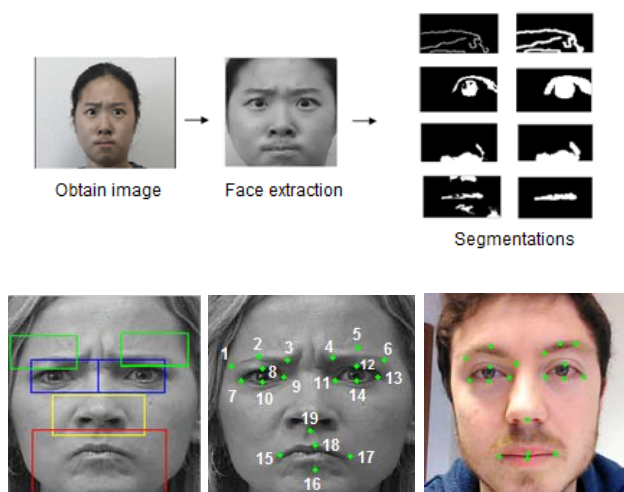


Fig. 1. Process to obtain 19 facial landmarks. Upper row: face extractions and segmentations. Lower row: regions, resulting points and real time implementation in mobile phone. Images are taken from JAFFE [5] and CK+ database [6].

The detection accuracy of our algorithm has been compared with [7] which is based on local information. Results are shown in Table I, as can be seen, they outperform the ones from [7] in most cases.

TABLE I: DETECTION ACCURACY OF 19 FACIAL POINTS

Face point	Proposed method	Silva [7]	Face point	Proposed method	Silva [7]
P1	90%	60%	P11	45%	60%
P2	95%	---	P12	70%	55%
P3	90%	70%	P13	90%	60%
P4	80%	55%	P14	100%	65%
P5	100%	---	P15	70%	30%
P6	90%	35%	P16	45%	70%
P7	85%	65%	P17	55%	40%
P8	80%	65%	P18	35%	75%
P9	50%	70%	P19	80%	---
P10	100%	70%			

### B. Action Unit selection and CAU definition

In FACS, 44 Action Units are defined, 12 of them in the eye region, 28 in the mouth region, 3 for the nose and 1 for the neck. For the here presented work, only a subgroup has been selected, because many are quite subtle and very few are needed to classify the six basic emotions. Furthermore, the few number of available facial landmarks restrict the number of detectable AUs, so finally only 18 AUs plus one newly defined one for *Happy* (AU-H) will be applied (see Table II).

Table II has been created by analyzing the opinion of experts presented in multiple publications. Concretely, the information from [6], FACS (cited in [8]), [9], [10], [11], [12], and [13] was unified, counting the coincidences. The table indicates the number of times each AU was mentioned by the cited authors. Some AUs were found to be uniquely presenting one emotion, or being dominant, they are marked as  $>n<$  (with  $n$  = number of citations) and outlined in grey. These AUs are the crucial ones, because, if they can be truly detected, all other emotions can be neglected. If AU7, 23 or 24 are detected, the emotion has to be anger, in case of AU9 and 10 it is *Disgust*. *Happy* is determined by AU6 and AU12 or AU-H, *Sadness* by AU15 and *Surprise* by AU26. Only *Fear* cannot be detected by an exclusive AU, but due to the former conditions possibly it could be the “left over” one, if all others don't match. *Fear*, *Sadness* and *Surprise* are the emotions that are most difficult to distinguish, because all of them include AU1, and AU2, 4 and 5 are shared by two of them.

TABLE II: AU SELECTION AND CAU DEFINITION

	AU	CAU	Suffering				Fun	
			Anger	Disgust	Fear	Sadness	Happy	Surprise
Eyes	1	1			9	9	1	8
	2				9			8
	5		6		8			7
	4			8		7	7	NOT
	6					4	>8<	
	7		>5<			1	1	1
	Mouth	9	2	2	>7<			
10		4		>6<			1	1
12								>10<
15		3		1			>8<	
17			3	2			5	
20			2		>6<		1	NOT
22				3				
23		4	>5<					
24			>5<					
25			5	4	5	2	2	1
26				5	3	4	1	1
27			1	1	3	1		>5<
H								New

To simplify the decisions even more and to eliminate ambiguities, 11 AUs were grouped together to form four Combined Action Units (CAUs), they appear grey shared in Table II. The idea and advantage of a CAU is that more points can be used to capture multiple facial movements in one step instead of detecting the individual movements one by one, which is more prone to errors. Also, the total amount of motions to detect is reduced. As an example, in case of CAU1,

the outer and inner brows raise as well as the eye lids. So, all first 14 facial points could be explored in relation to each other. In case of detecting AU1, 2 and 5 individually, only subsets of points would be explored, and the relation with others cannot be exploited as it depends on the expression. AU1, AU2 and AU5, which usually appear together and are present in *Fear* and *Surprise*, were joined to constitute CAU1. Nevertheless, in case of *Sadness* and *Anger* AU1 and AU5 are occurring without the others, so they also have to be detected individually. AU9 and AU10 are very similar and in our case are detected with the help of P18 and P19. We decided to join them to one which is called CAU2. CAU3 is built by AU15 and AU17, both being frowning mouths. Finally, AU23 and AU24, which are not even visually distinguishable, are joined together composing finally CAU4. Furthermore, a new AU has been defined, which does not appear in the FACS, as it is equal to the occurrence of AU12 (lip corner puller) + AU25 (lips part). As AU25 could occur in other expressions, the combined AU is defined and called AU-H (H for *Happy*). It represents an open smiling mouth which robustly identifies *Happy* and is easy to detect. It is not the same as AU27 which is a mouth opening without a smile and more likely to represent *Surprise*.

AU16 has been eliminated because it's too subtle to be distinguished from the neutral mouth, AU11 is neglected because it is secondary and only represents sadness. The grey written AU22 will also be excluded, as it is not decisive for any emotion.

### C. Action Unit detection

AU and CAU detections are made on the basis of a pattern comparison between the actual and a neutral image (which has to be obtained by an initial calibration process). The pattern formed by a group of points building an AU or CAU is characterized by 3 descriptors: enclosed area of especially defined triangles, point distances and motion vectors. Those descriptors are compared with predefined thresholds that identify the activation of an AU or CAU. The thresholds have been obtained through heuristic tests.

Figure 2 shows the definition of 10 triangles over the 19 points. Every AU or CAU is defined through one or more triangles and detected by analyzing one or more of the mentioned descriptors, depending on the facial movements. Figure 3 shows the influence of facial movements on the triangles for two examples: CAU1 (eyes) for Surprise and Fear compared with the neutral position and CAU3 (mouth) for Sadness compared with the neutral position.

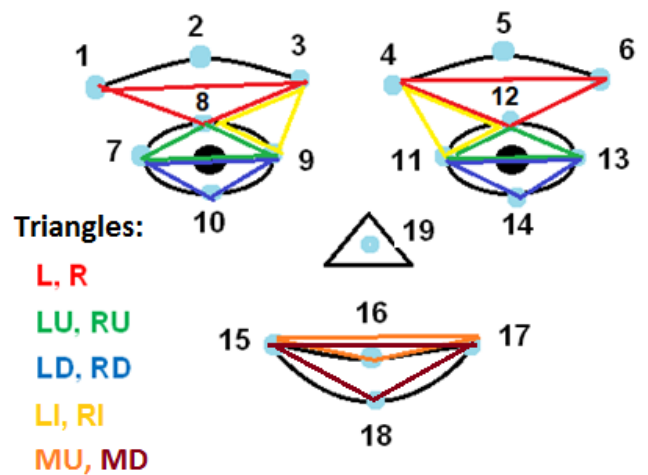


Fig. 2: Triangles used to detect AUs and CAUs

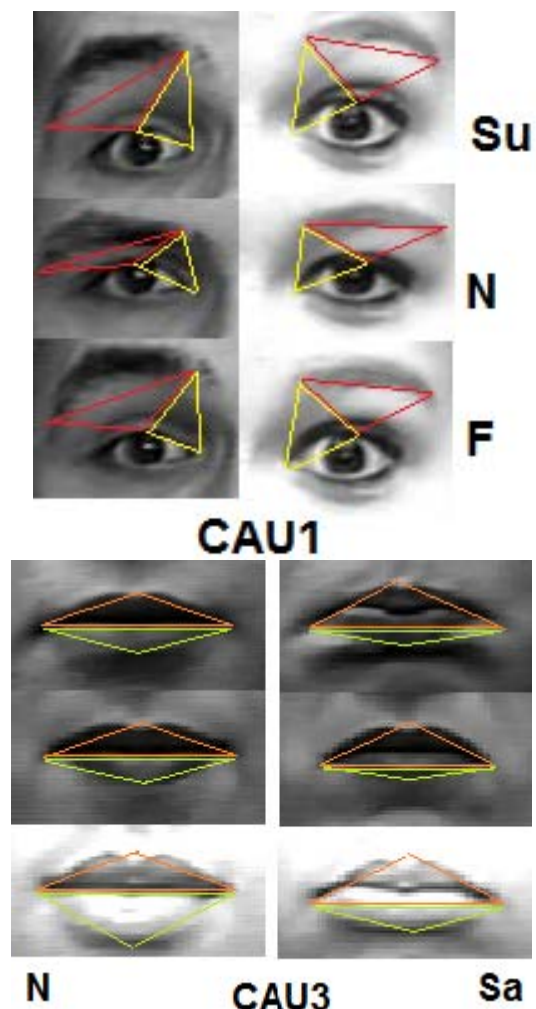


Fig. 3: Triangles used for CAU detection. Above: Effect of the emotions *Surprise* (Su), and *Fear* (F) on CAU1. Neutral expression (N) is in the middle center. Below: Three examples for the effect of *Sadness* (Sa) on CAU3. Neutral expression (N) on the left side.

#### D. Decision tree

For the AUs and CAUs found to be important (Table II), all combination rules published by the corresponding authors have been combined in a matrix to fit as input for building a decision tree, see Table III.

TABLE III: ESSENTIAL COMBINATIONS OF SELECTED AUS AND CAUS

Possible combinations for AUs and CAUs					
Anger	Suffering			Fun	
	Disgust	Fear	Sadness	Happy	Surprise
AU4+ CAU4	CAU2	CAU1+ AU4+	AU4+ CAU3	AU6+ AU12	CAU1+ AU26
AU4+ AU7+ CAU2	CAU2+ AU25	AU4+ AU25	AU6+ CAU3	AU6+ AU-H	CAU1+ AU27
AU4+ AU7+ CAU4		CAU1+ AU4+ AU26		AU12	CAU1
AU4+ AU7+ AU26		CAU1+ AU4+ AU27		AU-H	
AU4+ AU7		CAU1			
CAU4					

The idea was to create an essential tree out of the essential information and to avoid an excessive training process. Nevertheless, all available tree building algorithms, in our case we applied the ID3 implemented in Matlab 2012b are made for training over huge data sets and produce the best results for optimum size sets, which means that the amount of data neither has to be too small (which leads to under fitting) nor too big, which leads to overfitting. In spite of the fact that our matrix already contains all essential information, when fitting it in a tree building algorithm that explore data redundancies to find out more and less probable tree branches, the outcome is not optimal, as the matrix is too sparse. In our case, the tree obtained by the ID3 algorithm is losing elements which were thought to be essential, but as they are represented few times, they are falling out of the tree. The resulting tree is shown in figure 4. Instead of 15 AUs/CAUs, it only contains 5 and the results obtained are weak.

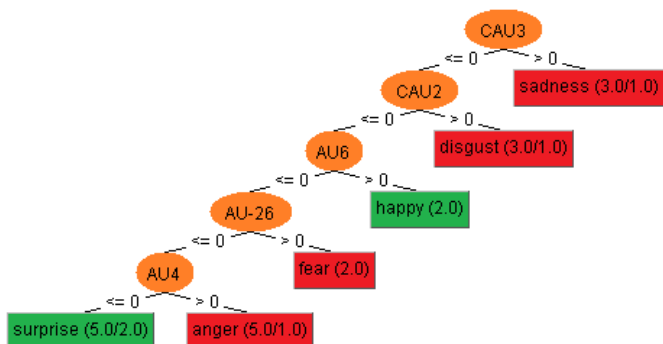


Fig. 4: Decision tree for two output states of moods

Consequently, this tree has to be built with help of a different algorithm which uses all AUs and CAUs, this will be the next step of our work.

Nevertheless, the matrix can be used with a different output vector that joins the positive and the negative moods. In

this way, a tree with just two output states can be build and serve as a global mood detector. Therefore, *Happy* and *Surprise* are joined to represent a positive mood (in Table II marked as “Fun”) whereas all others mean negative mood or “Suffering”. As the joined expressions carry redundant information, i.e. some AUs and CAUs are occurring in multiple emotions, the number of necessary AUs and CAUs is still reduced, now from 15 to 6, but the maintained AUs and CAUs are others. The corresponding decision tree is shown in figure 5, test results are presented in the following section.

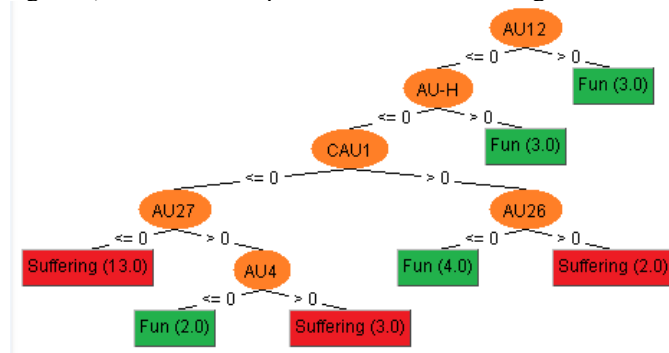


Fig. 5: Decision tree for two output states of moods

Finally, for each frame, the complete process is as follows: landmark detection, calculation of descriptors for the involved AUs and CAUs (CAU1, AU12, AU-H, AU27, AU4 and AU26), check if they are activated, create a 6<sup>th</sup> order vector with active/non-active AUs/CAUs, and feed it into the decision tree.

### III. TESTS AND RESULTS

Two different tests have been performed to show the efficiency of the tree, one in Matlab and with the CK+ image database [6] (Table IV) and one in an Android implementation with real faces (Table V). As can be seen, the real results are even better than the ones obtained with the database, although the same images were generally used by the authors who’s publications we analyzed to find out the rules.

TABLE IV: CONFUSION MATRIX FOR POSITIVE AND NEGATIVE MOOD WITH CK+ DATABASE [6]

	Anger	Disgust	Fear	Sadness	Happy	Surprise
Fun	13	5	38	41	<b>61</b>	<b>90</b>
Suffering	<b>87</b>	<b>95</b>	<b>63</b>	<b>59</b>	39	10

TABLE V: CONFUSION MATRIX FOR POSITIVE AND NEGATIVE MOOD WITH ANDROID PHONE IN REAL TIME

	Anger	Disgust	Fear	Sadness	Happy	Surprise
Fun	9	4	16	32	<b>92</b>	<b>97</b>
Suffering	<b>91</b>	<b>96</b>	<b>84</b>	<b>68</b>	8	3

Figure 5 shows the Android implementation with an example of the emotion *Surprise*, detected with three different intensities. The tool has two additional options: showing the facial regions used to extract the landmarks (center) and showing the 19 facial points obtained (right). The application can be downloaded for trials at [14].





Fig. 5: Real-time detection of the emotion “Surprise” in three different intensities. The center images shows the regions used for landmark detection and the right image shows the obtained landmarks.

#### IV. CONCLUSIONS

This work pretends to show a very simple and fast method to detect basic emotional states based on few facial landmarks, AUs, newly defined CAUs and rules. The method was created to provide emotion awareness to primary applications, which would use it to automatically adapt to the user’s requirements. On the one hand, results are weak for individual emotion recognition due to two problems: one is the difficulty to analyze action units robustly with help of few facial landmarks. The other is the sparse matrix used to train the decision tree. On the other hand, very good results have been obtained for a fast estimation of the mood of a person. This knowledge can be used as valuable knowledge by a primary application.

For the future work is planned to improve the AU/CAU recognition with help of more robust descriptors. Furthermore, the logic rules will be implemented in a new tree, without use of machine learning functions.

#### ACKNOWLEDGMENT

This work was sponsored by Spanish National Plan for Scientific Tech. Res. and Innov.: TEC2013-48453-C2-2-R.

#### REFERENCES

- [1] R. Verma and M.-Y. Dabbagh, "Fast FACial Expression Recognition Based on Local Binary Patterns", *26th IEEE Canadian Conference Of Electrical And Computer Engineering*, 2013.
- [2] J. Z. Zhang, Z. Qin, P. C. Tay and R. D. Adams, "A novel facial expression recognition method using fast BEMD based edge detection," *Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, 2010, pp. 204-208.
- [3] P. Ekman and W. Friesen, "Facial Action Coding System: A Technique for the Measurement of Facial Movement", *Consulting Psychologists Press*, Palo Alto, 1978.
- [4] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.
- [5] Japanese Female Facial Expression (JAFFE) Database [Online]. Available: <http://www.kasrl.org/jaffe.html>.
- [6] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," *Comput. Vis. Pattern Recognit. Work. (CVPRW), IEEE Comput. Soc. Conf.*, no. July, pp. 94–101, 2010.
- [7] C. Silva, L. Schmitman, L. Oliveira, "Detection of Facial Landmarks Using Local-Based Information", *19th Brazilian Conference on Automation*, Campina Grande, PB, Brazil, September 3, 2012.
- [8] M. F. Valstar and M. Pantic, "Biologically VS. Logic inspired encoding of facial actions and emotions in video," *IEEE Int. Conf. Multimed. Expo*, vol. 2006, pp. 325–328, 2006.
- [9] David Capino, "Facial Action Coding System - Khappucino’s Tutorial," 2009. [Online]. Available: <https://es.scribd.com/doc/18649644/Facial-Action-Coding-System-Khappucino-s-Tutorial>. [Accessed: 15-Apr-2016].
- [10] N. U. Khan, "A Comparative Analysis of Facial Expression Recognition Techniques," *3rd IEEE Int. Advance Comp. Conf.*, pp. 1262–1268, 2013.
- [11] Y. Li, S. Wang, Y. Zhao, and Q. Ji, "Simultaneous Facial Feature Tracking and Facial Expression Recognition," vol. 22, no. 7, pp. 2559–2573, 2013.
- [12] J. Hamm, C. G. Kohler, R. C. Gur, and R. Verma, "Automated Facial Action Coding System for dynamic analysis of facial expressions in neuropsychiatric disorders.," *J. Neurosci. Methods*, vol. 200, no. 2, pp. 237–56, Sep. 2011.
- [13] C. G. Kohler, T. Turner, N. M. Stolar, W. B. Bilker, C. M. Brensinger, R. E. Gur, and R. C. Gur, "Differences in facial expressions of four universal emotions," *Psychiatry Res.*, vol. 128, no. 3, pp. 235–244, 2004.
- [14] D. Zapatero, "Emorec - Emotion Recognition tool", <https://github.com/diegokr/emorec>