Real-time Emotion Recognition
Novel Method for Geometrical Facial Features Extraction

Claudio Loconsole[1], Catarina Runa Miranda[2], Gustavo Augusto[2], Antonio Frisoli[1] and Verónica Orvalho[2]

[1]*PERCRO Laboratory, Scuola Superiore Sant'Anna, Pisa, Italy*

[2]*Instituto Telecomunicacões, Faculdade de Ciências, Universidade do Porto, Porto, Portugal*

Keywords: Human-computer interaction, Emotion Recognition, Computer Vision.

Abstract: Facial emotions provide an essential source of information commonly used in human communication. For humans, their recognition is automatic and is done exploiting the real-time variations of facial features. However, the replication of this natural process using computer vision systems is still a challenge, since automation and real-time system requirements are compromised in order to achieve an accurate emotion detection. In this work, we propose and validate a novel methodology for facial features extraction to automatically recognize facial emotions, achieving an accurate degree of detection. This methodology uses a real-time face tracker output to define and extract two new types of features: *eccentricity* and *linear* features. Then, the features are used to train a machine learning classifier. As result, we obtain a processing pipeline that allows classification of the six basic Ekman's emotions (plus *Contemptuous* and *Neutral*) in real-time, not requiring any manual intervention or prior information of facial traits.

1 INTRODUCTION

Facial expressions play a crucial role in communication and interaction between humans. In the absence of other information such as speech interaction, facial expressions can transmit emotions, opinions and clues regarding cognitive states (Ko and Sim, 2010). A fully automatic real-time face features extraction for emotion recognition allows to enhance the communication realism between humans and machines. There are several research fields interested in developing automatic systems to recognize facial emotions. They mainly are represented by:

- Cognitive Human-Robot Interaction: the evolution of robots and computer animated agents bring a social problem of communication between these systems and humans (Hong et al., 2007);
- Human-Computer Interaction: facial expressions analysis is widely used for telecommunications, behavioural science, videogames and other systems that require facial emotion decoding for communication (Fernandes et al., 2011).

Several face recognition systems have been developed for real time facial features detection as well as (e.g. (Bartlett et al., 2003)). Psychological studies have been conducted to decode this information only using facial expressions, such as the Facial Action Coding System (FACS) developed by Ekman (Ekman and Friesen, 1978).

As stated on the recent survey (Jamshidnezhad and Nordin, 2012), among existing facial expression recognition systems, the common three-step pipeline for facial expressions classification (Bettadapura, 2009) is composed by:

1. the *Facial recognition* phase;
2. the *Features extraction* phase;
3. the *Machine learning classifier* phase (preliminary model training and on-line prediction of facial emotions).

As claimed in the same survey, the second pipeline phase (features extraction) strongly influences the accuracy and computational cost of the overall system. It follows that the choice of the type of the features to be extracted and the corresponding methods to be used for the extraction is fundamental for the overall performances.

The commonly used methods for feature extraction can be divided into *geometrical* methods (i.e. features are extracted from shape or salient point locations such as the mouth or the eyes (Kapoor et al., 2003)) and *appearance-based* methods (i.e. skin features like frowns or wrinkles, *Gabor Wavelets* (Fischer, 2004)).

Geometric features are selected from landmarks positions of essential parts of the face (i.e. eyes, eyebrows and mouth) obtained by a face features recognition technique. These extraction methods are characterized by their simplicity and low computational cost, but their accuracy is extremely dependent on the

Loconsole C., Runa Miranda C., Augusto G., Frisoli A. and Costa Orvalho V..
Real-time Emotion Recognition - Novel Method for Geometrical Facial Features Extraction.
DOI: 10.5220/0004738903780385
In *Proceedings of the 9th International Conference on Computer Vision Theory and Applications* (VISAPP-2014), pages 378-385
ISBN: 978-989-758-003-1

face recognition performances. Examples of emotion classification methodologies that use geometric features extraction are (Cheon and Kim, 2009; Niese et al., 2012; Gang et al., 2009; Hammal et al., 2007; Seyedarabi et al., 2004; Kotsia and Pitas, 2007). However, high accuracies on emotion detection usually require a calibration with a neutral face ((Kotsia and Pitas, 2007; Gang et al., 2009; Niese et al., 2012; Cheon and Kim, 2009; Hammal et al., 2007)), an increase of the computational cost ((Gang et al., 2009; Seyedarabi et al., 2004)), a decrease of the number of emotions detected ((Niese et al., 2012; Hammal et al., 2007)) or a manual grid nodes positioning (Kotsia and Pitas, 2007). On the other hand, appearance-based features work directly on image and not on single extracted points (e.g. *Gabor Wavelets* (Kotsia et al., 2008) and Local Binary Patterns (Shan et al., 2009) (Chatterjee and Shi, 2010)). They usually analyze the skin texture, extracting relevant features for emotion detection. Involving a higher amount of data, the appearance feature method becomes more complex than the geometric approach, compromising also the real-time feature required by the process (appearance-based features show high variability in performance time from 9.6 to 11.99 seconds (Zhang et al., 2012)). Hybrid approaches, that combine geometric and appearance extraction can be found (i.e. (Youssif and Asker, 2011)) with higher accuracies, but they are still characterized by a high computational cost. The aim of this research work is to propose a feature extraction method that provides performances comparable with appearance-based methods without compromising the real-time and automation requirements of the system. Nevertheless, we intent to solve the following main four facial emotion recognition issues (Bettadapura, 2009):

1. real-time requirement: communication between humans is a real time process with a time scale order of about 40 milliseconds (Bartlett et al., 2003);
2. capability of recognition of multiple standard emotions on people with different anthropometric facial traits;
3. capability of recognition of the facial emotions without neutral face comparison calibration;
4. automatic self-calibration capability without manual intervention.

(equivalent optimizations of these four issues can also be extracted from *Jamshidnezhad et al.*'s survey (Jamshidnezhad and Nordin, 2012)). *Real-time issue* is solved using a low complexity features extraction method without compromising the accuracy of emotion detection. In order to show the capacity of the *second issue*, we test our system on a multi-cultural database, the Radboud face database (Langner et al.,), featured with multiple emotions traits (Bettadapura, 2009). Additionally, we investigate all six universal facial expressions(Ekman and Friesen, 1978) (*Joy, Sorrow, Surprise, Fear, Disgust and Anger*) plus *Neutral* and *Contemptuous*. Regarding the *third issue*, though with slightly lower performance relative to neutral face comparison calibration, our method allows the recognition of eight different emotions without requiring any calibration process. To avoid any *manual intervention* in the localization of the seed landmarks required by our proposed geometrical features, we use as reference example in this work, a marker-less facial landmark recognition and localization software based on the *Saragih*'s *FaceTracker* (Saragih et al., 2011b). However, face recognition can be done with the use of different marker-based and marker-less systems which allow the localization of the basic landmarks defined in our system for emotion classification. Therefore, as main contribution, we defined facial features inherent to emotions and proposed a method for their extraction in real time, for further emotion recognition.

2 GEOMETRIC FACIAL FEATURES EXTRACTION METHOD

In this work, we propose a set of facial features suitable for marker-based and marker-less systems. In fact, we present an approach to extract facial features that are truly connected to facial expression. We start from a subset composed by 19 elements (see Fig. 1 and Table 1) of the 54 anthropometric facial landmarks set defined in (Luximon et al., 2011) that are usually localized using facial recognition methods.

The testing benchmark used for our extraction method is an existing marker-less system for landmark identification and localization by *Saragih et al.* (Saragih et al., 2011a). Their approach reduces detection ambiguities, presents low online computational complexity and high detection efficiency outperforming the other popular deformable real-time models to track and model non-rigid objects (Active Appearance Models (AAM) (Asthana et al., 2009), Active Shape Models (ASM) (Cootes and C.J.Taylor, 1992),

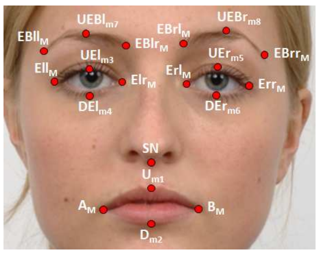

Figure 1: The subset composed by 19 points of the 66 *FaceTracker* facial landmarks used to extract our proposed geometric facial features.

Table 1: The subset of anthropometric facial landmarks used to calculate our proposed geometric facial features.

No.	Landmark	Label	Region
1	Right Cheilion	A_M	Mouth
2	Left Cheilion	B_M	Mouth
3	Labiale Superius	U_{m1}	Mouth
4	Labiale Inferius	D_{m2}	Mouth
5	Left Exocanthion	Ell_M	Left Eye
6	Right Exocanthion	Elr_M	Left Eye
7	Palpebrale Superius	UEl_{m3}	Left Eye
8	Palpebrale Inferius	DEl_{m4}	Left Eye
9	Left Exocanthion	Erl_M	Right Eye
10	Right Exocanthion	Err_M	Right Eye
11	Palpebrale Superius	UEr_{m5}	Right Eye
12	Palpebrale Inferius	DEr_{m6}	Right Eye
13	Zygofrontale	$EBll_M$	Left Eyebrown
14	Inner Eyebrown	$EBlr_M$	Left Eyebrown
15	Superciliare	$UEBl_{m7}$	Left Eyebrown
16	Inner Eyebrown	$EBrl_M$	Right Eyebrown
17	Zygofrontale	$EBrr_M$	Right Eyebrown
18	Superciliare	$UEBr_{m8}$	Right Eyebrown
19	Subnasale	SN	Nose

3D morphable models (Vetter,) and Constrained Local Models (CLMs) (Cristinacce and Cootes,)).

Saragih et al. (Saragih et al., 2011a) system identifies and localizes 66 2D landmarks on the face. Through the repetitive observation of facial behaviours during emotion expressions, we empirically choose a subset of 19 facial landmarks that better capture these facial changes among the 66 *FaceTracker* ones.

Using the landmark positions in the image space, we define two classes of features: *eccentricity* and *linear* features. These features are normalized to the range [0,1] to let the feature not affected by people anthropometric traits dependencies. So, we extract geometric relations among landmark positions during emotional expression for people with different ethnicities and ages.

2.1 Eccentricity Features

The *eccentricity* features are determined by calculating the eccentricity of ellipses constructed using specific facial landmarks. Geometrically, the *eccentricity* measures how the ellipse deviates from being circular. For ellipses the *eccentricity* is higher than zero and lower than one, being zero if it is a circle. As example, drawing an ellipse using the landmarks of the mouth, it is possible to see that while smiling the eccentricity is higher than zero, but when expressing surprise it is closer to a circle and almost zero. A similar phenomenon can be observed also in the eyebrow and eye areas. Therefore, we use the eccentricity to extract new features information and classify facial emotions. More in detail, the selected landmarks for this kind of features are 18 over 19 (see Table 1 and Fig. 1), whereas the total defined eccentricity features are eight: two in the mouth region, four in the eye region and two in the eyebrows region (more details can be found in Table 2). Now, we describe the *eccentricity* extraction algorithm applied to the mouth region. The same algorithm can be simply applied to the other face areas (eyebrowns and eyes) following the same guidelines.

With reference to Figure 2.a, let A_M and B_M be the end points of the major axis corresponding to the side ends of the mouth, while U_{m1} the upper end points of the minor axis (the distance between the major axis and U_{m1} corresponds to the semi-minor axis). Of course, the symmetry of U_{m1} with respect to A_M and B_M is not assured. For this reason, in the following, we will refer to each ellipse as the *best fitting ellipse* among the three points having the semi-minor axis equal to the distance between U_{m1} and the line $A_M B_M$.

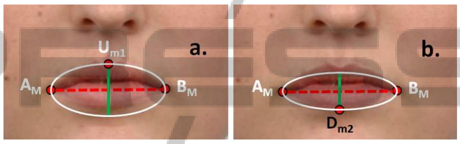

Figure 2: The definition of the first (a.), *"upper"* and the second (b.), *"lower"* ellipses of the mouth region using respectively the triple (A_M, B_M, U_{m1}) and (A_M, B_M, D_{m2}).

We construct the first ellipse E_1, named *"upper"* ellipse, defined by the triple (A_M, B_M, U_{m1}) and calculate its eccentricity e_1. The eccentricity of an ellipse is defined as the ratio of the distance between the two foci, to the length of the major axis or equivalently:

$$e = \frac{\sqrt{a^2 - b^2}}{a} \tag{1}$$

where $a = \frac{B_{Mx} - A_{Mx}}{2}$ and $b = A_{My} - U_{m1y}$ are respectively one-half of the ellipse E's major and minor axes, whereas x and y indicate the horizontal and the vertical components of the point in the image space. As mentioned above, for an ellipse, the eccentricity is in the range [0,1]. When the eccentricity is 0, the foci coincide with the center point and the figure is a circle. As the eccentricity tends toward 1, the ellipse gets a more elongated shape. It tends towards a line segment if the two foci remain a finite distance apart and a parabola if one focus is kept fixed as the other is allowed to move arbitrarily far away.

We repeat the same procedure for the ellipse E_2, named *"lower"* ellipse, using the lower end of the mouth (see Fig. 2.b). The other six ellipses are, then, constructed following the same extraction algorithm using the features summarized in Table 2 (for the landmark labels refer to Table 1 and Fig. 1). It is clear that for both eyebrows, it is not possible to calculate the lower ellipses due to their morphology. The final results of the ellipse construction can be seen in Figure 3.a, whereas in Figure 3.b it is possible to see how the eccentricities of the facial ellipses changes according to the person's facial emotion.

Table 2: The eight ellipses used to extract the eccentricity features (for the landmark labels please refere to Fig. 1).

Ellipse	Point Triple	Region
E_1	(A_M, B_M, U_{m1})	Upper mouth
E_2	(A_M, B_M, D_{m2})	Lower mouth
E_3	(Ell_M, Elr_M, UEl_{m3})	Upper left eye
E_4	(Ell_M, Elr_M, DEl_{m4})	Lower left eye
E_5	(Erl_M, Err_M, UEr_{m5})	Upper right eye
E_6	(Erl_M, Err_M, DEr_{m6})	Lower right eye
E_7	$(EBll_M, EBlr_M, UEBl_{m7})$	Left eyebrown
E_8	$(EBrl_M, EBrr_M, UEBr_{m8})$	Right eyebrown

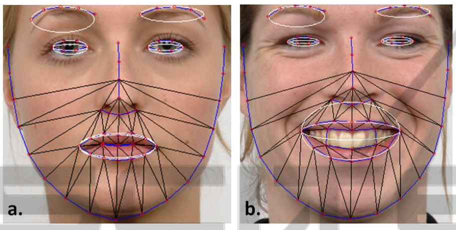

Figure 3: The final results of the eight ellipse construction (a). Eccentricities of the facial ellipses changes according to the person's facial emotion (b).

2.2 Linear Features

The *linear* features are determined by calculating linear distances between couples of landmarks normalized with respect to a physiologically greater facial inter-landmark distance. These distances intend to quantitatively evaluate the relative movements between facial landmarks while expressing emotions. The selected distances are those corresponding to the movements between eyes and eyebrows L_1, mouth and nose L_2 and upper and lower mouth points L_3. More in detail, with reference to Table 1 and Figure 1, indicating with $\_y$ only the vertical component of each point in the image space and selecting as $DEN = \overline{UEl_{m3y}SN_y}$ the normalizing distance, we calculate a total of three linear features as:

1. $L_1 = \overline{UEBl_{m7y}UEl_{m3y}}/DEN$;
2. $L_2 = \overline{U_{m1y}SN_y}/DEN$;
3. $L_3 = \overline{D_{m2y}SN_y}/DEN$;

3 EXPERIMENTAL PART

In this Section, we describe the conducted tests to evaluate the emotion recognition performances of the proposed facial geometrical features. More in detail, the *classifier validation* (Section 3.3), is related to investigate three classification methods and select the one that provides the best performances on emotion recognition using both for training and validation a particular subset of proposed features. The *feature evaluation* (Section 3.4), instead, is related to fully evaluate our proposed features using the classification method selected at the end of the first experiment. In Section 3.1, we report the organization of the defined features used in both tests, whereas, in Section 3.2, we illustrate the facial emotion database (the Radboud facial database) used to extract the defined features.

3.1 Extracted Features

In order to fully evaluate and compare our defined features, we consider five types of feature subsets:

1. only linear features (subset *S1*: 3 elements);
2. only eccentricity features (subset *S2*: 8 elements);
3. both eccentricity and linear features (subset *S3*: 11 elements);
4. differential eccentricity and linear features with respect to those calculated for neutral emotion face (subset *S4*: 11 elements);
5. all features corresponding to the union of *S3* and *S4* (subset *S5*: 22 elements).

(where the differential features are calculated as:

$$df_{i,x} = f_{i,x} - f_{i,neutral}$$

with i representing a subject of the database and x an emotion), resulting in a total number of calculated features for the entire database equal to (1.385 pictures $\times 22$ *S5* numerosity) 30.470. The five subsets can be grouped into two main classes:

1. the *intra-person-independent* or *non-differential* subsets *S1*, *S2* and *S3* that do not require any kind of calibration with other facial emotion states of the same person;
2. the *intra-person-dependent* or *differential* subsets *S4* and *S5* that require a calibration phase using the neutral expression of the same person.

3.2 Database Description

In order to demonstrate the capacity of recognition of multiple standard emotions on people with different anthropometric facial traits, we test our system on a multi-cultural database featured with multiple emotions elements. The selected testing platform is the Radboud facial database (Langner et al.,). It is composed by 67 real person's face models performing the six universal facial expressions (Ekman and Friesen, 1978) (*Joy, Sorrow, Surprise, Fear, Disgusted and Angry*) plus *Neutral* and *Contemptuous*. Even if the considered images are all frontal, for each couple person-expression, there are three pictures corresponding to slightly different angles of gaze directions, without changing head orientation. This leads to a total of 1608 (67*8*3) picture samples.

The pictures are coloured and contain both gender Caucasian and Moroccan adults and Caucasian kids. More specifically, in the database there are 39 Caucasian adults (20 males and 19 females); 10 Caucasian children (4 males and 6 females); 18 Moroccan male adults. Therefore, using this database we provide emotion expressions information relative to a population database that includes gender, ethnic and age variations combined with diverse facial positioning.This will allow us to create a model that will predict emotion expression even with this diverse changes.

To decouple the performances of our method's validation (in the scope) and those of the *FaceTracker* software (out of the paper scope), we adopt a preprocessing step. During this pre-processing we removed 223 elaborated picture samples in which the landmarks were not properly recognized by *FaceTracker* software, leading to a total number of tested pictures equal to 1385. With this outlier removal, we guarantee a correct training of the machine learning classifier, since we capture correctly the facial behaviors inherent to considered emotions.

3.3 Classifier Validation

The *classifier validation* test is subdivided into two parts:

1. the *training phase* of three emotion classification methods (k-Nearest Neighbours, Support Vector Machine and Random Forests that will be described in detail later);
2. the *classifier accuracy estimation* of the three methods in order to identify the best classification method to be used in the *second experiment*.

Both for training and for the accuracy estimation, we used only the subset *S5*, that is the most inclusive feature subset. In order to train a classifier according to supervising learning approach, we need an input dataset containing rows of features and an output class (e.g. the emotion). The trained classifier provides a model that can be used to predict the emotion corresponding to a set of features, even if the classifier did not use these combinations of features in the training process. According to (Zeng et al., 2009), the most significant classifiers that can be used for our experiment are *k-Nearest Neighbours* (Cover and Hart, 1967), *Support Vector Machine* (Amari and Wu, 1999) and *Random Forests* (Breiman, 2001).As mentioned above, as final result of the classifier validation, we will select the classification method that provides best performances on emotion recognition accuracy using only the subset *S5*.

Regarding the second part of the first test, to quantify the classification accuracy of the three presented methods, we use the K-Fold Cross Validation Method (K-Fold CRM).More in detail, the k-Fold CRM, after having iterated k times the process of dividing a database in k slices, trains a classifier with $k-1$ slices. The remaining slices are used as test sets on their respective $k-1$ trained classifier to calculate the accuracy and provides as final accuracy value the average of the k calculated accuracies.

In our case, we impose $K = 10$, because this is the number that provides statistical significance to the conducted analysis (Rodriguez et al., 2010). The accuracy estimations obtained with the three investigated methods, k-Nearest Neighbours (with $k = 1$), Support Vector Machine and Random Forests using the subset *S5* to recognize all eight emotions are the following, 85%, 88% and 89%, respectively.

Due to its better performances, we decided to use only the Random forests classifier to conduct the second experiment, that is a full analysis considering all the feature subsets and four different subsets of emotions with numerosity equal to 6, 7, 7 and 8 emotions.

3.4 Feature Evaluation

The results of the full analysis conducted using the Random Forests classifier (selected after the classifier validation test) are reported in Table 3. As expected, *S4* and *S5* provided better recognition performances with an overall accuracy increment of 6% (in the 6 emotions test) and of 9% (in the 8 emotions test) with respect to that obtained using *S3*. Furthermore, the *Neutral* expression calibration obviously increases the dissimilarity between other emotions.

Comparing the results obtained using the non-differential and differential subsets, in the latter case, it is possible to observe some improvements on the recognition of three particular emotions, *Anger*, *Neutral* and *Sorrow*. The increment of the recognition accuracy of the *Neutral* expression was expected due to the calibraton that uses the *Neutral* facial emotion. The increment in the *Anger* and *Sorrow* expressions recognition accuracy was a consequence of the better recognition of the *Neutral* expression since they were often mistaken as *Neutral*. However, we also noticed a decrease of accuracy for the *Disgust* expression recognition using the subset *S4*. In this case, the calibration reduced the *Disgust* dissimilarity in comparison with *Fear*, *Joy*, *Sorrow* and *Surprise*, resulting in misclassification towards *Surprise* expression.

An interesting result about the classifier performances using subset *S5*, is that it has proved its capacity to exploit the best aspects from the two *S5* subset's components, *S3* and *S4* to improve the emotion recognition accuracy. For example, the classifier used *S3* features to avoid the misclassification of the *Disgust* expression, typical misclassification when using only *S4* features. More in detail, we report in Table 5 and Table 6 the confusion matrices obtained with Random Forests classifier using respectively eight and six (without *Neutral* and *Contemptuous*) emotions for subsets *S3* | *S4* | *S5*. For sake of brevity, we do not report the confusion matrices obtained for the two seven-emotion tests (eight emotions except Neutral, eight emotions except •), because they provide intermediate results between those achieved for eight and six emotions.

Analysing the literature of the emotion facial recognition systems and comparing them with the obtained results reported in Table 3, we realized that the emotion recognition method based on our proposed features outperformed several alternative methods of feature extraction, presented in Table 3. We compare our method to:

- MPEG-4 FAPS (Pardàs and Bonafonte, 2002), Gabor Wavelets (Bartlett et al., 2003) and geometrical features based on vector of features displacements (Michel and El Kaliouby, 2003) methods with respect to the results obtained by Random Forests classifier using *S3*. These real time methods only classify the six universal facial expressions without using differential features with respect to *Neutral* face with an accuracy of 84%, 84% and 72%, respectively;
- three differential feature methods *Michel et al.* (Michel and El Kaliouby, 2003), *Cohen et al.* (Cohen et al., 2003) and *Wang et al.* (Wang and Yin, 2007) with respect to the results obtained by Random Forests classifier using *S5*. Also these State-of-the-Art (SoA) methods allow the detection of only six universal facial expressions with average accuracies of 73.22%, 88% and 93%, respectively.

To summarize, in Table 3, we report the performance comparison between the aforementioned emotion facial recognition methods considering only the six universal facial expressions emotions (for uniformity of comparison with SoA methods).

Finally, regarding the real-time issue of the emotion recognition system, we calculated that the mean required time (over 10^3 tries) to extract our complete proposed set of features (*S5*), once the position of facial landmarks is known, is equal to 1.9 *ms*. It follows that the working frequencies achievable for sampling and processing, especially when using marker-based landmark locators, are very high and do not compromise the real-time feature of the interaction process.

Table 3: Results using a Random Forests classifier for each dataset composed by a sub-set of features of a sub-set of emotions to classify. * means without considering contemptuous emotion, ** without considering neutral emotion, *** without considering neutral and contemptuous emotions

No. tested emotions	***S1[%]***	***S2[%]***	***S3[%]***	***S4[%]***	***S5[%]***
8	51	76	80	86	89
7*	61	80	84	88	**90**
7**	60	81	84	**90**	**92**
6***	67	87	89	**91**	**94**

Table 4: Accuracy comparison of emotion facial recognition methods(not differential or differential features) with six universal facial expressions.

Method	**Differential**	**Accuracy[%]**
Michel et al. (Michel and El Kaliouby, 2003)	No	72
Pardas et al. (Pardàs and Bonafonte, 2002)	No	84
Bartlett et al. (Bartlett et al., 2003)	No	84
Our method *S3*	No	**89**
Michel et al. (Michel and El Kaliouby, 2003)	Yes	84
Cohen et al. (Cohen et al., 2003)	Yes	88
Wang et al.(Wang and Yin, 2007)	Yes	93
Our method *S5*	Yes	**94**

4 CONCLUSIONS AND FUTURE WORK

In this paper, we propose a versatile and innovative geometric method that extracts facial features inherent to emotions. The proposed method solves the four typical emotion recognition issues and allows a high degree of accuracy on emotion classification, also when compared to complex appearance based methods. Moreover, our method versatility allows the use of different facial landmark localization techniques, both marker-based and marker-less, being a modular post-processing solution. However, it still requires that the face recognition technique presents as output a minimum number of landmarks associated to basic facial features, such as mouth, eyes and eyebrows.

Compared to traditional methods, our method allows, beyond the classification of the six universal facial expressions, the classification of two other emotions: *Contemptuous* and Neutral. Therefore, it can be considered as a complete tool that can be incorporated on facial recognition techniques for automatic and real time emotion classification of facial emotions. As concept proof, we incorporated this tool in a LIFEisGAME (Fernandes et al., 2011) game mode, where the user must match the expression asked by the game. His face is captured and emotion classified in real time. Regarding practical performance, we verified that it is more stable when we apply a neutral face calibration, classifying correctly the emotions expressed. However, it requires that the user knows how to make the expression properly. Problems regarding environment (background and illumination changes) were not addressed. Nevertheless,

Table 5: Confusion matrix with Random Forest using all eight emotions for subsets *S3* | *S4* | *S5*.

	Angry	**Cont.**	**Disgust**	**Fear**	**Joy**	**Neutral**	**Sorrow**	**Surprise**
Angry	**76 | 84 | 89**	09 | 06 | 03	02 | 03 | 02	00 | 00 | 00	00 | 00 | 00	07 | 01 | 00	05 | 06 | 06	00 | 00 | 00
Cont.	06 | 01 | 02	**73 | 77 | 82**	01 | 01 | 01	01 | 00 | 00	04 | 01 | 00	08 | 11 | 08	08 | 09 | 09	00 | 00 | 00
Disgust	05 | 03 | 01	01 | 00 | 00	**91 | 89 | 94**	00 | 01 | 01	02 | 04 | 00	01 | 01 | 01	01 | 02 | 05	00 | 01 | 00
Fear	01 | 00 | 00	00 | 02 | 01	00 | 00 | 00	**82 | 87 | 87**	00 | 00 | 00	05 | 02 | 01	05 | 04 | 05	08 | 07 | 07
Joy	02 | 00 | 00	01 | 01 | 00	01 | 04 | 02	00 | 00 | 00	**95 | 94 | 97**	01 | 00 | 00	01 | 02 | 00	00 | 00 | 00
Neutral	03 | 00 | 00	11 | 08 | 06	02 | 00 | 00	06 | 03 | 01	00 | 00 | 00	**69 | 84 | 87**	08 | 05 | 03	00 | 00 | 00
Sorrow	04 | 06 | 02	06 | 07 | 06	01 | 01 | 01	04 | 01 | 03	01 | 00 | 00	09 | 01 | 03	**75 | 84 | 85**	00 | 00 | 00
Surp.	00 | 00 | 00	00 | 00 | 00	00 | 00 | 00	10 | 07 | 06	00 | 00 | 00	01 | 00 | 00	00 | 00 | 00	**90 | 93 | 93**

Table 6: Confusion matrix with Random Forest using 6 emotions (without neutral and contemptuous) for subsets *S3* | *S4* | *S5*.

	Angry	**Disgust**	**Fear**	**Joy**	**Sorrow**	**Surprise**
Angry	**86 | 88 | 93**	03 | 05 | 01	00 | 00 | 00	00 | 00 | 00	10 | 07 | 05	00 | 00 | 00
Disgust	04 | 04 | 02	**94 | 92 | 96**	01 | 01 | 01	02 | 01 | 00	02 | 01 | 01	00 | 00 | 00
Fear	01 | 00 | 00	01 | 00 | 00	**86 | 88 | 91**	00 | 00 | 00	07 | 05 | 05	07 | 06 | 06
Joy	01 | 01 | 00	01 | 04 | 00	00 | 00 | 00	**95 | 96 | 98**	01 | 01 | 00	00 | 00 | 00
Sorrow	09 | 06 | 05	03 | 01 | 01	07 | 02 | 03	00 | 00 | 01	**82 | 91 | 90**	00 | 00 | 00
Surprise	00 | 00 | 00	00 | 00 | 00	08 | 08 | 06	00 | 00 | 00	00 | 00 | 00	**92 | 92 | 94**

our method is still restricted to emotion classification of frontal poses, being optimized for static pictures. As future work, we pretend to reduce the landmarks required for emotion classification and to automatize their detection when using unusual face recognition systems. At last but not least, we also pretend to explore sequences of images (including videos) to discover patterns that allow subtle emotions classification, overcoming the limitation of full emotion classification.

ACKNOWLEDGEMENTS

This work is supported by PERCRO Laboratory, Scuola Superiore Sant'Anna and Instituto de Telecomunicações, Fundação para a Ciência e Tecnologia (SFRH/BD69878/2010, SFRH/BD33974/2009), the projects GOLEM (ref:ref.251415, FP7-PEOPLE-2009-IAPP), LIFEisGAME (ref: UTA-Est/MAI/2009/2009) and VERE (ref:257695). The authors would like to thank to Xenxo Alvarez and Jacqueline Fernandes for their feedback and reviews.

REFERENCES

Amari, S. and Wu, S. (1999). Improving support vector machine classifiers by modifying kernel functions. *Neural Networks*, 12(6):783–789.

Asthana, A., Saragih, J., Wagner, M., and Goecke, R. (2009). Evaluating aam fitting methods for facial expression recognition. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pages 1–8. IEEE.

Bartlett, M., Littlewort, G., Fasel, I., and Movellan, J. (2003). Real time face detection and facial expression recognition: Development and applications to human computer interaction. In *Computer Vision and Pattern Recognition Workshop, 2003. CVPRW'03. Conference on*, volume 5, pages 53–53. IEEE.

Bettadapura, V. (2009). Face expression recognition and analysis: The state of the art. *Emotion*, pages 1–27.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Chatterjee, S. and Shi, H. (2010). A novel neuro fuzzy approach to human emotion determination. In *Digital Image Computing: Techniques and Applications (DICTA), 2010 International Conference on*, pages 282–287. IEEE.

Cheon, Y. and Kim, D. (2009). Natural facial expression recognition using differential-aam and manifold learning. *Pattern Recognition*, 42(7):1340 – 1350.

Cohen, I., Sebe, N., Garg, A., Chen, L., and Huang, T. (2003). Facial expression recognition from video sequences: temporal and static modeling. *Computer Vision and Image Understanding*, 91(1):160–187.

Cootes, T. and C.J.Taylor (1992). Active shape models - smart snakes. In *In British Machine Vision Conference*, pages 266–275. Springer-Verlag.

Cover, T. and Hart, P. (1967). Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on*, 13(1):21–27.

Cristinacce, D. and Cootes, T. Feature Detection and Tracking with Constrained Local Models. *Biomedical Engineering*, pages 1–10.

Ekman, P. and Friesen, W. (1978). *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, Palo Alto.

Fernandes, T., Miranda, J., Alvarez, X., and Orvalho, V. (2011). LIFEisGAME - An Interactive Serious Game for Teaching Facial Expression Recognition. *Interfaces*, pages 1–2.

Fischer, R. (2004). Automatic Facial Expression Analysis and Emotional Classification by. *October*.

Gang, L., Xiao-hua, L., Ji-liu, Z., and Xiao-gang, G. (2009).

Geometric feature based facial expression recognition using multiclass support vector machines. In *Granular Computing, 2009, GRC '09. IEEE International Conference on*, pages 318 –321.

Hammal, Z., Couvreur, L., Caplier, A., and Rombaut, M. (2007). Facial expression classification: An approach based on the fusion of facial deformations using the transferable belief model. *International Journal of Approximate Reasoning*, 46(3):542 – 567. <ce:title>Special Section: Aggregation Operators</ce:title>.

Hong, J., Han, M., Song, K., and Chang, F. (2007). A fast learning algorithm for robotic emotion recognition. In *Computational Intelligence in Robotics and Automation, 2007. CIRA 2007. International Symposium on*, pages 25–30. Ieee.

Jamshidnezhad, A. and Nordin, M. (2012). Challenging of facial expressions classification systems: Survey, critical considerations and direction of future work. *Research Journal of Applied Sciences*, 4.

Kapoor, A., Qi, Y., and Picard, R. W. (2003). Fully automatic upper facial action recognition. In *Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, AMFG '03, pages 195–, Washington, DC, USA. IEEE Computer Society.

Ko, K. and Sim, K. (2010). Development of a facial emotion recognition method based on combining aam with dbn. In *Cyberworlds (CW), 2010 International Conference on*, pages 87–91. IEEE.

Kotsia, I., Buciu, I., and Pitas, I. (2008). An analysis of facial expression recognition under partial facial image occlusion. *Image and Vision Computing*, 26(7):1052 – 1067.

Kotsia, I. and Pitas, I. (2007). Facial expression recognition in image sequences using geometric deformation features and support vector machines. *Image Processing, IEEE Transactions on*, 16(1):172 –187.

Langner, O., Dotsch, R., Bijlstra, G., and Wigboldus, D. Support material for the article : Presentation and Validation of the Radboud Faces Database (RaFD) Mean Validation Data : Caucasian Adult Subset. *Image (Rochester, N.Y.)*.

Luximon, Y., Ball, R., and Justice, L. (2011). The 3d chinese head and face modeling. *Computer-Aided Design*.

Michel, P. and El Kaliouby, R. (2003). Real time facial expression recognition in video using support vector machines. In *Proceedings of the 5th international conference on Multimodal interfaces*, pages 258–264. ACM.

Niese, R., Al-Hamadi, A., Farag, A., Neumann, H., and Michaelis, B. (2012). Facial expression recognition based on geometric and optical flow features in colour image sequences. *Computer Vision, IET*, 6(2):79 –89.

Pardàs, M. and Bonafonte, A. (2002). Facial animation parameters extraction and expression recognition using hidden markov models. *Signal Processing: Image Communication*, 17(9):675–688.

Rodriguez, J., Perez, A., and Lozano, J. (2010). Sensitivity analysis of k-fold cross validation in prediction error estimation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(3):569–575.

Saragih, J., Lucey, S., and Cohn, J. (2011a). Deformable model fitting by regularized landmark mean-shift. *International Journal of Computer Vision*, pages 1–16.

Saragih, J., Lucey, S., and Cohn, J. (2011b). Real-time avatar animation from a single image. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 117–124. IEEE.

Seyedarabi, H., Aghagolzadeh, A., and Khanmohammadi, S. (2004). Recognition of six basic facial expressions by feature-points tracking using rbf neural network and fuzzy inference system. In *Multimedia and Expo, 2004. ICME '04. 2004 IEEE International Conference on*, volume 2, pages 1219 –1222 Vol.2.

Shan, C., Gong, S., and McOwan, P. (2009). Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing*, 27(6):803–816.

Vetter, T. A Morphable Model For The Synthesis Of 3D Faces f g. *Faces*.

Wang, J. and Yin, L. (2007). Static topographic modeling for facial expression recognition and analysis. *Computer Vision and Image Understanding*, 108(1-2):19–34.

Youssif, A. A. A. and Asker, W. A. A. (2011). Automatic facial expression recognition system based on geometric and appearance features. *Computer and Information Science*, pages 115–124.

Zeng, Z., Pantic, M., Roisman, G., and Huang, T. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(1):39–58.

Zhang, L., Tjondronegoro, D., and Chandran, V. (2012). Discovering the best feature extraction and selection algorithms for spontaneous facial expression recognition. *2012 IEEE International Conference on Multimedia and Expo*.