Real-time Emotion Recognition

Novel Method for Geometrical Facial Features Extraction

Claudio Loconsole¹, Catarina Runa Miranda², Gustavo Augusto², Antonio Frisoli¹ and Verónica Orvalho²

¹PERCRO Laboratory, Scuola Superiore Sant’Anna, Pisa, Italy
²Instituto 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 (primary 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
Real-time Emotion Recognition - Novel Method for Geometrical Facial Features Extraction

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;

(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 univer-

sal 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.
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 (eyebrows 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 $U_{ml}$ the upper end points of the minor axis (the distance between the major axis and $U_{ml}$ corresponds to the semi-minor axis). Of course, the symmetry of $U_{ml}$ 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 $U_{ml}$ and the line $A_M B_M$.

We construct the first ellipse $E_1$, named “upper” ellipse, defined by the triple $(A_M, B_M, U_{ml})$ 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:

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

where $a = \frac{B_M - A_M}{2}$ and $b = A_M - U_{ml}$ 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.

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.

### 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

--

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

<table>
<thead>
<tr>
<th>No.</th>
<th>Landmark</th>
<th>Label</th>
<th>Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Right Cheilion</td>
<td>$A_R$</td>
<td>Mouth</td>
</tr>
<tr>
<td>2</td>
<td>Left Cheilion</td>
<td>$B_L$</td>
<td>Mouth</td>
</tr>
<tr>
<td>3</td>
<td>Labiale Superior</td>
<td>$U_{ls}$</td>
<td>Mouth</td>
</tr>
<tr>
<td>4</td>
<td>Labiale Inferius</td>
<td>$D_{in}$</td>
<td>Mouth</td>
</tr>
<tr>
<td>5</td>
<td>Left Exocanthion</td>
<td>$El_L$</td>
<td>Left Eye</td>
</tr>
<tr>
<td>6</td>
<td>Right Exocanthion</td>
<td>$El_R$</td>
<td>Right Eye</td>
</tr>
<tr>
<td>7</td>
<td>Palpebrale Superior</td>
<td>$UE_{ls}$</td>
<td>Left Eye</td>
</tr>
<tr>
<td>8</td>
<td>Palpebrale Inferius</td>
<td>$DE_{in}$</td>
<td>Left Eye</td>
</tr>
<tr>
<td>9</td>
<td>Left Exocanthion</td>
<td>$El_L$</td>
<td>Right Eye</td>
</tr>
<tr>
<td>10</td>
<td>Right Exocanthion</td>
<td>$El_R$</td>
<td>Right Eye</td>
</tr>
<tr>
<td>11</td>
<td>Palpebrale Superior</td>
<td>$UE_{rs}$</td>
<td>Right Eye</td>
</tr>
<tr>
<td>12</td>
<td>Palpebrale Inferius</td>
<td>$DE_{rn}$</td>
<td>Right Eye</td>
</tr>
<tr>
<td>13</td>
<td>Zygofrontale</td>
<td>$EB_{lg}$</td>
<td>Left Eyebrow</td>
</tr>
<tr>
<td>14</td>
<td>Inner Eyebrown</td>
<td>$EB_{lg}$</td>
<td>Left Eyebrow</td>
</tr>
<tr>
<td>15</td>
<td>Superciliarie</td>
<td>$UE_{lg}$</td>
<td>Left Eyebrow</td>
</tr>
<tr>
<td>16</td>
<td>Inner Eyebrown</td>
<td>$EB_{lg}$</td>
<td>Right Eyebrow</td>
</tr>
<tr>
<td>17</td>
<td>Zygofrontale</td>
<td>$EB_{lg}$</td>
<td>Right Eyebrow</td>
</tr>
<tr>
<td>18</td>
<td>Superciliarie</td>
<td>$UE_{lg}$</td>
<td>Right Eyebrow</td>
</tr>
<tr>
<td>19</td>
<td>Subnasale</td>
<td>SN</td>
<td>Nose</td>
</tr>
</tbody>
</table>

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.

---

**Figure 2**: The definition of the first (a.) and “upper” and the second (b.) “lower” ellipses of the mouth region using respectively the triple $(A_M, B_M, U_{ml})$ and $(A_M, B_M, B_{ml})$. 

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

<table>
<thead>
<tr>
<th>Ellipse</th>
<th>Point Triple</th>
<th>Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>E₁</td>
<td>(AS₁, BM₁, U₁m₁)</td>
<td>Upper mouth</td>
</tr>
<tr>
<td>E₂</td>
<td>(AS₂, BM₂, D₂m₂)</td>
<td>Lower mouth</td>
</tr>
<tr>
<td>E₃</td>
<td>(EH₃, ER₃, UE₃m₃)</td>
<td>Upper left eye</td>
</tr>
<tr>
<td>E₄</td>
<td>(EL₄, ER₄, DE₄m₄)</td>
<td>Lower left eye</td>
</tr>
<tr>
<td>E₅</td>
<td>(EL₅, ER₅, UE₅m₅)</td>
<td>Upper right eye</td>
</tr>
<tr>
<td>E₆</td>
<td>(EL₆, ER₆, DE₆m₆)</td>
<td>Lower right eye</td>
</tr>
<tr>
<td>E₇</td>
<td>(EB₁l₇, EB₁r₇, UE₁m₇)</td>
<td>Left eyebrow</td>
</tr>
<tr>
<td>E₈</td>
<td>(EB₂l₈, EB₂r₈, UE₂m₈)</td>
<td>Right eyebrow</td>
</tr>
</tbody>
</table>

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 L₁, mouth and nose L₂ and upper and lower mouth points L₃. 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 \( DEN = UEl_{m₃}ySN_y \) the normalizing distance, we calculate a total of three linear features as:

1. \( L₁ = UEB₁m₇y/\text{DEN} \);
2. \( L₂ = U_{m₈}ySN_y/\text{DEN} \);
3. \( L₃ = D_{m₉}ySN_y/\text{DEN} \);

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 \( \times 22 \times 35 \) 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 pre-processing 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 calibration 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 Ran-
dom 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 S5. 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 10³ 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.

### 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.

<table>
<thead>
<tr>
<th></th>
<th>Angry</th>
<th>Disgust</th>
<th>Fear</th>
<th>Joy</th>
<th>Neutral</th>
<th>Sorrow</th>
<th>Surprise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angry</td>
<td>186</td>
<td>88</td>
<td>93</td>
<td>01</td>
<td>00</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>Disgust</td>
<td>04</td>
<td>04</td>
<td>02</td>
<td>94</td>
<td>92</td>
<td>96</td>
<td>00</td>
</tr>
<tr>
<td>Fear</td>
<td>01</td>
<td>01</td>
<td>00</td>
<td>01</td>
<td>00</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>Joy</td>
<td>01</td>
<td>01</td>
<td>00</td>
<td>01</td>
<td>00</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>Sorrow</td>
<td>09</td>
<td>06</td>
<td>05</td>
<td>03</td>
<td>01</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>Surprise</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>Angry</th>
<th>Disgust</th>
<th>Fear</th>
<th>Joy</th>
<th>Sorrow</th>
<th>Surprise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angry</td>
<td>186</td>
<td>88</td>
<td>93</td>
<td>01</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>Disgust</td>
<td>04</td>
<td>04</td>
<td>02</td>
<td>94</td>
<td>92</td>
<td>96</td>
</tr>
<tr>
<td>Fear</td>
<td>01</td>
<td>01</td>
<td>00</td>
<td>01</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>Sorrow</td>
<td>09</td>
<td>06</td>
<td>05</td>
<td>03</td>
<td>01</td>
<td>00</td>
</tr>
</tbody>
</table>

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:251415, FP7-PEOPLE-2009-IAPP), LIFEisGAME (ref: UTA-Est/MAI/2009) and VERE (ref:257695). The authors would like to thank to Xenxo Alvarez and Jacqueline Fernandes for their feedback and reviews.

REFERENCES


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.)*.


Vejet, T. A Morphable Model For The Synthesis Of 3D Faces E g. Faces.


385