Facial Expressions: Discriminability of Facial Regions and Relationship to Biometrics Recognition

Elisa Barroso
University of Beira Interior
Department of Computer Science
6200, Covilhã, Portugal
ebarroso@di.ubi.pt

Gil Santos
University of Beira Interior
Department of Computer Science
6200, Covilhã, Portugal
gsantos@di.ubi.pt

Hugo Proença
University of Beira Interior
Department of Computer Science
IT - Instituto de Telecomunicações
Covilhã, Portugal 6200
hugomcp@di.ubi.pt

***Abstract*—Facial expressions result from movements of muscular action units, in response to internal emotion states or perceptions, and it has been shown that they decrease the performance of face-based biometric recognition techniques. This paper focuses in the recognition of facial expressions and has the following purposes: 1) confirm the suitability of using dense image descriptors widely known in biometrics research (e.g., local binary patterns and histograms of oriented gradients) to recognize facial expressions; 2) compare the effectiveness attained when using different regions of the face to recognize expressions; 3) compare the effectiveness attained when the identity of subjects is known / unknown, before attempting to recognize their facial expressions.**

***Index Terms*—Facial Expressions, Biometric Recognition, Performance Analysis.**

I. Introduction

The recognition of facial expressions has been motivating growing research efforts in recent years and benefited from advances in machine learning, image processing, and human cognition domains. Facial expressions constitute responses to internal emotion states, intentions, or social environment. They may be intentional or without conscious control and are produced by the synergistic or co-operative action of various facial muscles, as illustrated and described in fig. 1. Another interesting property is their universality: Paul Ekman studied the nature of facial expression and concluded that all humans are able to identify enjoyment, surprise, sadness, anger, fear, disgust. Also, when a set of volunteers was asked to make facial expressions to depict various scenarios, they were unmistakable [1].

The recognition of facial expressions is used to study facial behavior and several observational coding systems for that purpose were previously proposed, such as the *Facial Affect Scoring*($FAST$) [1], the *Facial Action Coding System* ($FACS$) [2], the *Emotional Facial Action Coding* system ($EMFACS$) [3] and *Facial Expression Coding* system ($FACES$) [2] Most of these are based in six discrete emotions: happiness / joy, sadness, anger, fear, surprise and disgust. Also, methods due to Matias *et* al. [4], Matsumoto *et* al. [5] and Coan and Gottman [6] are used in infants to detect and track their facial affect behavior.

The recognition of facial expressions mainly evolves two types of techniques: dense appearance descriptors and statistical machine learning techniques. In particular, local binary patterns (LBP) [8], histograms of oriented gradients (HoG) [9] and scale invariant feature transform (SIFT) [10] were successfully applied to this problem. HoG [9] describe local object appearances and shapes by distribution of local intensity gradients or edge directions. LBPs [8] describes the pixels of an image by thresholding the neighborhood of each pixel with the value of the centre point and using these binary numbers to construct a label. SIFT [10] is a widely used local descriptor that starts by localizing key points with the local scale-space maxima of difference-of-Gaussian (DoG), and - subsequently - uses such keypoints as reference to generate a 3D histograms of gradient locations and orientations. Also, various classifiers were used, such as neural networks (NN) [11], support vector machines (SVM) [12], linear discriminant analysis (LDA) [13], K-nearest neighbors (KNN), multinomial logistic ridge regression(MLR) and Hidden Markov models (HMM) [14].

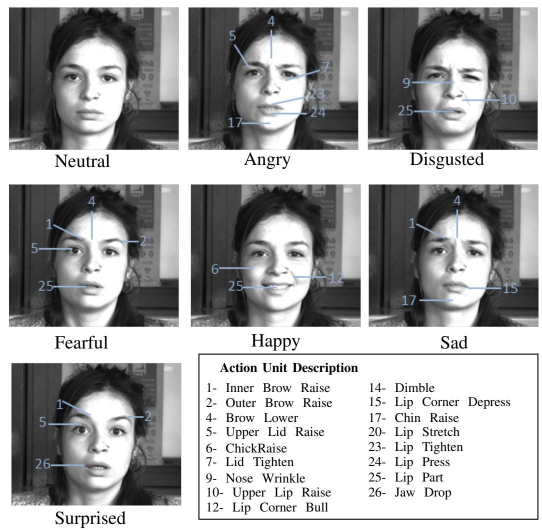

Fig. 1. Targeted action units for the emotional expressions considered in this work, as suggested by Root and Stephens [7]

According to the above, this paper mainly focuses in the recognition of facial expressions and in the suitability of using different facial regions for that task. Our work plan was divided into three main phases: 1) we started by confirming the suitability of fusing dense global and local image descriptors in the recognition of facial expressions; 2) we analyzed the effectiveness attained when using the *mouth*, the *periocular region*, the *whole face* and the *mouth* plus *periocular region* regions fused at the feature level was compared; and 3) we assessed the improvements in performance that are due to knowing subjects identity before recognizing their facial expressions. To accomplish this plan, we starting by defining the regions-of-interest (manually), and proceed for feature encoding according to the three feature extraction techniques. Then, for dimensionality reduction purposes, the analysis of data principal components (PCA) [13], [15] was carried out. Finally, feed-forward Neural Networks (NN) [11], [12] were used for classification purposes.

The remainder of this paper is organized as follows. A detailed description of the used dataset is given in Section II. Section III reports our experiments and discusses the results. Finally, Section IV presents the conclusions.

II. FaceExpressUBI Dataset

The *FaceExpressUBI* dataset was used as main data source for experiments. It contains 90 160 color images acquired using a video camera, from 184 subjects (490 per subject), with resolution of 2056 x 2452 pixels. Each file is associated to a text/annotation file that contains the coordinates of the face, periocular region, nose and mouth, respectively. Similarly to the majority of similar data sets, seven facial expressions were considered: happiness/joy, sadness, anger, fear, surprise and disgust plus the neutral expression. The dataset contains material from two imaging sessions: volunteers were 10 to 48 years of age, 35% female, 93% Caucasian European, 3% Latin-American, 1% Asian and 3% African. The number of participants with eyeglasses were 21 (12%). Furthermore, each expression was recorded during 5 seconds with a rate frame of 7 fps. Acquisition session were separated by at least two weeks for any subject on the data set. Also, from the first to the second session, the location and orientation of the acquisition device, and the artificial light sources were changed in order to increase the heterogeneity.

III. Experiments and Discussion

In our experiments 2 652 images of the FaceExpressUBI dataset were used. They include seven facial expressions and were selected according to the evidence of the facial expressions they correspond to. According to the annotation files, the regions of interest that comprise the *mouth*, *periocular region* and *face* were cropped and normalized for a constant size, using bi-cubic interpolation techniques. Then, due to the intrinsic properties of two of the feature encoding techniques used (LPBs and HOGs), data was sub-divided into square patches, as detailed in Table I.

The cohesive perspective of our experiments is given in figure 2. We used the LBP, HoG and SIFT descriptors to extract features from each region-of-interest. Then, PCA was used for dimensionality reduction and a feed forward neural network used for classification purposes. In this case, the problem was regarded as a binary classification task: for each pair of images that regard the same facial expression, a *positive* response from the neural network should be given, whereas pairs of images that regard different facial expressions should output a *negative* response.

TABLE I
Description of the pre-processing/size changes in the input images.

Anatomic Regions	Resize	Number of Blocks	Block Size
Mouth	45×54	5×6	9×9
Periocular	36×45	4×5	9×9
Face	54×54	6×6	9×9

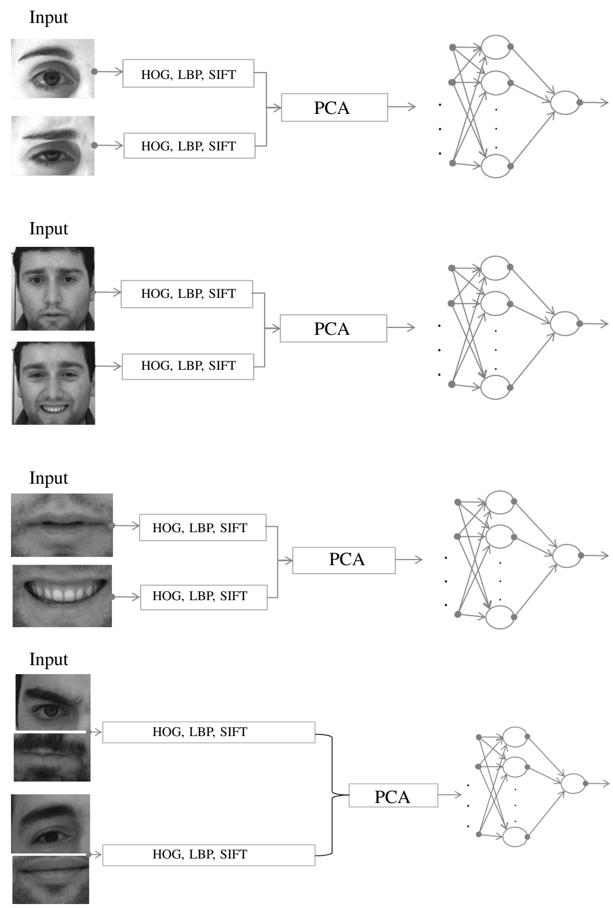

Fig. 2. Cohesive perspective of our experiments, in order to assess the discriminating ability of each region of the face to recognize facial expressions: a) using exclusively the periocular region; b) using the whole face; c) using mouth; and d) using mouth and periocular region fused at the feature level.

The used feature encoding strategies projected each region of interest into feature spaces of dimension 961 for the case of mouth, 641 for the periocular region, 1 153 for the face and

1 602 for mouth + periocular region. Then, as above stated, PCA was used for dimensionality reduction purposes, enabling projections to hyper-spaces of dimension 500, 330, 600, and 800 components respectively for the mouth, periocular, face, and mouth + periocular regions. The number of components used per region corresponds to the set that explained at least 98% of the information in the initial set (figure 3).

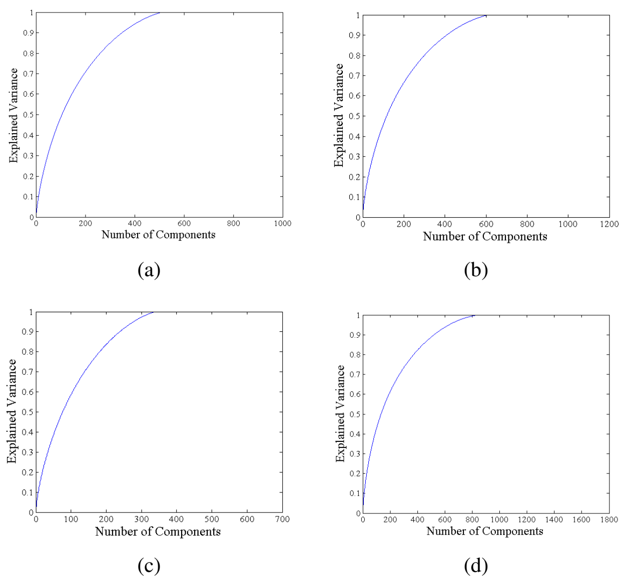

Fig. 3. Number of principal components selected for each region analyzed: a) Mouth; b) Face; c) Periocular; and d) Mouth + Periocular Region (d).

Finally, in the classification phase, data was divided into three disjoint subsets: training (70%), validation (15%) and test (15%) sets. A set of feed-forward neural networks with three layers was created, using Levenberg-Marquardt back-propagation, and varying the number of neurons in the hidden layer (between 50% and 200% of the dimension of the feature space). As stoping criterium for the learning process, a maximum number of 50 validation checks was used. Due to the non-deterministic property of neural networks, the best observed configuration was repeated twenty times for every region-of-interest and the median error rate taken.

In order to perceive the variance in performance when the identity of subjects is known / unknown, experiments were repeated in two different setups: 1) at first, only comparisons between facial expressions of the same subject were considered, corresponding to the setup where the identity of the subjects is known before attempting to recognize their facial expressions; 2) then, the identity constraint was relaxed, and comparison between facial expressions of different subjects were also considered.

A. *Setup 1: Knowing Subjects Identity*

By selecting the data frames where facial expressions are most evident in our data set, 33 306 pair wise image comparisons were considered, from which 5 124 were randomly sampled. This step was due to the computational burden of neural networks to learn in such high dimensional hyper-spaces and to the observation that results tend to maintain relatively stable when more than a few thousand of comparisons were used in the learning processes. Classes were balanced, meaning that the number of pair wise comparisons that regard the same facial expressions is equal to the number of comparisons that regard different facial expressions. Also, in order to perceive the discriminability of each facial region *per expression* this experiment was repeated when considering all facial expressions and each one separately, yielding seven variants of the experiment. At first, we attempted to recognize any facial expressions and then, exclusively attempted to recognize one at a time (among happy, anger, sad, surprise, disgust and fear). Results are given in Table II and the most interesting conclusions highlighted in bold font: the mouth outperformed in the recognition of the Happy expression, which is not too surprising due to the action units evolved in that expression. The whole face obtained the best error rates only twice: when considering all facial expressions and for the Anger expression. Interestingly, the fusion at the feature level of mouth + periocular attained the best results most times (three). In opposition, a surprising observation was the low levels of performance attained by mouth + periocular when attempting to recognize all facial expressions, which was explained due to the sparsity of instances in the feature space of higher dimension, when compared to the remaining ROIs. It should be stressed that in this experiment, only comparisons that regard facial expressions from the same subjects were selected, corresponding to the scenario where a biometric recognition system performs before the facial expression

TABLE II
MEDIAN RECOGNITION RATES OBSERVED, WHEN ATTEMPTING TO RECOGNIZE ALL FACIAL EXPRESSIONS (ALL COLUMN) AND EACH ONE SEPARATELY. IN THIS CASE, THE IDENTITY OF THE SUBJECTS IS ASSUMED TO BE KNOWN BEFORE ATTEMPTING TO RECOGNIZE THEIR FACIAL EXPRESSIONS.

Region	**All**	**Happy**	**Sad**	**Surprise**	**Fear**	**Anger**	**Disgust**
Mouth	86.5	**95.5**	93.5	**95.1**	94.2	92.7	**94.4**
Periocular	90	94.8	93.9	91.8	88.7	91.1	89.8
Face	**90.1**	**94.9**	92.9	94.8	93.7	**94.7**	94.3
Mouth + Periocular	69.6	94.8	**94**	95	**94.3**	94.6	94.2

B. *Setup 2: Unknowing Subjects Identity*

This section regards an empirical setup similar to the described above, with the exception that this time the identity of subjects was not known, meaning that pair wise image comparisons between different subjects were also considered. In this case, starting from an initial number of 6 561 282 pair wise comparisons, 5 124 were randomly selected, in order to obtain confidence intervals similar to the previous experiment. Table III gives the obtained results, where the best recognition rate was obtained for the face region and happy expression (95%). Overall, a slight decrease in the effectiveness (around 3-4%) was observed when attempting to recognize facial expressions separately. The most notorious decreases in performance occurred when all facial expressions were considered, in some circumstances up to 50% of the performance observed for the knowing identity setup. This leaded us to conclude that biometric recognition techniques contribute for consistent improvements in the analysis of subjects facial expressions.

In summary, based on the observed error rates, we concluded that positive expressions (happy and surprise) are easier to recognize than negative expressions (sad, anger, fear and disgust). Also, for most cases, the fusion at the feature level of both the mouth and periocular region did not contributed for consistent improvements in performance. Even though, using exclusively sub-parts of the face, as the mouth, lead to performance levels similar to the attained when the whole face is considered.

TABLE III
MEDIAN RECOGNITION RATES OBSERVED, WHEN RECOGNIZING ALL FACIAL EXPRESSIONS (ALL COLUMN) AND EACH ONE SEPARATELY. IN THIS CASE, THE IDENTITY OF THE SUBJECTS IS NOT KNOWN WHEN ATTEMPTING TO RECOGNIZE THEIR FACIAL EXPRESSIONS.

Region	**All**	**Happy**	**Sad**	**Surprise**	**Fear**	**Anger**	**Disgust**
Mouth	61.4	94.3	89.9	92.1	88.3	88	89.1
Periocular	58.4	93.4	87.1	86.7	84.4	82.1	84.1
Face	63.1	**95**	91.5	**93.3**	89.7	90.1	89.3
Mouth + Periocular	**65.2**	92.3	**91.9**	91	**91**	**91**	**90.7**

IV. CONCLUSIONS

This paper mainly focused on two types of analysis: 1) we compared the discriminating ability of regions of the face to the attained by using the whole face; and 2) compared the results obtained when subjects identity is previously known, in opposition to unknown identities. A data set of seven facial expressions was used and a set of regions of interest cropped, comprising the whole face, the mouth and the periocular region. Then, LBPs, HoGs and SIFTs were used for feature encoding purposes and PCA for dimensionality reduction. Finally, for each pairs of images, a feed-forward neural network binary discriminated between those that regard the same facial expression or not. This experimental setup was repeated in two different variants. At first, we assumed that the identity of subjects is previously known and only facial expressions that regard the same subject were considered. Then, this constraint was relaxed and facial expressions from different subjects were also taken into account.

Accordingly, our main conclusions are: 1) that LBP, HoG and SIFT are effective methods for feature encoding purposes in this specific purpose; 2) that the fusion of the mouth and periocular regions at the feature level does not lead to improvements in performance, when compared to using each region separately; 3) the use of the whole face in the recognition of most facial expressions does not provide better results than using exclusively regions of the face, such as the face and periocular region. Exceptions are the *happy* and *surprise*; and 4) by knowing subjects identity, consistent improvements in the recognition of their facial expressions are attained, giving support to the use of biometric recognition methods before attempting to recognize facial expressions.

ACKNOWLEDGMENT

The financial support given by “FCT - Fundação para a Ciência e Tecnologia”and “FEDER”in the scope of the PTDC/EIA/103945/2008 research project “NECOVID: Negative Covert Biometric Recognition”is acknowledge. Also, the support given by IT - Instituto de Telecomunicações in the scope of the “NOYSIRIS”research project is acknowledge too.

REFERENCES

[1] P. Ekman, “Facial Expressions,” in *Handbook of Cohnition and Emotion*, T. Dalgleish and M. Power, Eds. San Francisco, California, USA: John Wiley & Sons, 1999, ch. 16, pp. 301–320.

[2] A. M. Kring and D. M. Sloan, “The Facial Expression Coding System (FACES): development, validation, and utility.” *Psychological assessment*, vol. 19, no. 2, pp. 210–24, Jun. 2007.

[3] G. Koutlas, Anastasios (University of Ioannina and G. Fotiadis, Dimitrios (University of Ionnina, “Image Processing and Machine Learning Techniques for Facial Expression Recognition,” in *Handbook of Reasearch on Advanced Techniques in Diagnostic Imaging and Biomedical Applications*. Book News, 2009, pp. 1–16.

[4] R. Matias, J. F. Conn, and S. Ross, “A Comparison of Two Systems That Code Infant Affective Expression,” *Developmental Psychology - the American Psychological Association*, vol. 25, no. 4, pp. 483–489, 1989.

[5] D. Matsumoto, P. Ekman, and A. Fridlund, “Analyzing Nonverbal Behavior,” in *Practical Guide to Using Video in the Behavioral Sciences*. New York: John Wiley & Sons, 1991, ch. 10, pp. 153–165.

[6] J. A. Coan and J. M. Gottman, “The Specific Affect Coding System (SPAFF),” in *Handbook of Emotion Elicitation and Assessment*, J. A. Coan and J. J. B. Allen, Eds. New York: Oxford University Press, 2007, ch. 16, pp. 106–123.

[7] A. A. Root and J. A. Stephens, “Organization of the central control of muscles of facial expression in man.” *The Journal of physiology*, vol. 549, no. 1, pp. 289–98, May 2003.

[8] D. Huang and C. Shan, “Local Binary Patterns and Its Application to Facial Image Analysis: A Survey,” *IEEE Transactions on Systems, Man and Cybernetics: Part C*, vol. 41, no. 6, pp. 765–781, 2011.

[9] M. Dahmane and J. Meunier, “Emotion recognition using dynamic grid-based HoG features,” *Face and Gesture 2011*, pp. 884–888, Mar. 2011.

[10] H. Soyel and H. Demirel, “Facial expression recognition based on discriminative scale invariant feature transform,” *Electronics Letters*, vol. 46, no. 5, p. 343, 2010.

[11] L. Ma and K. Khorasani, “Facial expression recognition using constructive feedforward neural networks,” *IEEE Transactions on Systems, Man, and Cybernetics: Part B*, vol. 34, no. 3, pp. 1588–1595, 2004.

[12] I. Kotsia and I. Pitas, “Facial expression recognition in image sequences using geometric deformation features and support vector machines,” *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 172–187, 2007.

[13] G. Bai, W. Jia, and Y. Jin, “Facial Expression Recognition Based on Fusion Features of LBP and Gabor with LDA,” *2009 2nd International Congress on Image and Signal Processing*, pp. 1–5, Oct. 2009.

[14] Y.-l. Tian, T. Kanade, and J. F. Cohn, “Facial expression analysis,” *Handbook of Face Recognition*, vol. 3, no. 5, pp. 247–275, 2005.

[15] M. Uddin, J. Lee, and T.-S. Kim, “An enhanced independent component-based human facial expression recognition from video,” *IEEE Transactions on Consumer Electronics*, vol. 55, no. 4, pp. 2216–2224, Nov. 2009.