# Compound Facial Expressions in 3D Faces: Forced and Spontaneous

Daniel M. P. Parada, Flávio H. de B. Zavan, Olga R. P. Bellon and Luciano Silva
IMAGO Research Group – Universidade Federal do Paraná
Rua Francisco H. dos Santos, 100 – 81531-980 – Curitiba, Paraná, Brazil
{daniel.parada,flavio,olga,luciano}@ufpr.br

***Abstract*—This work investigates Compound Facial Expression (CFE) recognition in 3D images captured in two domains: forced and spontaneous. While no public datasets containing 3D faces and explicit CFE annotation data is available, such expressions can be inferred from the active actions units (AUs) present in the face. An well-known 3D AU detector, based on Local Depth Binary Patterns, was studied to develop a method for automatically determining the AUs of the face, allowing for CFEs to be recognized. Benchmarks are conducted on two public 3D datasets: Bosphorus and BP4D-Spontaneous. Results confirm the efficacy of the proposed approach, exploiting the robustness to variations in lighting, head pose and changes in facial appearance found in 3D images. Experiments indicate that CFEs are more likely to be found in spontaneous basic expressions, rather than on forced expressions. To the best of our knowledge, this is the first time that CFEs are studied in 3D images.**

## I. Introduction

The main non-verbal communication pathway in the human species happens through the interpretation and demonstration of facial muscle patterns [1], [2]. Individuals, regardless of their cultural background and characteristics, present non-verbal communication and express emotions through natural movements, which involve contracting groups of face muscles. These configurations are important non-verbal components in communication.

The psychological state of a person during communication can be inferred from their facial expression, playing a major role when conveying a message. Therefore, the study of these expressions is important in different areas, including psychotherapy, education, and graphics animation [3]. As a result, automatic methods for recognizing facial expression in images have been studied and devised [4].

While multiple works focus on analyzing and detecting basic facial expressions (happy, sad, fear, disgust, surprise and angry) [5], [6], [7], Du *et al.* [8], [9] defined the concept of Compound Facial Expressions (CFEs). CFEs represent a combination of the basic expressions (*e.g.* happily surprised, happily disgusted, and sadly fearful), resulting in a much larger set of emotions.

So far, most published works have focused on 2D images [4], [8], [10], [11], despite their limitations due to variations in pose, lighting and other changes in facial appearance [12]. In order to deal with these problems, the use of 3D images has been considered for analyzing expressions [13]. Sandbach *et al.* [14] proposed the detection of individual Action Units (AUs) that compose basic expressions for performing classification in depth images.

Another important distinction is between spontaneous and forced (posed) facial expressions. According to Zhang *et al.* [15], these differ in several dimensions, including complexity, time and intensity. Du *et al.* [9] suggest that the neural systems involved in the production of forced and spontaneous expressions are different, but that the standard AU configurations in CFEs are the same.

Recently, an increased interest in CFEs by the scientific community has became apparent. Publications include automatic recognition in 2D images [11], human recognition of computer generated 3D expressions [16], automatic annotation of 2D images [17], and an annotated 2D image dataset [18]. However, to the best of our knowledge, CFE recognition has not been explored in 3D images. Therefore, we study the presence of CFEs in publicly available 3D datasets, conceived with basic expressions in mind, and propose an approach for automatically recognizing CFEs in the 3D space, both for forced and spontaneous expressions.

The proposed approach is based on identifying individual AUs, applying a method based on [14], and inferring the CFE from their configuration. Both steps, detecting the AUs and recognizing the CFE, were tested on all publicly available 3D datasets with AU annotations, Bosphorus [3] and BP4D-Spontaneous [15], each representing a different a domain.

The rest of this paper is organized as follows: Section II describes the methodology for recognizing CFEs, detailing all steps; Section III presents the performed experiments, their results and a discussion on the findings; the final remarks are presented in Section IV.

## II. Proposed Approach

According to Liu *et al.* [19], facial expressions recognition methods can be classified into two groups: action unit based (those that consider individual components) and appearance based (those that consider the entire face region). We explore AU detection for the task, as there are, currently, no 3D datasets with explicitly annotated CFEs.

The proposed approach for CFE recognition in 3D faces is performed in two stages. First, a state-of-the-art AU detector is applied [14], then the resulting set of AUs is used for inferring the CFE. Both steps are explained in this section, including our implementation of the detector with all parameters and

ISBN 978-90-827970-1-5 © EURASIP 2018 

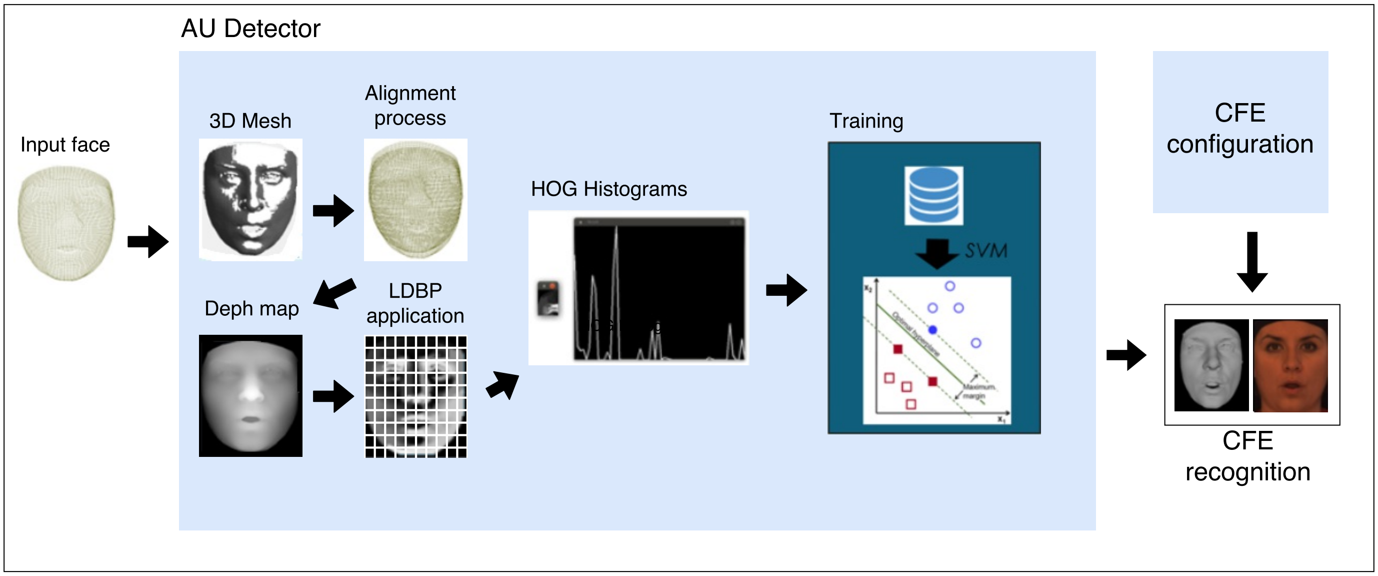


Fig. 1. The proposed CFE recognition approach, including the steps from Sandbach *et al.*'s AU detector [14]

decisions made to address the missing details in the original paper. A general overview of the approach is available in Figure 1.

### A. Local Depth Binary Patterns (LDBP) for AU detection

AU detection is performed according to Sandbach *et al.*'s method [14], via the following steps: noise-reduction; alignment; rendering; segmentation; noise-reduction; feature extraction; training or estimation.

Bosphorus [3] and BP4D-Spontaneous [15] each present the 3D information in a different format, therefore, the regular point cloud found in Bosphorus is triangulated, resulting in a mesh, similar to those found in BP4D-Spontaneous. To reduce noise, Laplacian smoothing [20] is applied to the meshes.

Alignment is performed by applying an affine transformation to the mesh. This transformation is defined according to the position of the annotated fiducial points of the face being processed and those of a, hand chosen, near frontal face (one for each dataset). While the faces in Bosphorus are already segmented, background and body data is removed from the images in BP4D-Spontaneous using Segundo *et al.*'s method [21]. The resulting mesh is rendered onto a depth image (220x300 pixel) using a planar projection, holes are filled using bicubic interpolation. Finally, the median filter of size five is applied to the image and the face bounding box is cropped. Examples of the resulting depth maps can be found in Figure 2.

Feature extraction is performed by applying the Local Depth Binary Patterns (LDBP) filter [14] on the depth images. The resulting image is divided into a 10x10 grid and the feature vector is defined by the concatenation of the normalized Histograms of Oriented Gradients (HOG) [22] of all subregions. Training and predicting is performed using numerous binary Support Vector Machines (SVM) [23], one for each AU. The histogram intersection kernel and 10-fold cross-validation were used.

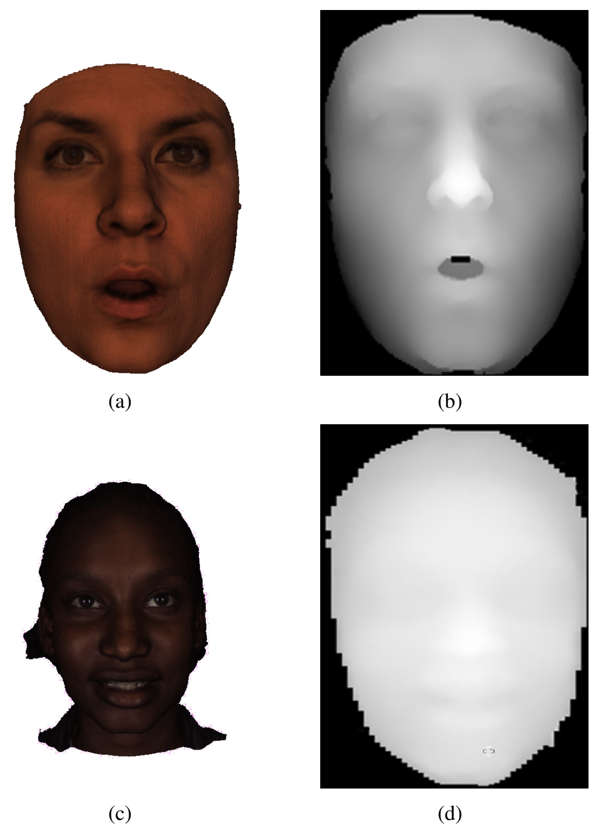


Fig. 2. Examples of 3D images from both datasets and resulting depth maps after preprocessing the 3D faces. (a) represents the original mesh from the Bosphorus dataset and (b), its resulting depth map. (c) is the original mesh from BP4D-Spontaneous and (d), its depth map

ISBN 978-90-827970-1-5 © EURASIP 2018

*B. Automatic recognition of Compound Facial Expressions in 3D images*

After the AUs are detected, CFEs can be easily inferred based on the AU configuration found in each face. Du *et al.* [8], [9] explicitly defined all CFEs in terms of AUs, which is used for translating a set of active AUs to the corresponding CFE. If the detected set or the ground-truth AU annotation does not configure an existing CFE, the face is classified as having an unknown expression, as neutral faces should not present active action units. Finally, the proposed method is unable to recognize between three CFEs, namely sadly disgusted, appalled and hateful, as they are all characterized by the presence of both AU 4 and 10, the only difference being their intensity, which is not addressed in this work. Therefore these three compound expressions are mapped to a single class.

## III. Experimental Results

Experiments were performed on the Bosphorus dataset [3], containing forced expressions, and on the BP4D-Spontaneous dataset [15], containing spontaneous expressions. Once the projection and all depth maps were finalized, a total of 2,900 frontally aligned depth maps from the Bosphorus dataset was obtained. On BP4D-Spontaneous, 10,000 random faces were chosen for the experiment and their corresponding depth map was generated. On both cases, training was performed with 90% of the faces and the rest was used for testing. When training each AU classifier, all positive training samples of that AU are used, and an equal number of negative samples is randomly chosen, preserving class balance.

To perform the task of recognizing the CFEs in 3D faces, all action units that compose them were considered for detection, namely 1, 2, 4, 5, 6, 10, 12, 15, 17, 20, 25 and 26. Ground-truth CFE data is generated by simply mapping the ground-truth set of active AUs to their corresponding CFE. For example: if the AUs 1, 2, 5 and 25 are present in the face, the awed expression is recognized.

AU detection and CFE recognition rates are reported separately, allowing for a deeper understanding of the performance and limitations of both our implementation and the proposed approach. Individual AU detection rates for both datasets are presented in Table I, the average detection rate on Bosphorus was 76.4% and 86.3% on BP4D-Spontaneous, suggesting a better performance with spontaneous expressions.

Because the datasets were not explicitly designed to include CFEs, an initial evaluation of the CFE inference step was performed using the ground-truth AU annotations. Results reveal that only approximately 1.5% of the all faces present compound expressions, with not all classes being represented (only eight out of 15), we believe this is due to the forced nature of the acquisition process, in which the subjects were asked to simulate basic expressions. In a spontaneous situations, the BP4D-Spontaneous dataset, subjects produced the expressions naturally, resulting in a wide range of nuances and over 41% of the captured faces containing CFEs, covering the whole possible range of combinations.

TABLE I
Individual hit rate for each AU on each dataset

| AU | Bosphorus | BP4D-Spontaneous |
|---|---|---|
| 1 | 75.0% | 84.7% |
| 2 | 67.2% | 83.9% |
| 4 | 76.9% | 85.1% |
| 5 | 73.7% | 85.0% |
| 6 | 79.6% | 88.7% |
| 10 | 76.3% | 87.8% |
| 12 | 82.2% | 92.9% |
| 15 | 68.4% | 84.1% |
| 17 | 76.8% | 79.0% |
| 20 | 75.8% | 90.9% |
| 25 | 92.4% | 95.9% |
| 26 | 72.3% | 84.1% |
| Average | **76.4%** | **86.3%** |

When inferring the CFEs directly from the automatically detected AUs, eight of the nine expression classes present in Bosphorus were found, and 11 of the 16 in BP4D-Spontaneous were represented. An overall accuracy of 84.83% was achieved on Bosphorus and of 78.5% on BP4D-Spontaneous. While inferior, the results on the spontaneous dataset provide better insight on the performance of the proposed approach, as the class distribution is not highly unbalanced. Table II presents all possible CFEs and their class numbers, Tables III and IV present the estimation results as confusion matrices.

TABLE II
All CFEs and their class numbers

| # | Expression |
|---|---|
| 0 | Unknown |
| 1 | Happily Surprised |
| 2 | Happily Disgusted |
| 3 | Sadly Fearful |
| 4 | Sadly Angry |
| 5 | Sadly Surprised |
| 6 | Special Case |
| 7 | Fearfully Angry |
| 8 | Fearfully Surprised |
| 9 | Fearfully Disgusted |
| 10 | Angrily Surprised |
| 11 | Angrily Disgusted |
| 12 | Disgustedly Surprised |
| 13 | Awed |
| 14 | Happily Fearful |
| 15 | Happily Sad |

## IV. Final Remarks

An approach for automatically recognizing CFEs in 3D faces was presented. AUs detected with a state-of-the-art method are used to infer the compound expressions, based on the configuration that defines them. While there are currently no public 3D datasets that include explicit CFE annotations, our experiments indicate that such expressions can appear in spontaneous situations, indicating that the BP4D-Spontaneous dataset [15] is suited for evaluating CFEs in 3D faces. In future work, the possibility of recognizing CFE directly from the 3D face can be explored, as well as the possibility of devising a

ISBN 978-90-827970-1-5 © EURASIP 2018

TABLE III
CONFUSION MATRIX WHEN AUTOMATICALLY RECOGNIZING CFES IN BOSPHORUS

| Truth \ Prediction | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 86.7 | 0.4 | 3.2 | | 1.8 | 0.4 | 2.2 | 0.4 | | | 0.7 | 0.7 | 0.4 | 2.5 | 0.7 | |
| 1 | | | | | | | | | | | | | | | | |
| 2 | | | 100 | | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | | | | | |
| 4 | 100 | | | | | | | | | | | | | | | |
| 5 | 50.0 | | | | | | | | | | | | | 50.0 | | |
| 6 | | | | | | | 66.7 | | | | | 33.3 | | | | |
| 7 | | | | 50.0 | | | | | | 50.0 | | | | | | |
| 8 | | | | | | | | | | | | | | | | |
| 9 | | | | | | | | | | | | | | | | |
| 10 | | | | | | | | | | | 100 | | | | | |
| 11 | 100 | | | | | | | | | | | | | | | |
| 12 | | | | | | | | | | | | | | | | |
| 13 | | | | | | | | | | | | | | | | |
| 14 | | | | | | | | | | | | | | | | |
| 15 | | | | | | | | | | | | | | | | |

TABLE IV
CONFUSION MATRIX WHEN AUTOMATICALLY RECOGNIZING CFES IN BP4D-SPONTANEOUS

| Truth \ Prediction | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 82.8 | 1.1 | 4.7 | | 1.9 | 1.0 | 1.0 | 0.8 | 0.2 | 0.6 | 1.3 | 3.4 | 0.5 | 0.2 | 0.3 | 0.3 |
| 1 | 16.7 | 62.5 | 20.8 | | | | | | | | | | | | | |
| 2 | 11.2 | 5.1 | 73.6 | 0.4 | | 1.1 | | 1.1 | | 1.4 | | | | | 3.6 | 2.5 |
| 3 | | | | | | | | | | | | | | | | |
| 4 | | | | | 100 | | | | | | | | | | | |
| 5 | 20.0 | | | | | 80.0 | | | | | | | | | | |
| 6 | 28.6 | | | | 28.6 | | 42.9 | | | | | | | | | |
| 7 | | | 33.3 | | | | | 66.7 | | | | | | | | |
| 8 | | | | | | | | | | | | | | | | |
| 9 | | | | | | | | | | | | | | | | |
| 10 | | | | | 14.3 | | | | | 14.3 | 57.1 | | | | | 14.3 |
| 11 | 26.5 | | 5.9 | | 2.9 | | | | | | | 64.7 | | | | |
| 12 | | | | | | | | | | | | | | | | |
| 13 | | | | | | | | | | | | | | | | |
| 14 | | | 16.7 | | | | | | 8.3 | | | | | | 75.0 | |
| 15 | 40.0 | | | | | | | | | | | | | | | 60.0 |

robust method for recognizing CFEs even when not all of the relevant AUs were detected.

## ACKNOWLEDGEMENTS

The authors would like to thank CAPES and CNPq for funding this research.

## REFERENCES

[1] Paul Ekman, Wallace Friesen, and JC Hager, "Facial action coding system: The manual on cd-rom. instructor's guide," *Network Information Research Co, Salt Lake City*, 2002.

[2] Lisa A Parr and Bridget M Waller, "Understanding chimpanzee facial expression: insights into the evolution of communication," *Social Cognitive and Affective Neuroscience*, vol. 1, no. 3, pp. 221–228, 2006.

[3] Arman Savran, Neşe Alyüz, Hamdi Dibeklioğlu, Oya Çeliktutan, Berk Gökberk, Bülent Sankur, and Lale Akarun, "Bosphorus database for 3d face analysis," in *European Workshop on Biometrics and Identity Management*. Springer, 2008, pp. 47–56.

[4] R. Jameel, A. Singhal, and A. Bansal, "A comprehensive study on facial expressions recognition techniques," in *2016 6th International Conference - Cloud System and Big Data Engineering (Confluence)*. IEEE, 2016, pp. 478–483.

[5] Guillaume-Benjamin Duchenne and R Andrew Cuthbertson, *The mechanism of human facial expression*, Cambridge university press, 1990.

[6] Paul Ekman and Wallace V Friesen, *Pictures of facial affect*, consulting psychologists press, 1975.

[7] Carl-Herman Hjortsjö, *Man's face and mimic language*, Studen litteratur, 1969.

[8] Shichuan Du and Aleix M Martinez, "Compound facial expressions of emotion: from basic research to clinical applications," *Dialogues in clinical neuroscience*, vol. 17, no. 4, pp. 443, 2015.

[9] Shichuan Du, Yong Tao, and Aleix M Martinez, "Compound facial expressions of emotion," *Proceedings of the National Academy of Sciences*, vol. 111, no. 15, pp. E1454–E1462, 2014.

[10] Jyoti Kumari, R. Rajesh, and K.M. Pooja, "Facial expression recognition: A survey," *Procedia Computer Science*, vol. 58, no. Supplement C, pp. 486 – 491, 2015.

[11] Zhiwen Liu, Shan Li, and Weihong Deng, "Recognizing compound emotional expression in real-world using metric learning method," in *Chinese Conference on Biometric Recognition*. Springer, 2016, pp. 528–536.

[12] Jeffrey F Cohn, Zara Ambadar, and Paul Ekman, "Observer-based measurement of facial expression with the facial action coding system," *The handbook of emotion elicitation and assessment*, pp. 203–221, 2007.

[13] Georgia Sandbach, Stefanos Zafeiriou, Maja Pantic, and Lijun Yin, "Static and dynamic 3d facial expression recognition: A comprehensive survey," *Image and Vision Computing*, vol. 30, no. 10, pp. 683–697, 2012.

[14] G. Sandbach, S. Zafeiriou, and M. Pantic, "Binary pattern analysis for 3d facial action unit detection," in *Proceedings of the British Machine Vision Conference, BMVC 2012, Surrey, UK*. BMVA Press, 2012, pp. 119.1–119.12.

[15] Xing Zhang, Lijun Yin, Jeffrey F Cohn, Shaun Canavan, Michael Reale, Andy Horowitz, Peng Liu, and Jeffrey M Girard, "Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database," *Image and Vision Computing*, vol. 32, no. 10, pp. 692–706, 2014.

[16] Juho Kastemaa, "Recognizing compound facial expressions of virtual characters in augmented reality," 2017.

[17] C Fabian Benitez-Quiroz, Ramprakash Srinivasan, and Aleix M Martinez, "Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2016, pp. 5562–5570.

[18] Kaili Zhao, Wen-Sheng Chu, Fernando De la Torre, Jeffrey F Cohn, and Honggang Zhang, "Joint patch and multi-label learning for facial action unit detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2015, pp. 2207–2216.

[19] Zhiwen Liu, Shan Li, and Weihong Deng, "Boosting-poof: Boosting part based one vs one feature for facial expression recognition in the wild," in *Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on*. IEEE, 2017, pp. 967–972.

[20] David A Field, "Laplacian smoothing and delaunay triangulations," *International Journal for Numerical Methods in Biomedical Engineering*, vol. 4, no. 6, pp. 709–712, 1988.

[21] Maurício Pamplona Segundo, Luciano Silva, Olga Regina Pereira Bellon, and Chauã C Queirolo, "Automatic face segmentation and facial landmark detection in range images," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 40, no. 5, pp. 1319–1330, 2010.

[22] Navneet Dalal and Bill Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. IEEE, 2005, pp. 886–893.

[23] Christopher JC Burges, "A tutorial on support vector machines for pattern recognition," *Data mining and knowledge discovery*, vol. 2, no. 2, pp. 121–167, 1998.