# Fully Automatic 3D Facial Expression Recognition using Differential Mean Curvature Maps and Histograms of Oriented Gradients

Pierre Lemaire, Liming Chen, Mohsen Ardabilian, Mohamed Daoudi

## ▶ To cite this version:

HAL Id: hal-00823903

https://hal.science/hal-00823903

Submitted on 19 May 2013

# Fully Automatic 3D Facial Expression Recognition using Differential Mean Curvature Maps and Histograms of Oriented Gradients

Pierre Lemaire, Mohsen Ardabilian,
Liming Chen
LIRIS
École Centrale de Lyon
UMR5205, F-69134, France
`{name}.{surname}@ec-lyon.fr`

Mohamed Daoudi
LIFL
Télécom Lille 1
UMR USTL/CNRS 8022, France
`daoudi@telecom-lille1.eu`

*Abstract*— **In this paper, we propose an holistic, fully automatic approach to 3D Facial Expression Recognition (FER). A novel facial representation, namely Differential Mean Curvature Maps (DMCMs), is proposed to capture both global and local facial surface deformations which typically occur during facial expressions. These DMCMs are directly extracted from 3D depth images, by calculating the mean curvatures thanks to an integral computation. To account for facial morphology variations, they are further normalized through an aspect ratio deformation. Finally, Histograms of Oriented Gradients (HOG) are applied to regions of these normalized DMCMs and allow for the generation of facial features that can be fed to the widely used Multiclass-SVM classification algorithm. Using the protocol proposed by Gong et al. [1] on the BU-3DFE dataset, the proposed approach displays competitive performance while staying entirely automatic.**

## I. INTRODUCTION

As a very natural mean of performing automatic emotion recognition, Facial Expression Recognition (FER) has grown a lot of interest in the past decades. It can be directly applied to the field of Human Computer Interfaces, including affective computing or the analysis of conversation structure, as well as in biometric systems, to enhance the performance of Face Recognition algorithms. Although research efforts have been mostly focused on the 2D image domain, 3D has emerged as a counterpart, allowing easier management of the recurrent pose variation and lighting condition issues. Hence the avaibility of 3D FER-dedicated public databases such as Bosphorus [2], BU-3DFE [3] or BU-4DFE [4], the latter being dedicated to dynamic 3D FER. In this paper, we will focus our efforts on static 3D FER.

To date, most of the works on FER have been influenced by the works of Ekman et al., who have stated that 6 facial expressions, namely Anger, Disgust, Fear, Happiness, Sadness and Surprise, regarded as prototypical, are universal among ethnicity [5]. Although those results have been recently partially discussed [6], a huge majority of the current research on this topic consists in identifying one of those 6 prototypical expressions on a probe face, usually without prior knowledge of the identity of the scanned person.

The current research on 3D FER has featured 2 main streams: feature-based approaches or model-based approaches [7]. Feature-based approaches are [8], [9], [10], [11], [12], [13] and [14]. They mostly require accurate localization of fiducial points, and analyze their configuration as well as the topology of their neighborhood for 3D FER. Such feature points are also called landmarks, and their definition is usually anatomical when used within the problem of 3D FER. [8] used 6 characteristic distances from 11 landmarks as inputs into a neural network used for classifying 7 expressions (the six prototypical expressions along with the neutral one). [9] suggested to use an automatic feature selection computed from the normalized Euclidian distance between landmarks, by using a multiclass-Adaboost classification algorithm. [10] proposed a surface descriptor derived from an estimation of primitive surface distribution, used on several regions determined with the help of several manually located fiducial points. [11] made use of SIFT descriptors in the neighborhood of landmark points. [12] used a framework able to compute the geodesic path between 3D surface paches around a set of selected landmark points. Those distances were then fed into a standard classifier like multiboosting or Support Vector Machines (SVM) to handle the classification. [13] proposed to use the *SFAM*, a statistical learning system for automatically determining the position of feature points. They later classified the expression using the position of those feature points through a Bayesian Belief Network. This automatic landmark localization method was later employed in [14] which claims that a rigid registration algorithm handles the imprecision of automatically localized landmarks.

Model-based approaches are [1], [15] and [7]. They generally fit a template 3D face model to analyze the deformations between a neutral state to an expressionnal state. [1] worked directly on the depth images, and encoded it as a sum of *BFSCs*, representing the main facial features, and *Expressional Shape Components (ESCs)* with the help of eigenvectors and the use of a few regional features. [15] proposed a joint approach between FER and Face Recognition using a bilinear model to represent the whole face. Using both asymmetric and symmetric formulations, they encoded identity and expression at the same time. [7] used the *AFM*, an annotated deformable Face Model. After a registration to the studied model, the face is represented into the conformal space, and a set of Point Distribution Models is established, using various surface descriptors (position, normals, curvature, wavelets...).

We can see several possible shortcomings for each of those streams. On the one hand, feature-based approaches tend to represent the facial components in a sparse manner, possibly resulting in an incomplete and imprecise description of facial surface deformations during an expression. Furthermore, the precise localization of a great number of landmarks on a 3D face remains an open problem. Despite the existence of several methods for automatic landmark localization, only few of the feature-based approaches have used them. On the other hand, model-based approaches often make use of rather complex fitting methods which can on some occasions severely suffer from topology changes (opening of the mouth for instance).

In this paper, we propose a holistic and non model-based approach. The key idea is to represent 3D face models, *i.e.* facial surfaces, through a set of 2D maps, thus making it possible to take advantage of the wealth of 2D-based image processing tools. For this work, we used the Histograms of Oriented Gradients (HOG) algorithm, which is already used in holistic approaches within the paradigm of the 2D FER domain [16]. This set of 2D facial maps must capture as accurately as possible facial surface deformations, whether ample or subtle, to enable facial expression analysis. For this purpose, we propose a novel facial representation, namely Differential Mean Curvature Maps (DMCMs), based on mean curvatures quantified at several scales with the help of an integral computation. To account for face morphology variations and ensure coherent facial features across people, we further normalize these facial representations using an aspect-ratio deformation. Facial features are then extracted from these normalized DMCMs using Histograms of Gradients (HOG) and fed to a multi-class SVM for classification. Using the experimental protocol as suggested in [1], the proposed approach displays competitive results as compared to the state of the art while remaining entirely automatic.

This paper is organized as follows. Section 2 overviews the proposed approach. Section 3 presents the mean curvature-based facial representations, namely Differential Mean Curvature Maps (DMCMs), through an integral computation. Then, we describe how such maps are integrated into a 3D FER scenario. Section 4 discusses the experimental results. Section 5 then concludes the paper.

## II. OVERVIEW OF THE PROPOSED APPROACH

The proposed approach can be decomposed in five main steps, which can be visualized on Figure 1.

The first step consists in aligning the 3D face models into a frontal pose, in order to generate depth images. Cropping the faces is also necessary. Incidentally, models provided in the testing database, BU-3DFE, are already rather well aligned, so we did not find appropriate to describe such a method in this paper. Numerous approaches have been proposed for this problem in 3D FER, as in [7] or [13].

The second step consists in extracting Representation Maps, called Differential Mean Curvature Maps (DMCMs), directly from the depth images. Those maps represent curvature-like data, and highlight the 3D surface topology
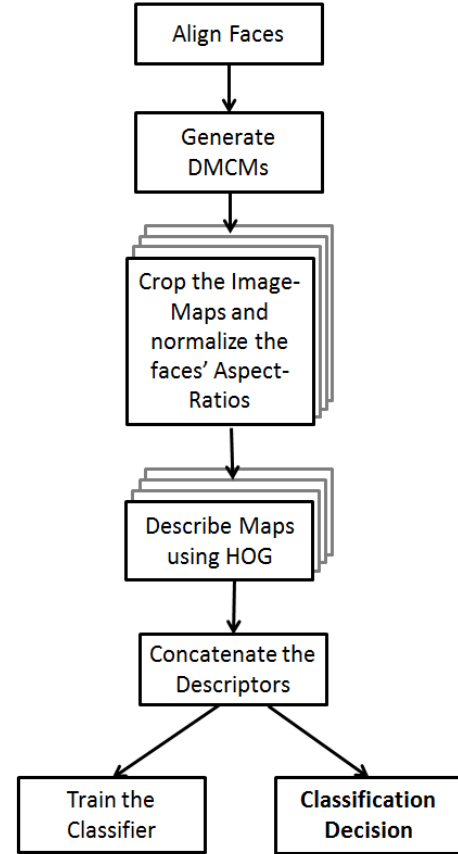


Fig. 1. Flowchart of the proposed method.

at various scales. Thus, several 2D representation images are generated for a single model, each corresponding to a different scale.

We further normalize each DMCM to only retain the informative part of the face, and discard as far as possible model boundaries. During this step, we distort the aspect ratio of face maps as well, since we found out that such a deformation helps the algorithm to better adapt to the various morphologies encountered in the dataset.

The fourth step consists in describing DMCMs thanks to the Histogram of Oriented Gradients (HOG) algorithm, as in [16]. We first decompose a normalized DMCM into several subdivisions according to a regular grid. Then, each subdivision is described using HOG. Finally, the descriptors of each subdivision are concatenated to form a global descriptor for the DMCM.

Finally, we proceed to the fusion and classification step. In our work, we chose a straightforward approach, and first performed an early fusion, consisting in the direct concatenation of the previously obtained descriptors of each different DMCM. Then, we used the classical multi-class SVM algorithm [17] for both learning and testing in our classification scheme.

## III. Representation of 3D Face model and Expression classification

[18] shows that popular holistic approaches can be very effective in 2D FER once 2D face images aligned. However, to our knowledge, similar approaches have not been used extensively in 3D FER, although 3D face scans allow for a full capture of facial surface deformations. Huang et al. [19] showed, in a context of 3D face recognition, that the direct application of usual 2D feature extraction methods like SIFT [20] on a facial depth image yields little to no relevant results. To overcome this issue, they generated several representation maps based on Local Binary Pattern (LBP) variants called MS-eLBP. Those maps are later fed into a SIFT feature point extraction step. Individual map similarity scores are computed mainly on the base of the number of matched feature points, and are finally fused to obtain a final matching score from one face to another.

To our understanding, compared to the original depth image, such maps retain only details of the 3D surface topology at various levels and scales. They discard the very global topology of the face, which is not relevant for the problem addressed by this method. LBP also stresses punctual specificities of the studied surface. Under certain circumstances, a simple bit difference yields a dramatic change in the value assigned to a pixel. Indeed, the aim of the method proposed in [19] is to enhance specificities on the 3D face, on a person-specific scale, in the paradigm of 3D facial recognition.

However, in the 3D FER domain, the purpose of the algorithms is to capture the deformations caused by expressions, *i.e.* the effects of muscles or groups of muscles on the facial shape. Which is, by essence, the principle of model-based approaches. Rather than enhancing local specificities on the surface, we made the asumption that in the paradigm of 3D FER, a representation based on the topology of the 3D model surface on a larger scale is more appropriate.

In this paper, we propose a map-extraction method that is inspired by curvature measures, computationnally efficient and which sensitivity to pose variations is only related to the projection step, while being computed directly from depth images. Such maps will be later exploited as described in section III-B in an holistic, fully automatic FER scenario.

In this section, we first introduce the integral computation for the estimation of mean curvatures. We then describe the mean curvature-based 3D facial representations, namely DM-CMs. Finally, we describe the HOG-based features extracted from these representations for the purpose of 3D FER.

### A. Estimation of Differential Mean Curvatures through Integral Computation

We made use of the integral computation method as proposed in [21] to approximate the computation of several curvature values on a 3D model. The computation of the mean curvature is particularly efficient. Given a point $p$ on the surface of a volume $V$, the intersection $V_b(r, p)$ of a sphere $b$ of radius $r$ centered in $p$ with $V$ gives :

$$V_b(r, p) = \frac{2\pi}{3}r^3 - \frac{\pi}{4}H(p)r^4 + O(r^5) \qquad (1),$$

where $H(p)$ corresponds to the mean curvature at the point $p$.

From equation (1), we can see that there is a direct correlation between $V_b(r, p)$ and $H(p)$. However, $H$ is dimensionally homogeneous to $r^{-1}$. Since we expect indices to produce an image map, we build $h(r, p)$ such as

$$h(r, p) = \frac{3}{4\pi r^3}V_b(r, p)$$

$h(r, p)$ is an index, which value is comprised between 0 and 1.

The smaller $r$ goes, the more accurate approximation of $H(p)$ we can get from $h(r, p)/r$ . However, as stated in [21], a smaller radius yields noisier results. We can actually take advantage of larger radiuses, for they emphasize on different scales and levels on the 3D object.

We further extended this approach by using an outer radius $r_o$ and an inner radius $r_i$, and define $h(r_i, r_o, p)$ as

$$h(r_i, r_o, p) = \frac{3}{4\pi(r_o^3 - r_i^3)}(V_b(r_o, p) - V_b(r_i, p))$$

with $0 \leq r_i < r_o$.

The idea is to be able to discard smaller radiuses, so that the behaviour of our descriptor is slightly more similar to that of a bandpass filter, regarding scales on the 3D face, than the original descriptor.

Computationnally, the volume $V_b(r, p)$ can be approximated by an set of unity cubic patches. When considering the depth image of a face, we regard any voxel behind the face as part of the volume of the face. Then, an approximation of $h(r_i, r_o, p)$ can be computed very efficiently directly from the depth image by matching the image resolution to the unity cubic patches. By construction, such a descriptor is invariant to pose, as long as we are not in the presence of self-occlusion phenomenons.

### B. Mean Curvature-based Facial Representations: DMCMs

Maps obtained by computing $h(r_i, r_o, p)$ at every point of a depth image are referred to as *Differential Mean Curvature Maps* (DMCMs) in the subsequent. Like in [19], our idea is to generate several DMCMs $M_I(r_i, r_o)$ using various values of $r_i$ and $r_o$ from a candidate depth image. Then, we generate a feature vector for each map individually.

First, we need to preprocess 3D face images before we can generate the various maps. Faces need to be aligned in order to generate frontal views. Then, we resample the 3D model to generate the corresponding depth image. The size of the grid applied to the X and Y coordinates is pretty important, being a tradeoff between the accuracy of the computation of DMCMs and their computational cost. In this work, we set the grid to 0.8 mm. Depth images were generated through a bilinear interpolation, which smoothes the surface and avoids holes. Before cropping the faces, we first generate DMCMs using various radiuses, in order to
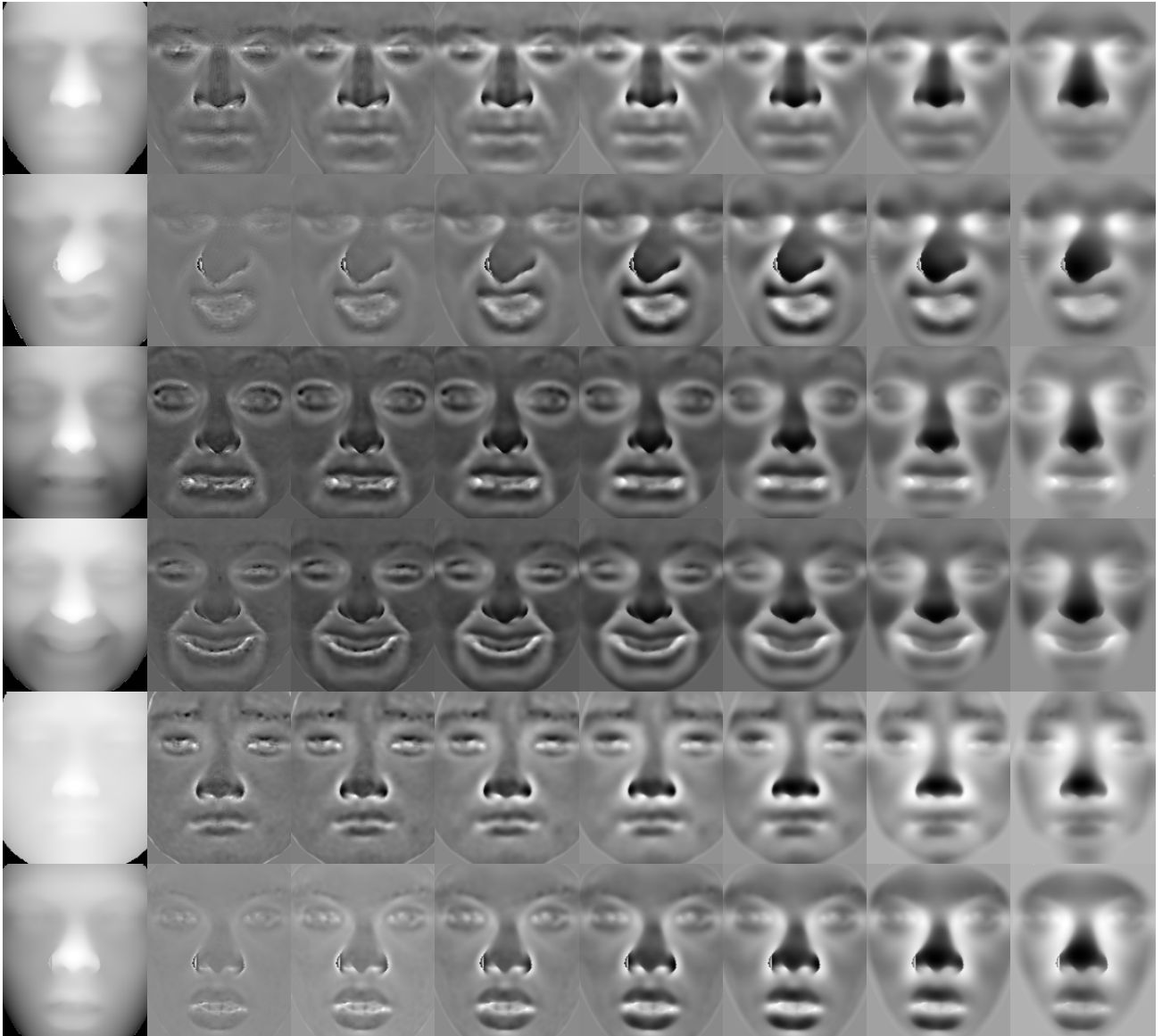
Fig. 2. Various examples of DMCMs after normalization, applied to 3D expressionnal faces. From top to bottom, expressions are Anger (AN), Disgust (DI), Fear (FE), Happinness (HA), Sadness (SA) and Surprise (SU). From left to the right, images are the original range image, then the DMCMs following sets of radiuses $S_1$, $S_2$, $S_3$, $S_4$, $S_5$, $S_6$ and $S_7$ according to section III-B.

avoid boundary artifacts. Sets of radiuses $S_k = (r_i^k, r_o^k)$ that we used are the following (in millimeters) : $S_1 = (0, 3\sqrt{2}); S_2 = (3, 6); S_3 = (3\sqrt{2}, 6\sqrt{2}); S_4 = (6, 12); S_5 = (6\sqrt{2}, 12\sqrt{2}); S_6 = (12, 24); S_7 = (12\sqrt{2}, 24\sqrt{2})$. Those were chosen as octaves, just like the radiuses picked by the SIFT algorithm when applying various gaussian filters. The chosen radiuses seem to highlight various features on the human face.

The following step consists in normalizing the DMCMs. Experimentally, we found out that distorting the aspect ratio of the projected image, so that facial features are roughly located in the same regions within the normalized maps, actually enhances consistently the performances of the algorithm. To our understanding, this is due to morphologies between individuals varying quite a lot in terms of both

head size and proportions. First, we crop the faces with a sphere of radius experimentally set to 80mm, centered at the nose-tip, which is the point with the highest Z value close to the center of the facial scan. At this stage, we set the boundaries of the image map so that they match the cropped face boundaries. As a result, and due to morphology varying from one person to another, different maps from different individuals are likely to display a different image size. Then, we resize cropped face views to a 240x200 image size, which modifies the aspect ratio of facial images. Finally, we crop the DMCMs to a 180x150 image size, centered at the nose-tip position plus an offset of 8 pixels over the X coordinates. This cropping allows us to retain most informative parts of the face while discarding its boundaries as much as possible. We set empirically the previously mentioned offset so to

focus better the information on the eyebrows region. In Figure 2, we present a set of normalized DMCMs, extracted from range image with the algorithm described above.

## C. HOG-based Facial features

After the normalization step, we need to generate a feature vector for each of these normalized DMCMs. For this purpose, each normalized DMCM is divided into several regions using a regular grid as in [22]. Similar to [16], Histograms of Oriented Gradients (HOG) are then extracted from each subdivision of the face. They are further concatenated to generate a facial feature vector, and then fed into a standard machine learning scheme. In our method, we directly concatenated the feature vectors extracted from several maps (corresponding to different values of $(r_i, r_o)$) before we fed them to our classifier.

## IV. EXPERIMENTS

In this section, we first describe the experimental protocol and setup and then discuss the results.

## A. Experimental protocol and setup

Most recent 3D FER works were experimented on the BU-3DFE database [3], the latter being used for both learning the parameters of the algorithm, learning the classifier and testing it. The BU-3DFE database consists in 2500 3D facial scans of 100 persons (56 female and 44 male subjects), acquired in constraint conditions (no visible clothes nor glasses nor hair, frontal pose). For each person, one neutral scan and 4 different scans of each of the 6 prototypical expressions, namely Anger (AN), Disgust (DI), Fear (FE), Happiness (HA), Sadness (SA) and Surprise (SU) as defined by Ekman [5]. The 4 different scans for each expression correspond to 4 increasing degrees of intensity in the expression performance. To date, most research works have only considered the last 2 levels of expressions intensity, and thus consider 12 expressionnal faces for each person. In this paper, we also comply to this setup.

Usually, 40 out of the 100 people of the database are used for defining the parameters of the algorithm if needed, while the 60 remaining persons are used in a cross-validation scheme for the classifier. In [1], the authors stated that peforming 10 times independently a 10-fold cross-validation over 60 different persons doesn't provide enough room for a stable mean recognition score. Thus, in our experimentation, we followed their recommendation and performed the 10-fold cross-validation scheme 1000 times independently. Results are presented in section IV-B. Since our method does not require a parameter learning step prior to the classification, the 60 individuals used in the cross-validation scheme were picked randomly. Our method also allowed us to perform the cross-validation experiment directly over the 100 people of the database, which results are exposed in section IV-C.

We chose the widely used multi-class SVM algorithm [17] as our classifier. In our experiments, we used 5x5 and 6x6 subdivisions of the face maps, and the HOG algorithm

| | Rate (%) | |
|---|---|---|
| DMCM set | 5x5 grid | 6x6 grid |
| range | 62.17 | 63.75 |
| $S_1$ | 68.63 | 68.68 |
| $S_2$ | 71.99 | 70.38 |
| $S_3$ | 72.94 | 71.06 |
| $S_4$ | 74.72 | 72.21 |
| $S_5$ | 73.74 | 71.38 |
| $S_6$ | 73.53 | 72.72 |
| $S_7$ | 73.28 | 72.53 |
| $S_{all}$ | 75.78 | 75.99 |
| $S_{all}$ | 76.61 | |

was set to 9 directions. In total, each of the 7 different face maps (which parameters have been exposed previously) was represented by a feature fector of size 225 (5x5 subdivision scenario) or 324 (6x6 subdivision scenario). The concatenation of all those different feature vectors generates a 3843-sized feature vector, which can still be handled by the publicly available multi-class SVM implementation without requiring a dimensionnality reduction step.

## B. Results using the usual experimental setup

We first present the average recognition rate for each invidual set of maps in table I, given the HOG subdivision parameter and the radiuses of mean maps. In this table $S_{all}$ corresponds to the concatenation of all $S_n$ maps (associated with varius radiuses, as in section III-B), while the last line gives the result for the concatenation of all combinations of radiuses and HOG grid subdivisions. As a comparison, *range* stands for results obtained with HOGs directly applied to the depth images.

Table II presents the average confusion matrix obtained by the concatenation of both 5x5 and 6x6 subdivisions and all radiuses (corresponding to the $S_{all}$ lines in Table I).

This experiment shows that all radiuses provide acceptable results, while remaining complementary towards a fusion scheme. They also prove to be more informative than the original depth image. The presence of a consistent maximum for both subdivision schemes at $S_4$ also leads us to believe that expressions affects the shape of the 3D face at a rather typical scale.

In Table III, we provide a comparison with other state of the art algorithms. We show that our method achieves comparable results, while being fully automatic, and providing a framework for potentially more ad-hoc methods within the 2D FER paradigm. Our method also has the advantage of not requiring any prior parameter learning step, which allowed us to conduct a similar experiment over the 100 persons included in the database, which is described in section IV-C.

## C. Extended experiments with the whole database

In this section, we present an experiment including all 100 persons in the cross-validation scheme. Again, we repeat a 10-fold cross-validation scheme and classify the level 3 and 4 of the 6 prototypical expressions in the BU-3DFE

TABLE II

AVERAGE CONFUSION MATRIX OBTAINED WITH $S_{all}$ AND BOTH 5X5
AND 6X6 GRIDS

| %  | AN    | DI    | FE    | HA    | SA  | SU    |
|----|-------|-------|-------|-------|-----|-------|
| AN | **72**    | 7.3   | 4.5   | 0     | 17.9 | 0.25  |
| DI | 8.3   | **74.9**  | 9.8   | 1.08  | 2.8 | 1.9   |
| FE | 4.17  | 10.5  | **62.25** | 11.17 | 5.5 | 3.67  |
| HA | 0.25  | 3     | 13.3  | **86.42** | 0.58 | 0.75  |
| SA | 14.33 | 2.5   | 5.17  | 0     | **72**  | 1.33  |
| SU | 0.92  | 1.75  | 4.92  | 1.3   | 1.17 | **92.08** |
| Average | | | **76.6** | | | |

TABLE III

AVERAGE RECOGNITION RATES FOR THE WORKS IN BERRETTI [11],
GONG [1], WANG [10], SOYEL [8], TANG [9], ONE OF OUR PREVIOUS
METHODS [14] AND THE METHOD PROPOSED IN THIS PAPER.

| %   | Berretti | Gong  | Wang  | Soyel | Tang  | [14]  | Ours  |
|-----|----------|-------|-------|-------|-------|-------|-------|
| Avg | 77.54    | 76.22 | 61.79 | 67.52 | 74.51 | 75.76 | 76.61 |

database. This time though, we repeated only 200 times the cross-validation scheme, since we noticed that the standard deviation was half the value observed on the first experiment. The content and organization of our corresponding Tables IV and V is similar to that of the previous section.

Interestingly, recognition rates are consistently higher at all radiuses within this extended experiment, while the machine learning scheme was using the same learning-testing ratio than in the previous experiment. Our interpretation is that this experiment highlights the issues pointed by [1], about BU-3DFE being too small a dataset for the standard 10-fold cross-validation performed over 60 different persons.

## V. CONCLUSION AND FUTURE WORKS

In this paper, we proposed a novel approach for representing 3D faces, which allowed us to apply a standard algorithm of the 2D FER domain to the problem of 3D FER. The generation of Differential Mean Curvature Maps (DMCMs), based on an integral, curvature-like and computationnally efficient calculation, enhances the distinctiveness of the facial surface topology at various scales. This method allows us to perform comparable performances to existing state of the art methods, while being fully automatic and holistic, and does not require the use of a single landmark point but the nose-

TABLE IV

AVERAGE RECOGNITION RATES WITH OUR EXTENDED EXPERIMENT

|           | Rate (%) | |
|-----------|----------|----------|
| DMCM set  | 5x5 grid | 6x6 grid |
| $S_1$     | 71.27    | 71.13    |
| $S_2$     | 72.77    | 72.4     |
| $S_3$     | 73.82    | 77.12    |
| $S_4$     | 73.58    | 75.43    |
| $S_5$     | 76.65    | 75.9     |
| $S_6$     | 75.42    | 76.57    |
| $S_7$     | 76.18    | 76.07    |
| $S_{all}$ | 76.68    | 78.1     |
| $S_{all}$ | 78.13    |          |

TABLE V

AVERAGE CONFUSION MATRIX OBTAINED WITH $S_{all}$ AND BOTH 5X5
AND 6X6 GRIDS IN OUR EXTENDED EXPERIMENT

| %  | AN   | DI    | FE   | HA    | SA   | SU   |
|----|------|-------|------|-------|------|------|
| AN | **74.1** | 7.7   | 3.6  | 0     | 15.7 | 0    |
| DI | 8    | **74.9**  | 12.3 | 1.7   | 3.6  | 1.3  |
| FE | 5.1  | 10.8  | **64.6** | 8.1   | 4.6  | 5.6  |
| HA | 0    | 3.1   | 10.7 | **89.8**  | 0.6  | 1    |
| SA | 12.4 | 2.3   | 5.1  | 0     | **74.5** | 1.2  |
| SU | 0.4  | 1.2   | 3.7  | 0.4   | 1    | **90.9** |
| Average | | | **78.13** | | | |

tip. Thus, it allowed us to perform an extended experiment based on the whole BU-3DFE database, that surprisingly displayed better performances.

Our experiments also showed that the face normalization step has a sensitive impact over the global performances of the 3D FER algorithm. Interestingly, deforming facial images by modifying their aspect ratio yields more accurate performances. We will investigate it further, as well as other 2D FER algorithms applied to DMCMs. We also want to study the performances involving other curvatures computed as integrals, as exposed in [21].

## REFERENCES

[1] B. Gong, Y. Wang, J. Liu, and X. Tang. Automatic facial expression recognition on a single 3d face by exploring shape deformation. In Proceedings of the 17th ACM international conference on Multimedia, MM 09, pages 569572, New York, NY, USA, 2009. ACM.

[2] A. Savran, N. Alyuz, H. Dibeklioglu, O. Celiktutan, B. Gokberk, B. Sankur, and L. Akarun. Biometrics and identity management. In B. Schouten, N. C. Juul, A. Drygajlo, and M. Tistarelli, editors, Biometrics and Identity Management, chapter Bosphorus Database for 3D Face Analysis, pages 4756. Springer-Verlag, Berlin, Heidelberg, 2008.

[3] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato. A 3d facial expression database for facial behavior research. In Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition, FGR 06, pages 211216, Washington, DC, USA, 2006.

[4] L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale. A high-resolution 3d dynamic facial expression database. In FG08, pages 16, 2008.

[5] P. Ekman and W. V. Friesen. Constants across cultures in the face and emotion. Journal of Personality and Social Psychology, 17(2):124129, 1971.

[6] R. Jack, O. Garrod, H. Yu, R. Caldara, P. Schyns. Facial expressions of emotion are not culturally universal, Proceedings. the National Academy of Sciences DOI: 10.1073/pnas.1200155109, 2012

[7] T. Fang, X. Zhao, O. Ocegueda, S.K. Shah, and I.A. Kakadiaris. 3D Facial Expression Recognition: A perspective on Promises and Challenges. In proceedings of IEEE International Conference on Automatic Face and Gesture Recognition and Workshops (FG 2011), 2011.

[8] H. Soyel and H. Demirel. Facial expression recognition using 3d facial feature distances. In ICIAR07, pages 831838, 2007.

[9] H. Tang and T. Huang. 3d facial expression recognition based on automatically selected features. In Computer Vision and Pattern Recognition Workshops, 2008. CVPRW 08. IEEE Computer Society Conference on, pages 1 8, june 2008.

[10] J. Wang, L. Yin, X. Wei, and Y. Sun. 3d facial expression recognition based on primitive surface feature distribution. In in Proc. Conf. Computer Vision and Pattern Recognition, pages 13991406, 2006.

[11] S. Berretti, A. Bimbo, P. Pala, B. Amor, and M. Daoudi. A set of selected sift features for 3d facial expression recognition. In Pattern Recognition (ICPR), 2010 20th International Conference on, pages 4125 4128, aug. 2010.

[12] A. Maalej, B. B. Amor, M. Daoudi, A. Srivastava, and S. Berretti. Shape analysis of local facial patches for 3d facial expression recognition. Pattern Recognition, 44(8):15811589, 2011.

[13] X. Zhao, E. Dellandréa, and L. Chen. A 3d statistical facial feature model and its application on locating facial landmarks. In J. Blanc-Talon, W. Philips, D. Popescu, and P. Scheunders, editors, Advanced Concepts for Intelligent Vision Systems, volume 5807 of Lecture Notes in Computer Science, pages 686697. Springer Berlin / Heidelberg, 2009.

[14] P. Lemaire, B. Ben Amor, M. Ardabilian, L. Chen, M. Daoudi, Fully automatic 3D facial expression recognition using a region-based approach. Proceedings of the 2011 joint ACM workshop on Human gesture and behavior understanding (J-HGBU '11), pages 53-58, 2011.

[15] I. Mpiperis, S. Malassiotis, and M. Strintzis. Bilinear models for 3-d face and facial expression recognition. Information Forensics and Security, IEEE Transactions on, 3(3):498 511, sept. 2008.

[16] M. Dahmane and J. Meunier, Emotion recognition using dynamic grid-based HoG features. In proceedings of IEEE International Conference on Automatic Face and Gesture Recognition and Workshops (FG 2011), 2011.

[17] V. Franc and V. Hlavac. Multi-class support vector machine. In Pattern Recognition, 2002. Proceedings. 16th International Conference on, volume 2, pages 236 239 vol.2, 2002.

[18] M. Pantic, S. Member, and L. J. M. Rothkrantz. Automatic analysis of facial expressions: The state of the art. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22:14241445, 2000.

[19] D. Huang; M. Ardabilian, W. Yunhong, and L. Chen, A novel geometric facial representation based on multi-scale extended local binary patterns. In proceedings of IEEE International Conference on Automatic Face and Gesture Recognition and Workshops (FG 2011), 2011.

[20] D.G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, Volume 60 Issue 2 pages 91-110, 2004.

[21] H. Pottmann, J. Wallner, Y.-L. Yang, Y.-K. Lai and S.-M. Hu, Principal curvatures from the integral invariant viewpoint. Computer Aided Geometric Design, volume 24 Issue 8-9, pages 428-442, 2007.

[22] T. Ahonen, A. Hadid, and M. Pietikäinen, Face Description with Local Binary Patterns: Application to Face Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28:2037-2041, 2006.