Triaxial Slicing for 3-D Face Recognition From Adapted Rotational Invariants Spatial Moments and Minimal Keypoints Dependence

Robson S. Siqueira, Gilderlane R. Alexandre, José M. Soares, and George A. P. Thé

***Abstract*—This letter presents a multiple slicing model for three-dimensional (3-D) images of human face, using the frontal, sagittal, and transverse orthogonal planes. The definition of the segments depends on just one key point, the nose tip, which makes it simple and independent of the detection of several key points. For facial recognition, attributes based on adapted 2-D spatial moments of Hu and 3-D spatial invariant rotation moments are extracted from each segment. Tests with the proposed model using the Bosphorus Database for neutral vs nonneutral ROC I experiment, applying linear discriminant analysis as classifier and more than one sample for training, achieved 98.7% of verification rate at 0.1% of false acceptance rate. By using the support vector machine as classifier the rank1 experiment recognition rates of 99% and 95.4% have been achieved for a neutral vs neutral and for a neutral vs non-neutral, respectively. These results approach the state-of-the-art using Bosphorus Database and even surpasses it when anger and disgust expressions are evaluated. In addition, we also evaluate the generalization of our method using the FRGC v2.0 database and achieve competitive results, making the technique promising, especially for its simplicity.**

***Index Terms*—Computer vision for automation, recognition, surveillance systems.**

I. Introduction

THE traditional 2D images have been widely used as data source for Face Recognition (FR) during the past decades and, inspite of that, 2D FR remains with important constraints. The major challenges associated to this task include strong inter-subject facial similarities and intra-subject variations [1]. The latter is greatly affected by illumination changes, variations in pose and expressions, variability of the background and occlusions [2], [3].

3D FR approaches have recently gained popularity as increasingly better 3D sensors have become available. 3D data provide more reliable geometric information and allow the extraction of certain features, which are resilient to changes in scale, rotation, and illumination. That capability of preserving geometric information in addition to powerful feature extraction methods allow overcoming degradation conditions, representing a clear advantage over 2D techniques [4].

On the other hand, 3D techniques are not as mature and well-established as 2D approaches are. Furthermore, some information may be more trivial to be extracted from 2D data than from 3D data (e.g., automatic landmark detection). For that reason, hybrid approaches that combine two types of data sources, 2D and 3D, are also employed.

Regarding the feature extraction methods, three main categories are addressed in the literature: holistic, feature-based and hybrid matching methods. Holistic methods find a set of global features from the entire 3D model while feature-based methods focus on the extraction of local features from the face or from regions of the face. Liu *et al.* [5] employ a holistic method by using spherical harmonic features. Principal Component Analysis (PCA) is also largely employed in this category. Hybrid matching methods combine holistic and feature-based approaches [4].

While holistic methods focus on finding overall similarities of faces, local-feature methods tend to be robust to partial models and occlusions, and are more suitable for matching, identification and verification purposes [4]. 3D local-feature based methods are divided into three groups, according to the descriptors: keypoint-based, curve-based and local surface-based methods.

3D keypoint-based methods rely on the detection of salient points of face and the description of a feature vector. According to [4], although these methods can deal with partial models and occlusions, since they use a large number of keypoints and consequently high dimensional feature vectors, they are computationally demanding. Indeed, recent works using keypoint-based methods compute a large number of keypoints in order to obtain satisfactory recognition rates, as reported in the literature [1], [6], [7], [8].

Curve-based methods, as the one employed in [9], use curves traced on faces according to certain criteria (e.g., contour, profile) and extract geometrical information from different areas of the face. That makes them more robust against facial expressions than keypoint-based methods [4].

Local surface-based methods, in turn, extract geometric information from several regions of the surface face that are invariant to facial expressions and robust to that condition [4]. Examples

Manuscript received February 24, 2018; accepted June 19, 2018. Date of publication July 9, 2018; date of current version August 2, 2018. This paper was recommended for publication by Associate Editor B. Morris and Editor W. K. Chung upon evaluation of the reviewers' comments. This work was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. *(Corresponding author: Robson S. Siqueira.)*

R. S. Siqueira is with the Faculty of Electrical Engineering, Instituto Federal de Educação, Ciência e Tecnologia do Ceará, Fortaleza, CE 60410-42, Brazil (e-mail: siqueira@ifce.edu.br).

G. R. Alexandre, J. M. Soares, and G. A. P. Thé are with the Department of Teleinformatics Engineering, Universidade Federal do Ceará, Fortaleza, CE 60020-181, Brazil (e-mail: gilderlane.ribeiro@gmail.com; marques@ufc.br; george.the@ufc.br).

Digital Object Identifier 10.1109/LRA.2018.2854295

include Ocegueda *et al.* in [10], which addresses the 3D FR problem by analyzing probabilistically the relevance of a certain region of the face for the current expression and considering the most discriminative regions for feature extraction.

In [11], Emambakhsh and Evans investigate the effect of the nasal region for expression robust 3D FR. Their method relies on the nose landmarking and performs the feature extraction based on surface normals over patches and curves on the nasal region to finally find the most stable patches under variations of expressions.

In [12], multiple keypoints are employed to derive 3D surface patches. In that paper, Al-Osaimi presents rotation invariant descriptor extraction techniques applicable for FR on 3D static images or 3D videos and does not address the partial occlusions of the local surface.

The approach proposed in the present work offers a novel strategy for the processing of 3D images based on tri-axial slicing with a minimal dependence on keypoints. In a nutshell, this technique includes a unique keypoint detection and the definition of surface regions between three sets of parallel planes, so called Frontal, Sagittal and Transverse. Planes are positioned in space based on the keypoint and attributes are then extracted from each surface region. The main contributions of this work are (1) the proposal of adapted Hu moments, originally applied in the 2D domain, (2) their extraction from surface regions along the three orthogonal axes and combinations (the literature usually considers single axis) and (3) minimal keypoint dependence. As a consequence, we achieved classification performance comparable to [1], in shorter times than those reported in that paper and in many of those referenced in the review [4], as well.

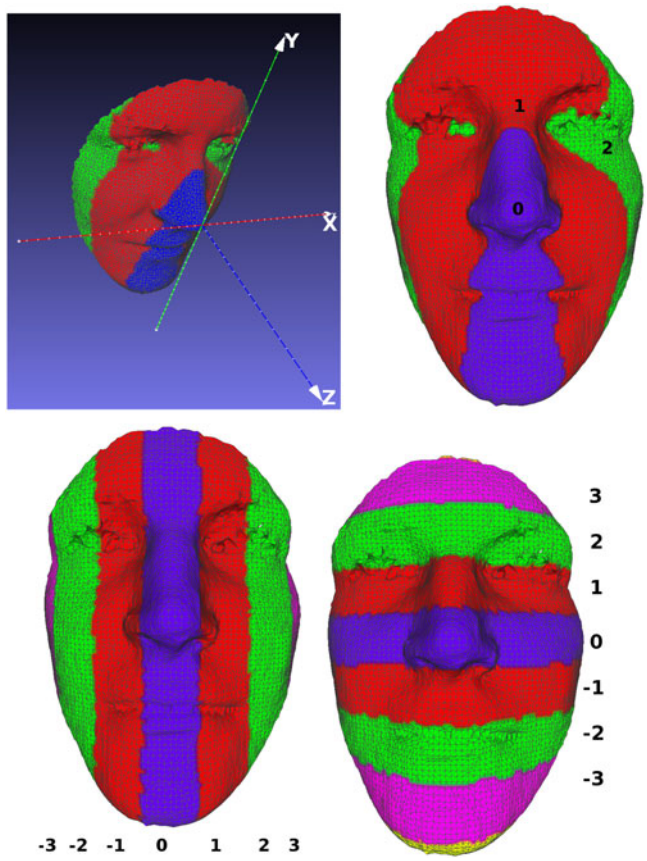

Fig. 1. Nose tip with respect to the XYZ axes and the slices in the three orthogonal planes, Frontal, Sagittal and Transverse. This example corresponds to a Neutral sample considering $\mathbf{c} = (c_f, c_s, c_t) = (20, 20, 20)$ in millimeters and a 80-mm wide cropping centered at the nose tip. Here the blue color represents the segment indexed as zero. Note: the visualization above as a surface is only for illustration.

II. Method

A. Tri-axial Face Slicing

Our method defines well-marked and reproducible regions, based on a unique and easily identifiable keypoint and extracts robust local descriptors to identify individuals using only 3D point clouds. We define regions of the face and descriptors from those regions that are able to identify the individuals even under large variation of facial expressions. We propose the slicing of the face by using planes orthogonal to each other, positioned based on one keypoint, in such manner that considerable surface changes caused by expressions concentrate in some of the regions but are minimal in others.

The intuition behind our method for feature extraction relates to the nature of deformations on the face surface under expression variations. Different patches suffer different deformation profiles along preferable directions while facial expression changes. We have observed that behaviour after estimating deformation as done in the field of nonrigid registration of 3D medical images [13]; from the eigenvalues of the covariance tensor of the 3D data, we proceeded to calculate anisotropy, omnidirectionality and other geometric features, and then investigated how these numbers vary among individuals and among different patches of a given individual. Result of this analysis revealed that some patches are less sensitive to deformation than others, thus confirming our expectation that patches that do not suffer great deformations along one or more of the three directions are robust enough to yet represent individuals.

The essence of the technique lies in point cloud slicing of a face image along the three spatial axes, giving rise to what is called Frontal, Sagittal and Transverse slices, as shown in Fig. 1. The method developed here requires some pre-processing steps: first, the point cloud has to undergo density normalization; this is followed by outliers removal; next by the keypoint detection, then a cropping is carried out based on the keypoint found; finally, a pose correction to align faces.

There are a number of reasons to perform density normalization as the very first pre-processing step. One is to decrease the number of points in a cloud, making the processing in the next steps faster. The other reason, and not less important, relates to the type of descriptors that are to be extracted here, since they can be sensitive to intra-subject and inter-subject density variations. The idea we follow here is to decrease the density of the point clouds until all of them have the same density of points. Ideally, a good acquisition apparatus should provide images with nearly the same point density, however this is not observed in practice. We have seen samples with denser clouds which

probably suggests the individual getting closer to the camera in some sessions. Those density variations can occur either for samples of the same subject or for distinct ones. Depending on the type of descriptors employed, the recognition process can be insensitive to that variation or be significantly affected by it. It is important to notice that the decrease of density can not be extreme, so the faces become uncharacterized, nor minimal, so the method does not benefit from it. Our technique uses the Point Cloud Library (PCL) 3D voxel grid algorithm, presented by Rusu and Cousins [14], to perform density normalization with a voxel leaf of 2 millimeters for all 3 dimensions.

Outliers removal is the most common step among methods that perform surface analysis, since outliers may cause peaks that uncharacterize facial surface. In our case, they can add noise to the face descriptors. Although outliers are prevalent in occluded or self-occluded images, in general outliers from any sources can prevent the detection of important keypoints, such as the nose tip.

Nose tip detection is performed by taking the closest point to the sensor (depending on the reference system adopted, it can be also considered the most distant point from the sensor). One reason why outliers removal is important in this case, is that outliers close to the sensor can compete with the actual nose tip.

Before pose correction, a cropping is carried out on the point cloud; starting from the nose tip, it extends up to an Euclidean distance $d = 80$ mm just enough to comprise the regions of the eyes, nose and mouth. Face cropping is performed by several authors, including [6], [7], [1] to restrict the region of the face that best characterizes the individual.

Pose correction with Iterative Closest Point (ICP) alignment is applied in order to iteratively correct the pose of the 3D faces. The probe faces are aligned with respect to the gallery faces as in Elaiwat *et al.* [6]. Elaiwat *et al.* [6] employ ICP algorithm only for the samples with great rotation and use PCA for the others, since ICP, being iterative, can make the whole procedure slower. In our work, we mitigate this drawback by reducing the cloud density and cropping the face around the nose, as well. We also limited the number of iterations to 100. In addition, we tested 3 possibilities of pose correction, two other methods for pose correction that aimed to find the face normal and align it with z-axis: RANSAC-Plane [15], finding the best plane that fits the face and using its normal for pose correction; and two-sided block-Jacobi SVD method [16] to calculate the face normal. None of them beat ICP. Two-sided block-Jacobi SVD is fast and highly parallelizable [16] while RANSAC-Plane has approximately the same processing time as ICP. The best arrangement we found was first perform a coarse alignment with two-sided block-Jacobi SVD and then apply ICP for fine correction, which minimizes the number of iterations required. The maximum number of iterations was set to 100, since superior values had no considerable effect on alignment.

Pose correction is an important step to mitigate head inclination during image acquisition sessions, thus preventing bad formation and indexation of slices. It allows the keypoint corresponding to the nose tip to be found ; in the following it is referred to as $\nu = (x_n, y_n, z_n)$. For each neutral or non-neutral sample of an individual expression, the point ν must be coincident or be at a minimal acceptable distance relative to the other samples.

The slicing is carried out in such a way that each slice contains only the points $p = (x, y, z)$ of the $\mathbf{C}$ point cloud lying between two planes of adjacent cut-sections (these are equally spaced along axes). Reference planes for the Frontal slices are those parallel to the XY plane of Fig. 1 and the slice thickness is c_f; reference planes for the Sagittal and Transverse slices are, in turn, those parallel to the YZ and XZ planes of Fig. 1, and the respective thicknesses are c_s and c_t. Thus the triple $\mathbf{c} = (c_f, c_s, c_t)$ defined in millimeters, along with the coordinates of the nose tip, x_k, y_k and z_k define the indices n of Frontal, Sagittal and Transverse slices as shown in Fig. 1.

The Frontal, Sagittal and Transverse sections are represented respectively by the sets of points F_n, S_n and T_n, $n \in Z$, as shown in Equations (1), (2) and (3) below:

$$p \in F_n \rightarrow \left\lfloor \frac{|z_k - z|}{c_f} \right\rfloor = n \tag{1}$$

$$p \in S_n \rightarrow \left\lfloor \frac{|x_k - x|}{c_s} \right\rfloor \cdot \text{sign}\,(x_k - x) = n \tag{2}$$

$$p \in T_n \rightarrow \left\lfloor \frac{|y_k - y|}{c_t} \right\rfloor \cdot \text{sign}\,(y_k - y) = n. \tag{3}$$

B. Extraction of Attributes

The combination of Hu Moments [17] and 3D Invariant Moments [18] can provide robust local descriptors to characterize face, even if there are rotations or translations between samples of a given individual acquired at different time instants. For FR, for example, the distance to the sensor may change significantly, in addition to rotations, even if minimal, around the three orthogonal axes. This requires the used descriptors to be invariant to these transformations. The spatial moments are defined in [17] and consider the xy-axis as the dependent terms. Equations (4)-(6) present variations of it for each pair of axis.

$$M_{pq}^{XY} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p\, y^q\, f(x,y)\, dx\, dy \tag{4}$$

$$M_{pq}^{XZ} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p\, z^q\, f(x,z)\, dx\, dz \tag{5}$$

$$M_{pq}^{YZ} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^p\, z^q\, f(y,z)\, dy\, dz \tag{6}$$

In the algorithm used here, the feature vector contains the Hu moments calculated from each spatial orientation. Thus, in the case of M^{XY}, the configuration that considers the z-axis as independent for the calculation of the 7 Hu moments is represented as HuXY. The discrete and centralized version of Equation (4) for the HuXY configuration duly normalized and with the independent term is shown in Equation (7), where x and y are the coordinates of a given pixel, F(x, y) is its intensity and $\overline{\text{x}}$ and $\overline{\text{y}}$ are the average values along X and Y axis, respectively. It is referred to as central moment of order i+j.

$$\mu_{ij}^{XY} = \sum_{\text{x}}\sum_{\text{y}} (\text{x} - \overline{\text{x}})^i\, (\text{y} - \overline{\text{y}})^j\, F(\text{x}, \text{y}) \tag{7}$$

The Hu moments implementations for 2D grayscale or binary images make use of the intensity of the pixels as F(x,y). Therefore, the function has values ranging between $[0, 255]$ for grayscale images and between $[0, 1]$ for binary images. When adapting this for 3D images, the above assignment loses sense, giving rise to new choices for that function.

We propose 2 possible implementations for F(x,y). In Equation (8), F(x, y) is simply the distance from the coordinate z to the average value along that coordinate, while in Equation (9), F(x,y) is the Euclidean distance between the point of coordinates (x, y, z) and the nose tip $(\mathrm{x_n}, \mathrm{y_n}, \mathrm{z_n})$.

$$\mathrm{F(x,y)} = (\mathrm{z} - \bar{\mathrm{z}}) \tag{8}$$

$$\mathrm{F(x,y)} = \sqrt[2]{(\mathrm{x}-\mathrm{x_n})^2 + (\mathrm{y}-\mathrm{y_n})^2 + (\mathrm{z}-\mathrm{z_n})^2} \tag{9}$$

From the above two definitions, we will consider HuXY and HuXYd as the Hu moments computed in XY plane that use the functions defined in Equations (8) and (9), (9), respectively. Similarly, we have for planes XZ e YZ, the following acronyms: HuXZ, HuXZd, HuYZ and HuYZd.

Hu moments undergo a normalization to be endowed with scale invariance, which is a very important property for 2D methods. In [17], the scale invariant moments of order i+j are presented as ψ_{ij} and calculated as Equation (10).

$$\psi_{ij} = \frac{\mu_{ij}}{\mu_{00}^{(1+\frac{i+j}{2})}} = \frac{\mu_{ij}}{\sqrt[2]{\mu_{00}^{(2+i+j)}}}, (i+j) \geq 2 \tag{10}$$

Since we are dealing with 3D point clouds of faces, scale transformations are not feasible. Instead, we deal with another case: density variation. For that reason, the 7 Hu moments proposed in [17] and computed based on ψ_{ij} from Equation (10) are here adapted and computed based on ψ_{ij} from Equation (11).

$$\psi_{ij} = \mu_{ij}^{XY} \tag{11}$$

An example for the first Hu moment is expressed in Equation (12). All the others follow the same idea. The complete list of moments can be reviewed in [17].

$$\Psi_1 = \psi_{20} + \psi_{02} \tag{12}$$

The 3D Invariant moments [18], which were also employed as descriptors in our approach, are based on an expanded version of Equation (4) to three dimensions. Thus, they are invariant to translation and rotation in 3-D. Its continuous version can be reviewed in Equation (13).

$$M_{pqr} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p\, y^q\, z^r\, f(x,y,z)\, dx\, dy\, dz \tag{13}$$

Similarly to Equations (8 and (9), we present two options for $f(x,y,z)$ in Equations (14) and (15) , where (x_n, y_n, z_n) are the nose tip coordinates.

$$f(x,y,z) = 1 \tag{14}$$

$$f(x,y,z) = \sqrt[2]{(x-x_n)^2 + (y-y_n)^2 + (z-z_n)^2} \tag{15}$$

The seven Hu moments and the three 3DIT moments are extracted from each of the slices. The total number of descriptors depends on the value set in c for the intervals that cut the three orthogonal axes.

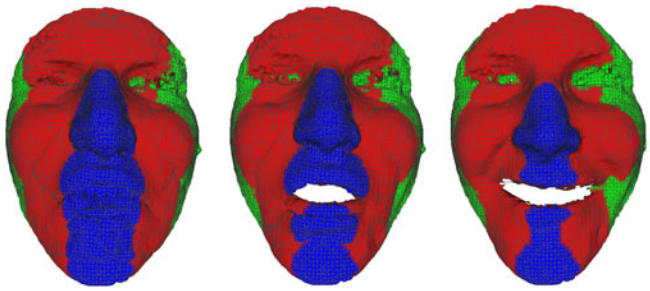

Fig. 2. Frontal slices from some emotional expressions samples: Anger, Fear, Disgust, Sad, Surprise, and Happy.

In summary, we adopted 6 forms to compute Hu moments: 3 forms with regard to the possible pairs of dependent axes XY, XZ and YZ and 2 possible functions for each pair. As for the 3DIT moments, there are 2 forms of calculation: one for each function. Therefore, we end up with 8 distinct groups of moments, each group of Hu moments with 7 moments and each group of 3DIT moments with 3 moments. Altogether, the feature vector comprises 48 measures . This number is dependent on parameter $\mathbf{c}$, because the thickness of a slice ultimately defines the amount of them. As we used $\mathbf{c} = (20, 20, 20)$, 4 frontal, 7 sagittal and 7 transversal slices were assigned 48 descriptors each, amounting 864 descriptors per face point cloud.

III. Preliminary Discussion

Several different databases have been used for the purpose of 3D facial recognition in the recent literature. We justify our choice of primarily working with Bosphorus 3D face database since it appears to be the one bringing the most challenging conditions. Additionally, we also evaluated the generalization of our method on FRGC v2.0 dataset.

Bosphorus Database was introduced by Savran *et al.* 2008 [19] and is now widely used 3D face recognition. The database provides 2D and 3D samples along with landmark coordinates placed on regions of Action Units (AU) [20]. It also provides the correspondence between 3D data and 2D image pixels, making it possible to work either in 2D or 3D spaces.

The database comprises 4666 face scans not equally distributed over 105 individuals in various expressions, poses and occlusion conditions. The subjects are aged between 25 and 35, and most of then are Caucasian. Among them, 29 actors integrate the database, giving it additional complexity since they can exaggerate expressions and thus push face deformation to high levels (see Fig. 2). Beard/mustache as well as short facial hair are present in some subjects.

The samples available in Bosphorus 3D face database can be divided in four wide categories: the ones performing expressions, poses, occlusions or with ignored label. Some subcategories are also derived from those and their descriptions are as follows: 34 Expressions (2902 samples); 13 Poses (1365 samples); 4 Occlusions (381 samples) and 1 Ignored (18 samples).

The FRGC v2.0 dataset, instead, comprises 466 subjects collected in 4,007 subject sessions [21] during the 2003-2004 academic year. It consists of 2,410 3D neutral samples and 1, 597 3D non-neutral samples of different facial expressions (disgust, happy, sadness, surprise, other).

As stated by Soltanpour et. al. [4], it is difficult to make a fair comparison between methods in the literature for FR, mainly because different experiments are presented in various situations and for different conditions and databases (i.e., number of subjects, samples and samples per expression vary significantly). To circumvent this, researchers frequently adopt similar figures-of-merit. For instance, the analysis of Receiver Operation Characteristic (ROC) curves, which consist of face verification experiments created by plotting the true positive rate (TPR) or verification rate (VR) against the false positive rate (FPR) at various threshold settings is a common choice for performance evaluation of FR algorithms. We proceeded similarly and decided to perform 2 kinds of experiments for FR, using Bosphorus 3D face database and FRGC v2.0, named ROC I and rank-1 experiments.

In ROC I experiment, the gallery set is composed of all Neutral samples while the probe set is composed of Non-Neutral samples. For Bosphorus database it means there are two sets: (a) Neutral gallery set with 299 samples and (b) Non-Neutral probe set with the rest 2603 samples. For FRGC v2.0, this experiment setup is (a) Neutral gallery set with 466 samples (b) Neutral probe set with 1, 944 samples and (c) Non-Neutral probe set with 1, 597 samples.

In rank-1 experiment, in turn, the gallery set comprises one Neutral sample per subject while there are two probe sets, one composed of the other Neutral samples and the other composed of Non-Neutral samples. For Bosphorus database those 3 sets are grouped as follows: (a) Neutral gallery set with 105 Neutral samples, one per subject; (b) Neutral probe set with the rest 194 Neutral samples and (c) Non-Neutral probe set with the rest 2603 expression samples. For FRGC v2.0, the gallery set has 466 Neutral samples, while the probe set has 3,451 Non-Neutral samples.

ROC I is the most typical experiment setup and the primary purpose of this paper. It allows a larger number of samples per subject to be included in the training set i.e., it includes variability per class in the input set. In Bosphorus 3D face database there is a small number of neutral samples. They represent around 6.4% of the whole database and 10% of all the Expression samples used in our work. Moreover, it has an unbalanced distribution of up to 4 samples per subject and the average is 2.8 samples per subject.

Rank-1 is a more challenging experiment setup, since it evaluates the classifier performance when only one sample per class is presented. With that restriction, the classifier has its ability to recognize individuals measured even when they perform expressions that cause significant deformations on face. As mentioned before in this paper, Bosphorus 3D face database included a relevant number of professional actors/actresses as subjects in order to guarantee high expressiveness.

Having defined the training and testing setup, the most suitable classifier should be selected. We selected two classifiers to perform our experiments: Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM), since both are frequently referred to in the literature as adequate to scenarios with few samples per class, which is specifically our case.

Linear Discriminant Analysis (LDA) has been commonly used for many applications such as FR as a dimensionality reduction technique [22] and as a classifier [23]. LDA is used in the present work as a classifier itself, and not for dimensionality reduction purposes prior to classification. LDA searches for basis vectors of a subspace onto which the data would be projected and that maximize the distinction between classes [24] and reduces the number of features down to the number of classes minus one. Various approaches have been proposed to solve this singularity problem, including applying Singular Value Decomposition (SVD) or eigenvalue decomposition to the data matrix and using iterative algorithms, such as LSQR to solve LDA as a least squares problem. The last one was the approach we have adopted in this work. But there is a constraint [24] while comparing PCA and LDA performances; Martinez [24] states that, in order the classifier not to become singular, the number of samples must be superior than the sum of number of classes and number of features. In summary, the number of samples must be, at least double the number of classes minus one. This is the reason why in this work LDA is used in ROC I, but not in rank-1 experiments.

Support Vector Machine (SVM) is originally a binary classifier that constructs an optimal linear decision surface based on a weighted combination of elements of the training set, the so called support-vectors [25], [26]. Training is formally defined by Cortes and Vapnik [25] as minimizing an error function which depends on two parameters (C, γ), where C is a constant and controls the trade-off between complexity of decision surface and frequency of misclassification in the training set [25]. Low values of C make the decision surface smooth and simple, while high values allow that more samples to be chosen as support vectors, creating a decision surface that tends to exactly separate classes in the training set. SVM can also be extended to construct a nonlinear decision surface by using a kernel function that projects the original space onto a high dimensional feature space. The kernel function used in this paper is the Radial Basis Function (RBF) kernel, where γ is a constant and in practice, adjusts how the decision surface is affected by each sample. Low values of γ mean that each sample has a far influence and therefore even far samples may affect the definition of the decision surface. Conversely, high values of γ mean that each sample has a close influence, which may force the decision surface to adapt to the closest samples. Parameters C and γ mainly influence the classifier's ability of generalization and therefore their appropriate tunning is fundamental. Unlike LDA classifier, no restrictions are imposed by the SVM in ROC I and rank-1 experiments. For what concerns the decision function shape, the implementation is twofold: *one vs all* (OvA) and *one vs one* (OvO); we use OvA in our experiments.

IV. Results and Discussion

The proposed method was developed using C/C++ OpenCV and PCL libraries on Ubuntu/Linux platform, and all the experiments are implemented on a PC with the CPU by Intel i7-5500U,

TABLE I
RR FOR NEUTRAL VS NONNEUTRAL RANK-1 EXPERIMENT WITH ST SLICES COMBINATION, USING SVM-SVC-RBF

Moments	4HuXY	4HuXZ	4HuYZ	3DIT
RR (%)	72.30	41.53	48.56	58.47
Moments (distance)	4HuXYd	4HuXZd	4HuYZd	3DITd
RR (%)	93.32	84.90	90.13	90.13

2.4 GHz, 8 GB RAM, using just one of the four cores. We found that the pose correction is the most time-consuming part, taking nearly up to 3 seconds (out of 4 seconds in total).

The computation of the HU moments and the 3DIT moments had time demand considerably decreased after the density normalization procedures, with PCL voxel grid filter of 2 mm leaf size for all the 3 dimensions.

To perform the tri-axial slicing, we tested the values $[5, 10, 15, 20, 25]$ in millimeters for the thickness $\mathbf{c} = (c_f, c_s, c_t)$. The best results were achieved with the triple $\mathbf{c} = (20, 20, 20)$. Lower values revealed to be very sensitive to rotation while higher values led to small quantity of slices and therefore insufficient amount of descriptors.

Next, we evaluated the descriptive capacity of the Hu moments and concluded that only the first 4 are essential for classification, for any function $f(x, y)$ tested. With that restriction, the maximum number of attributes was reduced down to 612. This improved the computation of descriptors, since the 3 last moments are more time-consuming than the first ones. Thus, we define 4HuXYd as the set of the first 4 Hu moments, taking in XY plane and function $f(x, y)$ computed as in Equation (9). Similar terminology was adopted for the other moments.

A. Best Combination of Slices and Moments

It is shown in Table I that features calculated according to the distance function of Equation (9) are individually better than those calculated from Equation (8). Moreover, the features HuXY are globally better and , regardless the function adopted in our experiments, the group of moments HuXZ performed worse, which suggests that the coordinate y as one of the dependent variables which considerably contributes to better results. Next, we also tested combinations of all the features and found that the best arrangement is 4HuXYd, 4HuXZd and 4HuYZd.

Once the best combination of moments has been defined, the next step would be to determine the combination of slices that yields the best recognition rate. We evaluated Frontal (F), Sagittal (S) and Transversal (T) slices individually and combined, forming the test set $[F, S, T, FS, FT, ST, FST]$.

Fig. 3 presents the performances for 4HuXYd + 4HuXZd + 4HuYZd attributes for all the possible combinations of Frontal (F), Sagittal (S) and Transverse (T) slicings. Frontal slicing alone is the worst scenario and, when combined to others, it degrades the VR. The best scenario is the combination Sagital-Transverse (ST). This trend repeats in the various cases investigated.

The behavior observed in Fig. 3 was also noticed for ROC I experiment with SVM-SVC-rbf classifier. For ROC I, we also employed LDA as classifier and the best combination of descriptors was 3DIT, 3DITF, 4HuXY, 4HuYZ, 4HuXYd and 4HuYZd. The SVM-SVC-rbf classifier performed worse than LDA.

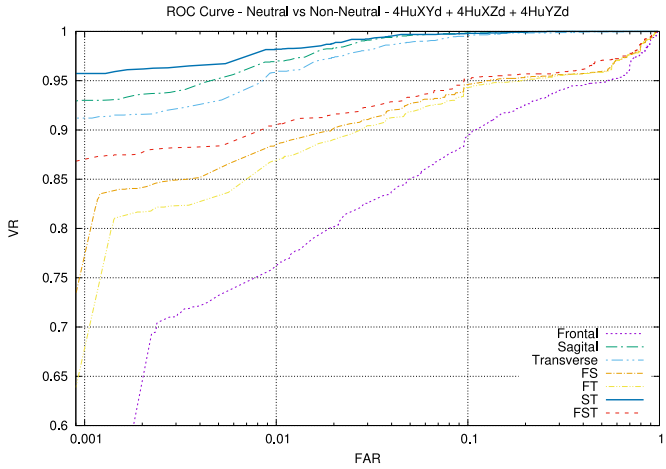

Fig. 3. VR for Neutral vs Non-Neutral rank-1 experiment with different slice combinations, using SVM-SVC-rbf.

TABLE II
VR VALUES AT 0.1% FAR

Databases	Bosphorus		FRGC v2.0	
Experiments	ROC I		ROC I	ROC III
References	RR (%)	VR (%)	VR (%)	VR (%)
This work (LDA*)	98.7	98.9	97.1	94.7
This work (SVM**)	85.3	87.0	84.2	87.9
Elaiwat et al [6]	n/a	91.0	98.0	97.8
Ocegueda et al [10]	n/a	93.8	98.1	97.9
Ming et al [27]	92.0	94.0	n/a	n/a
Liu et al [5]	n/a	95.6	93.1	n/a
Emambakhsh [11]	n/a	n/a	n/a	93.5
Al-Osaimi, Faisal R.[12]	n/a	n/a	97.8	94.1
Berreti et al [7]	n/a	n/a	91.4	86.6

Bosphorus ROC I experiment with VR and RR for Neutral vs NonNeutral, and FRGC v2.0 ROC I experiment with VR for Neutral vs NonNeutral and ROC III experiment with VR.

B. ROC I and rank-1 Experiments

In this subsection, we discuss the best experimental results for face recognition achieved in this work and compare them to the literature. Initially, results for ROC I are brought. To make fair the comparison the verification rate (VR) is analyzed. We then present results for rank 1 experiment, since it appears very often in recent works. In this case, recognition rate (RR) is used as performance indicator.

For ROC I, we utilized LDA classifiers. Our results are listed in Table II and refer to the ST slicing with 3DIT, 3DITF, 4HuXY, 4HuYZ, 4HuXYd and 4HuYZd descriptors forming a 308-long feature vector. It is important to mention that, eventhough LDA for ROC I experiment has a larger number of input descriptors, internally it selects a number of attributes equals to the number of classes minus 1, as explained in [24].

For what concerns Bosphorus database, Table II shows that the verification rate achieved in this work is superior to the ones reported by Elaiwat *et al.* [6], Liu *et al.* [5], Ocegueda *et al.* [10] and Ming *et al.* [27]. Ming *et al.* [27] also neglected samples with

occlusion and pose variation, but included IGN samples (only 18 samples). Elaiwat [6] and Liu [5], instead, included samples with occlusion and pose variation in the experiments; on doing this, symmetric filling [6] [5] or hole filling [5] techniques were used. Liu [5] only neglected 90° rotated samples. Although the testing sets employed in Elaiwat [6] e Liu [5] are very comprehensive, both works need to find the facial symmetry plane to achieve face completion, which ultimately adds complexity to their solution. Moreover, [6] uses ICP for fine alignment. In these experiments, the feature vector size is 252 units-long for LDA and SVM-SVC-rbf classifiers; LDA was solved by SVD method using the pack *sklearn* in python language, and SVM parameters were set $C = 8$ and $\gamma = 0.125$.

To assess the generalization of the proposed method, scans from the FRGC v2.0 dataset were used in the pretty same conditions of the experiments reported so far for the Bosphorus database; results are reported in Table II. Once again, the proposed method performs well, with rates comparable to important recent references and reaching the top 3 among those. Here we report ROC I experiment (as described earlier) and also a ROC III face verification experiment, in which the training includes the scans of the Fall (Autumn) 2003 sessions, whereas the probe scans contains shots from the Spring 2004 sessions. Due to this temporal lapse between the acquisition of probe and training scans, this experiment is regarded as the most difficult one of the FRGC protocol as pointed out by Berretti *et al.*[7].

As mentioned earlier, for rank 1 we only use SVM-SVC-rbf classifier where parameters were set $C = 8$ and $\gamma = 0.125$. This experiment is especially important to place the proposed method among those using training/testing strategy. A major advantage of this comparison is that it can be detailed, since the literature provides recognition rates for the subset of non-neutral samples. It allows to evaluate more specifically each method in regard to emotional expressions, neutral, LFAU, UFAU and CAU samples.

The results shown in Table III take into account only the ST slicings with moments 4HuXYd. 4HuXZd and 4HuYZd, for a total of 168 descriptors. In some some cases, our method reaches or surpasses the best results reported by Berreti *et al.* [7], Li *et al.* [1], Ming *et al.* [27], Emambakhsh and Evans [11] and Al-Osaimi and Faisal [12]. Those cases are highlighted in bold. Also for the set of expressions Anger and Disgust our method performed better than Li *et al.* results. We also beat Berreti *et al.* [7] for Neutral, Fear, Sadness, Surprise and CAU samples.

In addition to the results, some conceptual and practical aspects should be considered at this point. Unlike Berreti's method, which is not targeted for Kinect or low-resolution low-cost cameras, our method does well with this acquisition system. Indeed, in previous investigation, our method did well when using the 3DMAD [28] database of samples based on Kinect; using Multi Layer Perceptron Neural Network (MLP-NN) as classifier, we found recognition rates equivalent to those obtained for Bosphorus in neutral vs neutral experiment, which is consistent to the fact that 3DMAD only contains neutral expressions. This suggests how robust the method is to different data acquisition systems. Furthermore, as mentioned earlier in this paper, the present method has a cloud density regularization step based on voxel grid; this is especially important to overcome differences in the scanner resolution.

TABLE III
RR FOR DIFFERENT PROBE CATEGORIES ON BOSPHORUS RANK1 EXPERIMENT

Descriptors(#)	-	-	130	-	-	168
Key Points(#)	648	145	2	-	-	1
Probes(#)	Li et al [1]	Berreti et al [7]	[27]	[11]	[12]	This work (SVM)
Bosphorus (rank1 RR(%))						
Neutral (194)	100	98.5	94.3	99.0	98.5	**99.0**
NonNeutral (2603)	98.7	96.6	92.0	n/a	92.4	95.4
Emotions (453)	96.7	91.2	n/a	96.2	n/a	**95.2**
Anger (71)	97.2	88.7	n/a	94.1	n/a	**98.6**
Disgust (69)	86.9	81.2	n/a	88.2	n/a	**98.6**
Fear (70)	98.6	91.4	n/a	98.6	n/a	**95.7**
Happy (106)	98.1	94.3	n/a	98.1	n/a	87.7
Sad (66)	100	95.5	n/a	96.9	n/a	**98.5**
Surprise (71)	98.6	94.4	n/a	100	n/a	95.8
LFAU (1549)	98.8	97.5	n/a	n/a	n/a	94.7
UFAU (432)	100	99.1	n/a	n/a	n/a	98.4
CAU (169)	100	96.4	n/a	n/a	n/a	**98.8**

NOTE: Ming *et al.* [27], Emambakhsh and Evans [11] and Al-Osaimi and Faisal [12].

Finally, Li *et al.* [1] extended the SIFT-like matching framework to mesh data and proposed a novel approach using fine-grained matching of 3D keypoint descriptors, called Histogram of Multiple surface differential Quantities (HOMQ) by combining HOG, HOS, and HOGS at feature level. Although Li reported the best performances for non-neutral set, the high quantity of required keypoints resulted in significant computational efforts; his method takes on average 1.25 minutes on a PC with Intel 950, 3.07 GHz processor, 8 GB RAM. The method here introduced, running under similar conditions (Intel i7-5500U processor, 2.4 GHz) and employing OpenCv and PCL libraries for C/C++ and takes less than 4 seconds to perform the entire recognition process.

V. CONCLUSION AND FUTURE WORKS

In this work a technique for 3D face recognition relying on adapted Hu and 3D invariant moments of tri-axial point cloud slices was introduced. The combination of 4HuXY, 4HuXY, 4HuYZd, 4HuXYd, 3DIT and 3DITd attributes reached verification rate of 98.7% for 0.1% FAR in ROC I experiment using LDA classifier. For rank-1, using 4HuXYd, 4HuXZd and 4HuYZd, experiment the identification rates of 99% and 95.4% for the Neutral vs Neutral and Neutral vs Non-Neutral experiments, respectively, which reinforces that the use of function distance improves the characterization of individuals. Identification rates above 97% for the Neutral vs Non-Neutral in the subsets CAU, UFAU, Anger, Disgust and Sad were achieved. The technique only needs to identify a single key point to cut the face into 14 different cloud points in Sagittal and Transversal planes, which makes it computationally less complex than other approaches of the same nature. The Sagittal and

Transverse slices showed to be less sensitive to surface variations even for certain types of expressions, while the Frontal slices, in the vast majority of cases, degraded the recognition rate when associated with the other two orthogonal planes.

The method introduced here can be additionally improved and made more general if the following limitations are addressed: a) occlusion and self-occlusion: although were neglected in the current version, symmetry of human face can be exploited for dealing with this issue; b) the dependence on initial alignment: the method can be made faster as new approaches for the face orientation correction are made available. We emphasize that the essence of our method relies on the proposal of a new feature descriptor; and, finally, c) the dependence on cloud density: Hu moments are sensitive to density variation of point clouds. We addressed this issue in the current version by applying the voxel-grid algorithm in order to equalize the density for all samples, however other strategies may be tried.

REFERENCES

[1] H. Li, D. Huang, J.-M. Morvan, Y. Wang, and L. Chen, "Towards 3d face recognition in the real: A registration-free approach using fine-grained matching of 3D keypoint descriptors," *Int. J. Comput. Vis.*, vol. 113, no. 2, pp. 128–142, Jun. 2015.

[2] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, 2003.

[3] H. Hatem, Z. Beiji, and R. Majeed, "A survey of feature base methods for human face detection," *Int. J. Control Autom.*, vol. 8, no. 5, pp. 61–78, 2015.

[4] S. Soltanpour, B. Boufama, and Q. Jonathan Wu, "A survey of local feature methods for 3D face recognition," *Pattern Recognit.*, vol. 72, no. C, pp. 391–406, Dec. 2017.

[5] P. Liu, Y. Wang, D. Huang, Z. Zhang, and L. Chen, "Learning the spherical harmonic features for 3-D face recognition," *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 914–925, Mar. 2013.

[6] S. Elaiwat, M. Bennamoun, F. Boussaid, and A. El-Sallam, "A Curvelet-based approach for textured 3D face recognition," *Pattern Recognit.*, vol. 48, no. 4, pp. 1235–1246, 2015.

[7] S. Berretti, N. Werghi, A. del Bimbo, and P. Pala, "Selecting stable keypoints and local descriptors for person identification using 3D face scans," *Vis. Comput.*, vol. 30, no. 11, pp. 1275–1292, Nov. 2014.

[8] J. Gao and A. N. Evans, "Expression robust 3D face landmarking using thresholded surface normals," *Pattern Recognit.*, vol. 78, pp. 120–132, Jun. 2018.

[9] H. Drira, B. B. Amor, A. Srivastava, M. Daoudi, and R. Slama, "3D face recognition under expressions, occlusions, and pose variations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 9, pp. 2270–2283, Sep. 2013.

[10] O. Ocegueda, T. Fang, S. K. Shah, and I. A. Kakadiaris, "3D face discriminant analysis using Gauss-Markov posterior marginals," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 728–739, Mar. 2013.

[11] M. Emambakhsh and A. Evans, "Nasal patches and curves for expression-robust 3D face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 5, pp. 995–1007, May 2017.

[12] F. R. Al-Osaimi, "A novel multi-purpose matching representation of local 3D surfaces: A rotationally invariant, efficient, and highly discriminative approach with an adjustable sensitivity," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 658–672, Feb. 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7300438/

[13] E. B. Corrochano, *Handbook of Geometric Computing*. New York, NY, USA: Springer, 2005.

[14] R. B. Rusu and S. Cousins, "3D is here: Point cloud library (PCL)," in *Proc. 2011 IEEE Int. Conf. Robot. Automat.*, May 2011, pp. 1–4.

[15] D. Borrmann, J. Elseberg, K. Lingemann, and A. Nüchter, "The 3D Hough transform for plane detection in point clouds: A review and a new accumulator design," *3-D Research*, vol. 2, no. 2, pp. 1–13, Nov. 2011.

[16] M. Beka, G. Oka, M. Vajteric, and L. Grigori, "On iterative QR pre-processing in the parallel Block-Jacobi SVD algorithm," *Parallel Comput.*, vol. 36, no. 5, pp. 297–307, 2010.

[17] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Inf. Theory*, vol. 8, no. 2, pp. 179–187, Feb. 1962.

[18] T. Suk, J. Flusser, and J. Boldy, "3D rotation invariants by complex moments," *Pattern Recognit.*, vol. 48, no. 11, pp. 3516–3526, 2015.

[19] A. Savran *et al.*, "Bosphorus database for 3D face analysis," in *Biometrics and Identity Management*, B. Schouten, N. C. Juul, A. Drygajlo, and M. Tistarelli, Eds. Berlin, Germany: Springer, 2008, pp. 47–56.

[20] P. Ekman and W. V. Friesen, "Measuring facial movement," *Environ. Psychol. Nonverbal Behav.*, vol. 1, no. 1, pp. 56–75, 1976.

[21] P. J. Phillips *et al.*, "Overview of the face recognition grand challenge," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, vol. I, pp. 947–954.

[22] H.-M. Moon, D. Choi, P. Kim, and S. B. Pan, "LDA-based face recognition using multiple distance training face images with low user cooperation," in *Proc. IEEE Int. Conf. Consum. Electron.*, 2015, pp. 7–8.

[23] A. Tzalavra *et al.*, "Comparison of multi-resolution analysis patterns for texture classification of breast tumors based on DCE-MRI," in *Proc. Int. Workshop Mach. Learn. Med. Imag*. Springer, 2016, pp. 296–304.

[24] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 228–233, Feb. 2001.

[25] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[26] M. H. Mousavi, K. Faez, and A. Asghari, "Three dimensional face recognition using SVM classifier," in *Proc. 7th IEEE/ACIS Int. Conf. Comput. Inf. Sci.*, 2008, pp. 208–213.

[27] Y. Ming, "Rigid-area orthogonal spectral regression for efficient 3D face recognition," *Neurocomputing*, vol. 129, pp. 445–457, 2014.

[28] N. Erdogmus and S. Marcel, "Spoofing in 2D face recognition with 3D masks and anti-spoofing with kinect," in *Proc. 2013 IEEE 6th Int. Conf. Biometrics, Theory, Appl. Syst.* 2013, pp. 1–6.