A critical assessment of 2D and 3D face recognition algorithms

D. Giorgi, M. Attene, G. Patanè, S. Marini,
C. Pizzi, S. Biasotti, M. Spagnuolo, B. Falcidieno
IMATI-CNR
Via De Marini 6, I-16149 Genova, Italy
http://www.ge.imati.cnr.it

M. Corvi, L. Usai, L. Roncarolo, G. Garibotto
Elsag-Datamat
Via dei Pescatori 35, I-16128 Genova, Italy
http://www.elsagdatamat.com

***Abstract*—We present the results of a project aimed to evaluate 2D and 3D face recognition algorithms. In particular, we focused on the potentialities of 3D-based techniques to overcome typical limitations of 2D methods in non-controlled situations. According to the reference scenario of people identification at airport check points, we built a representative database on which we tested different face recognition algorithms. We implemented and tested an improved version of a well-known state-of-the-art 3D approach, and verified that on our dataset it performs better than a widely used commercial system.**

***Keywords*-isometry invariance; facial expressions; 3D meshes;**

I. Introduction

Biometric identification techniques are needed to build advanced security systems, for the surveillance of high-risk areas such as airports. At the same time, as observed by Max Snijder (European Biometrics Forum), "biometrics is entering our daily life": examples are keyless entry, and televisions that could use biometrics to identify people and match them to their personal settings [1].

Biometric identification techniques analyse physical or behavioural characteristics. Face recognition, in particular, is the recommended identification modality at the airports' gates, according to the ICAO (International Civil Aviation Organization): it is not invasive, it does not imply the contact with any objects, it does not require a passenger to remember a password. The most popular face recognition techniques, both in the academic and the commercial areas, analyse 2D data, i.e., images of faces acquired by means of cameras. Under controlled conditions, such systems attain very good results, whereas their trustworthiness is compromised by changes in the illumination; occlusions of landmark points such as pupils; changes in facial pose or expression [20]. In addition, image-based systems may be deceived by malevolent people using photos or videos [8].

For all these reasons, the research community has started focusing on the use of 3D data: 3D models are less sensitive to illumination conditions, are independent of the subject's pose, and carry information on fundamental face features [6].

In this paper, we present the early results within a joint project IMATI - Elsag Datamat. Our aim was the assessment of the potentialities of 3D models for face recognition at the airport check-points. We analysed the scenario and built a database to cover the different cases that may occur: changes in pose, expression, illumination, and time effects. We compared the results of a method to compare 3D face models, which we built upon the current state-of-the-art, with the results of a widely used commercial face recognition system by COGNITEC.

The paper is organized as follows: in Section II, we briefly review previous work and in Section III we describe the construction of the database used to test our methodology. Then, Section IV introduces the proposed approach and Section V discusses our results. Finally, Section VI concludes the paper and outlines future work.

II. Related work

We briefly summarize 3D face recognition techniques and methods integrating images and 3D data.

The earlier works on 3D face recognition date back to the nineties, whereas the number of published works has been significantly growing since 2000. Detailed surveys can be found in [6], [21], [14]. Many approaches make use of ICP (Iterative Closest Point) algorithms to align the surface meshes that represent the faces: the algorithms are run either on the whole face model [18] or on (possibly superimposing) face sub-regions [9].

The main drawback of many of these approaches is the assumption that human faces are rigid objects. Bronstein *et al.* deal with the problem of face deformation caused by different expressions [7]. They propose an isometry invariant comparison technique based on geodesic distances, which are supposed to be insensitive to expression changes.

The potentialities of combining 2D and 3D data for face recognition are well acknowledged in the literature [6], [10], [4]. In any case, most approaches are not truly multi-modal, because they limit themselves to combine *a posteriori* the outcome of 2D and 3D descriptions, used in an independent way. Beumier *et al.* [2] fuse the outcome at the score level, by a weighted sum of the distances obtained comparing the images and the distances obtained comparing the 3D models. Wang *et al.* [17] fuse the modalities at the feature level, by constructing a feature vector comprising both features computed on images and features computed on 3D face models.

III. Database constructions

During the first phase of our experimentation, we have created a database of 3D face models using a non-contact 3D digitizer VIVID 910 by Konica-Minolta, with a configuration of the acquisition parameters that guarantees a standard and homogeneous data quality. The data has been acquired using only one scan, imposing that the distance between the face and the scanner is in-between 50 and 100 cm, and providing an output in which the inter-pupil distance is at least 100 pixels (COGNITEC recommendation). Each scan took 2.5 seconds and generated a 640×480 bitmap image and a corresponding 640×480 distance map.

To provide a database able to portrait the variety of situations that may occur in the context of face recognition, for each face model we have acquired several poses and expressions, under different illumination conditions and at different time periods. Indeed, this choice is intended to simulate different scenarios and test the pros/cons of the methods. We acquired the face models during two different scanning sessions, producing the two sets S_1 and S_2:

1) $\mathcal{S}_1$ includes 8 faces with 4 different poses, namely frontal, up, down, left view, with a rotation of the face up to 15 degrees and a neutral expression;
2) $\mathcal{S}_2$ includes a set of face models that are typical of real situations, with different facial expression and worst illumination conditions. In detail, the same 8 subjects were requested to smile, to perform an arbitrary expression (e.g. being angry or disgusted) and to move their eyes down; the light source was put in a central or a lateral position. For each subject, 5 different poses have been acquired.

The final database is $\mathcal{S} := \mathcal{S}_1 \cup \mathcal{S}_2$. It has 72 models, i.e., 8 faces with 9 different poses. Besides pose and illumination changes, the model acquisition with two distinct scanning sessions allows us to simulate a scenario in which the recognition of a face must be performed between facial data acquired in different period of time, as it happens when comparing a just acquired face with the face stored in a document.

To compare 2D with 3D face recognition techniques, we have stored for each face both the bitmap image and the corresponding three-dimensional model, represented as a triangle mesh. To produce a smooth, manifold triangle mesh with only one boundary component, we have applied a set of processing techniques which include the *Laplacian smoothing* [24]; a *pose normalization* step [25] that aligns each face within a canonical coordinate system; and a *mask generation* [5] based on the geodesic cropping in such a way that the radius of the geodesic circle centered in the nose is 10 centimeters. The preprocessing pipeline, from raw data to the smooth mask, is shown in Figure 1.

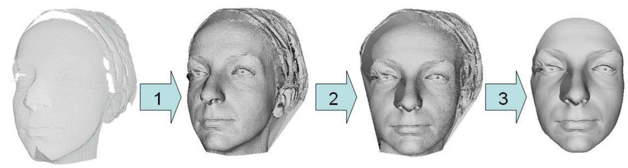

Figure 1. The various stages of the pre-processing. From raw point cloud to Delaunay triangulation (1), to pose-normalized mesh (2) and finally to cropped and smoothed mesh (3).

IV. 3D face description and comparison

The idea behind the method proposed by Bronstein *et al.* [7] is to convert the acquired 3D facial model into a new signature, which is invariant to facial expressions. Assuming that facial expressions are invariant to isometric transformations, it is enough to define an embedding that is isometry-invariant. To this end, the signature is computed as the optimal embedding of the face in $\mathbf{R}^3$ that best approximates the geodesic distances among a set of point samples on the input surface. These points are usually identified by simplifying the triangle mesh of the face model to a low resolution with a number of vertices in-between 1000 and 2000. A simple and effective way to compute the optimal embedding is provided by the *Multi-Dimensional Scaling* (MDS, for short) algorithm, which was applied to the analysis of 3D shapes in [15].

Once the embeddings $\mathcal{E}_i$ of each face $\mathcal{F}_i$ have been computed, we compare two embeddings $\mathcal{E}_i$ and $\mathcal{E}_j$ by using a metric $d(\mathcal{E}_i, \mathcal{E}_j)$. Among the several choices of the metric d, previous works [15] proposed the use of the Euclidean norm between the *geometric moments* [13], which might be unstable and have a low discriminant capacity, as shown in [11]. Hence, we have used the isometry-invariant embedding as the input for the computation of a new and more robust facial descriptor, which is achieved by applying the *spherical harmonics* (SH) introduced in [16]. In this way, we are able to include in the descriptor two main features: the invariance to isometries provided by the MDS, and a local characterization of the embedded geometry, given by the SH. The pipeline for 3D face description and comparison includes the following steps:

1) uniform simplification of the input facial model $\mathcal{F}_i$ to a lower resolution $\mathcal{F}_i^{'}$ with k points, and computation of the $k \times k$ matrix $A := (a_{ij})_{i,j=1}^{k}$ such that a_{ij} is the geodesic distance between the couple of points $\mathbf{p}_i$, $\mathbf{p}_j$ of $\mathcal{F}_i^{'}$;
2) computation of the embedding $\mathcal{E}_i^{'}$ of $\mathcal{F}_i^{'}$ using the MDS. In this case, $\mathcal{E}_i^{'}$ is computed using the couple $(\mathcal{F}_i^{'}, A)$ and running a fast iterative scheme. Since $\mathcal{F}_i^{'}$ is independent of isometric deformations of $\mathcal{F}_i$, we get that different expressions of the same face will have very similar embedded surfaces, as shown in Figure 2;
3) computation of the SH of $\mathcal{E}_i^{'}$. The resulting descrip-

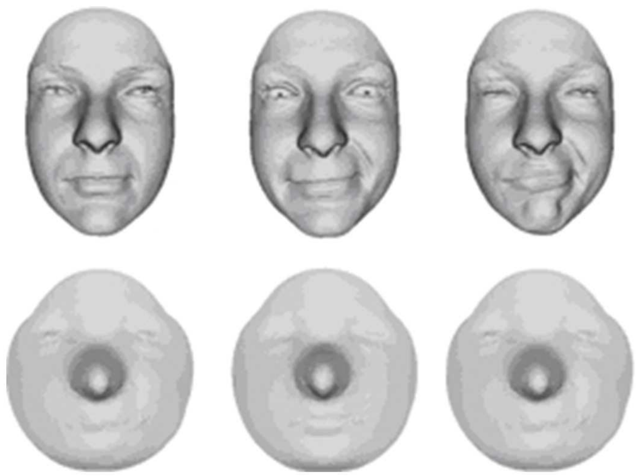

Figure 2. Three different expressions of the same person (top row) are transformed to three nearly identical shapes through isometric embedding (bottom row).

tor is a vector of coefficients, which correspond to different frequencies of a set of spherical functions defined on the volume around the input shape. Then, the frequencies of two models are compared using the Euclidean distance.

For the computation of the geodesic distances, we used the *Dijkstra* [12] and *fast marching* method [23]. For the computation of the multidimensional scaling, we have proposed an optimization with respect to the standard implementation, which stores the information necessary to compute the embedding in a unique $k \times k$ matrix that is used for the whole database without recomputing it for each face model. In this way, we use only $O(k^2)$ memory allocations and $O(k)$-time for each comparison. For the computation of the spherical harmonics, we used the implementation available at http://www.cs.jhu.edu/misha/.

V. Experimental Results

In this section, we present the results obtained through the method described in the previous section. Such results are compared against those obtained by the COGNITEC software available at Elsag-Datamat. For the comparison, COGNITEC was used in two different modalities: one using only 2D information (i.e. face bitmap images), and the other exploiting both 2D and 3D distance information captured by the scanner. The dataset S is made of 72 faces (8 individuals in 9 different poses each), as described in Section III. For each method, we built a similarity matrix in which the entry (i, j) corresponds to the similarity evaluation between the i-th and the j-th face. During the experiments we observed that the COGNITEC software failed to process 4 images of our dataset, shown in Figure 3. We argue that this is due to a failure in identifying the eye pupils in the photographs. Thus, we removed such images from

Figure 3. Four images rejected by the COGNITEC system.

the dataset, along with the corresponding 3D data. As a consequence, the results reported in the remainder are referred to a dataset containing 68 faces only, in which 4 of the 8 individuals are present in 9 different poses, whereas the remaining 4 are present in 8 poses.

The performance evaluation of the various methods is based on the analysis of the *precision-recall* graph. Such graph shows the standard precision (y axis) as a function of the standard recall (x axis). The *precision* represents the percentage of relevant elements within the total retrieved elements, whereas the *recall* represents the percentage of retrieved relevant elements (true positives) within the total relevant elements [22]. Thus, these parameters describe the ability of a method of identifying relevant models (in our case the various face models of a given single individual) and, at the same time, reducing the number of false positives. Better performances correspond to graph curves pushed towards the top-right of the diagram.

The graph in Figure 4 reports the curves related to the COGNITEC software in its two running modes (2D and 2D+3D), and the curves related to the method of the spherical harmonics computed on the isometric embedding ('SH embedded'), on the original model ('SH orig') and on a simplified version of the model uniformly re-sampled with 2000 vertices ('SH 2000v'). Note that the results obtained by comparing the geometrical moments of the isometric embeddings, as suggested in [7], led to very poor performances, not shown.

It can be observed that, on our dataset, when 3D information is provided (2D+3D modality) the COGNITEC software performs better than its 2D modality (2D only).

The experimentation with different vertex counts of the models is useful to verify possible performance degradations due to a varying resolution, with the objective of finding the best trade-off between quality of the results and request of computational time and resources. The performances of the

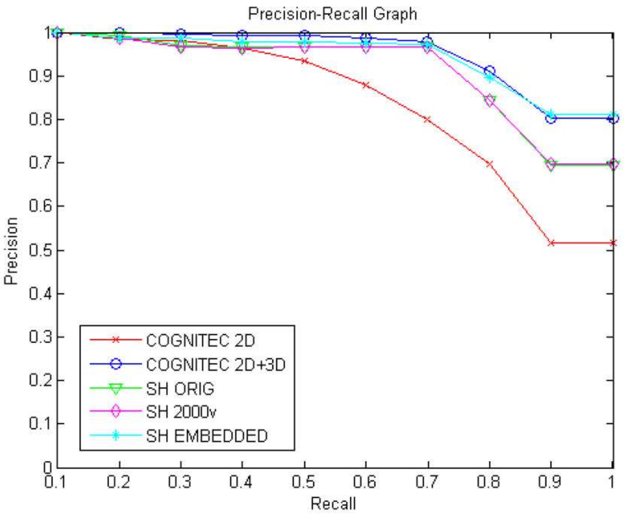

Figure 4. Precision-recall curves of the experimented algorithms.

COGNITEC 2D	COGNITEC 2D+3D	SH embedded
VR: 67.6%	88.1 %	91.6%

Table I
VERIFICATION RATE AT 0.1% FALSE ACCEPTANCE RATE

SH method on the original and simplified models are rather similar, indicating that the approach is stable to resolution and level-of-detail changes.

The overall best results were obtained by computing the SH descriptors on the isometric embeddings. Such embeddings are indeed built to achieve invariance with respect to face deformations, from small facial movements (face pose is never perfectly neutral [7]) to more evident expression changes, as shown in Figure 2. Hence, the combined use of an expression-invariant input and an effective comparison method led us to the best performances. The Verification Rate at 0.1% false acceptance Rate is given in Table I.

Our analysis leads to the idea that 2D recognition is mature enough to guarantee excellent results under controlled conditions; nonetheless, 2D recognition has intrinsic limitations such as the information loss due to the 2D representation of a 3D face surface and, most important, the conditions under which the digitization takes place. Extreme illumination conditions may spoil the performances of pure image-based methods (Figure 5), and in some cases may even lead the system to reject the input (Figure 3). Conversely, the analysis of 3D data appears to be robust even when 2D methods fail due to the sensorial gap, as 3D data is a more comprehensive representation of the real world. Notice that the impact of the research of novel methodologies for 3D recognition will depend also on the advancements of digitization systems, which will probably lead to accurate, fast and reliable sensors [19].

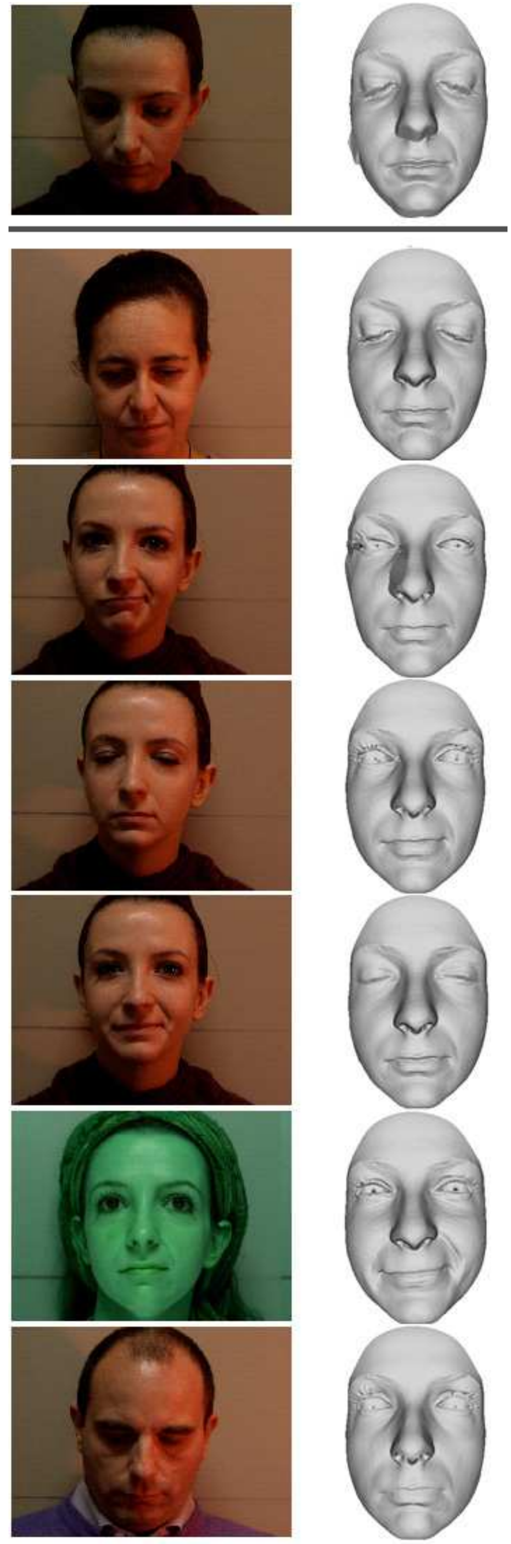

Figure 5. Results of a query given by the COGNITEC in 2D modality (left, 2 false positives) and by the SH method (right, no false positives).

VI. Conclusions

We have presented the early results of a study on face comparison via 3D shapes, and compared them with the results of a face recognition commercial system. Note that we built our own database to investigate well-focused problems: although very small, our database has proven to be useful to verify our assumptions. In the future, we plan to perform a deeper experimentation using widely used databases, such as the FRCG (Face Recognition Grand Challenge).

According to our study, a promising direction seems to be the integration of image analysis methodologies and 3D techniques, so as to combine the information coming from different sources and achieve a better degree of results trustworthiness in the various environmental conditions (illumination, pose and expression changes, malevolent subjects).

We will focus our future research on multi-modal algorithms for face recognition, integrating 2D and 3D data. First of all, we will study methods to use 3D models to support 2D techniques based on machine learning. Indeed, it is well known that the results of these algorithms depend upon the size of the training set and the shape variability there represented: deformed 3D shapes can be used to generate synthetic images of faces, to increase the number of faces in the training set without asking for the collaboration of humans. Then, 3D models can be used to integrate the 2D information which is lost due to the sensory gap, for example to reconstruct occluded or shady parts, or to detect the iris under adverse illumination conditions.

Finally, we plan to go beyond the integration of the 2D and the 3D information at the score level or the feature level, by defining a face from the beginning as a pair $(\mathcal{F}, \varphi)$ [3], where $\mathcal{F}$ represents the face geometry and φ is a function of either shape properties (e.g., curvature), or texture and color, or both.

Acknowledgment

This work is partly supported by the EU FP7 project *FOCUS K3D* and by the research contract Elsag - CNR-IMATI-GE on ”Multitrust: multi-biometry for applications to security”.

References

[1] Password - philips research technology magazine, 2007. issue 30.

[2] C. Beumier and M. Acheroy. Face verification from 3D and grey level clues. *Pattern Recognition Letters*, 22:1321–1329, 2001.

[3] S. Biasotti, L. De Floriani, B. Falcidieno, P. Frosini, D. Giorgi, C. Landi, L. Papaleo, and M. Spagnuolo. Describing shapes by geometrical-topological properties of real functions. *ACM Computing Surveys*, 40(4):12:1–12:87.

[4] S. Berretti, A. Del Bimbo, P. Pala, F. Silva-Mata. Face recognition by SVMS classification of 2D and 3D radial geodesics. *Proceedings of ICME 2008*, pages 93–96, 2008.

[5] V. Blanz, K. Scherbaum, and H.-P. Seidel. Fitting a morphable model to 3D scans of faces. In *International Conference on Computer Vision (ICCV 2007)*, 2007.

[6] K. W. Bowyer, K. Chang, and P. Flynn. A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition. *Computer Vision and Image Understanding*, 101(1):1–15, 2006.

[7] A. M. Bronstein, M. M. Bronstein, , and R. Kimmel. Threedimensional face recognition. *international Journal of Computer Vision*, 64(1):5–30, 2005.

[8] C. Busch and A. Nouak. 3D face recognition for unattended border control. In H. R. Arabnia and S. Aissi, editors, *Security and Management*, pages 350–356. CSREA Press, 2008.

[9] K. Chang, K. Bowyer, and P. Flynn. Face recognition using 2D and 3D facial data. In *Multimodal User Authentication, IEEE Workshop on Face Recognition Grand Challenge Experiments*, pages 25–32, 2005.

[10] A. Colombo, C. Cusano, R. Schettini. Face3 a 2D+3D robust face recognition system. *Proceedings of 14th IEEE International Conference on Image Analysis and Processing (ICIAP) 2007*, pages 393–398, 2007.

[11] A. Del Bimbo and P. Pala. Content based retrieval of 3D models. *ACM Transactions on Multimedia Computing, Communications and Applications*, 2(1):1–24, 2006.

[12] E. W. Dijkstra. A note on two problems in connection with graphs. *Numerical Mathematics*, 24(1):269–271, 1959.

[13] M. Elad, A. Tal, and S. Ar. Content based retrieval of VRML objects - An iterative and interactive approach. In *EG Multimedia*, pages 97–108, 2001.

[14] R. I. Hammoud, B. R. Abidi, M. A. Abidi. Face Biometrics for Personal Identification. Springer, ISBN:3540493441, 2007.

[15] A. E. Elbaz and R. Kimmel. On bending invariant signatures for surfaces. *IEEE Trans. on PAMI*, 25(10):1285–1295, 2003.

[16] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz. Rotation invariant spherical harmonic representation of 3D shape descriptors. In *Proceedings of SGP 2003: Eurographics Symposium on Geometry Processing*, pages 156–165, 2003.

[17] K. I. Kim, K. Jung, and H. J. Kim. Face recognition using kernel principal component analysis. *IEEE Signal Processing Letters*, 9(2):40–42, 2002.

[18] G. Medioni and R. Waupotitsch. Face recognition and modeling in 3D. In *IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG 2003)*, pages 232–233, 2003.

[19] J. P. Neugebauer. Research on 3D sensors - minimizing the environmental impacting factors. In *Workshop Biometrics and eCards*, 2007.

[20] A. Razdan, G. Farin, M. Soo-Bae, and M. Chaudhari. State of 3D face biometrics for homeland security applications, 2007. In Chen, Santanam, Ramesh, Vinze and Zeng, Eds.: Handbooks in Information Systems, Vol. 2.

[21] A. Ruifrok, A. Scheenstra, and R. C. Veltkamp. A survey of 3D face recognition methods. In *Audio- and Video-based Biometric Person Authentication (AVBPA 2005)*, volume LNCS 3546, pages 891–899, 2005.

[22] G. Salton and M. McGill. *Introduction to modern information retrieval*. McGraw Hill, 1983.

[23] J. A. Sethian. A fast marching level set method for monotonically advancing fronts. *Proceedings of the National Academy of Science*, 93(4):1591–1595, 1996.

[24] G. Taubin. Geometric signal processing on polygonal meshes. In *Proc. Eurographics 2000 State of the Art Report*, 2000.

[25] F. B. ter Haar and R. C. Veltkamp. A 3D face matching framework. Technical Report Technical Report UU-CS-2007-047, 2007. http://www.cs.uu.nl/research/techreps/repo/CS-2007/2007-047.pdf.