Boosting 3D LBP-Based Face Recognition by Fusing Shape and Texture Descriptors on the Mesh

Naoufel Werghi, *Senior Member, IEEE*, Claudio Tortorici, Stefano Berretti, *Member, IEEE*, and Alberto Del Bimbo, *Senior Member, IEEE*

***Abstract*— In this paper, we present a novel approach for fusing shape and texture local binary patterns (LBPs) on a mesh for 3D face recognition. Using a recently proposed framework, we compute LBP directly on the face mesh surface, then we construct a grid of the regions on the facial surface that can accommodate global and partial descriptions. Compared with its depth-image counterpart, our approach is distinguished by the following features: 1) inherits the intrinsic advantages of mesh surface (e.g., preservation of the full geometry); 2) does not require normalization; and 3) can accommodate partial matching. In addition, it allows early level fusion of texture and shape modalities. Through experiments conducted on the BU-3DFE and Bosphorus databases, we assess different variants of our approach with regard to facial expressions and missing data, also in comparison to the state-of-the-art solutions.**

***Index Terms*— Mesh-LBP, feature and score fusion, 3D face recognition.**

I. Introduction

THE LAST decade has seen an extensive investigation of 3D face image usage for human identification. Adding to shape information the intrinsic features characterizing facial image, such as universal acceptance and non-invasiveness, 3D face image has emerged as promising modality addressing the limitations of its 2D counterpart, such as pose and luminance variation, while opening-up new horizons for enhancing the reliability of face-based identification systems. This trend has been further fueled by the advances in 3D scanning technology, which provides now 3D textured scans encompassing aligned shape and photometric data.

Since their introduction in the mid '90, Local Binary Patterns (LBP) [2] have been extensively used in 2D face description and representation, and rapidly have been extended to the 3D modality. 3D-LBP approaches advanced the state of the art, and proved to be competitive with other classes of methods. However, their application is hindered by the intrinsic limitations of the 2D image support. Indeed, most if not all 3D-LBP approaches operate on depth images, in which depth is mapped to a gray level via 2D projection. As such, depth images require normalization to accommodate with pose variation. Yet, they still remain vulnerable to self-occlusions (caused for instance by the nostrils).

To address these problems, we propose a novel LBP-based face representation that can be constructed over triangular mesh manifolds. This representation, which is based on the recently proposed mesh-LBP concept [1], preserves the full 3D geometry of the shape, thus relieving the recognition process from the need for pose normalization (i.e., since mesh-LBP descriptors are computed on the 3D mesh triangulation, they are intrinsically independent from the mesh orientation in the 3D space). In another hand, given the consensus on the advantageous aspects of multi-modal face recognition [3], LBP construction on the mesh allows boosting recognition by offering an elegant framework for fusing, over a mesh support, texture and shape information at data and feature level, in addition to score and decision level, noticeably. To the best of our knowledge, this work is the first one to propose texture and shape fusion for face recognition using LBP constructed on the mesh. We also point out that our contribution in this work focuses mostly on the aspect related to the face description and, as a matter of fact, we are employing a very basic minimum distance classifier in the recognition pipeline. In the remaining of this Section, we first summarize the works that are most related to our solution (Sect. I-A), then we outline the proposed approach and its main contributions (Sect. I-B).

A. Related Work

Many 3D face recognition approaches have been proposed in the literature, and going through all of them is out of the scope of this summary. Instead, in the following, we will focus on existing methods that are relevant for the proposed solution, which can be categorized according to three different aspects: *a)* Methods that use local representations of the face, and thus are capable of supporting partial face matching, as can occur in the case of expression variations or missing parts (many recent methods achive this goal relying on *fiducial points* of the face); *b)* Approaches featuring face recognition by extending the LBP framework to depth images and 3D modalities; *c) Multi-modal* solutions that fuse together the 3D geometry and the 2D photometric appearance of the face to improve recognition. A more general and comprehensive review of 3D face recognition can be found in [3]–[6].

Manuscript received August 20, 2015; revised November 27, 2015; accepted December 21, 2015. Date of publication January 6, 2016; date of current version February 24, 2016. This work was supported by the National Research Foundation, United Arab Emirates, under Grant UIRCA 2013-24877. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. P. C. Yuen.

N. Werghi and C. Tortorici are with the Electrical and Computer Engineering Department, Khalifa University, Abu Dhabi 127788, United Arab Emirates (e-mail: naoufel.werghi@kustar.ac.ae; claudio.tortorici@kustar.ac.ae).

S. Berretti and A. Del Bimbo are with the Department of Information Engineering, University of Florence, Florence 50139, Italy (e-mail: stefano.berretti@unifi.it; alberto.delbimbo@unifi.it).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2016.2515505

1) Local Methods Based on Fiducial Points: At a very broad level, solutions for 3D face recognition can be grouped as *global*, performing face matching based on the whole face, and *local* that partition the face surface into regions and extract appropriate descriptors for each of them [7]. Methods in this latter category have recently gained an increasing credit, mainly thanks to their capability of natively supporting partial face match, as occurring in the case of scans with missing parts or occlusions (the case of facial expressions is often managed in a similar way, by excluding from the match the parts of the face that are most affected by expression variations). Among local approaches, effective results have been reported by methods that detect *fiducial points* of the face (being them either anthropometric landmarks, points of a predefined grid, or sparse keypoints), and compute local descriptors of surface patches centered at the fiducial points. One of the first approaches following this framework was proposed by Mian et al. [8], which designed a 3D keypoints detector and descriptor inspired by SIFT [9]. This detector/descriptor was used to perform 3D face recognition through a multi-modal 2D+3D approach that also used SIFT to index 2D images of the face. However, results reported for the method did not account for face scans with pose variations and missing parts. In [10] and [11], the framework of SIFT keypoints detector has been reformulated to operate on 3D face meshes by defining the mesh-SIFT detector and local descriptor. A scale-space analysis of the mesh is first performed through subsequent smoothing of the 3D geometry, then 3D keypoints are identified as the local extrema of the mean curvature extracted from the smoothed versions of the original mesh through the scales. Local descriptors are defined at the keypoints using nine local regions (arranged according to a daisy-like pattern), and computing for each of them a pair of histograms (the shape-index and the angle between surface normal descriptors are used). Effective local solutions based on fiducial points have been recently reported also in [12]–[14]. In [12], Lin et al. used mesh-SIFT to detect feature points on 3D face scans; Then, the quasi-daisy local shape descriptor [15] at each feature point was obtained using multiple order histograms of differential quantities extracted from the surface; Finally, these local descriptors were matched by computing their orientation angles. The same authors extended this work in [13], by boosting the keypoints matching with the Sparse Representation based Classifier (SRC) [16]. In [14], Berretti et al. used a similar paradigm by considering different varieties of histogram descriptors computed at mesh-DOG 3D keypoints [17]. The keypoints matching was also improved using the RANSAC algorithm.

2) LBP-Based Solutions: Since the seminal work of Ahonen et al. [18], [19], LBP-based solutions have shown their effectiveness in face recognition from 2D still images [20]. Inspired by these works, the idea of extending LBP to the 3D geometry of the face has been explored in several studies. Most, if not all, the LBP-based face recognition methods in the literature operate on depth images. This format allowed a straightforward application of the 2D-LBP operator as it was demonstrated in the pioneering work of Li et al. [21]. Later, Huang et al. [22], [23] proposed the multi-scale extended LBP (eLBP), which consists of several LBP codes in multiple layers accounting for the exact gray value differences between the central pixel and its neighbors. Sandbach et al. [24] introduced the local normal binary pattern (LNBP), which used the angle between normals at two points, rather than the depth value to obtain the local binary code. This novel LNBP concept has been adopted in subsequent works in different variants. Li et al. [25] extracted surface normals in 3D, then the values of the normal components along the direction of the three coordinate axes are interpreted as depth values, and LBP is computed on these depth maps reporting the values of the normal components. In a further extension, Sandbach et al. [26] constructed images of azimuthal projection distance. The azimuthal equidistant projection is able to project normals onto points in an Euclidean space according to the direction. Though the projected information is not the depth, depending on the normals of the 3D surface, 2D LBP are still computed on the projection images. The 3D-LBP method proposed in [27] computed the difference of the depth value or the angle between the normal of a central vertex and the eight neighboring vertices on a mesh. Using this descriptor, a region based representation of the face similar to the one developed in [19] for 2D face recognition is derived. This work includes the idea of using normals computed on the mesh, but the mesh requires an elaborated preprocessing in order to extract LBP constrained to the eight vertices near to a central one. Also, the circular ordering procedure of these vertices, necessary to perform LBP computation is not revealed. In addition, multi-resolution LBP is not supported, and the partitioning of the face into regions is defined based on a set of 48 landmarks manually annotated. More recently, Bayramoglu et al. [28] combined a central symmetric variant (CS-3DLBP) pattern, and a set of geometrical features in a decision-level fusion using a robust random forest classifier. This method operates on depth images and adopted also surface normal orientation as a shape function. All the aforementioned LBP-based methods, except [27], operate on depth images, and therefore when dealing with a mesh model as input have to convert it into a depth image via assiduous normalization procedures. This makes handling incomplete face scan resulting, for instance, from pose variation and occlusion, quite problematic for these methods. Finally, while the method of Tang et al. [27] constructs LBP patterns on the mesh, it requires intense mesh preprocessing and lacks the multi-resolution aspect of the original LBP.

3) Multi-Modal 2D-3D Solutions: Multi-modal methods try to combine multiple processing paths (typically in 2D and 3D) into a coherent architecture to solve critical aspects of individual methods. In [29], Chang et al. proposed applying PCA to face depth images and 2D face images separately, and then fusing the results together. In the work of Lu et al. [30], ICP registration of the 3D face models was combined with LDA applied to 2D face images to improve the robustness of 2D face matching in the presence of pose and illumination variations. Beumier et al. [31] extracted central and lateral profiles of the face and compared them in both 3D and 2D. In the approach of Hüsken et al. [32], landmark positions used to define the face regions were also detected on 2D texture

images obtained with the 3D face scan. Mian et al. [33] assembled a fully automated system performing: pose correction, automatic region segmentation to account for local variations of the face geometry, quick filtering of distant faces using SIFT and 3D Spherical Face Representation, and matching of the remaining faces applying a modified ICP to a few regions of the face (eyes, forehead, and nose) that are less sensitive to face expressions. The similarity scores provided by the two matching engines were fused into a single similarity measure. An in-depth study of fusion strategies for 3D face recognition was carried out by Gökberk et al. [34] that discussed and compared various techniques for classifier combination, such as fixed rules, voting- and rank-based fusion schemes, by fusing several off-the-shelf 3D and 2D features. Soltana et al. [35] through extensive experimentation show that individual 2D and 3D features are far from being distinctive for discriminating human faces. They propose an adaptive score level fusion strategy for multi-modal 2D-3D face recognition. The strategy consists of an offline and an online weight learning process, which automatically selects the most relevant weights of all the scores for each probe face in each modality.

B. Contribution and Paper Organization

From the above analysis, it emerges that solutions locally describing the face around fiducial points can perform 3D face recognition in difficult conditions, thanks to their intrinsic capability of managing partial match. On another side, there is evidence that LBP is an effective descriptor of the face capable of capturing local information. Last, multi-modal solutions that fuse together shape and photometric information can be used to boost further the recognition. In light of these considerations, we propose in this work a method capable of supporting recognition in the presence of missing parts, occlusions and expressions. Our method encompasses the following stages: 1) Computation of LBP descriptors using both shape and photometric information of the face mesh surface; 2) Construction of a grid of points on the face surface to obtain an ordered set of regions (equivalent to blocks in the 2D case); 3) Computing a histogram at each region, then concatenating the regional histograms into a structure encoding either a global or partial description of the face; 4) Performing the face matching by exploiting different fusion modalities. Our work presents the following innovative aspects:

- We introduce an LBP-based face representation constructed over triangular mesh manifolds;
- Our method relieves the recognition process from face pose normalization, while preserving the full geometry of the facial shape;
- Operating on the mesh, with our approach the photometric appearance is processed directly attached to the mesh and not on a separated planar image as in other multi-modal methods, thus allowing an early level-fusion of the texture and shape information;
- Our method uses a fixed set of fiducial points based on a sampling grid of the face. The points of the grid are obtained according to a predefined arrangement with respect to three reference facial landmarks. This avoids the need for elaborated processing required by keypoints detectors.

The results obtained on the BU-3DFE and Bosphorus datasets show the proposed method competes, and in some cases overcomes, the state of the art solutions.

The rest of the paper is organized as follows: In Sect. II, we give an overview on the mesh-LBP concept, focussing on the descriptor computation and its properties; In Sect. III, we present our face representation based on mesh-LBP; The fusion modalities used to combine geometric and photometric descriptors attached to the mesh are discussed in Sect. IV; Experimental evaluation in comparison to state of the art methods with results on two datasets is reported in Sect. V; Finally, we discuss the main positive aspects of our framework together with its current limitations in Sect. VI, where we also draw possible directions for future work.

II. LBP Descriptors on the Mesh

LBP construction on triangular mesh manifolds is a recent concept introduced by Werghi et al. [1], [36]. Before describing it, let us briefly remind about the standard LBP construction. In its simplest form, an LBP is an 8-bit binary code obtained by comparing a pixel's value (e.g., gray level, depth) with each pixel's value in its 3×3 neighbour. The outcome of this comparison is 1 if the difference between the central pixel's value and its neighbour pixel's counterpart is less or equal than a certain threshold, and 0 otherwise. The so obtained local description can be refined and extended at different scales by adopting circular neighbourhoods at different radii and using pixel sub-sampling.

Werghi et al. [36] elegantly extended the LBP concept to 2D-mesh manifolds by proposing a simple yet efficient technique for constructing sequences of facets ordered in a circular fashion around a central facet. The principle of the approach consists in categorizing the facets on the contour defined by a central facet's edges in two categories, namely, the $Fout$ facet and the $Fgap$ facets. An $Fout$ facet (respectively, an $Fgap$ facet) shares an edge (respectively, a single vertex) with a central facet (referred by f_c in Fig. 1).

Starting with three—clockwise or anticlockwise—ordered $Fout$ facets ($fout_1$, $fout_2$, and $fout_3$ in Fig. 1), the construction algorithm iteratively extracts the $Fgap$ facets located between each pair of consecutive $Fout$ facets following the same order in which the $Fout$ facets have been initially arranged, and closing the loop at the pair composed by the last $Fout$ facet (the third one) and the first one. The outcome of this procedure is a ring of ordered facets arranged clockwise or anticlockwise around the central facet. From this ring, a new sequence of ordered $Fout$ facets located on the ring's outer-contour can be extracted, thus allowing the ring construction procedure to be iterated, and to generate a sequence of concentric rings around the central facet (see the illustrations on the bottom of Fig. 1).

Algorithms 1 and 2 summarize the computation of ordered rings of facets.

The so obtained structure of ordered and concentric rings around a central facet forms an adequate support for computing LBP operators (referred as mesh-LBP in [36]) at different

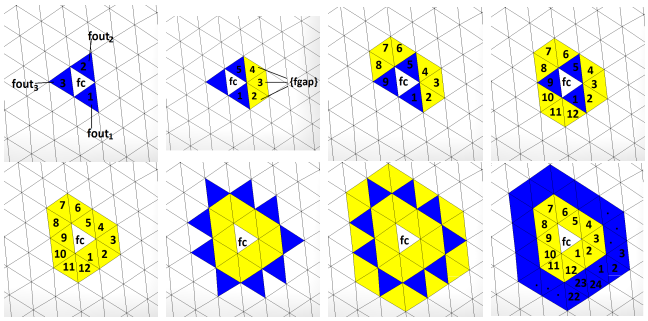


Fig. 1. Ordered ring construction: From the initial $Fout$ facets formed by the three ordered facets $fout_1$, $fout_2$, and $fout_3$ that are adjacent to the central facet f_c, a sequence of $Fgap$ facets located between each pair $\langle fout_1, fout_2\rangle$, $\langle fout_2, fout_3\rangle$, and $\langle fout_3, fout_1\rangle$ are extracted. The $Fgap$ facets have exactly one vertex on the initial 3-edge contour of the central facet f_c, and they are dubbed so because they look like filling the gap between the $Fout$ facets. This procedure produces a ring of facets ordered in a circular fashion around the central facet f_c. By iterating this procedure, using as new set of $Fout$ facets the sequence of facets that share an edge on the outer contour of the current ring, a sequence of rings of ordered facets can be generated.

Algorithm 1 Bridge

Require: $fout_i$, $fout_{i+1}$ two consecutive $Fout$ facets sharing a vertex; fin_i facet that shares an edge with $fout_i$

Ensure: $Fgap_i$ set of consecutive $fgap$ facets bridging the gap between $fout_i$ and $fout_{i+1}$

procedure BRIDGE($fout_i$, $fout_{i+1}$, fin_i)
  $Fgap_i = [\ ]$
  $v \leftarrow$ vertex shared by $\langle fout_i, fout_{i+1}\rangle$
  $gf \leftarrow$ facet adjacent to $fout_i$, different from fin_i and containing v
  $prev \leftarrow fout_i$
  while $gf \neq fout_{i+1}$ **do**
    append gf to $Fgap_i$
    $new\_gf \leftarrow$ facet adjacent to gf, different from $prev$ and containing v
    $prev \leftarrow gf$
    $gf \leftarrow new\_gf$
  end while
end procedure

radial and azimuthal resolutions, while preserving the simplicity of the original LBP. Let $h(f)$ be a scalar function defined on the mesh, incarnating either a geometric (e.g., curvature) or photometric (e.g., color or gray level) information. The mesh-LBP operator at the facet f_c is defined as follows [36]:

$$meshLBP_m^r(f_c) = \sum_{k=0}^{m-1} s(h(f_k^r) - h(f_c)) \cdot \alpha(k),$$
$$s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0, \end{cases} \quad (1)$$

where r is the ring number, and m is the number of facets uniformly spaced on the ring. The parameters r and m control, respectively, the radial resolution and the azimuthal quantization. In practice, in our implementation, we used a number of rings going from 1 to 7, with 12 facets per ring

Algorithm 2 GetRing

Require: $Fout$, set of n ordered facets, $fout_1$, $fout_2, \ldots$, $fout_n$, lying on a convex contour; Fin, set of n ordered facets, $fin_1, fin_2, \ldots, fin_n$, one-to-one adjacent to the $Fout$ facets and located inside the region delimited by the convex contour (depending on the contour, Fin might include duplicates)

Ensure: $Ring$, ring of ordered facets

procedure GETRING($Fout$, Fin)
  $Ring = [\ ]$
  for all $\langle fout_i, fout_{i\%n+1}\rangle$, $i \leftarrow 1, \ldots, n$ **do**
    append $fout_i$ to $Ring$
    $Fgap_i \leftarrow$ BRIDGE($fout_i$, $fout_{i\%n+1}$, fin_i)
    append $Fgap_i$ to $Ring$
  end for
end procedure

for computing mesh-LBP descriptors. This choice reflects the fact we have 12 facets in the first ring (regular mesh), and we keep this number of samples in any subsequent ring of the facet's support. The discrete function $\alpha(k)$ is introduced for the purpose of deriving different LBP variants. In this work, we will consider two variants of $\alpha(k)$: for $\alpha(k) = 2^k$, we obtain the mesh counterpart of the basic LBP operator firstly suggested by Ojala et al. [2]; for $\alpha(k) = 1$, we obtain the sum of the digits equal to 1 composing the binary pattern. In the experiments, we will refer to these two functions by α_2 and α_1, respectively. For the discrete surface function $h(f)$, in this work we experimented the *mean curvature* (H), the *curvedness* (C), the *Gaussian curvature* (K) and the *shape index* (SI), as shape descriptors, plus the *gray level* value (GL) as photometric characteristic of the facets.

With reference to the computation of mesh-LBP, it is relevant to note that the facets of the first ring can be ordered in three different ways, depending on which of the three $Fout$ facets adjacent to the central facet f_c is considered as the initial one. To solve this ambiguity, the closest facet to the center of mass of the f_c's neighbourhood is elected as the initial facet of the ring. Subsequent rings inherit the ordering of the facets from that established for the first ring. It can be also observed that, by construction, patterns computed with the α_1 function do not depend on the choice of the initial facet of the ring (i.e., the pattern value is determined just from the number of digits set to 1, rather than from their position as instead occurs for α_2). In the ideal case of a regular mesh, the number of facets v at ring i is computed according to the arithmetic progression $v_{i+1} = v_i + 12$ ($v_0 = 0$). In the real case, to cope with mesh tessellation irregularities as produced by 3D scanner acquisitions, the scalar function $h(f)$ is interpolated and sub-sampled across each ring, allowing thus to maintain a constant azimuthal quantization. The authors in [36] showed that this technique copes to a large extent with mesh irregularity.

A. Constructing and Comparing Mesh-LBP Descriptors

As for their 2D counterpart, the outputs of mesh-LBP operators of Eq. (1) computed across a mesh surface are not usually

directly used in shape matching, but rather accumulated into a discrete histogram constructed over a given neighborhood. The size of the histogram depends on the radial and azimuthal parameters r and m, as well as on the discrete function α. For example, with $r = 7$ and $m = 12$ we will obtain the histogram encompassing 7×13 and 7×4096 bins, for the α_1 and α_2, respectively. In fact, in the first case, 13 different values of the patterns are possible, being them coincident with the possible number of digits set to 1 in the binary code (i.e., the number of bit from 0 to 12 that to 1, which is also equal to the sum of the bit values); in the case of α_2, each digit in the pattern is weighted according to its position, so that 4096 different binary codes are possible (i.e., from 0 to 4095). The obtained histogram bins can be arranged in a 1-D or 2-D accumulator, and compared using χ^2 distance:

$$d(H_1, H_2) = \frac{1}{2} \cdot \sum_i \frac{(H_1(i) - H_2(i))^2}{H_1(i) + H_2(i)}, \tag{2}$$

where H_1 and H_2 are two normalized histogram descriptors. Good results have been obtained also using the *cosine* distance, especially for the α_1 variant.

Considering histograms obtained with the α_1 and α_2 functions, it is evident the different size of the respective descriptors. In particular, with an azimuthal quantization $m = 12$, 4096 mesh-LBP patterns are possible for α_2, compared to the 13 different patterns for α_1. This aspect has been investigated in [36], showing that the majority of the α_2 patterns have a number of 0-1 transitions below 4. These patterns have been called "uniform" following a similar property noticed first by Ojala et al. [37] for 2D-LBP (in that case, for patterns of eight bits, the uniformity was assumed for a number of 0-1 transitions not greater than 2).

In this work, we re-investigated the presence of uniform patterns on face scan samples from the Bosphorus database, using the *mean curvature*, *curvedness* and the *gray level* as scalar functions. Again, we found that the mesh-LBP with a number of 0-1 transitions less or equal than 4 form more than 95% of the total number of patterns across seven rings. The detailed statistics are reported in Fig. 2, whereby we can see the frequencies of the different 0-1 transitions in the mesh-LBP patterns and the percentage of the transitions below or equal to 4, across all the rings. In the bottom of Fig. 2, we also visualize the facets corresponding to non-uniform patterns. It is evident, there are a few non-uniform patterns, and they are located mostly in non-rigid parts of the face, which change with facial expressions. These results seem to suggest that considering uniform patterns is sufficient. Thus, considering four 0-1 transitions as the threshold for uniform patterns, it results in exactly 1124 uniform patterns against 2972 non-uniform ones. Following the same partition scheme of [37], where all the non-uniform patterns are grouped into a single label, whereas a separate label is assigned to each non-uniform pattern, the number of histogram bins (or classes) is reduced to 1125 for our mesh-LBP. We will adopt this partition in the rest of the paper for the α_2 function. For α_1, the distinction into uniform/non-uniform patterns does not make too much sense, since in this case the sum of the number of digits set to 1 in the binary code is computed, rather than

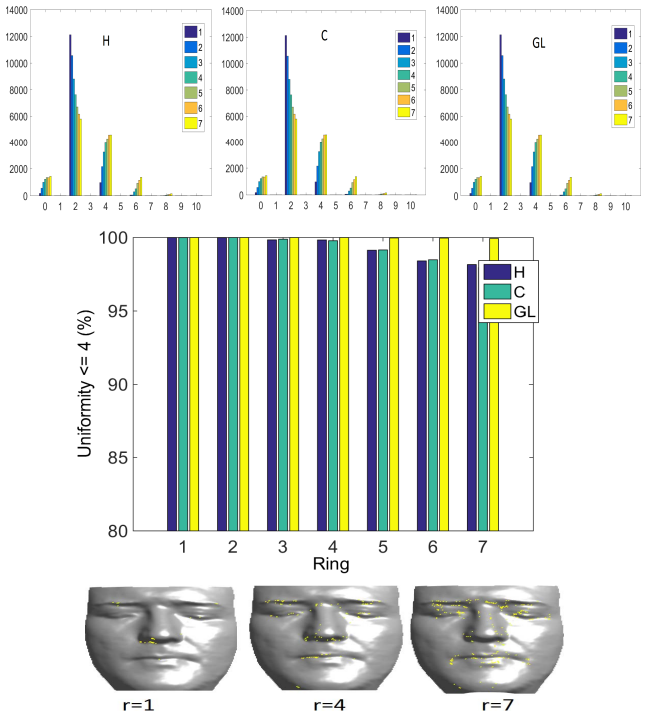


Fig. 2. Top: distribution of the number of 0-1 transitions in the mesh-LBP patterns using α_2 and the scalar functions *mean curvature* (H), *curvedness* (C) and *gray level* (GL). The mesh-LBP patterns have been computed for the seven radial resolutions $r = 1 : 7$ (i.e., seven rings), and for an azimuthal resolution $m = 12$ across all the rings. Note that number of odd transitions is always zero because what is counted actually is both the 0-1 and 1-0 transitions, and considering a circular arrangement of the binary digits. Middle: Percentage of the mesh-LBP patterns, in the same variants, showing a number of 0-1 transitions below or equal to four. Bottom: Facets on an example face scan having a non-uniform pattern obtained with *mean curvature*, for the radial resolutions $r = 1$, 4 and 7.

the binary value given by the polynomial expansion of the digits, as for α_2. This results in only 13 possible different patterns.

To have a visual insight on the capacity of mesh-LBP to capture and discriminate local shape information, we considered first five fundamental shapes, namely, *valley*, *ridge*, *pit*, *peak*, and *saddle* (see Fig. 3(a)), and computed their mesh-LBP histograms using the *mean curvature* as scalar surface function. Results are reported in Fig. 3(b)-(c) for the α_1 and α_2 (adopting the uniform/non-uniform pattern partition) variants, respectively. We can notice that the pairs *valley-ridge*, *pit-peak* show similar histograms, because of their symmetry relationships, while they are quite distinguishable from each other and from their *saddle* counterpart. For the facial shape, we report in Fig. 4 representative mesh-LBP variants computed with the geometric and photometric functions H and GL, at seven different radial resolutions ($r = 1, \ldots, 7$). We can easily observe, across these different variants, patterns reflecting facial features. Also we notice that, as the radial resolution increases, these patterns exhibit a fine to coarse evolution common to multi-resolution operators. In Fig. 5, we extend this analysis to the case of within and between class variation of the mesh-LBP descriptors, by reporting examples computed on two sets of four instances corresponding to a same and

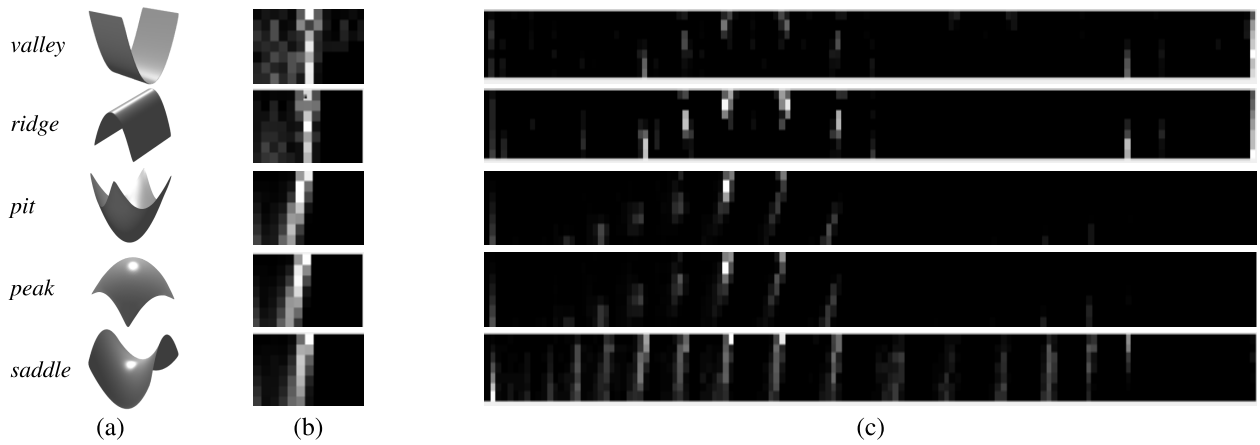


Fig. 3. Fundamental shapes (a) and their mesh-LBP histograms obtained using the mean curvature descriptor with α_1 (b), and α_2 (c).

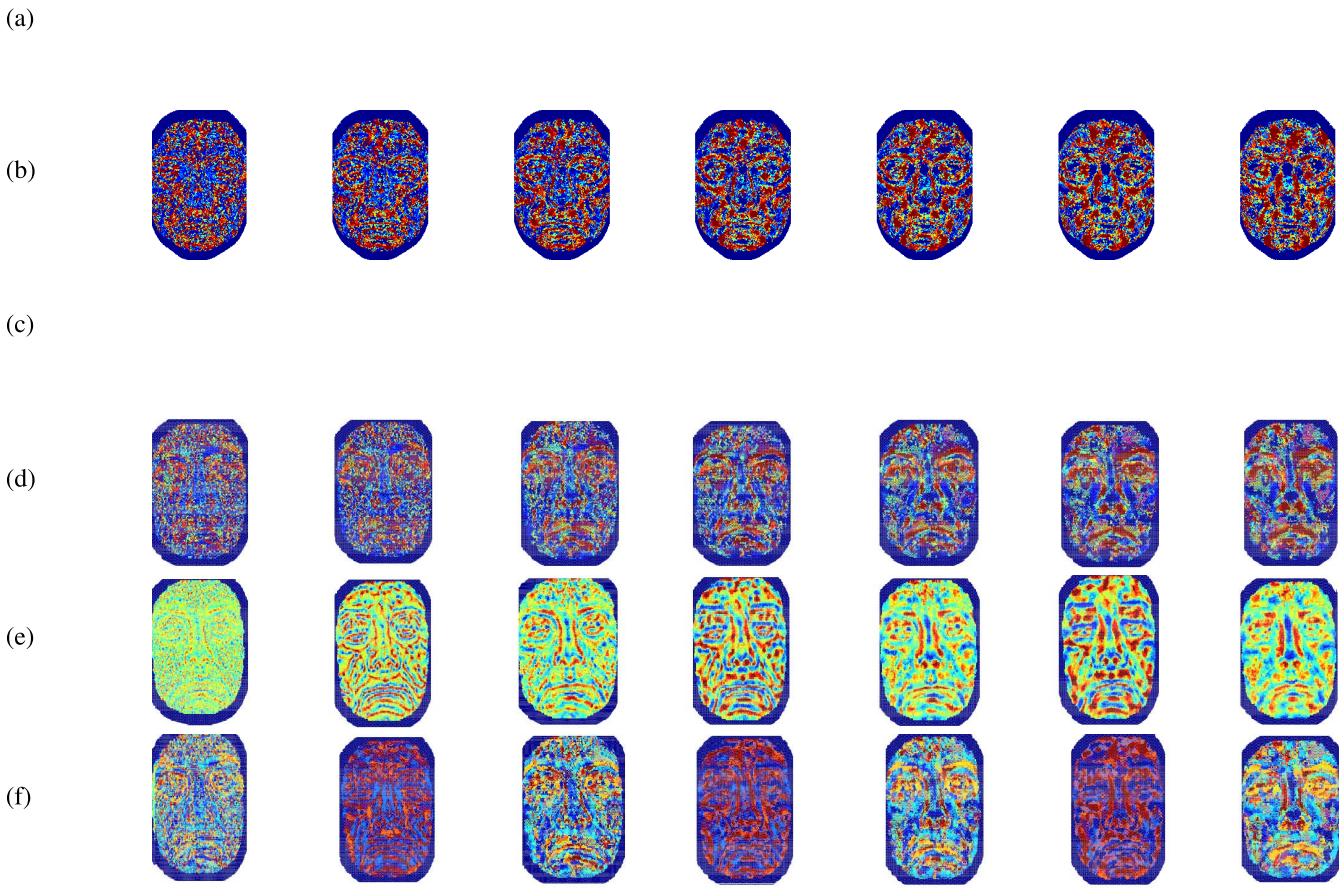


Fig. 4. Examples of mesh-LBP computed using the *mean curvature* (H), and the *gray level* (GL) in combination with α_1 and α_2: (a) $\langle H, \alpha_1 \rangle$; (b) $\langle H, \alpha_2 \rangle$; (c) $\langle GL, \alpha_1 \rangle$; (d) $\langle GL, \alpha_2 \rangle$; (e) $FF3\langle (H, GL), \alpha_1 \rangle$; (f) $FF3\langle (H, GL), \alpha_2 \rangle$. From left to right, the radial resolution r changes from 1 to 7 in each case.

different subjects. We can easily appreciate the stability of the patterns across the sibling instances as opposed to the neat variability observed across the non-related ones.

From Figs. 3, 4 and 5, both α_1 and α_2 categories exhibit great potential to be employed in facial surface description. While rotation invariance and low size properties of α_1 give it more favor than α_2, there are no prior indicators that can objectively indicate whether it can equate or outperform α_2 in terms of discriminating ability. The accentuated level of details and granularity exhibited by the examples of α_2 mesh-LBP descriptors displayed in Fig. 4 and 5 seem rather to indicate the opposite.

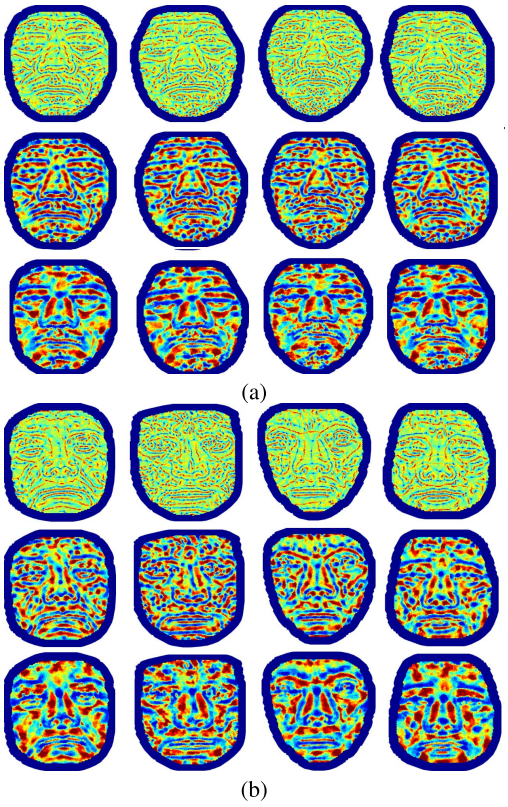


Fig. 5. Examples of mesh-LBP descriptors for four instances of the same subject (a), versus their counterparts related to four different subjects (b). The mesh-LBP used here is $\langle H, \alpha_1 \rangle$ computed at three radial resolutions $r = 1$, 3 and 7, from top to bottom of (a) and (b).

III. Face Representation

The previous analysis indicates that the mesh-LBP has some useful properties that make it attractive for capturing the shape and photometric information of a 3D surface. In order to exploit such potential for deriving a suitable face representation, we have taken inspiration from 2D face recognition methods that use standard LBP, and 3D methods based on fiducial points of the face that showed their appropriateness in supporting face recognition in the presence of facial expressions, occlusions and missing parts of the face. In particular, in the standard LBP-based face representation [19], a 2D face image is divided into a grid of rectangular blocks, then histograms of LBP descriptors are extracted from each block and concatenated afterwards to form a global description of the face. In so doing, image partitioning is performed easily due to the natural ordering of image pixels.

To extend this scheme to the face manifold, we need firs t to partition the facial surface into a grid of regions (the counterpart of the blocks in the 2D-LBP), compute their corresponding histograms, and then group them into a single structure. Since partitioning of the 2D mesh manifold is not straightforward, we rely on the idea of extracting a grid of fiducial points of the face with predefined position, and then use their neighborhood regions as local supports for computing mesh-LBP. In more details, this is performed with the following steps. First, the plane formed by the nose tip and the two eyes inner-corner landmark points is initially

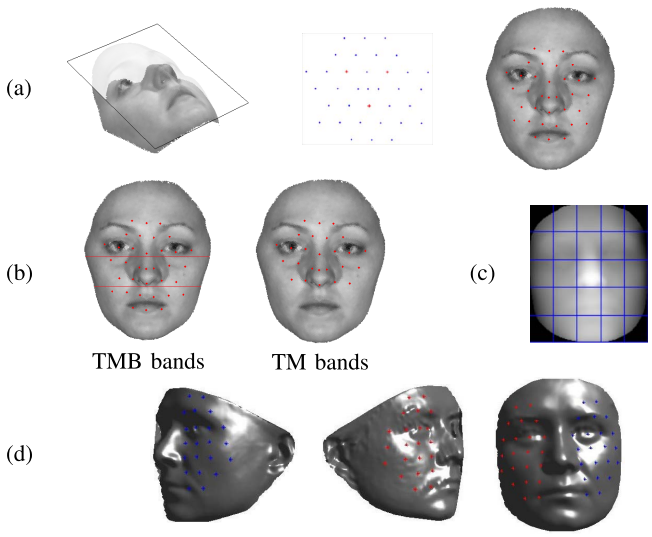


Fig. 6. (a) Construction of the face grid on the mesh; (b) On the left scan, all the grid points are shown and partitioned into three bands, namely, top (T), middle (M) and bottom (B), whereas on the right scan only the points in the top and middle bands (TM) are shown; (c) Grid partition of a depth image as used for the LBP method applied to depth images (3D-LBP); (d) Construction of the partial grid on a two rotated probe scans and a gallery scan.

computed (see Fig. 6(a), left). We used these three landmarks as they are the most accurate detectable landmarks on the face, and they are also quite robust to facial expressions. From these landmarks we derive, via simple geometric calculation, an ordered and regularly spaced set of points on that plane (see Fig. 6(a), middle). Afterwards, the plane is tilted slightly, by a constant amount, to make it more aligned with the face orientation, and then we project this set of points on the face surface, along the plane's normal direction. The outcome of this procedure is an ordered grid of points, which defines an atlas for the facial regions that will divide the facial surface (see Fig. 6(a), right). To account for the effects of facial expressions, we segmented the grid points into three bands, dubbed *top* (T), *middle* (M) and *bottom* (B). The TM option allows us to neutralize to some extent the shape changes manifesting at the lower part of the face, and caused by the mouth in particular. The TMB and the TM grids contain 35 and 26 points, respectively. The three TMB bands are shown on the left of Fig. 6(b), while the points comprised by the TM bands only are shown on the right.

For a yaw rotated pose resulting on a partial scan that does not allow the extraction of one of the two eyes inner-corner landmarks, we adopted a lateral grid, constructed upon the plane defined by one eye inner-corner, an eye outer-corner and the nose ridge. The grid covers one side of the face and contains 22 points. For the gallery scans, the TMB grid and both the left and right lateral grids are constructed (see Fig. 6(d)). Figure 6(c) instead, shows the partitioning of a depth image into a grid of 5×5 blocs, which is used to compare our method with the 3D-LBP counterpart operating on depth images, as detailed in Sect. V-B.

Once the grid of points has been defined, we extract a neighbourhood of facets around each point of the grid. Each neighbourhood can be defined by the set of facets confined

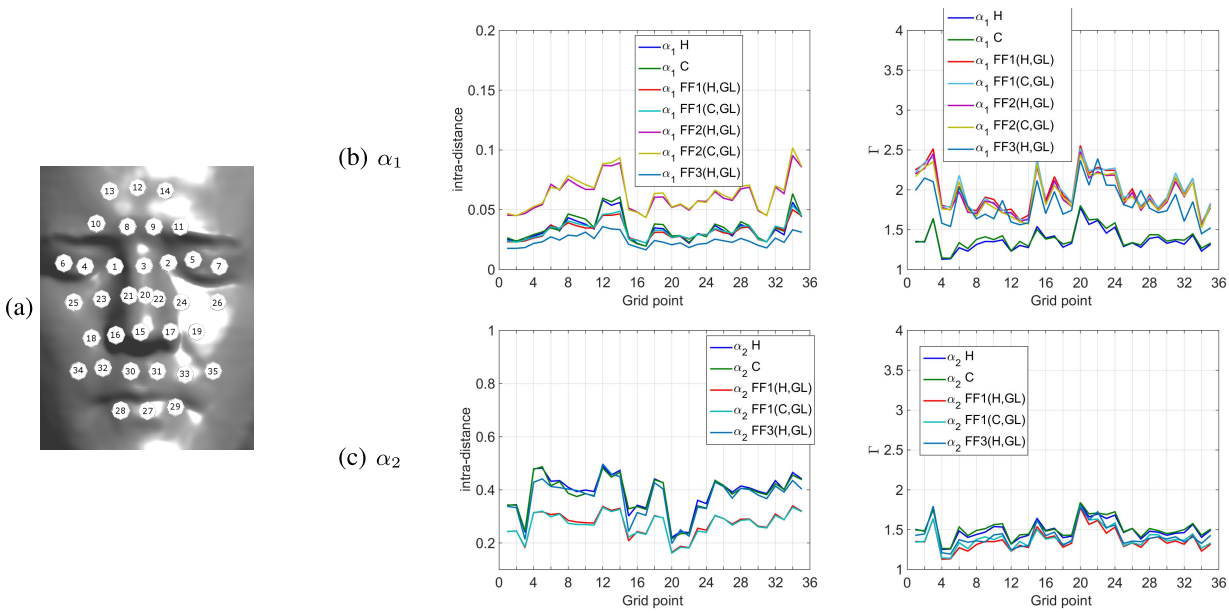


Fig. 7. (a) The numbered grid points; (b) and (c) report the *intra-class* distance and the criterion Γ computed at each grid point for the α_1 and α_2 mesh-LBP variants, respectively.

within a geodesic disc or a sphere, centered at a grid point. Then, we compute the multi-resolution mesh-LBP descriptor using Eq. (1) at each grid point region, considering both shape-valued (i.e., H, K, C, SI) and texture-valued (i.e., GL) functions. In the final step, the histograms of these descriptors are computed and integrated into a single histogram describing either the whole face or part of it (see Fig. 9(a)).

As primary assessment of the repeatability and the discrimination capacity of the different grid points in face matching we computed, for each grid point, the *inter-class* distance and the *intra-class* distance of the corresponding histogram. These two quantities have been obtained from, respectively, 35 pairs of scans, each corresponding to the same subject, and 35 scans corresponding to different subjects. Here, we adopted the *intra-class* distance and the ratio Γ=*inter-class distance/intra-class distance* as *repeatability* and *discrimination* indicators, respectively. Figure 7(b) and Fig. 7(c) depict the plot of these two indicators for each point of the grid (numbered according to Fig. 7(a)), and for the α_1 and α_2 mesh-LBP variants, respectively. Each plot compares a group of different descriptors including single and fusion variants (these will be described in Sect. IV). We can notice that the repeatability indicator shows virtually the same pattern across the different histogram descriptors. The best repeatability (i.e., lowest value) is observed at grid points around the nose and inner-eye corners (grid points {1, 2, 15, 16, 17, 22}). A similar behaviour is observed for the criterion Γ, whereby the grid points {1, 2, 3, 15, 20, 22} exhibit the most discriminative histograms (note that in this case the maximum of the curves correspond to the most discriminative points).

When we examined the distributions of the *intra-class* and the *inter-class* distances across the different grid points, we found that those in the α_2 variants exhibit more compact and separated distributions when compared to their α_1 counterparts. Figure 8 depicts some distribution examples illustrating this aspect. This suggested us that the α_2 variants have a higher discrimination, superior than α_1, as it will be confirmed in the experiments.

IV. Fusion Schemes

As a contribution of the proposed face representation, we propose the fusion of shape and photometric descriptors computed on the mesh. We further emphasize that the photometric channel is elaborated on the mesh as gray level attached to the triangles. No information is extracted from the 2D domain of gray (or depth) images of the face, but all the information is directly processed on the mesh manifold domain. Therefore, rather than being a multi-modal solution, the proposed approach can be regarded as a particular case of 3D methods, where the gray level plays an interchangeable role with standard shape surface descriptors.

In biometry applications, there are four levels of fusion considered, namely, *data*, *feature*, *score*, and *decision* [38]. As mentioned by Al-Osaimi et al. [5], it is believed that low-level fusion (data and feature) performs better than its higher level counterparts (score and decision) [39]. Looking at the spectrum of region methods fusing texture and 3D shape modalities, we found much concentration in the score-level category [21], [29], [33], [40], [41], as compared to the feature-level [8], [21], [42]. The work of Li et al. [21] in particular, fused LBP features derived from depth and texture image.

In our approach, we have investigated a score-level fusion and three variants of feature-level fusion. We have chosen the sum rule for the score-level, as it has been proven to be the optimal one [43]. In the first variant of the feature-level fusion, we concatenate the two mesh-LBP regional histograms, corresponding to the shape and the texture functions. For example, considering an azimuthal quantization $m = 12$ and α_1,

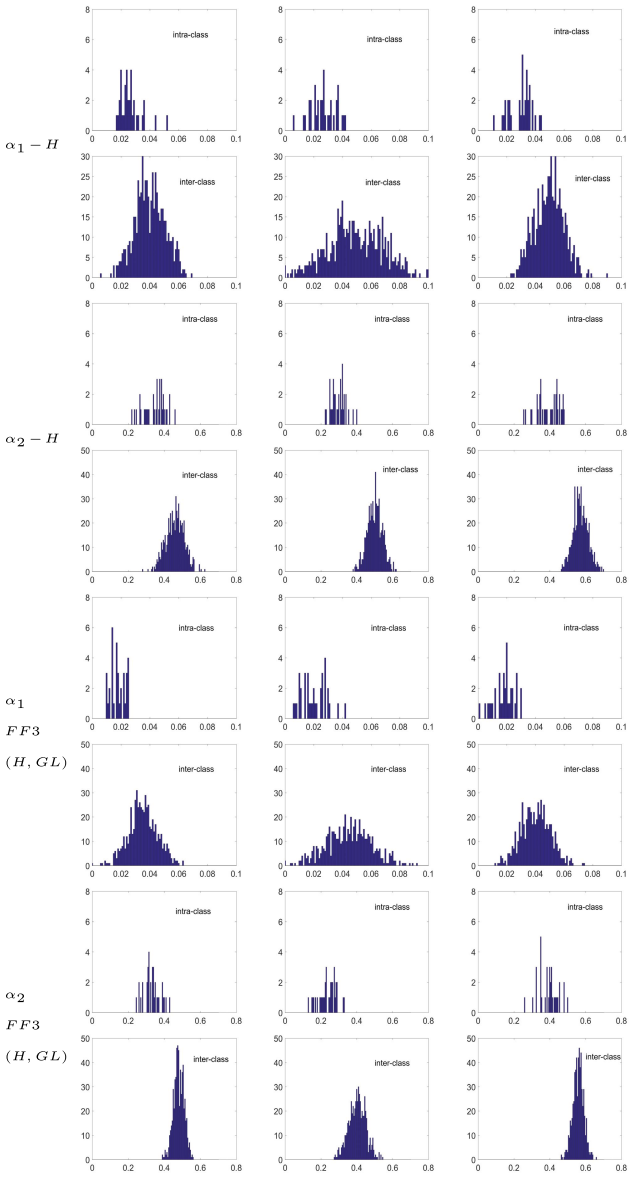


Fig. 8. Examples of *intra-class* and *inter-class* distribution computed at grid points at the nose tip (leftmost plots), right inner eye corner (middle plots), and mouth area (grid point 29 in Fig. 7, rightmost plots), for four different mesh-LBP variants. We notice that α_2 distributions exhibit more separation and compactness than their α_1 counterparts. The number of inter-class looks larger than its intra-class counterparts, as it encompasses all the pair combinations in the 35 subjects ($34 \times 35/2$).

we obtain a 13-bins histogram for each function, thus leading to a one-dimensional 26-bins histogram for each radial resolution r, that is a $r \times 26$ histogram. In the second feature-level fusion variant, we used a 2-D accumulator that counts for the co-occurrences of the mesh-LBP corresponding to the shape and the texture functions. For the same aforementioned parameters' values, we obtain an $r \times 13 \times 13$ histogram (Fig. 9(b) depicts some examples). In the third variant, the fusion is performed at the LBP pattern level, rather than the histogram level, as for the first two. Here, the mesh-LBP pattern is constructed by interleaving digits from the shape mesh-LBP with a texture mesh-LBP. So, for an azimuthal

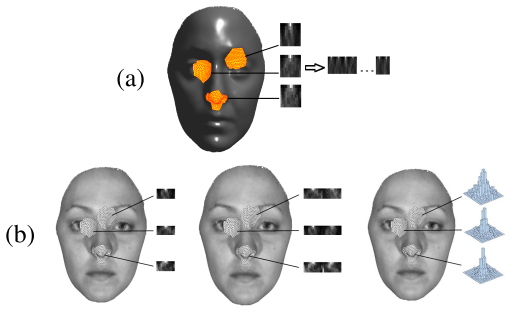


Fig. 9. (a) Global histogram construction: Region histograms are computed and then concatenated into a global histogram; (b) Examples of regional histogram variants obtained with $m = 12$ and $r = 7$ and α_1: (left) A 7×13 unimodal histogram corresponding to a shape function; (middle) A 7×26 histogram obtained by concatenating two 7×13 histograms corresponding to a shape function and a photometric function (gray level). This corresponds to the first variant of feature-level fusion ($FF1$); (right) A 2D section of a $7 \times 13 \times 13$ histogram obtained with a shape function and a photometric function. This is the second variant we used of feature-level fusion ($FF2$).

quantification $m = 12$, the mesh-LBP pattern sequence is $b_1^s\, b_2^t\, b_3^s\, b_4^t\, b_5^s\, b_6^t\, b_7^s\, b_8^t\, b_9^s\, b_{10}^t\, b_{11}^s\, b_{12}^t$. The last variant has the advantage to keep the related histogram to the same size than its mono-feature counterpart. In the rest of the paper, we will refer to these first, second and third feature-level fusion variants by $FF1$, $FF2$, and $FF3$, respectively, whereas the score-level fusion will be referred by SF.

V. Experiments

We conducted a series of experiments aiming at studying the behavior and performance of our fusion framework with respect to facial expressions, missing face data resulting from pose variation and occlusion, and the extent it improves the recognition over the classic fusion performed on the depth image. Our framework is assessed in comparison with the best methods in the literature, adopting similar experimental settings.

A. BU-3DFE Database

A first series of experiments was conducted with the BU-3DFE database from Binghamton University [44]. This database contains scans of 56 males and 44 females, acquired in a neutral plus six different expressions (*anger*, *disgust*, *fear*, *happiness*, *sadness*, and *surprise*). Apart of the neutral expression, all the other facial expressions have been acquired at four levels of intensity. This combination results in a total of 2500 scans. We considered as gallery and probe the sets of neutral scans and the expression scans, respectively. Scans in this database contain both texture and shape data. Figure 10 depicts samples of the 3D facial expression instances, and a 2D image used for texture mapping in that database. The image encompasses two face sides acquired from the two stereo pods composing the face scanner used in the data collection.

The purpose of using the BU-3DFE is to assess the performance of our method, in particular our fusion schemes, with respect to facial expressions. On this dataset, we set the radial resolution r and the azimuthal quantization m used in computing mesh-LBP equal to 7 and 12, respectively.

Fig. 10. BU-3DFE: (a) 3D face scans (with texture) of a sample subject showing, from left, the *neutral*, *anger*, *disgust*, *fear*, *happy*, *sad*, and *surprise* expressions (the level-1 of intensity is shown in each case); (b) The appearance image acquired by the scanner with two 45° side views of the face.

The choice of 12 for m is justified by the fact that given a generic central facet, the number of facets in its first ring is always equal to 12 for regular meshes, regardless of the resolution, as demonstrated in [36]. Choosing this value allows then to account for all facets in the first ring. This number is used for the subsequent rings, so as to have patterns taking values in the same range. The number of rings r is related to the resolution of the mesh. The rationale behind the choice of r is to cover an area around a point of the sampling grid wide enough to capture local surface information. With the mesh of the BU-3DFE we found that $r = 7$ covers about $7mm$ around the point making a good compromise between computation efficiency and effectiveness of the description.

To account for the effects of facial expressions, we considered the grid points partition into three bands, dubbed *top* (T), *middle* (M) and *bottom* (B), as introduced in Sect. III. Then, we tested our recognition approach considering the full grid (TMB) and the top and middle bands (TM) only (see Fig. 6(b)). The TM option allows us to neutralize to some extent the shape changes manifesting at the lower part of the face, and caused by the mouth in particular. The TMB and the TM grids contain 35 and 26 points, respectively. For the choice of the local descriptors we tested, in a preliminary experimentation, a variety of descriptors that include the *mean* (H) and the *Gaussian* (K) curvatures, the *curvedness* (C), and the *shape index* (SI), in combination with the α_1 and α_2 functions. We found that the H and C descriptors perform best than the rest, so we will report results related to these descriptors, mainly.

In the first experiment, we considered two grid configurations, namely, the full grid encompassing the top, middle and bottom band (TMB), and the partial grid including the top and the middle band only (TM). The goal is to assess to what extent excluding the bottom region of the face can neutralize the facial expressions for different descriptors and fusion modes. In order to emphasize this effect, we considered only the first level of expression intensity (referred to as *level-1*) of the BU-3DFE. Table I reports the *rank*-1 recognition rates obtained for different combinations of α_1, α_2, H, C, and the gray level GL as texture function, in both a unimodal and a fusion scheme. The table shows also the recognition rate for two types of histogram distances, namely, the cosine distance (cos), and the chi-squared distance (χ^2). First, we notice that the TM grid produces better results across most of the variants. This confirms the capacity of the TM grid matching of reducing the effects of facial shape variation caused by the mouth, while ensuring an overall acceptable recognition accuracy. Looking at the combination between the operator α and the histogram distance, we observe that α_1 and α_2 are best coupled with cos and χ^2, respectively. So, in the subsequent experiments, we used each variant with its best distance (i.e., cos with α_1 and χ^2 with α_2). Regarding the fusion aspect, we can notice the improvement induced by fusing shape and texture at each instance of the aforementioned combinations. In this context, we reported also results related to SI to show the ample improvement brought by the fusion, which is illustrated, for instance, in a jump in the accuracy from 82.43% to 95.65%, and from 78.96% to 95.13% in the $\langle TMB, cos\rangle$ and $\langle TM, \chi^2\rangle$ variants, respectively. We also observe that feature-fusion variants perform better than their score-level counterparts. The variant using $\langle FF1, TM, C, \chi^2\rangle$, in particular, scored the best performance of 97.74%.

TABLE I
BU-3DFE: RANK-1 RECOGNITION ACCURACY (IN PERCENTAGE) OBTAINED WITH DIFFERENT VARIANTS OF OUR METHOD FOR LEVEL-1 EXPRESSION INTENSITY

		TMB		**TM**	
		cos	χ^2	cos	χ^2
α_1	H	90.61	88.00	92.52	89.57
	C	89.57	87.13	90.43	89.04
	SI	82.43	80.00	82.96	80.00
	GL	92.35	91.48	93.22	93.22
	$FF1$ H	95.65	95.13	**96.70**	95.65
	$FF1$ C	**96.35**	94.96	96.35	96.35
	$FF1$ SI	94.78	93.91	96.17	94.96
	$FF2$ H	96.17	95.13	**96.70**	95.65
	$FF2$ C	95.30	95.30	96.00	96.00
	$FF2$ SI	95.65	93.74	96.17	94.96
	SF H	96.00	95.13	96.00	95.13
	SF C	**96.35**	94.96	96.35	94.96
	SF SI	95.48	93.91	96.17	94.96
α_2	H	82.96	92.70	87.83	95.48
	C	85.22	93.57	89.39	95.13
	SI	70.61	82.96	66.26	78.96
	GL	75.93	90.09	78.61	92.35
	$FF1$ H	81.22	96.17	85.22	97.39
	$FF1$ C	80.35	**96.35**	84.35	**97.74**
	$FF1$ SI	78.96	94.96	83.13	95.13
	SF H	89.22	96.17	89.22	96.17
	SF C	88.87	**96.35**	88.87	96.35
	SF SI	86.61	94.96	86.43	95.13

Referring to the computational cost and pattern repeatability, the α_1 variant is more appealing than α_2. This also motivated us to not include the $FF2$ fusion modality for α_2, since this would result in a high dimensionality of the fused descriptor with a consequently high computational cost. Nevertheless, α_2 takes advantage, theoretically, in its discriminative power given the wider range of its related patterns. While the results confirm the superiority of the α_2 variant overall, we notice that at some instances, α_1 performs better than α_2. While we do not have a definitive postulate explaining this consistency, we believe that the most plausible one is the intrinsic repeatability of the α_1 variant.

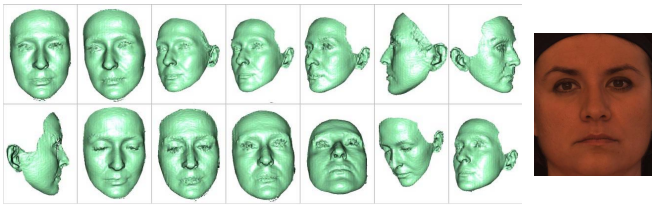
In Table II, the probe scans are categorized into the six different facial expressions, and recognition rates are reported for each category separately. We also included results obtained

TABLE II
BU-3DFE: RANK-1 RECOGNITION RATE (IN PERCENTAGE) OBTAINED FOR THE DIFFERENT EXPRESSION SUBSETS COMPARED TO [14]

		Level 1 & Level 2 Expressions							Level 3 & Level 4 Expressions						
Descriptors		**AN**	**DI**	**FE**	**HA**	**SA**	**SU**	**All**	**AN**	**DI**	**FE**	**HA**	**SA**	**SU**	**All**
[14]	HOG	90.00	87.50	88.80	88.10	90.60	85.00	88.30	81.30	75.60	78.80	80.60	82.50	76.90	79.30
	SHOT	93.80	90.60	91.90	90.00	94.40	88.80	91.60	87.50	78.80	85.60	79.40	90.00	79.40	83.40
	GH	90.60	85.00	84.40	85.60	90.60	82.50	86.50	86.30	79.40	80.00	79.40	85.00	78.80	81.50
α_1	H	89.00	74.50	83.50	89.00	96.50	93.00	87.58	73.50	48.00	68.00	79.00	85.00	84.00	72.92
	C	88.50	68.00	79.50	86.00	93.00	91.50	84.42	69.00	43.50	65.50	73.50	82.50	83.50	69.58
	GL	83.00	72.50	83.00	83.50	86.00	81.50	81.58	67.00	49.50	71.00	69.00	76.50	71.50	67.42
	FF1 H	95.50	86.50	92.50	94.50	97.00	97.50	93.92	82.50	67.50	86.00	86.50	94.00	94.00	85.08
	FF1 C	94.50	83.00	91.00	94.50	96.50	97.50	92.83	82.00	59.50	86.00	87.00	92.00	93.50	83.33
	FF2 H	94.00	85.50	91.50	95.00	97.50	96.50	93.33	83.00	66.50	85.50	87.50	93.00	93.50	84.83
	FF2 C	94.00	83.00	90.50	92.50	97.50	97.50	92.50	82.50	62.00	85.00	86.50	93.50	93.00	83.75
	SF H	95.00	86.50	92.50	95.00	97.00	98.00	94.00	83.00	67.50	86.50	87.00	94.00	93.50	85.25
	SF C	94.50	85.50	92.50	94.50	97.00	97.50	93.58	82.50	61.00	86.50	88.00	92.50	93.50	84.00
α_2	H	96.50	90.00	95.50	**98.00**	**99.00**	**99.50**	96.42	92.50	72.50	90.50	92.50	98.00	**99.50**	90.92
	C	97.00	89.00	95.50	**98.00**	**99.00**	**99.50**	96.33	92.00	69.00	89.50	92.50	97.50	**99.50**	90.00
	GL	88.00	82.00	87.50	89.50	91.00	87.50	87.58	72.50	58.00	80.00	77.50	81.00	85.50	75.75
	FF1 H	**98.00**	**93.50**	**96.50**	**98.00**	98.50	**99.50**	**97.33**	**94.50**	**80.00**	**92.50**	**95.50**	**98.50**	**99.50**	**93.42**
	FF1 C	97.50	92.50	96.00	**98.00**	**99.00**	**99.50**	97.08	94.00	73.50	91.00	94.50	97.50	**99.50**	91.67
	SF H	**98.00**	**93.50**	**96.50**	**98.00**	98.50	**99.50**	**97.33**	**94.50**	**80.00**	**92.50**	**95.50**	**98.50**	**99.50**	**93.42**
	SF C	97.50	92.50	96.00	**98.00**	**99.00**	**99.50**	97.08	94.00	73.50	91.00	94.50	97.50	**99.50**	91.67

with three variants of the interest-points method proposed in [14] and which have been applied on the same database. Methods in [42] and [45] also used the BU-3DFE database for 3D face recognition, but they are not directly comparable with our due to the different settings. The work in [45] limited the analysis to consistently labeled scans with expression intensities 3 and 4, that do not show large variations in illumination and geometry (total of just 212 scans of 81 subjects out of 2500 scans of 100 subjects). The approach in [42] is based on training multiple SVMs, thus dividing the dataset into two halves of 1200 scans each, one used for training and the other for test. Depending on the fact the intensities 1-2 or 3-4 are used for training, the rank-1 recognition rate is 97.7% and 98.7%, respectively.

From Table II, we first notice the α_2 variant of mesh-LBP outperforms in all the cases the α_1 variant. Compared to the results of Table I, where at *level*-1 expression α_1 and α_2 score similar results. This seems to indicate a major robustness of this latter variant to large and exaggerated expressions. Secondly, we observe that our method outperforms [14] even with variants using single modality (see scores related to H, C and GL with α_2). We notice, in particular, the almost full recognition rate obtained for the *surprise* category. The *disgust* category, which is the most radical expression, exhibits the lowest rate (93.50% for lower level distortions). The distribution of the best scores, highlighted in bold, clearly indicates the recognition enhancement brought by the fusion schemes. Also, we can observe that most of the best scores have been obtained with the feature-level fusion variants, though the score level fusion $\langle \alpha_2, SF, H \rangle$ achieves similar results. This observation is confirmed in the over-all results, whereby the configurations using $\langle \alpha_2, FF1, H \rangle$ and $\langle \alpha_2, SF, H \rangle$ score the best performance.



(a) (b)

Fig. 11. Bosphorus: (a) Samples from the different categories of Bosphorus scans; (b) A sample of the 2D image obtained with the single view scanner used for this database.

B. Bosphorus Database

The Bosphorus database [46], contains 4666 scans of 105 subjects scanned in different poses, action units, and occlusion conditions. Figure 11 shows some scan instances of this database. Notice here that scans are obtained with a single-view scanner, that is one stereo-pod. In particular, the dataset is divided in multiple subsets corresponding to neutral and expressive scans (the six fundamental expressions are considered, namely, *anger*, *disgust*, *fear*, *happy*, *sad*, *surprise*), scans with Action Units (*Lower Face Action Unit* (LFAU), *Upper Face Action Unit* (UFAU), and *Combined Action Unit* (CAU)), scans with rotations (*Yaw Rotation* (YR), *Pitch Rotation* (PR), and *Cross Rotation* (CR)), and scans with *occlusions* (O). Most of the face instances are provided with a set of landmarks that also includes the inner corner landmarks and the nose tip. These three landmarks are those used to define the plane on which the sampling grid of the face is defined (see Sect. III). For the scans with rotation, the inner corner of one of the eyes can be missing. In that case, the partial grid of points is constructed (as illustrated in Fig. 6(d)).

TABLE III

BOSPHORUS: RANK-1 RECOGNITION ACCURACY OBTAINED WITH A SELECTION OF OUR METHOD VARIANTS COMPARED TO [12] AND [14], AND THE BEST TWO VARIANTS REPORTED [13]. THE MAXIMUM OBTAINED RECOGNITION RATE IN EACH SUBSET IS HIGHLIGHTED IN BOLD

	Depth		[12]	[14]	HOMQ [13]		mesh-LBP α_1		mesh-LBP α_2		
	Hist	Score			CGM	FGM	$FF2$ H+GL	SF H+GL	SF H+GL	SF C+GL	$FF3$ H+GL
Neutral	93.30	93.30	**100**	97.9	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Anger	73.24	73.24	88.73	85.9	88.73	**97.18**	**97.18**	**97.18**	94.37	92.96	91.55
Disgust	73.91	73.91	76.81	81.2	76.81	86.96	85.51	89.86	**92.75**	**92.75**	88.41
Fear	62.86	62.86	92.86	90.0	92.86	**98.57**	**98.57**	**98.57**	**98.57**	**98.57**	**98.57**
Happy	87.74	87.74	95.28	92.5	95.28	**98.11**	88.68	91.51	97.17	97.17	97.17
Sad	89.39	89.39	95.45	93.9	94.45	**100**	96.97	98.48	98.48	98.48	98.48
Surprise	56.34	56.34	98.59	91.5	98.59	98.59	97.18	98.59	**100**	**100**	**100**
LFAU	88.19	88.19	97.22	96.5	97.22	98.84	97.09	97.55	99.10	99.03	**99.16**
UFAU	91.67	91.67	99.07	98.4	99.07	**100**	99.77	99.77	**100**	**100**	**100**
CAU	84.02	84.02	98.82	95.6	98.82	**100**	**100**	**100**	**100**	**100**	**100**
Yaw	6.39	6.39	77.96	81.6	77.96	**84.08**	72.0	74.47	56.57	56.19	59.24
Pitch	48.21	48.21	98.81	98.3	98.81	**99.52**	97.85	98.09	94.51	93.32	92.84
Cross	3.32	3.32	94.31	93.4	94.31	**99.05**	92.68	90.24	80.48	80.49	85.37
Occlusion	77.43	77.43	99.21	93.2	99.21	99.21	96.29	95.68	98.76	98.77	**99.38**

Experiments on this dataset aim to test the proposed approach on a larger dataset and in the presence of action units, missing parts and occlusions, in addition to expressions. On this dataset, we can also compare our approach with respect to state of the art methods. In particular, we compared with Li et al. [12], Berretti et al. [14], and Li et al. [13], which share the idea of using keypoints matching, and use the same experimental protocol. Actually, differently from our solution, in these methods keypoints are regarded as points on the mesh-manifold, which are stable over multi-scale differentiation, and which are usually detected using the mesh-DOG operator [17]. Local descriptors constructed at these keypoints are compared in order to find the best matches. In [12], multiple order histograms of differential quantities constructed at each face keypoint and its immediate neighbourhood points are used. In [14], a similar paradigm is used by considering different variety of histogram descriptors. The keypoints matching is also improved using the RANSAC algorithm. In their second version, Li et al. [13], boosted the keypoints matching with the Sparse Representation based Classifier (SRC) [16]. The approaches of Sandbach [26] and Bayramoglu [28] used also the Bosphorus database, but their purpose and setting are different from ours. First, these works assess expression recognition; and second, they employed, respectively, AdaBoost and Random Forest classifiers, and a 10-fold cross-validation scheme, whereas our method used a simple minimum-distance classifier. Besides, they do not consider pose scans in their experiments because of the limitation of the depth images with regard to this category. Therefore, to assess our fusion paradigm on the mesh over its counterpart on the depth images, we compared our method with the 3D-LBP operating on depth images, considering the same aforementioned fusing schemes, namely, score fusion (SF) of the depth and gray-level data, the first and third feature fusion ($FF1$ and $FF3$) of Sect. IV. For the setting of the 3D-LBP face description, the LBP patterns have been computed on 5 rings (radii from 1 to 5) and with an azimuthal resolution of 8. The global histogram is constructed over a grid of 5×5 blocs in the depth image, as shown in Fig. 6(c).

Table III depicts the comparison results. First we notice that, despite the fusion scheme, the 3D-LBP on the depth image scores quite below the other methods, for both *histogram* and *score* fusion variants. We can notice that our method neatly outperforms [12], [14], while it competes well with [13], equating and outperforming it at several subsets, noticeably at the *Disgust* and *Surprise* for expressions, *LFAU* for action units, and at the *Occlusion* subset.

For the *Pitch*, and *Occlusion* subsets our scores are reasonably close to [13], whereas the *Cross* subset score is a bit distant. The most critical case for our solution is represented by the *Yaw* subset, where we obtain an accuracy of about 75% for the $\langle SF, H + GL \rangle$ variant of α_1. In order to investigate more this most critical case, we broken-down the *Yaw* rotation subset results, and we found that our method scores well up to 20 degrees rotation as reported in Table IV. If we exclude the 45-degrees results, we obtain an overall score of 86.66%. Interestingly, the α_1 variant resulted more robust than the α_2 for rotation angles of 30 and 45 degrees. Examining the 45-degree scans, we found that the recognition failures in this category are probably due to surface corruption noticed at many instances (Fig. 12 shows some samples). While they do affect the global facial shape, such surface corruptions alter mesh-LBP patterns, which are by principle sensitive to surface artifacts, and consequently will be reflected on the grid histograms.

For the intra-comparison side, referring to the different fusion schemes and the α_1 and α_2 variants of our approach, some considerations can be drawn. As emerged also in the experiments on the BU-3DFE, fusion techniques combined with the α_2 variant seem more robust to expressions than the corresponding α_1 variants, though with a lower gap than in Table II. This is motivated also by the lower intensity of expressions in the Bosphorus dataset. The α_2 variants also show very high accuracy, equal or very close to 100%, on the action unit subsets. The α_1 variants, instead, are neatly more competitive than their α_2 counterparts in the case of rotated scans (as also emerged for the larger rotation angles in Table IV), with the much marked progress observed for the

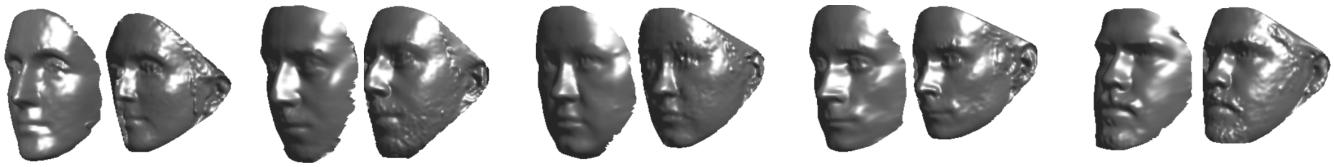


Fig. 12. Samples of the 45-yaw rotated scans failure cases. For each pair, the gallery scan is reported on the left, in the same pose of the probe scan shown on the right.

TABLE IV

BOSPHORUS: RANK-1 RECOGNITION ACCURACY OBTAINED FOR DIFFERENT *Yaw* ROTATION SUBSETS

Descriptors		10°	20°	30°	45°
α_1	FF1 H+GL	98.10	89.52	73.33	46.67
	FF2 H+GL	98.10	87.62	73.33	45.71
	SF H+GL	98.10	89.52	76.19	46.67
α_2	FF1 H+GL	99.05	93.33	64.76	12.38
	SF H+GL	99.05	93.33	64.76	13.33

Yaw subset. This can be mainly due to the intrinsic rotation invariance, and thus repeatability of the patterns obtained with α_1, which is expected to be much relevant in this case. Last, for the occlusion subset, comparable performance is obtained with a slight prevalence of the α_2 variants.

Table V reports an algorithmic complexity comparison between our method and the best variant of [13] (HQMQ FGM). We can notice that up to the mesh-LBP computation (for our method) and the keypoints detection (for [13]) both methods have a same linear complexity. The keypoint description and the grid construction have both constant complexity. The last two stages, however, show some difference. For the keypoint matching in [13], assuming all galley subjects have a same number of keypoints K, and considering the descriptor size as constant, the algorithmic complexity can be approximated by $O(KIG)$, where I is the number of iterations in Orthogonal Matching Pursuit algorithm (OMP) [47], involving a non-linear minimization used in the keypoint matching, and G is the number of subjects in the gallery. Considering the typical values of $K = 350$ the algorithmic complexity can be estimated as to $O(350IG)$. Using a simple minimum distance classifier, the algorithmic complexity of our method at the mesh-LBP histogram matching is $O(G)$. This indicates that the iterative nature of the OMP algorithm, and the individual keypoint matching in the last stage of (HQMQ FGM) variant in [13] is quite computationally more demanding than its counterpart in our method. For what concerns the size of the face signature, in Li's method it is $m \times K = 261 \times 350 = 75600$, being m the keypoint descriptor size for the HQMQ variant. In ours, it is $35 \times 13 \times 7 = 3185$, and $35 \times 1125 \times 7 = 275625$ for the α_1 and α_2 variants, respectively. These figures, give advantage to Li's method when compared to our α_2 variant.

TABLE V

COMPARISON OF THE ALGORITHM COMPLEXITY OF THE METHOD HQMQ FGM IN [13] AND OUR METHOD

HQMQ FGM [13]		
Stage	**Complexity**	**Details**
Smoothing	O(N)	N: number of vertices
Difference of Gaussians	O(N)	–
Keypoints detection	O(aN)	a: number of scales (3)
Keypoints description	O(bK)	K: number of keypoints (350 average) b: number of neighborhood points (400)
Keypoints matching (Orthogonal Matching Pursuit algorithm)	O(KIG) [51]	I: number of iterations G: number of subjects in the gallery
Our method		
Stage	**Complexity**	**Details**
Geometric descriptors	O(N)	N: number of vertices
mesh-LBP computation	O(cF)	F: number of mesh facets c: number of sampled facets in the mesh-LBP rings (84)
Grid construction	O(de)	d: number of grid points (35) e: number of facets per grid region (336)
mesh-LBP histogram matching minimum distance classifier	O(G)	G: number of subjects in the gallery

VI. DISCUSSION AND CONCLUSIONS

In this paper, we presented an original approach for constructing a multi-modal LBP-based face representation on a triangular mesh-model. It is the first approach of its kind that integrates texture and shape information in LBP-patterns derived from a mesh support. This marriage between mesh-model and LBP-based face recognition will open-up new horizons that go quite beyond the limits imposed by the depth image constraints. We proposed a face representation that encompasses a face-centric grid to which is attached, at each point of it, LBP histograms constructed using geometric and photometric data. Contrary to its depth-image counterpart, this representation supports partial facial matching, and does not require normalization. In addition, it preserves the full geometry of the facial shape, which might be partially lost in depth images because of self-occlusion. In addition, we have showed that our framework can be easily adapted to different fusion schemes, in particular the early stage fusion.

Despite having used a basic minimum distance classifier, we showcased the performance enhancement brought by our novel 3D face representation, and demonstrated that it can compete to a reasonable extent with the best methods of the state of

the art. Indeed, The experiments conducted with BU-3DFE database showcased the boosting of the recognition performance brought by our fusion framework, and its superiority with regard to the most closest approach. Results obtained on the Bosphorus database report competitive accuracy compared to the state of the art solutions, with an increment for some specific subsets.

Regarding the different variants of our method, including different shape descriptors in the mesh-LBP computation, the α_1 and α_2 weighting functions, and the varying fusion schemes, some summary comments can be drawn. Among the different surface descriptors we tested, the mean curvature (H) resulted the most suited to be combined with mesh-LBP across almost all the experiments. The mean curvature also resulted the optimal option for fusing with the gray level appearance of the surface's facets, using either low-level fusion at the feature level, or late fusion at the score level. The comparison between the α_1 and α_2 variants of mesh-LBP does not come to a univocal conclusion: the α_1 variant is intrinsically invariant to rotation and more efficient from a computational point of view; the α_2 variant, instead, takes advantage from the large gamut of possible values, which makes it more discriminative in most of the cases.

Looking at the performance of our method in the presence of facial expressions, one valid question might raise on how the methods achieve elevated scores for facial expression cases, where the facial surface might undergo significant changes compared to the neutral expression. We believe that this robustness lies first on the choice of TM grid, which discards the lower part of the face that is affected the most by deformation. Also, we think that the small size we choose for the grid regions ($r = 7$) made the representation fine enough to preserve local variability up to large extent. Discarding non-uniform-patterns for α_2 contributes further to the robustness to expressions, since these patterns are mostly located in non-rigid parts of the face.

For what concerns the matching procedure, our method has been employed in a global way, that is considering all the grid points in the matching, without assessing the plausibility of individual pairs of corresponding grid-points. Such procedure is a fundamental part of the methods in [12]–[14], where the plausibility of a pair of potential matching keypoints is evaluated by comparing their related local descriptors. In fact, the boosting of the performance in Li et al. method [13] as compared to their first work in [12] is due to the Sparse Representation based Classifier employed in keypoints matching. However this is without compromising the computation cost, as we have demonstrated it in the algorithmic complexity comparison.

As future work, there are several aspects worth to explore. First, the feature fusion methods we employed used two descriptors, while the numerous descriptors we can derive from the mesh, in addition to the texture, are appealing for investigating a multiple-descriptor fusion. However, we think that this needs to go beyond the standard concatenation and co-occurrence schemes, and that a sound theoretical framework would be necessary for accommodating such a fusion. Second, we believe that integrating a robust mechanism, rather than the simple minimum distance classifier, at the level of grid point matching would considerably boost our method's performance. Third, investigating keypoints framework with mesh-LBP as local descriptors would be novel blending worthwhile to investigate. Finally, optimizing further the size of our face signature, while keeping its discrimination power, noticeably for the α_2 variant.

Overall, we think that our contribution will pave the way for applying the other techniques and methods developed within the LBP-based face recognition directory.

ACKNOWLEDGMENT

Preliminary ideas and experiments of the face mesh-LBP based face recognition approach appeared in [49]. The mesh-LBP code is publicly available at Matlab Central: http://www.mathworks.com/matlabcentral/fileexchange/48875-mesh-lbp.

REFERENCES

[1] N. Werghi, S. Berretti, A. del Bimbo, and P. Pala, "The mesh-LBP: Computing local binary patterns on discrete manifolds," in *Proc. ICCV Int. Work. 3D Represent. Recognit.*, Sydney, NSW, Australia, Dec. 2013, pp. 562–569.

[2] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, Jan. 1996.

[3] K. W. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition," *Comput. Vis. Image Understand.*, vol. 101, no. 1, pp. 1–15, Nov. 2006.

[4] S. Berretti, A. del Bimbo, and P. Pala, "3D face recognition using isogeodesic stripes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 12, pp. 2162–2177, Dec. 2010.

[5] F. R. Al-Osaimi, M. Bennamoun, and A. Mian, "Spatially optimized data-level fusion of texture and shape for face recognition," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 859–872, Feb. 2012.

[6] H. Drira, B. Ben Amor, M. Daoudi, A. Srivastava, 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.

[7] 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, Dec. 2003.

[8] A. S. Mian, M. Bennamoun, and R. Owens, "Keypoint detection and local feature matching for textured 3D face recognition," *Int. J. Comput. Vis.*, vol. 79, no. 1, pp. 1–12, Aug. 2008.

[9] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[10] C. Maes, T. Fabry, J. Keustermans, D. Smeets, P. Suetens, and D. Vandermeulen, "Feature detection on 3D face surfaces for pose normalisation and recognition," in *Proc. IEEE Int. Conf. Biometrics, Theory, Appl. Syst. (BTAS)*, Washington, DC, USA, Sep. 2010, pp. 1–6.

[11] D. Smeets, J. Keustermans, D. Vandermeulen, and P. Suetens, "meshSIFT: Local surface features for 3D face recognition under expression variations and partial data," *Comput. Vis. Image Understand.*, vol. 117, no. 2, pp. 158–169, Feb. 2013.

[12] H. Li, D. Huang, P. Lemaire, J.-M. Morvan, and L. Chen, "Expression robust 3D face recognition via mesh-based histograms of multiple order surface differential quantities," in *Proc. 18th IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 3053–3056.

[13] H. Li, L. Chen, D. Huang, Y. Wang, and J.-M. Morvan, "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.

[14] S. Berretti, N. Werghi, A. del Bimbo, and P. Pala, "Matching 3D face scans using interest points and local histogram descriptors," *Comput. Graph.*, vol. 37, no. 5, pp. 509–525, 2013.

[15] E. Tola, V. Lepetit, and P. Fua, "A fast local descriptor for dense matching," in *Proc. Int. Conf. Comput. Vis. Pattern Recognit.*, Anchorage, AK, USA, Jun. 2008, pp. 1–8.

[16] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[17] A. Zaharescu, E. Boyer, K. Varanasi, and R. Horaud, "Surface feature detection and description with applications to mesh matching," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Miami, FL, USA, Jun. 2009, pp. 373–380.

[18] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *Proc. Eur. Conf. Comput. Vis.*, Prague, Czech Republic, May 2004, pp. 469–481.

[19] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.

[20] D. Huang, C. Shan, M. Ardabilian, Y. Wang, and L. Chen, "Local binary patterns and its application to facial image analysis: A survey," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 41, no. 6, pp. 765–781, Nov. 2011.

[21] S. Z. Li, C. Zhao, M. Ao, and Z. Lei, "Learning to fuse 3D+2D based face recognition at both feature and decision levels," in *Proc. Int. Work. Anal. Modeling Faces Gestures*, Beijing, China, Oct. 2005, pp. 44–54.

[22] Y. Huang, Y. Wang, and T. Tan, "Combining statistics of geometrical and correlative features for 3D face recognition," in *Proc. Brit. Mach. Vis. Conf.*, Edinburgh, Scotland, Sep. 2006, pp. 879–888.

[23] D. Huang, M. Ardabilian, Y. Wang, and L. Chen, "3-D face recognition using eLBP-based facial description and local feature hybrid matching," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 5, pp. 1551–1565, Oct. 2012.

[24] G. Sandbach, S. Zafeiriou, and M. Pantic, "Local normal binary patterns for 3D facial action unit detection," in *Proc. IEEE Int. Conf. Image Process.*, Orlando, FL, USA, Sep. 2012, pp. 1813–1816.

[25] H. Li, L. Chen, D. Huang, Y. Wang, and J. Morvan, "3D facial expression recognition via multiple kernel learning of multi-scale local normal patterns," in *Proc. Int. Conf. Pattern Recognit.*, Nov. 2012, pp. 2577–2580.

[26] G. Sandbach, S. Zafeiriou, and M. Pantic, "Binary pattern analysis for 3D facial action unit detection," in *Proc. Brit. Mach. Vis. Conf.*, Guildford, U.K., Sep. 2012, pp. 1–12.

[27] H. Tang, B. Yin, Y. Sun, and Y. Hu, "3D face recognition using local binary patterns," *Signal Process.*, vol. 93, no. 8, pp. 2190–2198, Aug. 2013.

[28] N. Bayramoglu, G. Zhao, and M. Pietikäinen, "CS-3DLBP and geometry based person independent 3d facial action unit detection," in *Proc. Int. Conf. Biometrics (ICB)*, Madrid, Spain, Jun. 2013, pp. 1–6.

[29] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "An evaluation of multimodal 2D+3D face biometrics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 4, pp. 619–624, Apr. 2005.

[30] X. Lu and A. K. Jain, "Deformation modeling for robust 3D face matching," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, New York, NY, USA, Jun. 2006, pp. 1377–1383.

[31] C. Beumier and M. Acheroy, "Face verification from 3D and grey level clues," *Pattern Recognit. Lett.*, vol. 22, no. 12, pp. 1321–1329, Oct. 2001.

[32] M. Hüsken, M. Brauckmann, S. Gehlen, and C. Von der Malsburg, "Strategies and benefits of fusion of 2D and 3D face recognition," in *Proc. IEEE Workshop Face Recognit. Grand Challenge*, San Diego, CA, USA, Jun. 2005.

[33] A. S. Mian, M. Bennamoun, and R. Owens, "An efficient multimodal 2D-3D hybrid approach to automatic face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 11, pp. 1927–1943, Nov. 2007.

[34] B. Gökberk, H. Dutağaci, A. Ulas, L. Akarun, and B. Sankur, "Representation plurality and fusion for 3-D face recognition," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 38, no. 1, pp. 155–173, Feb. 2008.

[35] W. B. Soltana, D. Huang, M. Ardabilian, L. Chen, and C. B. Amar, "Comparison of 2D/3D features and their adaptive score level fusion for 3D face recognition," in *Proc. 3D Data Process., Visualizat. Transmiss. (3DPVT)*, Paris, France, May 2010, pp. 1–8.

[36] N. Werghi, S. Berretti, and A. del Bimbo, "The mesh-LBP: A framework for extracting local binary patterns from discrete manifolds," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 220–235, Jan. 2015.

[37] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[38] A. Ross and A. Jain, "Information fusion in biometrics," *Pattern Recognit. Lett.*, vol. 24, no. 13, pp. 2115–2125, 2003.

[39] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 4–20, Jan. 2004.

[40] T. Maurer *et al.*, "Performance of geometrix ActiveID 3D face recognition engine on the FRGC data," in *Proc. IEEE CVPR Workshop Face Recognit. Grand Challenge Experim.*, Jun. 2005, p. 154.

[41] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "Multimodal 2D and 3D biometrics for face recognition," in *Proc. IEEE Int. Workshop Anal. Modeling Faces Gestures (AMFG)*, Oct. 2003, pp. 187–194.

[42] Y. Lei, M. Bennamoun, and A. A. El-Sallam, "An efficient 3D face recognition approach based on the fusion of novel local low-level features," *Pattern Recognit.*, vol. 46, no. 1, pp. 24–37, Jan. 2013.

[43] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.

[44] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3D facial expression database for facial behavior research," in *Proc. IEEE 7th Int. Conf. Autom. Face Gesture Recognit.*, Southampton, U.K., Apr. 2006, pp. 211–216.

[45] Y. V. Venkatesh, A. A. Kassim, J. Yuan, and T. D. Nguyen, "On the simultaneous recognition of identity and expression from BU-3DFE datasets," *Pattern Recognit. Lett.*, vol. 33, no. 13, pp. 1785–1793, Oct. 2012.

[46] N. Alyüz, B. Gökberk, and L. Akarun, "A 3D face recognition system for expression and occlusion invariance," in *Proc. IEEE Int. Conf. Biometrics, Theory, Appl., Syst.*, Washington, DC, USA, Sep./Oct. 2008, pp. 1–7.

[47] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, vol. 1. Pacific Grove, CA, USA, Nov. 1993, pp. 40–44.

[48] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[49] N. Werghi, C. Tortorici, S. Berretti, and A. Del Bimbo, "Representing 3D texture on mesh manifolds for retrieval and recognition applications," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 2521–2530.

Naoufel Werghi received the Ph.D. degree in computer vision from the University of Strasbourg. He was a Visiting Professor with the Department of Electrical and Computer Engineering, University of Louisville, in 2002, and the Media Integration and Communication Center, University of Florence, in 2012. He has been a Research Fellow with the Division of Informatics, University of Edinburgh, and a Lecturer with the Department of Computer Sciences, University of Glasgow. He is currently an Associate Professor with the Electrical and Computer Engineering Department, Khalifa University of Science, Technology and Research, UAE. He published over 100 journal and conference papers. His main research area is 2-D/3-D image analysis and interpretation, where he has been leading several funded projects in the areas of biometrics, medical imaging, geometrical reverse engineering, and intelligent systems.

Claudio Tortorici received the bachelor's and master's degrees in computer engineering from the University of Florence, in 2011 and 2013, respectively. He was with the Media Integration and Communication Center as an Internee from 2013 to 2014. He is currently a Research Associate with the Visual Signal Analysis Processing Center, Khalifa University of Science, Technology and Research, UAE. He has been working on research projects related to multimedia information retrieval and 3-D face recognition. He is a Reviewer of the *Scientific World Journal*.

Stefano Berretti received the Ph.D. degree in information and telecommunications engineering from the University of Florence, Italy, in 2001. He has been a Visiting Researcher with IIT Mumbai, India, and a Visiting Professor with the Institute TELECOM, TELECOM Lille 1, Lille, France, and Khalifa University, Sharjah, UAE. He is currently an Associate Professor with the Department of Information Engineering and the Media Integration and Communication Center, University of Florence. He has authored over 100 papers in conference proceedings and international journals in the area of pattern recognition, computer vision, and multimedia. His main research interests focus on 3-D object retrieval and partitioning, face recognition and facial expression recognition from 3-D and 4-D data, and 3-D face superresolution, human action recognition from 3-D data. He is on the program committee of several international conferences and serves as a frequent reviewer of many international journals. He was the Cochair of the Fifth Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment (2012) in conjunction with ECCV 2012.

Alberto Del Bimbo was the Deputy Rector for Research and Innovation Transfer with the University of Florence from 2000 to 2006. He is a Full Professor of Computer Engineering, the Director of the Master in Multimedia, and the Director of the Media Integration and Communication Center with the University of Florence. His scientific interests are multimedia information retrieval, pattern recognition, image and video analysis, and natural human–computer interaction. From 1996 to 2000, he was the President of the IAPR Italian Chapter and the Member-at-Large of the IEEE Publication Board from 1998 to 2000. He was the General Chair of IAPR ICIAP97, the International Conference on Image Analysis and Processing, IEEE ICMCS99, the International Conference on Multimedia Computing and Systems and Program Cochair of ACM Multimedia 2008. He was the General Cochair of ACM Multimedia 2010 and the European Conference on Computer Vision in 2012. He is an IAPR Fellow, and an Associate Editor of *Multimedia Tools and Applications*, *Pattern Analysis and Applications*, the *Journal of Visual Languages and Computing*, and the *International Journal of Image and Video Processing*, and was an Associate Editor of *Pattern Recognition*, the IEEE TRANSACTIONS ON MULTIMEDIA, and the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE. He serves as the Editor-in-Chief of the *ACM Transactions on Multimedia Computing, Communications, and Applications*.