

# Representing 3D Texture on Mesh Manifolds for Retrieval and Recognition Applications

Naoufel Werghi, Claudio Tortorici  
Khalifa University of Science, Technology & Research  
Sharjah, UAE

Naoufel.Werghi@kustar.ac.ae

Stefano Berretti, Alberto Del Bimbo  
University of Florence  
Florence, Italy

stefano.berretti@unifi.it

## Abstract

In this paper, we present and experiment a novel approach for representing texture of 3D mesh manifolds using local binary patterns (LBP). Using a recently proposed framework [37], we compute LBP directly on the mesh surface, either using geometric or photometric appearance. Compared to its depth-image counterpart, our approach is distinguished by the following features: a) inherits the intrinsic advantages of mesh surface (e.g., preservation of the full geometry); b) does not require normalization; c) can accommodate partial matching. In addition, it allows early-level fusion of the geometry and photometric texture modalities. Through experiments conducted on two application scenarios, namely, 3D texture retrieval and 3D face recognition, we assess the effectiveness of the proposed solution with respect to state of the art approaches.

## 1. Introduction

The advancement of 3D imaging technologies resulted in a new generation of acquisition devices capable of capturing the geometry of 3D objects in the three-dimensional physical space. The geometric information captured by such 3D acquisition devices is typically in the form of a cloud of points, which represents the three-dimensional coordinates of a set of samples of the 3D object surface. However, the direct processing of these point clouds is not convenient or even possible, so that other representation formats have been established. Depth images are one of the most commonly used imaging modality, since they permit a direct extension to the depth dimension of many computer vision and pattern recognition solutions developed for analyzing the photometric information in 2D images. Though the possibility of a straightforward extension of 2D techniques is attractive, this modality loses the full 3D geometry, by reducing it to a 2.5D projection. The full 3D shape information is instead preserved and encoded in a simple, compact

and flexible format by the triangular mesh manifold modality. This is widely used in many fields, such as animation, medical imaging, computer-aided design, terrain modeling, etc. The recent advances in shape scanning and modeling have also allowed the integration of both photometric and geometric information into a single support defined over a 3D mesh-manifold. However, despite the abundance and the richness of the mesh manifold modality, the number of solutions for representing the geometry of 3D objects is still limited, and not comparable with the large variety of methods available in 2D.

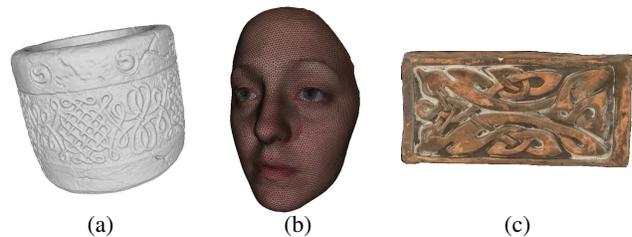


Figure 1. Example 3D objects with different 3D textures: (a) 3D *geometric texture*, characterized by repeatable patterns of the mesh surface; (b) 3D *photometric texture* attached to the triangular mesh. In this case, the textural information is most present in the photometric appearance of the mesh, rather than in the geometric appearance; (c) Combination of 3D geometric and photometric texture on a 3D mesh manifold.

An evidence of this is given by the lack of efficient descriptors to represent the texture component associated to 3D objects. This motivated us to focus on this aspect that can reveal new possibilities in 3D objects retrieval and recognition. In particular, two different meanings are associated here to the broad term “texture”: On the one hand, we consider the 3D *geometric texture* as a property of the surface, distinct from the shape, which is characterized by the presence of repeatable geometric patterns (see Fig. 1(a)). These patterns can be seen as geometric corrugations of the surface that do not alter the overall 3D shape, but rather change the local smoothness and appearance of the surface.

This can result in 3D objects that show similar or equal shape, but very different 3D geometric texture; On the other, the 3D *photometric texture* attached to the mesh is related to the photometric appearance of the surface as captured by a 2D imaging sensor (see Fig. 1(b)). Being attached to the triangular mesh, this property of the 3D surface is represented in the RGB domain. This represents a different concept from the 2D texture, since it is represented on the mesh, rather than on the image plane, though using the same RGB domain of the 2D counterpart. These two properties may also be present together, for a 3D object characterized by both the combined 3D geometric and photometric texture (see Fig. 1(c)). For capturing these two aspects of the 3D object appearance on a triangular mesh support, consolidated approaches do not exist. In this work, we present a novel approach for holding both aspects within a single framework. We believe it is the first of its kind to address this problem, to the best of our knowledge.

### 1.1. Related work

In the literature, the problem of representing the 3D geometric texture has been not addressed directly; rather, the 3D geometric texture has been managed as a component of the surface shape either recurring to 3D shape descriptors [15, 29], that in the large part are not adequate to capture 3D geometric texture, or resorting to the 2D case by applying 2D descriptors to planar projections of the 3D surface, in the form of depth images. In addition, the photometric properties of the surface have either been not considered or have been addressed following the obvious solution of processing the 2D texture image separately in the planar domain. But, similarly to the case of depth images, this loses the possibility to represent occluded parts or closed surfaces as a whole, and hinders the possibility of an early fusion of geometric and photometric descriptors.

In this paper, we address the above shortcomings building on the framework of Local Binary Pattern (LBP). Since its first formal definition by Ojala *et al.* [27, 28], the LBP has established itself as one of the most effective local shape descriptors for image representations. It has been originally introduced for representing 2D textures in still images, but its computational simplicity and discriminative power attracted the attention of the image processing and pattern recognition community for other tasks. Rapidly, LBP has found applications in visual inspection [10, 23], remote sensing [13, 22, 34], face recognition [2, 3, 38, 40], facial expression recognition [33], and motion analysis [9, 36]. However, the LBP-based methods developed so far operate either on photometric information provided by 2D color images or on geometric information in 2D depth images. The few solutions that extract surface features directly in 3D (typically in the form of surface normals), resort to the 2D case by converting the 3D extracted features to depth

values, and then use ordinary LBP processing on 2D images [19, 32, 31]. 3D-LBP approaches advanced the state of the art and proved to be competitive with other classes of methods. However, their applications 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 lateral areas of the nose). Recently, LBP construction on triangular mesh manifolds has been introduced by Werghi *et al.* [37]. The mesh-LBP framework keeps the simplicity and the elegance characterizing the original LBP, while relieving the recognition process from the need for normalization, and preserving the full 3D geometry of the shape.

### 1.2. Our contribution and paper organization

In this paper, we target the problem of representing the texture properties of 3D mesh manifolds for retrieval and recognition applications. As a main contribution of this work, we propose a solution, which is based on the recently proposed mesh-LBP concept [37], to address the above challenges. In particular, to the best of our knowledge, this paper is the first one to present and apply a unified framework, which enables an elegant and effective representation of 3D geometric and photometric texture. Two novel solutions have been proposed and investigated: 1) A retrieval approach of 3D objects based on the 3D geometric texture of the surface using the mesh-LBP representation. To the best of our knowledge, this is the first attempt to exploit 3D geometric texture rather than 3D shape to perform 3D retrieval; 2) A mesh-LBP based face representation that can be constructed over triangular mesh manifolds. This representation relieves the recognition process from the need for normalization, while it preserves the full 3D geometry of the shape. Furthermore, mesh-LBP construction allows boosting recognition by offering an elegant framework for fusing, over a mesh support, photometric texture and shape information at data and feature level, in addition to score and decision level. To the best of our knowledge, this work is the first one to propose texture and shape fusion for face recognition using LBP patterns constructed on the mesh.

The rest of the paper is organized as follows: In Sect. 2, we give an overview on the mesh-LBP concept; In Sect. 3, the 3D texture retrieval scenario is introduced and the mesh-LBP results are presented in comparison to other state of the art solutions; In Sect. 4, we describe how mesh-LBP is used for constructing face representation, and for fusing shape and appearance information. Experimental results for face recognition are exposed in Sect. 4.2 and Sect. 4.3, respectively, for the BU-3DFE and the Bosphorus datasets; Concluding remarks are discussed in Sect. 5.

## 2. LBP descriptor on the mesh

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  $3 \times 3$  neighborhood. 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.* [37] elegantly extended the LBP concept to the 3D mesh manifold 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. 2).

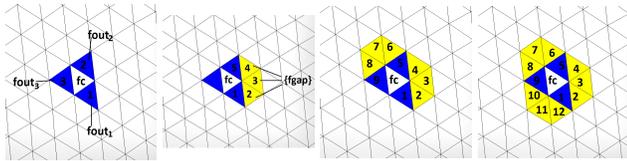


Figure 2. Ordered ring construction: From the initial *Fout* facets formed by the three ordered facets  $f_{out_1}$ ,  $f_{out_2}$ , and  $f_{out_3}$  that are adjacent to the central facet  $f_c$ , a sequence of *Fgap* facets located between each pair  $\langle f_{out_1}, f_{out_2} \rangle$ ,  $\langle f_{out_2}, f_{out_3} \rangle$ , and  $\langle f_{out_3}, f_{out_1} \rangle$  are extracted. The *Fgap* facets have exactly one vertex on the initial three-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  $f_c$ .

Starting with three–clockwise or anticlockwise–ordered *Fout* facets ( $f_{out_1}$ ,  $f_{out_2}$ , and  $f_{out_3}$  in Fig. 2), 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 Fig. 3). 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 [37]) at different radial and azimuthal resolutions, while preserving the simplicity of the original LBP.

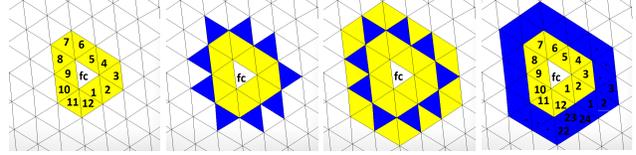


Figure 3. By iterating the procedure of Fig. 2, using as new set of *Fout* facets the sequence of facets that share an edge on the outer contour of the current ring, a sequence of rings of ordered facets can be generated.

Let  $h(f)$  be a scalar function defined on the mesh, which can incarnate either a geometric (*e.g.*, curvature) or photometric (*e.g.*, color) information. The mesh-LBP operator, as proposed in [37], is defined as follows:

$$meshLBP_m^r(f_c) = \sum_{k=0}^{m-1} s(h(f_k^r) - h(f_c)) \cdot \alpha(k), \quad (1)$$

$$s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases},$$

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. The discrete function  $\alpha(k)$  is introduced for the purpose of deriving different LBP variants. For example,  $\alpha(k) = 2^k$  results into the mesh counterpart of the basic LBP operator firstly suggested by Ojala *et al.* [27]; with  $\alpha(k) = 1$ , we obtain the sum of the digits composing the binary pattern. In the experiments, we will refer to these two functions by  $\alpha_1$  and  $\alpha_2$ , respectively.

To cope with mesh tessellation irregularities, the scalar function  $h(f)$  is interpolated and sub-sampled across each ring, allowing thus to maintain a constant azimuthal quantization. In [37] it is shown that this technique copes to a large extent with mesh irregularity. Invariance with respect to the position of the first facet in the ring is assured by selecting the closest face to the center of mass of the central facet’s neighbourhood. The presence of uniform patterns has been also investigated in [37], and it has been found that the majority of patterns have a number of 0-1 transitions below 4.

In the rest of the paper, we exploit the mesh-LBP in two different surface analysis applications, which involve both local and global representations, namely, 3D texture retrieval (Sect. 3), and 3D face matching (Sect. 4).

## 3. Retrieval based on 3D geometric texture

This experiment aims to assess the mesh-LBP potential for detecting a specific type of texture in a given surface. Such capacity is useful in “3D texture retrieval” applications, where a sample of specific 3D texture (probe texture) is available, and we want to automatically detect regions, in

a gallery surface, matching that particular sample. Potential applications of this scenario can be found in 3D medical imaging, or 3D terrain relief inspection. To the best of our knowledge, we are the first to attempt retrieving 3D geometric texture on a mesh manifold. However, elaborating a full scheme for 3D texture retrieval is beyond the scope of this paper. Rather, we aim to showcase the potential of the mesh-LBP and its performance for such a task in comparison with other standard descriptors. Therefore, we just used a naive template-matching-like method, where the gallery mesh surface is browsed, and at each facet a texture descriptor is computed and compared to its probe texture model counterpart using a given metric (i.e., the Bhattacharyya distance in this application). Facets exhibiting a low distance (i.e., below a certain threshold) are selected as a potential match. Of course, this exhaustive search approach is prone to inefficiency issues, but addressing this aspect is out of the scope of the present treatment.

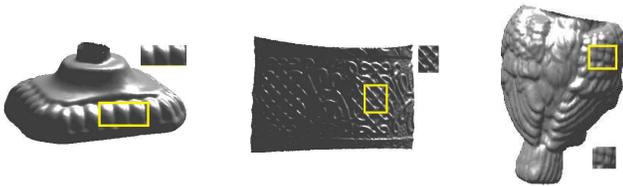


Figure 4. Surfaces extracted from the *bird*, *pot* and *owl* models, and their corresponding position, highlighted with a rectangle, in the probe models.

In the experiments, we considered as gallery a representative set of three surfaces (Fig. 4), exhibiting different global and local shape characteristics. These surfaces were extracted from the *bird*, *pot*, and *owl* objects in the “MIT CSAIL textured 3D models database” [1]. The order of the aforementioned objects reflects an ascending level of surface complexity and 3D texture richness. The *bird* instance contains basically two free-from surfaces, one is smooth and the other is textured. The *pot* surface is composed of a single cylindrical surface exhibiting different types of 3D texture patterns. The *owl* surface, which is the most complex one, encompasses different free-from textured surfaces. The set of probe textures is composed of three instances extracted from the three surfaces, as shown in Fig. 4. The experiment consists in searching each probe within its corresponding surface, and then assessing the detection and retrieval capacity of the different descriptors. In so doing, we computed the  $\alpha_1$  and  $\alpha_2$  mesh-LBP variants, using the *Gaussian curvature* ( $K$ ), *mean curvature* ( $H$ ) and *angle between facets normal* ( $D$ ) as surface functions. In addition, we compared the mesh-LBP with other standard 3D surface descriptors including: the *Shape Distribution* variants [29], namely, the distance between a fixed point and one random point on the surface ( $D1$ ), the distance between two ran-

dom points on the surface ( $D2$ ), the square root of the area of the triangle between three random points on the surface ( $D3$ ), the cube root of the volume of the tetrahedron between four random points on the surface ( $D4$ ), the angle between three random points on the surface ( $A3$ ); and the *Spin-Images* [15]. In the case of the shape-distribution descriptor, results are reported only for the  $D4$  variant, since it provided the best result when compared with the  $D1$ ,  $D2$ ,  $D3$  and  $A3$  variants. In addition, we also tested the recently proposed *Intrinsic Shape Context* (ISC) [17].

Figures 5(a), (b) and (c) show the maps of the Bhattacharyya distance computed at each facet and the related retrieval results for the *bird*, *pot* and *owl* objects, respectively. Referring to the distance maps (first row of each case), we can assert that the *shape-distribution*  $D4$  practically shows no possibility for detecting the searched texture. The *spin-images* looks partially spotting the textured regions in the distance maps for the *bird* and the *pot* surfaces, whereas its corresponding *owl* map indicates neat incapacity. The ISC does not indicate a particular ability for spotting the probe texture apart for the *owl* model, but with a significant false detection rate. These observations are confirmed in the texture retrieval results (second row of the cases (a), (b) and (c)), which indicate a nearly total failure in recovering the searched texture.

On the opposite, the mesh-LBP distance maps indicate a neat superior performance across the three surfaces, though with different levels. For the  $\alpha_1$  variant (shown in the three middle columns), the  $K$  and  $D$  results clearly indicate an ability of detecting the searched texture for the *bird* and the *pot* surfaces, as compared with the  $H$ . The same is noticed for the *owl* surface. These observations are confirmed by the retrieval results, whereby  $K$  and  $D$  instances show the best performance across the three surfaces. For the  $\alpha_2$  operator variant (shown in the three rightmost columns), the distance maps show an overall improvement, particularly visible in the *pot* and the *owl* results. We can observe that for the three surfaces, the regions in the related maps look more compact and localised when compared with their  $\alpha_1$  counterparts. The appearances of these maps suggest an even more ability in texture retrieval, which has been confirmed in the detection results depicted in the last row of each case.

#### 4. Face recognition by fusing shape and 3D photometric texture

The effectiveness of LBP in 2D face recognition has been demonstrated in [3] and in several subsequent works. Motivated by these results, in the following, we investigate if a similar capability exists for the mesh-LBP counterpart. In so doing, we propose an original face representation approach that in addition to exploiting the mesh-LBP capability of capturing 3D texture, it also includes in a unified framework a mesh-LBP descriptor derived by the photomet-

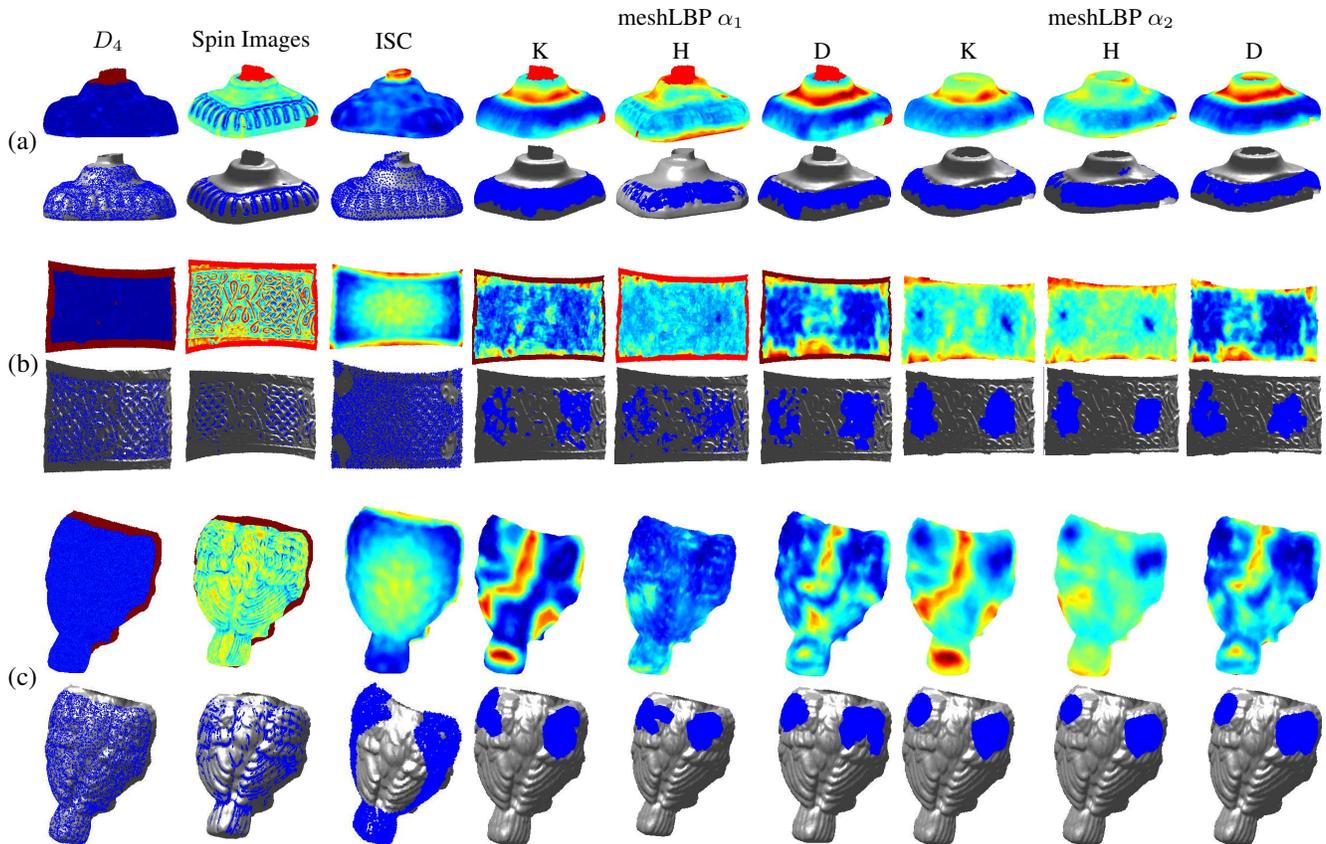


Figure 5. Results for the *bird*, *pot* and *owl* surface are reported in (a), (b) and (c), respectively. Two rows are reported in each case: the upper row represents the distance map obtained with the Bhattacharyya distance; in the lower row, the region on the mesh where the probe texture is best identified is highlighted in blue. For mesh-LBP, the  $\alpha_1$  and  $\alpha_2$  operators are evaluated, in combination with the surface functions *Gaussian curvature*  $K$ , *mean curvature*  $H$ , and *angle between facets normal*  $D$ . Comparison with *Shape Distribution* variant  $D_4$ , *Spin Images* and *ISC* are also reported in the first three columns. (Figure best viewed in soft-copy).

ric (appearance) component attached to the mesh surface.

In the standard LBP-based face representation [3], 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. To extend this scheme to the 3D face manifold, we need first 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. Our method encompasses the following stages: 1) Construction of a grid of points on the face surface, to obtain an ordered set of regions; 2) Computation of an histogram of the mesh-LBP descriptors computed on the surface region centered at each point of the grid; 3) Concatenation of the regional histograms into a structure encoding either a global or partial description of the face; 4) Performing the face matching.

In the first step (see Fig. 6(a)), the plane formed by the nose tip and the two eyes inner-corner landmark points is initially computed. We choose these three landmarks be-

cause they are the most accurate detectable landmarks of the face, in addition to be robust to facial expressions. From these landmarks we derive, via simple geometric calculation, an ordered and regularly spaced set of points on that plane. 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 (see Fig. 6(b)), which defines an atlas for the regions that divide the facial surface. Around each grid point, we extract a neighbourhood of facets. These can be defined by the set of facets confined within a geodesic disc or a sphere, centered at that grid point. In the second step, we compute a multi-resolution mesh-LBP descriptor, using (1), for each facet in a region, considering both shape-valued and appearance-valued functions. In the third 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. 7(a)). In the last step, we employ a basic minimum dis-

tance classifier using  $\chi^2$  distance, as we aim at showcasing the performance enhancement that our 3D face representation can bring to face recognition, rather than demonstrating its superiority over other competitive schemes.

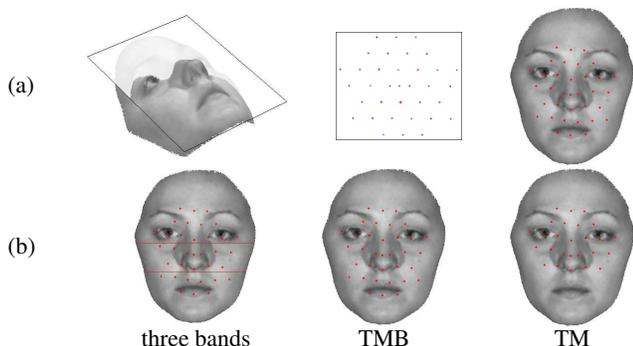


Figure 6. (a) Construction of the face grid; (b) On the left scan, partition of the grid points into a top (T), middle (M) and bottom (B) band is reported, whereas in the center and right scans all the points in the three bands (TMB), and only the points in the top and middle (TM) bands are shown, respectively.

#### 4.1. Fusion of mesh-LBP descriptors

Four levels of fusion are typically considered in biometry applications, namely, *data*, *feature*, *score*, and *decision* [30]. As mentioned by Osaimi *et al.* [4], it is believed that low-level fusion performs better than its higher level counterparts (score and decision) [14]. Looking at the spectrum of region methods fusing 2D appearance and 3D shape modalities, we found much concentration in the score-level category [11, 12, 21, 24, 25], as compared to the feature-level [18, 21, 26]. The work of [21], in particular, fused LBP features derived from depth and texture image.

Given the consensus on the advantageous aspects of multi-modal face recognition [8], mesh-LBP 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 patterns constructed on the mesh.

In our approach, we have investigated a score-level fusion (*score\_fus*) and two variants of feature-level fusion (*feat\_fus*). We have chosen the sum rule for the score-level, as it has been proven to be the optimal one [16]. In the first variant of the feature-level, we simply concatenate the two mesh-LBP regional histograms, corresponding to the shape and the appearance functions. For example, considering an azimuthal quantization  $m = 12$  and  $\alpha_1$ , 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 variant, we used

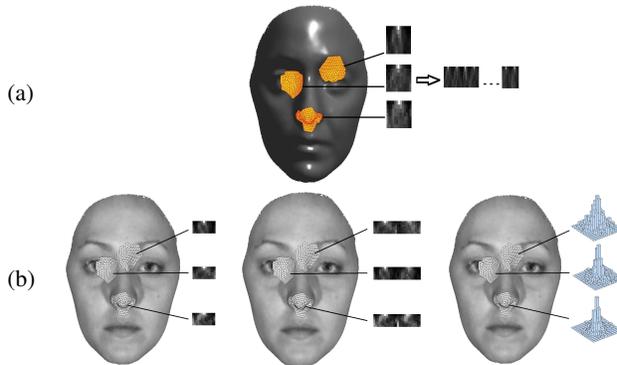


Figure 7. (a) Global histogram construction: Region histograms are computed and then concatenated into a global histogram. (b) Examples of regional histogram variants obtained with  $m = 12$ ,  $r = 7$  and  $\alpha_1$ : (left) A  $7 \times 13$  unimodal histogram corresponding to a shape function; (middle) A  $7 \times 26$  *feat\_fus1* histogram obtained by concatenating two  $7 \times 13$  histograms corresponding to a shape function and a photometric function (gray level); (right) A 2D section of a *feat\_fus2*  $7 \times 13 \times 13$  histogram obtained with a shape function and a photometric function.

a 2-D accumulator that counts for the co-occurrences of the mesh-LBP corresponding to the shape and the appearance functions. For the same aforementioned parameters' values, we obtain an  $r \times 13 \times 13$  histogram (Fig. 7(b) depicts some examples). In the rest of the paper, we will refer to these first and second variants by *feat\_fus1* and *feat\_fus2*, respectively.

#### 4.2. Face recognition results on BU-3DFE dataset

The BU-3DFE database from Binghamton University [39] 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. Actually, we note that level-1 scans already exhibit significant disparity from the neutral expression, especially for the disgust, fear, happy and surprise expressions, as it can be noticed in Fig. 8(a). We considered as gallery the set of neutral scans, while the expressive scans are used as probes. Scans in this database contain both shape and appearance data. 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. In addition, the appearance image captured by the scanner comprises two views of the face at about  $+45^\circ$  and  $-45^\circ$ . This gives to the 2D texture image the necessary angle to expose the side parts of the face, thus avoiding auto-occlusions, mainly due to the nose, that can occur in the case a single frontal image of the face is captured.

We set the radial resolution  $r$  and the azimuthal quan-



Figure 8. 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.

tization  $m$  to 7 and 12, respectively. To account for the effects of facial expressions, we segmented the grid points into three bands, dubbed *top* (T), *middle* (M) and *bottom* (B). 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  $\alpha_1$  and  $\alpha_2$ . We found that the  $H$  and  $C$  descriptors perform best than the rest, so we will report results related to these descriptors. Results for the mesh-LBP descriptor based on the *gray-level* (GL) on the mesh are also reported.

Table 1 shows the probe scans categorized into the six different facial expressions, and recognition rates are reported for each category separately. We also included results obtained with three variants of the interest-points method proposed in [7], and which have been applied on the same database. Methods in [18, 35] 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 [35] 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 [18] 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 1, we first notice that our method outperforms [7] even with variants using single modality (see scores related to  $H$  and  $C$ ). 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 ob-

tained with the feature-level fusion variants. This observation is confirmed in the over-all results, whereby the configuration using  $\langle \alpha_2, feat\_fus_1, H \rangle$  scores the best performance. Also, we can notice that our framework preserves an overall reasonable performance for the higher expression amplitudes *level-3* and *level-4*.

### 4.3. Face recognition results on Bosphorus dataset

The Bosphorus database [5] contains 4666 scans of 105 subjects. The subjects were scanned in different poses, action units, and occlusion conditions. We assess the multi-modal variant  $\langle TMB, H, \chi^2 \rangle$  of our method on the database’s subsets corresponding to the seven facial expressions, *Lower Face Action Unit* (LFAU), *Upper Face Action Unit* (UFAU), and *Combined Action Unit* (CAU). Scans with *Yaw Rotation* (YR), *Pitch Rotation* (PR), and *Cross Rotation* (CR) are instead excluded from our analysis at this stage. We compared our method with Berretti *et al.* [7] and Li *et al.* [20], which used the same experimental protocol. Sandbach *et al.* [31] and Bayramoglu *et al.* [6] used also the same database, but for facial expression and action unit recognition purpose rather than face recognition. Their experimental setting is also different than ours. They employed, respectively, AdaBoost and Random Forest classifiers, and a 10-fold cross-validation scheme, whereas our method used a simple minimum-distance classifier.

An overview of the comparison results depicted in Table 2 shows that our framework outperforms the methods in [7] and [20]. In more details, the  $H$  mesh-LBP shape descriptor (columns 4 and 9) is quite competitive for the  $\alpha_2$  variant, whereby it scores either better or slightly less than these two methods at nearly all the subset instances. The same can also be observed for  $\alpha_1$  fusion variants, which show a better overall scores. The  $\alpha_2$  fusion scores better across all the subset instances, reflecting thus a neatly superior performance.

## 5. Discussion and conclusions

In this paper, we presented two original contributions that derive from the application of the mesh-LBP framework on triangular mesh manifolds:

- A 3D retrieval paradigm based on the 3D texture of a mesh surface. While qualitative, and despite employing basic techniques for descriptor comparison, the experiments on 3D texture retrieval revealed the great potential of the mesh-LBP descriptors and the incapacity of the standard descriptors for such a task. This paradigm brings an adequate solution for 3D texture analysis and retrieval in imaging modalities for which photometric information is lacking or not reliable;
- An approach for constructing a multi-modal LBP-based face representation on a triangular mesh-model.

Descriptors		Level-1 & Level-2 Expressions							Level-3 & Level-4 Expressions						
		An	Di	Fe	Ha	Sa	Su	All	An	Di	Fe	Ha	Sa	Su	All
[7]	HOG	90.0	87.5	88.8	88.1	90.6	85.0	88.3	81.3	75.6	78.8	80.6	82.5	76.9	79.3
	SHOT	93.8	90.6	91.9	90.0	94.4	88.8	91.6	87.5	78.8	85.6	79.4	90.0	79.4	83.4
	GH	90.6	85.0	84.4	85.6	90.6	82.5	86.5	86.3	79.4	80.0	79.4	85.0	78.8	81.5
$\alpha_1$	GL	83.0	72.5	83.0	83.5	86.0	81.5	81.6	67.0	49.5	71.0	69.0	76.5	71.5	67.4
	H	89.0	74.5	83.5	89.0	96.5	93.0	87.6	73.5	48.0	68.0	79.0	85.0	84.0	72.9
	C	88.5	68.0	79.5	86.0	93.0	91.5	84.4	69.0	43.5	65.5	73.5	82.5	83.5	69.6
	<i>feat_fus<sub>1</sub></i> H	95.5	86.5	92.5	94.5	97.0	97.5	93.9	82.5	67.5	86.0	86.5	94.0	94.0	85.1
	<i>feat_fus<sub>1</sub></i> C	94.5	83.0	91.0	94.5	96.5	97.5	92.8	82.0	59.5	86.0	87.0	92.0	93.5	83.3
	<i>feat_fus<sub>2</sub></i> H	94.0	85.5	91.5	95.0	97.5	96.5	93.3	83.0	66.5	85.5	87.5	93.0	93.5	84.8
	<i>feat_fus<sub>2</sub></i> C	94.0	83.0	90.5	92.5	97.5	97.5	92.5	82.5	62.0	85.0	86.5	93.5	93.0	83.8
	<i>score_fus</i> H	95.0	86.5	92.5	95.0	97.0	98.0	94.0	83.0	67.5	86.5	87.0	94.0	93.5	85.3
<i>score_fus</i> C	94.5	85.5	92.5	94.5	97.0	97.5	93.6	82.5	61.0	86.5	88.0	92.5	93.5	84.0	
$\alpha_2$	GL	88.0	82.0	87.5	89.5	91.0	87.5	87.6	72.5	58.0	80.0	77.5	81.0	85.5	75.8
	H	96.5	90.0	95.5	<b>98.0</b>	<b>99.0</b>	<b>99.5</b>	96.4	92.5	72.5	90.5	92.5	98.0	<b>99.5</b>	90.9
	C	97.0	89.0	95.5	<b>98.0</b>	<b>99.0</b>	<b>99.5</b>	96.3	92.0	69.0	89.5	92.5	97.5	<b>99.5</b>	90.0
	<i>feat_fus<sub>1</sub></i> H	<b>98.0</b>	<b>93.5</b>	<b>96.5</b>	<b>98.0</b>	98.5	<b>99.5</b>	<b>97.3</b>	<b>94.5</b>	<b>80.0</b>	<b>92.5</b>	<b>95.5</b>	<b>98.5</b>	<b>99.5</b>	<b>93.4</b>
	<i>feat_fus<sub>1</sub></i> C	97.5	92.5	96.0	<b>98.0</b>	<b>99.0</b>	<b>99.5</b>	97.1	94.0	73.5	91.0	94.5	97.5	<b>99.5</b>	91.7
	<i>score_fus</i> H	<b>98.0</b>	<b>93.5</b>	<b>96.5</b>	<b>98.0</b>	98.5	<b>99.5</b>	<b>97.3</b>	<b>94.5</b>	<b>80.0</b>	<b>92.5</b>	<b>95.5</b>	<b>98.5</b>	<b>99.5</b>	<b>93.4</b>
	<i>score_fus</i> C	97.5	92.5	96.0	<b>98.0</b>	<b>99.0</b>	<b>99.5</b>	97.1	94.0	73.5	91.0	94.5	97.5	<b>99.5</b>	91.7

Table 1. BU-3DFE: Rank-1 recognition rate obtained for the different expression subsets compared with [7].

	[7]	[20]	$\alpha_1$					$\alpha_2$				
			GL	H	<i>feat_fus<sub>2</sub></i>	<i>feat_fus<sub>1</sub></i>	<i>score_fus</i>	GL	H	<i>feat_fus<sub>1</sub></i>	<i>score_fus</i>	
Neutral	97.9	<b>100.0</b>	<b>100.0</b>	99.5	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>
Anger	85.9	88.7	95.8	81.7	<b>97.2</b>	95.8	<b>97.2</b>	93.0	93.0	94.4	94.4	94.4
Disgust	81.2	76.8	85.5	53.6	85.5	89.9	89.9	91.3	76.8	<b>92.8</b>	<b>92.8</b>	<b>92.8</b>
Fear	90.0	92.9	92.9	81.4	<b>98.6</b>	<b>98.6</b>	<b>98.6</b>	<b>98.6</b>	92.9	<b>98.6</b>	<b>98.6</b>	<b>98.6</b>
Happy	92.5	95.3	83.0	68.9	88.7	89.6	91.5	92.5	87.7	<b>97.2</b>	<b>97.2</b>	<b>97.2</b>
Sad	93.9	95.5	97.0	83.3	97.0	97.0	<b>98.5</b>	<b>98.5</b>	97.0	<b>98.5</b>	<b>98.5</b>	<b>98.5</b>
Surprise	91.5	98.6	95.8	93.0	97.2	98.6	98.6	<b>100.0</b>	98.6	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>
LFAU	96.5	97.2	97.1	87.9	97.1	97.4	97.6	<b>99.2</b>	96.2	99.1	99.1	99.1
UFAU	98.4	99.1	98.6	95.8	99.8	99.5	99.8	<b>100.0</b>	98.8	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>
CAU	95.6	98.8	<b>100.0</b>	94.7	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	98.8	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>
All	95.7	96.9	96.7	88.3	97.1	97.6	97.9	98.8	96.1	<b>99.0</b>	<b>99.0</b>	<b>99.0</b>

Table 2. Bosphorus: Rank-1 recognition accuracy obtained for some subsets of the database in comparison to [7] and [20].

It is the first approach that integrates shape and appearance information in LBP-patterns derived from a mesh support. The combination of 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-occlusions. The experiments conducted with BU-3DFE and Bosphorus database showcased the boosting of the recognition performance brought by our fusion framework, and its superiority with regard to the most closest approaches. We have showed that our framework can be

easily adapted to different fusion schemes, in particular the early stage fusion.

The comparison of the  $\alpha_1$  and  $\alpha_2$  mesh-LBP functions across the different experiments, gives more credentials to the second one. However, the compactness of the  $\alpha_1$  descriptor, and the resulting lower computational complexity required to compare descriptors, vote for this solution when time constraints are relevant. In both applications, despite using a basic minimum distance classifier, we showcased the powerfulness of the mesh-LBP descriptors for shape analysis, and the performance enhancement that the mesh-LBP fusion framework can bring in recognition tasks.

Finally, we want to point out the mesh-LBP is more appropriate for the class of manifold objects exhibiting texture or shape variation. So, it is more related to intra-class classification/retrieval, where objects in the class have a similar global structure and different local shape characteristics. While the mesh-LBP is not meant for simplified and optimized mesh, where the mesh uniformity is corrupted.

## Acknowledgments

This work is supported by a research grant NRF, Ref: UIRCA 2013-24877. The mesh-LBP code is publicly available at Matlab Central: <http://www.mathworks.com/matlabcentral/fileexchange/48875-mesh-lbp>

## References

- [1] MIT CSAIL database, 2008. 4
- [2] T. Ahonen, A. Hadid, and M. Pietikäinen. Face recognition with local binary patterns. In *European Conf. on Computer Vision*, pages 469–481, Prague, Czech Republic, May 2004. 2
- [3] T. Ahonen, A. Hadid, and M. Pietikäinen. Face description with local binary patterns: Application to face recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(12):2037–2041, Dec. 2006. 2, 4, 5
- [4] F. R. Al-Osaimi, M. Bennamoun, and A. Mian. Spatially optimized data-level fusion of texture and shape for face recognition. *IEEE Trans. Image Processing*, 21(2):859–872, Feb. 2012. 6
- [5] N. Alyz, B. Gkberk, and L. Akarun. 3D face recognition system for expression and occlusion invariance. In *IEEE Int. Conf. on Biometrics: theory, applications, and systems*, pages 1–7, Washington, DC, USA, 2008. 7
- [6] N. Bayramoglu, G. Zhao, and M. Pietikäinen. CS-3DLBP and geometry based person independent 3d facial action unit detection. In *Int. Conf. on Biometrics*, pages 1–6, Madrid, Spain, June 2013. 7
- [7] S. Berretti, N. Werghi, A. Del Bimbo, and P. Pala. Matching 3D face scans using interest points and local histogram descriptors. *Computers & Graphics*, 37(5):509–525, Aug. 2013. 7, 8
- [8] K. 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*, 101:1–5, Nov. 2006. 6
- [9] L. Cai, C. Ge, Y. Zhao, and X. Yang. Fast tracking of object contour based on color and texture. *Int. Journal of Pattern Recognition and Artificial Intelligence*, 23(7):1421–1438, Nov. 2009. 2
- [10] L. Cao, J. Luo, F. Liang, and T. Huang. Heterogeneous feature machines for visual recognition. In *Int. Conf. on Computer Vision*, pages 1095–1102, Kyoto, Japan, Sept. 2009. 2
- [11] K. Chang, K. Bowyer, and P. Flynn. An evaluation of multimodal 2-D and 3-D face biometrics. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(4):619–624, apr 2005. 6
- [12] K. I. Chang, K. Bowyer, and P. J. Flynn. Multimodal 2-d and 3-d biometrics for face recognition. In *IEEE Int. Workshop on Analysis and Modeling of Faces and Gestures*, pages 187–194, Nice, France, Oct. 2003. 6
- [13] D. Guo, V. Atluri, and N. Adam. Texture-based remote-sensing image segmentation. In *Int. Conf. on Multimedia and Expo*, pages 1472–1475, Amsterdam, The Netherlands, July 2005. 2
- [14] A. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Trans. on Circuits Systems and Video Technology*, 14(1):4–20, Jan. 2004. 6
- [15] A. E. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21(5):433–449, May 1999. 2, 4
- [16] J. Kittler, M. Hatef, R. Duin, and J. Matas. On combining classifiers. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(3):226–239, Mar. 1998. 6
- [17] I. Kokkinos, M. Bronstein, R. Littman, and A. Bronstein. Intrinsic shape context descriptors for deformable shapes. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 159–166, Providence, Rhode Island, USA, June 2012. 4
- [18] 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 Recognition*, 46(1):24–37, Jan. 2013. 6, 7
- [19] 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 *Int. Conf. on Pattern Recognition*, pages 2577–2580, Tsukuba, Japan, Nov 2012. 2
- [20] 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 *IEEE Int. Conf. on Image Processing*, pages 3053–3056, Brussels, Belgium, Sept. 2011. 7, 8
- [21] S. Li, C. Zhao, M. Ao, and Z. Lei. Learning to fuse 3D+2D based face recognition at both feature and decision levels. In *Int. Work. on Analysis and Modeling of Faces and Gestures*, pages 44–54, Beijing, China, Oct. 2005. 6
- [22] A. Lucieer, A. Stein, and P. Fisher. Multivariate texture-based segmentation of remotely sensed imagery for extraction of objects and their uncertainty. *Int. Journal of Remote Sensing*, 26(14):2917–2936, 2005. 2
- [23] T. Mäenpää, J. Viertola, and M. Pietikäinen. Optimizing color and texture features for real-time visual inspection. *Pattern Analysis and Applications*, 6(3):169–175, March 2003. 2
- [24] T. Maurer, D. Guigonis, I. Maslov, B. Pesenti, A. Tsaregorodtsev, D. West, and G. Medioni. Performance of Geomatrix ActiveID 3D face recognition engine on the FRGC data. *IEEE CVPR Workshop on Face Recognition Grand Challenge Experiments*, pages 154–159, June 2005. 6
- [25] A. S. Mian, M. Bennamoun, and R. Owens. An efficient multimodal 2D-3D hybrid approach to automatic face recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 29(11):1927–1943, Nov. 2007. 6
- [26] A. S. Mian, M. Bennamoun, and R. Owens. Keypoint detection and local feature matching for textured 3D face recognition. *Int. Journal of Computer Vision*, 79(1):1–12, Aug. 2008. 6

- [27] T. Ojala, M. Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on featured distribution. *Pattern Recognition*, 29(1):51–59, Jan. 1996. [2](#), [3](#)
- [28] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(7):971–987, July 2002. [2](#)
- [29] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin. Shape distributions. *ACM Trans. on Graphics*, 21(4):807–832, Oct. 2002. [2](#), [4](#)
- [30] A. Ross and A. Jain. Information fusion in biometrics. *Pattern Recognition Letters*, 24(13):2115–2125, 2003. [6](#)
- [31] G. Sandbach, S. Zafeiriou, and M. Pantic. Binary pattern analysis for 3D facial action unit detection. In *British Machine Vision Conf.*, pages 1–12, Guildford, UK, Sept. 2012. [2](#), [7](#)
- [32] G. Sandbach, S. Zafeiriou, and M. Pantic. Local normal binary patterns for 3D facial action unit detection. In *IEEE Int. Conf. on Image Processing*, pages 1813–1816, Orlando, FL, Sept. 2012. [2](#)
- [33] C. Shan, S. Gong, and P. W. McOwan. Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing*, 27(6):803–816, May 2009. [2](#)
- [34] C. Song, F. Yang, and P. Li. Rotation invariant texture measured by local binary pattern for remote sensing image classification. In *IEEE Int. Work. on Education Technology and Computer Science*, volume 3, pages 3–6, Wuhan, Hubei, China, Mar. 2010. [2](#)
- [35] 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 Recognition Letters*, 33(13):1785–1793, Oct. 2012. [7](#)
- [36] X. Wang and M. Mirmehdi. Archive film restoration based on spatiotemporal random walks. In *European Conf. on Computer Vision*, pages 478–491, Crete, Greece, 2010. [2](#)
- [37] N. Werghi, S. Berretti, and A. Del Bimbo. The mesh-lbp: a framework for extracting local binary patterns from discrete manifolds. *IEEE Trans. on Image Processing*, 24(1):220–235, Jan. 2015. [1](#), [2](#), [3](#)
- [38] L. Wolf, T. Hassner, and Y. Taigman. Descriptor based methods in the wild. In *ECCV Work. on Faces in Real-Life Images*, pages 1–14, Marseille, France, Oct. 2008. [2](#)
- [39] L. Yin, X. Wei, Y. Sun, J. Wang, and M. Rosato. A 3D facial expression database for facial behavior research. In *IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 211–216, Southampton, UK, Apr. 2006. [6](#)
- [40] G. Zhang, X. Huang, S. Li, Y. Wang, and X. Wu. Boosting local binary pattern (lbp)-based face recognition. In *Int. Work. on Advances in Biometric Person Authentication*, pages 179–186, Beijing, China, Oct. 2005. [2](#)