

# Early Features Fusion over 3D Face for Face Recognition

Claudio Tortorici and Naoufel Werghi<sup>(✉)</sup>

Electrical and Computer Engineering,  
Khalifa University, Abu Dhabi, UAE  
{claudio.tortorici,naoufel.werghi}@kustar.ac.ae

**Abstract.** In this paper, a novel approach for fusing shape and texture Local Binary Patterns (LBP) for 3D Face Recognition is presented. Using the recently proposed *mesh-LBP* [23], it is now possible to compute LBP directly on a mesh manifold, allowing *Early Feature Fusion* to enhance face description power. Compared to its depth image counterparts, the proposed method (a) inherits the intrinsic advantages of mesh surfaces, (such as preservation of full geometry), (b) does not require face registration, (c) can accommodate partial or rotation matching, and (d) natively allows early-level fusion of texture and shape descriptors. The advantages of early-fusion is presented together with an experimentation of two merging schemes tested on the Bosphorus database.

**Keywords:** Local Binary Pattern · Early feature-fusion · LBP · 3D face recognition

## 1 Introduction

The last years have seen an extensive investigation of image usage for human identification and authentication. Even though biometric technologies, such as fingerprint and iris scan, seem to be more accurate, they require more human collaboration than face recognition techniques. Moreover, the creation of 3D imaging technologies has brought a further boost in the development of face recognition. In fact, the new generation of acquisition devices is now capable of capturing the geometry of 3D objects in the three-dimensional physical space.

Besides shape information, face imaging, in general, has emerged as promising modality with respect to other biometrics recognition techniques, such as universal acceptance and non-invasiveness. Moreover, 3D face imaging addresses some limitations of its 2D counterpart, like pose and luminance variation, while opening-up new horizons for enhancing the reliability of face-based identification systems [5]. This trend has been further fueled by the advances of 3D scanning technology, which provides now 3D textured scans encompassing aligned shape and photometric data.

In this paper, after a brief literature review, an explanation of Mesh-LBP framework is given (Sect. 2), then an outline of the proposed approach (Sect. 3.1), analyzing its potentialities; finally, some preliminary results are presented (Sect. 4) to support our proposal.

## 1.1 Related Works

The state of the art is plenty of 3D face recognition approaches, making impossible to analyze all of them. Instead, we are going to present the works that guided our decisions categorizing them into three categories.

First we have approaches that base their strength in the local description given by *Fiducial Points*. Such methods use local representation of a face natively supporting partial matching, and in the last years are gaining credit in the community. In fact, a face can be described as a whole (*global representation*), or as combination of *local* partitions. Each partition, or region, is represented by a descriptor [25], and the combination of such descriptors is the representation of the face. Using fiducial points, it is also possible to get a face matching that handle face expressions distortions; it is in fact possible isolate and discard regions that are highly affected by such expression deformation, mouth and eyebrows above all. One of the first proposed approach [16], presented a *keypoints* detector based on SIFT [12]. However, it did not account partial scan and face rotation. Later on [14, 19], presented a SIFT based method modeled to work on mesh manifold instead of standard flat images. That new born mesh-SIFT has been used in [9] together with the Sparse Representation based Classifier (SRC) [24] to boost the keypoint matching.

As second category is composed by all the Local Binary Pattern based approaches. LBP has been proposed in [2] as a 2D descriptor that well performed in texture retrieval problems. Given its success it has been applied to face recognition problem in [7], and later in 3D face recognition. In fact, LBP is now widely used on depth images [8, 11], performing very well both from precision than performance perspectives. Moreover, LBP's versatility allowed building several variants. In [18] it has been introduced the Local Normal Binary Pattern (LNBP) that uses normals angle instead of depth values. 3D-LBP [20] works on a mesh computing the code using two kinds of values, one is the depth values and the other is the angle between normals of vertex of the mesh. Such approach however, requires an elaborated processing on the mesh in order to obtain the neighborhood of a central vertex. Moreover, 3D-LBP does not support multiple scale resolution like other previous LBP variants.

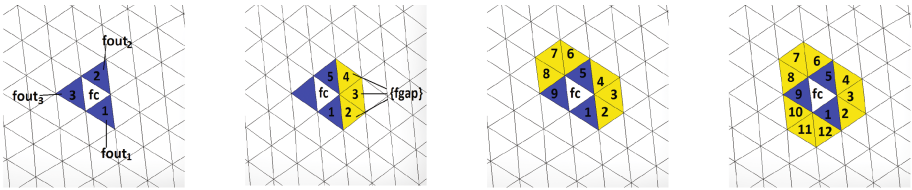
Finally, there is the group of Multi-modal 2D-3D approaches. Multi-modal solutions aim to combine different processing paths, usually 2D and 3D, into a single framework in order to overcome criticisms of individual approaches. In [6] Principal Component Analysis (PCA) is applied to depth images and standard images separately, then the outcomes are combined to get the final result. In [13] Iterative Closest Point (ICP) is used to register the 3D face model, and combined with Linear Discriminant Analysis (LDA) applied to the 2D image to avoid illumination and pose variation problems. Finally, [15] performs face registration, to avoid pose variations, region segmentation, to account local geometry changes, a filtering of the scans using SIFT and 3D Spherical Face Representation (SFR), and then a region wise matching with the remaining faces focusing on region robust to expression distortions.

## 2 Mesh-LBP

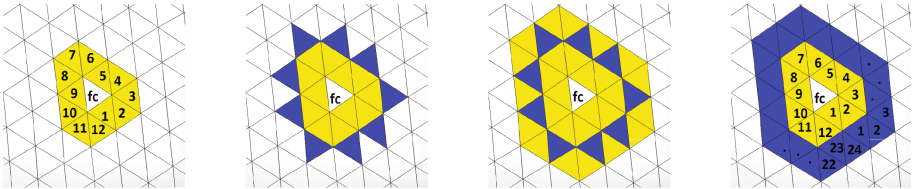
Our reference work generates Local Binary Patterns (LBP) over a real 3D support represented by triangular mesh manifold. In fact, LBP has been recently refiled in [21,23]. Since its definition [17] and its simplest application in face recognition [1], LBP is an 8-bit code obtained comparing pixels' values inside a  $3 \times 3$  window; the outcome of this comparison can be 1 or 0, whether the difference with neighbors' values is grater or less than zero. This pattern can be extended at different scales by changing the windows dimension and adopting circular neighborhoods at different radii.

In [23], the LBP idea has been broadened to 2D-mesh manifolds implementing power and elegance of LBP on a real 3D support.

Instead of pixels, the mesh is composed by facets. In order to obtain an ordered ring around a generic central facet  $f_c$ , the algorithm searches adjacent facets  $f_{out}$  and iteratively concatenate them as shown in Fig. 1. In such elegant way, it is now possible to generate a *ring-like* pattern at different radius scales. In fact, a new sequence of ordered  $f_{out}$  facets on the ring outer corner can be extracted allowing the ring construction procedure to be iterated (as shown in Fig. 2), generating concentric rings around the initial central facet  $f_c$ .



**Fig. 1.** From left to right the rings construction pipeline for mesh-LBP framework: starting from the adjacent facets, then the construction of the ring [22].



**Fig. 2.** Concentric rings construction sequence [22].

The concentric rings generated form an adequate structure for Local Binary Pattern computation. The mesh-LBP operator<sup>1</sup>, around a generic central facet  $f_c$ , is defined as:

<sup>1</sup> The LBP descriptor complied with the mesh manifold.

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

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

where parameters  $r$  and  $m$  control respectively the radial resolution and the azimuth quantization (see Fig. 2). Furthermore, a function  $\alpha(k)$  has been introduced to derive different LBP variants. In this work two variants have been studied:

- $\alpha_2(k) = 2^k$ , as originally suggested in [17];
- $\alpha_1(k) = 1$ , to obtain a simplified form that sums the binary pattern digits.

In Sect. 4 we will refer to these two functions with  $\alpha_2$  and  $\alpha_1$  respectively.  $h(f)$  function can be any desired feature; it can represent shape or appearance information, depending on the feature used. For example, as shape descriptor a geometric feature can be extracted from the mesh surface, such as *mean curvature* or *curvedness*, rather than *gray level* values to represent appearance information. Such photometric values come from 2D flat images, acquired with standard cameras, and subsequently projected over the mesh using a mapping scheme embedded in the mesh itself.

### 3 Fusion Schemes

In order to proceed, a brief description of Face Recognition pipeline has to be presented. Mesh-LBP framework presented in [23] can be summarized in 5 main steps:

**Features extraction**, since a mesh manifold is a structure, some features have to be extracted in order to describe the shape of the mesh surface.

**Local Binary Pattern computation**, applying Eq. 1 using the features beforehand extracted as input data.

**3D grid construction**, a grid is constructed and projected on the mesh manifold focusing on some stable region of the face.

**Histograms computation and concatenation**, for each point of the grid, a region is defined and an histogram computed inside it; the concatenation of all the region histograms form a signature for the examined face scan.

**Face matching**, checks differences between probe scan and a defined gallery.

As this framework operates at different levels over the same structure, it is possible to perform descriptors fusion at each level of the pipeline. In [22] has been shown how a simple score fusion, between geometric and photometric descriptors, fits or sometimes even outperforms the state of the art [4, 10]. Furthermore, it presents two fusion schemes at histograms computation level: one concatenates two different histograms derived from geometric and photometric features (*region histograms concatenation*); while the other one counts the co-occurrences of the

two features (*2D-histogram*). Such fusions show the potentiality of climbing the face matching pipeline to merge different descriptors.

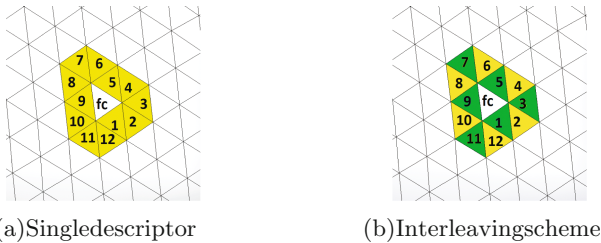
The idea proposed in this paper is to do a step forward and make the fusion at Mesh-LBP computation level. Even if the results displayed in [22] show high accuracy rate, the histograms fusion introduces an increment of the face descriptor size. In fact, the more simple *region histograms concatenation* doubles the original histogram size, while *2D histogram*, that adds one dimension to the standard histograms, sees a geometric increment of size. Instead, if the fusion is performed during, or even before, the mesh-LBP computation, it is possible to use both geometric and photometric data, keeping dimension and size equal to a single descriptor. Our aim is to produce a descriptor that holds the same size obtained with a single feature, but the information of two features (shape and appearance in our case<sup>2</sup>).

### 3.1 Early-Fusion

In this paper two kinds of early-fusion are presented. The first is a very basic fusion scheme that use logic operators (*AND*, *OR* and *XOR*). In order to get the LBP code, such operators have been added to the original formula:

$$meshLBP = \begin{cases} AND(s_g(x), s_p(x)) \\ OR(s_g(x), s_p(x)) \\ XOR(s_g(x), s_p(x)) \end{cases} \quad (2)$$

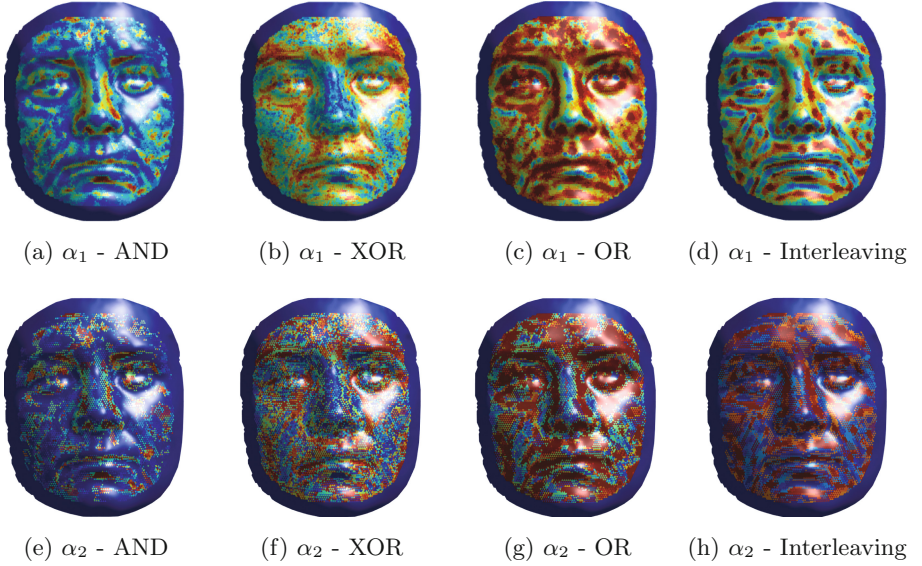
where  $s_g(x)$  and  $s_p(x)$  are computed as  $s(x)$  in Eq. 1 respectively for geometric and photometric information.



**Fig. 3.** Graphic comparison between standard ordering of a ring with single descriptor and the interleaving scheme with two descriptors.

In the second variant the mesh-LBP pattern is generated replacing the single feature function  $h(f)$ , shown in Eq. 1, with a combination of extracted features  $h_g(f)$  and  $h_p(f)$ . In particular, such new descriptor, named  $h_{g,p}(f)$ , is composed by interleaved values from geometric and photometric data,

<sup>2</sup> The framework allows to use any kind of features.



**Fig. 4.** Visual representation of *early-fusion* mesh-LBP code generated with radius  $r = 4$  and azimuth quantization  $m = 12$  in both  $\alpha_1$  and  $\alpha_2$  variants.

respectively  $d^g$  and  $d^p$  (Fig. 3). For example, for an azimuth quantization  $m = 12$ , the  $h_{g,p}(f)$  sequence would be

$$h_{g,p}(f) = d_1^g, d_2^p, d_3^g, d_4^p, d_5^g, d_6^p, d_7^g, d_8^p, d_9^g, d_{10}^p, d_{11}^g, d_{12}^p \quad (3)$$

Successively, the mesh-LBP code is obtained from the new combination  $h_{g,p}(f)$  applying Eq. 1 (Fig. 4).

From now on, these two variants will be referred to as *Logic Fusion* AND/OR/XOR and *Interleaving Fusion* respectively.

## 4 Experimentation

Experiments have been conducted on Bosphorus database [3], that is composed by 4666 scans of 105 subjects scanned in different poses, action units, and occlusion conditions. In addition to the shape structure, represented as mesh manifold, the database contains bitmap images of the scanned subject to provide appearance information as well. Since the aim of the project is to build a new LBP-like descriptor that can embed the strong points of a 3D environment, we did not focus on the matching algorithm. A naive template-matching-like method has been used, where each face probe descriptor is compared with a reference gallery using  $\chi^2$  distance.

Comparing our results with [22], the same features have been chosen to be merged. In particular in Table 1 we show results obtained using the *mean curvature* to represent shape information, and the *graylevel*, got from the bitmap mapped on the mesh surface, for the appearance.

**Table 1.** Overall outcomes of Bosphorus database showing the accuracy of *Logic* operators (AND, XOR, OR) and *Interleaving scheme* compared with [22] single descriptors *mean curvature* (H) and *graylevel* (GL).

	Accuracy	
	$\alpha_1$	$\alpha_2$
H	86.82%	90.70%
GL	90.89%	90.26%
Fusion <sub>1</sub> H+GL	93.84%	–
Fusion <sub>2</sub> H+GL	93.29%	92.60%
Logic AND H+GL	86.24%	89.14%
Logic XOR H+GL	77.30%	90.35%
Logic OR H+GL	85.99%	89.79%
<b>Interleaving H+GL</b>	<b>92.76%</b>	<b>93.03%</b>

**Table 2.** Histogram sizes (per region) for each variant reported in Table 1 in number of bins.

Hist size	
$\alpha_1$	$\alpha_2$
13	1125
13	136
169	–
26	1261
13	1125
13	1125
13	1125
<b>13</b>	<b>1125</b>

Results from logic fusions show an accuracy rate close to the original single descriptor. Even if the size of logic descriptor is equal to a single one, the outcomes are not satisfying: this scheme shows a decrease in its descriptive power respect to what has been achieved in our reference paper. In fact, logic operators seem to annihilate the mutual information provided by the couple of features.

Interleaving scheme, instead, preserves the descriptive power of both geometric and photometric information, outperforming single descriptor precision and above mentioned histograms fusions. In particular,  $\alpha_1$ , even if a bit lower in accuracy compared with *Fusion<sub>1</sub>* and *Fusion<sub>2</sub>* schemes, sees a drastic decrease of descriptor size (Table 2): half respect to *region histograms concatenation* schema, and even of the order of root square respect to the *2D-histogram* (13 times smaller).  $\alpha_2$ , instead, does not only keep the same size of single feature histogram, but also outperforms the region histograms concatenation fusion scheme.

The effectiveness of *Interleaving* early-fusion approach become clear if we think that 2D-histogram fusion scheme, cannot be computed for  $\alpha_2$ , that is the original LBP variant. In that case, in fact, the 2D-histogram would have had  $1125 \times 136 = 153000$  bins instead of the 1125 of our proposed fusion scheme.

## 5 Conclusion

In this paper a novel early level fusion approach for actual 3D face recognition has been presented. The proposed method exploits mesh manifold potentialities as support structure. In particular, we extended mesh-LBP, a framework that enables to generate LBP-like codes directly on a triangular mesh. Our aim is to fuse different features during, or even before, the LBP descriptor computation. For this purpose logic operators and interleaving schemes have been

used to generate a pattern comprehensive of photometric texture and geometric shape information. The experimentation, conducted on Bosphorus database, shows promising results, raising the curtains on the potentiality held by early feature fusion among real 3D support, like mesh manifolds. It is in fact now possible to consider more refined early-fusion techniques directly employed on a mesh manifold. In this manner, we can hold the descriptive power of two, or even more, descriptor, improving performances without increasing the descriptor size.

## References

1. Ahonen, T., Hadid, A., Pietikäinen, M.: Face recognition with local binary patterns. In: European Conference on Computer Vision, Prague, pp. 469–481, May 2004
2. Ahonen, T., Hadid, A., Pietikäinen, M.: Face recognition with local binary patterns. In: Pajdla, T., Matas, J. (eds.) ECCV 2004. LNCS, vol. 3021, pp. 469–481. Springer, Heidelberg (2004). doi:[10.1007/978-3-540-24670-1\\_36](https://doi.org/10.1007/978-3-540-24670-1_36)
3. Alyüz, N., Gökberk, B., Akarun, L.: 3D face recognition system for expression and occlusion invariance. In: IEEE International Conference on Biometrics: Theory, Applications, and Systems, Washington, DC, pp. 1–7, September 2008
4. Berretti, S., Werghe, N., Del Bimbo, A., Pala, P.: Matching 3D face scans using interest points and local histogram descriptors. *Comput. Graph.* **37**(5), 509–525 (2013)
5. Bowyer, K.W., Chang, K.I., Flynn, P.J.: A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition. *Comput. Vis. Image Underst.* **101**(1), 1–15 (2006)
6. Chang, K., Bowyer, K., Flynn, P.: An evaluation of multimodal 2-D and 3-D face biometrics. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(4), 619–624 (2005)
7. Huang, D., Shan, C., Ardabilian, M., Wang, Y., Chen, L.: Local binary patterns and its application to facial image analysis: a survey. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **41**(6), 765–781 (2011)
8. Huang, Y., Wang, Y., Tan, T.: Combining statistics of geometrical and correlative features for 3D face recognition. In: British Machine Vision Conference, Edinburgh, pp. 879–888, September 2006
9. Li, H., Chen, L., Huang, D., Wang, Y., Morvan, J.: Towards 3D face recognition in the real: a registration-free approach using fine-grained matching of 3D keypoint descriptors. *Int. J. Comput. Vis.* **113**(2), 128–142 (2015)
10. Li, H., Huang, D., Lemaire, P., Morvan, J.M., Chen, L.: Expression robust 3D face recognition via mesh-based histograms of multiple order surface differential quantities. In: IEEE International Conference on Image Processing, pp. 3053–3056, September 2011
11. Li, S., Zhao, C., Ao, M., Lei, Z.: Learning to fuse 3D+2D based face recognition at both feature and decision levels. In: International Workshop on Analysis and Modeling of Faces and Gestures, Beijing, pp. 44–54, October 2005
12. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **60**(2), 91–110 (2004)
13. Lu, X., Jain, A.K.: Deformation modeling for robust 3D face matching. In: IEEE International Conference on Computer Vision and Pattern Recognition, New York, pp. 1377–1383, June 2006

14. Maes, C., Fabry, T., Keustermans, J., Smeets, D., Suetens, P., Vandermeulen, D.: Feature detection on 3D face surfaces for pose normalisation and recognition. In: 2010 Fourth IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS), pp. 1–6. IEEE (2010)
15. Mian, A.S., Bennamoun, M., Owens, R.: An efficient multimodal 2D-3D hybrid approach to automatic face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(11), 1927–1943 (2007)
16. Mian, A.S., Bennamoun, M., Owens, R.: Keypoint detection and local feature matching for textured 3D face recognition. *Int. J. Comput. Vis.* **79**(1), 1–12 (2008)
17. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distribution. *Pattern Recognit.* **29**(1), 51–59 (1996)
18. Sandbach, G., Zafeiriou, S., Pantic, M.: Local normal binary patterns for 3D facial action unit detection. In: IEEE International Conference on Image Processing, Orlando, pp. 1813–1816, September 2012
19. Smeets, D., Keustermans, J., Vandermeulen, D., Suetens, P.: meshSIFT: local surface features for 3D face recognition under expression variations and partial data. *Comput. Vis. Image Underst.* **117**(2), 158–169 (2013)
20. Tang, H., Yin, B., Sun, Y., Hu, Y.: 3D face recognition using local binary patterns. *Signal Process.* **93**(8), 2190–2198 (2013)
21. Werghi, N., Berretti, S., Del Bimbo, A., Pala, P.: The mesh-LBP: computing local binary patterns on discrete manifolds. In: ICCV International Workshop on 3D Representation and Recognition, Sydney, pp. 562–569, December 2013
22. Werghi, N., Tortorici, C., Berretti, S., Del Bimbo, A.: Representing 3D texture on mesh manifolds for retrieval and recognition applications. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, pp. 2521–2530, June 2015
23. Werghi, N., Tortorici, C., Berretti, S., del Bimbo, A.: Local binary patterns on triangular meshes: concept and applications. *Comput. Vis. Image Underst.* **139**, 161–177 (2015)
24. Wright, J., Yang, A., Ganesh, A., Sastry, S., Ma, Y.: Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(2), 210–227 (2009)
25. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: a literature survey. *ACM Comput. Surv. (CSUR)* **35**(4), 399–458 (2003)