MESH LBP FEATURES FOR 3D CONSTRAINED LOCAL MODEL

Marwa Chendeb El Rai, Naoufel Werghi, Claudio Tortorici, Hassan Al-Muhairi, Habiba Al Safar

Khalifa University of Science, Technology and Research
Electrical and Computer Engineering Department
United Arab Emirates

ABSTRACT

We propose an automatic facial landmarks detection in 3D mesh manifold. The method is based on 3D Constrained Local Model (CLM) which learns both global variations in 3D face scan and local ones around every vertex landmark. Differently from the other approaches of CLM, our contribution is a full 3D mesh. The framework exploits the intrinsic 3D features around the mesh vertices by utilizing histogram-based mesh Local Binary Patterns (mesh-LBP). The experiments are conducted on publicly available 3D face scans Bosphorus database.

***Index Terms*—** Constrained local model, Histograms-based mesh LBP, 3D Facial landmarks, Landmarks detection

1. INTRODUCTION

The human face contains a rich source of information used for example for identification, communication, face alignment, registration, face recognition, facial expression and for genetic screening [1, 2, 3, 4, 5]. The landmarking is a crucial initial step for most of the face processing and analysis methods. It consists in locating facial feature points such as nose tip, inner eye corners, outer eye corners, mouth corners, etc. The Active Shape Models (ASMs) [6], Active Appearance Models (AAMs) [7], Constrained Local Models (CLM) [2, 8] have shown a great attention in the literature of facial feature detection. CLM is a model closely related to AAM that has shown improved robustness and accuracy. CLM doesn't cover the whole face, it uses local descriptors located in specific feature points that are characterized jointly by their own properties in terms of local feature information and their structural relationships which result from the global face morphology [9]. 2D images are sensitive to variations caused by expressions, illumination, pose, occlusions while 3D data is robust to these variations and can be inspected from any viewpoint. The advanced techniques in 3D capturing tools, laser scanners and high-speed stereo machines have been increased. There have been many attempts of varying success at tackling the problem of face alignment. The landmark detector ($known$ as patch expert in CLM) has to be robust to the variations of facial data and discriminative, to distinguish between different anatomical landmarks. In the case of CLM, many features like raw intensity, Linear Binary pattern and Gabor wavelets are used in the training of local detectors especially in the case of 2D images [1, 4]. The main contribution of our work comes from the use of the robust mesh-LBP features recently developed in [10]. In this paper, we present the framework of CLM that utilizes histograms based on mesh LBP. It takes the full advantage of mesh and provide more discriminative local information. This paper is organized as follows: related work in 3D CLM is presented in section 2. The framework of mesh-LBP is described in section 3. An overview of patch experts and CLM fitting is shown in section 4. The last one concerns the experimentation done using samples of Bosphorus database.

Thanks to Khalifa University of Science, Technology and Research for funding.

2. RELATED WORK

With regard to 3D facial annotations, most if not all the approaches have been developed in the concept of the standard depth images where the z coordinate is mapped to a gray-level value. This format allowed a straightforward employment of 2D local features as the intensity information, Linear Binary Patterns (LBP), Gabor wavelets, etc. [4, 8] and [1]. CLM-Z is a recent 3D CLM proposed by [4] for face alignment. This method combines both depth and intensity data for response maps. However, the geometric information is not fully exploited since the depth image is used in almost the same manner as the intensity image. [11] presented a robust 3D CLM based on exploiting the geometric information in depth data by using the histogram-based 3D facial geometry features HONV (Histogram of Oriented Normal Vectors) and LNBPs (Local Normal Binary Patterns). In [5] the input data is a mesh, after generating the 3D Point distribution model, the 3D feature points are transformed into 2D and then the CLM proposed was applied. LBP descriptor has attracted great interest for the analysis of 2D images, mainly for its simple and efficient computation and for the effective results that can be achieved relying on the LBP theory. Recently, various attempts have been done for extending the LBP framework to the case of 3D meshes, but none of them succeeded in addressing all the issues posed by the need for a simple and effective processing directly performed on a mesh-manifold. Indeed, existing solutions address the LBP extraction on 3D meshes by resorting to the easier 2D case, through the projection of 3D meshes on 2D depth maps. Werghi et al. [10] addressed an original framework to compute the mesh LBP around the vertices. The histograms calculated using the mesh LBP are used in our work as local descriptors. Therefore, the novelty of our work comes from the use of the 3D mesh LBP features during the training and fitting phases of CLM. We thoroughly evaluated the proposed approach on publicly available Bosphorus datasets.

3. 3D MESH LBP FEATURES

Mesh LBP is a framework developed recently by Werghi et al. [10] for designing and extracting local binary patterns directly from a 2D mesh-manifold. The construction of LBP-like patterns on a mesh, requires a scheme for constructing rings of facets around a central one related to the vertex landmark and for traversing them in an ordered fashion. Consider $S = \langle h, V \rangle$ the triangular mesh representation of an open or closed surface, where V and F are, respectively, the sets of vertices and facets of the mesh. Each vertex has a valence of six, consider the facets that have an edge on the contour Fig.1(a). (F_{out}) are the facets which seem pointing outside the contour. F_{in} are the facets that are one to one adjacent to the F_{out} facets and which are located inside the convex contour. Each facet F_{in}, shares with its corresponding F_{out} facet an edge located on the convex contour. Suppose that the F_{out} facets are initially ordered in a circular fashion across the contour. Given that initial arrangement, the gap between each pair of consecutive F_{out} facets, is bridged. Thus, the sequence of adjacent facets, located between the two consecutive F_{out} facets and which share their common vertex (the vertex on the contour). These facets are called F_{gap} (see Fig.2(b)). By iterating the process of bridging the gap between two consecutive F_{out} facets with the F_{gap} facets results in a ring of facets that are ordered in a circular fashion (see Fig.1(c)). The resulting arrangement of the ring facets inherits the same direction (clock-wise or anti-clockwise) of the initial sequence of F_{out} facets (for more details about this method please refer to [10]). In this work, the F_{out} set includes the three facets adjacent to the central one, and the obtained ring is composed of 12 ordered facets (i.e., the three F_{out} facets, plus the nine F_{gap} facets bridging the gap between consecutive F_{out} facets), as shown in Fig.1(d). Let $h(f) = S- > R$ be a scalar function defined on the mesh S (e.g., photometric data or curvature). The circular ordering of the facets obtained with Ordered Ring Facets (ORF) allows to derive a binary pattern (i.e., sequence of 0 and 1 digits) from it, and thus to compute a local binary operator in the same way as in the standard LBP. The basic mesh-LBP operator at a central facet f_c by thresholding its ordered ring neighbourhood constituted by the 12 facets in the ORF is defined by the following equation:

$$meshLBP(f_c) = \sum_{k=0}^{11} s(h(f_k) - h(f_c)).\alpha(k) \quad (1)$$

$$s(x) = \{ \begin{array}{c} x \geq 0 \\ x \prec 0 \end{array} \quad (2)$$

where $\alpha(k)$ is a weighting function. Different definitions of the function $\alpha(k)$ permit to obtain different binary patterns, and thus different mesh-LBP values can be derived from the central facet and its ring neighborhood. For example, with $\alpha(k) = 2^k$ the basic LBP operator firstly suggested by Ojala et al. [12] is obtained; for $\alpha(k) = 1$, the sum of the digits of the pattern is computed (i.e., the number of digits equal to 1). Given a multi-ring constructed around a central facet f_c, a multi-resolution mesh-LBP operator is computed as [10]:

$$meshLBP_m^r(f_c) = \sum_{k=0}^{m-1} s(h(f_k^r) - h(f_c)).\alpha(k) \quad (3)$$

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 of the operator. In principle, any predefined number of samples per ring can be used. In this work, we considered, in almost all the cases, a number of samples per ring $m = 12$.

4. 3D CLM WITH HISTOGRAM BASED MESH LBP

The shape model is defined by a 3D mesh and in particular the 3D vertex locations of the mesh, called landmark points. Consider the shape as the coordinates of 3D vertices of the M landmark points: $\mathbf{x} = (x_1, y_1, z_1, ..., x_M, y_M, z_M)^t$. Apart from the global transformations: scale, rotation, and translation, $\mathbf{x}$ can be approximated by means of the linear principal component analysis (PCA). Our model in 3D mesh CLM can be described by the vector parameters $\mathbf{p} = [s, \mathbf{R}, \mathbf{q}, \mathbf{t}]$ that can be varied to obtain variance instances of the model: scaling factor s, rotation **R**, translation **t** and a vector described the non-rigid variation of the shape **t**. The point distribution model (PDM) is used for a parametric representation of the rigid and non-rigid shape variation as follows:

$$x_j(\mathbf{p}) = s\mathbf{R}(\bar{x}_j + \phi_j \mathbf{q}) + \mathbf{t}) \quad (4)$$

x_j is the 3D coorindates of the j^{th} feature vertex in a given face scan. ϕ_j is the j^{th} eigenvector obtained from the training set that describes the linear variations of non-rigid shape of this feature vertex.

4.1. PATCH EXPERTS

In the training phase, the histograms of mesh LBP are computed around every vertex v. The positive and negative samples selected from training data are extracted. The linear classifier Support Vector Machine (SVM) proposed in [13] is used due to its rapid computations. The patch model is described as a group of linear classifiers $L = \{w_j, b_j\}_{j=1}^n$, where w_j and b_j are the weights and bias of the j^{th} patch expert. The probability for the j^{th} landmark being correctly aligned ($l_j = 1$) at vertex v of mesh *M* is modeled by logistic function [8]:

$$p(l_j = 1/v, M) = (\frac{1}{e^{-\alpha w_j C(v_j;M)+b_j}}) \quad (5)$$

where α is the logistic regressor intercept, $C(v; M)$ denotes the histogram of the mesh LBP local descriptor computed from a set of rings around the vertex v_j as described in section 3.

4.2. SEARCHING WITH CLM FITTING

For the fitting process, an exhaustive search around the present estimate of spatial location of 3D PDM leading to a response map around every facial landmark is firstly performed, and then iteratively updating the model parameters to minimize errors until a convergence metric is reached. The state of art Regularized Landmark Mean Shift (RLMS) [8] was used. CLM fitting aims to estimate and update the PDM parameters **p**, that jointly minimizes the misalignment error over all landmarks [8]. It can be interpreted as maximizing the likelihood of the model parameters that all its points are aligned with their corresponding vertices on the mesh. The independence between the detection for each landmark vertex, permits the probabilistic interpretation to take the form:

$$p(\mathbf{p}|\{l_j = 1\}_{j=1}^{n}, M) = argmax\{p(\mathbf{p}) \prod_{j=1}^{n} p(l_j = 1/\mathbf{v}_j, M)\} \quad (6)$$

Regularized Landmark Mean-Shift (RLMS) [8] is used for the fitting step. By assuming a homoscedastic isotropic Gaussian kernel density estimate in a set of fixed locations $\{\psi_j\}_{j=1}^{n}$ for every part j. The equation can be reformulated as:

$$\mathbf{p} = argmax\{p(\mathbf{p}) \prod_{j=1}^{n} \sum_{y_j \in \psi_j} p(l_j = 1|y_j, M) N(\mathbf{v}_j(\mathbf{p})|y_j, \rho\mathbf{I}\} \quad (7)$$

In 3D PDM, the non rigid shape parameters **q** is assumed to follow Gaussian distribution which permits to the regularisation term $p(\mathbf{p})$. The above optimization problem was solved using Expectation-Maximization (EM) algorithm. The Expectation step concerns the computation of $p(y_j|l_j = 1, v_i, I)$, given the parameters p, while the Maximization step involves the minimization of: which can be solved using a Gauss-Newton optimization [8].

5. EXPERIMENTS

To validate our 3D CLM based histograms of mesh-LBP approach, the publicly database Bosphorus is used [14]. Fig. 2 shows a mesh with the landmarks. The data is split into two groups: training (70 face scans) and testing (30 face scans). 22 landmarks are annotated manually to generate the shape and patch models. However only 11 of them have correspondents in our landmarks sets (e.g. have similar locations). Therefore we used this subset as groundtruth data (can be seen in Fig. 4). To remove the effect of rigid transformation, all training shapes are aligned by Procrustes analysis. After making the mean of PCA a guess of the initial position, the ICP (iterative closest point) based rigid registration (rotation and translation) method is applied to create the correspondence between the template mesh and the current 3D estimated landmarks. This step is essential to find the closest vertices on the mesh indispensable to calculate the response map $C(v_j, M)$ based on the mesh LBP around every feature vertex j. In our experiment, we have used 7 rings and 13 samples per ring (i.e., histograms with 7 rows and 13 columns). Each histogram bin cumulates the frequency of a mesh-LBP pattern computed for all the facets of a sample surface. The performance was evaluated using two metrics, namely, the success rate (SR) and the absolute distance error (ADE). The success rate is defined as the percentage of the successful detection of landmark. Here we consider that a landmark is detected successfully if the distance to its ground truth counterpart is less than 7 mm. The absolute distance error is the Euclidean distance between the detected landmark and its ground truth counterpart. In the experiments we considered as scalar function $h(f)$ (equation 1) a photometric function (the gray level function G) and two shape functions, namely, the mean curvature H and Curvedeness C. Fig. 3 depicts the SR and the RMS of the ADE obtained with unimodal variant and corresponding to each of G, H and C functions. We note that the shape functions performs better than the texture function across all the landmarks and achieves 100% SR for some of them.

6. REFERENCES

[1] Q. Zhao, K. Okada, K. Rosenbaum, L. Kehoe, D. Z. R. Sze, M. Summar, and M. Linguraru, "Digital facial dysmorphology for genetic screening: Hierarchical constrained local model using ica," *Medical Imge Analysis*, vol. 18, pp. 699–710, 2014.

[2] D. Cristinacce and T. Cootes, "Feature detection and tracking with constrained local models," *BVMC*, vol. 18, 2006.

[3] M. Yu and B. Tiddeman, "Facial feature detection and tracking with a 3d constrained local model," *18th International Conference in Central Europe on Computer Graphics*, pp. 181–188, 2010.

[4] T. Baltruaitis, P. Robinson, and L. Morency, "3d constrained local model for rigid and non-rigid facial tracking," *IEEE Computer Vision and Pattern Recognition*, 2012.

[5] A. Lorincz, L. Jeni, Z. Szab, and J. Cohny, "Emotional expression classification using time-series kernels," *IEE conference Computer Vision and Pattern Recognition Workshops*, 2013.

[6] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models - their training and application," *Pattern Recognition*, vol. 38, pp. 38–59, 1995.

[7] T. F. Cootes, G. J. Edwards, , and C. J. Taylor, "Active appearance models," *European Conference on Computer Vision*, vol. 2, pp. 484–498, 1998.

[8] J. Saragih, S. Lucey, and J. Cohn, “Deformable model fitting by regularized landmark mean-shift,” *International Journal of Computer Vision*, vol. 91, pp. 200–215, 2011.

[9] X. Zhao, E. Dellandra, L. Chen, and I. Kakadiaris, “Digital facial dysmorphology for genetic screening: Hierarchical constrained local model using ica,” *IEEE Transcations on systems, man and cybernetics*, vol. 41, pp. 1417–1428, 2012.

[10] N. W. S. Berretti and A. del Bimbo, “The mesh-lbp: a framework for extracting local binary patterns from discrete manifolds,” *IEEE Transactions on Image Processing*, 2015.

[11] Sh.Cheng, S. Zafeiriou, A. Asthana, and M. Pantic, “3d facial geometric features for constrained local model,” *IEEE International Conference on Image Processing*, October 2014.

[12] T. Ojala, M. Pietikainen, , and D. Harwood, “A comparative study of texture measures with classification based on featured distribution,” *Pattern Recognition*, vol. 29, pp. 51–59, January 1996.

[13] C. Chang and J. L. Ch, “Libsvm: A library for support vector machines,” *CM Transactions on Intelligent Systems and Technology*, vol. 3, pp. 27–127, 2011.

[14] bosphorus.ee.boun.edu.tr

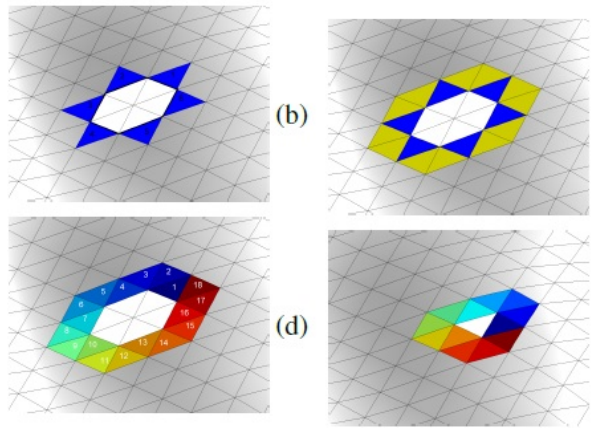

Fig. 1. Construction of an ordered ring: (a) Initial F_{out} facets on a convex contour; (b) Bridging the gap between the pairs of consecutive F_{out} facets with the F_{gap} facets; (c) The obtained ordered ring; (d) Ordered ring constructed around a central facet [10].

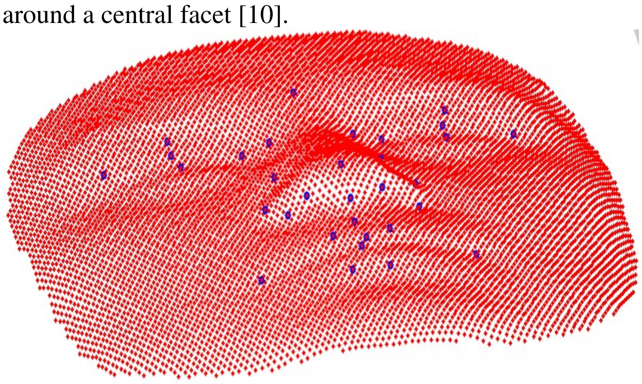

Fig. 2. Mesh of a sample Bosphorus database with 32 landmarks obtained after applying 3D CLM based on mesh LBP.

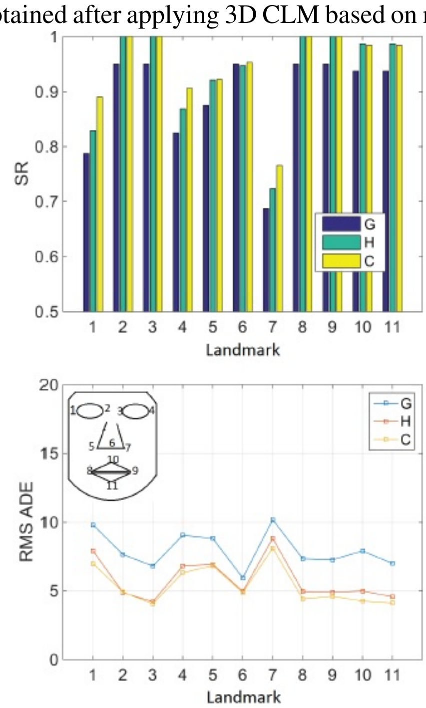

Fig. 3. SR and RMS ADE obtained with the unimodal variants.