EXPRESSION ROBUST 3D FACE RECOGNITION VIA MESH-BASED HISTOGRAMS OF MULTIPLE ORDER SURFACE DIFFERENTIAL QUANTITIES

Huibin Li[1,2], *Di Huang*[1,2], *Pierre Lemaire*[1,2], *Jean-Marie Morvan*[1,3,4], *Liming Chen*[1,2]

[1]Université de Lyon, CNRS
[2]Ecole Centrale de Lyon, LIRIS UMR5205, F-69134, Lyon, France
[3]Université Lyon 1, Institut Camille Jordan,
43 blvd. du 11 Nov. 1918, F-69622 Villeurbanne - Cedex, France
[4]King Abdullah University of Science and Technology, GMSV Research Center,
Bldg 1, Thuwal 23955-6900, Saudi Arabia

ABSTRACT

This paper presents a mesh-based approach for 3D face recognition using a novel local shape descriptor and a SIFT-like matching process. Both maximum and minimum curvatures estimated in the 3D Gaussian scale space are employed to detect salient points. To comprehensively characterize 3D facial surfaces and their variations, we calculate weighted statistical distributions of multiple order surface differential quantities, including histogram of mesh gradient (HoG), histogram of shape index (HoS) and histogram of gradient of shape index (HoGS) within a local neighborhood of each salient point. The subsequent matching step then robustly associates corresponding points of two facial surfaces, leading to much more matched points between different scans of a same person than the ones of different persons. Experimental results on the Bosphorus dataset highlight the effectiveness of the proposed method and its robustness to facial expression variations.

Index Terms— mesh-based 3D face recognition, histograms of multiple order surface differential quantities, 3D shape descriptor

1. INTRODUCTION

Face is potentially one of the best biometrics for people identification related applications, since it is non-intrusive, contactless and socially well accepted. The past several decades have witnessed tremendous efforts firstly focused on 2D face images [1] and more recently on 3D face scans [2]. Despite the great progress achieved so far in the field [1], 2D images are still not reliable enough [3], especially in the presence of pose and lighting changes [4]. Along with the development of 3D imaging systems, 2.5D or 3D scans have emerged as a major solution to deal with these unsolved issues in 2D face recognition, i.e. pose and illumination variations. Meanwhile, although 3D face scans capture accurate facial surface shape, thereby theoretically reputed to be robust to lighting variations, they are likely to be more sensitive to facial expression variations.

Many methods have been proposed to address the problem caused by expression changes. They can be roughly categorized into three main streams, i.e. region based, model based and learning based. The first category claims that facial distortions due to expressions do not have impact on the entire surface, and proposes to segment a facial surface into relative rigid and non-rigid regions. The rigid regions were generally adopted for an improved matching result [5] [6]. Meanwhile, automatic segmentation of facial surface into rigid and mimic regions is still problematic [2]. The second stream builds a generic face model and for a given non-neutral face, a virtual face is generated by this model with reduced facial expression effects. Recognition is then performed on virtual 3D faces [7] [8] [9]. These techniques improve the performance to some extent but they are rather computationally expensive. The last stream proposes to learn the distributions of intra-class and inter-class variations across expression changes [10] [11], and generally leads to better performance. The downside is that it requires a large training dataset with at least several expressive face models for each subject.

In this paper, we make the assumption that, when a facial expression occurs, there are always some small local areas that vary slightly or keep invariant as compared to the neutral expression, and these local areas can be found in both relative rigid and elastic facial regions. Once located and characterized, these local regions can be used to achieve 3D face recognition, which is robust to facial expressions and partial occlusion through a proper matching process.

This paper proposes to characterize these local regions as salient points, which are local extrema in terms of maximum and minimum curvatures within a 3D Gaussian scale space. Once located, the local region around each salient point is then described by histograms of multiple order surface differential quantities, including histograms of mesh gradient (HoG, 1^{st} order), shape index (HoS, 2^{nd} order) and gradient of shape index (HoGS, 3^{rd} order), to comprehensively describe the local facial surface. The descriptors of detected

local regions are further used in local matching for 3D face recognition.

The whole framework is thus a region-based approach but in contrast to the previous works, all these local regions are not only in static regions but can be distributed over the entire face including elastic regions. The proposed framework is similar to SIFT [12] and their extensions in 3D [13] [14]. Meanwhile, our approach differs from the previous ones by the use of surface differential quantities both in salient point localization and their geometric description. As salient point detection and their description are directly operated on 3D meshes, the proposed approach is pose independent, thus avoids costly alignment which is widely required by facial range image based methods. Moreover, the use of local matching also endows the proposed approach with some tolerance to occlusions. The effectiveness of the proposed approach was demonstrated on the Bosphorus dataset.

The remainder of this paper is organized as follows: multiple order differential quantities estimated on surface is introduced in section 2, and section 3 presents the proposed shape descriptor. Section 4 describes the matching scheme. Experimental results are discussed in section 5. Section 6 concludes the paper.

2. ESTIMATING MULTIPLE ORDER DIFFERENTIAL QUANTITIES ON TRIANGULAR MESHES BY LOCAL SURFACE FITTING

There are three main approaches to calculate curvature on triangular meshes [15]: a. local fitting; b. discrete estimation of curvature directly on triangular meshes; c. estimation of curvature tensor. In this paper, we make use of the local cubic-order fitting method as in [16] which has better behaviour than discrete estimation method for facial expression analysis [17].

For each vertex p of a triangular mesh, this method first defines a local 3D coordinate frame with its origin at the vertex and z axis along the normal vector of the vertex. The vertex normal can be computed as the average of normal vectors of the faces adjacent to the vertex. Given two orthogonal axes, x and y, randomly chosen in the tangent plane perpendicular to the normal vector, vertices in the neighborhood (two-ring) of the vertex p are then transformed and rotated to the local coordinate system, in which a cubic polynomial function:

$$z(x,y) = \frac{A}{2}x^2 + Bxy + \frac{C}{2}y^2 + Dx^3 + Ex^2y + Fxy^2 + Gy^3 \quad (1)$$

is approximated by the coordinates of the vertices within the local neighborhood. And it's normal :

$$(z_x, z_y, -1) = (Ax + By + 3Dx^2 + 2Exy + Fy^2, + Bx + Cy + Ex^2 + 2Fxy + 3Gy^2, -1) \quad (2)$$

is approximated by the normal vectors of the vertices within the local neighborhood. By using least-square fitting method to solve the approximated equations (1) and (2), we can get the local fitting function $z(x,y)$. The maximum curvature κ_{max} and minimum curvature κ_{min} ($\kappa_{max} \geq \kappa_{min}$) can be estimated as the eigenvaules of the Weingarten matrix (shape operator). Then the shape index can be estimated as:

$$\mathbf{S} = \frac{1}{2} - \frac{1}{\pi} arctan(\frac{\kappa_{max} + \kappa_{min}}{\kappa_{max} - \kappa_{min}}) \quad (3)$$

After the local coordinate system is transformed to the original global one, we estimate the surface gradient by the normal vector of each point. Let $z(x,y)$ be the fitting function of a local surface patch at a point p, the normal vector $n_p = (n_x, n_y, n_z)^T$ can be written as $(-\frac{n_x}{n_z}, -\frac{n_y}{n_z}, -1)^T$. Thus, the surface gradient direction θ equals to $arctan(\frac{n_y}{n_x})$, and gradient magnitude can be estimated as follows:

$$||\nabla z(x,y)|| = \sqrt{{z_x}^2 + {z_y}^2} = \sqrt{(-\frac{n_x}{n_z})^2 + (-\frac{n_y}{n_z})^2} \quad (4)$$

We compute the gradient of shape index instead of derivative of curvature as the 3rd order differential quantity, which is a quantitative measurement of the shape variation of a surface. Shape index can be taken as a scalar field defined on 3D discrete surface as proposed in [13]. For a triangular mesh, the gradient of scale field $\mathbf{S}(p)$ at p defined in its local tangent plane is generally approximated by solving an optimization problem using the finite element method (FEM) [18]. Let $q_i \in N_1(p)$ be the one-ring neighborhood of vertex p, the gradient $\nabla\mathbf{S}(p)$ can be estimated by minimizing the following error:

$$\nabla\mathbf{S}(p) = argmin \sum_{q_i \in N_1(p)} |\nabla\mathbf{S}(p)^T \mathbf{P}(\overrightarrow{pq_i}) - \frac{\mathbf{S}(p) - \mathbf{S}(q_i)}{||p - q_i||}| \quad (5)$$

where $\mathbf{P}(\overrightarrow{pq_i})$ is the projected unit vector of $\overrightarrow{pq_i}$ in the local tangent plane T_pS.

3. MULTIPLE ORDER DIFFERENTIAL QUANTITIES BASED LOCAL DESCRIPTORS

Similar to the mesh-SIFT algorithm [14], salient points are detected in the 3D Gaussian scale space. The original mesh is firstly smoothed by a Gaussian filter with different values of standard deviation σ. Curvature is then estimated for each vertex on the original and smoothed meshes. Given a vertex p, we compute the difference of estimated curvatures of the same vertex for each pair of adjacent scales. Finally, this difference of curvature at a vertex on a given scale is compared to the differences of its one-ring neighbors on its own scale as well as the corresponding ones on upper and lower scales. The vertex is selected as a salient point only if its value is a local extrema within the neighborhood. In this paper, we propose to make use of maximum and minimum curvatures instead of the mean curvature as in [14], since using both the maximum and minimum curvatures characterizes the shape of a local region more accurately than only using their average value i.e. mean curvature. Fig.1 shows the detected salient points by maximum and minimum curvature respectively.

We extract multiple order surface differential quantities based local descriptor at the scale where salient points are detected. Around each salient point, a local geodesic disk with a radius R is considered. As the same scheme in section 2, we

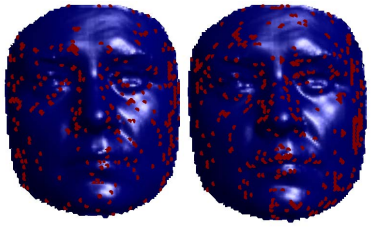

Fig. 1. Salient points detected by κ_{max} (left) and κ_{min} (right).

firstly transform the local patch points to the local coordinate system, where the salient point is the origin, and its normal vector is along the positive z axis. In order to make the descriptor invariant to rotation, each salient point is assigned one or several canonical orientations according to the dominant direction(s) of gradients in the local tangent plane with 360 bins. Once the canonical orientations are assigned, the local coordinate system rotates in the local tangent plane, making each canonical orientation as new x axis. New y axis can be computed by cross product of z and x. In this new local coordinate system, we project all the neighbors of a salient point to its tangent plane. Eight projected points along to eight quantized directions starting from canonical orientation with a distance of r_1 to the salient point are fixed. Nine circles centered at the salient point and its eight neighbors with a radius r_2 can be further located. Fig.2 shows this arrangement. In each circle, we calculate three histograms including surface gradient (HoG), shape index (HoS) and gradient of shape index (HoGS). For HoG and HoGS, we compute histogram of gradient angle weighted by gradient magnitude. This histogram is with 8 bins representing 8 main orientations ranging form 0 to 360 degree. For HoS, the values of shape index ranging from 0 to 1 are also quantized to 8 bins. Then, all the values of histograms are weighted by Gaussian with the Euclidian distance to the center point of the circle as the standard deviation. Every histogram is then normalized, and the feature vectors of the three histograms are formed as follows:

$$HoG = (hog_1, hog_2, \ldots, hog_9) \tag{6}$$

$$HoS = (hos_1, hos_2, \ldots, hos_9) \tag{7}$$

$$HoGS = (hogs_1, hogs_2, \ldots, hogs_9) \tag{8}$$

Finally, we concatenate HoG, HoS and HoGS as:

$$HoG + HoS + HoGS = (hog_1 hos_1 hogs_1, hog_2 hos_2 hogs_2, \ldots, hog_9 hos_9 hogs_9) \tag{9}$$

which is normalized and used as local descriptor for the following facial surface matching.

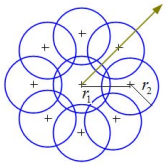

Fig. 2. Canonical orientation (arrow), salient point and its 8 neighborhood vertices (+) assigned with 9 circles.

4. SURFACE MATCHING BY COUNTING THE CORRESPONDING POINTS

In order to find the corresponding salient points between two surfaces, we compare the set of local descriptors by computing their angles. Let FV_i and FV_j be the same kind of feature vectors defined in section 3, the angle of each pair of feature vectors is computed as follows:

$$\alpha = arccos(\frac{< FV_i, FV_j >}{||FV_i||||FV_j||}) \tag{10}$$

The angles of all candidates are ranked in ascending order. We set two vertices as matched only if the rate between their first and second angles is smaller than a pre-defined threshold μ. After this matching step, the number of matched points between two facial surfaces can be counted. This number is then used as the similarity measurement.

5. EXPERIMENTAL RESULTS

We tested our method on the Bosphorus database [19], which contains 4666 textured 3D face models of 105 subjects in various facial expression, pose and occlusion conditions. In our experiments, we extracted the depth maps from all face models and down-sampled them by a factor of 2 for speeding up. Then, they were converted to 3D triangular meshes simply by connecting each pair of neighborhood. Fig.3 shows some samples of the database with six different expressions.

Fig. 3. Some samples with six different expressions, from left to right, anger, disgust, fear, happy, sadness and surprise.

For each subject, the first neutral face was used as gallery. For salient point detection, we set $\sigma_i, (i = 1, 2, 3)$ to 1.83, 2.5 and 4.8 respectively. Using this configuration of parameters, the average number of salient points detected based on κ_{max} and κ_{min} is 293 and 355 respectively. To calculate descriptors, we set R, r_1 and r_2 equal to 22.50 mm, 15 mm and 7 mm respectively. Matching threshold μ was set to 0.70 for HoG and 0.75 for the others.

The matching process was carried out based on the salient points detected by κ_{max} and κ_{min} respectively, and the sum of their matched point numbers was then used as the final decision score. Fig.4 shows two matching results: the left one is between two neutral faces of the same subject while the right one is between a neutral and a non-neutral face of the same person.

Rank one recognition rates of different experiments are given in Table 1. It shows that the performances of HoG, HoS and HoGS are 82.50%, 92.11% and 81.93% respectively on the whole database, and the performance is further improved to 94.10% by fusing all of them, i.e. HoG+HoS+HoGS. This result is slightly better than mesh-SIFT [14] which is 93.66%.

In the case of neutral probes, all the descriptors perform very well, and their recognition rates are 99.48%,

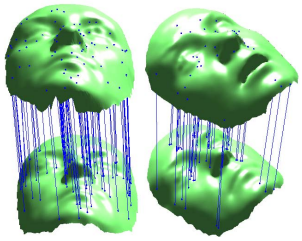

Fig. 4. Matched points between neutral to neutral(left) and neutral to non-neutral (right), salient points detected by κ_{max}.

Table 1. Recognition rates of different descriptors

	HoG	HoS	$HoGS$	$HoG + HoS + HoGS$
All meshes(4561)	82.50 %	90.11 %	81.93 %	94.10 %
Neutral(194)	99.48%	100%	100%	100 %
Anger(71)	69.01 %	87.32 %	76.06 %	88.73 %
Disgust(69)	50.72 %	56.52 %	60.87 %	76.81 %
Fear(70)	71.43%	88.57 %	84.29 %	92.86 %
Happy(106)	79.25 %	85.85 %	75.47 %	95.28 %
Sadness(66)	80.30%	90.91 %	86.36 %	95.45 %
Surprise(71)	81.69%	97.18 %	84.51 %	98.59 %
LFAU(1549)	88.96 %	95.09 %	90.70%	97.22%
UFAU(432)	94.21 %	98.15 %	95.14 %	99.07%
CAU(169)	92.90 %	97.04 %	95.86 %	98.82%
YR(735)	55.37 %	69.25 %	51.84 %	77.96%
PR(419)	94.27%	96.90%	84.73 %	98.81%
CR(419)	60.66%	79.62%	48.34 %	94.31%
O(381)	92.65%	97.38%	93.96 %	99.21%

100%,100%, and 100% respectively. When only considering the probes with six different expressions, HoG and HoS achieved the best results (81.69% and 97.18%) in the subset of surprise expression and HoGS obtained the best one (86.36%) for face scans displaying sadness. However, the subset with disgust expression is the most difficult for all of them. When considering face scans displaying lower facial action unit (LFAU), upper facial action unit (UFAU), and combined action unit (CAU), the performances of HoG+HoS+HoGS are 97.22%, 99.07% and 98.82% respectively. All these results thus demonstrate the robustness of our descriptor to facial expression variations. Additionally, the results on the subset of occlusions (O) show that the proposed method also has some tolerance to occlusions.

6. CONLUSION

In this paper, we presented a mesh-based 3D face recognition approach and evaluated it on the Bosphorus database. Our approach is based on a novel shape descriptor named multiple order surface differential quantities, (HoG+HoS+HoGS) and a SIFT-like matching scheme. The surface differential quantities are extracted on the local neighborhoods of salient points, which are detected by maximum and minimum curvatures respectively. The matching scores based on both salient point detection methods are fused for final decision. Experimental results show the effectiveness of the proposed method and its robustness to facial expression variations.

7. REFERENCES

[1] W. Zhao et al, "Face recognition: a literature survey," *ACM Computing Survey*, vol. 35, pp. 399–458, 2003.

[2] K. W. Bowyer, K. Chang, and P. J. Flynn, "A survey of approaches and challenges in 3d and multi-modal 3d+2d face recognition," *CVIU*, vol. 101, pp. 1–15, 2006.

[3] A. F. Abate et al, "2d and 3d face recognition: a survey," *PRL*, vol. 28, pp. 1885–1906, 2007.

[4] P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi, "The feret evaluation methodology for face recognition algorithms," *PAMI*, vol. 22, pp. 1090–1104, 2000.

[5] K.I. Chang, W. Bowyer, and P.J. Flynn, "Multiple nose region matching for 3d face recognition under varying facial expression," *PAMI*, vol. 28, pp. 1695–1700, 2006.

[6] C. S. Chua, F. Han, and Y. K. Ho, "3d human face recognition using point signature," *FG*, 2000.

[7] X. Lu and A. K. Jain, "Deformation modeling for robust 3d face matching," *PAMI*, vol. 30, pp. 1346–1357, 2008.

[8] A. M. Bronstein et al, "Three dimensional face recognition," *IJCV*, vol. 64, pp. 5–30, 2005.

[9] Y. Wang, G. Pan, and Z. Wu, "3d face recognition in the presence of expression: A guidance-based constraint deformation approach," *CVPR*, 2007.

[10] F. Al-Osaimi, M. Bennamoun, and A. Mian, "An expression deformation approach to non-rigid 3d face recognition," *IJCV*, vol. 81, pp. 302–316, 2009.

[11] Y. Wang, J. Liu, and X. Tang, "Robust 3d face recognition by local shape difference boosting," *PAMI*, vol. 32, pp. 1858–1870, 2010.

[12] D. G. Lowe, "Distinctive image features from scale invariant keypoints," *IJCV*, vol. 60, pp. 91–110, 2004.

[13] A. Zaharescu, E. Boyer, K. Varanasi, and R. Horaud, "Surface feature detection and description with applications to mesh matching," *CVPR*, 2009.

[14] C. Maes et al, "Feature detection on 3d face surfaces for pose normalisation and recognition," *BTAS*, 2010.

[15] T. D. Gatzke, "Estimate curvature on triangular meshes," *IJSM*, vol. 12, pp. 1–28, 2006.

[16] J. Goldfeather and V. Interrante, "A novel cubic-order algorithm for approximating principal direction vectors," *ACM Trans. Graph.*, vol. 23, pp. 45–63, 2004.

[17] J. Wang et al, "3d facial expression recognition based on primitive surface feature distribution," *CVPR*, 2006.

[18] T. H. Meyer, M. Eriksson, and R. C. Maggio, "Gradient estimation from irregularly spaced data sets," *Mathematical Geology*, vol. 33, pp. 693–717, 2001.

[19] A. Savran et al, "Bosphorus database for 3d face analysis," *BIOID*, 2008.