REGIONAL CONFIDENCE SCORE ASSESSMENT FOR 3D FACE

Nesli Erdoğmuş, Jean-Luc Dugelay

Multimedia Communications Department, EURECOM
Sophia-Antipolis, France
{nesli.erdogmus, jean-luc.dugelay}@eurecom.fr

ABSTRACT

3D shape data for face recognition is advantageous to its 2D counterpart for being invariant to illumination and pose. However, expression variations and occlusions still remain as major challenges since the shape distortions hinder accurate matching. Numerous algorithms developed to overcome this problem mainly propose region-based approaches, where similarity scores are calculated separately by local regional matchers and fused for recognition. In this paper, we present a regional confidence score assessment scheme that estimates the expression or occlusion induced distortions in different facial regions. Thereby, reliability scores are obtained which can be used in fusion step for recognition. For 7 regions of face, primitive shape distributions are extracted and the surface quality is measured automatically by an Iterative Closest Point (ICP) based method. Using these measurements, an Artificial Neural Network (ANN) is trained and utilized to estimate regional reliability scores. Experiments have been conducted on FRGC v2 3D face database and results demonstrate a high accuracy in surface quality estimation.

Index Terms— 3d face, surface quality, region-based

1. INTRODUCTION

3D face recognition becomes more popular each day as it offers superiority over 2D face recognition by being intrinsically robust against illumination and pose variations. However, despite many algorithms developed, the deterioration in the recognition accuracy due to intra subject deformations that are introduced by facial expressions and occlusion still needs to be handled.

Approaches aimed to solve these issues are seemed to be gathered around two main directions: extraction of distortion-invariant shape descriptors and region-based matching.

Among the methods to extract distortion-insensitive features, [1] proposes the geodesic distance matrix as a representation of 3D face and claims that the set of largest singular values is an expression-robust descriptor. In [2], expression-insensitive low-level geometric features are ranked by their *confidence* which is calculated in a fashion similar to Fisher's Linear Discriminant. An expression-insensitive descriptor is formed by the higher-ranked features.

Region-based face recognition methods can be again subdivided into two classes. The first class relies on the facial regions that are least affected by the distortions. For example in [3] , [4] and [5], a relative robust area is selected around the nose and eyes. In [6], a preliminary analysis is proposed to weigh the regions according to their stability by studying the anatomical face. The second class separately matches all regions of face and makes a decision based on the fusion of all scores. In [7], Linear Discriminant Analysis is utilized for optimal linear fusion of scores from 10 different pre-defined regions. In [8], face segments are registered individually to specific Average Region Models and point-set differences of all regions are combined for decision-making. Among many different fusion techniques evaluated, the product rule is concluded to be performing best.

In this paper, we present a scheme to evaluate different regions of face according to the present distortions due to expressions and occlusions. First, the face is divided into 7 components (forehead, left eye, right eye, left cheek, right cheek, mouth, and nose) and for each component; every vertex is labeled as one of the 12 primitive shape categories. The distributions of these labels in each region are taken as shape descriptors to determine the quality. Here, quality refers to presence of occlusions or expressions, rather than mesh resolution, noisiness, etc. Next, for every face model, each region is registered to its neutral and clean equivalent which belongs to the same person, using the ICP algorithm [9]. Accepting ICP registration errors as the approximations of volumetric differences between two regions and hence, as the deviations from the undistorted versions, these errors are taken as metrics of imperfection. Finally, an ANN is trained to map between primitive shape distributions (input set) and ICP registration errors (target set).

The remainder of this paper is organized as follows. In Section 2, the algorithm to obtain primitive shape distributions is described in details. In Section 3, we present a method to obtain surface quality metrics automatically for training. Experimental results are given in Section 4. Finally, the paper is concluded in Section 5.

2. PRIMITIVE SHAPE LABELING

Utilization of primitive shape distributions within different facial regions is suggested in [10] for facial expression recognition. In this study, we use similar features not to recognize expressions but to assess the quality of the face segments.

To determine the primitive shape class of each vertex, firstly, gradient magnitudes for all vertices on the facial triangular mesh are computed. For this purpose, initially normals of neighboring triangles are calculated for each point. Then, based on the distances between the vertex and the centroids of the triangles, the weighted average of these normal vectors is taken and assigned as the normal of that vertex. Finally, for a vertex with a normal vector of $n_p = (a,b,c)^T$, the gradient magnitude is calculated as:

$$G = \sqrt{\left[\left[-\frac{a}{c}\right]^2 + \left[-\frac{b}{c}\right]^2\right.} \quad (1)$$

After that, in order to compute the maximum and minimum curvatures, for each vertex on the facial surface, the adjacent points are found and transformed to a local coordinate system, where that vertex becomes the origin and its normal becomes the unit vector along the positive *z* axis. To fit a smooth second order polynomial patch on (X, Y, Z), where X, Y and Z are column vectors of *x, y, z* coordinates of the transformed adjacent points, a least-square solution that minimizes the residual error is computed for the following equation:

$$A \times \left[\frac{1}{2}X^2\ XY\ \frac{1}{2}Y^2\ X^3\ X^2Y\ XY^2\ Y^3\right] = Z \quad (2)$$

Lastly, the eigenvalues of the Weingarten matrix W=[A(1) A(2); A(2) A(3)] are found. The first eigenvalue (the one with the largest absolute value) gives us the amount of greatest curvature whereas the other one gives the least. Therefore, the first and the second eigenvalues are taken as maximum and minimum curvatures, respectively.

Table I Classification rules for the vertices according to magnitudes of their gradients and principal curvatures

		G = 0	G ≠ 0
$k_1 = 0$	$\mathbf{k_2 = 0}$	flat	slope hill
$k_1 = 0$	$\mathbf{k_2 < 0}$	ridge	convex hill
$k_1 = 0$	$\mathbf{k_2 > 0}$	ravine	concave hill
$k_1 < 0$	$\mathbf{k_2 < 0}$	peak	convex hill
$k_1 < 0$	$\mathbf{k_2 > 0}$	ravine saddle	concave saddle hill
$k_1 > 0$	$\mathbf{k_2 < 0}$	ridge saddle	convex saddle hill
$k_1 > 0$	$\mathbf{k_2 > 0}$	pit	concave hill

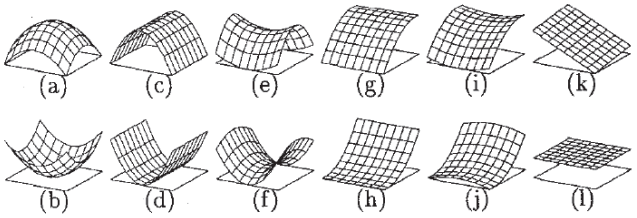

Figure 1 Topographic labels for the center vertex in each example: (a) peak; (b) pit; (c) ridge; (d) ravine; (e) ridge saddle; (f) ravine saddle; (9) convex hill; (h) concave hill; (i) convex saddle hill; (j) concave saddle hill; (k) slope hill; and (I) flat.[11]

Once the gradient magnitude (G – the steepness of the surface) and maximum and minimum curvatures (k_1 and k_2 – maximum and minimum degrees of bending, respectively) for a vertex are obtained, they are utilized to classify it as one of the twelve primitive shapes shown in Figure 1, according to the rules given in Table I.

The decision for the gradient and principal curvature magnitudes to be equal to zero is made based on a threshold $\varepsilon = 0.001 \cdot s$, where s is the average distance from each vertex to the mean of whole vertices. If the magnitudes of the gradients and curvatures are smaller than this threshold, they are accepted as zero.

Finally, histogram distributions of primitive shapes for each region are calculated, resulting in 7 shape descriptors of size [12x1].

3. SURFACE QUALITY COMPUTATION

In this part of the study, an automatic system is developed to measure surface qualities of the facial regions. The outputs of this system together with the corresponding primitive shape distributions are utilized to train an ANN and to evaluate the final scheme.

To this end, each region of all facial scans of a subject is registered to the corresponding region of a neutral (without expression) and clean (without occlusion) scan of the same person (reference model). The final registration error obtained is accepted to measure the distortions on the surface, in view of the fact that it gives us the deviations from an undistorted equivalent.

In order to have more accurate registrations, firstly, the faces are aligned rigidly based on 3 manually-marked fiducial point pairs: outer corners of the eyes and nose tip. There exist many algorithms capable of automatically detecting these landmarks, e.g. [14]. The rotation and translation are computed using Horn's quaternion-based method [12].

Subsequently, the faces are registered using ICP algorithm which iteratively adjusts the transformation needed to minimize the distance between two point clouds until the improvement is stabilized. The final distance obtained is adopted as the surface quality metric where a smaller value indicates a more reliable region. Two example regions of a bad and a good quality are given in Figure 2.

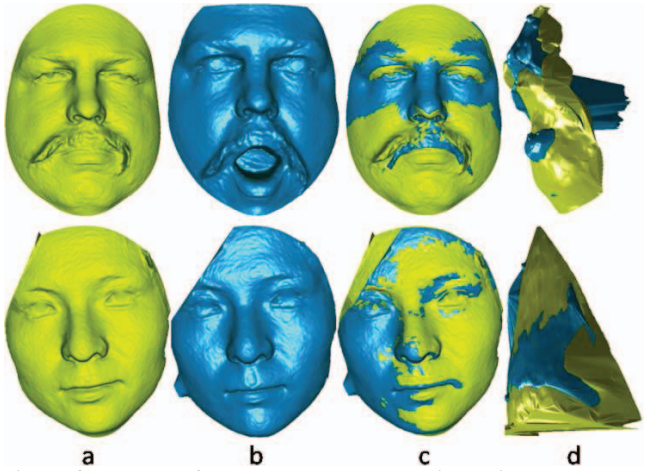

Figure 2 Examples for a bad and a good quality region: (a) neutral and clean reference model; (b) models to be evaluated; (c) model pairs after initial registration; (d) close-up to mouth to region for the 1st example and to forehead region for the 2nd example. The scores computed are 7.67 and 0.96, respectively.

4. EXPERIMENTS AND RESULTS

The proposed surface quality assessment scheme is tested on FRGC v2 database [13]. After eliminating the subjects without a neutral and clean model, primitive shape histograms are computed for 7 regions of 3123 face models from 343 subjects. Next, by ICP-based registration of each model to the corresponding neutral and clean sample from the same person, surface qualities are measured.

The distributions of the obtained scores over different regions are given in Figure 4. In FRGC, most of the existing expressions are around the mouth and the cheeks. Larger deviations can be observed from the corresponding histograms. For the forehead, the distortions are mainly sourced from occlusions by hair. The most scarcely distorted region is the nose.

Lastly, an artificial neural network to map between primitive shape distributions (12 inputs) and quality measurements (1 output) with a hidden layer of 10 neurons is constructed for each region.

For the experiments, the whole sample set is divided randomly into 3 subsets: 50% for training (1561), 25% for validation (781) and 25% for testing (781). The regional ANNs are trained according to Levenberg-Marquardt optimization until the mean square error (MSE) of the validation samples stops decreasing. Finally, the network outputs for the test set are obtained and the MSE is calculated between the given output values and the real automatic measurements.

These tests are repeated for 1000 times for each region, each time with different subsets. The obtained average MSE for all regions and their standard deviations are presented in Table II. Additionally, for the purpose of having a reference point to evaluate the results, minimum scores that are automatically measured; in other words, best registration errors for each region of all face models are given. It is noted that different results are due to not only different training and test subsets but also different initial conditions for the ANN.

Table II Average MSE for regional quality estimations and their standard deviations and minimum scores measured for each region

	Forehead	Left Eye	Right Eye	Left Cheek	Right Cheek	Mouth	Nose
Avg.	0.2058	0.1771	0.1519	0.2386	0.2930	0.5741	0.0552
Std.	0.1581	0.0790	0.0946	0.1312	0.1242	0.1024	0.0063
Min.	0.7219	0.7846	0.7645	0.7964	0.8400	0.8027	0.8146

For further analysis, training, validation and test sets are created by dividing the sample set in blocks; first 50% of the samples for training, next 25% for validation and next 25% for testing. 7 ANNs are trained for all regions and MSE values of **0.1861, 0.1665, 0.1359, 0.2038, 0.2098, 0.4916** and **0.0399** are obtained for forehead, left eye, right eye, left cheek, right cheek, mouth and nose, respectively.

The accuracy of the ANN's with respect to the magnitude of distortions present on the surface is analyzed. For this purpose, MSE values and success rates are calculated for different bins, where each bin consists of regions with a quality measurement less than a score threshold. A quality estimation is accepted as successful if the squared error is lower than 0.5. As can be seen in Figure 3, the performance tends to decrease at extremities (possibly due to small number of samples); however the accuracy remains very high.

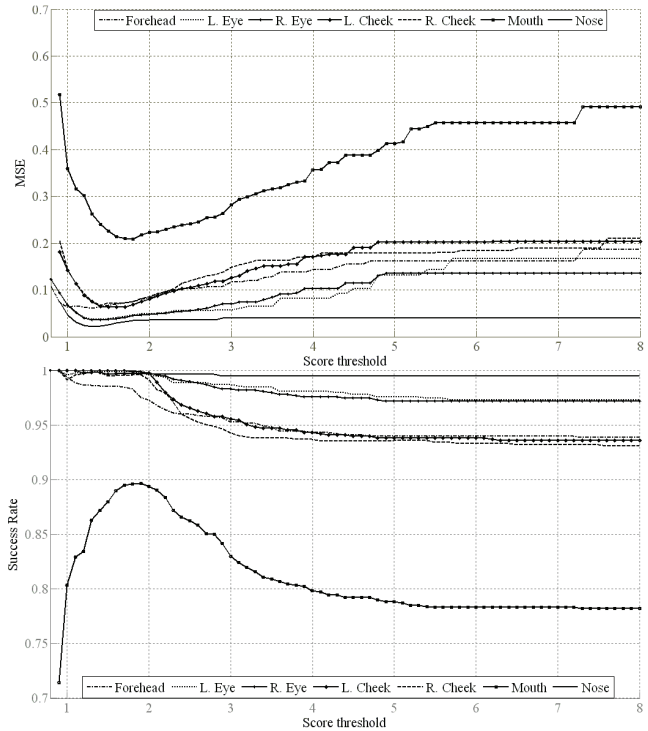

Figure 3 MSEs and success rates calculated separately for different bins, where each bin consists of regions with a quality measurement less than a score threshold

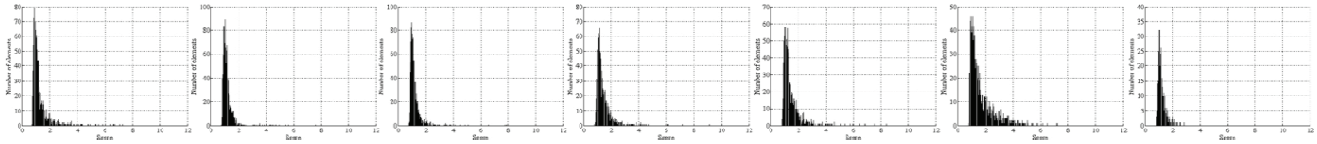

Figure 4 Distributions of scores for each region: forehead, left eye, right eye, left cheek, right cheek, mouth and nose

The least accurate estimations (78.21%) are done for the mouth region. This is mainly because the possible deformations for mouth are much more diverse than the other regions.

5. CONCLUSION

In this paper, we presented a novel scheme to estimate facial surface qualities within different regions. Distortions in the face mesh due to expressions and occlusions are a prime challenge in 3D face recognition. We think the suggested method will help to evaluate different regions and improve the recognition performances by discarding or giving less importance to areas with low quality.

The regional shape information of face is extracted based on primitive shape labeling of facial points. After calculating the gradient magnitude and principal curvatures for each point, one of the 12 primitive shapes is assigned. The histograms of these labels in each region are utilized to estimate the surface quality.

The ground truth for quality measurements is determined by registering each facial region to its neutral and clean equivalent from the same person. ANN-based fitting is adopted for assessment of regional qualities.

Experiments are conducted on FRGC v2 database. Among 3123 sample models from 343 subjects, training, validation and testing sets are created by taking 50%, 25% and 25% of the samples, respectively.

The results reveal that the surface qualities of different facial regions can be estimated with high accuracy. Over 93% success rate is achieved for all regions except mouth. For mouth region, the average MSE (maximum among all regions) is calculated to be 0.5741, where minimum scores (registration error for best quality surfaces) of all regions range between 0.72-0.84. The performances are observed to deteriorate slightly for low quality regions, which we think is due to small sample set size.

Possible future work includes adding the surface quality estimations to a 3D face recognition algorithm, to observe the increase in robustness against expressions and occlusions.

6. REFERENCES

[1] D. Smeets, T. Fabry, J. Hermans, D. Vandermeulen, P. Suetens, “Isometric Deformation Modeling Using Singular Value Decomposition For 3D Expression-Invariant Face Recognition,” *3rd IEEE Int. Conference on Biometrics: Theory, Applications, and Systems,* Washington, pp.1-6, September 2009.

[2] L. Xiaoxing, J. Tao, Z. Hao, "Expression-Insensitive 3D Face Recognition Using Sparse Representation," *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, pp.2575-2582, June 2009.

[3] Z. Cheng, S. Zhenan, T. Tieniu, H. Zhaofeng, "Robust 3D Face Recognition In Uncontrolled Environments," *IEEE Conference on Computer Vision and Pattern Recognition*, Alaska, pp.1-8, June 2008.

[4] I. C. Kyong, K. W. Bowyer, P. J. Flynn, "Adaptive Rigid Multi-region Selection for Handling Expression Variation in 3D Face Recognition," *IEEE Conference on Computer Vision and Pattern Recognition - Workshops*, San Diego, pp.157, June 2005.

[5] G. Gunlu, H. S. Bilge, "3D Face Decomposition and Region Selection against Expression Variations," *20th Int. Conference on Pattern Recognition*, Istanbul, pp.1298-1301, August 2010.

[6] B. B. Amor, M. Ardabilian, C. Liming, "Toward A Region-Based 3D Face Recognition Approach," *IEEE Int. Conference on Multimedia and Expo*, Hannover, pp.101-104, June 2008.

[7] L. Wei-Yang, W. Kin-Chung, N. Boston, H. H. Yu, "3D Face Recognition Under Expression Variations using Similarity Metrics Fusion," *IEEE Int. Conference on Multimedia and Expo*, Beijing, pp.727-730, July 2007.

[8] N. Alyuz, B. Gokberk, L. Akarun, "A 3D Face Recognition System for Expression and Occlusion Invariance," *2nd IEEE Int. Conference on Biometrics: Theory, Applications and Systems*, Washington, pp.1-7, October 2008.

[9] P. J. Besl, H. D. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.14, no.2, pp.239-256, February 1992.

[10] J. Wang, L. Yin, X. Wei, Y. Sun, "3D Facial Expression Recognition Based on Primitive Surface Feature Distribution," *IEEE Conference on Computer Vision and Pattern Recognition*, New York, pp.1399-1406, June 2006.

[11] O. D. Trier, T. Taxt, A. K. Jain, "Data capture from maps based on gray scale topographic analysis," *3rd Int. Conference on Document Analysis and Recognition*, Montreal, pp.923-926, August 1995.

[12] B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," *Optical Society of America A*, vol. 4, pp.629-642, 1987.

[13] P. J. Phillips, P. J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, “Overview of the face recognition grand challenge” *IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, pp.947-954, June 2005.

[14] N. Erdogmus, J.-L. Dugelay, “Automatic extraction of facial interest points based on 2D and 3D data”, *Electronic Imaging Conference on 3D Image Processing (3DIP) and Applications, SPIE, San Francisco,* vol. 7864, January 2011.