A STUDY ON REGION-BASED RECOGNITION OF 3D FACES WITH EXPRESSION VARIATIONS

Helin Dutağacı[1], Berk Gokberk[2], Bülent Sankur[1], Lale Akarun[2].
[1]*Department of Electrical-Electronics Engineering,*
[1]*Department of Computer Engineering,*
Boğaziçi University, Istanbul
{dutagach, bulent.sankur, akarun}@boun.edu.tr, gokberk@gmail.com}

ABSTRACT

In this paper, we propose the application of masks as a means to mitigate expression-distortions on 3D faces and to enhance their recognition performance. Masking becomes necessary to de-emphasize the face regions that deform under expression. We have conducted experiments with various masks, namely, ellipse-shaped binary masks, Gaussian, super-Gaussian and raised-cosine masks. The design issues of the masks, such as the mask size, the centre, the support region, the decay rate of the tails, etc. are studied and adjusted with respect to their recognition performances. We show first that warping the depth values of corresponding face points onto the same spatial coordinates while obtaining the 2D depth images is beneficial, and second, that proper masking can add several percentage points to the recognition performance.

1. INTRODUCTION

There is increasing interest in 3D face recognition, not only due to advances in sensor technologies, but also due to the fact that 3D face recognition possesses certain benefits over intensity-based 2D face recognition: The two crucial advantages are the illumination-invariance and the ease of detection and cropping of the face region from the background. However expression variation remains as a challenge for 3D face recognition systems. This point is illustrated in Fig. 1 where we show face scans of three subjects, each with three varying facial expressions, from the Face Recognition Grand Challenge (FRGC) database [1].

A number of algorithms were proposed to deal with the deformation of the geometric structure of the face due to expression. One approach is to model the face as a deformable object. Lu and Jain [2] propose person-specific deformable models, where a small group of subjects is used to learn the deformations. The learned deformation model is then transferred to the 3D neutral model of each subject via warping with thin-plate splines. Bronstein et al. [3] propose multidimensional-scaling to embed a deformed 3D face onto another face. The multidimensional-scaling is used to determine the geometrical correspondences between two deformed surfaces.

** This work was partially supported by TÜBİTAK project 104E080.*

Another approach to deal with expression variations is to adopt a region-based scheme. Chang et al. [4] use three overlapping face regions around the nose. These regions are assumed to be less deformable under expressions as compared to those facial parts including eyes and mouth. The corresponding facial region pairs from the gallery and probe images are matched with Iterative Closest Point (ICP) algorithm, and the matching scores are combined with the product rule. Any other region that is deemed deformable under expressions is ignored. Faltemier et al. [5] describe a system, where one pre-determined facial region in the gallery image is compared with multiple regions in the probe image, and then their outcomes are combined through committee voting.

A more general treatment of local region-based face recognition system is presented in [6] and [7] for 2D and 3D face modalities, respectively. The underlying principle is the automatic determination of discriminative parts of facial regions via feature subset selection heuristics. These authors show that, even without prior knowledge on the importance of facial subregions, one can learn informative facial parts from the data itself which leads eventually to better identification rates.

In this paper, we investigate schemes to compensate for facial expressions in order to enhance 3D face recognition. The scheme is based on the de-emphasis of the deformable face regions through masking, instead of totally discarding them. The design of the masks is an important issue, since the mask parameters determine the relative contributions of different facial regions to the recognition performance. We test various masking schemes and compare the performances at different parameter settings on the FRGC v2.0 face database.

In Section 2, we describe the registration procedure and the generation of 2D depth images via spatial warping. In Section 3 we introduce the masks used in this work. We briefly describe the feature extraction techniques in Section 4. Experimental results are provided in Section 5, and finally we conclude in Section 6.

2. 2D DEPTH IMAGE GENERATION

The common approach for registration is alignment of the 3D point cloud of a probe image onto each gallery image separately via the Iterative Closest Point (ICP) algorithm

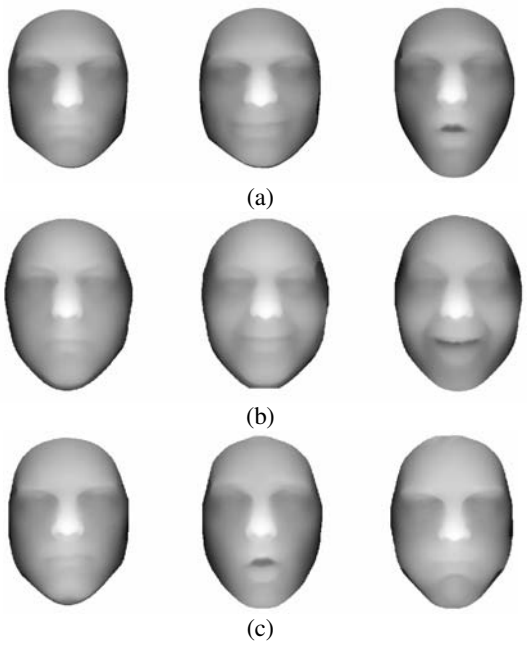


Figure 1 – Unwarped depth images from three different subjects (a, b, c). Faces on the same row correspond to the same person with different facial expressions.

[8]. Since ICP is a time-consuming procedure, the alignment of an input face to all the faces in the database precludes real-time operation. Therefore we use an Average Face Model (AFM) obtained from a set of training face samples and align the 3D point cloud of each face only to AFM via ICP. Fig. 2 shows an Average Face Model mapped onto a 2D depth image. This scheme allows us to rapidly build correspondences among faces.

ICP alignment is a rigid transformation that yields aligned point set correspondence of a face. We first use the ICP algorithm to best match the fiducial points of a given face to those of the AFM. The seven fiducial points used are the four inner- and outer-eye corners, nose tip and the two mouth corners. Then we apply spatial warping to relocate (x,y) face coordinates on top of the regular grid of the AFM. Finally, the registered depth image of a face is formed with the z-coordinates of the input face image located at the (x, y) coordinates of the Average Face Model to yield the depth function $H(x, y)$.

This idea is similar to the Active Appearance Model of Cootes et al. [9], where 2D intensity faces are warped on an average shape model of the faces in order to establish correspondences. In our study, we treat the depth of each point as the appearance of a face. The model will be complete if we also consider the (x, y) coordinates of the face points and model the spatial arrangement of the points. However in this study, we limit ourselves to the depth values only.

Fig. 3 shows the warped depth images of the faces depicted in Fig. 1. The faces look very similar to each other, because the spatial arrangements of the pixels belong to the average face. However, the geometric information represented by the depth values is preserved. Fig. 4 shows the profiles of three face images of a subject in dashed curves and three profiles of another subject in solid and black curves. With this single profile, two classes seem to be separable from each other.

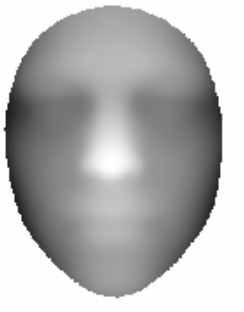


Figure 2 – Depth view of the AFM: Average Face Model

This warping scheme not only moves corresponding face points to the same spatial locations in the depth image, but also reduces the deformation caused by expression variation. A visual inspection of Figures 1 and 3 shows that the within-class variations due to expression are reduced after warping. This result is coherent with the Active Appearance Model of faces [9], where by warping intensity values on to an average shape model, one can decouple expression from the appearance of the face.

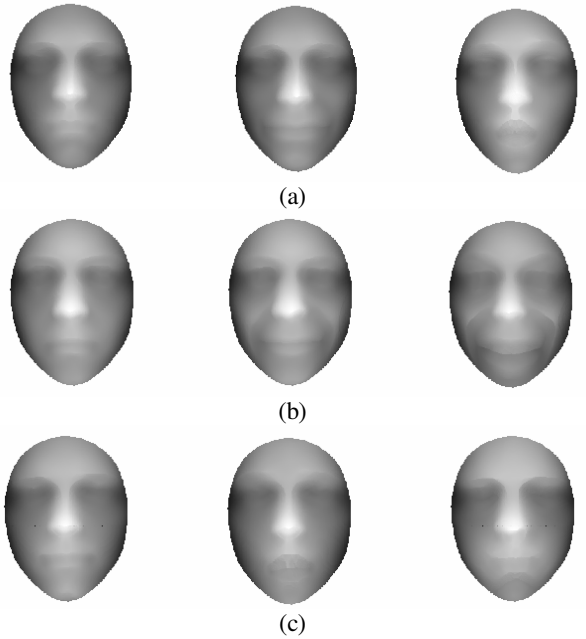


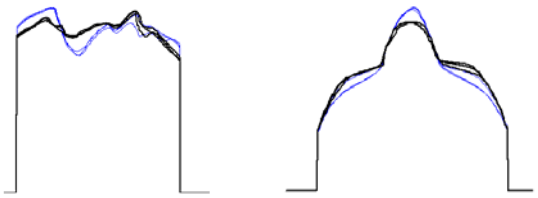
Figure 3 – Warped depth images of the faces shown in Figure 1.

3. MASKING SCHEMES

Since regions of the depth map $H(x, y)$ have varying reliability, we can privilege certain regions over others by multiplying with masks $W(x, y)$:

$$I(x, y) = W(x, y)H(x, y)\,. \qquad (1)$$

The two issues that must be addressed are the shape and the location of the mask functions.



(a) (b)

Figure 4 – Vertical (a) and horizontal (b) profiles of faces from two subjects.

We have tested four different masks: Ellipse-shaped binary mask, Gaussian mask, super-Gaussian mask and raised-cosine mask (Fig. 5).

The ellipse-shaped binary masking can be considered as a parts-based approach, where one particular region of the face is matched with the corresponding region of another face. We have chosen ellipse-shaped regions in order to make a fair comparison with the Gaussian and super-Gaussian counterparts based on similar control parameters, such as centre, size, support region, etc. The general form of the ellipse-shaped binary mask is as follows:

$$W_E(x,y) = \begin{cases} 1 & if \quad \left(\frac{x-X_c}{a}\right)^2 + \left(\frac{y-Y_c}{b}\right)^2 \le 1 \\ 0 & otherwise \end{cases}. \qquad (2)$$

We have selected three parameters of the ellipse as variables: The vertical centre point of the ellipse, Y_c, along the symmetry axis of the face, the horizontal radius, a and the vertical radius, b. The centre of the ellipse is constrained to be at the symmetry axis of the face. The Gaussian mask has the following form:

$$W_G(x,y) = \exp\left\{-\left(\frac{x-X_c}{a}\right)^2 - \left(\frac{y-Y_c}{b}\right)^2\right\}. \qquad (3)$$

The Gaussian mask is applied to the whole face; hence this scheme does not discard any face region. Instead, we weight the face points smoothly, with the points near the centre of the mask contributing more as compared to further points. This is controlled by the aperture parameters of the Gaussian mask.

To manipulate the decay regime of the Gaussian mask, so that it remains flat over a larger region and then drops more rapidly to zero, we propose the use of a super-Gaussian mask of order 3. Higher powers of the super-Gaussian will make the mask similar to an ellipse-shaped mask.

$$W_{SG}(x,y) = \exp\left\{-\left|\frac{x-X_c}{a}\right|^3 - \left|\frac{y-Y_c}{b}\right|^3\right\}. \qquad (4)$$

The fourth type of mask is the raised-cosine mask, which can provide a flat value over a controlled support region. The raised-cosine mask can be obtained from the multiplication of raised-cosine windows along rows and columns of the image:

$$W_{RC}(x,y) = w_{RC}^a(x) w_{RC}^b(y), \qquad (5)$$

where,

$$w_{RC}^K(t) = \begin{cases} 1 & |t| \le \frac{K(1-\beta)}{2} \\ g_{RC}^K(t) & \frac{K(1-\beta)}{2} < |t| \le \frac{K(1+\beta)}{2} \\ 0 & \text{otherwise} \end{cases} \qquad (6)$$

and

$$g_{RC}^K(t) = \frac{1}{2}\left[1+\cos\left(\frac{\pi K}{\beta}\left[|t| - \frac{(1-\beta)}{2K}\right]\right)\right]. \qquad (7)$$

We have set β to 0.5. The raised-cosine mask provides a region-based representation similar to the ellipse-shaped binary mask. However, with the raised-cosine mask we have a smoother transition between the support region and other regions of the face.

(a) (b) (c) (d)

Figure 5 – Ellipse-shaped (a), Gaussian (b), super-Gaussian (c) and raised-cosine (d) masks.

4. FEATURES

We have tested the performance of masking schemes with two well-known feature extraction schemes: Discrete Fourier Transform (DFT), and Principal Component Analysis (PCA). DFT-based techniques were previously applied to 3D face recognition [10] and they are common in region-based image recognition problems. On the other hand, PCA techniques are suitable for problems where correspondent points among images are well established.

We apply 2D-DFT on the registered and masked depth function and extract the first *M*x*M* complex DFT coefficients. The real and imaginary parts of these coefficients are concatenated in a one-dimensional vector, which forms the DFT-based feature vector of a masked face.

PCA-based feature extraction is a well-studied technique, and has been widely applied to both 2D and 3D face recognition problems [1]. The values of each of the masked faces in the training set are concatenated to form a single vector. Part of these vectors are used as training vectors to constitute the PCA bases, while the remaining ones are projected onto these bases to form the feature vectors of the test faces.

Furthermore, the DCT and PCA coefficients are re-weighted through QR-decomposition in order to make use of the class information available in the training set:

$$\mathbf{F} = \mathbf{QR}, \qquad (8)$$

where the columns of $\mathbf{F}$ contain the difference of the feature vector of each face sample to its class mean. $\mathbf{R}$ is the upper

diagonal matrix obtained from the QR-decomposition of the training features. We multiply all feature vectors in the training and test sets with the pseudoinverse of **R** to re-weight features. Two re-weighted features are then compared using L_2 distance.

5. EXPERIMENTAL RESULTS

We have tested the performance of masking-based 3D face recognition on the FRGC v2.0 database. We have considered the case where there is only one gallery image in the database. There are 410 subjects hence, 410 gallery images. The remaining 3542 face scans are used as test images.

In order to train the PCA basis and obtain QR decomposition we have used a separate dataset: The FRGC v1.0 database. This database consists of 854 face scans of 194 subjects and does not contain the face scans present in FRGC v2.0. The PCA basis and the transformation matrix **R** are calculated and fixed on the v1.0 database, and then used to weight the features of the gallery and test images of the v2.0 database.

As a baseline, we have used both warped and unwarped depth images without masking. Table 1 shows the performances obtained with unmasked face images using DFT and PCA-based features. The best performance on unmasked images is obtained with warping and PCA-coefficients. This is much higher than the best performance obtained from the DFT coefficients with masking (Fig.6 and Table 2). This is not surprising, since the 2D-DFT is sensitive to spatial structure of the depth values, whereas PCA only considers the variations among corresponding points regardless of their position. After warping, the spatial structure of the depth values does not carry class information since they are arranged with respect to the average face.

Table 1 – Recognition performances (%) of unmasked faces.

	DFT	PCA
Unwarped	71.71	74.20
Warped	80.66	87.15

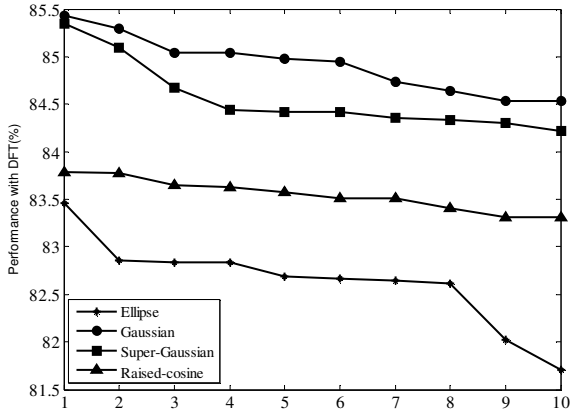


Figure 6 – Performances with best 10 mask parameter sets for each masking scheme, obtained with DFT coefficients.

By varying the vertical centre, the support regions and the decay rates, we have experimented with 128 variations of each of the four masks. The depth image is of size 201x161. We varied the centres of the masks between 30 and 160, with an increment of 10. The *a* and *b* parameters for the elliptic, Gaussian and super-Gaussian windows are taken in the range of 20 to 80 with an increment of 20. For the raised-cosine mask, *a* and *b* vary between 60 to 240 with an increment of 60.

Table 2 gives the best 10 performances of the four masks among their different parameterizations with DFT features. While unmasked image performance is 80.66 %, all masked versions register a few percentage point improvement. The Gaussian mask has the highest gain, followed closely by super-Gaussian. Both raised-cosine and elliptic windows fall about two percentage points behind. Both Gaussian and raised-cosine masks are quite insensitive to parameter adjustments while the elliptic mask necessitates fine-tuning. (**Fig. 6**)

Table 2 – Recognition performances (%) of best masks with DFT. Gain: 4.77 percentage points.

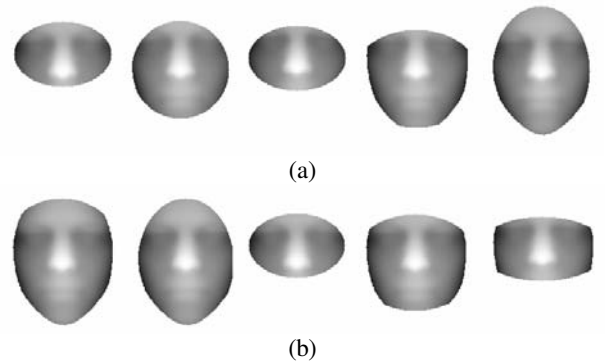
Unmasked	Ellipse-shaped	Gaussian	Super-Gaussian	Raised-cosine
80.66	83.46	85.43	85.35	83.79

Table 3 shows the best recognition performances among mask parameterizations, obtained via PCA coefficients. Compared to DFT features, the gains with the PCA features are less impressive. The best performance is again achieved with the Gaussian mask. Elliptic binary masking yields little improvement. Fig. 7 (a) and (b) illustrates the best five ellipse-masked faces with for DFT and PCA techniques. For the DFT technique, the best ellipse includes only the nose and eye regions. The second runner ellipse includes also the mouth. This result shows that discarding the mouth and chin for the sake of expression invariance causes a loss in the class information available to a recognition system. Actually, when we observe the best ellipse-masked face for the PCA case (Fig. 7 (b), first face) we see that almost all face regions contribute to this performance.

Table 3 – Recognition performances (%) of best masks with PCA. Gain: 0.94 percentage points.

Unmasked	Ellipse-shaped	Gaussian	Super-Gaussian	Raised-cosine
87.15	87.32	88.09	87.89	87.63

The PCA coefficients derived from the Gaussian masked faces give the best performance, and the performance is relatively insensitive to the parameterization. Fig. 8 shows the Gaussian-masked faces giving the best five performances, all of which are around 88%. Their centres are all located around the nose tip. However their aperture parameters are different.



(a)

(b)

Figure 7 –Ellipse-masked faces giving the best five performances with (a) DFT coefficients (b) PCA coefficients

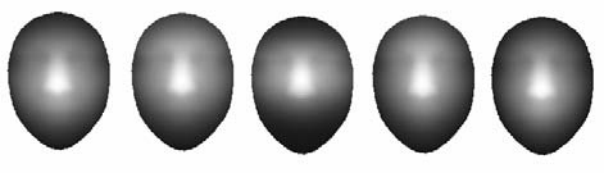


Figure 8 –Gaussian-masked faces giving the best five performances with PCA coefficients

6. CONCLUSIONS AND FUTURE WORK

In this paper, we have proposed the use of smooth masks to deal with expression variations in 3D faces. We have conducted experiments with an ellipse-shaped binary mask, a Gaussian mask, a super-Gaussian mask and a raised-cosine mask with a large number of possible parameters for each. We have also experimented with the use of warping depth fields into an average face in order to reduce the deformation due to facial expressions.

Warping depth images so that the depth values at the same location come from the corresponding points of the 3D point clouds is beneficial for reducing the effect of facial expression. This scheme is of great advantage especially with the PCA-based technique, since PCA models the data better when correspondences are well-established.

Another important observation is that avoiding expression-prone face regions such as mouth and chin results in class information loss. Weighting the face regions smoothly via a Gaussian mask, with high weights assigned to the rigid regions such as nose tip, results in an improved performance. Furthermore one does not need to fine-tune the parameters of the Gaussian mask in order to get the best region.

The best performance in the literature with the same database and the same experimental setting is 94.9 % [5]. Here various face regions are compared with each of the gallery images via ICP and the results are fused with committee voting. We have obtained 88.09 % recognition performance with warping and Gaussian masking. We have implemented only one ICP procedure for a probe image, which makes the system much faster and we have used a single masked image. The proposed schemes are open to improvements.

As future work, we will consider methods to regain the spatial arrangement of the points, i.e. the geometric structure lying in (x, y) coordinates of the point cloud. One alternative is to model (x, y) coordinates separately with PCA and to combine the results with some fusion scheme. Another alternative is to apply warping on class basis. Warping a probe face onto all the depth images in the gallery may give better recognition performance in expense with increasing processing time.

Another future work that follows naturally is the fusion of the results obtained with different masking schemes or with different parameterizations of the same masking scheme. The best parameter/region set may be selected with some automatic algorithm performed over a training set.

REFERENCES

[1] K. Chang, K. W. Bowyer, and P. J. Flynn, “Face recognition using 2D and 3D facial data,” in ACM Workshop on Multimodal User Authentication, 2003, pp. 25–32.

[2] X. Lu and A. K. Jain, “Deformation modelling for robust 3D face matching,” in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR2006), 2006.

[3] A. Bronstein, M. Bronstein, and R. Kimmel, “Three-dimensional face recognition,” International Journal of Computer Vision, vol. 64, no. 1, pp. 5–30, 2005.

[4] K. Chang, K. Bowyer, and P. Flynn, “Adaptive rigid multi-region selection for handling expression variation in 3D face recognition,” in IEEE Workshop on Face Recognition Grand Challenge Experiments, 2005.

[5] T. Faltemier, K. Bowyer, and P. Flynn, “3D face recognition with region committee voting,” in Proceedings of the Third International Symposium on 3D Data Processing, Visualization and Transmission, 2006.

[6] B. Gökberk, M. O. İrfanoğlu, L. Akarun and E. Alpaydın, “Learning the best subset of local features for face recognition”, Pattern Recognition, Vol 40, No. 5, pp. 1520-1532, 2007

[7] B. Gökberk and L. Akarun “Selection and extraction of patch descriptors for 3D face recognition,” Proceedings of ISCIS’05. Lecture Notes in Computer Science, Vol. 3733, pp. 718-727.

[8] X. Lu, A. Jain, and D. Colbry, “Matching 2.5D face scans to 3D models,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 1, pp. 31–43, 2006.

[9] T. F. Cootes, G. J. Edwards and C. J. Taylor. "Active appearance models", IEEE PAMI, Vol.23, No.6, pp.681-685, 2001.

[10] H. Dutağacı, B. Sankur, and Y. Yemez, “3D face recognition by projection-based features,” in Proc. SPIE Conf. on Electronic Imaging: Security, Steganography, and Watermarking of Multimedia Contents, 2006.