

3D Facial Landmarking under Expression, Pose, and Occlusion Variations

Hamdi Dibeklioglu, Albert Ali Salah and Lale Akarun

Abstract—Automatic localization of 3D facial features is important for face recognition, tracking, modeling and expression analysis. Methods developed for 2D images were shown to have problems working across databases acquired with different illumination conditions. Expression variations, pose variations and occlusions also hamper accurate detection of landmarks. In this paper we assess a fully automatic 3D facial landmarking algorithm that relies on accurate statistical modeling of facial features. This algorithm can be employed to model any facial landmark, provided that the facial poses present in the training and test conditions are similar. We test this algorithm on the recently acquired Bosphorus 3D face database, and also inspect cross-database performance by using the FRGC database. Then, a curvature-based method for localizing the nose tip is introduced and shown to perform well under severe conditions.

I. INTRODUCTION

Advances in sensing technology and the availability of sufficient computing power has made 3D face recognition one of the attractive modalities for biometrics. 2D face recognition traditionally suffers from pose, illumination, occlusion and expression variations that affect facial images more than changes due to identity. However illumination problems can be avoided to a large degree by using 3D. As in the 2D case, 3D data must be properly pose normalized and registered to enable recognition or expression analysis. 3D registration remains the crucial part in the recognition chain as in 2D. As most existing registration approaches require landmarks (feature points) to be located on the surface as initialization, a robust estimation of facial feature points is of prime importance.

We distinguish between two types of landmark detection for face images, *coarse* and *fine*, respectively. Coarse landmarking aims at indicating a rough position for a particular landmark, whereas fine landmarking aims at finding the precise location. Depending on the application requirements, detecting several landmarks at a coarse scale may be sufficient. A typical example would be fast initial alignment of face scans in an iterative closest point (ICP)-based registration scheme. Fine landmarking, on the other hand, is preferably performed in facial areas greatly restricted by coarse landmarking. Since the local feature information and the structural relations between landmarks are jointly relevant

for fine landmarking, the computational cost becomes an important issue. Fine landmarking is used for expression analysis, face recognition from structural landmark distributions, and non-rigid registration methods like thin-plate splines.

Most facial landmarking approaches work with 2D information (see [1] for an extensive review). Under consistent illumination conditions, 2D is richer in discriminative information, but 3D methods are found to be more robust under changing illumination conditions [2]. In 3D, the prominence of the nose makes it a relatively easy candidate for fast, heuristic-based approaches [3]. For instance in [4] a coarse-to-fine heuristic procedure is described to locate the tip of the nose, and subsequent pose correction relies on this point. D'Hose et al. use Gabor filter responses and ICP to find the tip of the nose, and use it to restrict the search for other landmarks [5]. In [6], the nose is located first, and candidates for the remaining landmark points are encoded according to their position with respect to the nose. Even though relying on a single landmark is a potential threat to the robustness of the system, nose localization can produce very good results in practice, since it is the easiest landmark to locate and has the greatest effect in registration [7].

Apart from the nose, eye and mouth corners are the most frequently used facial landmarks. If the facial symmetry axis can be found, locating the eye and mouth corners is straightforward for neutral faces [8]. However, the search for the symmetry axis can be costly without the guiding landmarks. Curvature-based features are promising in 3D, but they suffer from a number of problems [9], [10]. Reliable estimation of curvature requires a strong pre-processing that eliminates surface irregularities, especially near eye and mouth corners. Two problems are associated with this pre-processing: The computational cost is high, and the smoothing destroys local feature information to a great extent, producing many points with similar curvature values in each local neighbourhood. However, curvature can be used to greatly reduce the area for the more costly fine-tuned search.

One issue that makes consistent landmarking difficult is that the anatomical landmarks are defined in structural relations to each other, and the local feature information is sometimes not sufficient to determine them correctly. For flat-nosed persons, the “tip of the nose” is not a point, but a whole area of points with similar curvature. This is the natural version of the smooth-areas problem artificially created with pre-processing. More elaborate 3D methods, like spin images, are very costly in practice, but may become feasible with fast embedded hardware or with sufficient

This work was supported by the European FP6 BIOSECURE and the Dutch BRICKS/BSIK projects.

H. Dibeklioglu and L. Akarun are with the Department of Computer Engineering, Boğaziçi University, Bebek, Istanbul, Turkey hamdi.dibeklioglu@cmpe.boun.edu.tr, akarun@boun.edu.tr

A.A. Salah is with Centrum Wiskunde & Informatica, 1090 GB, Amsterdam, The Netherlands a.a.salah@cwi.nl

downsampling [11]. Recently, Kakadiaris et al. proposed matching of spin images to align faces prior to ICP, and obtained very good results [12]. The reported time delay to convert raw scanner data to registered metadata is 15 seconds.

In this paper we develop statistical and heuristic approaches for 3D facial landmarking. Section II introduces a generic, coarse-to-fine algorithm for 3D facial landmarking. In Section III we focus on the tip of the nose, and introduce a robust curvature-based method for its localization. Section IV summarizes our experimental results on the newly acquired Bosphorus 3D face database. This database contains more than 3,000 images from 81 subjects, and has ground truth locations for 22 landmarks. We also inspect cross-database performance of the proposed algorithm by applying the learned models to 3D faces from the FRGC dataset. We conclude in Section V.

II. AUTOMATIC LANDMARK LOCALIZATION: STATISTICS

The method we use for statistical landmark localization is based on an analysis of local features to determine the most likely location for each landmark separately. If the local feature information is poor for some landmarks (e.g. moustache hiding the mouth corners), an optimization-based method that attempts to localize features jointly will suffer. The structural information can be integrated at a later stage, with more control on the relative contributions of local features and structural information, respectively.

To detect the landmarks individually, we model the distribution of features sampled from around the landmark locations in the training set. We use a mixture of factor analysers (MoFA) as an unsupervised model. A MoFA is similar to a mixture of Gaussians, but the parameter space for the covariance matrix is better explored. During the training stage, clustering and dimensionality reduction are performed in conjunction.

For a training set and a given MoFA model (i.e. by specifying the number of components and the number of factors per component) maximum likelihood parameters can be computed with the Expectation-Maximization (EM) algorithm [13]. Here we use an incremental algorithm that searches the model space along with the parameter space [14]. The incremental MoFA algorithm (IMoFA) starts with a one-factor, one-component mixture and adds new factors and new components until a stopping condition is satisfied. At each iteration, fast heuristics are used to constrain possible actions. The incremental approach is especially feasible when the dimensionality is high, as opposed to starting with a large number of components, and eliminating components one by one [15].

A. Local Features

We locate the facial features using the depth map (or range images) and its gradient information, with a coarse-to-fine method. It was shown in [1] that a coarse-to-fine approach increases the accuracy of feature detection by making more

discriminatory information available to a single patch in the coarse stage. We downsample the range images to a size that allows low dimensional modeling of local features. When adapting models learned on one dataset to a new dataset, a similar scale must be chosen, as the range images are not scale independent. For the FRGC dataset, the downsampling ratio was selected to be one to eight (i.e. 480×640 to 60×80). A z -normalization was applied to depth values of valid pixels by subtracting the mean and dividing by the standard deviation. Then the gradient images in horizontal and vertical directions are extracted from the depth map, to produce two supplementary information channels. During training, 7×7 neighbourhoods are cropped from around each landmark, from each of those maps. The ensuing 49-dimensional features are (min-max) normalized to the range $[0,1]$ before modeling with IMoFA. The neighbourhood size is empirically determined.

In the test phase, the model produces one likelihood map of the image for each landmark per channel. The channels can be combined by a simple SUM rule, or by assigning different weights for different landmark types, since the relative importance of horizontal and vertical features depend on the landmark appearance. The best set of weights in this case is determined on the training set.

Working on the downsampled images, we determine the landmark locations on the coarse level. Fig. 1 shows the likelihood maps for seven landmarks on the FRGC dataset, and the final detection result in the coarse scale. In [2], the coarse localization is complemented with a structural correction algorithm (GOLLUM) and a fine-level search for greater accuracy. Due to restricted space, we omit these steps from the present work.

III. AUTOMATIC LANDMARK LOCALIZATION: HEURISTICS

The statistical methods are successful for faces where the local feature information is sufficient, but they do not cope well with severe pose variations, occlusions and novel expressions. Especially for greater rotations, a robust method for determining and correcting the pose of the face would be necessary. Once the pose is corrected, it is possible to consult statistical models to determine a large number of landmarks on the face. In this section we describe a heuristic method to localize the tip of the nose under severe conditions. The proposed algorithm finds discrete candidate regions for the nose, and selects the most plausible location based on curvature values, which makes it rotation and scale invariant.

A. Estimation of Candidate Regions

In the first stage of the algorithm, we estimate Gaussian (K) and Mean curvature (H) values of the 3D facial surface, by computing the principal curvatures of the range image. Principle curvatures are found with the method proposed by Besl and Jain [16], by sliding a 3×3 window on the range image and using a set of discrete orthogonal polynomials to fit a quadratic surface to the contents of each window. The coefficients that minimize the mean squared error are used

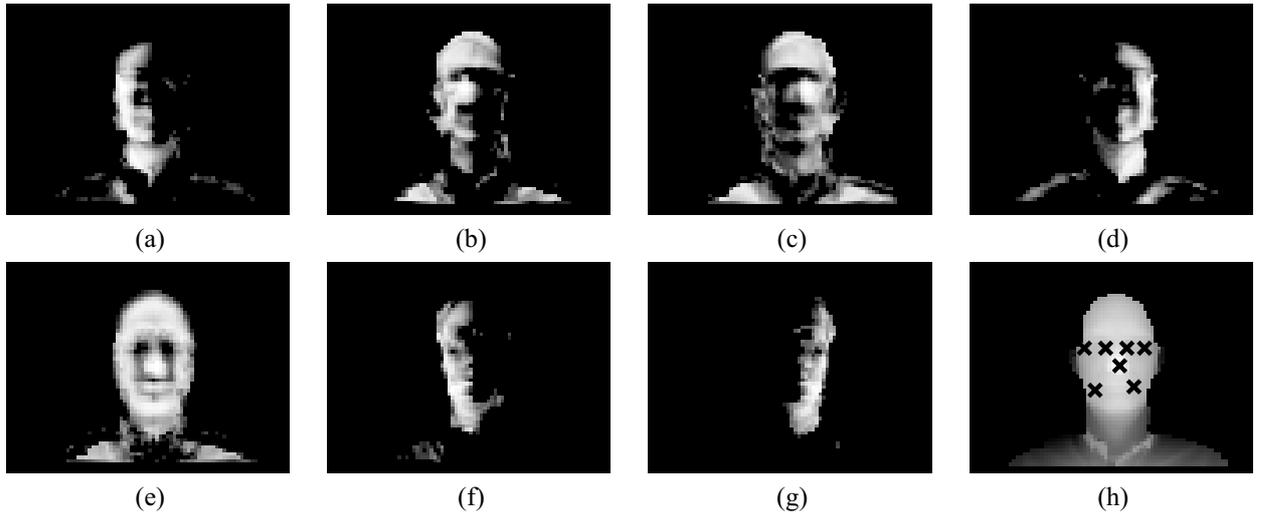


Fig. 1. (a)-(g) The likelihood maps for the individual landmarks (left eye out, left eye in, right eye in, right eye out, nose tip, mouth left, mouth right) where brighter values denote higher likelihoods. (h) The locations with highest likelihood for each landmark are the coarse-scale localizations.

TABLE I
CURVATURE ANALYSIS OF FACIAL SURFACE BASED ON HK MAP COMPONENTS.

| | $K < 0$ | $K = 0$ | $K > 0$ |
|---------|---------------|---------|------------|
| $H < 0$ | Saddle Ridge | Ridge | Peak |
| $H = 0$ | Minimal | Flat | Impossible |
| $H > 0$ | Saddle Valley | Valley | Pit |

to estimate the first and second partial derivatives. After this process, the curvatures are used to determine the fundamental elements on the facial surface by calculating the HK map. Mentioned elements and HK map details can be seen in Table I.

The HK map can be conceived of as a label map, and as providing a discretization of the facial surface. Although this kind of a labeling would be useful to restrict the search area for statistical methods, they entail a certain loss of information. We base our approach on the more informative soft values of H and K maps.

We first produce a difference map \mathcal{D} , such that

$$\mathcal{D}_{i,j} = K_{i,j} - H_{i,j} \quad (1)$$

Fig. 2 shows the difference map for one sample. To



Fig. 2. The mean and Gaussian curvature maps, and the difference map \mathcal{D} , respectively.

overcome the sensitivity of curvature values to surface irregularities, the difference map is next smoothed by a Gaussian filter. The filter size is experimentally set to 15×15 pixels, with one standard deviation corresponding to three pixels. To further reduce noise and the adverse effect of outliers (and depth artifacts in particular) the smoothed values are mapped into a logarithmic scale:

$$\mathcal{D}'_{i,j} = \ln(\mathcal{D}_{i,j} + 1) \quad (2)$$

This transformation accentuates details on the facial surface, and compresses depth differences that are removed from the camera. Finally, the map is min-max normalized to the range $[0, 1]$, and thresholded. The normalization helps in defining a single threshold to determine possibly relevant locations. A larger threshold reduces the search area, but increases the probability of missing the nose. Therefore, it is advisable to keep the threshold value low. In what follows, this value is set to 0.05. A binary image is produced after thresholding.

B. Narrowing down the candidate regions

We combine two indicators to narrow down the candidate regions for the nose. First we compose a change map \mathcal{C} of the face image I by:

$$\mathcal{C} = m(\ln \nabla(I))^2 \quad (3)$$

where ∇ is the gradient operator, and m denotes the min-max normalization function:

$$m(x) = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (4)$$

The shape index (SI) of the face image is the quantitative measure of the shape of a surface at a point [17], and it is defined as:

$$S_I(p) = \frac{1}{2} - \frac{1}{\pi} \tan^{-1} \frac{\kappa_1(p) + \kappa_2(p)}{\kappa_1(p) - \kappa_2(p)} \quad (5)$$

where $\kappa_1(p)$ and $\kappa_2(p)$ are maximum and minimum principal curvature values at point p , respectively. Then, the following point P_c is a relatively close point to the nose tip:

$$P_c = \frac{1}{ij} \sum_i \sum_j \mathcal{C}_{i,j} \mathcal{S}'_{i,j} \quad (6)$$

which stands for the weighted centroid of points described by the Gaussian smoothed shape index \mathcal{S}' (See Fig. 3) multiplied by the corresponding change map value at each point.



Fig. 3. (a) Shape index (b) Smoothed shape index. Lighter colors denote higher values.

A circular region is cropped with a dynamic radius r around P_c (See Fig. 4). The radius is selected to encompass all remaining components found after thresholding \mathcal{D}' . For this purpose, the three-dimensional covariance matrix of the 3D points is computed, and its eigenvectors and eigenvalues are derived. The ratio of eigenvalues is used to scale r to include roughly the same proportion of data, regardless of scale.

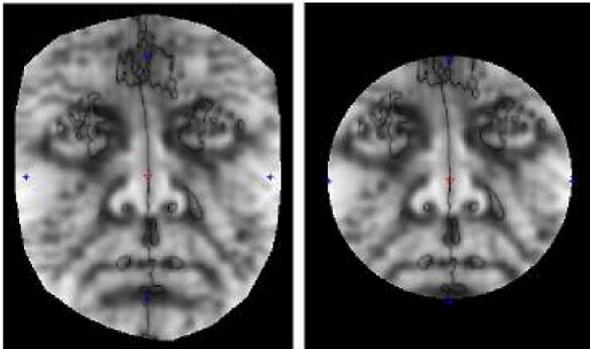


Fig. 4. Cropping a circular region around P_c .

After narrowing the interest region thus, small candidate region components are eliminated, components are connected via dilation and erosion operations. This processing step is shown in Fig. 5.

C. Selecting the nose tip region

Once we are left with a small number of candidates to contain the nose region, we update P_c by the weighted centroid of the smallest circular region that contains all

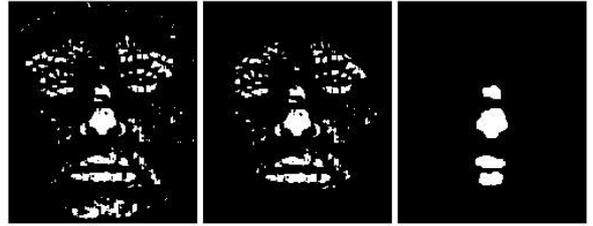


Fig. 5. Initial, cropped and pre-processed candidate regions, respectively.

components. The squared shape indices are used to give weight to each point within this circle. Then, one component is singled out to contain the nose tip by a voting process, where the roundness of the region, prominence of protrusion and the distance to the centroid are taken into account. Nose tip is the point with highest protrusion within the nose region, so we use the SI and \mathcal{D} maps to compose a protrusion map P which is defined as:

$$P_{i,j} = SI_{i,j} \mathcal{D}_{i,j} \quad (7)$$

The weighted centroid coordinates of the protrusion map of selected region gives the nose tip. Fig. 6 shows the located nose tips in various samples from BOS database.

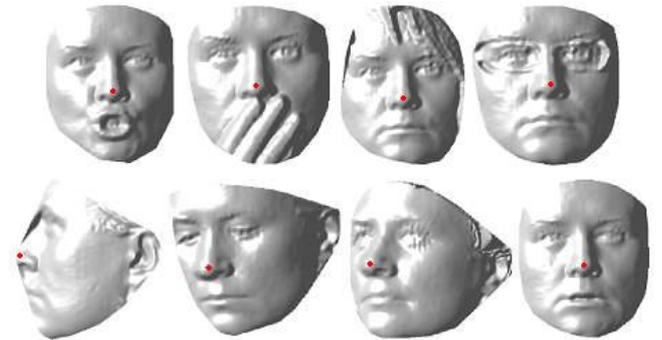


Fig. 6. Correctly localized nose tips in various samples from BOS database.

IV. EXPERIMENTAL RESULTS

A. Database

We have used the recently collected Bosphorus 3D face database in our experiments [18], [19]. The subject variation in this database incorporates various expressions, poses, but also realistic occlusions like eye glasses, hair tassel and eye rubbing. The pose variations take rotations systematically into account. The facial expressions include six universal expressions (happiness, surprise, fear, sadness, anger and disgust), as well as expressions based on facial action units (AU) of the Facial Action Coding System (FACS) [20]. This group is composed of three sets: i) 20 lower face AUs, ii) five upper face AUs and iii) three AU combinations. In Fig. 7 (a), subject variations containing facial expressions, rotations, and occlusions are shown. This database differs from others with its ground-truthed action units. For more

realistic expressions, professional actors and actresses have been enrolled in the Bosphorus database. The manually located landmarks are shown in Fig. 7 (b).

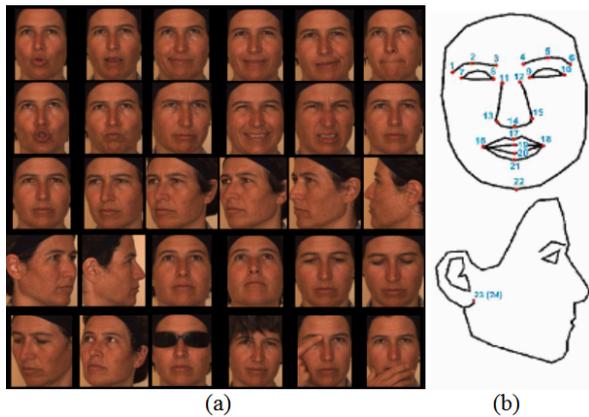


Fig. 7. (a) Subject variations containing facial expressions, rotations, and occlusions. (b) Manually located landmark points for the Bosphorus database.

The Bosphorus database contains 3396 facial scans of 81 subjects. There are 51 men and 30 women in the database. Majority of the subjects are Caucasian and aged between 25 and 35. The database has two parts: the first part (denoted BOSv1), contains 34 subjects with 31 scans each: 10 types of expressions, 13 different poses, four occlusions, and four neutral/frontal scans. The second part (BOSv2) has more expression variations, with 47 subjects having each 53 scans. Each subject has 34 scans for different expressions, 13 scans for pose variations, four occlusions and one or two frontal/neutral face. 30 of these 47 subjects are professional actors/actresses.

Additionally, we have tested the cross-database performance of the statistical approach with the help of the FRGC database [21]. We have used FRGCv1 that contains 943 face scans from 275 subjects, separated into training, validation (i.e. to control model complexity) and test sets of equal size. FRGCv1 contains frontal and neutral faces. An extension of experiments on FRGCv2 is being conducted.

B. Statistical Method and Cross-Database Evaluation

Statistical modeling of facial features can be a blessing and a curse. On the one hand, a statistical approach and a general treatment of the local features for localizing landmarks obviates the need for designing heuristics for each landmark. On the other hand, the local features may depend on acquisition and preprocessing for a particular database, and cross-database performance of the statistical method may be poor. In this section, we apply the statistical models obtained on the Bosphorus database to the FRGC database (and vice versa) to assess the generalization capabilities of such models. See [2] for further tests with this approach.

Table II shows the correct localization accuracies for different training and test conditions. The threshold of accepted landmarks is set to at most three pixels distance in the coarse

TABLE II
CROSS-DATABASE ACCURACY OF THE STATISTICAL APPROACH.

| Database | | Landmark Type | | | |
|----------|--------|---------------|----------|-------|------------|
| Training | Test | Out. Eye | Inn. Eye | Nose | Mouth Cor. |
| FRGCv1 | FRGCv1 | 90.09 | 99.55 | 99.10 | 87.16 |
| BOSv1 | BOSv1 | 95.05 | 98.53 | 99.58 | 67.26 |
| FRGCv1 | BOSv1 | 78.00 | 99.47 | 99.79 | 48.42 |
| BOSv1 | FRGCv1 | 70.10 | 96.28 | 95.61 | 88.12 |
| BOSv1 | BOSv2 | 92.83 | 100.00 | 99.87 | 63.33 |

scale. Four different landmark types are taken into consideration, for which we have manually labeled ground truth in both databases. For these experiments, images with pose variations and occlusions were not taken into consideration. Thus, 475 images on BOSv1 and 1576 images on BOSv2 were used. Nonetheless, different acquisition conditions and novel expression variations make this a challenging task.

The first two lines of Table II show that the mouth corners are the most difficult landmarks to localize, and BOSv1 is more difficult in comparison to FRGCv1, as it contains expression variations. As for the remaining landmarks; the accuracy is close to 100 percent for the nose tip and the inner eye corners, and above 90 per cent for the outer eye corners. Lines 3-4 of the table show the cross-database accuracy of the method. While the nose tip and the inner eye corners are well-localized, there is a marked reduction in the outer eye corners (close to 20 per cent). One reason for this decrease is that the outer eye corners are more effected by expression deformations in BOSv1. For the mouth corners, training with BOSv1 and testing with FRGCv1 does not decrease the accuracy, as FRGCv1 has only closed-mouth images. However, the opposite is not true; training on FRGCv1 results 20 per cent accuracy loss for the mouth corners on BOSv1. The last line of Table II shows the result of training on BOSv1 and testing on BOSv2, which has the same acquisition and preprocessing conditions with BOSv1, but with different subjects. The accuracy is close to within-database performance, demonstrating that the statistical method works well as long as acquisition conditions are stable.

C. Curvature Based Localization

Table III shows the results obtained with the curvature-based heuristic approach (shown by MC) on BOSv1 and BOSv2 datasets. As a second approach, a simple nose tip heuristic from the literature is used (shown by MD); namely, to take the point with the highest depth value as the tip of the nose. The reported accuracies are obtained by accepting points within 10 per cent inter-eye distance to the ground truth. The results show that the nose tip is localized with good accuracy even under strong rotations and occlusions. The simple nose tip heuristic and the statistical methods cannot operate under these conditions.

V. CONCLUSIONS AND FUTURE WORK

We have evaluated a statistical and a heuristics-based method to detect facial landmarks on 3D range images. The statistical approach is useful for frontal faces, to detect

TABLE III

THE ACCURACY OF THE CURVATURE-BASED METHOD (MC) AND THE DEPTH-BASED NOSE TIP HEURISTIC (MD) ON BOSV1 AND BOSV2 DATASETS.

| Type: | BOSv1 | | | BOSv2 | | |
|--------------------------------------|---------------|---------------|-----------|---------------|---------------|-----------|
| | MC - Accuracy | MD - Accuracy | # Samples | MC - Accuracy | MD - Accuracy | # Samples |
| Neutrals and Facial Expressions | 100.00 | 90.11 | 475 | 100.00 | 91.56 | 1576 |
| +Yaw Rotations (10,20,30,45 degrees) | 87.41 | 25.19 | 135 | 95.74 | 27.13 | 188 |
| +90 degrees Yaw Rotation | 73.53 | 0.00 | 34 | 97.87 | 0.00 | 47 |
| -45 degrees Yaw Rotation | 76.47 | 8.82 | 34 | 93.62 | 10.64 | 47 |
| -90 degrees Yaw Rotation | 50.00 | 0.00 | 34 | 95.74 | 2.13 | 47 |
| Strong and Slight Upwards | 97.06 | 57.35 | 68 | 98.94 | 61.70 | 94 |
| Strong and Slight Downwards | 97.06 | 41.18 | 68 | 98.94 | 54.26 | 94 |
| -20 degrees Pitch and +45 Yaw | 85.29 | 0.00 | 34 | 100.00 | 2.13 | 47 |
| +20 degrees Pitch and +45 Yaw | 91.18 | 5.88 | 34 | 100.00 | 13.04 | 46 |
| Occlusion of Eye with Hand | 91.18 | 23.53 | 34 | 100.00 | 21.28 | 47 |
| Occlusion of Mouth with Hand | 94.12 | 11.76 | 34 | 100.00 | 14.89 | 47 |
| Eye Glasses Occlusion | 93.94 | 69.70 | 33 | 100.00 | 68.09 | 47 |
| Hair Occlusion | 100.00 | 34.78 | 23 | 100.00 | 34.48 | 29 |

different landmarks with a single method. We demonstrate through cross-database studies that a statistical method can generalize to different acquisition and preprocessing conditions successfully, but not to different expression (or pose and occlusion) conditions.

For these reasons, we propose a heuristic method that is tailored specifically for the nose tip, and designed to operate under severe pose, occlusion and expression variations. This approach can serve as a starting point for pose normalization and more thorough landmarking, be it statistical or heuristics-based. We demonstrate the success of the approach on the challenging Bosphorus 3D face database.

Currently, more extensive experiments with both Bosphorus and the full FRGC databases are being conducted. The statistical method is being extended for handling 22 landmarks, and high-resolution refinements are being implemented, as well as combinations of the proposed methods for robust operation.

REFERENCES

- [1] A. A. Salah, H. Çınar-Akakin, L. Akarun, and B. Sankur, "Robust facial landmarking for registration", *Annals of Telecommunications*, vol. 62, no. 1-2, pp. 1608–1633, 2007.
- [2] A. A. Salah and L. Akarun, "3D facial feature localization for registration", *Proc. Int. Workshop on Multimedia Content Representation, Classification and Security LNCS*, vol. 4105, pp. 338–345, 2006.
- [3] C. Xu, T. Tan, Y. Wang, and L. Quan, "Combining local features for robust nose location in 3D facial data", *Pattern Recognition Letters*, vol. 27, no. 13, pp. 1487–1494, 2006.
- [4] A.S. Mian, M. Bennamoun, and R. Owens, "An efficient multimodal 2D-3D hybrid approach to automatic face recognition", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 11, pp. 1927–1943, 2007.
- [5] J. D'Hose, J. Colineau, C. Bichon, and B. Dorizzi, "Precise Localization of Landmarks on 3D Faces using Gabor Wavelets", *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on*, pp. 1–6, 2007.
- [6] E. Akagündüz and İ. Ulusoy, "3D Face Representation Using Scale and Transform Invariant Features", *Proc. 16th IEEE Signal Processing, Communication and Applications Conference*, 2008.
- [7] A.A. Salah, N. Alyüz, and L. Akarun, "Registration of three-dimensional face scans with average face models", *Journal of Electronic Imaging*, vol. 17, pp. 011006, 2008.
- [8] M.O. İrfanoğlu, B. Gökberk, and L. Akarun, "3d shape-based face recognition using automatically registered facial surfaces", vol. 4, pp. 183–186, 2004.
- [9] P. Li, B.D. Corner, and S. Paquette, "Automatic landmark extraction from three-dimensional head scan data", *Proceedings of SPIE*, vol. 4661, pp. 169, 2002.
- [10] D. Colbry, G. Stockman, and A.K. Jain, "Detection of anchor points for 3D face verification", 2005.
- [11] C. Conde, A. Serrano, L.J. Rodríguez-Aragón, and E. Cabello, "3D facial normalization with spin images and influence of range data calculation over face verification", 2005.
- [12] I. Kakadiaris, G. Passalis, G. Toderici, N. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis, "3D face recognition in the presence of facial expressions: an annotated deformable model approach", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 640–649, 2007.
- [13] Z. Ghahramani and G.E. Hinton, "The EM algorithm for mixtures of factor analyzers", *University of Toronto Technical Report CRG-TR-96-1*, 1997.
- [14] A.A. Salah and E. Alpaydin, "Incremental Mixtures of Factor Analysers", *Proceedings of the Pattern Recognition, 17th International Conference on (ICPR 04) Volume 1-Volume 01*, pp. 276–279, 2004.
- [15] M.A.T. Figueiredo and A.K. Jain, "Unsupervised Learning of Finite Mixture Models", *IEEE Trans. Pattern Analysis and Machine Intelligence*, pp. 381–396, 2002.
- [16] Paul J. Besl and Ramesh C. Jain, "Invariant surface characteristics for 3d object recognition in range images", *Comput. Vision Graph. Image Process.*, vol. 33, no. 1, pp. 33–80, 1986.
- [17] C. Dorai and A.K. Jain, "COSMOS - A Representation Scheme for 3D Free-Form Objects", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 10, pp. 1115–1130, 1997.
- [18] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, "Bosphorus database for 3D face analysis", in *Proceedings of the First COST 2101 Workshop on Biometrics and Identity Management (BIOD)*, Denmark, May 2008.
- [19] N. Alyüz, B. Gökberk, H. Dibeklioğlu, A. Savran, A.A. Salah, L. Akarun, and B. Sankur, "3D face recognition benchmarks on the bosphorus database with focus on facial expressions", in *Proceedings of the First COST 2101 Workshop on Biometrics and Identity Management (BIOD)*, Denmark, May 2008.
- [20] P. Ekman and W.V. Friesen, *Facial action coding system*, Consulting Psychologists Press, Palo Alto, Calif., 1978.
- [21] 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", vol. 1, pp. 947–954, 2005.