# Nasal Region-Based 3D Face Recognition under Pose and Expression Variations

Hamdi Dibeklioğlu[1], Berk Gökberk[2], and Lale Akarun[1]

[1] Boğaziçi University, Computer Engineering Department, Turkey
{hamdi.dibeklioglu,akarun}@boun.edu.tr
[2] University of Twente, Department of Electrical Engineering, The Netherlands
b.gokberk@utwente.nl

**Abstract.** In this work, we propose a fully automatic pose and expression invariant part-based 3D face recognition system. The proposed system is based on pose correction and curvature-based nose segmentation. Since the nose is the most stable part of the face, it is largely invariant under expressions. For this reason, we have concentrated on locating the nose tip and segmenting the nose. Furthermore, the nose direction is utilized to correct pose variations. We try both one-to-all and Average Nose Model-based methodologies for registration. Our results show that the utilization of anatomically-cropped nose region increases the recognition accuracy up to 94.10 per cent for frontal facial expressions and 79.41 per cent for pose variations in the Bosphorus 2D/3D face database.

## 1 Introduction

An effective face recognition system has to be fully automatic and robust enough for real life conditions where illumination, rotation, and expression variations are present. Although state-of-the-art 3D face recognition systems provide identification rates up to 99 per cent, they are mostly constrained by certain assumptions such as slight expression and pose variations [1]. Overcoming the problems due to *extreme* expression and pose changes still remains as a challenging problem.

A recently proposed approach for expression invariant 3D face recognition is a part-based system [2]. Part-based approach is useful to alleviate the pose, facial expression and partial occlusion effects on the recognition performance. To deal with these effects, the entire 3D model of the face is split into several regions and these regions are processed separately. Alyüz *et al.* [2] split the face into patches, and carry out an exhaustive search of all possible combinations of these surface to find the best subset of all the patches around the whole facial surface.

Moreno *et al.* [3] segment the 3D facial surface using signs of mean, Gaussian curvatures and several three dimensional descriptors. Cook *et al.* [4] use Log-Gabor Templates on range images to deal with occlusions, distortions and facial expressions. Recently, Kakadiaris et al. proposed matching of spin images prior to ICP for alignment. They used wavelet analysis for feature extraction and obtained good recognition results in the presence of facial expressions [5]. In [6], Chang *et al.* use multiple regions selected around the nose area which have the

least shape variation due to facial expressions. In [7] Faltemier *et al.* extend the use of multiple regions. Seven overlapping regions around the nose are extracted and the ICP algorithm [8] is used to align each segment separately. On the FRGC v.2 database, a rank-one accuracy of 94.9 per cent is reported using a voting mechanism. In their later work, Faltemier *et al.* [9] use 38 overlapping regions to find the best committee. By selecting multiple small regions on the face, any errors caused by a single region can be compensated for when combining the matching scores. These regions are extracted by spherical cropping of facial surface and aligned with the ICP. The best performance is provided by the fusion of 28 regions with a modified Borda count method, and the recognition accuracy is reported as 97.2 per cent on the FRGC v.2.

Most of these part-based systems first localizes some salient feature points over the facial surface such as nose tip and eye pits for component formation. There is a very broad range of facial landmark localization algorithms, ranging from statistical techniques to heuristics. For a comprehensive overview, the reader is referred to a recent survey [1]. In this paper we propose an automatic nose tip localization and nose segmentation method that can be applied to cases where significant pose variations are present. Extreme rotations create difficulties both for nose tip localization and nose segmentation because a significant part of the nose is missing due to self-occlusion. Our recognition scheme achieves significant success even under these conditions.

## 2   Nose Tip Estimation and Nose Segmentation

We propose a heuristic method to localize the nose tip and segment the nose region under pose and expression variations. The proposed algorithm estimates the regions with high protrusion and selects the most probable location based on curvature values. Principle normal curvatures are used since they are invariant to rotation and translation. We use an analytical method to estimate the principal curvatures by fitting a quadratic order surface. Details of the curvature estimation algorithm are given in [2].

### 2.1   Initial Estimation of the Nose Tip

An initial estimation of nose tip is required for segmentation of the nose region. First, Gaussian ($K$) and mean curvature ($H$) values of the 3D facial surface are estimated from the the principal curvatures as: $K = \kappa_1(p) \times \kappa_2(p)$, $H = \frac{\kappa_1(p) + \kappa_2(p)}{2}$ where, $\kappa_1$ and $\kappa_2$ denote the minimum and maximum principal curvatures, respectively. In theory, a $HK$ map can be comprehended as a label map. Although this kind of a labeling would be useful to restrict the search area for statistical methods, it entails a certain loss of information. Our approach is based on the more informative continuous values of $H$ and $K$ maps [10]. We compose a difference map $\mathcal{D}_{i,j} = K_{i,j} - H_{i,j}$. In the proposed $\mathcal{D}$ map, the nose tip is revealed as a peaked cap. Then $\mathcal{D}$ map is smoothed by a Gaussian filter to cope with curvature sensitivity and surface noise. However, the peaks on

the sides of the nose may be confused with the nose tip when there are large rotations. Therefore, we need a rotation invariant feature which yields higher values around the nose tip. To handle the rotation conditions, we estimate an Eigen map, $\mathcal{E}$. A sliding $3 \times 3$ pixels block is used for $\mathcal{E}$ map estimation. 3D coordinates of nine points on each block are taken as inputs and their covariance matrix is calculated. Then the eigenvalues and eigenvectors are estimated for the calculated covariance matrix. The biggest eigenvalue is selected and the ratio of the selected eigenvalue to sum of all eigenvalues is calculated. This is the value of the $\mathcal{E}$-map for the center point of the related block. Estimated $\mathcal{E}$-map gives relatively higher values for points with steep depth changes such as points on nose slopes. Additionally, we know that $\mathcal{D}$ has higher values for ridges and peaks. Eventually, all the nose region can be highlighted by the combination of these two indicators (see Fig. 1). A combination map $\mathcal{C}$ is formed such that:

$$\mathcal{C}_{i,j} = \frac{\mathcal{D}'_{i,j} + \mathcal{E}'_{i,j}}{2} \tag{1}$$

where $\mathcal{D}' = m(\mathcal{D})$, $\mathcal{E}' = m(\mathcal{E})$, and $m$ denotes the min-max normalization function $m(x) = \frac{x - min(x)}{max(x) - min(x)}$.
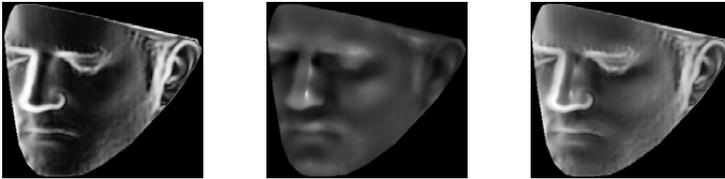


**Fig. 1.** The eigen, difference and the combination maps, respectively

We crop the difference and combination maps by 20 per cent on the top and bottom to focus on the center of the face. Left and right sides of the $\mathcal{D}$ map are eliminated by using $\mathcal{C}$ map. The vertical projection of the squared $\mathcal{D}$ map reveals a peak which corresponds to the nose region. We calculated the shape index $(SI)$ in this interest region to locate the nose tip precisely. The shape index of the face image is the quantitative measure of the shape of a surface at a point [11], and it is defined as:

$$SI(p) = \frac{1}{2} - \frac{1}{\pi} tan^{-1} \frac{\kappa_1(p) + \kappa_2(p)}{\kappa_1(p) - \kappa_2(p)} \tag{2}$$

Then the interest region is thresholded by the mean value of the whole shape index map. Lastly, horizontal projection of the $\mathcal{D}$ map is calculated for remaining points and the point with the maximum value is selected as a close point to the nose tip.

## 2.2   Coarse Nose Cropping

After the estimation of the nose tip, points on the initial $\mathcal{D}$ map are cropped by a rectangular mask. Then, the minimum left and right side $\mathcal{D}$ values are found which are on the same y-coordinate with the nose tip. These minimum points are used as starting points for nose cropping by greedy search. Coarse cropping is based on finding the minimum energy path on the $\mathcal{D}$ map, by greedy search, as shown in Fig. 2.
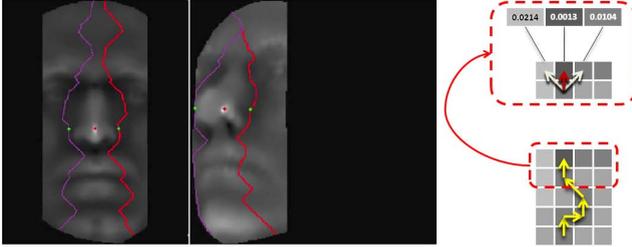


**Fig. 2.** Minimum energy path estimation: Green points show the initial minimum points. Red points denote the estimated nose tip.

## 2.3   Yaw Rotation Estimation and Pose Correction

Yaw rotations have to be estimated to rotate the face into a frontal pose. It can be assumed that the relative horizontal position of the nose tip gives a coarse estimation of the yaw rotation angle. Hence, normalized nose tip positions for subjects are determined relative to the leftmost valid point of the facial surface. In other words, minimum $x$ coordinate of the facial surface is subtracted from the $x$ coordinate of the nose tip, and the ratio of the calculated value to the width of the facial surface gives the normalized nose tip position. Since the nose tip position of the frontal poses is approximately at the middle, its value is 0.5. 90 degrees to the left and 90 degrees to the right rotated surfaces' nose tip positions are 0 and 1, respectively. Then, the estimated yaw angle is used to correct the rotation and to bring the facial points to a frontal pose. It should be noted that rotation compensation using nose tip is just a preliminary step. In the following registration steps, the ICP algorithm is utilized for more accurate pose normalization.

## 2.4   Nose Tip Correction and Fine Level Cropping

After coarse rotation compensation phase, $\mathcal{D}$ map is generated again, as explained in 2.1. Rotated nose region is cropped vertically in order to eliminate artifacts that may emerge around the previously invisible side of the nose. We prefer to eliminate these regions since interpolation is almost impossible considering the amount of missing surface data. In the next step, minimum energy
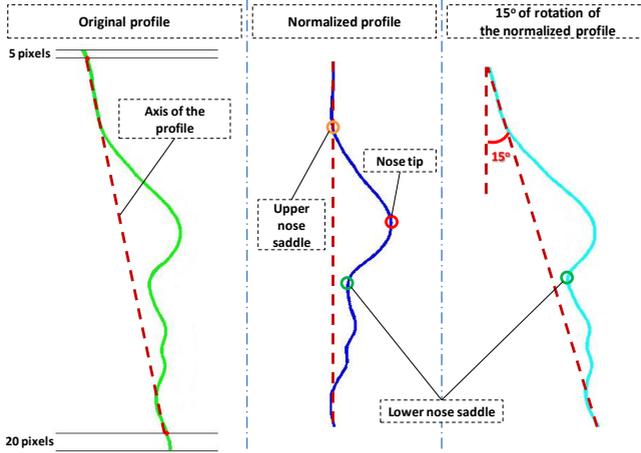
**Fig. 3.** Upper and lower limits estimation procedure of the nose region

paths are refined by updating the starting points and repeating the search. At the last step, upper and lower bounds of the nose are found as illustrated in Fig. 3. The profile is normalized, rotated, and local minima which corresponds to the upper and lower limits of the nose are found.

## 3  Registration and Recognition Methodology

Our registration method is based on the ICP algorithm. We compare two different approaches, namely, *one-to-all ICP* and Average Nose Model-based ICP (*ANM-based ICP*). In the one-to-all ICP approach, probe nose regions are registered to each gallery nose. In the ANM-based method, all of the gallery nose regions are registered to an ANM by the ICP. ANM is a generic 3D nose model computed from the gallery set. A modified version of the average face model construction algorithm presented in [12] is used to generate the ANM. Given all gallery nose regions, they are aligned to a consensus shape using manually located five fiducial points by the Procrustes analysis. After surface resampling, average surface is determined by taking the mean of depth coordinates. Details of the algorithm can be found in [12]. In the ANM-based approach, a probe nose is registered to the ANM at the recognition phase which provides dense point-to-point correspondence to every gallery nose. After registration, we resample the final nose surfaces by Delaunay trangulation-based interpolation and employ point set difference approach for dissimilarity calculation. If $\mathcal{N}$ and $\mathcal{N}'$ are the re-sampled depth maps of a registered training and test nose, respectively, a mean error function, $E_{PSD}$ can be formulated as:

$$E_{PSD} = \frac{1}{G_h \times G_h} \sum_{i=1}^{G_h} \sum_{j=1}^{G_w} w_{i,j} \times \left| \mathcal{N}_{i,j} - \mathcal{N}'_{i,j} \right| \qquad (3)$$

where, $G_h$ and $G_w$ indicate height and width of the re-sampled depth maps, respectively. $w_{i,j}$ denotes the intersection of valid points in $\mathcal{N}$ and $\mathcal{N}'$. $w_{i,j}$ is zero for invalid points and one otherwise.

## 4    Experimental Results

We have used the Bosphorus 3D face database [13] in our experiments. The Bosphorus database is designed specifically to study extreme pose, expression, and occlusion variations (see Fig. 4). Additionally, the database incorporates realistic occlusions like eye glasses, hair tassel and eye rubbing. The Bosphorus database includes $+10°$, $+20°$, $+30°$, $±45°$, $±90°$ of yaw rotations, strong/slight upwards, strong/slight downwards, combinations of $±20°$ pitch and $+45°$ yaw rotations. Facial expressions include six basic emotions (happiness, surprise, fear, sadness, anger and disgust), as well as expressions based on facial action units. Database has two versions as v.1 and v.2. Bosphorus v.1 includes 34 subjects with only 10 expressions, 13 poses, four occlusions and four neutral faces, thus resulting in a total of 31 scans per subject. In version 2, there are 47 people with 53 different face scans per subject. Totally there are 34 expressions, 13 poses, four occlusions and one/two neutral faces. expression and pose variations in the Bosphorus Database are shown in Fig. 4.
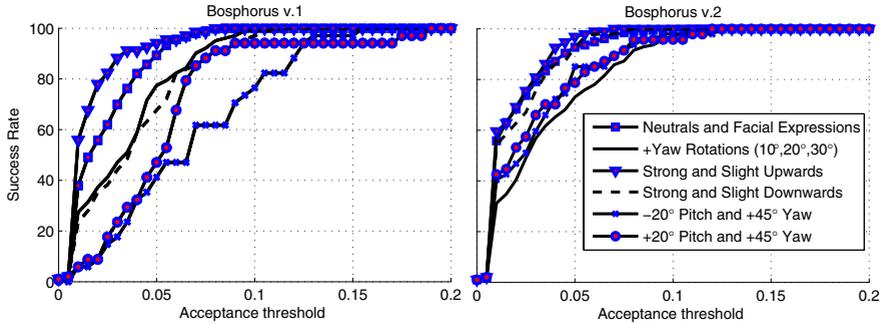


**Fig. 4.** Samples of expression and pose variations in the Bosphorus database

### 4.1    Nose Tip Localization and Segmentation

The proposed nose tip estimator has not been tested for facial images having $45°$ and $90°$ yaw rotations because these samples are not considered in our identification experiments. The reported success rates are obtained by accepting points within 10 per cent inter-eye distance to the ground truth (see Table 1 and Fig. 6). In the Bosphorus database, the average inter-eye distance is 64.9 millimeters. To control the generality of the used heuristics, we have also tested our algorithm on FRGC v.1 and achieved 100 per cent success with 10 per cent inter-eye distance acceptance. The main factors about the robustness and performance of our algorithm are the pose correction step and the refinement of the nose tip location during the segmentation process. Segmented nose regions are shown in Figure 6.

**Table 1.** The performance of the proposed node tip localization method

| Type | Bosphorus v.1 | | Bosphorus v.2 | |
|---|---|---|---|---|
| | Success Rate | # Samples | Success Rate | # Samples |
| Neutrals and Facial Expressions | 100.00 | 475 | 99.68 | 1574 |
| +Yaw Rotations (10°, 20°, 30°) | 99.02 | 102 | 97.16 | 141 |
| Strong and Slight Upwards | 100.00 | 68 | 100.00 | 94 |
| Strong and Slight Downwards | 97.06 | 68 | 100.00 | 94 |
| -20° Pitch and +45° Yaw | 76.47 | 34 | 97.87 | 47 |
| +20° Pitch and +45° Yaw | 94.12 | 34 | 95.75 | 47 |



**Fig. 5.** The performance of the proposed node tip localization method on several acceptance thresholds of inter-eye distance



**Fig. 6.** Segmented nose regions under several expression and pose variations

## 4.2 Identification Experiments

In our identification experiments, the gallery set is composed of one neutral scan per subject. The remaining faces, probe images, are divided into two groups as i) frontal scans and ii) rotated faces. Poses with rotation of $\pm45°$ and $\pm90°$ are discarded in the rotation probe set because these images have insufficient amount of surface information for our identification algorithm. Number of samples in the gallery and probe sets for each version are given in Table 2.

Part-based identification rates are given in Table 3 and Table 4. Two different types of gallery sets are used in frontal tests of one-to-all methods. We refer to the one-to-all ICP and the ANM-based one-to-one ICP methods as $Natural_{one:all}$ and $Natural_{anm}$, respectively. The first one is composed of automatically cropped nose models, denoted as *Original*. The other one includes

**Table 2.** The gallery and probe sets for the Bosphorus v.1 and v.2

| Version | Type | Gallery | Frontal Probe Set | Rotation Probe Set |
|---|---|---|---|---|
| Bosphorus v.1 | Neutral | 34 | 102 | 306 |
| | Expression | - | 339 | - |
| | Total | 34 | 441 | 306 |
| Bosphorus v.2 | Neutral | 47 | 19 | 423 |
| | Expression | - | 1508 | - |
| | Total | 47 | 1527 | 423 |

**Table 3.** Rank-1 recognition rates of the one-to-all ICP approach for the frontal poses with expression in the Bosphorus database (* denotes the highest recognition rates)

| Pose | Method | Gallery Type | Bosphorus v.1 | Bosphorus v.2 |
|---|---|---|---|---|
| Frontal | $Natural_{one:all}$ | *Enlarged* | 92.97 | 88.41 |
| | | *Original* | 94.10* | 89.19* |
| | $Spherical_{one:all}^{P1}$ | *Enlarged* | 73.71 | 65.62 |
| | | *Original* | 89.57 | 80.48 |
| | $Spherical_{one:all}^{P2}$ | *Enlarged* | 85.94 | 78.26 |
| | | *Original* | 91.84 | 81.79 |

**Table 4.** Rank-1 recognition rates for the frontal poses with expression and poses with rotation in the Bosphorus database (* denotes the highest recognition rates)

| Pose | Method | Gallery Type | Bosphorus v.1 | Bosphorus v.2 |
|---|---|---|---|---|
| Frontal | $Natural_{one:all}$ | *Original* | 94.10* | 89.19* |
| | $Natural_{anm}$ | *Original* | 93.88 | 89.19* |
| Rotation | $Natural_{one:all}$ | *Original* | 79.41* | 60.99 |
| | $Natural_{anm}$ | *Original* | 73.20 | 62.55* |

enlarged nose models, denoted as *Enlarged*, which are constructed by dilation with 10 pixels of the bounding boxes of *Original* set.

The region extraction method which is suggested in [9] is also implemented and tested for two different settings for comparative purposes. In this method, the nose is cropped by a sphere with a radius $r$. The center of the sphere is defined by offsets with respect to the tip of the nose, namely x-axis offset $(x_o)$ and y-axis offset $(y_o)$. Nose cropping parameters have been selected as in [9] which provide the best results in terms of rank-one recognition rate in the FRGC v.2. These tests are denoted as $Spherical_{one:all}^{P1}$ and $Spherical_{one:all}^{P2}$, respectively. The related parameters are set as $x_o = 0$ , $y_o = +30$, $r = 40$ millimeters for $Spherical_{one:all}^{P1}$ and $x_o = 0$, $y_o = +20$ and $r = 35$ millimeters for $Spherical_{one:all}^{P2}$. Nose tip coordinates which are used in this method, are provided by our nose tip estimator. One-to-all ICP is used for the registration of the baseline method.

Test results indicate that proposed methods, $Natural_{one:all}$ and $Natural_{anm}$ with *Original* gallery type, provide the highest recognition rates for the frontal

expressions in the Bosphorus v.1, 94.10 per cent and in the Bosphorus v.2, 89.19 per cent. The nose region which is segmented by using its natural boundaries is more discriminative than the point set of the manually cropped coarse nose region. As explained above, $Spherical_{one:all}^{P1}$ and $Spherical_{one:all}^{P2}$ use nose models which also covers an extra portion of the facial surface outside of the nose region. Because of that, the recognition success rates of the mentioned methods are degraded. If we compare the results for *Original* and *Enlarged* types of gallery set usages, it can be seen that *Enlarged* gallery set usage degrades the accuracy because of the same reason.

According to the results of the poses with rotations, the best rates achieved by $Natural_{one:all}$ on the Bosphorus v.1 and $Natural_{anm}$ on the Bosphorus v.2 are 79.41 per cent and 62.55 per cent, respectively. The rates are obviously decreased with respect to frontal poses by rotation variations. However the performance of the system is accurate enough for rotation conditions. It is needed to be noted that a very small part of a nose can be extracted under rotation variations. One final observation is related to one-to-one versus one-to-all methods: there is not a big performance difference between them despite very high computation complexity of the one-to-all method, i.e, registering a probe image with every gallery image without the use of the ANM.

## 5   Conclusion

A fully automatic part-based 3D face recognition system has been proposed in this work. Curvature-based nose localization/segmentation and part-based registration methods have been used to cope with rotations and facial expressions. The nose tip estimator is a curvature based heuristic method which includes pose correction. The results show that the nose tip is localized with a very good accuracy even under strong rotations and expressions. Our nose segmentation algorithm automatically finds the nose borders. According to the experimental results, recognition rates are significantly increased by using natural regions of the noses as opposed to enlarged nose regions. This is due to i) the shape of the segmented nose is discriminative and ii) the muscle deformations around the nose region increases the intra-class variations under expressions. Therefore fine segmentation of the nose region is crucial for identification. In our experiments, we show that high recognition performances can be attained under facial expression and rotation variations by a part-based method. Using only nose region, 94.10 and 79.41 per cent recognition accuracies can attained for expression and pose variations, respectively. As future work, the proposed system should be extended for pose variations larger than 45 degrees and other facial parts should also be consulted under such conditions.

## References

1. Bowyer, K.W., Chang, K., Flynn, P.: A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition. Computer Vision and Image Understanding 101, 1–15 (2006)

2. Alyüz, N., Gökberk, B., Dibeklioğlu, H., Akarun, L.: Component-based registration with curvature descriptors for expression insensitive 3D face recognition. In: International Conference on Automatic Face and Gesture Recognition (2008)
3. Moreno, A.B., Sanchez, A., Velez, J.F., Diaz, F.J.: Face recognition using 3D surface-extracted descriptors. In: Irish Machine Vision and Image Processing Conference (IMVIP) (2003)
4. Cook, J., Chandran, V., Fookes, C.: 3D face recognition using log-gabor templates. In: Biritish Machine Vision Conference, pp. 83–92 (2006)
5. Kakadiaris, I., Passalis, G., Toderici, G., Murtuza, N., Lu, Y., Karampatziakis, N., Theoharis, T.: 3D face recognition in the presence of facial expressions: an annotated deformable model approach. IEEE Trans. Pattern Analysis and Machine Intelligence 29, 640–649 (2007)
6. Chang, K.I., Bowyer, K.W., Flynn, P.J.: Adaptive rigid multi-region selection for handling expression variation in 3D face recognition. In: IEEE CVPR 2005, pp. 157–164 (2005)
7. Faltemier, T., Bowyer, K.W., Flynn, P.J.: 3D face recognition with region committee voting. In: Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT 2006), pp. 318–325 (2006)
8. Besl, P., McKay, N.: A method for registration of 3-D shapes. IEEE Trans. Pattern Analysis and Machine Intelligence 14, 239–256 (1992)
9. Faltemier, T., Bowyer, K., Flynn, P.: A region ensemble for 3-d face recognition. IEEE Transactions on Information Forensics and Security 3, 62–73 (2008)
10. Dibeklioğlu, H., Salah, A.A., Akarun, L.: 3D facial landmarking under expression, pose, and occlusion variations. In: IEEE BTAS 2008, Washington, DC, USA (2008)
11. Dorai, C., Jain, A.K.: Cosmos - a representation scheme for 3d free-form objects. IEEE Trans. PAMI 19, 1115–1130 (1997)
12. İrfanoğlu, M.O., Gökberk, B., Akarun, L.: 3D shape-based face recognition using automatically registered facial surfaces. In: International Conference on Pattern Recognition (ICPR 2004), vol. 4, pp. 183–186 (2004)
13. Savran, A., Alyüz, N., Dibeklioğlu, H., Çeliktutan, O., Gökberk, B., Sankur, B., Akarun, L.: Bosphorus database for 3D face analysis. In: The First COST 2101 Workshop on Biometrics and Identity Management (BIOID) (2008)