Faceprint: Fusion of Local Features for 3D Face Recognition

Guangpeng Zhang and Yunhong Wang

School of Computer Science and Engineering, Beihang University
gpzhang@cse.buaa.edu.cn, yhwang@buaa.edu.cn

Abstract. 3D face recognition is a very active biometric research field. Due to the 3D data's insensitivity to illumination and pose variations, 3D face recognition has the potential to perform better than 2D face recognition. In this paper, we focus on local feature based 3D face recognition, and propose a novel Faceprint method. SIFT features are extracted from texture and range images and matched, the matching number of key points together with geodesic distance ratios between models are used as three kinds of matching scores, likelihood ratio based score level fusion is conducted to calculate the final matching score. Thanks to the robustness of SIFT, shape index, and geodesic distance against various changes of geometric transformation, illumination, pose and expression, the Faceprint method is inherently insensitive to these variations. Experimental results indicate that Faceprint method achieves consistently high performance comparing with commonly used SIFT on texture images.

Keywords: 3D face recognition, local feature, fusion.

1 Introduction

As an important biometrics face recognition has a long history, and enormous amount of face recognition algorithms have been proposed. Most of these algorithms are 2D-image-based, that is to identify or verify subjects through their face images captured by cameras or video recorders. A literature survey of these methods can be found in [1]. Despite its popularity, 2D face recognition is sensitive to illumination, pose and expression variations. In order to deal with these limitations, researchers are paying more and more attentions to 3D face recognition [2], as 3D model captures the exact shape of facial surface, thus is invariant to illumination and pose variations. But 3D-based methods are more sensitive to expression variations than 2D-based, and influenced by occlusion. Actually 2D images and 3D models are two different modalities, they are complementary, and fusion of these two modalities may benefit face recognition [2].

Zhao et al. [1] divided still 2D face recognition techniques into holistic, feature-based, and hybrid categories. This categorization can also be extended to 3D face recognition. Holistic method uses the whole face region as the input to face recognition algorithm. Typical holistic methods include PCA [3] and LDA [4] based methods. Most of these popular 2D face recognition methods were also

M. Tistarelli and M.S. Nixon (Eds.): ICB 2009, LNCS 5558, pp. 394–403, 2009.
© Springer-Verlag Berlin Heidelberg 2009

extended to 3D face recognition. Many 3D face recognition methods such as ICP-based matching [22], annotated deformable model [23], isometry-invariant representation [17] also belong to this category. In feature-based methods, local (geometric or appearance) features are extracted from facial feature points or regions such as eyes, nose, and mouth, and fed into classifiers. Many earlier works are feature-based face recognition [5]. They measured distances and angles between fiducial points to do matching. The most successful feature-based method is graph matching [6], which is based on Dynamic Link Architecture (DLA). It's also extended to 3D recognition [7]. Point signature method proposed by Chua et al. [8] belongs to this category. Hybrid methods use both local features and the whole face region to do recognition. Typical methods include modular eigenfaces [10], component-based [11] et al.

Among these three categories, we pay special attention to feature-based methods. This kind of methods extracts features from local points or regions, and has the potential to deal with expression variations and occlusion. As no matter when expression variation or occlusion occurs, there always exists some local points or regions remain invariant. If we can match these invariant points or regions, recognition is done. SIFT is used in this paper to extract robust local features from feature points. SIFT proposed by Lowe [12] is a robust feature extraction and matching method against image scaling, translation, and rotation, and partially invariant to illumination changes and affine or 3D projection. Because of its great property, SIFT spreads rapidly into various applications since proposed, including face recognition. M. Bicego et al [9] first utilized SIFT on 2D face recognition, matching scores of feature descriptors was used for verification. The method was further improved by incorporating graph composed of the detected SIFT points [16]. Mian et al [13] utilized SIFT on 2D texture images together with SFR (Spherical Face Representation) on 3D data to form a rejection classifier. As far as we know, all the proposed method of using SIFT for face recognition are performed on 2D domain (texture images), and none of the results reported in the above publications is very high. We think the main reason for the unsatisfying performance is that, unlike artificial objects such as buildings, face is a smooth surface, and there are not so many distinctive feature points. As reported in [13], the average number of detected feature points is 80. Another reason is that images change greatly with illumination variations, while SIFT is only partially insensitive to illumination variations.

In order to deal with the limitations of utilizing SIFT for face recognition, we resort to 3D data, and shape index is extracted from range images, as shape index is derived from 3D curvatures, so it's invariant to illumination and pose variations. Compared with 2D texture images, shape index images have much more details as shown in Fig. 2(c), thus more feature points can be extracted from them. In order to deal with sensitivity of 3D data to expression variations, geodesic distances in between matched feature points are also computed and matched. The final matching score is obtained by fusion of the number of matched 2D and 3D SIFT feature points, and geodesic distance ratio between matched models using likelihood ratio test strategy. We call this proposed method ”Faceprint”. Just like minutia in fingerprint, Faceprint describes local facial features, and is robust against many kinds of variations.

The paper is organized as follows. SIFT, shape index and geodesic features are briefly introduced in section 2, and then Faceprint method is described in detail in section 3. Experiments are carried out on FRGC v2.0 database [24], and results are shown in section 4. Section 5 concludes the paper.

2 SIFT, Shape Index and Geodesic distance

2.1 Scale Invariant Feature Transform

SIFT (Scale Invariant Feature Transform) was proposed by Lowe [12], [14] for object recognition. The original image is repeatedly convolved with Gaussians of different scales separated by a constant factor k to produce an octave in scale space. Once an octave is processed, the coarsest image of this octave is down sampled by 4 to be the start image of next octave. After convolution with Gaussian, neighboring images in each octave are subtracted to get DOGs (Difference of Gaussian). DOG function provides a close approximation to scale-normalized Laplacian of Gaussian [15]. Extremas are detected by comparing a pixel with its 26 neighbors in $3 * 3$ regions at the current and adjacent scales.

$$G(x, y, k\sigma) - G(x, y, \sigma) \approx (k-1)\sigma^2\nabla^2 G \tag{1}$$

After detection of extremas, a threshold is used to eliminate key points with low contrast. Later key points that have a ratio between the principle curvatures greater than a threshold are removed, because these points are located along edges and poorly defined.

For each key point, an orientation histogram is calculated from the gradient orientations of its neighboring sample points. Each sample is weighted by its gradient magnitude and a Gaussian-weighted circular window. Highest local peak and peak that’s within 80% of the highest peak are detected, and the corresponding orientations are assigned to the key point.

For each orientation of a key point, a feature vector is extracted as a descriptor from the gradients of its neighboring sample points. In order to achieve orientation invariance, the coordinates and gradient orientations or neighboring points are rotated relative to the key point orientation. Then a Gaussian function is used to assign a weight to the gradient magnitude of each point. Points that are close to the key point are given more emphasis than those far from it. Orientation histograms of $4 * 4$ sample regions are calculated, each with eight orientation bins. Thus a feature vector with dimension of $4 * 4 * 8 = 128$ is formed. To achieve illumination insensitive, the feature vector is normalized to unit length and thresholded such that no element is larger than 0.2, and normalized to unit length again. This normalized feature vector is the descriptor of the central key point.

Given the descriptors, matching between key points on two images can be done. A matching of one key point to another is only accepted if the matching distance is less than a threshold t times the distance to the second closest match, t is set to 0.6 in this work. The SIFT based key point detection and matching is done with the help of Lowe's code.

2.2 Shape Index

Shape Index feature was proposed by Dorai et al. [25] to represent surface concave and convex attributes. The Shape Index at point p is defined as:

$$S(p) = \frac{1}{2} - \frac{1}{\pi} \arctan \frac{\kappa_1(p) + \kappa_2(p)}{\kappa_1(p) - \kappa_2(p)} \tag{2}$$

where k_1 and k_2 represent maximum and minimum principal curvatures respectively. As principal curvatures are invariant to pose variations, so is Shape index.

The value of Shape Index lies in the range of $[0, 1]$. Local shape at point p is a spherical cup when $S(p) = 0$, and a spherical cap when $S(p) = 1$. When Shape Index changes from 0 to 1, local shape changes from spherical cup to spherical cap.

Fig. 2(c) shows an example of shape index image, in which dark pixels represent lower shape index, while bright pixels represent higher shape index.

2.3 Geodesic Distance

In the specific case of 3D mesh, geodesic distance between two vertices is defined as the shortest path connecting them. As shown in [17], geodesic distance is invariant to deformation of 3D model caused by expression variations. A numerically consistent algorithm for distance computation on triangulated mesh was proposed by Kimmel and Sethian [18] as a generalization of the fast marching method [19]. Using this method, the geodesic distances between a surface vertex and the rest of the n surface vertices can be computed in $O(n)$ operations. We use this method for computing geodesic distances between SIFT key points on 3D mesh.

3 Faceprint

The flow chart of Faceprint method is shown in Fig. 1. A 3D model of one subject contains one texture image and one range image, which are densely registered. Unlike the careful normalization including face pose correction, resampling, and segmentation performed in reference [13], faces are simply cropped out of the background using a bounding box on both texture and range images, by which computational complexity is reserved. As SIFT is very robust to image scaling, translation and rotation, we can fully benefit from advantages of SIFT, and do not need to do pose correction and resampling. After normalization, the size of the texture and range images is different from one model to another. Histogram equalization is carried out on normalized texture images.

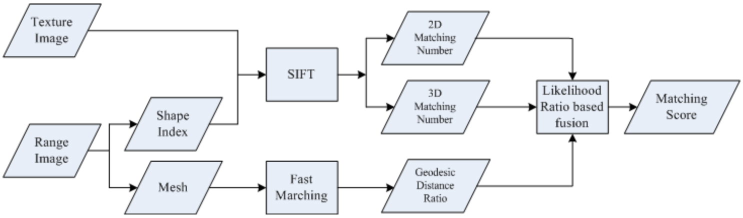

Fig. 1. Flow chart of Faceprint method

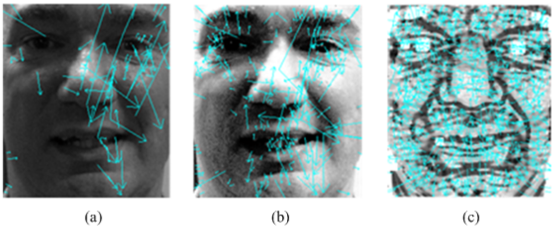

Fig. 2. SIFT features extracted from (a) texture images, (b) histogram equalized texture images, and (c) shape index images

From range images, shape index images as shown in Fig. 2(c) are calculated, SIFT is performed not only on texture images but also on shape index images. Shape index images provide robust features against pose and illuminations, thus repeatability is guaranteed, and these features are more sufficient and with more diversity than those of texture images. In our experiments, the average number of descriptors extracted from shape index images is 862, while that from texture images is 37, and that from histogram equalized texture images is 128. Fig. 2 shows the SIFT features extracted from texture images, histogram equalized texture images, and shape index images, all of which were captured at the same time for the same subject. From Fig. 2, we can see that compared with texture and histogram equalized texture images, shape index image provides the most evenly spaced and largest number of descriptors.

Then the range image is triangulated and mesh is reconstructed. Triangulation is done by dividing each square in the image grid with the shortest diagonal edge. As texture and range images are densely registered, so each detected SIFT key point on texture image corresponds to a vertex on the generated mesh, unless those located in regions such as eyes where no range data is captured. Fast marching is performed for each key point on the mesh, and the geodesic distances between key points are obtained after that.

On matching, detected SIFT key points from texture images and shape index images are matched separately, and the matching number of key points is treated as matching score. The other matching score is the ratio of geodesic distance between corresponding pairs of matched points.

Given a set of key points $\{p_i|i = 1, \cdots, N\}$ on mesh X, there is a set of corresponding matching points $\{p'_i|i = 1, \cdots, N\}$ on mesh X', then the geodesic distance ratio $GDR(X, X')$ between X and X' is defined as below.

$$GDR(X, X') = \frac{1}{N(N-1)/2} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{\min(gd(p_i, p_j), gd(p'_i, p'_j))}{\max(gd(p_i, p_j), gd(p'_i, p'_j))} \quad (3)$$

where $gd(p_i, p_j)$ means the geodesic distance between points p_i and p_j. According to its definition, the geodesic distance ratio lies in the range of $[0, 1]$.

Therefore, after the above process, we get three kinds of matching scores, namely, the matching number of key points from texture images MNT and shape index images MNS, and the geodesic distance ratio GDR between models. The final matching score is obtained by fusion of these three kinds of matching scores.

Score level fusion can be divided into three categories [20]: transformation-based, classifier-based, and density-based fusion. A comparison of eight biometric fusion methods conducted by NIST [21] favors the density-based fusion, specifically the likelihood ratio based fusion. In order to reduce the computational complexity, Gaussian Mixture Model (GMM) is used in [20] instead of kernel density estimator (KDE) for density estimation during likelihood ratio based fusion. However, in our case, the matching scores MNT and MNS are both of discrete values, which are inconvenient to be modeled with GMM. For the models with no matching key points, the GDR is set to be zero, and as a result there are many zero valued GDR, mostly for those matching between different subjects' models, which is also hard to be handled by GMM. Considering these inconvenience, we choose KDE for density estimation instead of GMM.

The likelihood ratio based fusion is formulated as below. Given a vector of K matching scores $\mathbf{s} = [s_1, s_2, \cdots, s_K]$, and estimated genuine density $\widehat{f}_{gen}(\mathbf{s})$ and impostor density $\widehat{f}_{imp}(\mathbf{s})$, compute the likelihood ratio $LR(\mathbf{s}) = \widehat{f}_{gen}(\mathbf{s})/\widehat{f}_{imp}(\mathbf{s})$, assign $\mathbf{s}$ to the genuine class if $LR(\mathbf{s}) \geq \eta$. We assume the independence of the three matching scores, therefore the density function $\widehat{f} = \prod_{k=1}^{3} \widehat{f}_k(s_k)$.

4 Experiments and Results

4.1 Dataset

Experiments are conducted on FRGC v2.0 database [24], which contains 4007 3D models of 466 subjects, and each model contains a pair of registered texture image and range image. Among these subjects, 57% are male and 43% are female, with the age distribution: 65% 18-22 years old, 18% 23-27 and 17% 28 years or over. The database was collected during 2003-2004 academic year, thus contains time variations. It also contains neutral and non-neutral expression variations.

All the texture and range images are simply cropped using a bounding box determined by two eye-outer corners and chin point, so that faces are separated from the background. These simply cropped images have pose, illumination, resolution, and expression variations, and also contain occlusions and holes.

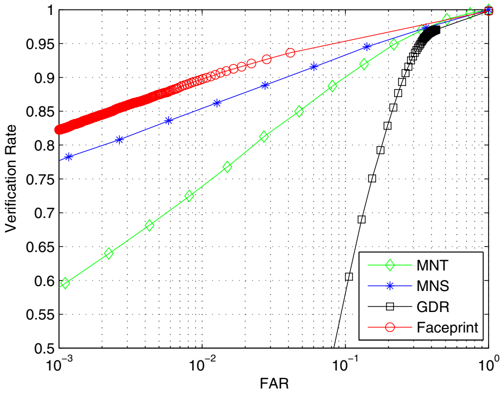

Fig. 3. ROC curves of All vs. All experiment

The database is divided into a training set and a testing set. The training set contains 977 models of randomly selected 100 subjects, and 3030 models of the rest 366 subjects form the testing set. The training set is used to estimate densities for likelihood ratio based fusion.

Four experiments are carried out, namely, All vs. All, and Neutral vs. Neutral, Neutral vs. Non-Neutral, and Neutral vs. All. All vs. All experiments are conducted to be consistent with FRGC experiment 3. All possible pairs of 3D models int the testing set are compared, so that a 3030*3030 similarity matrix is calculated. In Neutral vs. Neutral, Neutral vs. Non-Neutral, and Neutral vs. All experiments, 366 neutral expression faces are selected as the gallery set. The rest 2664 images are treated as probe set in Neutral vs. All experiment, while only neutral and non-neutral face models are used in Neutral vs. Neutral and Neutral vs. Non-Neutral experiments respectively. The reason to do Neutral vs. Neutral, Neutral vs. Non-Neutral, and Neutral vs. All experiments is to analyze robustness against expression variations.

4.2 All vs. All

As described above, in the All vs. All experiment, Faceprint method is used to match all possible pairs of 3030 3D models, and a 3030*3030 similarity matrix is obtained. From the similarity matrix ROC curves can be derived. We compare the ROC curves of the proposed Faceprint (fusion) method with those of MNT, MNS, and GDR in Fig. 3.

From Fig. 3, we can see that the proposed Faceprint method performs the best, verification rate of 82.3% is achieved at false accept rate of 0.1%. MNS performs much better than MNT which demonstrates the robustness of Shape index images is better than that of texture images. GDR performs not so well, but still shows some distinguishability.

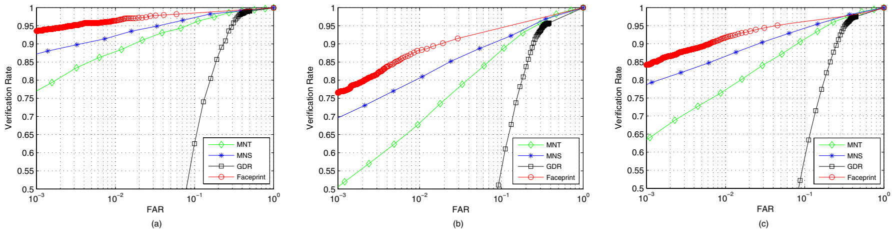

Fig. 4. ROC curves of experiments (a) Neutral vs. Neutral, (b) Neutral vs. Non-Neutral, and (c) Neutral vs. All

4.3 Neutral vs. Neutral, Non-Neutral, and All

In order to evaluate the robustness against expression variations, three experiments are conducted, namely, Neutral vs. Neutral, Neutral vs. Non-Neutral, and Neutral vs. All. ROC curves of these three experiments are shown in Fig. 4.

As can be seen from Fig. 4, no matter with or without expression variations, the order of the performance of Faceprint, MNT, MNS, and GDR remains the same. The performance degradation when non-neutral expressions are included in the experiment is calculated, and shown in Table 1. From Table 1, we can see that performance of MNT degrades most seriously (34.3%), followed by MNS (20.2%), and Faceprint (18.3%), while performance of GDR degrades by only 11.5%. To be noted, as it's hard and unreasonable to get the verification rate at FAR of 0.1%, the performance degradation of GDR is calculated at FAR of 1%, which is different from the other three matching scores. The results show that SIFT features extracted from shape index images are more robust than those from texture images, and geodesic distance ratio is most insensitive to expression variations. Fusion of these three kinds of matching scores increases the performance while retaining most of the robustness against expression variations.

Table 1. Performance degradation of Faceprint, MNT, MNS, and GDR when expression variations are included in the experiments

	Faceprint FAR=0.1%	MNT FAR=0.1%	MNS FAR=0.1%	GDR FAR=1%
Neutral vs. Neutral	93.6%	77.0%	87.2%	62.6%
Neutral vs. Non-Neutral	76.5%	50.6%	69.6%	55.4%
Performance degradation	18.3%	34.3%	20.2%	11.5%

From all the above four experiments, we can see that SIFT features extracted from shape index images (MNS) perform much better than the commonly used texture SIFT features (MNT), and the proposed Faceprint method performs the best by fusion the three kinds of matching scores.

5 Conclusions

In this paper, we pay special attention to local feature based 3D face recognition, and propose a novel Faceprint method. SIFT features are extracted from texture and range images and matched, the matching number of key points together with geodesic distance ratios between models are used as three kinds of matching scores, likelihood ratio based score level fusion is conducted to calculate the final matching score. Because of the invariance of SIFT to image scaling, translation and rotation, the robustness of shape index against illumination and pose changes, and the insensitivity of geodesic distance to expression variations, the Faceprint method is inherently robust against all these variations. Systematical experiments are carried out on FRGC v2.0 database, and the experimental results indicate that Faceprint method achieves consistently high performance comparing with commonly used SIFT on texture images.

Acknowledgements

This work was supported by Program of New Century Excellent Talents in University, National Natural Science Foundation of China (No. 60575003, 60332010, 60873158), Joint Project supported by National Science Foundation of China and Royal Society of UK (60710059), and Hi-Tech Research and Development Program of China (2006AA01Z133).

References

1. Zhao, W., Chellappa, R., Rosenfeld, A., Phillips, P.J.: Face Recognition: A Literature Survey. ACM Computing Surveys, 399–458 (2003)
2. 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), 1–15 (2006)
3. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscience 3(1), 71–86 (1991)
4. Belhumeur, P., Hespanha, J., Kriegman, D.: Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. IEEE Transactions on pattern analysis and machine intelligence 19(7), 711 (1997)
5. Kanade, T.: Computer Recognition of Human Faces. Interdisciplinary Systems Research 47 (1977)
6. Wiskott, L., Fellous, J., Kruger, N., von der Malsburg, C.: Face Recognition by Elastic Bunch Graph Matching. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7), 775–779 (1997)
7. Husken, M., Brauckmann, M., Gehlen, S., von der Malsburg, C.: Strategies and benefits of fusion of 2D and 3D face recognition. In: IEEE Workshop on Face Recognition Grand Challenge Experiments (2005)
8. Chua, C., Han, F., Ho, Y.K.: 3D human face recognition using point signature. In: Proc. IEEE International Conference on Automatic Face and Gesture Recognition, pp. 233–238 (2000)

9. Bicego, M., Lagorio, A., Grosso, E., Tistarelli, M.: On the use of SIFT features for face authentication. In: Proc. IEEE International Conference on Computer Vision and Pattern Recognition Workshop, pp. 35–41 (2006)
10. Pentland, A., Moghaddam, B., Starner, T.: View-Based and modular eigenspaces for face recognition. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, pp. 84–91 (1994)
11. Huang, J., Heisele, B.: Blanz. V.: Component-based Face Recognition with 3D Morphable Models. In: Proc. of the 4th International Conference on Audio- and Video-Based Biometric Person Authentication, pp. 27–34 (2003)
12. Lowe, D.G.: Object recognition from local scale-invariant features. In: Proc. of the International Conference on Computer Vision 1999, pp. 1150–1157 (1999)
13. Mian, A.S., Bennamoun, M., Owens, R.A.: An Efficient Multimodal 2D-3D Hybrid Approach to Automatic Face Recognition. IEEE Transactions on pattern analysis and machine intelligence 29(11), 1927–1943 (2007)
14. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(4), 91–110 (2004)
15. Lindeberg, T.: Scale-space theory: A basic tool for analysing structures at different scales. Journal of Applied Statistics 21(2), 224–270 (1994)
16. Kisku, D.R., Rattani, A., Grosso, E., Tistarelli, M.: Face Identification by SIFT-based Complete Graph Topology. In: 5th IEEE Workshop on Automatic Identification Advanced Technologies, Alghero, Italy (2007)
17. Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Three-dimensional face recognition. International Journal of Computer Vision 64(1), 5–30 (2005)
18. Kimmel, R., Sethian, J.A.: Computing geodesic on manifolds. Proc. US National Academy of Science 95, 8431–8435 (1998)
19. Sethian, J.A.: A review of the theory, algorithms, and applications of level set method for propagating surfaces. Acta numerica (1996)
20. Nandakumar, K., Chen, Y., Dass, S.C., Jain, A.K.: Likelihood Ratio Based Biometric Score Fusion. IEEE Transactions on Pattern Analysis and Machine Intelligence 30(2), 342–347 (2008)
21. Ulery, B., Hicklin, A.R., Watson, C., Fellner, W., Hallinan, P.: Studies of Biometric Fusion. NIST, Tech. Rep. IR 7346. (2006)
22. Lu, X., Jain, A.K., Colbry, D.: Matching 2.5D Face Scans to 3D Models. IEEE Transactions on pattern analysis and machine intelligence 28(1), 31–43 (2006)
23. Kakadiaris, I.A., Passalis, G., Toderici, G., Murtuza, N., Lu, Y., Karampatziakis, N., Theoharis, T.: Three-Dimensional Face Recognition in the Presence of Facial Expressions: An Annotated Deformable Model Approach. IEEE Transactions on pattern analysis and machine intelligence 29(4), 640–649 (2007)
24. Phillips, P.J., et al.: Overview of the Face Recognition Grand Challenge. In: Proc. Of IEEE Conf. on Computer Vision and Pattern Recognition, pp. I:947–954 (2005)
25. Dorai, C., Jain, A.K.: COSMOS - A Representation Scheme for 3D Free- Form Objects. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(10), 1115–1130 (1997)