Keypoint Identification and Feature-Based 3D Face Recognition

Ajmal Mian, Mohammed Bennamoun, and Robyn Owens

School of Computer Science and Software Engineering,
The University of Western Australia,
35 Stirling Highway, Crawley, WA 6009, Australia
{ajmal,bennamou}@csse.uwa.edu.au, robyn.owens@uwa.edu.au

Abstract. We present a feature-based 3D face recognition algorithm and propose a keypoint identification technique which is repeatable and identifies keypoints where shape variation is high in 3D faces. Moreover, a unique 3D coordinate basis can be defined locally at each keypoint facilitating the extraction of highly descriptive pose invariant features. A feature is extracted by fitting a surface to the neighbourhood of a keypoint and sampling it on a uniform grid. Features from a probe and gallery face are projected to the PCA subspace and matched. Two graphs are constructed from the set of matching features of the probe and gallery face. The similarity between these graphs is used to determine the identity of the probe. The proposed algorithm was tested on the FRGC v2 data and achieved 93.5% identification and 97.4% verifiction rates.

1 Introduction

The human face is a socially acceptable and non-intrusive biometric. It requires minimal or no cooperation from the subject making it ideal for surveillance. However, machine recognition of faces is very challenging because the distinctiveness of facial biometrics is quite low compared to other biometrics [6]. Moreover, changes caused by expressions, illumination, pose, occlusions and facial makeup (e.g. beard) impose further challenges on accurate face recognition.

Zhao et al. [18] categorize face recognition algorithms as holistic, feature-based and hybrid. Holistic algorithms use global features (e.g. Eigenfaces [17]) of the complete face whereas feature-based algorithms use local features [7] or regions [11] of the face for recognition. Hybrid matching methods use a combination of global and local-features for recognition e.g. [4]. A limitation of holistic matching is that it requires accurate normalization of the faces according to pose, illumination and scale. Variations in these factors can affect the global features of the face leading to inaccuracies in the final recognition. Moreover, global features are also sensitive to facial expressions and occlusions. Feature-based matching algorithms have the advantage that they are robust to variations in pose, illumination, scale, expressions and occlusions.

Bowyer et al. [2] give a survey of 3D face recognition algorithms and argue that 3D face recognition has the potential to overcome the limitations of its

S.-W. Lee and S.Z. Li (Eds.): ICB 2007, LNCS 4642, pp. 163–171, 2007.

2D counterpart. However, there is a need for better algorithms which are more tolerant to the above mentioned variations. Many 3D face recognition approaches are based on the ICP algorithm [1] or its modifications because of two major advantages. Firstly, perfect normalization of the faces is not required as ICP iteratively corrects registration errors while matching. Secondly, a partial region can be matched with a complete face. The latter has been exploited to avoid facial expressions [11][14] and to handle pose variations [8]. On the downside, ICP is a computationally expensive algorithm and does not extract any feature from the face. This rules out any possibility of indexing to speed up the matching process. Unless another algorithm and/or modality is used to perform indexing or prior rejection of unlikely faces [12], ICP based algorithms must perform a brute force matching thereby making the recognition time linear to the gallery size. Selecting expression insensitive regions of the face for matching is a potentially useful approach to overcome the sensitivity of ICP to expressions. However, deciding upon such regions is a problem worth exploring as such regions may not only vary between different persons but between different expressions as well.

We propose a novel keypoint identification and feature extraction algorithm for 3D face recognition. The identification of keypoints is repeatable and allows for the extraction of highly descriptive 3D features. Each feature is extracted by fitting a surface to the neighbourhood of a keypoint and sampling it on a uniform grid. Multiple features are extracted from each gallery face and projected to a PCA subspace. During recognition, features are extracted at keypoints on the probe and projected to a similar PCA subspace before matching. The set of matching features from a probe and gallery face are individually meshed to form two graphs. A spatial constraint is used to remove false matches (nodes) and the remaining graph is used to calculate the similarity measure between the faces.

Preliminary results of our algorithm have been published [13]. However, a number of extensions have been done since then including keypoint identification, local 3D coordinate derivation from single keypoints, projection of the features to a subspace, use of a more sophisticated graph matching approach and results from experiments on the complete FRGC v2 data.

2 Keypoint Identification

The input to our algorithm is a point cloud of a face $\mathbf{F} = [x_i \ y_i \ z_i]^T$ (where $i = 1 \ldots n$) which is sampled at uniform intervals. At each sample point p, a local region is cropped from the face using a sphere of radius r_1 centered at p. The value of r_1 is a trade off between descriptiveness of the feature and its sensitivity to variations. The smaller the value of r_1, the less will be the sensitivity to variations but this will also decrease the descriptiveness of the feature.

Let $\mathbf{L} = [x_j \ y_j \ z_j]^T$ (where $i = 1 \ldots n_l$) be the points in the region cropped by the sphere of radius r_1 centered at p. The mean vector $\mathbf{m}$ and the covariance matrix $\mathbf{C}$ of $\mathbf{L}$ are given by

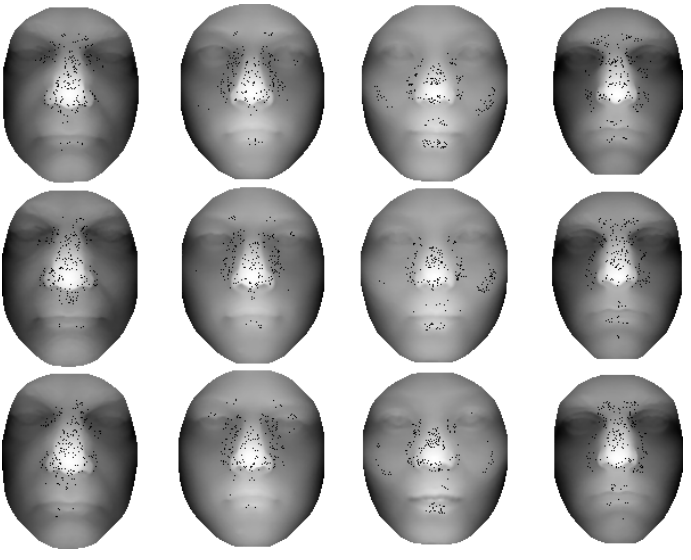


Fig. 1. Illustration of keypoint repeatability. Each column contains three range images of the same individual. Keypoints are repeatably identified for the same individual.

$$\mathbf{m} = \frac{1}{n_1}\sum_{j=1}^{n_1}\mathbf{L}_j \ , \quad \text{and} \tag{1}$$

$$\mathbf{C} = \frac{1}{n_1}\sum_{k=1}^{n_1}\mathbf{L}_j\mathbf{L}_j^T - \mathbf{m}\mathbf{m}^T \ , \tag{2}$$

where $\mathbf{L}_j$ is the jth column of $\mathbf{L}$. Performing PCA on the covariance matrix $\mathbf{C}$ gives the matrix $\mathbf{V}$ of eigenvectors such that Eqn. 3 is satisfied (where $\mathbf{D}$ is a diagonal matrix of the eigenvalues of $\mathbf{C}$). The matrix $\mathbf{L}$ can be aligned with its principal axes using Eqn. 4.

$$\mathbf{C}\mathbf{V} = \mathbf{D}\mathbf{V} \ , \tag{3}$$

$$\mathbf{L}' = \mathbf{V}(\mathbf{L} - \mathbf{m}). \tag{4}$$

$$\delta = \max(\mathbf{L}'_x) - \min(\mathbf{L}'_x) - (\max(\mathbf{L}'_y) - \min(\mathbf{L}'_y)) \tag{5}$$

In Eqn. 5, δ is the difference between the first two principal axes of the local region $\mathbf{L}'$ and $\mathbf{L}'_x$ is the vector of the x coordinates of $\mathbf{L}'$. If δ is greater than a threshold (i.e. $\delta \geq t_1$) p is selected as a keypoint. The total number of keypoints is determined by t_1. As the value of t_1 increases the total number of keypoints will decrease. The values of r_1 and t_1 are empirically chosen as $r_1 = 20mm$ and $t_1 = 2mm$. However, our algorithm is not sensitive to these parameters. Fig. 1 shows keypoints identified on different range images of four individuals. The keypoints are repeatably identified for a given individual but vary between individuals

because they have different facial shapes. The latter also enhances recognition accuracy. In the first column, the keypoints cluster mostly on the nose. In column three, some keypoints are also detected on the cheek bones. Experiments showed that the keypoints have 86% and 75.6% repeatability within the sampling interval for faces with neutral and non-neutral expressions respectively.

3 3D Feature Extraction

The neighbourhood $\mathbf{L}'$ of each keypoint is used to extract a 3D feature which is an extension of [10] and [13]. The major difference is that in this paper, the local coordinate basis for extracting the feature is derived from the principal directions of the neighbourhood $\mathbf{L}'$ of a single keypoint. This avoids the C_2^n combinatorial problem [10] without the knowledge of the nose tip [13]. Since the keypoints are selected such that there is no ambiguity in the principal directions of the surface patch, the derived 3D coordinate bases are stable and so are the features.

A surface is fitted to the points in $\mathbf{L}'$ using approximation [3] as opposed to interpolation so that it is robust to noise and outliers. Each point in $\mathbf{L}'$ pulls the surface towards itself and a stiffness factor controls the flexibility of the surface. The surface is sampled on a uniform 20×20 lattice (see Fig. 2-a). In order to avoid the effects of boundaries, a larger region is cropped first using r_2 (where $r_2 > r_1$) and a surface is fitted to it. This surface is then sampled on a bigger lattice and only the central 20×20 samples covering the r_1 region are concatenated to form a vector of dimension 400.

A constant value of t_1 will result in different numbers of keypoints identified for each face and bias the recognition results in favor of faces with more features. Therefore, an upper limit of 200 is imposed on the total number of features per face. The feature vectors are projected to a subspace defined by the eigenvectors of their largest eigenvalues using PCA. Let $\digamma = [\mathbf{f}_1 \dots \mathbf{f}_{200N}]$ (where N is the gallery size) be the $400 \times 200N$ matrix of all the feature vectors in the gallery. Each column of $\digamma$ contains a feature vector of dimension 400. The mean feature vector (Eqn. 6) is subtracted from all the feature vectors using Eqn. 7.

$$\overline{\mathbf{f}} = \frac{1}{200N} \sum_{i}^{200N} \mathbf{f}_i \tag{6}$$

$$\mathbf{f}'_i = \mathbf{f}_i - \overline{\mathbf{f}} \tag{7}$$

The mean subtracted feature matrix is given by Eqn. 8 and its covariance matrix by Eqn. 9 (where $\mathbf{C}$ is a 400×400 matrix). The eigenvalues and eigenvectors of $\mathbf{C}$ are calculated using Singular Value Decomposition (SVD) in Eqn. 10.

$$\digamma' = [\mathbf{f}'_1 \dots \mathbf{f}'_{200N}] \tag{8}$$

$$\mathbf{C} = \digamma'(\digamma')^T \tag{9}$$

$$\mathbf{USV}^T = \mathbf{C}\ , \tag{10}$$

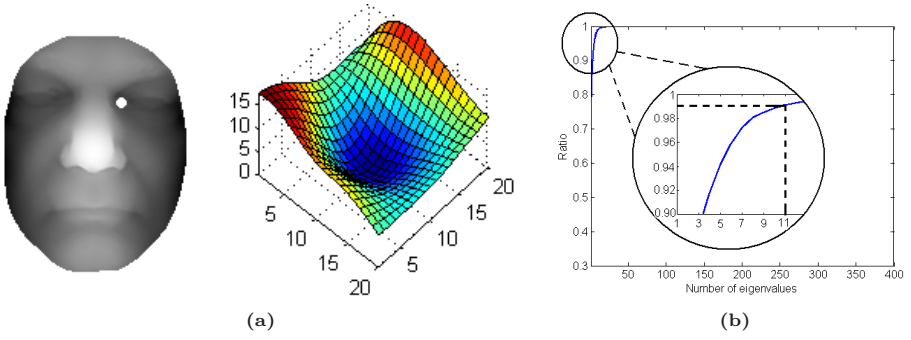


Fig. 2. (a) A keypoint displayed (in white colour) on a 3D face and a local surface fitted to the neighbourhood of the keypoint using a 20×20 lattice. (b) A plot of the ratio ψ as a function of the number of eigenvalues k.

where $\mathbf{U}$ is a 400×400 matrix of the eigenvectors and $\mathbf{S}$ is a diagonal matrix of the eigenvalues, both sorted in decreasing order. The dimension of the PCA subspace is governed by the amount of required accuracy (fidelity) in the projected space. Plotting the ratio of the first k eigenvalues to the total eigenvalues (i.e. $\psi = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{400} \lambda_i}$, where λ_i is the ith eigenvalue) as a function of the number of eigenvalues k (Fig. 2-b) shows that 99% accuracy is reached at only $k = 11$. This means a compression ratio of $\frac{(400-11)}{400} = 97.3\%$ which is not surprising given that all human faces have a similar topological structure and are roughly symmetric on either side of the nose. The first k eigenvectors are taken as $\mathbf{U}_k = \mathbf{U}_i$ (where $i = 1 \ldots k$ and $\mathbf{U}_k$ is a $400 \times k$ matrix of the first k eigenvectors). The mean subtracted feature matrix is projected to the eigenspace

$$F^{\lambda} = (\mathbf{U}_k)^T F' \quad , \tag{11}$$

where F^{λ} is a $k \times 200N$ matrix of the 3D feature vectors of the gallery faces. F^{λ} is normalized so that its variance along each of the k dimensions is equal

$$F^{\lambda}_{rc} = \frac{F^{\lambda}_{rc}}{\lambda_r} \quad \text{where} \quad r = 1 \ldots k \quad \text{and} \quad c = 1 \ldots 200N. \tag{12}$$

In Eqn. 12, r stands for the dimension or row number and c stands for the feature or column number. The feature vectors in F^{λ} (i.e. the columns) are normalized to unit magnitude and saved in a database along with $\overline{f}$ and $\mathbf{U}_k$ for online feature-based face recognition. The representation of gallery faces is quite compact as each face is represented by only 200 vectors of dimensionality 11.

4 Feature Matching

Indexing or hashing can speed up the matching process however, they are not the focus of this paper. Moreover, matching the probe with every gallery face results in many impostor scores useful for drawing the Receiver Operating Characteristic

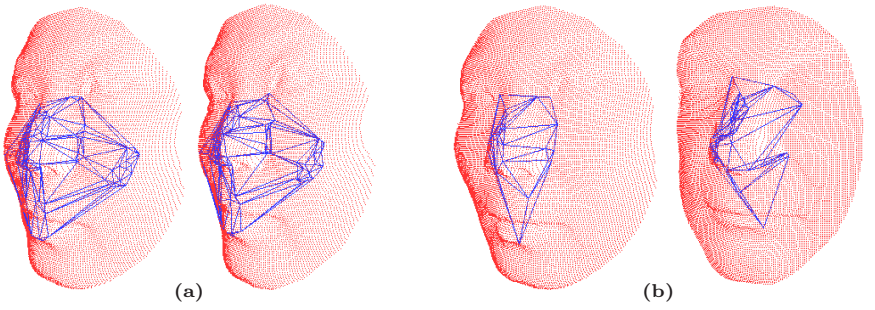


Fig. 3. (a) Correct match (b) Incorrect match

(ROC) curves. During recognition, features are extracted from the probe using the same parameters as the gallery. A probe feature vector $\mathbf{f}_p$ is first projected to the PCA subspace (Eqn. 13) and matched with a gallery feature (Eqn. 14).

$$\mathbf{f}_p^{\lambda} = (\mathbf{U}_k)^T(\mathbf{f}_p - \overline{\mathbf{f}}) \tag{13}$$

$$e = \cos^{-1}(\mathbf{f}_p^{\lambda}(\mathbf{f}_g^{\lambda})^T) \tag{14}$$

The value e measures the matching error between the probe and gallery features ($\mathbf{f}_p^{\lambda}$ and $\mathbf{f}_g^{\lambda}$) in the PCA subspace. For a given probe feature, the feature from the gallery face that has the minimum error with it is taken as its match. Once all the features are matched, the list of matches is sorted according to e. If a gallery feature matches more than one probe feature, only the one with the minimum value of e is considered. This allows for only one-to-one matches and the total number of matches m is different for every probe-gallery pair.

The keypoints of the matching features on the probe are projected on the xy-plane, meshed using Delaunay triangulation and projected back to the 3D space resulting in a 3D graph. The edges of this graph are used to construct a graph from the corresponding nodes (keypoints) of the gallery face using the list of matches. If the matches are correct i.e. the matching pairs of features correspond to the same location on the probe and gallery face, the two graphs will be similar (Fig. 3). The similarity measure between the graphs is given by

$$\gamma = \frac{1}{n_\varepsilon}\sum_{i}^{n_\varepsilon}(\varepsilon_{pi} - \varepsilon_{gi}) \ , \tag{15}$$

where ε_{pi} and ε_{gi} are the lengths of the corresponding edges of the probe and gallery graphs respectively and n_ε is the total number of edges. Eqn. 15 is an efficient way of measuring the spatial error between the matching pairs of features. The measure γ is pose invariant because the edge lengths of the graphs remain constant under pose variation. Another similarity measure is the mean Euclidean distance d between the nodes of the two graphs after least squared error minimization.

The matching algorithm results in four measures of similarity between the two faces i.e. $\overline{e}$, the total number of matches m, γ, and d. Excluding m, all similarity measures have a negative polarity (i.e. a smaller value means a better similarity). A probe is matched with every gallery face resulting in four vectors $\mathbf{s}_q$ of similarity measures (where q corresponds to a similarity measure). Each vector is normalized on the scale of 0 to 1 using

$$\mathbf{s}'_q = \frac{\mathbf{s}_q - \min(\mathbf{s}_q)}{\max(\mathbf{s}_q - \min(\mathbf{s}_q)) - \min(\mathbf{s}_q - \min(\mathbf{s}_q))} \ , \tag{16}$$

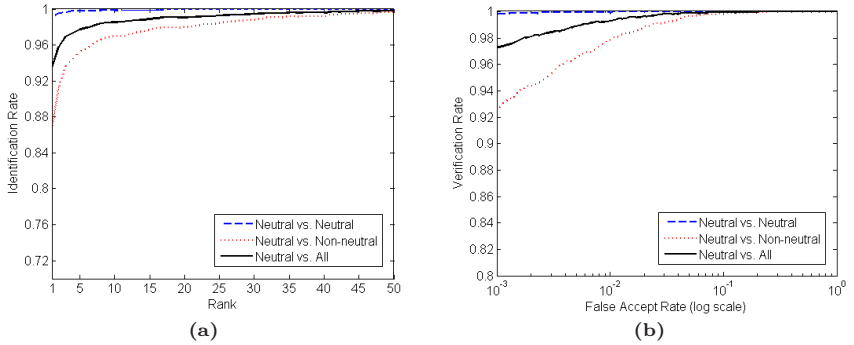
where $\mathbf{s}'_q$ contains the normalized similarity measures. The elements of $\mathbf{s}'_m$ are subtracted from 1 in order to reverse their polarity. The overall similarity is calculated using a confidence weighted sum rule

$$\mathbf{s} = \kappa_e \mathbf{s}'_e + \kappa_m (1 - \mathbf{s}'_m) + \kappa_\gamma \mathbf{s}'_\gamma + \kappa_d \mathbf{s}'_d \ , \tag{17}$$

where κ_q is the confidence in a similarity measure which can be calculated offline from training data or dynamically during online recognition as $\kappa_q = \frac{\overline{s}_q - \min(\mathbf{s}_q)}{\overline{s}_q - \min_2(\mathbf{s}_q)}$ (where $\overline{s}_q$ is the mean value of $\mathbf{s}_q$ and the operator $\min_2(\mathbf{s}_q)$ produces the second minimum value of the vector $\mathbf{s}_q$). Note that κ_m is calculated from $1 - \mathbf{s}'_m$. The gallery face which has the minimum value in $\mathbf{s}$ is declared as the probe's identity.

5 Results and Analysis

The FRGC v2 [16] validation set comprises 4007 3D scans of 466 subjects along with their texture maps. We only used the 3D shape of the faces and selected one face per individual under neutral expression to make a gallery of 466. The remaining faces (4007 − 466) are treated as probes and divided into neutral expressions (1944) and non-neutral expressions (1597).



(a) (b)

Fig. 4. (a) Identification results. The rank one identification rate for neutral vs. all is 93.5%. (b) ROC curves. The verification rate at 0.001 FAR for neutral vs. all is 97.4%.

Fig. 4-a shows our identification results. Our algorithm achieved rank one identification rates of 99.0% and 86.7% for probes with neutral and non-neutral expressions respectively. Under neutral expressions, only one probe is above rank 17 (100% recognition rate at rank 17). The identification rate drops under non-neutral expressions. However, it should be kept in mind that 3D face recognition is generally more sensitive to expressions. For example, the 3D face recognition rate of Lu et al. [8] dropped by 30%. In our case, the recognition rate drops by 12.3%. Moreover, the steep rise in the identification rate (i.e. 86.7% to 95%) from rank 1 to rank 5 indicates that the rank one identification rate will significantly improve by fusing other features e.g. global.

Fig. 4-b shows the ROC curves of our algorithm. At 0.001 FAR, we achieved verification rates of 99.9% and 92.7% respectively for probes with neutral and non-neutral expressions. In the neutral expressions case, a 100% verification rate is achieved at 0.01 FAR. It is not the aim of this paper to report the most accurate results on the FRGC v2 data and we believe that better results can be obtained by using a multi-algorithm approach. However, to give some idea of the performance of our algorithm, we compare our results to others. At 0.001 FAR we achieved a verification rate of 97.4% (neutral versus all) which can be compared to the results of Passalis et al. [15], Maurer et al. [9] and Huskën et al. [5] who achieved 85.1%, 86.5% and 89.5% verification rates respectively on the same dataset.

6 Conclusion

We presented a novel keypoint identification and feature extraction algorithm for 3D face recognition. These keypoints; (1) have 86% repeatability in the range images of the same individual, (2) vary between individuals, (3) are identified at locations where the shape variation is high, and (4) provide stable and repeatable local 3D coordinate frames for the computation of highly descriptive features. We also presented a graph based feature matching algorithm and reported experiments on the largest publicly available database of 3D faces. Our algorithm has an equal error rate (EER) of 0.75%(neutral vs. all) and has the potential for further improvement in a multi-algorithm setup.

Acknowledgment

We acknowledge the FRGC [16] organizers for providing data and J. D'Erico for the surface fitting code [3]. This research is sponsored by ARC Grant DP0664228.

References

1. Besl, P.J., McKay, N.D.: Reconstruction of Real-world Objects via Simultaneous Registration and Robust Combination of Multiple Range Images. IEEE TPAMI 14(2), 239–256 (1992)
2. Bowyer, K.W., Chang, K., Flynn, P.: A Survey Of Approaches and Challenges in 3D and Multi-modal 3D + 2D Face Recognition. CVIU 101, 1–15 (2006)

3. D'Erico, J.: Surface Fitting using Gridfit. MATLAB Central File Exchange (2006)
4. Huang, J., Heisele, B., Blanz, V.: Component-based Face Recognition with 3D Morphable Models. AVBPA (2003)
5. Huskën, M., Brauckmann, M., Gehlen, S., Malsburg, C.: Strategies and Benefits of Fusion of 2D and 3D Face Recognition. In: IEEE Workshop on FRGC Exp. (2005)
6. Jain, A.K., Ross, A., Prabhakar, S.: An Introduction to Biometric Recognition. IEEE TCSVT 14(1), 4–20 (2004)
7. Jones, M., Viola, P.: Face Recognition using Boosted Local Features. IEEE ICCV (2003)
8. Lu, X., Jain, A.K., Colbry, D.: Matching 2.5D Scans to 3D Models. IEEE TPAMI 28(1), 31–43 (2006)
9. Maurer, T., Guigonis, D., Maslov, I., Pesenti, B., Tsaregorodtsev, A., West, D., Medioni, G.: Performance of Geometrix ActiveID^{TM} 3D Face Recognition Engine on the FRGC Data. In: IEEE Workshop on FRGC Exp. (2005)
10. Mian, A.S., Bennamoun, M., Owens, R.A.: A Novel Representation and Feature Matching Algorithm for Automatic Pairwise Registration of Range Images. IJCV 66, 19–40 (2006)
11. Mian, A.S., Bennamoun, M., Owens, R.A.: 2D and 3D Multimodal Hybrid Face Recognition. ECCV 3, 344–355 (2006)
12. Mian, A.S., Bennamoun, M., Owens, R.A.: Automatic 3D Face Detection, Normalization and Recognition. 3DPVT (2006)
13. Mian, A.S., Bennamoun, M., Owens, R.A.: Face Recognition Using 2D and 3D Multimodal Local Features. ISVC, 860–870 (2006)
14. Mian, A.S., Bennamoun, M., Owens, R.A.: An Efficient Multimodal 2D-3D Hybrid Approach to Automatic Face Recognition. IEEE TPAMI 2007 (to appear)
15. Passalis, G., Kakadiaris, I., Theoharis, T., Tederici, G., Murtaza, N.: Evaluation of 3D Face Recognition in the Presence of Facial Expressions: An Annotated Deformable Model Approach. In: IEEE Workshop on FRGC Experiments. IEEE Computer Society Press, Los Alamitos (2005)
16. Phillips, P.J., Flynn, P.J., Scruggs, T., Bowyer, K., Chang, J., Hoffman, K., Marques, J., Min, J., Worek, W.: Overview of the Face Recognition Grand Challenge. IEEE CVPR (2005)
17. Turk, M., Pentland, A.: Eigenfaces for Recognition. JOCN 3 (1991)
18. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face Recognition: A Literature Survey. ACM Computing Survey, 399–458 (2003)