

Automatic Feature Extraction for Multiview 3D Face Recognition

Xiaoguang Lu and Anil K. Jain
Dept. of Computer Science & Engineering
Michigan State University
East Lansing, MI 48824
{Lvxiaogu, jain}@cse.msu.edu

Abstract

Current 2D face recognition systems encounter difficulties in recognizing faces with large pose variations. Utilizing the pose-invariant features of 3D face data has the potential to handle multiview face matching. A feature extractor based on the directional maximum is proposed to estimate the nose tip location and the pose angle simultaneously. A nose profile model represented by subspaces is used to select the best candidates for the nose tip. Assisted by a statistical feature location model, a multimodal scheme is presented to extract eye and mouth corners. Using the automatic feature extractor, a fully automatic 3D face recognition system is developed. The system is evaluated on two databases, the MSU database (300 multiview test scans from 100 subjects) and the UND database (953 near frontal scans from 277 subjects). The automatic system provides recognition accuracy that is comparable to the accuracy of a system with manually labeled feature points.

1. Introduction

Current 2D face recognition systems encounter difficulties in handling facial variations due to head poses and lighting condition [1], which introduce large amount of intra-class variations. Range image based 3D face recognition has been demonstrated to be effective in enhancing the face recognition accuracy [5, 17, 2]. Since each range image provides only a single view point of the face, instead of the full 3D view (see Fig. 1), we use 3D face models to recognize 2.5D face (range) images [12].

In both 2D and 3D face recognition systems, alignment (registration) between the query and the template is necessary [1]. Registration based on feature point correspondence is one of the most popular methods [10]. To make the face recognition system fully automatic, robust facial feature extraction is one of the crucial steps. Facial features can be of different types: region [15, 8], landmark [19, 16],

and contour [7, 20]. Generally, landmarks provide more accurate and consistent representation for alignment purposes than region-based features and have lower complexity and computational burden than contour feature extraction. We select a subset of the facial landmarks (or the fiducial points), as defined in anthropometry [9], including nose tip, inner eye corners, outside eye corners, and mouth corners. In the presence of large head pose variations, heuristics used for frontal scans may not hold, e.g., the nose tip is not the closest point to the sensor as in frontal scans. With the head pose unknown, the configuration models of the facial feature points, such as EGM [19] and AAM [7], are difficult to apply without a good initialization. Therefore, head pose is also considered as a feature to be extracted.

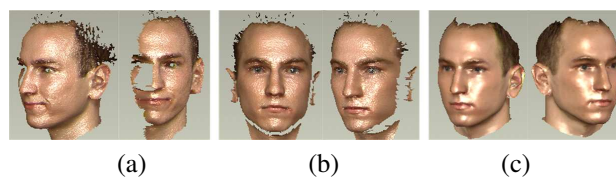


Figure 1. 2.5D scans and 3D model. (a) A 2.5D scan with large pose changes. (Notice the missing data due to pose change.) (b) A 2.5D frontal scan. (c) A 3D face model. Each pair is the same scan/model but displayed from different viewpoints.

A number of approaches have been proposed for feature extraction from (near) frontal facial scans [18, 4]. Wang et al. [18] used the point signature [6] and the stacked Gabor filter responses [19] to identify 3D and 2D features. Boehnen and Russ [4] explored 2D color information to extract skin tone regions and identify the eyes and the mouth. The 3D information contained in the range image was utilized to compute the geometry constraint. However, few of these studies address feature extraction in the presence of large pose changes.

We have focused on automatically extracting feature points and estimating the head pose in the presence of large pose variations. The extracted features are used for face alignment in three-dimensional space. Utilizing the automatic feature extraction module, a fully automatic 3D face recognition system is developed and evaluated.

2. Feature Extraction

The overall feature extraction process is shown in Fig. 2. Each 2.5D scan provides 4 matrices, $X(r, c)$, $Y(r, c)$, $Z(r, c)$, and $M(r, c)$ ¹, where X , Y , and Z are the spatial and depth coordinates in the units of millimeters and M is the mask, indicating which point is valid; $M(r, c)$ equals 1 if the point $p(r, c)$ is valid and 0 otherwise.

2.1. Face Segmentation

The first step in the face recognition system is to extract the facial area from the background. Given a facial scan, the invalid points in X , Y , and Z are filtered out by matrix M . The facial area is segmented by thresholding the horizontal and vertical integral projection curves of M . The face segmentation result of the facial scan in Fig. 2 is provided in Fig. 3.

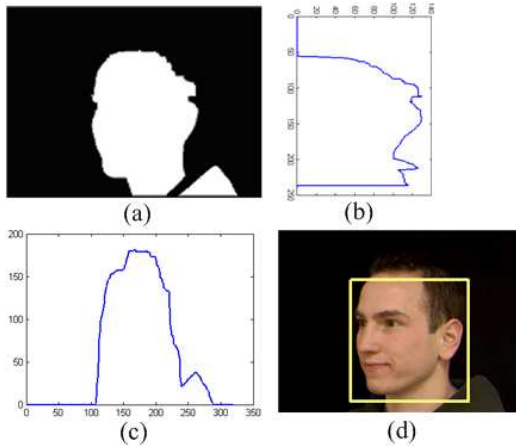


Figure 3. Segmentation of facial scan. (a) Mask image; (b) horizontal integral projection of M ; (c) vertical integral projection of M ; (d) face segmentation result.

2.2. Nose Tip and Pose Estimation

The nose tip is a distinctive point of the human face, especially in the range map. It is also insensitive to the facial

¹ r and c are the row and column indices, respectively.

expression changes. The pose of a face scan is represented by the angle of rotation w.r.t. the frontal pose (zero degree). For a frontal facial scan, the nose tip usually has the largest z value. But, in the presence of large pose changes, e.g., rotation along the yaw² direction, this heuristic does not hold. However, if the original coordinate system is rotated with the same pose change as the non-frontal scan, the nose tip will have the largest value along the rotated Z -axis. See Fig. 4. In other words, the nose tip still has the largest depth value if projected onto the corrected pose direction. We call it the *directional maximum*. Since the nose tip and the pose angle are coupled, we estimate them simultaneously.

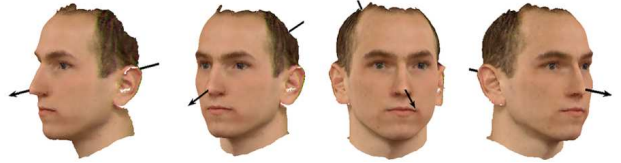


Figure 4. Directional maximum of the nose tip.

We illustrate the proposed algorithm based on an example with yaw angle changes. After the raw face scan is centered at its centroid, the nose tip extraction and pose estimation algorithm follows five steps: pose quantization, directional maximum, pose correction, nose profile extraction, and nose profile identification.

1. **Pose quantization.** The yaw angle change ranges from -90 degrees (full right profile) to 90 degrees (full left profile) in the X - Z plane. This 180 degree range (R_{pose}) is quantized into N_{pose} angles with equal angular interval ($\Delta\theta$). ($\Delta\theta$ and N_{pose} values are 2 degrees and 91 , respectively, in our experiments.) See Fig. 5.

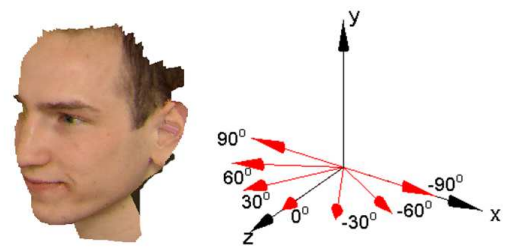


Figure 5. Pose angle quantization.

2. **Directional maximum.** At each pose angle θ_j ($j = 1, \dots, N_{pose}$), find the point with the maximum projection value along the corresponding pose direction as the nose tip candidate. The (x_i, y_i, z_i) coordinate of each face point p_i

²The rotation w.r.t. the Y -axis.

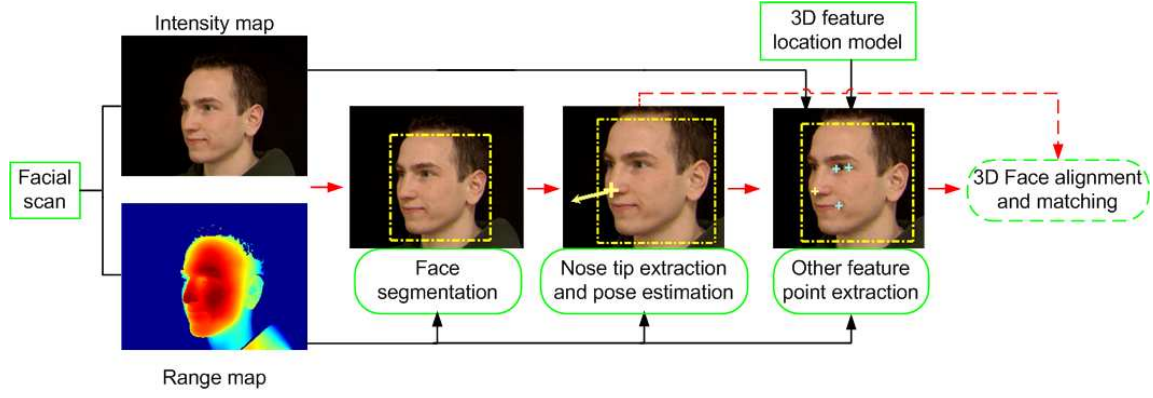


Figure 2. Automatic feature extraction for 3D face matching.

($i = 1, \dots, N$, where N is the total number of valid face points) is rotated to the new position $(x_i^{\theta_j}, y_i^{\theta_j}, z_i^{\theta_j})$ according to the pose angle θ_j , calculated as

$$\begin{pmatrix} x_i^{\theta_j} \\ y_i^{\theta_j} \\ z_i^{\theta_j} \end{pmatrix} = \begin{pmatrix} \cos \theta_j & 0 & \sin \theta_j \\ 0 & 1 & 0 \\ -\sin \theta_j & 0 & \cos \theta_j \end{pmatrix} \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix}. \quad (1)$$

The point p_k for which $z_k^{\theta_j} = \max(z_i^{\theta_j}, i = 1, \dots, N)$ is used as a nose tip candidate with the corresponding pose angle θ_j . By repeating this for every θ_j , M candidate pairs (nose tip candidate p and associated pose angle θ) are obtained (see Fig. 6). Since the directional maximum may happen with the same face point p at multiple θ_j s, $M \leq N_{pose}$. In such case, the angle with the largest projection value is selected as the pose angle to be associated with the point p . In the example of Fig. 6, M is 18. To determine the best candidate from M pairs, the nose profile will be utilized from the pose-corrected face scan.

3. **Pose correction.** For each candidate pair (p, θ) , the coordinates (x, y, z) of all the original face points are transformed to (x', y', z') so that point p is at the origin, and the face points are rotated according to the pose angle θ as follows:

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{pmatrix} \begin{pmatrix} x - p_x \\ y - p_y \\ z - p_z \end{pmatrix}. \quad (2)$$

The pose-corrected scans based on candidates 1 and 2 in Fig. 6 are shown in Figs. 7(a) and (b), respectively.

4. **Nose profile extraction.** From the pose-corrected scans based on each candidate (p, θ) , extract the nose profile at p (the origin of the coordinate system after pose correction), i.e., the intersection between the facial surface and the Y - Z plane. Let $X'(r, c)$, $Y'(r, c)$, and $Z'(r, c)$ denote the point coordinate matrices after pose correction. For each row r_i , find the point closest to the Y - Z plane, i.e.,

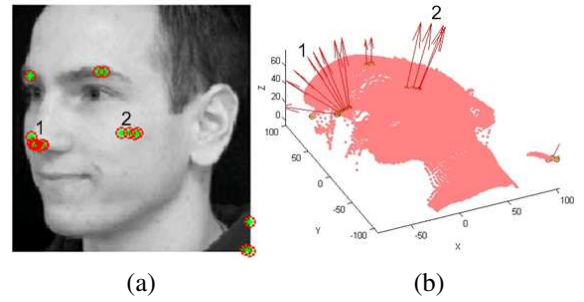


Figure 6. Example of directional maximum. The markers in (a) are the positions of the directional maximum with the associated pose direction plotted in (b). The pose angles of candidates 1 and 2 are 40 and -16 degrees, respectively.

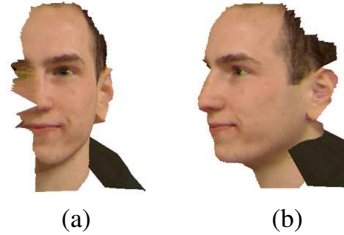


Figure 7. Pose corrected scans based on (a) candidate 1 and (b) candidate 2 in Fig. 6.

$(r_i, c^i) = \arg \min_c (|X'(r_i, c)|)$, resulting in a sequence of point pairs $(Y'(r_i, c^i), Z'(r_i, c^i))$. To make all the profiles comparable, each profile is normalized by centering it at the nose tip candidate and resampling it with equal interval along Y -axis, resulting in a nose profile vector. Linear

interpolation is applied for resampling.

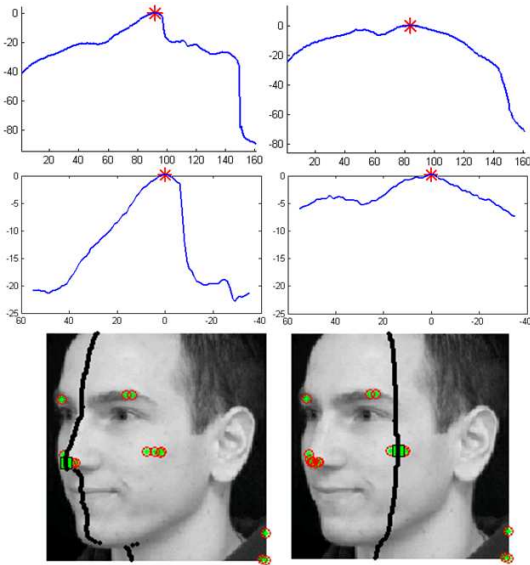


Figure 8. Top: extracted nose profiles; middle: normalized and resampled nose profile; bottom: extracted profiles overlaid on the original scan. The left (right) column is based on candidate 1 (2) in Fig. 6.

5. Nose profile identification. To identify the nose tip from the candidate pairs (p, θ) , we apply the subspace analysis on the nose profile vector space. A number of nose profiles from an independent group of subjects are extracted with manually labeled nose tip and pose, aligned at the nose tip, and resampled in the same way described in Step 4, resulting in a training set $\{V_i\}$. These (training) nose profiles are used to construct the nose profile subspace based on PCA. Given a test profile vector V , the distance-from-feature-space ($DFFS$) [13] is used as the distance metric. The nose tip candidates with the smallest $DFFS$ is identified as the nose tip and the associated pose angle is used as the pose estimation result. In the example of Fig. 6, candidate 1 has the smallest $DFFS$ among all the candidates.

2.3. Extracting Eye and Mouth Corners

A statistical model of the facial feature point locations is used as a prior constraint to reduce the search area for the feature points. The model contains constraints (in terms of interpoint distance and geometric relationship) between facial feature points. This not only enhances the accuracy of feature extraction, but also improves the computational efficiency. In our experiments, 145 frontal facial scans were used to compute the model.

Given the estimated nose tip and the pose angle, the feature point location model can be overlaid onto the given scan, and the search region for each feature point is constrained. The eye and mouth corners are then determined by using shape index from the range image and cornerness from the intensity image. The final score $F(p)$ for each feature point is a weighted sum of scores from range and intensity modalities. Figure 9 shows an example of the extracted corners by fusing range and intensity results. See [11] for details.



Figure 9. Feature extraction using fusion.

3. Automatic Face Recognition

A fully automatic 3D face recognition system has been developed, which matches stored 3D face models to 2.5D face scans in the presence of large pose changes. Using the nose tip position and the corresponding pose angle, the pose of the test face scan can be normalized up to a rigid transformation, i.e., translation and rotation. An iterative closest point (ICP) scheme [3], is applied to further improve the alignment results. The performance of ICP algorithms depends on the initial alignment. Since the nose profile subspace is built on a limited number of training samples, there is a possibility that the second best nose tip candidate may provide better matching results. Therefore, we keep the top- K nose tip candidates. The minimum distance among all the obtained K matching distances generated by ICP is chosen as the final matching distance.

To account for some errors in nose tip position and pose angle, we also used the eye and mouth corners to obtain a better initialization for the ICP algorithm.

4. Experiments and Discussion

Experiments are conducted on two databases. One was collected at our laboratory with large pose variations, called the MSU database. The other one is provided by the University of Notre Dame (UND) [5], which is also used in FRGC [14]. UND database contains (near) frontal facial scans. Both databases were collected using Minolta VIVID series scanners.

Experiment on the MSU database

There are 100 subjects in the MSU database with corresponding 100 3D face models stored in the gallery. For each subject, three 2.5D scans were captured with pose angles of less than -45 , 0 (frontal), and more than 45 degrees along the yaw direction, respectively. In total, the test database consists of 300 multiview scans of the same 100 subjects. Representative test scans are shown in Fig. 10.



Figure 10. Multiview 2.5D test scan examples.

Our feature extractor correctly located the nose tip region in 98% of the test scans. Using the manually labeled position as the ground truth, the localization displacement is computed as the Euclidean distance between the position of the automatically extracted feature point and the ground truth position, see Table 1. Figure 11 provides examples of the feature extraction results. The large displacement of

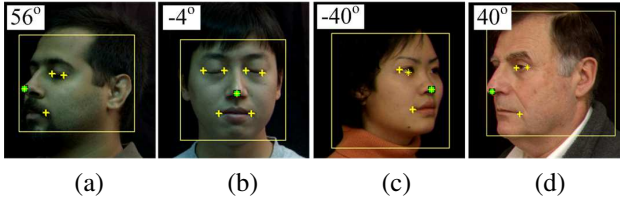


Figure 11. Feature extraction results which lead to correct 3D face matches on the MSU database. The number in the top-left corner is the estimated pose angle. The inner eye corner of (c) and the outside eye corner of (d) are not considered as valid feature points for matching due to low feature score F .

nose tip localization is often due to facial hair.

Fig. 12 shows the identification results for matching 300 multiview test scans to the 100 3D face models. The identification results using manually labeled feature points are also plotted for comparison. The fully automatic system provides an identification accuracy close to the system using manually labeled feature points by taking two (or more) feature candidate sets into consideration.

Experiment on the UND database

The UND database contains 953 facial scans from 277 subjects. Representative facial scans along with automatically extracted feature points are given in Fig.13. Table 1 provides the statistics of the localization displacement on

the UND database compared with the ground truth positions. If the head pose (near frontal) is provided, a more accurate algorithm can be designed [11]. For each subject, the scan with the earliest time stamp is used as the template and the remaining scans are used for queries. In total, there are 676 query scans and 277 templates. Unlike the MSU database, there is no 3D face model stored in the gallery for each subject. Both templates and queries are 2.5D scans. The identification results are provided in Fig. 14, compared with those obtained by using manually labeled feature points.

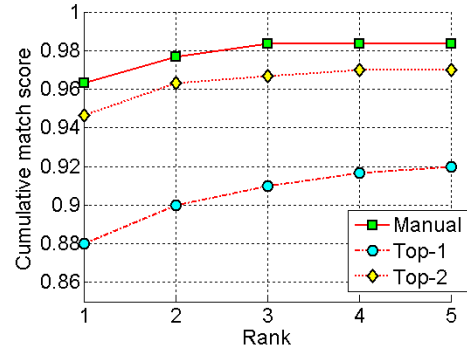


Figure 12. CMC curves of experiments on the MSU database. ‘Top- K ’ indicates the number of feature candidate sets used for matching.

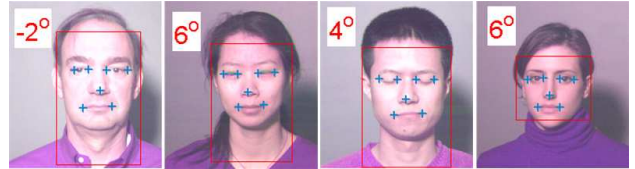


Figure 13. Examples of feature extraction results on the UND database.

5. Conclusions and Future Work

We have proposed an automatic feature extraction scheme to locate the nose tip and estimate the head pose, along with other facial feature points in multiview 2.5D facial scans. With the estimated pose, the system rejects the feature points that are not valid due to self-occlusion. The extracted features are used to align the multiview face scans with stored 3D face models to conduct surface matching. As a result, a fully automatic 3D face recognition system has been developed, which can recognize 2.5D facial scans in the presence of large pose changes. Our automatic face recognition system achieves an identification accuracy close to the system with manually labeled feature points.

Table 1. Statistics of the distance (in 3D) between the automatically extracted and manually labeled feature points. (distance between two pixels in x and y directions is $\sim 1mm$.) NT: nose tip; LE: inner left eye corner; RE: inner right eye corner; ORE: outside right eye corner; OLE: outside left eye corner; RM: right mouth corner; LM: left mouth corner.

Experiment	MSU database							UND database						
	NT	LE	RE	ORE	OLE	RM	LM	NT	LE	RE	ORE	OLE	RM	LM
Mean (mm)	6.4	7.1	9.0	13.6	13.3	6.7	5.2	8.3	8.2	8.3	9.5	10.3	6.0	6.2
Std (mm)	13.4	9.2	13.1	11.9	10.1	12.9	9.0	19.4	17.2	17.2	17.1	18.1	16.9	17.9
Median (mm)	4.3	5.3	6.0	12.7	11.7	3.8	3.2	5.3	5.8	5.4	5.5	7.4	2.9	3.3

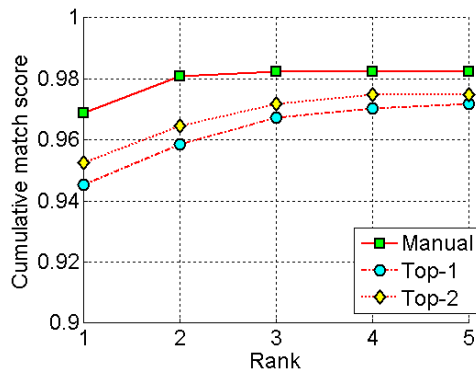


Figure 14. CMC curves of experiments on the UND database.

The proposed algorithm is designed to estimate the nose tip and head pose change by angle space quantization. The computational cost to handle the entire 3D space including three directions (i.e., yaw, pitch, and roll) would be expensive using brute force search. Therefore, a more efficient search scheme is being pursued. We are also exploring to utilize the matching score as a confidence measure to robustly select the most reliable points for registration or design a reject option to make the system generate fewer incorrect decisions.

References

- [1] *Face Recognition Vendor Test (FRVT)*. <<http://www.frvt.org/>>.
- [2] C. BenAbdelkader and P. Griffin. Comparing and combining depth and texture cues for face recognition. *Image and Vision Computing*, 23:339–352, 2005.
- [3] P. Besl and N. McKay. A method for registration of 3-D shapes. *IEEE Trans. PAMI*, 14(2):239–256, 1992.
- [4] C. Boehnen and T. Russ. A fast multi-modal approach to facial feature detection. In *Proc. 7th IEEE WACV*, pages 135–142, Breckenridge, CO, Jan. 2005.
- [5] K. I. Chang, K. W. Bowyer, and P. J. Flynn. Multi-modal 2D and 3D biometrics for face recognition. In *Proc. IEEE Workshop on Analysis and Modeling of Faces and Gestures*, pages 187–194, France, Oct. 2003.
- [6] C. Chua and R. Jarvis. Point signature: A new representation for 3D object recognition. *International Journal of Computer Vision*, 25(1):6385, 1997.
- [7] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. *IEEE Trans. PAMI*, 23(6):681–685, Jun. 2001.
- [8] D. Cristinacce and T. Cootes. Facial feature detection using adaboost with shape constraints. In *Proc. 14th BMVC*, pages 231–240, Norwich, UK, Sep. 2003.
- [9] L. Farkas. *Anthropometry of the Head and Face*. Raven Press, 2nd edition, 1994.
- [10] S. Li and A. Jain (Eds.). *Handbook of Face Recognition*. Springer, 2005.
- [11] X. Lu and A. K. Jain. Multimodal facial feature extraction for automatic 3D face recognition. Technical Report MSU-CSE-05-22, Department of Computer Science, Michigan State University, East Lansing, Michigan, August 2005.
- [12] X. Lu, A. K. Jain, and D. Colbry. Matching 2.5D face scans to 3D models. *IEEE Trans. PAMI*, 28(1):31–43, 2006.
- [13] B. Moghaddam and A. Pentland. Probabilistic visual learning for object representation. *IEEE Trans. PAMI*, 19(7):696–710, Jul. 1997.
- [14] P. Phillips. Face recognition grand challenge. *Biometric Consortium Conference*, 2004. <<http://www.frvt.org/FRGC/>>.
- [15] Y. Ryu and S. Oh. Automatic extraction of eye and mouth fields from a face image using eigenfeatures and multiplayer perceptrons. *Pattern Recognition*, 34(12):2459–2466, 2001.
- [16] K. Toyama, R. Feris, J. Gemmell, and V. Kruger. Hierarchical wavelet networks for facial feature localization. In *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, pages 118–123, Washington D.C., 2002.
- [17] F. Tsalakanidou, S. Malassiotis, and M. Strintzis. Face localization and authentication using color and depth images. *IEEE Transactions on Image Processing*, 14(2):152–168, 2005.
- [18] Y. Wang, C. Chua, and Y. Ho. Facial feature detection and face recognition from 2D and 3D images. *Pattern Recognition Letters*, 23:1191–1202, 2002.
- [19] L. Wiskott, J. Fellous, N. Kruger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Trans. PAMI*, 19(7):775–779, 1997.
- [20] J. Xiao, S. Baker, I. Matthews, and T. Kanade. Real-time combined 2D+3D active appearance models. In *Proc. IEEE CVPR*, pages 535–542, 2004.