

Matching Tensors for Pose Invariant Automatic 3D Face Recognition

A. S. Mian, M. Bennamoun and R. A. Owens
School of Computer Science and Software Engineering
The University of Western Australia, Crawley WA 6009
email: {ajmal, bennamou, robyn}@csse.uwa.edu.au

Abstract

The face is an easily collectible and non-intrusive biometric used for the authentication and identification of individuals. 2D face recognition techniques are sensitive to changes in illumination, makeup and pose. We present a fully automatic 3D face recognition algorithm that overcomes these limitations. During the enrollment, 3D faces in the gallery are represented by third order tensors which are indexed by a 4D hash table. During online recognition, tensors are computed for a probe and are used to cast votes to the tensors in the gallery using the hash table. Gallery faces are ranked according to their votes and a similarity measure based on a linear correlation coefficient and registration error is calculated only for the high ranked faces. The face with the highest similarity is declared as the recognized face. Experiments were performed on a database of 277 subjects and a rank one recognition rate of 86.4% was achieved. Our results also show that our algorithm's execution time is insensitive to the gallery size.

1. Introduction

The science of biometrics is defined as the automated use of physiological and behavioral characteristics to determine or verify an identity [2]. Verification (also known as authentication) is a one-to-one matching process whereby an individual's biometric signature is matched with the template of the identity claimed. The output is a binary decision (accept or reject). Identification on the other hand is a one-to-many matching process whereby an identity is associated with an individual by matching the biometric signature with the template of every identity in the database. The result in this case is a list of candidate identities sorted according to their degree of match. In this paper, our main focus will be on the more challenging problem of identification.

The face is an attractive biometric because of its universality, ease of collectability and the non-intrusive nature of measurement [10]. 2D face recognition has been around for a while in the research community however 3D face recognition is a relatively niche area and has gained popularity only recently. Many of the limitations of 2D face recogni-

tion, for example its sensitivity to illumination, makeup and pose, can be overcome using 3D face recognition. Moreover, 3D face recognition algorithms can be combined with 2D face recognition algorithms to achieve multimodal face recognition for higher accuracy (see Section 2 for a brief literature review).

In this paper, we propose a novel pose invariant 3D face recognition algorithm. Our algorithm is fully automatic and does not require any prior normalization or registration of the probe or gallery faces as opposed to our prior work [17]. As a matter of fact, registration of the probe comes as a by product of our recognition algorithm. Briefly our algorithm proceeds as follows. During enrollment, the gallery¹ is built by representing each face with multiple third order tensors [15] (approximately 400 tensors per face). These tensors are then indexed by a 4D hash table for quick referencing. During recognition, tensors are computed for the probe and are used to cast votes to the tensors in the gallery using the hash table. A similarity measure is then calculated between each probe tensor and with only those gallery tensors which receive votes above a certain threshold. The top 10 faces with the highest similarity are then registered to the probe face by aligning their matching pair of tensors. This registration is refined with the ICP algorithm [1] and the gallery faces are rearranged by fusing the registration error with the similarity measure. We performed experiments on the biometrics database of the University of Notre Dame [5][9][18] comprising 3D probe and gallery faces of 277 subjects and achieved a rank one recognition rate of 86.4%.

2. Previous Work

Related work in the area of 3D face recognition includes the following. Chua et al. [7] extracted point signatures [6] of the rigid parts of the face for expression invariant face recognition. However, they only performed experiments on a small database of 6 subjects and reported a 100% recognition rate. Blanz and Vetter [3] estimated the 3D shape of a face from its 2D images using morphable models. The

¹The database of enrolled faces is referred to as "gallery" and the face to be matched during online recognition is referred to as "probe".

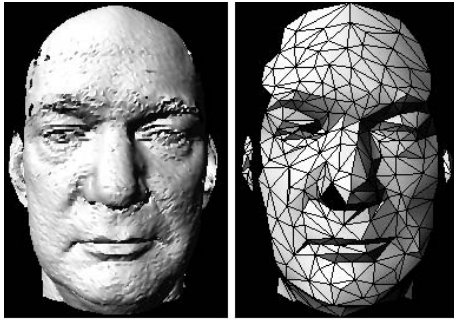


Figure 1: (a) A shaded view of a 3D face (24000 triangles). (b) After simplification (400 triangles).

3D faces were then used for pose and illumination invariant recognition and a recognition rate of 95% was reported. However their algorithm requires the manual identification of approximately seven feature points each on the front, side and profile views of a subject. Xu et al. [20] combined global geometric features with local shape variation to perform automatic 3D face recognition. They report 96.1% and 72.6% recognition rates when using a gallery of 30 and 120 subjects respectively². Lu et al. [13][14] used feature detection and registration with the ICP [1] algorithm for 3D face recognition. In [14], they reported a recognition rate of 96.5% but with a manual identification of control points and on a small gallery size of only 18 subjects. In [13], they reported a recognition rate of 90% with automatic feature point extraction but on a smaller gallery size of 10 subjects. Lao et al. [12] used stereo vision to acquire range images of faces and performed pose invariant 3D face recognition. They achieved a maximum of 96% recognition rate on a gallery size of 10 subjects.

In addition to the individual 2D and 3D face recognition approaches, some researchers have also investigated fusion of these two approaches. Such approaches generally perform matching on the basis of 2D images and 3D range images separately and the results are then fused. Some example techniques which use this hybrid approach include the following. Chang et al. [5] used a PCA-based approach for separate 2D and 3D face recognition and fused the results using a certainty-weighted sum-of-distance. They achieved a recognition rate of 93% using 3D facial data alone and a recognition rate of 99% for multimodal face recognition using 2D and 3D facial data. Their algorithm however requires the manual identification of control points for registering the faces. Wang et al. [19] used Gabor Wavelet filters in the 2D domain and point signatures [6] in the 3D domain for feature extraction. The results of the 2D and 3D feature matching were fused using a support vector machine (SVM). They report a recognition rate of above 90% on a gallery of 50 subjects and conclude that the integrated features (2D and 3D) perform better than the individual fea-

²Notice the significant effect of gallery size on the recognition rate.

tures alone. Bowyer et al. [4] give a detailed survey of 3D face recognition algorithms and conclude that there is still a need for the development of an improved face recognition algorithm.

In this paper, we focus on the *automatic* 3D face recognition problem. We perform face recognition using the 3D faces only and without using any texture information. Our algorithm however can be fused with any 2D recognition algorithm for improved multimodal recognition.

3. 3D Face Representation

In this section, we shall briefly explain our tensor representation. For a more elaborate explanation of the representation the reader is referred to [15]. The range image of a face (in the form of a point cloud) is first triangulated and then decimated using Garland's algorithm [8] to 400 triangles per face. Note that this is an extremely low resolution compared to the original resolution of the database of the University of Notre Dame [18]. Normals are then calculated for each vertex of the mesh. The vertices are then *paired* according to a distance and angle constraint. According to the distance constraint the distance between two vertices in a pair should be within 75mm and 110mm. According to the angle constraint the angle θ_d between the vertices in a pair should be within 5° and 60° . Since the number of possible vertex pairs is still likely to be large, only 400 valid vertex pairs are randomly selected.

Each valid vertex pair is then used to define a 3D basis with origin at the center of the two vertices. The average of the vertex normals defines the z -axis, their cross product defines the x -axis and the cross product of the z -axis with the x -axis defines the y -axis. This coordinate basis is then used to define a $15 \times 15 \times 15$ grid centered at the origin. Next, the surface area of the face crossing each bin of the grid is recorded in a third order tensor. Each element of the tensor is the face surface area that intersects the bin which corresponds to that tensor element. Since most of the bins of the grid are likely to be empty, the resultant tensor will have many zero elements. Therefore, the tensors are compressed into a sparse form to cut down on memory utilization.

The tensors of gallery faces are indexed by a 4D hash table. Three dimensions of the hash table correspond to the i, j, k indices of the tensor elements whereas the fourth dimension corresponds to θ_d . θ_d is quantized into bins of 5° . For each tensor of each gallery face, the tuple (face number, tensor number) entry is made in all the bins of the hash table corresponding to the i, j, k indices of the non-zero elements of the tensor and θ_d of the tensor.

4. 3D Face Recognition

During recognition, tensors are calculated for the probe face as described in Section 3. The i, j, k indices of the non-

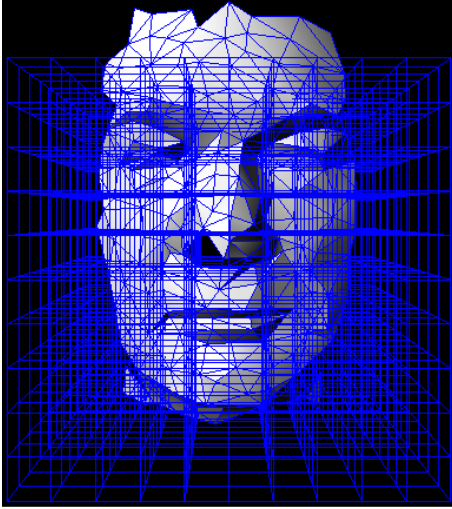


Figure 2: Illustration of the tensor. A $10 \times 10 \times 10$ grid defined over the face. The surface area of the face intersecting each bin of the grid is the value of the tensor element corresponding to the bin. (This figure is best viewed in colour.)

zero elements of each tensor and its θ_d are then used to cast votes to the tuples (face number, tensor number) which are present at the corresponding index positions (i, j, k, θ_d) in the hash table. The tensors (not the faces) that receive fewer than $0.7n_t$ votes are discarded (n_t is the number of non-zero elements of the probe tensor). Next, the similarity measure between the remaining pairs of tensors is calculated. To cater for holes (missing data as in the bottom right image of Fig. 4) in the probe and gallery, the two tensors are matched only in their overlapping regions (i.e. index positions where both tensors have nonzero elements). A vector is extracted from each tensor which comprises only the overlapping index elements of the respective tensor. We denote these vectors by \mathbf{p}_i for probe and \mathbf{g}_i for gallery (where $i = 1, \dots, n_p$ and n_p is the number of overlapping nonzero indices of the two tensors). The linear correlation coefficient C_c of \mathbf{p}_i and \mathbf{g}_i is then calculated using Eqn. 1. C_c is then weighted by the number of tensor elements n_p used to calculate it using Eqn. 2 to calculate the similarity between the probe and the gallery face.

$$C_c = \frac{n_p \sum p_i g_i - \sum p_i \sum g_i}{\sqrt{n_p \sum p_i^2 - (\sum p_i)^2} \sqrt{n_p \sum g_i^2 - (\sum g_i)^2}} \quad (1)$$

$$S = \frac{n_p}{n_t} C_c \quad (2)$$

In Eqn. 1, all summations are performed from $i = 1, \dots, n_p$. In Eqn. 2, S is the similarity measure between the two tensors. The top 10 gallery faces (not tensors) with the highest individual similarity measures and highest average similarity measures are then registered one by one to

the probe face. This is performed by transforming the coordinate basis of the gallery tensor to the coordinate basis of its matching probe tensor (see Eqn. 3 and 4).

$$\mathbf{R} = \mathbf{B}_g^\top \mathbf{B}_p \quad (3)$$

$$\mathbf{t} = \mathbf{O}_p - \mathbf{O}_g \mathbf{R} \quad (4)$$

In Eqn. 3, \mathbf{B}_g and \mathbf{B}_p are the 3×3 matrix of x, y, z coordinate vectors of the gallery and probe tensors respectively. In Eqn. 4, \mathbf{O}_g and \mathbf{O}_p are the 1×3 vectors of coordinates of the origins of the gallery and probe tensors respectively. \mathbf{R} and \mathbf{t} are the 3×3 rotation matrix and 1×3 translation vector that align the two coordinate bases and hence the gallery with the probe. This transformation provides a coarse registration of the probe and gallery face which is refined with the ICP algorithm [1]. The registration error between the probe and the gallery face is then calculated using Eqn. 5 and fused with the similarity measure S using Eqn. 6.

$$e = \frac{1}{n} \sum_{i=1}^n \|\mathbf{G}_i \mathbf{R} + \mathbf{t} - \mathbf{P}_i\| \quad (5)$$

$$S_r = \frac{S d_r}{e} \quad (6)$$

In Eqn. 5, \mathbf{G}_i and \mathbf{P}_i are the corresponding points of the gallery and probe respectively. \mathbf{R} and \mathbf{t} are the rotation matrix and translation vector (after refinement with the ICP algorithm) that align the gallery face to the probe. n is the total number of corresponding points between the two faces and e is the average error. In Eqn. 6, d_r is the resolution (mean edge length of the mesh) of the gallery face. The gallery faces are finally ranked according to the fused similarity measure S_r and the face with the highest value of S_r is declared as the recognized face.

5. Experiments

We performed our experiments on the 3D face database of the University of Notre Dame [5][18]. This database consists of multiple facial range images (and their corresponding 2D images) of 277 subjects acquired at different times. For details of the database the reader is referred to [5]. The resolution of this database is very high (480×640); therefore we down sampled it to 240×320 in order to gain memory efficiency. Although each subject in this database was asked to look directly into the camera and have a normal expression [5], the database still contains variations in pose and expression. Moreover, the database also has variations in the spatial resolution of the faces and illumination conditions. Fig. 3 shows an example of pose variation in the database. The pose variation is small, however it is still significant for 3D face recognition and would require a prior normalization of the pose unless a pose invariant 3D face



Figure 3: Example of pose variation in the database. The pose variation is small however it is still significant for 3D face recognition.

recognition algorithm is used. Fig. 4 shows an example of expression variation in the database (the left face has normal expression while the right one is smiling). Notice that there is a significant change in the 3D facial surface as a result of the expression change (Fig. 4 second row).

Fig. 5 shows an example of spatial resolution variation in the database. The left face in Fig. 5 has a point cloud of 9,760 points (resolution = 2.81mm) whereas the right one has a point cloud of 24,704 points (resolution = 1.46mm). The 3D faces corresponding to the 2D faces of Fig. 5 are not shown because it is not possible to visually differentiate between the two resolutions (as both of them are high). Fig. 6 shows an example of illumination variation in the database. Notice that despite the significant illumination variation in the 2D images (Fig. 6 first row), there is virtually no variation in their corresponding 3D images (Fig. 6 second row) except for a few holes in the right 3D image.

6. Results

We tested 242 probes in our experiments and did not perform any normalization of the probes or the gallery and nor did we register the faces by manually identifying feature points on the faces. Our gallery comprised a single 3D face for each of the 277 subjects. Recognition was also performed on the basis of a single probe per trial. Fig. 7 shows our recognition results. We achieved a rank one recognition of 86.4% which is less than the rank one recognition rate reported in [5]. However, it should be noted that our algorithm is pose invariant and does not require any prior regis-

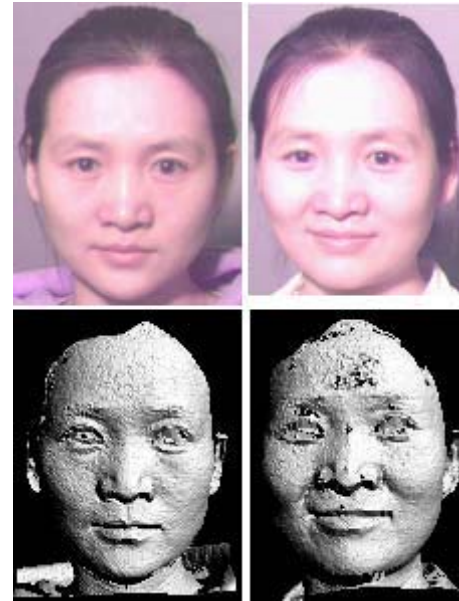


Figure 4: Example of expression variation in the database.

tration of the faces which was performed by manual selection of feature points in [5]. Moreover, we did not process the range images to remove spikes and interpolate holes as our algorithm is robust to such anomalies in the range data.

To evaluate the performance of our algorithm when used for authentication purposes, we calculated the probability distributions of the genuine and impostor probes (see Fig. 8). The genuine probes are the ones that are matched with their corresponding gallery faces. The impostor probes are the ones that are matched with a different gallery face that has the highest similarity measure. These two distributions can be used to select a threshold and then calculate the false acceptance and false rejection rates or vice versa [11]. Selecting a threshold of $S_r = 1.3$ (where the two curves cross), the false acceptance and false rejection rates each become approximately equal to 0.84%. It may be noted that the distributions of Fig. 8 have been calculated on the basis of the top 10 matches only.

Fig. 9 shows the performance of our recognition (identification) algorithm as a function of the gallery size. The gallery size was varied from 25 to 277 in steps of 25 and the matching time of our algorithm was recorded. The implementation of our algorithm was in Matlab 6.5 on a Pentium 4 PC with 1.2 GB RAM. The upper straight line in Fig. 9 shows the matching time as a linear function of the size of the gallery for comparison purposes. The lower curve in Fig. 9 shows the matching time of our algorithm. Note that this curve is almost flat. When the gallery size increases 11 times (from 25 to 277) the matching time increases only 2.7 times (30.0 to 82.4 seconds). This is mainly because our algorithm uses a simultaneous one-to-many matching approach (i.e. the hash table based voting scheme). On this



Figure 5: Example of spatial resolution variation in the database. The left face has a point cloud of 9,760 points whereas the right one has a point cloud of 24,704 points.

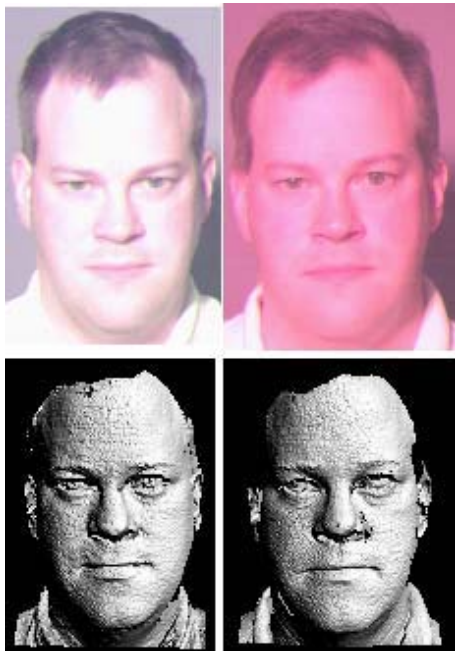


Figure 6: Example of illumination variation in the database.

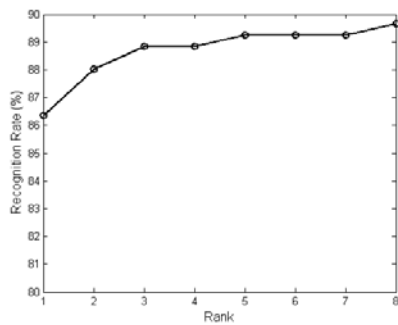


Figure 7: Recognition rate up to rank 10. Note that the recognition beyond rank 10 was not considered.

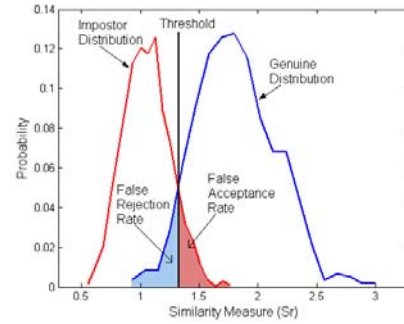


Figure 8: Probability distributions of the similarity measures of the genuine and impostor probes. The shaded areas show the false rejection and false acceptance rates.

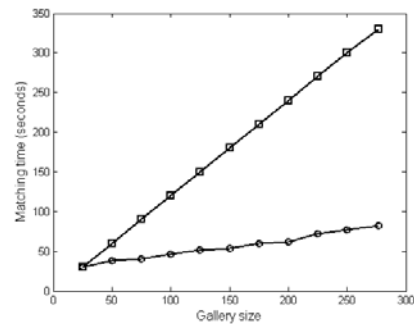


Figure 9: Matching time as a function of gallery size. The upper line shows the matching time as a linear function of the size of the gallery. The matching time (lower curve) of our algorithm is almost flat.

basis, we can conclude that our algorithm is not sensitive to the size of the gallery and can be scaled to even greater gallery sizes.

7. Discussion

Registration or pose normalization has seemingly been a prerequisite of most 3D face recognition algorithms. However, accurate registration of 3D faces is a challenging task mainly because the identity of the probe is unknown and there is no accurate matching template to which the face can be registered. This is in contrast to 3D modeling in which case one range image is registered to another overlapping range image of the same object. The pose invariant 3D face registration problem resembles the classic chicken and egg problem i.e. recognition requires registration and registration requires recognition. Chang et al. [5] performed registration of 3D faces by manually identifying feature points. However, the results of Lu et al. [13] show that even with manual identification of feature points, the 3D face recognition rate on a small gallery size of 10 subjects was 90%. Without manual identification of feature points, the 3D face recognition rate of [13] was 79%. This also indicates that

the recognition rates generally drop once the manual identification of feature points is replaced with an automatic feature identification algorithm. This is not surprising as no current feature identification algorithm has yet reached an accuracy comparable to the capability of humans.

Excluding [5] and [3], most of the work cited in Section 2 performed experiments on very small gallery sizes. It should be noted that recognition rates tend to drop when the gallery size is increased. The results of Xu et al. [20] support this fact. The face recognition algorithms cited in Section 2 also do not consider the effect of increasing gallery size on the performance of the algorithm. This is a very important criterion as in practical biometric applications, the gallery size is expected to be very large. From that perspective, our algorithm is unique in the sense that its performance is relatively insensitive to the gallery size. The performance of our algorithm can be further improved by reducing the number of tensors to be calculated (and hence matched) for each face. This can be done by using an automatic feature identification algorithm with our approach and calculating tensors for these identified feature points only. In the ideal case, if three feature points can be reliably identified on the faces, then our algorithm can represent each face with a single tensor and achieve a realtime performance independent of the gallery size. To validate our claim, we performed a limited number of experiments using a small gallery size. We used our tensor matching algorithm after manually registering the probes to their respective gallery faces and achieved 100% recognition rate. An important point to be noted is that our automatic 3D face recognition results reported in this paper are not as good as our automatic 3D object recognition results [16] because face is a non-rigid object.

8. Conclusion

We presented a 3D face recognition algorithm capable of recognizing faces from arbitrary pose. Our algorithm does not require the manual identification of feature points on the faces for their registration or any preprocessing of the 3D faces to remove spikes and interpolate between holes. The algorithm was tested on a large gallery consisting of 277 subjects and a recognition rate of 86.4% was achieved. In the future, we intend to reduce the number of tensors required per face and fuse 2D recognition with our algorithm for multimodal face recognition.

9. Acknowledgements

We would like to acknowledge Prof. Flynn, University of Notre Dame for providing the face data and Carnegie Mellon University for providing the mesh simplification software. This research is sponsored by ARC Grant DP0344338.

References

- [1] P. J. Besl and N. D. McKay, "Reconstruction of Real-world Objects via Simultaneous Registration and Robust Combination of Multiple Range Images," *IEEE TPAMI*, vol. 14(2), pp. 239–256, 1992.
- [2] Biometric Consortium, <http://www.biometrics.org>, 2004.
- [3] V. Blanz and T. Vetter, "Face Recognition Based on Fitting a 3D Morphable Model," *IEEE TPAMI*, vol. 25, pp. 1063–1074, 2003.
- [4] K. W. Bowyer, K. Chang and P. J. Flynn, "A Survey of Approaches to Three-Dimensional Face Recognition", *IEEE ICPR*, pp. 358–361, 2004.
- [5] K. I. Chang, K. W. Bowyer and P. J. Flynn, "Face Recognition Using 2D and 3D Facial Data," *MMUA*, pp. 25–32, 2003.
- [6] C. S. Chua and R. Jarvis, "Point Signatures: A New Representation for 3D Object Recognition," *IJCV*, vol. 25(1), pp. 63–85, 1997.
- [7] C. Chua, F. Han and Y. Ho, "3D Human Face Recognition Using Point Signatures," *IEEE AMFG*, pp. 233–238, 2000.
- [8] M. Garland and P. S. Heckbert, "Surface Simplification using Quadric Error Metrics", *SIGGRAPH*, pp. 209–216, 1997.
- [9] P. J. Flynn, K. W. Bowyer and P. J. Phillips, "Assessment of time dependency in face recognition: An initial study", *AVBPA*, pp. 44–51, 2003.
- [10] A. Jain, L. Hong and S. Pankanti, "Biometric Identification," *Communications of the ACM*, vol. 43(2), pp. 91–98, 2000.
- [11] A. Jain, A. Ross and S. Prabhakar, "An Introduction to Biometric Recognition", *IEEE TPAMI*, vol. 14(1), pp. 4–20, 2004.
- [12] S. Lao, Y. Sumi, M. Kawade and F. Tomita, "3D Template Matching for Pose Invariant Face Recognition Using 3D Facial Model Built with Isoluminance Line Based Stereo Vision," *IEEE ICPR*, vol. 2, pp. 911–916, 2000.
- [13] X. Lu, D. Colbry and A. K. Jain, "Matching 2.5D Scans for Face Recognition," *ICBA, LNCS 3072*, pp. 30–36, 2004.
- [14] X. Lu, D. Colbry and A. K. Jain, "Three-dimensional Model Based Face Recognition," *IEEE ICPR*, 2004.
- [15] A. S. Mian, M. Bennamoun and R. A. Owens, "Matching Tensors for Automatic Correspondence and Registration", *ECCV*, part 2, pp. 495–505, 2004.
- [16] A. S. Mian, M. Bennamoun and R. A. Owens, "A Novel Algorithm for Automatic 3D Model-based Free-form Object Recognition", *IEEE SMC*, pp. 6348–6353, 2004.
- [17] A. S. Mian, D. Matherr, M. Bennamoun, R. A. Owens and G. Hingston, "3D Face Recognition by Matching Shape Descriptors", *IVCNZ*, pp. 23–28, 2004.
- [18] University of Notre Dame Biometrics Database, available at <http://www.nd.edu/~cvrl/UNDBiometricsDatabase.html>, Comp ID 10, 2004.
- [19] Y. Wang, C. Chua and Y. Ho, "Face Recognition From 2D And 3D Images," *AVBPA*, pp. 26–31, 2001.
- [20] C. Xu, Y. Wang, T. Tan and L. Quan, "Automatic 3D Face Recognition Combining Global Geometric Features with Local Shape Variation Information," *IEEE ICPR*, pp. 308–313, 2004.