2D and 3D Multimodal Hybrid 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, robyn}@csse.uwa.edu.au
http://web.csse.uwa.edu.au

Abstract. We present a 2D and 3D multimodal hybrid face recognition algorithm and demonstrate its performance on the FRGC v1.0 data. We use hybrid (feature-based and holistic) matching for the 3D faces and a holistic matching approach on the 2D faces. Feature-based matching is performed by offline segmenting each 3D face in the gallery into three regions, namely the eyes-forehead, the nose and the cheeks. The cheeks are discarded to avoid facial expressions and hair. During recognition, each feature in the gallery is automatically matched, using a modified ICP algorithm, with a complete probe face. The holistic 3D and 2D face matching is performed using PCA. Individual matching scores are fused after normalization and the results are compared to the BEE baseline performances in order to provide some answers to the first three conjectures of the FRGC. Our multimodal hybrid algorithm substantially outperformed others by achieving 100% verification rate at 0.0006 FAR.

1 Introduction

Machine recognition of human faces has fascinated many researchers because of its potential applications in scenarios where fingerprinting or iris scanning are impractical (e.g. surveillance) or undesirable due to problems of social acceptance [7]. Considerable work has been done in this area for over three decades [14] which has resulted in a number of face recognition algorithms. These algorithms are categorized from two different perspectives, namely the type of data and the type of approach they use. From the first perspective, face recognition algorithms are divided into (1) 2D face recognition (which use 2D greyscale or colour images), (2) 3D face recognition (which use 3D range images or pointclouds of faces) and (3) multimodal face recognition algorithms (which use both 2D and 3D facial data) e.g [8]. Bowyer et al. [4] give a detailed survey of 3D and multimodal face recognition algorithms and state that multimodal face recognition outperforms both 2D and 3D face recognition alone. A comprehensive survey of 2D face recognition algorithms is given by Zhao et al. [14]. They also categorize face recognition into (1) holistic, (2) feature-based (referred to as region-based in this paper) and (3) hybrid matching face recognition algorithms. Holistic algorithms match the faces as a whole whereas region-based algorithms match local regions of the faces e.g. eyes and nose. Hybrid algorithms perform recognition on the

A. Leonardis, H. Bischof, and A. Pinz (Eds.): ECCV 2006, Part III, LNCS 3953, pp. 344–355, 2006.

Table 1. First three experiments of FRGC. Gallery and probe correspond to the database face and the face to be tested respectively. "Controlled" means controlled illumination and normal expression.

	Gallery	Probe
Experiment 1	Single controlled 2D image	Single controlled 2D image
Experiment 2	Four controlled 2D images	Four controlled 2D images
Experiment 3	Single 3D image (shape & texture)	Single 3D image (shape & texture)

basis of both holistic and region-based matching. It is argued that the hybrid methods "could potentially offer the best of the two types of methods" [14].

One of the major limitations in comparing different face recognition algorithms is that most researchers perform their experiments on different datasets. In most cases these datasets are very sparse and insufficient to provide statistically significant inference. To overcome this problem the Face Recognition Grand Challenge (FRGC) [10] was designed, with an objective to pursue the development of face recognition algorithms by providing sufficient datasets, challenge problems and standard benchmarks so that the performance of different algorithms can be compared on similar benchmarks. The first three challenge problems (or experiments) of FRGC which are related to this paper are listed in Table 1. FRGC also states five conjectures [10], the first three of which regard 2D versus 3D face recognition. For completeness, the relevant conjectures are summarized below (> stands for "performance will be better than" and "texture" means the 2D luminance image acquired by a 3D sensor).

Conjecture I-A: Exp3 (shape only) > Exp3 (texture only). 3D face recognition will perform better than 2D face recognition at equal resolution.

Conjecture I-D: Exp3 (shape only) > Exp1. 3D face recognition will perform better than higher resolution 2D face recognition.

Conjecture I-E: Exp3 (shape + texture) > Exp1. Multimodal (2D and 3D) face recognition will perform better than higher resolution 2D face recognition.

Conjecture II: The opposite of I-D and I-E.

Conjecture III-A: Exp2 > Exp3 (shape + texture). 2D face recognition using four high resolution images will perform better than multimodal (2D and 3D) face recognition at lower resolution.

Conjecture III-B: Exp2 > Exp3 (shape only). 2D face recognition using four high resolution images will perform better than 3D face recognition at lower resolution.

In this paper, we present a multimodal hybrid face recognition approach and perform Experiment 3 (see Table 1) on the FRGC v1.0 dataset (frontal views of faces). Our algorithm is multimodal as it utilizes both the shape and texture data of a face. At the same time it is hybrid as it performs recognition on the basis of region-based and holistic matching. In order to provide some answers

to the above listed conjectures, we compare our results to those of Experiment 1, 2 and 3 (see Table 1) when using the BEE (Biometric Experimentation Environment) baseline algorithms (PCA-based face recognition [12]). Comparison is performed using the FRGC criterion i.e. verification rate at 0.001 FAR (False Acceptance Rate). Our multimodal hybrid algorithm significantly outperforms the BEE baseline performance by achieving 100% verification rate at 0.0006 FAR which is well below the FRGC benchmark of 0.001 FAR. Our results clearly support Conjecture I-A, I-D and I-E and go against Conjectures II and III. In other words, 3D face recognition using our region-based matching algorithm performs better than PCA based 2D face recognition.

1.1 Overview of Multimodal Hybrid Face Recognition

Most sensors, including the Minolta scanner which was used to acquire the FRGC data, give a 3D pointcloud of the face along with its registered coloured texture map. We use the texture map for 2D holistic face recognition using the BEE baseline PCA algorithm. The 3D facial data consist of the x, y and z components of the pointcloud of the face. Taking the z component alone results in the range image of the face where each pixel value represents the depth of the corresponding facial point. The range image of the face is used separately for holistic 3D face recognition using the BEE baseline PCA algorithm. Additionally, the 3D pointcloud of the face is segmented into three disjoint regions, namely the eyes-forehead, the nose, and the cheeks, in order to perform a region-based matching [9]. For region-based matching, a modified version of the ICP algorithm [1] is used. An advantage of using this algorithm is that a partial region (e.g. nose) from the gallery can be matched directly with a complete probe face without segmenting the probe. Only the eyes-forehead (referred to as "forehead" hereafter) and the nose are used for region-based matching in order to avoid facial expressions and artifacts resulting from facial hair. These matching processes result in four similarity matrices which are normalized and subsequently fused. A min-max rule is used for normalization and a multiplication rule is used for fusion. The resulting similarity matrix is normalized once again and used to calculate the verification and identification rates of our algorithm. Fig. 1 shows the block diagram of our multimodal hybrid face recognition algorithm. The region-based 3D matching (left blocks of Fig. 1) algorithm was initially proposed in [9] however it is explained in Section 2 for completeness.

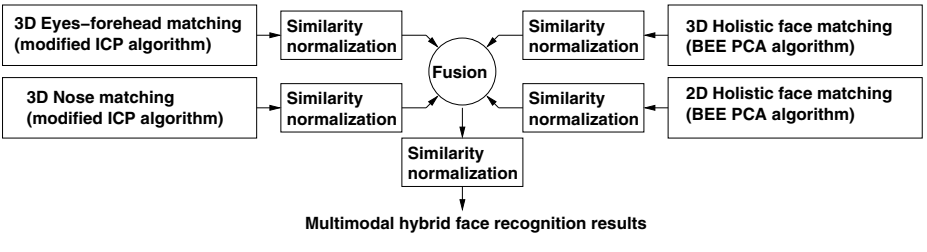

Fig. 1. Illustration of our multimodal hybrid face recognition algorithm

2 3D Region-Based Matching Algorithm

2.1 Offline Preprocessing

A face is first detected in its 2D image using Viola and Jones' algorithm [13] and both the 2D and its corresponding 3D images are cropped. The resolution of FRGC 3D faces is very high (480×640). Therefore, we downsampled the spatial data (by a factor of $\frac{1}{4}$) by eliminating alternate rows and columns. Each 3D face is preprocessed to remove spikes and noise. Spikes are removed by converting the 3D pointcloud of a face into a triangular mesh and removing triangles with long edges. This is followed by the elimination of disconnected points. The resulting mesh is smoothed using Taubin's algorithm [11] (50 iterations using the mesh toolbox of The Robotics Institute, CMU). Unlike the BEE baseline algorithms, which rely on prespecified landmarks on the gallery as well as the probe faces for their normalization, **our region-based matching approach is fully automatic** and requires no user intervention during the online recognition phase. However, during offline preprocessing, we manually identify six landmarks on each gallery face for its segmentation (Fig. 2). Note that this does not affect the automatism of our approach since this operation is only performed during the offline preprocessing phase. Moreover, this identification of landmarks can be replaced with an automatic feature detection algorithm [2] in order to automate the offline process as well.

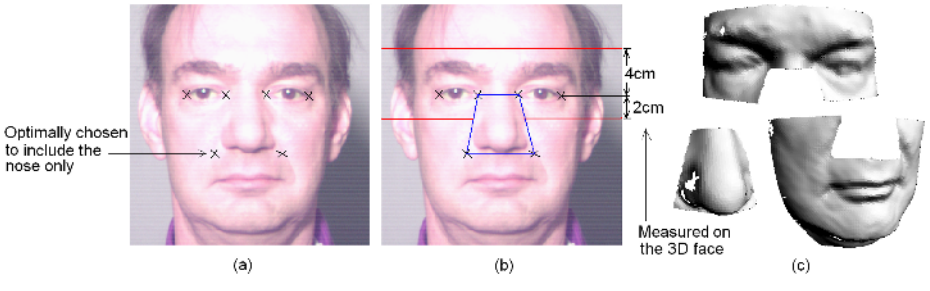

Fig. 2. Six points are manually identified on a gallery face (i.e. off-line) to segment its corresponding 3D face into three disjoint regions i.e. eyes-forehead, nose and cheeks (reproduced from [9]).

Fig. 3. (a) 2D coloured image of an eyes-forehead region. (b) Skin map after skin detection. (c) The corresponding 3D eyes-forehead region before skin detection and after skin detection (d). Note that small holes have been interpolated (reproduced from [9]).

The identified landmarks are used to segment the gallery face into three disjoint regions, namely the forehead, the nose, and the cheeks (Fig. 2). The forehead region may contain artifacts caused by the eyes and hair (see Fig. 3-a and c). The latter causes more problems since it covers more area. To remove these artifacts a skin detection algorithm [3] is used to detect the skin pixels in the 2D coloured image of the forehead region. Points in the 3D forehead region which do not correspond to the skin pixels are removed (see Fig. 3-d). The threshold for skin detection is tuned to minimize false positives so that the non-skin pixels are rejected with a high probability.

2.2 Online Nose Matching

During online recognition, the probe face is first detected using Viola and Jones' algorithm [13] and preprocessed as explained in Section 2.1. However, a prior segmentation of the probe is not required. This makes the online matching process fully automatic. Next, each gallery nose is registered to the probe for matching. Registration is performed in two steps. First, a gallery nose is coarsely registered to the probe nose by aligning their ridge lines and points of maximum slope. Next, the registration is refined with our modified version of the ICP algorithm [1] (explained in Section 2.3). Advantages of using this algorithm are that the gallery and probe need not cover exactly the same area of the face nor are they required to have the same resolution. A gallery nose, for example, can be registered to the nose of a complete probe face without having to segment the probe (see Fig. 5). The only requirement is that the probe nose must first be detected in order to coarsely register the gallery nose to it. For this purpose, the ridge line of the probe nose and its point of maximum slope are detected as follows. First, the 3D probe face is horizontally sliced at different vertical positions and a cubic spline is then passed through the points of each slice in order to accurately detect the peak of the slice. Next, a line is passed through the peak points of all the slices using RANSAC. This line forms the nose ridge. Since not every slice may contain the nose, a decision is made on the basis of the spline curvatures, the side lengths and area of the triangle (Fig. 4) on whether or not a slice contains the nose. Once the nose ridge is detected, the probe is vertically sliced along the nose ridge, a cubic spline is passed through the slice

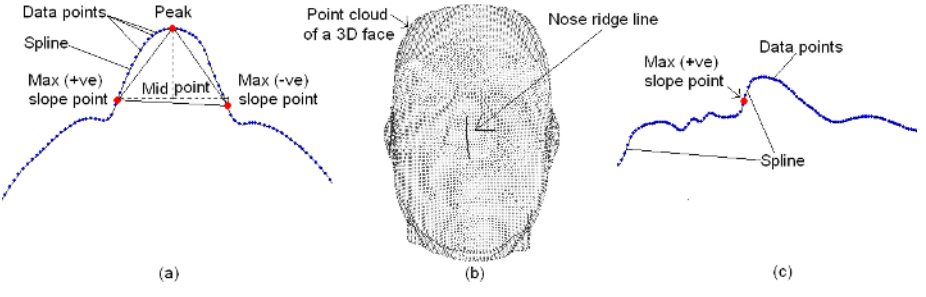

Fig. 4. Online nose detection in the 3D probe face (reproduced from [9])

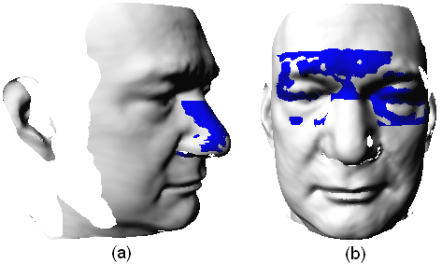

Fig. 5. A correct match between a 3D probe face (shaded white) and a gallery (shaded dark blue) (a) nose and (b) forehead. Note that some holes may have been interpolated.

data points and the point of maximum slope is detected on the spline. The ridge lines and points of maximum slope of the probe and gallery noses are aligned and the registration is then refined with our modified version of the ICP algorithm [1]. The average registration error e_N normalized with the resolution of the probe is taken as the matching score between the two (a lower value of e_N means a better match). Fig. 5-a shows a gallery nose registered to a complete probe face.

2.3 Online Forehead Matching

Forehead matching is performed by registering each gallery forehead to the probe without segmenting the probe. This process is also fully automatic. Since, the forehead and nose of each gallery face exist in the same coordinate basis, the rigid transformation resulting from the nose matching can be used as the coarse registration for the forehead matching. This registration is further refined with our modified version of the ICP algorithm. Our modified ICP algorithm establishes correspondences between the nearest points of the probe and a gallery region whose mutual distance is below a threshold t_c. Correspondences whose mutual distance is more than t_c are considered outliers and are therefore removed. A high initial value (four times the resolution of the probe) is chosen for t_c which is then reduced as the registration is refined. To speed up the correspondence search, a kd-tree data structure is used. After few iterations (when the registration error falls below a threshold), a region of interest (within the neighborhood of the gallery region) is cropped in the probe face to gain further computational efficiency. A conservative threshold is chosen for this purpose to avoid removing the overlapping region of the probe. The registration is refined iteratively until the correspondences between the probe and the gallery reaches a maximum saturation value. Next, t_c is further reduced and the above process is repeated. At the final stage, the stopping criterion of the algorithm is changed to the minimization of the registration error. Moreover at this stage, in the case of the forehead only, the correspondences are established between points which are close in the xy plane. In other words, correspondences are established along the approximate viewing direction. Points which are close in the xy plane but far in the z-dimension are still considered corresponding points as they provide

useful information regarding the similarity or dissimilarity of the gallery and the probe. Fig. 5-b shows a gallery forehead registered to a complete probe face.

3 Online Holistic 3D and 2D Face Matching

Holistic 3D and 2D face matching was performed using the BEE (Biometric Experimentation Environment) baseline PCA-based algorithm [12]. For 2D face matching, the texture maps accompanied with the 3D face data were used, whereas for 3D face matching, the range image (see Section 1.1) of the 3D faces was used. The parameters of the PCA algorithm were separately tuned in the case of 2D and 3D face matching in order to maximize their individual performance (i.e. verification rate at 0.001 FAR). The BEE algorithm utilizes prespecified landmarks on the faces, which are provided along with the data in a metadata file, for their normalization. Note that we did not use these landmarks for the online region-based matching as it is fully automatic and does not require manually specified landmarks. The BEE algorithm normalizes the faces with respect to pose and illumination and scales them to 150×130 spatial pixels each. A mask is used to crop out unwanted pixels. The normalized faces are then projected onto the PCA space and matched using the Mahalanobis distance.

4 Fusion

Each matching process results in a similarity matrix $\mathbf{S}_i$ (where i denotes a modality) of size $P \times G = 668 \times 275$ (where P is the number of tested probes and G is the number of faces in the gallery). An element s_{rc} (at row r and column c) of a matrix $\mathbf{S}_i$ denotes the similarity score between probe number r and gallery face number c. Each $\mathbf{S}_i$ has a negative polarity in our case i.e. a smaller value of s_{rc} means high similarity. The similarity matrices resulting from the 3D nose matching, 3D forehead matching, 3D holistic face matching and 2D holistic face matching are normalized before fusion. Since none of the similarity matrices had outliers a simple min-max rule (Eqn. 1) was used for normalizing each one of them on a scale of 0 to 1.

$$\mathbf{S}'_i = \frac{\mathbf{S}_i - \min(\mathbf{S}_i)}{\max(\mathbf{S}_i - \min(\mathbf{S}_i)) - \min(\mathbf{S}_i - \min(\mathbf{S}_i))} \quad \text{where} \quad i = 1 \ldots n \tag{1}$$

$$\mathbf{S} = \prod_{i=1}^{n} \mathbf{S}'_i \tag{2}$$

$$\mathbf{S}' = \frac{\mathbf{S} - \min(\mathbf{S})}{\max(\mathbf{S} - \min(\mathbf{S})) - \min(\mathbf{S} - min(\mathbf{S}))} \tag{3}$$

In Eqn. 1 and Eqn. 3, $\max(\mathbf{S}_i)$ and $\min(\mathbf{S}_i)$ mean the overall minimum and maximum value (i.e. a scalar) of the entries of matrix $\mathbf{S}_i$ respectively. In Eqn. 2, n is the number of modalities used. The normalized similarity matrices $\mathbf{S}'_i$ are then fused using a multiplication rule (Eqn. 2) to get a combined similarity

matrix **S** which is normalized using the min-max rule (Eqn. 3) once again. **S′** is used to calculate the combined performance of the used modalities. We also tested a weighted sum rule for fusing the similarity matrices with the weights adjusted according to the confidence in each modality. However, this technique gave slightly worse results compared to the multiplication rule.

5 Results

The FRGC v1.0 [10] data contain multiple high resolution 2D images of 275 subjects. For Experiment 1 and 2, 2D images of subjects are acquired with a high resolution camera (4 megapixel) under controlled illumination and expressions. For Experiment 3, multiple 3D snapshots (3D shape and texture) of 275 subjects are acquired under a controlled environment. For the sake of comparison, we fixed the gallery size to 275 subjects and the number of tested probes to 668 for each experiment. For each experiment, a brute force matching approach was used i.e. every probe was matched with every gallery face to get a 668×275 similarity matrix. Although brute force matching could have been avoided through the use of some indexing scheme, it was still performed to get dense non-match scores [10] in order to derive statistically more significant inferences.

5.1 Multimodal Hybrid Face Recognition Results

Fig. 6 shows the individual and combined results Receiver Operating Characteristic (ROC) curve and identification rate) of our multimodal hybrid face recognition on Experiment 3. The region based matching was performed using our modified ICP algorithm whereas the holistic face recognition was performed using the BEE baseline PCA algorithm. Our results show that the forehead shape matching has the best performance which indicates that this region contains the most discriminating information regarding a face. These results support the findings of Zhao et al. [14] which state that the upper part of the face is more

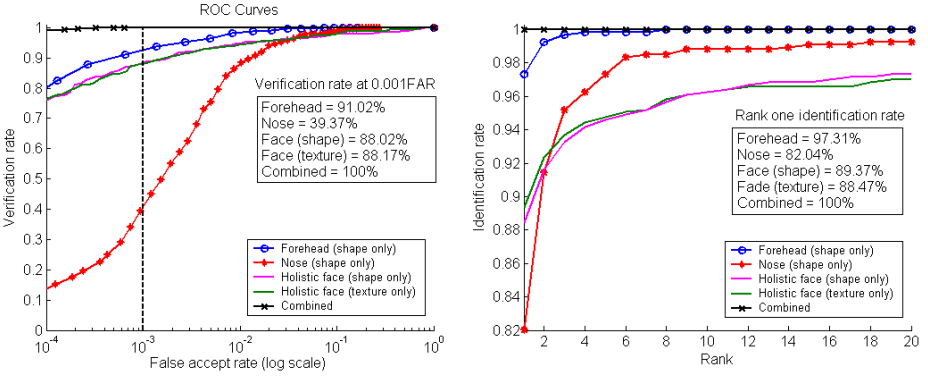

Fig. 6. Individual and combined results of the multimodal hybrid face recognition. The holistic face recognition is performed using the BEE baseline PCA algorithm.

important for recognition compared to the lower part. As expected, the performance of the nose shape matching is the lowest due to the sparsity of information in this region. However, a significant improvement in performance is achieved when the results of the nose are fused with those of the forehead (see Fig. 8). The overall performance of our multimodal hybrid face recognition algorithm is very high with a combined verification rate of 100% reached at 0.0006 FAR (well below the 0.001 FAR benchmark of FRGC).

One of the important findings of our results is that when the region-based (forehead and nose) shape matching results are fused with the holistic face matching (shape only) results, no significant improvement in performance is achieved. Therefore, we excluded the holistic 3D face matching while comparing our results to the BEE baseline performances on Experiment 1 and 2.

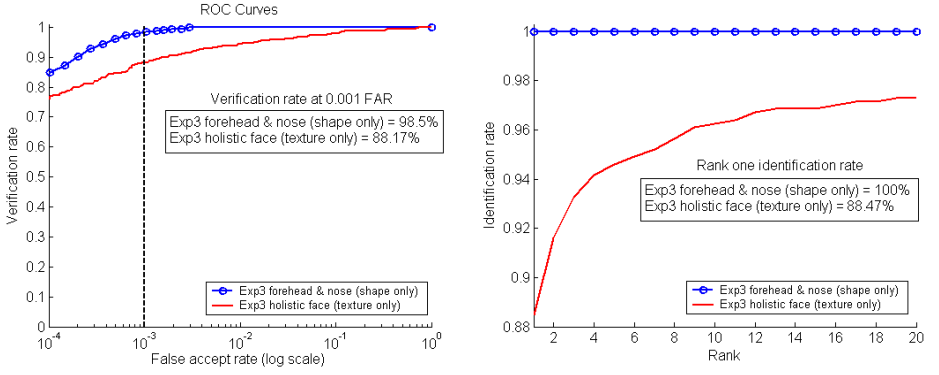

Fig. 7. Conjecture 1-A (Supported): Experiment 3 (shape only) performance is better than Experiment 3 (texture only) when region-based matching of shape is performed with our modified ICP algorithm and the BEE baseline PCA algorithm is used for holistic matching of texture.

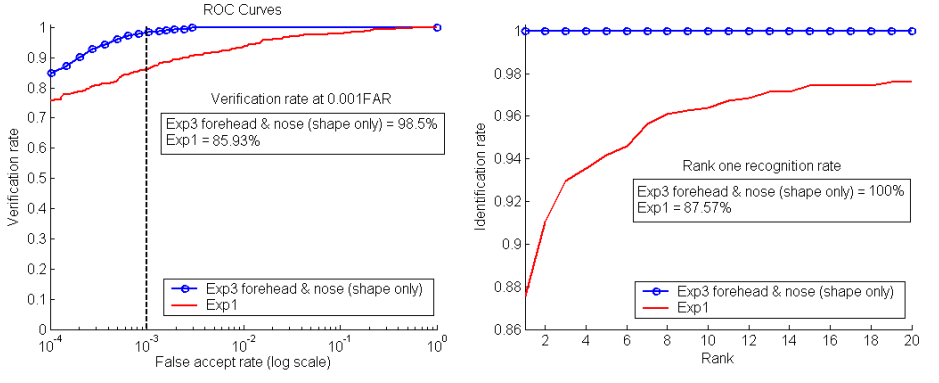

Fig. 8. Conjecture 1-D (Supported): Experiment 3 (shape only) performance is better than Experiment 1 when region-based shape matching is performed with our modified ICP algorithm and BEE baseline PCA algorithm is used for holistic 2D face matching.

5.2 Conjecture I and II

Fig. 7 and 8 show that Experiment 3 (shape only) using the region-based approach gives far better verification and identification performance compared to Experiment 3 (texture only) as well as Experiment 1 using the BEE baseline PCA algorithm. Moreover, Fig. 9 shows that Experiment 3 (shape and texture) using our multimodal hybrid matching approach gives better performance compared to Experiment 1 using the BEE baseline PCA algorithm. These results support Conjecture I-A, I-D and I-E respectively and oppose Conjecture II. Put another way, 3D face recognition (with or without texture) significantly outperforms PCA based 2D face recognition (at equal or higher resolution).

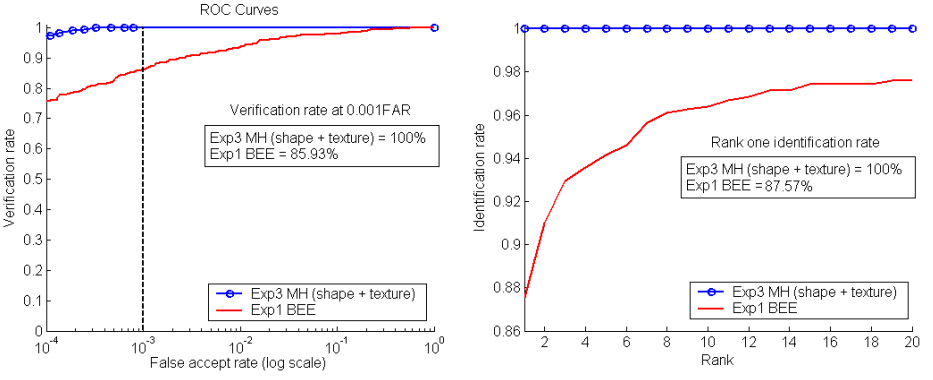

Fig. 9. Conjecture 1-E (Supported): Experiment 3 (shape and texture) performance using our multimodal hybrid (MH) face recognition algorithm is better compared to Experiment 1 using the BEE baseline PCA algorithm.

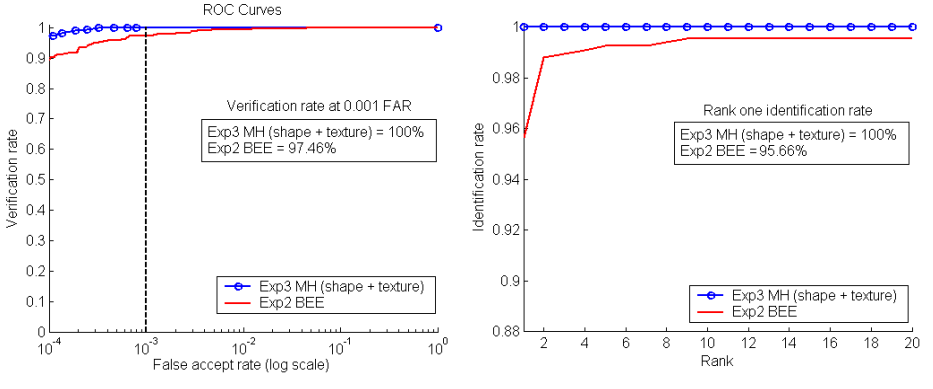

Fig. 10. Conjecture 3-A (Opposed). Experiment 3 (shape and texture) performance using our multimodal hybrid (MH) face recognition algorithm performance is better than the BEE baseline performance on Experiment 2.

5.3 Conjecture III

Fig. 10 and Fig. 11 show that regardless of whether texture information is used, our algorithm outperforms the BEE baseline performance on Experiment 2 where four high resolution images per gallery and probe are used. These results oppose Conjecture III-A and III-B. In other words, our results show that 3D face recognition (with or without texture) outperforms PCA based 2D face recognition even when multiple high resolution images per face are used in the latter case.

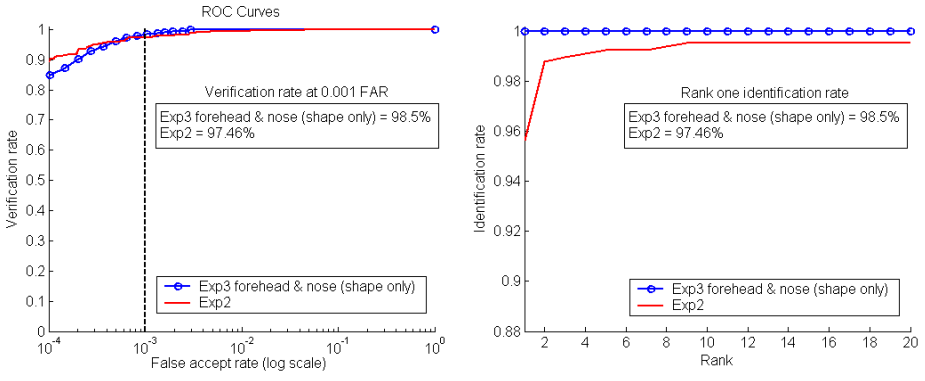

Fig. 11. Conjecture 3-B (Opposed). Experiment 3 (shape only) performance using region-based shape matching with our modified ICP algorithm is better than the BEE baseline performance on Experiment 2.

6 Analysis and Conclusion

We presented a multimodal hybrid face recognition algorithm and demonstrated its performance on the FRGC v1.0 dataset. The results were compared to the BEE baseline performances on Experiment 1, 2 and 3 in order to answer the important conjectures of the FRGC which compare 2D and 3D face recognition. Our results show that our multimodal hybrid algorithm significantly outperforms others by achieving 100% verification rate at 0.0006 FAR. The fact that our region-based 3D face recognition performance alone is better than 2D face recognition using multiple high resolution images per face in the latter case is a strong indicator of the potential of 3D face recognition. A number of additional important conclusions can be drawn from our results. Firstly, the eyes-forehead region of a face contains the maximum discriminating features important for face recognition. Secondly, the nose which apparently plays an insignificant role in face recognition from 2D frontal views [14], plays a more significant role in the 3D case. Finally, fusing the results of an inferior classifier (e.g. holistic 3D face matching using PCA) with that of a superior classifier (e.g. region-based 3D face matching using ICP) when operating on the same modality (i.e. 3D face in our example) does not improve performance as much as when the results of classifiers which operate on different modalities (e.g. 3D and 2D face) are fused. Recall that the region-based 3D face

matching was performed at $\frac{1}{4}$ of the original resolution of the FRGC data. Using full resolution is likely to further improve the performance. Based on our findings we would like to add the following conjecture to the FRGC.

Conjecture MBO: Exploiting multimodal hybrid matching techniques has the potential to give the best face recognition performance.

Our conjecture gives rise to a number of questions. What combination of modalities (e.g. 2D face, 3D face and IR image of the face) should be used? For each modality, what is the best possible segmentation of the face to perform region-based matching? What is the best matching algorithm for each modality and region? Finally, what fusion technique will produce the best results? These questions give directions for focusing future research of the FRGC.

Acknowledgments

We would like to thank Ashley Chew for setting up the Biometrics Experimentation Environment, CMU for the Mesh Toolbox and F. Boussaid for the skin detection code. This research is sponsored by ARC grant DP0664228.

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. C. Boehnen and T. Russ, "A Fast Multi-Modal Approach to Facial Feature Detection", *IEEE WACV*, 2005.
3. F. Boussaid, D. Chai and A. Bouzerdoum, "A Current-mode VLSI Architecture for Skin Detection", *ISSPA*, Vol. 1, pp. 629 – 632, 2003.
4. K. W. Bowyer, K. Chang and P. 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. J. Huang, B. Heisele and V. Blanz, "Component-based Face Recognition with 3D Morphable Models", *AVBPA*, 2003.
7. A. K. Jain, A. Ross and S. Prabhakar, "An Introduction to Biometric Recognition," *IEEE TCSVT*, Vol. 14(1), pp. 4–20, 2004.
8. X. Lu, A. K. Jain and D. Colbry, "Matching 2.5D Face Scans to 3D Models," *IEEE TPAMI*, Vol. 28(1), pp. 31–43, 2006.
9. A. S. Mian, M. Bennamoun and R. A. Owens, "Region-based Matching for Robust 3D Face Recognition," *BMVC*, Vol. 1, pp. 199–208, 2005.
10. P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min and W. Worek, "Overview of the Face Recognition Grand Challenge", *IEEE CVPR*, 2005.
11. G. Taubin, "Curve and Surface Smoothing without Shrinkage," *ICCV*, pp. 852–857, 1995.
12. M. Turk and A. Pentland, "Eigenfaces for Recognition", *JOCN*, Vol. 3, 1991.
13. P. Viola and M. J. Jones, "Robust Real-Time Face Detection", *IJCV*, Vol. 57(2), pp. 137–154, 2004.
14. W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld, "Face Recognition: A Literature Survey", *ACM Computing Survey*, pp. 399-458, 2003.