A Region Ensemble for 3-D Face Recognition

Timothy C. Faltemier, Kevin W. Bowyer, and Patrick J. Flynn, *Senior Member, IEEE*

***Abstract*—In this paper, we introduce a new system for 3-D face recognition based on the fusion of results from a committee of regions that have been independently matched. Experimental results demonstrate that using 28 small regions on the face allow for the highest level of 3-D face recognition. Score-based fusion is performed on the individual region match scores and experimental results show that the Borda count and consensus voting methods yield higher performance than the standard sum, product, and min fusion rules. In addition, results are reported that demonstrate the robustness of our algorithm by simulating large holes and artifacts in images. To our knowledge, no other work has been published that uses a large number of 3-D face regions for high-performance face matching. Rank one recognition rates of 97.2% and verification rates of 93.2% at a 0.1% false accept rate are reported and compared to other methods published on the face recognition grand challenge v2 data set.**

***Index Terms*—Biometric, expression variation, fusion methods, multiinstance, range image, 3-D face recognition.**

I. Introduction

FACE recognition in 3-D has been addressed using a variety of methods, including alignment, subregion matching, mapping, and principal component analysis (PCA). The Face Recognition Grand Challenge (FRGC) v2 data set [1] is the largest publicly available data set for 3-D face recognition research. This set contains images exhibiting substantial expression variation, which can cause problems for many recognition algorithms. Our approach exploits subregions on the face that remain relatively consistent in the presence of expressions and uses a committee of classifiers based on these regions to improve performance in the presence of expression variation. This concept is called region ensemble for face recognition (REFER). This paper extends the work in [2] by increasing the number of facial regions considered to 38, selecting the best-performing subset of size 28, and discussing results achieved through the use of additional fusion methods found in the literature [3], [4]. Each region matches independently to a gallery surface using the iterative closest point (ICP) algorithm [5], resulting in a committee of error distances for a single probe-to-gallery comparison. Based on a threshold for each match, regions are selected to vote for a given identity. The results presented in this paper significantly outperform those in our previous work [2].

The FRGC v2 [1] data set includes a 3-D shape for each of 4007 face scans of 466 unique subjects. This data set has been distributed to organizations across the world to promote consistency in experimental evaluation of face recognition systems [6]. The FRGC program sought to document an increase in face recognition performance by an order of magnitude from the previous face recognition vendor test (FRVT 2002) [7]. The top FRVT 2002 performance on 2-D face recognition was an 80% verification rate at a false accept rate (FAR) of 0.1%. The goal of the FRGC program was to increase the verification rate to 98% at a fixed FAR of 0.1%. For all verification experiments in this paper, we report results using an operating point of $\mathrm{FAR} = 0.1\%$. We show improved 3-D face recognition performance over previously published papers [8]–[12] on the FRGC v2 data set, using the same experiments as defined in those papers.

This paper is organized as follows. Section II gives an overview of related work in the area of face recognition. Section III discusses experimental materials and methods. Next, the REFER method for increasing face recognition performance in the presence of nonneutral facial expressions is introduced in Section IV. Section V discuses our experimental results and compares them to other published results on the FRGC v2 data set [1]. Finally, Section VI provides conclusions and discussion.

II. Related Work

A recent broad survey of face recognition research is given in [13] and a survey focusing specifically on face recognition using 3-D data is given in [14]. This section focuses on selected prior work that is most closely related to our current work.

Chang *et al.* [8] use multiple overlapping nose regions and obtain increased performance relative to using one whole-frontal-face region. These regions include a nose circle, nose ellipse, and a region composed of just the nose itself. This method uses the ICP algorithm to perform image matching and reports results on a superset of the FRGC v2 data set containing 4485 3-D face images. Two-dimensional skin detection is performed for automated removal of hair and other nonskin-based artifacts on the 3-D scan. They report results of 97.1% rank one recognition on a neutral probe matched to neutral gallery images and 87.1% rank one recognition on nonneutral probes matched to a neutral gallery. The product rule was used to process the results from multiple regions. When the neutral probe was matched to the neutral gallery set, maximum performance was reported when

Manuscript received March 21, 2007; revised August 30, 2007. This work was supported in part by the Central Intelligence Agency, in part by the National Science Foundation under Grant CNS01-30839, in part by the US Department of Justice/National Institute for Justice under Grants 2005-DD-CX-K078 and 2006-IJ-CX-K041, in part by the National Geo-spatial Intelligence Agency, and in part by UNISYS Corp. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Davide Maltoni.

The authors are with the Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46545 USA (e-mail: tfaltemier@progeny.net; kwb@cse.nd.edu; flynn@cse.nd.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2007.916287

only two of the three regions were combined. The authors mention that increased performance may be gained by using additional regions, but they did not explore anything beyond three overlapping nose regions.

Lu *et al.* [15] combined multiple 2.5-D scans to create a single 3-D image for gallery enrollment. For their experiments, they employed 598 2.5-D probe models of various poses and expressions that were matched to 200 gallery models. They found that by using the full 3-D image of a subject in the gallery and implementations of the ICP and linear discriminant analysis (LDA) algorithms, they were able to achieve a 90% rank one recognition rate using a probe data set consisting of arbitrary poses. The authors report that nearly all of the errors in recognition were caused by a change in expression between the probe and the gallery images. In [16], Lu and Jain present an algorithm for matching 2.5-D scans in the presence of expressions and pose variation using deformable face models. A small control set is used to synthesize a unique deformation template for a desired expression class (smile, surprise, etc.). A thin-plate-spline (TPS) mapping technique drives the deformation process. The deformation template is then applied to a neutral gallery image to generate a subject-specific 3-D deformation model. The model is then matched to a given test scan using the ICP algorithm. The authors report results on three different types of experiments. The first data set contains ten subjects with three different poses and seven different expressions. Rank one results of 92.1% are reported when deformation modeling is used compared to 87.6% when it is not. The second data set consists of 90 subjects in the gallery and 533 2.5-D test scans and similar results are reported. Data for the first two experiments were gathered at the authors' institution. The data for the final experiment were taken from the FRGC v2 data set and consisted of 50 randomly chosen subjects in the gallery and 150 2.5-D test scans. When deformation modeling is employed, a rank one recognition rate of 97% is reported compared to 81% when it is not.

Martinez [17] uses multiple local region patches to perform 2-D face recognition in the presence of expressions and occlusion. The motivation for this is that different facial expressions influence different parts of the face more than others. His algorithm addresses this belief by weighting areas that are less affected by the current displayed emotion more heavily. Reported results show that up to one-sixth of the face can be occluded without a loss in recognition, and one-third of the face can be occluded with minimal loss. This work used the well-known AR database of 2-D images.

Heisele and Koshizen [18] demonstrate a novel approach for face recognition that combines the techniques of 3-D morphable models and component-based recognition. The authors use three 2-D face images to generate a 3-D head model of each subject. That subject is then rendered under varying illumination and pose conditions to build a large gallery of synthetic images. Recognition is performed on a single global image as well as 35 individual facial components. Results show that when fused, the individual facial components perform better than the single global image. For this experiment, results were reported on 2000 images (ten subjects, 200 images per subject) and were collected at the authors' institution.

Gökberk and Akarun [3] perform a comparative evaluation of five face-shape representations (point clouds, surface normals, facial profiles, PCA, and LDA) using the well-known 3-D-RMA data set [19] of 571 images from 106 subjects. They find that the ICP and LDA approaches offer the best average performance. They also perform various fusion techniques for combining the results from different shape representations to achieve a rank-one recognition rate of 99.0%.

As with Chang [8] and Lu *et al.* [15], we are specifically interested in 3-D face recognition in the presence of varying facial expression between gallery and probe images. Since Martinez [17] and Heisele and Koshizen [18] found good performance in 2-D face recognition by matching a large number of face regions, we consider something similar for 3-D. Whereas Chang considered just three regions, all overlapping the nose area, we initially consider 38 regions, representing various areas of the face. We consider a variety of approaches for fusion of the committee results, and find that different methods than previously used [8], [18] give the best results. We also find better performance than a number of previous papers that also use the FRGC v2 data set.

III. Experimental Materials and Methods

A. FRGC v2 3D Data Set

The data used in our experiments come from the FRGC v2 data set. This data set includes 4007 3-D face images of 466 distinct persons, with as many as 22 images per subject. The images were acquired with a Minolta Vivid 910 [20]. The Minolta 910 scanner uses triangulation with a laser stripe projector to build a 3-D model of the face. Both color (r, g, b) and 3-D location (x, y, z) coordinates are captured, but not perfectly simultaneous, and the laser stripe requires a few seconds to cross the face. The resolution on the Minolta camera is 640×480, yielding approximately 300 000 possible sample points. The number of 3-D points on a frontal image of the face taken by the Minolta camera is around 112 000, and depends on the lens used as well as standoff. Additional vertices arise from hair, clothing, and background objects. A limitation of this data set is that only minor lighting and pose variations are present. Sample images from this sensor can be seen in Fig. 1.

Our experiment uses all 4007 3-D face images in the validation partition of the FRGC v2 data set. One-thousand five-hundred thirty-eight of the 4007 images have nonneutral expressions (e.g., disgust, happiness, surprise). Examples of images displaying neutral and nonneutral expressions can be seen in Fig. 2. A source of complications is found in images containing shape artifacts due to subject motion during scanning. Examples of these images containing artifacts can be seen in Fig. 3. Visual inspection of the FRGC v2 data set found a total of 13 images that contain significant shape artifacts. Texture misregistration can also occur but it is not relevant to this work.

Maurer *et al.* manually separated the FRGC v2 data set into three different categories [21] based on the subject's expression.

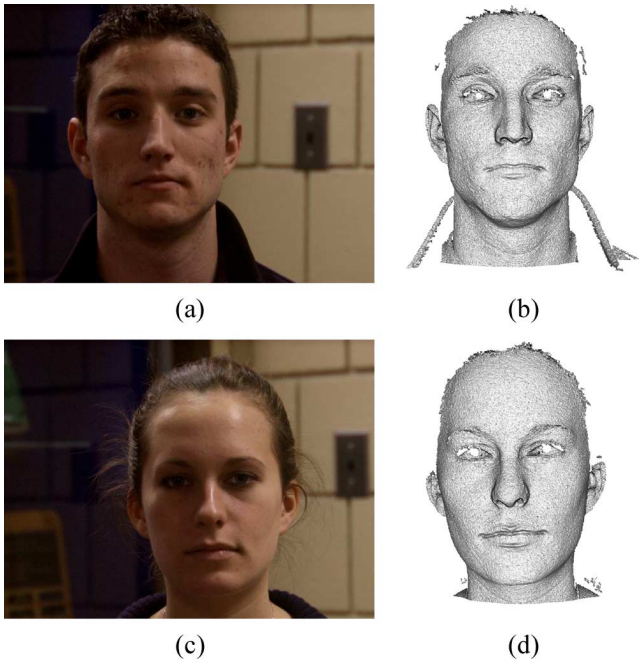


Fig. 1. (a), (c) Samples of images captured with the Vivid 910 by Minolta. (b), (d) 2-D data for two different subjects associated with 3-D shape information. (a) 04831d152. (b) 04831d152. (c) 04701d157. (d) 04701d157.

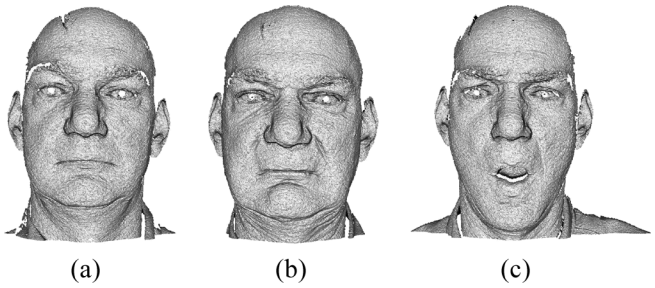


Fig. 2. Examples of image categories manually labeled by Geometrix. (a) Neutral image. (b) Small expression. (c) Large expression. (a) 02463d550. (b) 02463d560. (c) 02463d666.

These categories were based on visual inspection of each image for the amount of expression present. They classified 2469 images as "neutral," 796 images as "small expressions," and 742 images as "large expressions." They report results [9] based on these three image categories.

B. Data Preprocessing

Fig. 4 shows a diagram of the REFER algorithm when it is used in a verification scenario. Specific details of the algorithm will be given. Table I shows the running time of the REFER algorithm in a verification scenario. From preprocessing to final decision, the verification process takes less than 10 s on a 2.4-GHz Pentium IV processor with 1 GB of memory. This suggests that a feasible execution time may be achieved for use at airport checkpoints (X-rays or check in), bank ATMs, or building security scenarios, where the user may expect a slight authentication delay.

Our algorithm operates automatically by using only the 3-D shape from a frontal view of the face. First, small holes in the range image are filled by locating "missing" points that are surrounded by four or more "good" points. The x, y, and z coordinates of the missing point are interpolated from its valid neighbors. Boundary points are initially determined by a sweep through the range image, row by row, to find the first and last valid 3-D point. This process is repeated until no additional points are created.

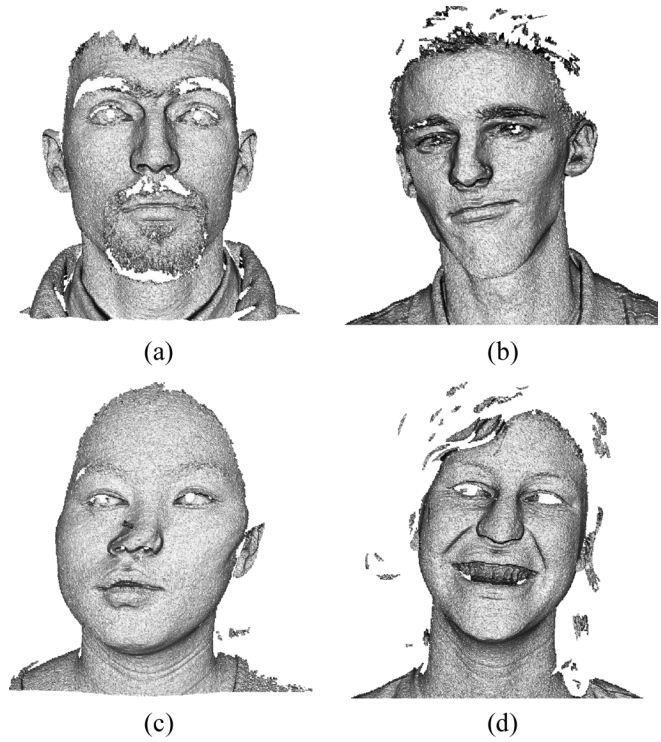


Fig. 3. Examples of images that contain artifacts in the FRGC v2 data set. (a) 04475d114. (b) 04749d72. (c) 04760d76. (d) 04812d42.

Once hole filling is complete, a final pass over the range image with a 3×3 median filter smoothes the data, removing spikes in the z-coordinate.

Finally, the nose tip point is detected using a consensus of two or three methods previously discussed [2]. The first method calculates the curvature and shape index [22] at each point on the face in the range image to find possible nose tip candidates labeled c_n. The second method aligns the input image to a template using the ICP algorithm. The position of the nose tip is the highest Z value in the image after alignment and is labeled p_n. If the candidate nose tip points found in these methods are within 20 mm of one another, then we report the final nose tip location b_n as the average of c_n and p_n. Otherwise, a tiebreaker step is performed. The position of the nose tip is known on the template image and when properly aligned, this point should be the nose tip of the input image as well and is labeled m_n. Distances between m_n, p_n, and c_n are calculated and the pair with the smallest distance is averaged and reported as b_n, the best nose tip. The majority of images causing this algorithm to fail were due to scanner misregistration and hair artifacts. Visual analysis of this algorithm applied to the 4007 images in the FRGC v2 data set determined that 3935 of the nose tips were less than 10 mm away from the manually marked nose tips.

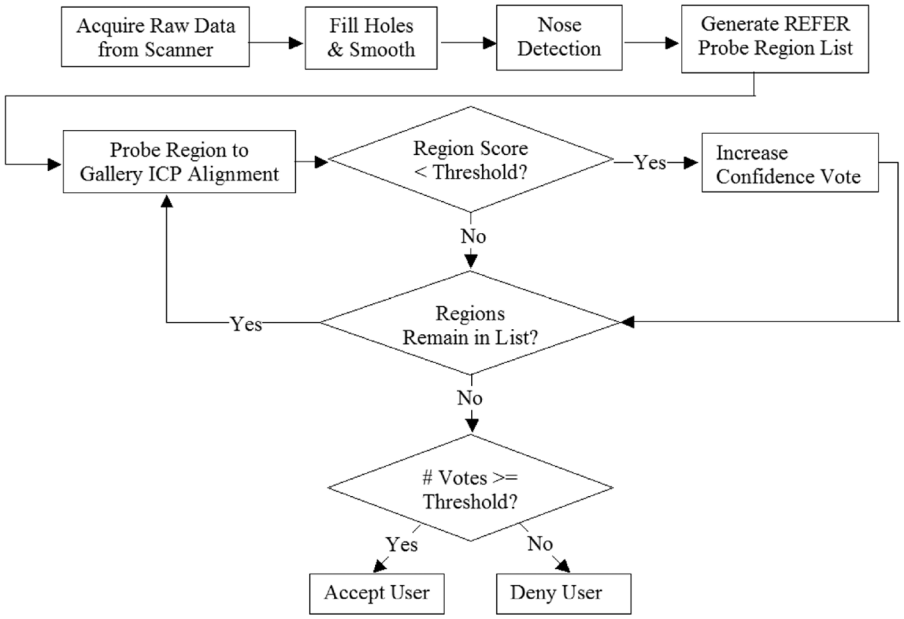


Fig. 4. Diagram representing the required operations for the REFER algorithm.

TABLE I
RUNNING TIMES OF THE REFER ALGORITHM

Step	Time (in ms)
Data Preprocessing	7,520
REFER Matching (28 regions)	2,380
Result Fusion (CV)	20

IV. REGION COMMITTEE VOTING APPROACH

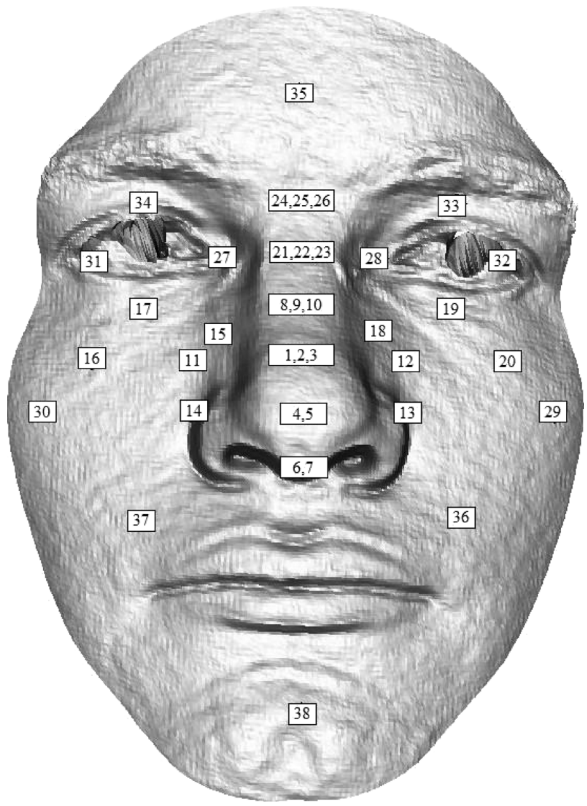
A. Region Extraction

Once the nose tip is successfully found in the preprocessing step, we translate the incoming image to the origin and crop a gallery region, which is defined by a sphere radius of 100 mm centered at the nose tip. To find the best committee of local regions for maximum results, we consider 38 regions on the face, some of which overlap, whose centroids are shown in Fig. 5. Examples of the relative radius sizes used in our experiments are seen in Fig. 6. Together, this collection of regions densely covers the face. We experimented with additional regions surrounding the mouth; however, none led to additional performance gains.

Table II shows the cropping parameters used to generate each region. OffsetX and OffsetY determine the new sphere center in relation to the origin. SphereRadius determines the new sphere radius. By selecting multiple small regions on the face, any errors caused by a single region can be compensated for when combining the matching scores from the other regions making the system more robust to image artifacts, wrinkles, facial hair, or expression variations.

B. Individual Region Matching

Once all of the regions have been cropped and the ICP algorithm is run on each probe-to-gallery combination, we determined how well each individual region performed in a verification experiment with the FAR fixed at 0.1% (the FRGC performance point [1]) and an identification experiment reporting the rank-one recognition rate. A baseline region (a sphere cropped at 100 mm from the nose tip) is also included to show the effects of matching the entire face to the gallery rather than using smaller regions. The results in Table III show that no individual

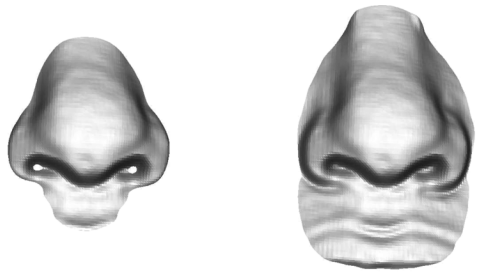


(a)

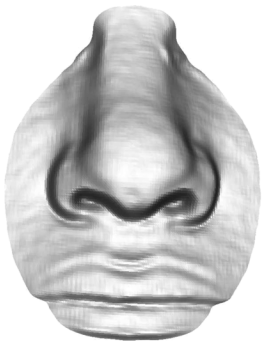
Fig. 5. Image of probe sphere centroids used in this experiment labeled by region number. Multiple region numbers at a centroid indicate that more than one radius was used for cropping, yielding multiple region probes with the same centroid. (a) 04514d324.

TABLE II
ICP PROBE REGION INFORMATION

Region	offsetX	offsetY	SphereRadius	Region	offsetX	offsetY	SphereRadius
1	0	10	25	20	40	10	45
2	0	10	35	21	0	30	40
3	0	10	45	22	0	30	35
4	0	0	25	23	0	30	45
5	0	0	45	24	0	40	40
6	0	-10	25	25	0	40	35
7	0	-10	35	26	0	40	45
8	0	20	35	27	-15	30	35
9	0	20	25	28	15	30	35
10	0	20	45	29	50	0	45
11	-20	10	25	30	-50	0	45
12	20	10	25	31	-40	30	45
13	20	0	25	32	40	30	45
14	-20	0	25	33	30	40	45
15	-15	15	45	34	-30	40	45
16	-40	10	45	35	0	60	35
17	-30	20	45	36	30	-20	35
18	15	15	45	37	-30	-20	35
19	30	20	45	38	0	-55	35
Full Frontal	0	0	100				



(a) 25mm Sphere Radius (b) 35mm Sphere Radius



(c) 45mm Sphere Radius

Fig. 6. Radius size examples (25 mm, 35 mm, 45 mm).

region is able to demonstrate performance greater than 84.8% TAR at 0.1% FAR, or 90.2% rank-one recognition. This motivates the use of fusion to further increase performance.

C. Combining Results From a Committee of Regions

Choosing the best fusion method when processing a large number of regions is a challenging task. In this paper, we have experimented with many of the fusion techniques described by Gökberk and Akarun [3] and Jain *et al.* [4].

The sum rule computes the final fusion score based on the formula $s_i = \sum_{k=1}^{K} s_{ki}, i = 1, \ldots, N$, where K is the number of regions and N is the number of images to process. Similarly, the product rule is based on the formula $s_i = \prod_{k=1}^{K} s_{ki}$, $i = 1, \ldots, N$. These methods can result in a large score range depending on the input values. Chang *et al.* [8] uses the product rule to generate their results. The min rule simply finds the smallest value in each of the K regions and reports it as the final score. This method is highly dependent on data normalization. Without this step, the regions with overall smaller similarity values will dominate the final scores. The consensus-voting (CV) method returns a vote for the closest match in the gallery for each region K. The image with the highest number of votes is declared the match. The Borda count (BC) technique [3], [4], [23]–[25] sorts all of the similarity scores in the gallery and adds the rank for each region; the image with the lowest sum of ranks is declared a match. The CV and BC methods both require knowledge of similarity scores from other individuals in the gallery to make a decision. Hence, the standard implementations of these two methods are unacceptable for use with a verification experiment, since only the match score(s) of the current subject are known.

In this paper, we use a modified version of the BC method. Unlike the standard BC method that provides a rank score (first place, second place, third place, ..., nth place) to each probe-to-gallery similarity score entry in the set from $1 \ldots n$, where n is the number of gallery images, we only give a rank score to the first α (in our experiments, $\alpha = 4$) entries in the gallery. A limitation of the BC method is that a probe region must be matched to each gallery image before the first α entries can be awarded a score. The traditional BC method takes the sum of the ranks to determine a final score. Our version distributes points such that the first place gets $(\alpha)^2$ points, the second gets $(\alpha-1)^2$ points, until the number of points is equal to 0. This modification yielded experimentally higher results than the traditional BC method. We believe this result is due to the larger weight given to rank-one matches. In our modification, a rank-one match and a rank three match outperform two rank two matches. An example of this method can be seen in Fig. 7. In this example, G1 is the gallery image that is a match to the incoming probe image. Regions R1, R2, and R3 are individual regions that are matched to each gallery. After distributing the points using the method

TABLE III
INDIVIDUAL REGION MATCHING PERFORMANCE

Region	TAR at 0.1% FAR	Rank-One Recognition Rate	Region	TAR at 0.1% FAR	Rank-One Recognition Rate
1	79.1%	79.2%	20	64.2%	73.1%
2	76.5%	83.2%	21	83.8%	**90.2%**
3	71.0%	82.6%	22	82.6%	89.6%
4	73.1%	80.5%	23	81.1%	89.8%
5	57.2%	74.2%	24	78.3%	87.9%
6	68.4%	78.1%	25	74.0%	84.8%
7	56.0%	72.0%	26	77.9%	88.2%
8	**84.8%**	89.1%	27	73.5%	86.3%
9	73.7%	70.9%	28	74.2%	87.3%
10	76.4%	86.2%	29	54.1%	67.5%
11	68.0%	64.6%	30	53.6%	66.8%
12	68.7%	66.2%	31	66.4%	82.0%
13	70.7%	70.4%	32	64.7%	80.7%
14	70.3%	73.6%	33	68.3%	81.9%
15	74.7%	85.1%	34	70.9%	83.4%
16	62.9%	75.4%	35	72.9%	83.0%
17	72.6%	84.1%	36	35.8%	40.6%
18	73.3%	84.8%	37	37.2%	42.9%
19	72.1%	83.4%	38	43.0%	48.8%
Full Frontal	60.9%	73.4%			

		Gallery Match Scores											
		G1 (Match)			G2			G3			G4		
		Similarity	Rank	Pts	Similarity	Rank	Pts	Similarity	Rank	Pts	Similarity	Rank	Pts
Probe Regions	R1	0.35	1	16	0.75	4	1	0.56	3	4	0.54	2	9
	R2	0.62	2	9	0.72	3	4	0.61	1	16	0.83	4	1
	R3	0.75	2	9	0.74	1	16	0.79	3	4	0.86	4	1
	Modified Borda Count		34			21			24			11	
	Final Rank		**1**			**3**			**2**			**4**	

Fig. 7. Example of the BC algorithm on a sample set of data.

TABLE IV
FUSION RULE RESULTS

Fusion Rule	Best Rank-One Recognition Rate	Committee Size	Overall Rank-One Recognition Rate (All 38 Regions)
Sum Rule	93.0%	19	61.8%
Product Rule	93.9%	6	72.8%
Min Rule	93.5%	11	75.2%
Consensus Voting	96.7%	27	96.2%
Borda Count (Traditional)	93.2%	5	90.1%
Borda Count (Modified)	**97.5%**	29	97.1%

TABLE V
ALL VERSUS ALL REFER ROC VALUES AT INCREASING CONFIDENCE LEVELS

Confidence Votes	False Accept Rate	Verification Rate
1	0.08	99.4%
2	0.009	96.9%
3	**0.001**	**93.2%**
4	0.0003	90.2%
5	0.0001	87.4%
6	0.00006	84.9%
7	0.000031	82.3%
8	0.000016	80.1%
9	0.000008	77.5%
10	0.000004	74.7%

described before, the final fusion ranks are displayed in the last row in bold.

For verification experiments, we train the set of region-specific thresholds $\{t_1, t_2, \ldots, t_i\}$ by using images from the FRGC v1 data set. The FRGC v1 data set consists of 943 images containing neutral expressions and is completely disjoint from the FRGC v2 data set. We believe that verification results presented in this paper could be further increased by using a more representative training set (that contains expressions).

Our committee-based algorithm is similar to the consensus voting method. There is a set of region-specific thresholds t_i that are independently tuned to optimize performance on each region. The t_i for each region is fixed once the desired operating point of 0.1% FAR is reached (by the tuning procedure). For each probe-to-gallery face match, we have m matching scores. We compare each matching score m_j to its corresponding region threshold t_i. If any matching score is below the threshold, we report this region as a correct match (1). If the distance is above the eligibility threshold, then we output this match as incorrect (0). If we wish to decrease the FAR, we simply increase the vote threshold required for a match to be considered correct. The number of votes for a particular match can be seen as a relative measure of confidence (the more votes an image receives, the greater confidence in our final decision).

V. EXPERIMENTAL RESULTS

In this paper, we perform two types of experiments. The first is a verification experiment, in which the system's performance is quoted as a true accept rate (TAR) at a given FAR. The second type of experiment is an identification experiment for which performance is quoted as a rank-one recognition rate.

TABLE VI
FUSION RECOGNITION RATE AS THE NUMBER OF REGIONS INCLUDED IN THE COMMITTEE INCREASES ONE BY ONE STARTING WITH THE HIGHEST INDIVIDUAL RECOGNITION RATE

Number of Regions	Region Added	Rank-One Recognition Rate	Number of Regions	Region Added	Rank-One Recognition Rate
1	21	90.2%	20	33	95.1%
2	23	90.7%	21	32	95.4%
3	22	91.2%	22	4	95.5%
4	8	91.6%	23	1	95.8%
5	26	92.0%	24	6	96.2%
6	24	92.4%	25	16	96.5%
7	28	92.7%	26	5	96.7%
8	27	92.9%	27	14	96.8%
9	10	93.2%	28	20	**97.2%**
10	15	93.5%	29	7	97.5%
11	18	93.7%	30	9	97.5%
12	25	93.8%	31	13	97.4%
13	17	93.9%	32	29	97.5%
14	34	94.0%	33	30	97.5%
15	19	94.1%	34	12	97.3%
16	2	94.2%	35	11	97.3%
17	35	94.5%	36	38	97.2%
18	3	94.8%	37	37	97.2%
19	31	95.0%	38	36	97.1%

TABLE VII
EXPERIMENT 1: VERIFICATION RESULTS

Paper	Experiment	Stated Performance Verification Rate at 0.1% FAR	Our Performance Verification Rate at 0.1% FAR
Mian *et al.* [11]	All Expressions *vs.* All Expressions	86.6%	93.2%
Maurer *et al.* [9]	All Expressions *vs.* All Expressions	87.0%	93.2%
Cook *et al.* [12]	All Expressions *vs.* All Expressions	92.31%	93.2%

The FRGC v2 [1] protocol defines a set of standard verification experiments for face recognition. For pure 3-D face recognition, Experiment 3 is most applicable. In the "ROC III" experiment, the gallery images come from the Fall 2003 semester and the probe entries come from the Spring 2004 semester. Thus, gallery images are guaranteed to have been taken before probe images and there is at least a month of time lapse between the gallery and the probe. The result of this experiment is a receiver operating characteristic (ROC) curve.

The second experiment ("All versus All") employs the entire FRGC v2 data set. All images appear in both the gallery and the probe sets. For this setup, we perform both verification and identification experiments. In the identification experiment, we take the earliest image of every participant and use it as the gallery image for that subject. All subsequent images for each subject are used as probes. The result is a cumulative match characteristic (CMC) curve. For the verification experiment, we match all 4007 images in the FRGC v2 data set to all 4007 images regardless of expression, which can be summarized in an ROC curve. Unlike the identification experiment, the gallery images in the all versus all experiment may not have been taken before the probe images.

In order to accurately compare the performance between the methods and experiments described in this section, a standard z-test [26] is performed on the respective results to determine if there is a statistically significance difference. The rank-one recognition and verification rate comparisons can be viewed as a binomial distribution problem. We denote the probability of a correct match (either correctly identifying or verifying a subject) by p and the probability of an incorrect match by $(1-p)$. With a sufficiently large sample size N, the binomial distribution converges to a normal distribution. For identification result comparisons, $\hat{p}$ is the probability that a subject is correctly identified from a gallery set. For verification result comparisons, $\hat{p}$ is the probability that a probe-to-gallery match results in a true accept at a given false accept rate (i.e., 93.2% at a 0.1% FAR would result in $\hat{p} = 0.932$ for that comparison). In general, verification experiments will have a significantly higher N as each comparison is included in the total number of samples. Identification experiments only contribute to a single sample for each probe-to-gallery set match. This causes experimental verification results to require a much smaller difference than identification results to be statistically significant. Given two results, with sample sizes N_1 and N_2, and the percentage of observed correct matches as $\hat{p_1}$ and $\hat{p_2}$, the test statistic using a 0.05 level of significance is

$$z = \frac{\hat{p_1} - \hat{p_2}}{\sqrt{\left(\frac{N_1+N_2}{N_1 N_2}\right)\left(\frac{X_1+X_2}{N_1 N_2}\right)\left(1 - \frac{X_1+X_2}{N_1 N_2}\right)}}.$$

$$X_1 = \hat{p_1} \times N_1, X_2 = \hat{p_2} \times N_2$$

To test the performance of each fusion method presented in the previous section, we created an experiment where the first image of each participant is in the gallery set, and all subsequent images are in the probe set. This yields 466 images in the gallery set and 3541 in the probe set. This ensures that every image in the probe set will have a possible correct match in the gallery. The rank-one recognition results of the fusion methods discussed in the previous section are listed in Table IV.

TABLE VIII
EXPERIMENT 2: IDENTIFICATION RESULTS

Paper	Experiment	Stated Performance Rank One Recognition Rate	Our Performance Rank One Recognition Rate
Mian *et al.* [11]	Earliest neutral scan as gallery remainder as probe	96.2%	98.1%
Chang *et al.* [27]	Earliest neutral scan as gallery remainder as probe (FRGC Superset)	91.9%	98.1%
Cook *et al.* [12]	Earliest neutral scan as gallery remainder as probe	94.6%	98.1%
Cook *et al.* [12]	Earliest scan (neutral or non-neutral) as gallery remainder as probe	92.9%	97.2%

TABLE IX
EXPERIMENT 3: ROC III RESULTS

Paper	Experiment	Stated Performance Verification Rate at 0.1% FAR	Our Performance Verification Rate at 0.1% FAR
Husken *et al.* [10]	First semester scans as galleries second semester scans as probes	86.9%	94.8%
Kakadiaris *et al.* [28]	First semester scans as galleries second semester scans as probes	97.0%	94.8%

To demonstrate how the number and quality of regions affects the different fusion rules, the rank-one recognition rates of the individual regions (seen in Table III) are initially sorted from best to worst. Each fusion rule is then performed until all of the sorted regions have been included. Table IV shows each fusion method, its best rank-one recognition rate achieved, associated committee size, and overall recognition rate when all 38 regions are employed. The sum, product, and min rules each show high recognition with a limited number of regions; however, as the quality (based on the individual rank-one recognition rate) of additional regions decreases, so does the recognition rate. The fusion rules are effective only if the individual regions are prescreened for quality and contribution to the final match. Unlike the modified BC method, the traditional BC method can be strongly affected by regions that perform poorly. For example, if five regions each report a rank-one match for a given subject, and a sixth region is occluded and reports a 200th rank, the total score is 205 which is likely to result in a mismatch. The modified BC method avoids this problem by rewarding only the top α matches (maximum performance was achieved for our experiments when $\alpha = 4$) with scores.

While the peak performance of the CV method is similar to the modified BC method, a statistically significant difference in recognition rates still exists. The difference between 97.1% and 96.2% is statistically significant at the 0.05 level of significance. Unlike the modified BC method, the CV method does not reward a region if it is not the best match in the set. This distinction is likely the result of the discrepancy in best reported rank-one recognition rates. Based on the results shown in Table IV, we are able to conclude that maximum performance for our REFER algorithm is achieved when using the modified BC fusion method.

Table V shows the tradeoff between confidence (number of votes for a particular match) and the number of false accepts. These results show the number of false accepts divided by the total number of matches performed in this experiment. The total number of votes for an individual probe-to-gallery match will range from 0 to n, where n is equal to the number of regions that are employed in the fusion technique. The votes are calculated using the CV method, where a vote is awarded if a region match score is below the predetermined region match threshold t_i.

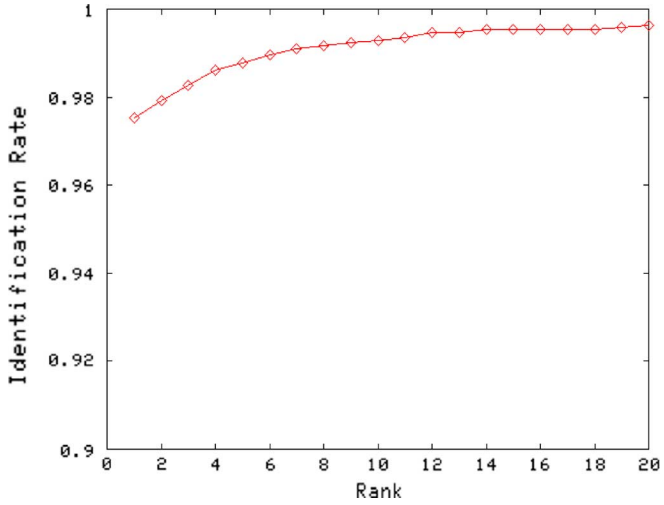


Fig. 8. Experiment 2 CMC curve for the first versus all gallery and probe configuration.

TABLE X
EXPERIMENT 4: EXPRESSION VARIATION RESULTS

Gallery Set	Probe Set	Rank-One Recognition Rate
Neutral (369 images)	Neutral (1889 images)	99.2%
Neutral (369 images)	Non-Neutral (1206 images)	96.3%
Neutral (369 images)	All (3095 images)	98.4%
Non-Neutral (369 images)	Neutral (1889 images)	95.9%
Non-Neutral (369 images)	Non-Neutral (1206 images)	95.0%
Non-Neutral (369 images)	All (3095 images)	95.6%

Table VI shows how the number (and quality) of regions included in the fusion affects the rank-one performance when using the modified BC method. The bold value in Table VI signifies that there is no statistically significant difference between the performance of that experiment and the peak rank one recognition rate for the table. The results show that although the peak recognition rate is 97.5% when employing

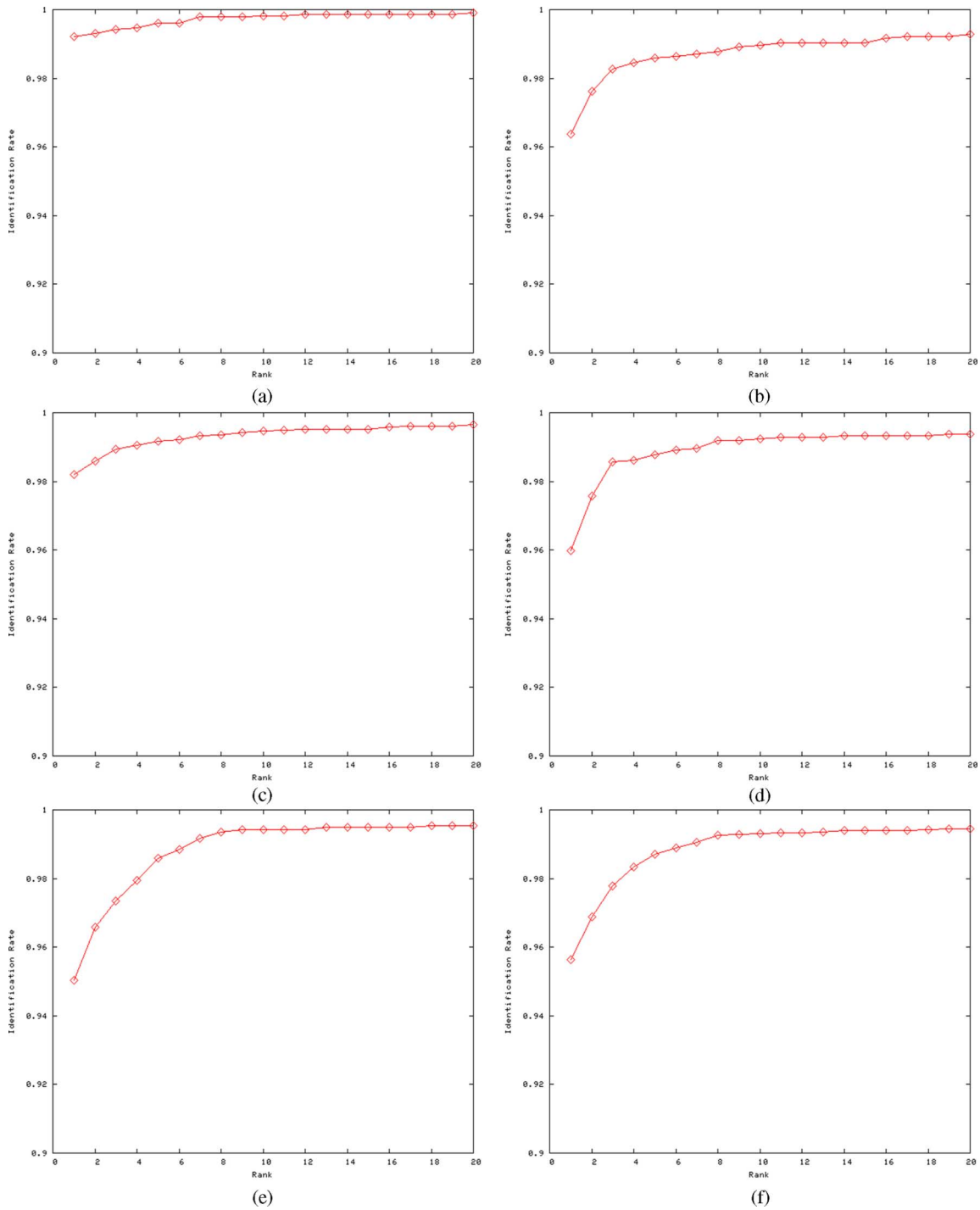


Fig. 9. Experiment 4 CMC curves. (a) Neutral versus neutral. (b) Neutral versus nonneutral. (c) Neutral versus all. (d) Nonneutral versus neutral. (e) Nonneutral versus nonneutral. (f) Nonneutral versus all.

29 regions, there is no statistically significant difference in the score achieved by using 28 regions (97.2%). There is, however, a statistically significant difference between 29 regions (97.5%) and the score achieved when using 27 regions (96.8%) and fewer. These results allow us to conclude that maximum performance can be achieved in future experiments by creating a 28-region ensemble.

A. *Experiment 1: Verification*

Our first experiment is the "All versus All" verification experiment previously reported on by others [9], [11], [12]. Our REFER algorithm using 28 regions and the consensus voting fusion approach of matching multiple 3-D regions of the face was able to achieve a verification rate of 93.2% at an FAR of 0.1%, as shown in Table VII.

B. *Experiment 2: Identification*

Our second experiment is an identification experiment also performed by other authors. Mian *et al.* [11] and Cook *et al.* [12] use a gallery comprised of the first image of each subject (466) and set the remaining images as probes (3581). Chang *et al.* [8] report results using a gallery comprised of the first neutral image

of each subject and set the remaining as probes. In this paper, we reproduce both experiments using the BC fusion method and the results can be seen in Table VIII and the corresponding CMC curve can be seen in Fig. 8.

C. Experiment 3: FRGC ROC III

Our third experiment described in the FRGC program [1] as "Experiment 3" was also reported on previously by others [10], [28]. In this verification experiment, the gallery images come from one semester and the probe entries come from the following semester. This ensures that the time sequence between gallery and probe is maintained, and the average time lapse is increased over the "All versus All" experiment. An increased time lapse between gallery and probe is generally considered to make the experiment more difficult. Since this experiment was designed to determine the verification rate at 0.1%, we employed the CV fusion method technique to produce results. Our algorithm was able to achieve a verification rate of 94.8% at an FAR of 0.1% as shown in Table IX.

D. Experiment 4: Expression Variation

Experiment 4 examines the performance of our algorithm in the presence of nonneutral facial expressions. For this test, we create two gallery sets—one containing the first neutral image of each subject and one containing the first nonneutral image of each subject. Only subjects with at least one neutral and one nonneutral expression are considered for this experiment. Three probe sets are formed in a similar manner: one containing only the remaining neutral images, one containing only the remaining nonneutral images, and one containing all remaining images. This experiment uses the expression classification data provided with the FRGC v2 [1] data set to determine the probe and gallery image separation.

For this experiment, we run a standard identification test (employing the BC fusion method) on each of six possible scenarios. The results from this experiment are seen in Table X and their associated CMC curves can be seen in Fig. 9. The results show only a minimal difference in the recognition rate between the gallery subsets. Our algorithm is able to perform well when matching a neutral expression probe to the neutral expression gallery. In neutral-to-neutral matching, there is a minimal change in the 3-D shape of the face. Our algorithm performed less well when matching the gallery and probe across different facial expressions. This is primarily due to the fact that the selected region of the face is not perfectly rigid across facial expressions.

E. Experiment 5: Missing Data

Finally, Experiment 5 explores how the performance of our algorithm is affected when limited to only using certain regions on the face. This experiment uses a gallery containing the first image of each subject (regardless of expression) and a probe set containing the remaining images. We simulate missing data on the face by manually excluding regions of interest. Table XI lists the regions of the face segmented into different categories based on the location of the probe centroid and their individual rank-one recognition rates achieved for each subset. This experiment employs the BC method for score-based fusion. The best locality performance (95.2%) is found using only the regions located in the center of the face. This recognition rate is similar to the overall performance that resulted when using the best 28 regions. The results also show that both the "left only" and "right only" sets are able to perform well (88.0% and 89.2%, respectively) even though they only contain regions on the side of the face. In the future, we plan to further investigate the robustness of the REFER algorithm on nonfrontal 3-D pose data.

TABLE XI
EXPERIMENT 5: SIMULATED MISSING DATA RESULTS

Subset Label	Included Regions	Rank-One Recognition Rate
Best 28 Regions	21, 23, 22, 8, 26, 24, 28, 27, 10, 15, 18, 25, 17, 34, 19, 2, 35, 3, 31, 33, 32, 4, 1, 6, 16, 5, 14	**97.2%**
Left Only	30, 16, 31, 34, 17, 37	88.0%
Left + Partial Nose	30, 16, 31, 34, 17, 37, 27, 15, 11, 14	92.5%
Right Only	29, 20, 32, 33, 19, 36	89.2%
Right + Partial Nose	29, 20, 32, 33, 19, 36, 28, 18, 12, 13	92.9%
Center Only	38, 6, 7, 4, 5, 1, 2, 3, 8, 9, 10, 21, 22, 23, 24, 25, 26, 35	95.2%
Non-Center	30, 16, 31, 34, 17, 37, 27, 15, 11, 14, 29, 20, 32, 33, 19, 36, 28, 18, 12, 13	94.7%
Eye	31, 34, 27, 21, 22, 23, 24, 25, 26, 28	94.5%
Below Eye	30, 16, 15, 11, 14, 37, 1, 2, 3, 4, 5, 6, 7, 38, 18, 12, 13, 20, 29, 36	92.7%

F. Comparison to Results Published on the FRGC v2 Data Set

Although Phillips *et al.* [1] defined a set of standard experiments for FRGC v2, researchers are free to define their own experiments and use the FRGC data set. Using the recognition and image refinement methods previously discussed, we replicated the experiments from recent publications. Results can be seen in Tables VII–IX. Using the standard z-test previously discussed [26], we are able to confirm that the results achieved by our algorithm statistically significantly outperform each method presented except for that of Kakadiaris *et al.* [28].

Kakadiaris *et al.* [28] present an algorithm for 3-D face recognition that uses an annotated face model (AFM) to create a unique representation of a human face. This algorithm consists of five steps. First, a preprocessing step constructs a 3-D mesh of the geometry of the face from the data. Second, an AFM is fitted to the mesh. Third, a 2-D parameterization is used on the AFM to generate a three-channel deformation image encoding of the shape information. Fourth, the deformation data are processed with two different wavelet transforms (Pyramid and Haar) to extract a signature of the participant. Finally, the signature is matched to other signatures by using an L1 distance metric (for the Haar wavelet) and a complex wavelet structural similarity index algorithm [29] (for the Pyramid wavelet). They reported a 97.0% verification rate at an FAR of 0.1% on experiment 3 from the FRGC v2 program [1]. These results were achieved by fusing the individual scores from the Pyramid and Haar wavelets. The results of our algorithm are similar to those reported by Kakadiaris *et al.* [28]. Relative to Kakadiaris' algorithm, we use a simpler approach that does not require an annotated face model. While the results reported by the author outperform those presented in this paper by a small margin, our algorithm shows the additional potential of being capable to deal with large holes and missing data in images, which is typical in realistic biometric applications.

Maurer *et al.* [9] created an algorithm that uses fusion of 2-D and 3-D face data for multimodal face recognition. Their algorithm first cleans each mesh, extracts relevant face data, and then performs ICP on the 3-D set to generate a distance map between the two aligned meshes, which allows a score to be generated from the results. The 2-D component of their algorithm uses the recognition system created by Neven Vision [30] and fuses the results with those of the 3-D matcher based on the quality of each match. If the 3-D match was very good, then the match is considered correct and the 2-D score is not used. If this is not the case, then the results are fused together to return a combined score. They report results on their 3-D algorithm, as well as reporting the 2-D component's contribution to the 3-D performance. They achieved an 87.0% verification rate at an FAR of 0.1% using 3-D face information, based on the complete 4007 $\times$ 4007 matrix of matching scores compiled from the images in the FRGC v2 data set.

Husken *et al.* [10] created an algorithm that operates primarily on 2-D face data. Their approach uses a version of hierarchal graph matching (HGM) created by Wiskott *et al.* [31]. It is based on creating an elastic graph that holds texture information and position of facial landmarks on an image. Distances are then taken between graphs to determine the similarity between models. When using 3-D shape data alone, they reported an 86.9% verification rate at an FAR of 0.1% on experiment 3 that is defined in the FRGC v2 program [1] to match images across academic semesters.

Mian *et al.* [11] automatically detect the nose, perform pose correction and normalization in both 2-D and 3-D, create a rejection classifier to reduce the overall processing time, and finally segment the 3-D images into two regions (nose and eye/forehead) and match them independently to increase the overall recognition rate. They report verification of 98.5% at 0.1% FAR and rank-one identification of 96.2%, based on a neutral gallery and a probe set comprising the remaining images using their R3D algorithm. In addition, the authors report that the eye/forehead and nose regions of the face contain maximum discriminating features necessary for expression-invariant face recognition.

Cook *et al.* [12] present a novel method based on Log-Gabor Templates for handling expression variation in 3-D face recognition. The authors apply 18 Log-Gabor filters on 49 square windows to generate 147 feature vectors comprising 100 dimensions. After matching is performed, they report results using the FRGC v2 data set. When the 4007 $\times$ 4007 similarity matrix is calculated, they report a 92.31% verification rate at 0.1% FAR. In the identification scenario, the authors employ the first image of a subject in the gallery set (466 images) and the remainder in the probe set (3581) for a rank-one recognition rate of 92.93%. They also discuss how the best performance is achieved when using windows surrounding the upper nose area while the inclusion of outer areas adversely affects the accuracy of the system.

Bowyer *et al.* [27] perform four experiments on a superset of the FRGC v2 experiment 3 data containing 4485 total scans, 2798 neutral image sets (449 subjects), and 1590 nonneutral expression sets (355 subjects). The authors' first experiment examines the change in rank-one recognition rate when observing a time lapse between gallery and probe images. The second looks at the change in the rank-one rate when the images are separated by expression. The third experiment shows how their adaptive rigid multiregion selection (ARMS) algorithm performs when presented with an increasing set of data to demonstrate scalability. Chang *et al.* also report results [8] on two neutral gallery experiments on the same data superset. When the probe set contains only neutral expressions, a rank-one recognition rate of 97.1% is achieved. When the probe set is restricted to only images containing nonneutral expressions, the best reported rank-one recognition rate drops to 87.1%. Table X shows that the REFER algorithm significantly outperforms the results reported by Chang *et al.* and shows a 99.2% rank-one recognition rate for a neutral gallery matched to a neutral probe set and a 96.3% rank-one recognition rate for a neutral gallery matched to a nonneutral probe set for the same experiments.

While the concept of independently matching multiple regions mentioned by Chang *et al.* is similar to the technique mentioned in this paper, our work has many differences that extend the approach. Chang *et al.* limit their experiments to three regions overlapping the nose area. In this paper, we explore the efficacy of including additional regions (38 total) that overlap the entire face rather than simply the center. Chang *et al.* found that maximum performance occurred when only two of the three regions were combined. We believe this is a result of region saturation. When there is too much overlap on a selected region (the nose in this case), the descriptiveness will begin to drop. When we add a region from an alternate location (the forehead for example), the addition of independent results boosted the overall performance. Finally, the authors found that using the product rule for fusion provided the highest results. Our results suggest that this method is inferior to the fusion techniques used in this paper (BC and CV) especially when a large number of regions are used.

VI. Conclusion and Future Work

We have presented the results of an approach to 3-D face recognition designed to handle variations in facial expression between gallery and probe images. Our approach automatically finds the nose tip and selects 28 different regions around the face. The algorithm has been evaluated on the FRGC v2 data set containing 4007 3-D face scans from 466 unique subjects, representing a variety of facial expressions. For the identification experiment, we achieve a rank-one recognition rate of 97.2%. The results of our algorithm using only the single 3-D shape modality outperform those reported by many other published organizations [8]–[12] on the same data set.

Incomplete facial data and artifacts are still a major issue in realistic biometric experiments. We have performed an extensive study on how individual regions across the face affect the recognizability of a subject, and how manually removing certain regions simulates large amounts of missing data or facial pose variation. In the presence of both obstacles, our REFER algorithm still provides a high level of recognition due to its piecewise robustness. In addition, we found that variants of the BC and CV techniques provide maximum performance when fusing results from multiple 3-D face regions.

REFERENCES

[1] 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," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 2005, pp. 947–954.

[2] T. Faltemier, K. Bowyer, and P. Flynn, "3D face recognition with region committee voting," in *Proc. 3rd Int. Symp. 3D Data Processing, Visualization Transmission*, 2006, pp. 318–325.

[3] B. Gokberk and L. Akarun, "Comparative analysis of decision-level fusion algorithms for 3D face recognition," in *Proc. 18th Int. Conf. Pattern Recognition*, 2006, pp. 20–24.

[4] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognit.*, vol. 38, pp. 2270–2285, 2005.

[5] P. Besl and N. McKay, "A method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, Feb. 1992.

[6] J. Phillips and P. Grother, Facial Recognition Vendor Test 2002 NISTIR 6965, 2002. [Online]. Available: www.frvt2002.org/FRVT2002/documents.htm.

[7] Face recognition vendor test. [Online]. Available: http://www.frvt.org. 2006.

[8] K. Chang, K. W. Bowyer, and P. Flynn, "Multiple nose region matching for 3D face recognition under varying facial expression," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1–6, Oct. 2006.

[9] T. Maurer, D. Guigonis, I. Maslov, B. Pesenti, A. Tsaregorodtsev, D. West, and G. Medioni, "Performance of geometrix ActiveID 3D face recognition engine on the FRGC data," in *Proc. Face Recognition Grand Challenge Workshop*, 2005, III: 154-154.

[10] M. Husken, M. Brauckmann, S. Gehlen, and C. von der Malsburg, "Strategies and benefits of fusion of 2D and 3D face recognition," in *Proc. IEEE Workshop Face Recognition Grand Challenge Experiments*, 2005, p. 174.

[11] A. S. Mian, M. Bennamoun, and R. Owens, "An efficient multimodal 2D-3D hybrid approach to automatic face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 11, pp. 1584–1601, Nov. 2007.

[12] J. Cook, V. Chandran, and C. Fookes, "3D face recognition using log-gabor templates," in *Proc. British Machine Vision Conf.*, 2006, pp. 83-83.

[13] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, 2003.

[14] K. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition," *Comput. Vis. Image Understanding*, vol. 101, no. 1, pp. 1–15, 2006.

[15] X. Lu, A. K. Jain, and D. Colbry, "Matching 2.5D face scans to 3D models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 31–43, Jan. 2006.

[16] X. Lu and A. Jain, "Deformation modeling for robust 3D face matching," in *Proc. IEEE Computer Society Conf. Computer Vision Pattern Recognition*, 2006, pp. 1377–1383.

[17] A. Martinez, "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample perl class," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 6, pp. 748–763, Jun. 2002.

[18] B. Heisele and T. Koshizen, "Components for face recognition," in *Proc. IEEE Int. Conf. Face and Gesture Recognition*, 2004, pp. 153–158.

[19] C. Beumier and M. Acheroy, "Face verification from 3D and grey level cues," *Pattern Recognit. Lett.*, vol. 22, no. 12, pp. 1321–1329, 2001.

[20] Konica-Minolta USA, Minolta Vivid 910 Non-Contact 3D Laser Scanner. [Online]. Available: http://www.minoltausa.com/vivid/.

[21] Geometrix Inc., Image segmentation categories. [Online]. Available: http://www.beebiometrics.org/_frgc05/. 2005.

[22] A. Jain, X. Lu, and D. Colbry, "Three-dimensional model based face recognition," in *Proc. 17th Int. Conf. Pattern Recognit.*, 2004, vol. I, pp. 362–366.

[23] O. Melnik, Y. Vardi, and C. Zhang, "Mixed group ranks: Preference and confidence in classifier combination," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 8, pp. 973–981, Aug. 2004.

[24] S. Jeong, K. Kim, B. Chun, J. Lee, and Y. Bae, "An effective method for combining multiple features of image retrieval," in *Proc. IEEE Region 10 Conf.*, 1999, pp. 982–985.

[25] T. Ho, J. Hull, and S. Srihari, "Decision combination in multiple classifier systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 1, pp. 66–75, Jan. 1994.

[26] P. Yan and K. Bowyer, "Empirical evaluation of advanced ear biometrics," in *Proc. IEEE Comput. Soc. Conf. Computer Vision Pattern Recognition Workshop*, 2005, pp. 41–48.

[27] K. W. Bowyer, K. Chang, and P. Flynn, "Adaptive rigid multi-region selection for handling expression variation in 3D face recognition," in *Proc. IEEE Workshop Face Recognition Grand Challenge Experiments*, 2005, pp. 157-157.

[28] I. Kakadiaris, G. Passalis, G. Toderici, N. Murtuza, and T. Theoharis, "3D face recognition," in *Proc. British Machine Vision Conf.*, 2006, pp. 200–208.

[29] Z. Wang and E. Simoncelli, "Translation insensitive image similarity in complex wavelet domain," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, 2006, pp. 573–576.

[30] Neven Vision. [Online]. Available: http://www.nevenvision.com/. 2004.

[31] L. Wiskot, J. Fellous, N. Kruger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 775–779, Jul. 1997.

Timothy C. Faltemier received the B.S. degree in computer science from the University of Illinois at Champaign-Urbana in 2003, the M.S. degree in computer science from Texas A&M University, College Station, and the Ph.D. degree in computer Science and engineering from the University of Notre Dame, Notre Dame, IN.

His research interests include 2-D face recognition, 3-D face recognition, and data mining. He is currently with Progeny Systems Corporation, Manassas, VA.

Kevin W. Bowyer is the Schubmehl-Prein Professor and Chair of the Department of Computer Science and Engineering at the University of Notre Dame, Notre Dame, IN. His recent efforts have concentrated on the areas of biometrics and data mining. His biometrics research has been supported by a number of agencies, and his research group has been active in support of the government's Face Recognition Grand Challenge program and Iris Challenge Evaluation program. His data mining research, aimed at ensemble methods for "extreme" problems, has been supported by Sandia National Labs.

Prof. Bowyer's paper "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition," published in *Computer Vision and Image Understanding*, was number one on the CVIU most-downloaded list for two quarters and in the top eight for five straight quarters. His paper "Face Recognition Technology: Security Versus Privacy" published in *IEEE Technology and Society* was recognized with a 2005 "Award of Excellence" from the Society for Technical Communication.

Patrick J. Flynn (SM'96) received the Ph.D. degree in computer science from Michigan State University, East Lansing, in 1990

Currently, he is Professor of Computer Science and Engineering and Concurrent Professor of Electrical Engineering at the University of Notre Dame, Notre Dame, IN, and held faculty positions at Washington State University from 1991 to 1998, The Ohio State University from 1998 to 2001, and has been on the University of Notre Dame faculty since 2001. His research interests include computer vision, biometrics, computer graphics and visualization, and image processing.

Dr. Flynn is a Fellow of IAPR, a past Associate Editor and Associate Editor-in-Chief of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, and a past Associate Editor of *Pattern Recognition Letters*. He has received outstanding teaching awards from Washington State University and the University of Notre Dame.