An Evaluation of Multimodal 2D+3D Face Biometrics

Kyong I. Chang, Kevin W. Bowyer, and Patrick J. Flynn

Abstract—We report on the largest experimental study to date in multimodal 2D+3D face recognition, involving 198 persons in the gallery and either 198 or 670 time-lapse probe images. PCA-based methods are used separately for each modality and match scores in the separate face spaces are combined for multimodal recognition. Major conclusions are: 1) 2D and 3D have similar recognition performance when considered individually, 2) combining 2D and 3D results using a simple weighting scheme outperforms either 2D or 3D alone, 3) combining results from two or more 2D images using a similar weighting scheme also outperforms a single 2D image, and 4) combined 2D+3D outperforms the multiimage 2D result. This is the first (so far, only) work to present such an experimental control to substantiate multimodal performance improvement.

Index Terms—Biometrics, face recognition, three-dimensional face, multimodal, multisample.

1 Introduction

Studies of multimodal biometrics commonly report that performance improves over that of a single modality. However, such results typically mix effects of: 1) improvement due to multiple image samples and 2) improvement due to multiple sensing modes. However, it is not sufficient for the 2D+3D face recognition to show improvement over single-sample 2D face recognition. Instead, since 2D+3D recognition uses two image samples to represent a person, it should show improvement over using two 2D samples to be considered superior. We evaluate a 2D+3D recognition scheme using this standard to determine how much of the "multimodal improvement" comes from combining results from different sensing modes versus simply from multiple images.

Section 2 briefly summarizes some related work in multimodal face recognition. Section 3 outlines the experimental methods and materials used. Section 4 presents the experimental results obtained by different multibiometrics. Last, Section 5 summarizes and discusses the results.

2 Related Studies in Multimodal Face Recognition

Various multimodal 2D+3D face recognition schemes are proposed in [1], [2], [3], [4] and [5], [6] (for additional survey detail on 2D+3D and 3D face recognition, see [7]). In all of these 2D+3D studies, using different data sets and algorithmic approaches, the multimodal approaches are shown to outperform either mode alone. However, the comparison is always made between a multimodal result and a result obtained from one sample from an individual mode rather than multiple samples from that mode.

An interesting approach in [8] can be classified as "hybrid multiple biometrics." Five samples of face and voice were collected for each person. Two interesting points can be observed from the results. As the number of probe samples increases, the accuracy improves faster for face than for speech, and improvement is more rapid for multimodal samples than for multiple samples of the same mode.

• *The authors are with the Computer Science & Engineering Department, University of Notre Dame, Notre Dame, IN 46556. E-mail: {kchang, kwb, flynn}@cse.nd.edu.*

Manuscript received 28 Apr. 2004; revised 5 Oct. 2004; accepted 8 Oct. 2005; published online 10 Feb. 2005.
Recommended for acceptance by M. Pietikainen.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0201-0404.

3 Image Set, Recognition Algorithm, and Multimodal Fusion

Two four-week sessions were conducted for data collection, with approximately a six weeks time lapse between the two. The "gallery" (enrollment) representation of a person is selected from the earliest session in which valid images were acquired. A "probe" representation of a person is taken from a later session to be matched against the gallery for recognition. In our *single probe study*, there are at least six and as many as a 14 weeks time lapse between the gallery and the probe images. In our *multiple probe study*, there are as many as seven probe images corresponding to a given gallery image and there is at least a one week time lapse between gallery and probe.

For image acquisition, persons stood approximately 1.5 meters from the sensors, against a neutral gray background. The 3D images were acquired using a Minolta Vivid 900,[1] with either its "Medium" or "Tele" lens and with the scanner height adjusted to that of the person's face if needed. For the 3D images, one central spotlight was used to light the face (LT) and subjects were asked to have a normal facial expression ("FA" in FERET terminology [9]) and to look directly at the camera. Because 3D image acquisition takes more time, just one 3D image was acquired for each person at each acquisition session. The 2D images were acquired with a Canon PowerShot G2[1] digital camera. Each subject was asked to have one normal expression (FA) and one smile expression (FB), once with three spotlights (LM) and a second time with two side spotlights (LF). A 640×480 range image is produced by the 3D scanner and $1,704 \times 2,272$ color images are produced by the 2D camera. Thus, at each image acquisition session, each person had one 3D image acquired (image condition FALT) and four different 2D images acquired (conditions FALM, FBLM, FALF, and FBLF). See Fig. 1 for an example.

A total of 275 different persons participated in one or more sessions. Of these, 198 had two or more sessions of usable data. Thus, the single probe study has 198 individuals in the probe set, the same 198 individuals in the gallery and 77 individuals in the training set. For the multiple probe study, 472 probes are added to the single probe data set, yielding 670 in total.

Normalization steps for geometry and brightness are applied to the 2D images. The 2D images are treated as having pose variation only around the Z axis, the optical axis. Two control points (**1** and **2**) at the centers of the eyes are selected manually for geometric normalization to correct for rotation, scale, and position of the face, as shown in Fig. 2a. Finally, median filtering is applied with a 7×7 kernel. The face region is interpolated into a 130×150 template that masks out the background. This scales the original image so that the pixel distance between the eye centers is 80. Histogram equalization is applied to standardize the intensity distribution. This attempts to minimize the variation caused by illumination changes between images.

Each point defined in 3D space for a range image has a depth value along the Z-axis. Only the geometric normalization is needed to correct the pose variation. Four control points are manually selected to accomplish the task, as shown in Fig. 2b. We standardize the pose in a 3D face image as follows: A transformation matrix is first computed based on the surface normal angle difference in X (roll) and Y (pitch) between manually selected landmark points (**1**, **2**, and **3** in Fig. 2b) and predefined reference points of a standard face pose and location. The outer eye corners

1. Specific manufacturer and model are mentioned only to specify the work in detail and do not imply an endorsement of the equipment or vendors.

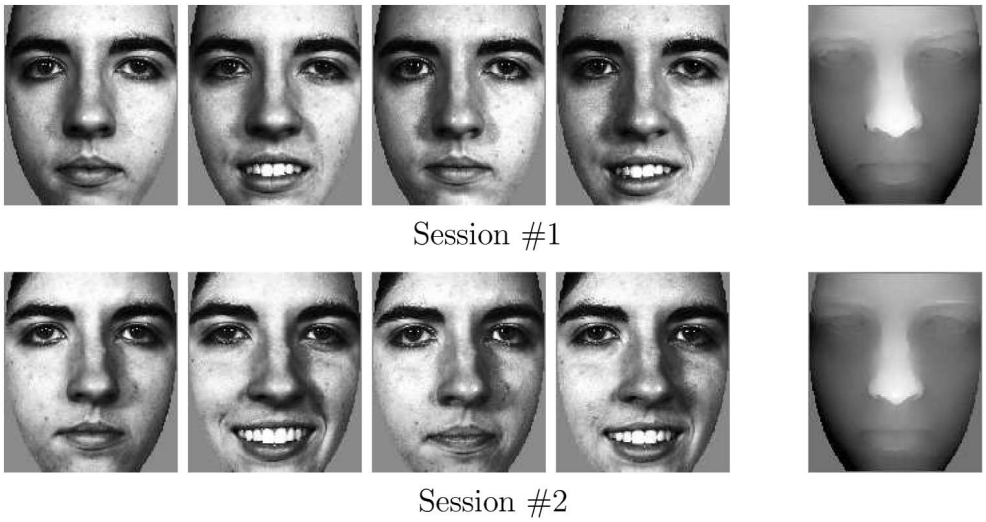

Fig. 1. Two different sessions acquired in six weeks apart. Four 2D images (left to right: FALM, FBLM, FALF, and FBLF) and one 3D image (rightmost: FALT) of a person are acquired in each session.

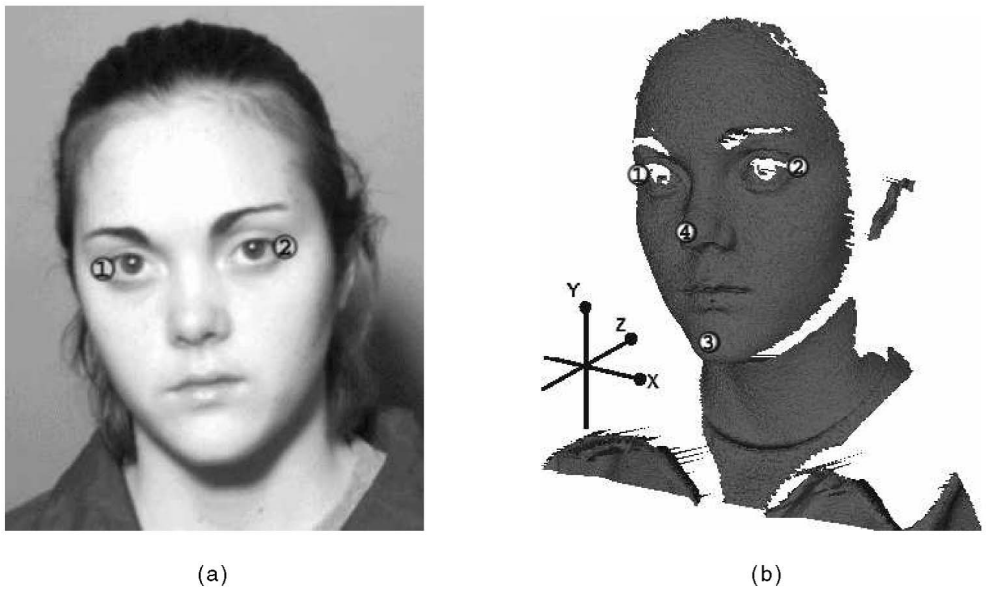

Fig. 2. Landmark (control) points specified in (a) a 2D image and (b) a 3D image.

rather than eye centers are used as landmark points because the eyeball is an artifact-prone region for the range sensor, whereas the eye corners marked on the skin are more reliable. The landmark points for the eye corners and the center of chin are used to place the raw 3D image in a standard pose. Pose variation around the Z axis (yaw) is corrected by measuring the angle difference between the line across the two eye points and a horizontal line. At the end of the pose normalization, the nose tip (point **4** in Fig. 2b) of every subject is translated to the same point in 3D relative to the sensor. The geometric normalization in 2D gives the same pixel distance between eye locations to all faces. This is necessary because the absolute scale of the face is unknown in 2D. However, this is not the case with a 3D face image and, so, the eye locations may naturally be at different pixel locations in depth images of different faces. The Minolta sensor produces registered 2D and 3D images. Thus, in principle, it is possible to create a fully pose corrected 2D image by projecting the color texture from the pose corrected 3D. However, there are missing data points in the 3D image. In an initial study, we found that missing data problems with fully pose-corrected 2D outweighed the gains from the additional pose correction [5] and, so, we use the typical Z-rotation corrected 2D.

Problems with the 3D data are alleviated to some degree by preprocessing to fill in holes (a region where there is missing 3D data during sensing) and remove "spikes." The 640×480 raw 3D image is converted to a 130×150 range image by the following process. The outer eye corners, nose tip, and the center of chin are marked as landmark points on the 640×480 raw image, as shown in Fig. 2b. Then, a 21×21 region around the marked nose tip is searched to refine the nose tip location, if needed. The refined nose tip gives the centerline for cropping a 130×150 region from the raw 3D image to create a range image from the depth values. The next step attempts to remove spike artifacts that can occur in the 3D image. The variance in the Z value of the 3D is computed for an 11×11 window around each pixel. If the variance is larger than a threshold value, then the current pixel is considered to be part of a spike artifact and is eliminated, leaving holes in the data. Last, these holes and any originally occurring holes are removed by

TABLE 1
SMSS Rank-One Recognition Rates

Results with one time-lapse probe per subject					
	FALM	*FALF*	*FBLM*	*FBLF*	*FALT(3D)*
FALM	**90.9(86.7)**%	81.3(75.8)%	63.6(41.9)%	62.6(34.9)%	N/A
FALF	83.3(81.3)%	82.3(83.3)%	60.6(40.4)%	66.7(38.4)%	N/A
FBLM	68.7(44.4)%	62.6(43.9)%	86.4(82.8)%	83.8(73.7)%	N/A
FBLF	62.6(40.9)%	66.7(47.5)%	86.4(75.8)%	83.8(79.3)%	N/A
FALT (3D)	N/A	N/A	N/A	N/A	**88.9**%

Results with one or more time-lapse probes per subject					
	FALM	*FALF*	*FBLM*	*FBLF*	*FALT(3D)*
FALM	**92.4(88.4)**%	87.9(81.8)%	68.7(48.5)%	64.1(38.9)%	N/A
FALF	91.4(85.4)%	86.9(86.9)%	69.2(47.5)%	71.2(45.5)%	N/A
FBLM	72.7(49.0)%	68.7(47.5)%	92.4(87.4)%	91.9(85.9)%	N/A
FBLF	71.2(47.5)%	71.2(51.0)%	93.4(83.3)%	92.9(87.4)%	N/A
FALT (3D)	N/A	N/A	N/A	N/A	**87.4**%

linear interpolation of missing values from good values around the edges of the hole. The process of creating the 130×150 range image is fully automated after eye, nose, and chin points are marked. In this experiment, the Mahalanobis cosine distance metric was used during the matching process [10]. It is our experience that the Mahalanobis cosine distance metric consistently outperforms other metrics, such as the L_1 and L_2 norms, for both 2D and 3D face recognition. The "face space" is created from a training set of 2D and 3D images for 77 subjects. Initially, eigenvectors are dropped starting with the one corresponding to the largest eigenvalue, then the next largest, and so on, and the rank-one recognition rate for each modality computed each time, continuing until the rank-one recognition rate drops. The number of the "top" eigenvectors dropped is denoted as M. Then, a similar process is followed to drop eigenvectors starting with the one corresponding to the smallest eigenvalue, then the second smallest, and so on. The number of the eigenvectors corresponding to smallest eigenvalues dropped is denoted N. This tuning step is done separately for the 2D face space and the 3D face space. When images from all four acquisition conditions for 2D are used for training, there are 308 training images. Tuning the face space resulting from these 308 training images gives $M = 26$ and $N = 62$ for the 2D image face space, and $M = 3$ and $N = 6$ for the 3D image face space. Previous researchers have reported dropping fewer than 26 of the largest eigenvectors in tuning the face space. However, note that the set of training images used here explicitly incorporates variation in lighting condition and facial expression. This naturally leads to an increased number of eigenvectors being dropped among the largest eigenvalues that represent image variation, yet irrelevant to subject identity.

The multimodal decision is made by combining the match scores for each person across the different biometrics and ranking the subjects based on the combined scores. Scores from each modality are linearly normalized to a range of [0, 100] before combining. We explore confidence-weighted versions of the sum, product, and minimum rules in this work. Among the fusion rules that we tested, the sum rule with linear score transformation considering weighting scheme provides the best performance overall. Both the sum rule and product rule consistently show good performance across different score normalization methods. The minimum rule, however, shows lower performance than the others. For each probe, a "confidence" weight is computed for each modality's decision, as follows:

$$weight = \frac{distance_2 - distance_1}{distance_3 - distance_1},$$

where $distance_i$ is the ith smallest distance from the probe to one of the gallery elements in the given modality's space. If the difference between the first and second distance metric is large compared to the typical distance, then this value will be large.

4 Experiments

Our "baseline" experiment looks at recognition performance from a single modality (SM), either 2D alone or 3D alone. For each modality, a single sample (SS) is used to represent a person, both for enrollment in the gallery and as a probe into the gallery. There are four possible single-image-per-subject gallery sets for the 2D images (FALM, FALF, FBLM, and FBLF), and one for 3D (FALT). The same is true of possible probe image sets acquired in a later session(s). Thus, there are 16 possible recognition results for 2D and one for 3D, summarized in Table 1. Within this experiment, we found the highest 2D face recognition performance in the case of using FALM images in both the gallery and the probe set. We refer to this as "*FALM*:FALM" where the labels are interpreted as a match between "*GALLERY*" and "PROBE."

Taking advantage of our image acquisition conditions for 2D, we explore two options for creating the face space for 2D recognition. The first option is that training for the 2D face space is done with the same image condition as used in the gallery set. For instance, when *FALF* images are used in the gallery, 77 FALF training images can be used to create the face space. The other option is to use all four image conditions in creating the face space, for a total of 77×4 images. The rank-one recognition rates are generally higher in the case of training with the larger number of images, and these are the values listed outside the parentheses in Table 1. The recognition rates for the smaller training set with the uniform image condition are the ones listed in parentheses.

The main result of the baseline experiment is that similar recognition performance can be obtained using either 2D or 3D in a single-modality and single-sample (SMSS) scenario. Not surprisingly, we find that recognition performance with 2D images is generally higher when the gallery and probe images are matched for image condition (lighting and expression) than when they are not. The lighting variation used in our image acquisition does not cause as large a drop in performance as the facial expression variation. However, because there is no concept of a "unit variation" across lighting and expression, this does not support any general conclusion about the relative difficulty of variation in the two conditions.

The generally higher performance obtained with the larger training set may be due in part to the larger number of images and in part to the variation in lighting and expression in the images.

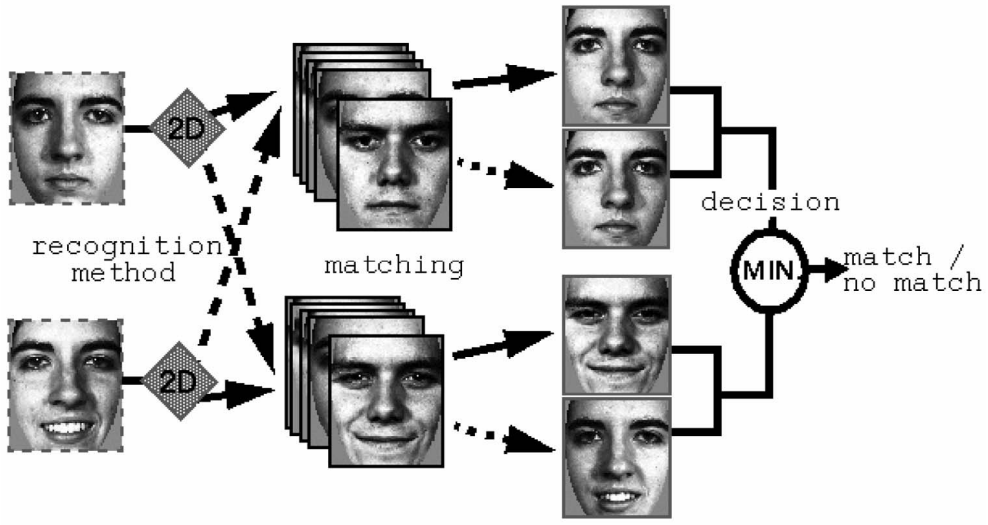

Fig. 3. SMMS decision based on four matches—two subject images for gallery and for probe.

The relative importance of these two effects is confounded in our experiment. However, the results may suggest that face spaces created from training images of diverse conditions show a more robust performance, especially when the gallery and probe images conditions cannot be controlled. In either case, because it results in higher performance, the face space created by using the larger (varied image condition) training set is used for the remaining 2D experiments.

4.1 Single-Modality and Multiple-Sample (SMMS)

Recognition performance can be improved by increasing the number of images used to represent a person [11]. Therefore, this experiment examines the performance that can be gained by a single-modality and multiple-sample (SMMS) scheme. This SMMS approach illustrated in Fig. 3 makes a decision based on four matches when a subject is represented by two images in the gallery and the probe. With four image conditions available, there are $\binom{4}{2} = 6$ ways of choosing two image conditions to use to represent a person. Thus, we have six possible two-image representations of a person to use in a gallery and, similarly, six to use for a probe representation, for a total of $6 \times 6 = 36$ experiments. In these experiments, the result for a given two-image probe is created by matching each of the pair of probe images against each of the pair of gallery images representing each person. This results in $2 \times 2 = 4$ individual match distances for each person in the gallery and the probe is recognized as the gallery person with the minimum sum of four match distances.

The results of these 36 SMMS experiments are summarized in Table 2. For one time-lapse probe (pair of images) per subject, the rank-one recognition rate ranges from a low of 74 percent to a high of 96 percent. This compares to a range of 61 percent to 91 percent for the corresponding set of 16 single-sample experiments in Table 1. It seems clear that this multiple-sample approach results in a general and substantial improvement over the single-sample approach. The improvement can be interpreted as coming from the fact that an individual is represented more robustly by using a pair of varying-condition images than by using a single image.

Performance generally continues to increase as more than two images are used to represent a person. When the analogous set of 16 experiments is performed using three images to represent a person ($\binom{4}{3} = 4$ choices for gallery and for probe), the rank-one recognition rates for the single time-lapse probe version of the experiments ranges between 92 percent and 96 percent. We can also perform a single experiment in which we use all four images to represent a person in the gallery and as a probe. For this experiment, the rank-one recognition rate is 96 percent. Thus, for the size and composition of data set that we use, performance improvement appears to plateau at about the range of using four images to represent a person.

4.2 Multiple-Modality and Single-Sample (MMSS)

In this experiment, a person is represented by the combination of a one 2D image and one 3D image. The two images that represent a given person in the gallery are restricted to come from the same acquisition session, and the two images that represent a person as a probe are restricted to come from a later session. The result for a given two-image probe is created by matching the 2D probe image against each of the 2D gallery images, matching the 3D probe image against each of the 3D gallery images, and taking the sum of two normalized match distances.

The results of these experiments are summarized in Table 3. For one time-lapse probe per subject, the rank-one recognition rate ranges between 90 percent and 97 percent. This compares to a

TABLE 2
SMMS Rank-One Rates in 2-Gallery and 2-Probe Mode

Results with one time-lapse probe per subject						
	FALM–FALF	*FALM–FBLM*	*FALF–FBLM*	*FALM–FBLF*	*FALF–FBLF*	*FBLM–FBLF*
FALM–FALF	**92.9**%	93.4%	86.9)%	93.9%	87.4%	74.2%
FALM–FBLM	93.4%	93.9%	91.4%	93.9%	90.9%	91.9%
FALF–FBLM	88.9%	94.4%	91.9%	91.4%	86.9%	91.4%
FALM–FBLF	92.9%	95.5%	91.9%	95.0%	90.9%	90.4%
FALF–FBLF	86.4%	91.4%	89.4%	89.4%	88.9%	88.9%
FBLM–FBLF	74.2%	91.9%	91.9%	89.4%	88.4%	91.9%

Results with one or more time-lapse probes per subject						
	FALM–FALF	*FALM–FBLM*	*FALF–FBLM*	*FALM–FBLF*	*FALF–FBLF*	*FBLM–FBLF*
FALM–FALF	**94.4**%	94.4%	90.9%	96.0%	90.4%	76.8%
FALM–FBLM	94.4%	95.0%	93.4%	95.0%	93.9%	96.5%
FALF–FBLM	92.9%	96.0%	93.4%	95.5%	91.9%	96.0%
FALM–FBLF	94.4%	95.5%	95.5%	95.5%	94.4%	97.0%
FALF–FBLF	90.9%	96.0%	93.4%	95.5%	93.4%	96.0%
FBLM–FBLF	75.3%	97.0%	95.5%	95.0%	95.0%	97.5%

TABLE 3
MMSS Rank-One Recognition Rates

Results with one time-lapse probe per subject				
	FALM–FALT(3D)	*FALF–FALT(3D)*	*FBLM–FALT(3D)*	*FBLF–FALT(3D)*
FALM–FALT(3D)	**95.0%**	95.0%	90.9%	90.4%
FALF–FALT(3D)	95.0%	95.0%	92.4%	88.9%
FBLM–FALT(3D)	93.9%	92.4%	96.5%	96.0%
FBLF–FALT(3D)	91.4%	91.9%	96.5%	98.5%

Results with one or more time-lapse probe per subject				
	FALM–FALT(3D)	*FALF–FALT(3D)*	*FBLM–FALT(3D)*	*FBLF–FALT(3D)*
FALM–FALT(3D)	**97.5%**	97.5%	95.5%	94.4%
FALF–FALT(3D)	98.0%	97.0%	96.0%	96.0%
FBLM–FALT(3D)	97.0%	96.0%	99.0%	99.0%
FBLF–FALT(3D)	97.0%	97.0%	98.5%	99.5%

range of 74 percent to 96 percent (see Table 2) when using two 2D image samples to represent a person, and 61 percent to 91 percent (see Table 1) when using a single 2D image. For a more detailed comparison of SMMS with two 2D images and MMSS with one 3D image and one 2D image, consider a smaller set of results with matched image conditions. The four MMSS single-probe results that have "FA**" in both gallery and probe in MMSS (Table 3) all achieve 95 percent recognition. This is higher than 23 of the 25 SMMS results listed in Table 2 that have "FA**" in both the gallery and probe, which range from 86.4 percent to 95.5 percent. Thus, it seems clear that multimodal 2D+3D face recognition achieves real improvement over 2D face recognition, even when the comparison is controlled for the number of image samples used to represent a person. However, it also appears possible that multisample 2D face recognition using a larger number of samples could achieve performance essentially equivalent to multimodal single-sample 2D+3D.

4.3 CMC and ROC Curves

The cumulative match characteristic (CMC) curves in Fig. 4a are created from the results of the one-or-more-time-lapse-probe versions of the experiments, in order to sample finer differences in recognition rate. The best rank-one correct identification rate for the baseline SMMS scheme is 94.4 percent, versus 97.5 percent for the MMSS scheme. The result of McNemar's test [12] for significance of the difference in the rank-one match between the integrated biometrics (both MMSS and SMMS schemes) and either the 2D face or the 3D face alone shows that multimodal performance is significantly greater ($\alpha = 0.05$). However, we found no significant difference between 2D alone and 3D alone in SMSS recognition.

To present the results in the context of a verification scenario, the False Acceptance Rate (FAR), False Rejection Rate (FRR), and Equal Error Rate (EER) are summarized in the ROC curve in Fig. 4b. Similar to what the ROC curves show, the multimodal approach (0.019) achieves significantly lower EER than either SMSS approach (0.043 for 2D, 0.045 for 3D) and the multiple-sample 2D approach (0.048) performed close to single-sample 2D, but not as good as multimodal in verification mode.

The EER of 3D SMSS shows very similar accuracy to that of 2D SMSS. However, the 2D rank-one match rate is greater than the 3D rank-one match rate, as shown in the CMC curves. It is important to note that the results presented in EERs should be carefully analyzed because the EERs represent only *one* operating point on the ROC curves for the comparison. The operating points in the function of FAR and FRR will be changed to meet the requirements of an application.

5 Discussion

We have presented results from the largest experimental study to date of 3D and multimodal 2D+3D face recognition, with 198 persons in the gallery. We present results for 1) recognition with one time-lapse probe per person, for 198 probes, and 2) recognition with as many time-lapse probes as are available for each person, for 670 total probes. For each image acquisition

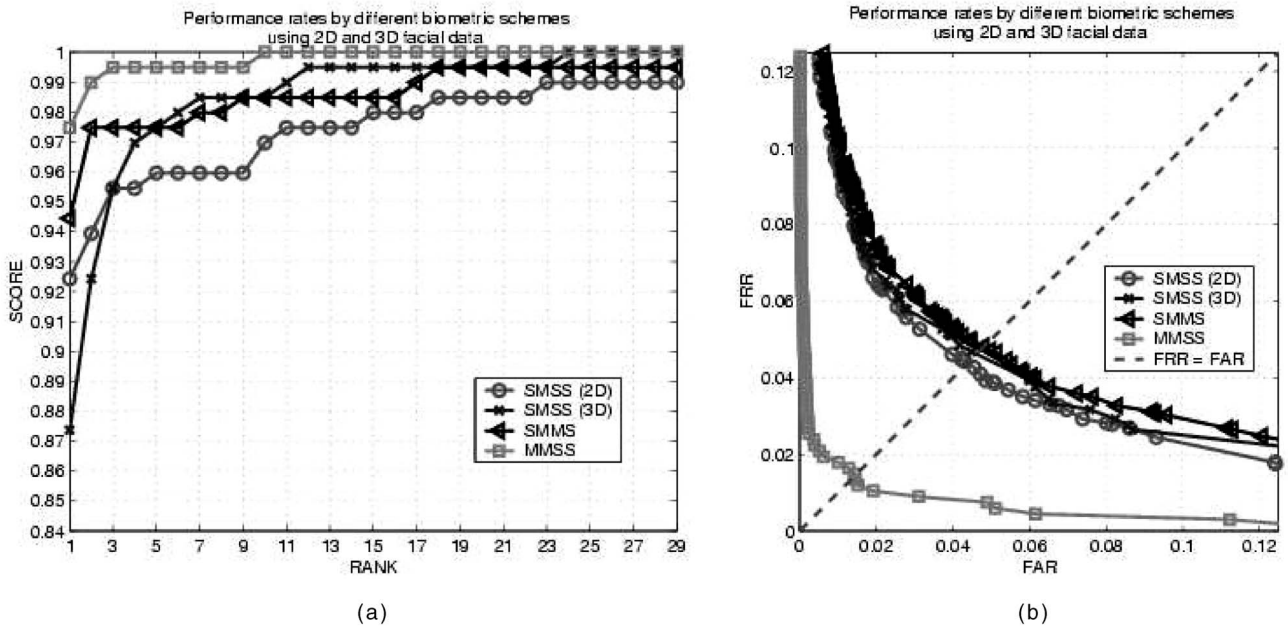

Fig. 4. Baseline CMC and ROC performance of the multiple probe study. (a) Performance rates in CMC. (b) Performance rates in ROC. SMSS result with *FALM*:FALM, SMMS result with *FALM–FALF*:FALM–FALF based on four matches, and MMSS result with *FALM–FALT(3D)*:FALM–FALT(3D) for a multiple probe study are reported.

session, multiple 2D images were acquired under different lighting and facial expression conditions. Therefore, we are able to consider 2D recognition results over a range of experimental conditions.

Our results support four basic conclusions:

1. Similar recognition performance is obtained using a single 2D image or a single 3D image.
2. Multimodal 2D+3D face recognition performs significantly better than using either 3D or 2D alone.
3. Combining results from two or more 2D images using a similar fusion scheme as used in multimodal 2D+3D also improves performance over using a single 2D image.
4. Even when the comparison is controlled for the same number of image samples used to represent a person, multimodal 2D+3D still outperforms multisample 2D, though not by as much; also, it may be possible to use more 2D samples to achieve the same performance as multimodal 2D+3D.

For item 1, about the relative power of 2D and 3D for face recognition, the conclusion should be interpreted cautiously. Our results reported here use the same basic recognition engine for both 2D and 3D. It is possible that some other algorithm that exploits information in 2D images in some ideal way that cannot be applied to 3D images would result in 2D face recognition being more powerful than 3D face recognition, or vice-versa.

Overall, we are led to conclude that improved face recognition performance will result from 1) the combination of 2D+3D imaging and also 2) representing a person by multiple images taken under varied lighting and facial expression. Both of these topics should be the subject of substantial additional future work. The topic of 3D face recognition has been only lightly explored so far [7]. The topic of multiimage representations of a person for face recognition is even less well explored. Also, we should note that the results reported in this paper are obtained using manually marked eye locations. Thus, these are in a sense "best possible" results since an automatic eye-finding procedure is almost certain to introduce errors. Algorithms for automatically locating landmark points on the face is another area in which more research is needed.

Currently, 3D scanners do not operate with the same flexibility of conditions of lighting, depth of field, and timing as normal 2D cameras. Thus, 3D face imaging requires greater cooperation on the part of the subject. Also, some 3D sensing technologies, such as the Minolta, are "invasive" in the sense that they project light of some type onto the subject. Clearly, another important area of future research in 3D face recognition is the development of better 3D sensing technology.

The image data set used in this research is available for noncommercial research use. See http://www.nd.edu/~cvrl for additional information.

ACKNOWLEDGMENTS

This work is supported by US National Science Foundation grant EIA 01-20839 and the Department of Justice grant 2004-DD-BX-1224. The authors would like to thank the associate editor and the anonymous reviewers for their helpful suggestions to improve this paper.

REFERENCES

[1] 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 System," *Proc. Int'l Conf. Pattern Recognition*, vol. 2, pp. 911-916, 2000.

[2] Y. Wang, C. Chua, and Y. Ho, "Facial Feature Detection and Face Recognition from 2D and 3D Images," *Pattern Recognition Letters*, vol. 23, pp. 1191-1202, 2002.

[3] C. Beumier and M. Acheroy, "Automatic Face Verification from 3D and Grey Level Clues," *Proc. 11th Portuguese Conf. Pattern Recognition*, pp. 95-101, 2000.

[4] F. Tsalakanidou, S. Malassiotis, and M. Strintzis, "Integration of 2D and 3D Images for Enhanced Face Authentication," *Proc. Sixth Int'l Conf. Automated Face and Gesture Recognition*, pp. 266-271, May 2004.

[5] K. Chang, K. Bowyer, and P. Flynn, "Face Recognition Using 2D and 3D Facial Data," *Proc. ACM Workshop Multimodal User Authentication*, pp. 25-32, Dec. 2003.

[6] K. Chang, K. Bowyer, and P. Flynn, "Multi-Biometrics Using Facial Appearance, Shape, and Temperature," *Proc. Sixth IEEE Int'l Conf. Face and Gesture Recognition*, pp. 43-48, 2004.

[7] K. Bowyer, K. Chang, and P. Flynn, "A Short Survey of 3D and Multi-Modal 3D+2D Face Recognition," *Proc. Int'l Conf. Pattern Recognition*, 2004.

[8] N. Poh, S. Bengio, and J. Korczak, "A Multi-Sample Multi-Source Model for Biometric Authentication," *Proc. IEEE Workshop Neural Networks for Signal Processing*, pp. 375-384, 2002.

[9] J. Phillips, H. Moon, S. Rizvi, and P. Rauss, "The FERET Evaluation Methodology for Face-Recognition Algorithms," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090-1104, Oct. 2000.

[10] W. Yambor, B. Draper, and R. Beveridge, "Analyzing PCA-Based Face Recognition Algorithms: Eigenvector Selection and Distance Measures," *Proc. Second Workshop Empirical Evaluation in Computer Vision*, 2000.

[11] J. Min, P. Flynn, and K. Bowyer, "Using Multiple Gallery and Probe Images Per Person to Improve Performance of Face Recognition," Technical Report TR-03-7, Univ. of Notre Dame, 2003.

[12] G. Givens, R. Beveridge, B. Draper, and D. Bolme, "A Statistical Assessment of Subject Factors in the Pca Recognition of Human Faces," *Proc. Workshop Statistical Analysis in Computer Vision (CVPR)*, 2003.