

Face Recognition Using 2D and 3D Facial Data

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

Computer Science & Engineering Department
University of Notre Dame
Notre Dame, IN 46556
{kchang, kwb, flynn}@cse.nd.edu

Abstract

Results are presented for the largest experimental study to date that investigates the comparison and combination of 2D and 3D face recognition. To our knowledge, this is also the only such study to incorporate significant time lapse between gallery and probe image acquisition, and to look at the effect of depth resolution. Recognition results are obtained in (1) single gallery and a single probe study, and (2) a single gallery and multiple probe study. A total of 275 subjects participated in one or more data acquisition sessions. Results are presented for gallery and probe datasets of 200 subjects imaged in both 2D and 3D, with one to thirteen weeks time lapse between gallery and probe images of a given subject yielding 951 pairs of 2D and 3D images. Using a PCA-based approach tuned separately for 2D and for 3D, we find that 3D outperforms 2D. However, we also find a multi-modal rank-one recognition rate of 98.5% in a single probe study and 98.8% in multi-probe study, which is statistically significantly greater than either 2D or 3D alone.

1. Introduction

The identification of the human face in 2D has been investigated by many researchers, but relatively few 3D face identification studies have been reported [1, 2, 3, 4, 5]. One of the main motivations of 3D face recognition is to overcome the problems in general 2D recognition methods resulting from illumination, expression or pose variations.

This study deals with face recognition using 2D and 3D. Each modality captures different aspects of facial features, 2D intensity representing surface reflectance and 3D depth values representing face shape data. Even though each modality has its own advantages and disadvantages depend-

ing on certain circumstances, there is often some expectation that 3D data should yield better performance. However, no rigorous experimental study has been reported to validate this expectation. The experiments reported in this study are aimed at (1) examining the spatial / depth resolution needed for 3D face recognition (2) testing the hypothesis that 3D face data provides better biometric performance than 2D face data, using the PCA-based method, and (3) exploring whether a combination of 2D and 3D face data may provide better performance than either one individually in both a single probe study and a multiple probe study.

This is an extension of our earlier work [6]. We have expanded the size of the dataset and have improved the method of geometric normalization used in the 2D and 3D PCA algorithms, resulting in improved recognition performance, both individually and in combination. We have also examined the effect of depth resolution on performance of 3D recognition.

2. Previous Work

In this section, methods that use multiple types of biometric sources for identification purposes, multi-modal biometrics, are reviewed. The term “multi-modal biometrics” is used here to refer to the use of different sensor types without necessarily indicating that different parts of the body are used. The important aspects of these multi-modal studies are summarized in Table 1. Due to the effectiveness of combining multiple biometrics, such studies are included as well to review their data fusion methods, types of biometric sources and the size of experimental dataset. In addition to recognition methods based solely on the human face, there are other recognition methods using multiple biometric sources in addition to face data. One commonality of the studies described in Table 1 is that identification based on multiple

Table 1: Multi-biometrics studies for personal identification

Source (year)	Biometric sources	Fusion methods	Set size
Wang ('03) [7]	Face, Iris	metric-based	90
Chang ('03) [8]	Face, Ear	pixel-based	111
Shakhnarovich('02) [9]	Face, Gait	metric-based	26
Ross ('01) [10]	Face, Hands Fingerprint	metric-based	50
Frischholz ('00) [11]	Face, Voice, Lip Movement	metric-based	150
Ben-Yacoub ('99) [12]	Face, Voice	metric-based	37
Hong ('98) [13]	Face, Fingerprint	metric-based	64
Bigun ('97) [14]	Face, Voice	metric-based	40
Kittler ('97) [15]	Face, Profile Voice	metric-based	37
Brunelli ('95) [16]	Face, Voice	metric- / rank-based	89

Studies that integrate multiple types of <i>facial data</i>			
Chang ('03) [6]	2D frontal& 3D shape	metric-based	278
Wang ('02) [17]	2D frontal & 3D shape	metric-based	50
Beumier ('00) [18]	2D frontal & 3D profiles	metric-based	120
Achermann ('96) [19]	2D frontal& 2D profile	metric- / rank-based	30

sensors / biometrics sources provides overall performance improvement.

3. Methods and Materials

3.1. 2D and 3D Face Recognition Using PCA

Extensive work has been done on face recognition algorithms based on PCA, popularly known as “eigenfaces” [20]. A standard implementation of the PCA-based algorithm [21] is used in the experiments reported here.

3.2. Normalization

The main objective of the normalization process is to minimize the uncontrolled variations that occur during the acquisition process and to maintain the variations observed in

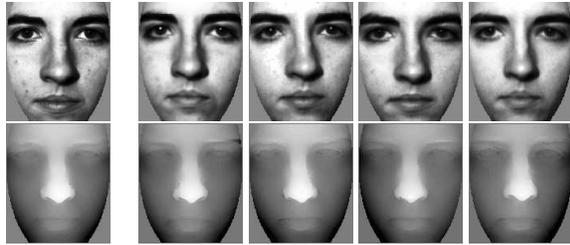
facial feature differences between individuals. The normalized images are masked to omit the background and leave only the face region (see Figure 1). While each subject is asked to gaze at the camera during the acquisition, it is inevitable to obtain data with some level of pose variations between acquisition sessions.

The 2D image data is typically treated as having pose variation only around the Z axis, the optical axis. The PCA software [21] uses two landmark points (the eye locations) for geometric normalization to correct for rotation, scale, and position of the face for 2D matching. However, the face is a 3D object, and if 3D data is acquired there is the opportunity to correct for pose variation around the X, Y, and Z axes.

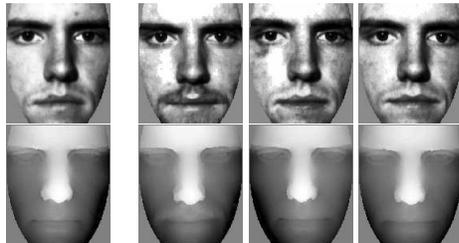
A transformation matrix is first computed based on the surface normal angle difference in X (roll) and Y (pitch) between manually selected landmark points (two eye tips and center of lower chin) and predefined reference points of a standard face pose and location. Pose variation around the Z axis (yaw) 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 of every subject is transformed to the same point in 3D relative to the sensor (see Figure 2). 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. Thus, the geometric scaling was not imposed to 3D data points as it was in 2D. We found that missing data problems with fully pose-corrected 2D outweighed the gains from the additional pose correction [6], and so we use the typical Z-rotation corrected 2D. Problems with the 3D are alleviated to some degree by preprocessing the 3D data to fill in holes and remove spikes (see Figure 3). This is done by median filtering followed by linear interpolation using valid data points around a hole.

3.3. Data Collection

A gallery image is an image that is enrolled into the system to be identified. A probe image is a test image to be matched against the gallery images. Images were acquired at the University of Notre Dame between January and May 2003. Two four-week sessions were conducted for data collection, approximately six weeks apart. The first session is to collect gallery images and the second session is to collect probe images for a single probe study in mind. For a study with multiple probes, an image acquired in the first week is used as a gallery and images acquired in later weeks are used as probes. Thus, in the single probe study, there are at least six and as many as thirteen weeks time lapse between the acquisition of gallery image and its probe image, and at



A study of one gallery with four probes

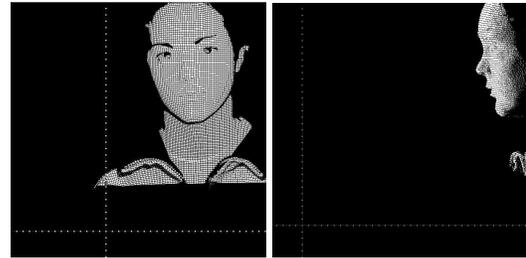


A study of one gallery with three probes

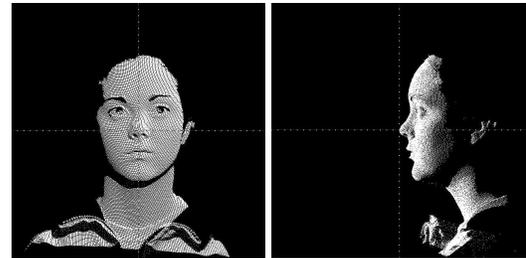
Figure 1: Examples of masked images in 2D and 3D

least one and as many as thirteen weeks time lapse between the gallery and the probe in the multiple probe study. All subjects completed an IRB-approved consent form prior to participating in each data acquisition session. A total of 275 different subjects participated in one or more data acquisition sessions. Among 275 subjects, 200 participated in both a gallery acquisition and a probe acquisition. Thus, there are 200 individuals in the single probe set, the same 200 individuals in the gallery, and 275 individuals in the training set. The training set contains the 200 gallery images plus an additional 75 for subjects whom good data was not acquired in both the gallery and probe sessions. And for the multiple probe study, 476 new probes are added to the 200 probes, yielding 676 probes in total. The training set of 275 subjects is the same as the set used in the single probe study.

In each acquisition session, subjects were imaged using a Minolta Vivid 900 range scanner. Subjects stood approximately 1.5 meter from the camera, against a plain gray background, with one front-above-center spotlight lighting their face, and were asked to have a normal facial expression (“ F_A ” in FERET terminology [22]) and to look directly at the camera. Almost all images were taken using the Minolta’s “Medium” lens and a small number of images was taken with its “Tele” lens. The height of the Minolta Vivid scanner was adjusted to the approximate height of the subject’s face, if needed. The Minolta Vivid 900 uses a projected light stripe to acquire triangulation-based range data. It also captures a color image near-simultaneously with the range data capture. The result is a 640 by 480 sampling of range data and a registered 640 by 480 color image.



(a) X-Y plane (b) Y-Z plane
Initial pose of a subject in 3D space



(a) X-Y plane (b) Y-Z plane
Corrected pose of a subject in 3D space

Figure 2: Pose normalization

3.4. Distance Metrics

2D data represents a face by intensity variation whereas 3D data represents a face by shape variation. It is obvious that the “face space” could be very different between modalities. Thus, during the decision process, certain metrics might perform better in one space than in the other. In this experiment, the Mahalanobis distance metric was explored during the decision process for the gallery matching [23].

3.5. Data Fusion

The pixel level provides perhaps the simplest approach to combining the information from multiple image-based biometrics. The images can simply be concatenated together to form one larger aggregate 2D-plus-3D face image. Metric level fusion combines the match distances that are found in the individual spaces. Having distance metrics from two or more different spaces, a rule for combination of the distances across the different biometrics for each person in the gallery can be applied. The ranks can then be determined based on the combined distances.

One of the early tasks in data fusion is to normalize the scores that result from the metric function. Scores from each space need to be normalized to be comparable. There are several ways of transforming the scores including linear, logarithm, exponential and logistic [19]. The scores from different modalities are normalized so that the distribution and the range are mapped to the same unit interval.

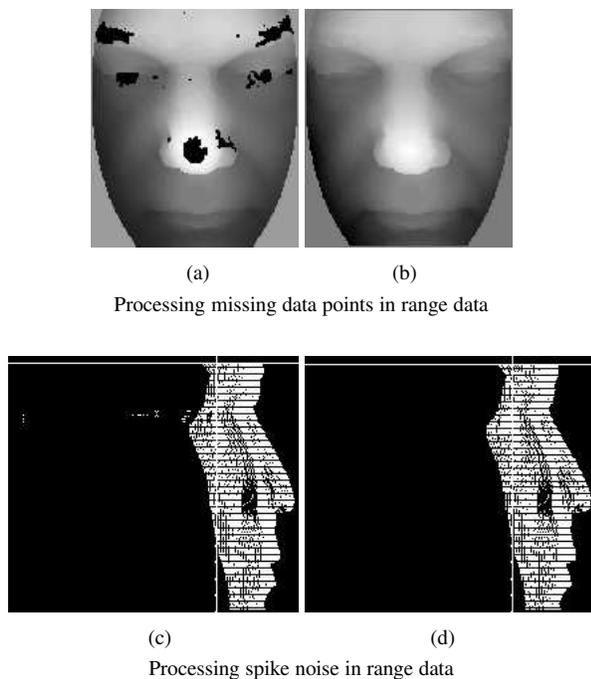


Figure 3: Preprocessing in 3D data points

There are many ways of combining different metrics to achieve the best decision process, including majority vote, sum rule, multiplication rule, median rule, min rule, average rule and so on. Depending on the task, a certain combination rule might be better than others. It is known that the sum rule and multiplication rule generally provide plausible results [24, 19, 9, 7, 6, 18].

In our study, a weight is estimated based on the distribution of the top three ranks in each space. The motivation is that a larger distance between first- and second-ranked matches implies greater certainty that the first-ranked match is correct. The level of the certainty can be considered as a weight representing the certainty. The weight can be applied to each metric as the combination rules are applied. The multi-modal decision is made as follows. First the 2D probe is matched against the 2D gallery, and the 3D probe against the 3D gallery. This gives a set of N distances in the 2D face space and another set of N distances in the 3D face space, where N is the size of the gallery. A plain sum-of-distances rule would sum the 2D and 3D distances for each gallery subject and select the gallery subject with the smallest sum. We use a confidence-weighted variation of the sum-of-distances rule. For each of 2D and 3D, a “confidence” is computed using the three distances in top ranks as $(\text{second distance} - \text{first distance}) / (\text{third distance} - \text{first distance})$. If the difference between the first and second match is large compared to the typical distance, then this confidence value will be large. The confidence values are

used as weights in distance metric. A simple product-of-distances rule produced similar combination results, and a min-distance rule produced slightly worse combination results.

4. Experiments

There are three main parts to this study. The first part is to examine how the recognition performance is affected by the X - Y in both 2D and 3D and depth resolution in 3D data. The second part is to evaluate the performance of 2D and 3D independently in both single and multiple probe studies. Data fusion is considered, in the third part, to combine results at the metric level with different fusion strategies.

The eigenvectors for each face space are tuned by dropping the first M and last N eigenvectors to obtain an optimum set of eigenvectors. Thus, in general we expect to have a different set of eigenvectors 2D face space versus representing 3D face space. The cumulative match characteristic (CMC) curve is generated to present the results.

4.1. Experimental Results: X - Y resolution

This experiment looks at the performance rate changes while the spatial resolution is varied in texture and shape images. One average pixel in X axis produced by the Minolta Vivid 900 covers $0.9765mm$ and one pixel in Y axis covers $0.9791mm$ of surface area. A typical template size that we initially used was 130×150 pixels (a face coverage area of approximately $12.7cm \times 14.7cm$). Figure 4-(a) shows example of both 2D (top row) and 3D (bottom row) images used for this experiment, starting from the right most, 25%, 50%, 75%, 100% of the original dimension. Thus, every pixel is retrieved in the step of $3.97mm$, $1.96mm$, $1.31mm$ and $0.98mm$ from the original X and Y data points in each image set.

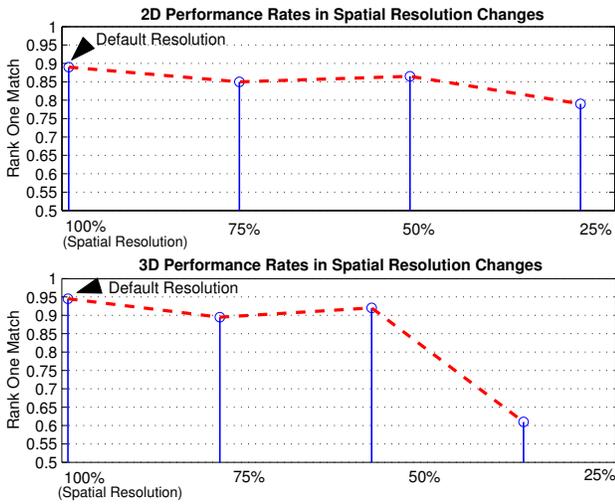
The performance results are shown in Figure 4-(b). The graph is plotted using the first rank match performance rate. Both performance curves begin to drop at the resolution of $1.31mm$ in X - Y , (in 2D, 89.0% to 85.0%, and 94.5% to 89.5% in 3D). However, the spatial resolution changes attempted in both 2D and 3D suggest that there is no significant difference in performance rates from the original resolution. We believe that performance degradation results from undersampling the face and missing differentiating features. The stiff performance drop has been shown in between 50% and 25% due to the insufficient facial features to be differentiated between subjects in PCA method.

4.2. Experimental Results: Depth resolution

This experiment has a similar purpose as the previous one. However, this examines the depth resolution required to



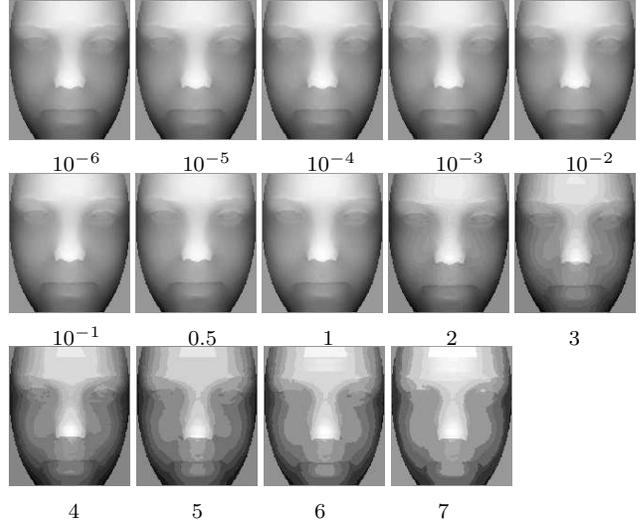
(a) Example of images in different spatial resolutions



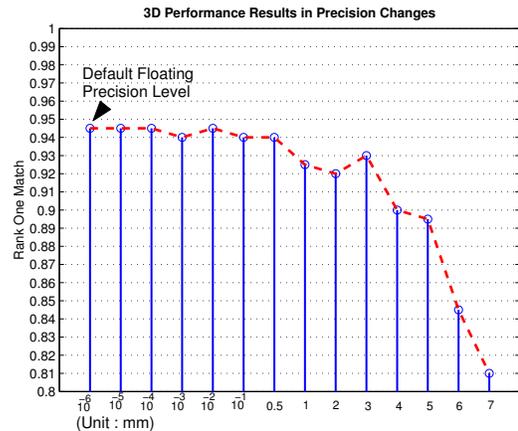
(b) Different spatial resolutions

Figure 4: Experiment in spatial resolutions changes

maintain the performance rate from the original depth resolution. According to the Minolta Vivid 900 specification, its depth accuracy level may be obtained at $0.35mm$. One way to vary the original resolution is to change the precision level in floating point values of the Z coordinate. A lower limit on precision could be $10^{-6}mm$. However, the camera-to-subject distance and lens combination used in our acquisition likely support an actual depth resolution of no better than about $0.5mm$ on average. Fourteen different resolutions were examined so that every pixel value representing the actual coordinate is retrieved in the unit of $10^{-6}mm$, $10^{-5}mm$, $10^{-4}mm$, $10^{-3}mm$, $10^{-2}mm$, $10^{-1}mm$, $0.5mm$, $1mm$, $2mm$, $3mm$, $4mm$, $5mm$, $6mm$ and $7mm$ as shown in Figure 5-(a). As shown in Figure 5-(b), the overall performance rate decreases as the depth



(a) Example of images in different depth resolutions (in mm)



(b) Performance results in different depth resolutions

Figure 5: Experiment in depth resolution changes

resolution gets coarser. It becomes prominent after $3mm$.

However, it is interesting to note that the performance rates between $0.5mm$ and $3mm$ maintain remarkably close to the original resolution (within 2.5%). This may be partially because as the resolution gets coarser, random noise would be suppressed. As it gets even coarser, a face surface becomes overly contoured and identification suffers from such coarsely quantized surfaces.

4.3. Experimental Results: 2D versus 3D face - Single probe study

This experiment is to investigate the performance of individual 2D eigenface and 3D eigenface methods, given (1) the use of the same PCA-based algorithm implementation,

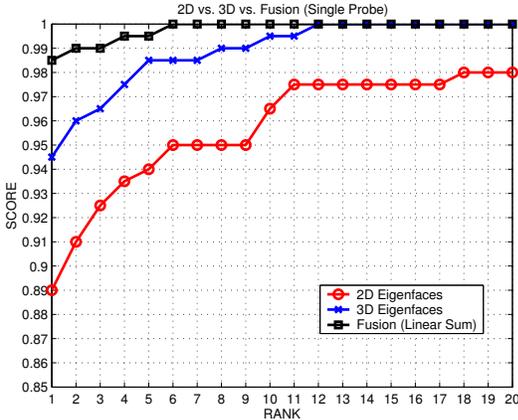


Figure 6: Performance results in single probe study.

(2) the same subject pool represented in training, gallery and probe sets, and (3) the controlled variation in one parameter, time of image acquisition, between the gallery and probe images. A similar comparison experiment between 2D and 3D acquired using stereo-based system was also performed by Medioni *et al.*[25].

There can be many ways of selecting eigenvectors to accomplish the face space creation. In this study, at first, one vector is dropped at a time from the eigenvectors of largest eigenvalues, and the rank-one recognition rate is computed using the gallery and probe set again each time, and continue until a point is reached where the rank-one recognition rate gets worse rather than better. We denote the number of dropped eigenvectors of largest eigenvalues as M . Also, one vector at a time is dropped from the eigenvectors of the smallest eigenvalues, and the rank-one recognition is computed using the gallery and probe set again each time, and continue until a point is reached where the rank-one recognition rate gets worse rather than better. We also denote the number of dropped eigenvectors of smallest eigenvalues as N .

During the eigenvector tuning process, the rank-one recognition rate remains basically constant with from one to 20 eigenvectors dropped from the end of the list. This probably means that more eigenvectors can be dropped from the end to create a lower-dimension face space. This would make the overall process simpler and faster. The rank-one recognition rate for dropping some of the first eigenvectors tend to improve at the beginning but it start to decline as M gets larger.

After the eigenvectors are tuned, both 2d and 3D are coincided at $M = 3$, and $N = 0$ to create the face spaces. With the given optimal set of eigenvectors in 2D or 3D, the results show that rank-one recognition rate is 89.0% for 2D, and 94.5% for 3D (see Figure 6).

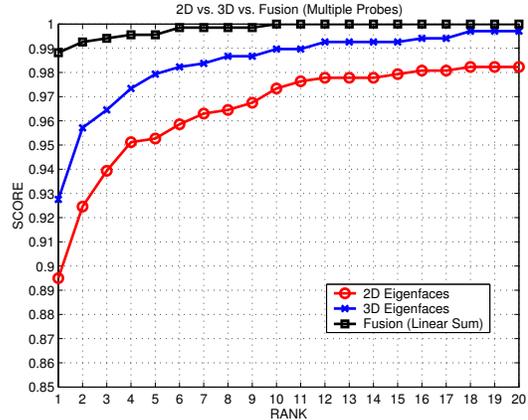


Figure 7: Performance results in multiple probe study.

4.4. Experimental Results: Multi-modal biometrics using 2D and 3D

The purpose of this experiment is to investigate the value of a multi-modal biometric using 2D and 3D face images, compared against individual biometrics. The null hypothesis for this experiment is that there is no significant difference in the performance rate between uni-biometrics (2D or 3D alone) and multi-biometrics (both 2D and 3D together). According to Hall [26], a fusion can be usefully done if an individual probability of correct inference is between 50% and 95% with one to seven classifiers. From our results in the previous experiment, it is reasonable to fuse the two individual biometrics which meet this fusion criteria. Figure 6 shows the CMC with the rank-one recognition rate of 98.5% for the multi-modal biometric, achieved by combining modalities at the distance metric level. In the fusion methods that we considered, the multiplication rule showed the most consistent regardless of the particular score transformation. However, the min rule showed lower performance than any other rules in different score transformations (see Figure 8). Also, when the distance metrics were weighted based on the confidence level during the decision process, all the rules result in significantly better performance than the individual biometric. A McNemar's test for significance of the difference in accuracy in the rank-one match between the multi-modal biometric and either the 2D face or the 3D face alone shows that multi-modal performance is significantly greater, at the 0.05 level.

4.5. Experimental Results: 2D face versus 3D face in biometrics - multiple probe study

In these experiments, there will be one or more probes for a subject who appears in the gallery, with each probe being acquired in a different acquisition session separated by

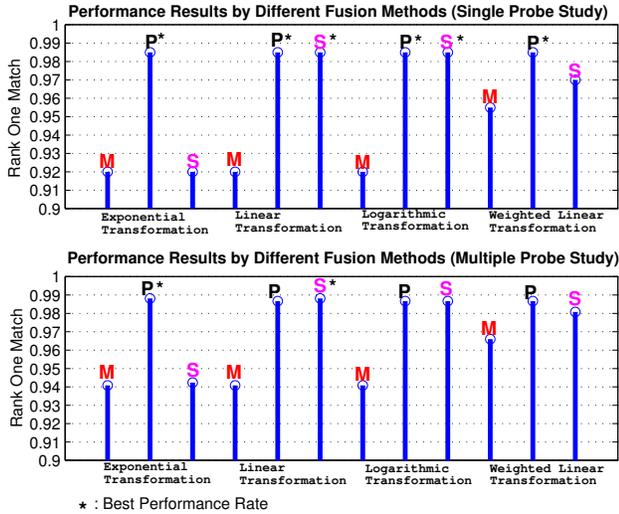


Figure 8: Performance results of fusion schemes used.

a week or more. We are attempting to retrieve more practical use of face identification method by incorporating multiple probes to be matched against the gallery images. The multiple probe dataset consists of 676 probes in total. Subjects might have a different number of probes. For example, there are 200 subjects with 1 or more probes, 166 subjects with 2 or more probes and so on. In the probe dataset, the number of probes can be up to 7 per subject. There might be different rules to determine a correct match given several probes to a gallery. In this experiment, a correct match is measured based on an each individual probe rather than on some function of all probes per subject.

By using the same set of eigenvectors tuned in the single probe study, we achieved similar results as in the previous sections. While 3D performance dropped a little, 92.8%, 2D performance maintains slightly better than the previous experiment, 89.5% (see Figure 7).

After combining these two biometrics in the multiple probes, we also were able to obtain significantly better performance, at 98.8%, than for either 2D or 3D alone. The results of 2D and 3D combination show very similar performance behavior as the single probe study. Product rule performs better than minimum rule regardless of score transformation (see Figure 8). Most combined methods consistently perform significantly better than the single biometrics. A McNemar’s test for significance of the difference in accuracy in the rank-one match between the multi-modal biometric and either the 2D face or the 3D face alone shows that multi-modal performance is significantly greater, at the 0.05 level. Thus, significant performance improvement has been accomplished by combining 2D and 3D facial data in both single and multiple probe studies.

5. Summary and Discussion

The value of multi-modal biometrics with 2D intensity and 3D shape of facial data in the context of face recognition is examined in a single probe study and a multiple probe study. This is the largest experimental study (in terms of number of subjects) that we know of to investigate the comparison and combination of 2D and 3D data for face recognition. In our results, each modality of facial data has roughly similar value as an appearance-based biometric. The combination of the face data from both modalities results in statistically significant improvement over either individual biometric. In general, our results appear to support the conclusion that *the path to higher accuracy and robustness in biometrics involves use of multiple biometrics* rather than the best possible sensor and algorithm for a single biometric.

We also have investigated the effect of spatial and depth resolution on recognition performance. This was done by producing successively coarser versions of the original image. The original image has a depth accuracy at $0.35mm$. We found that performance drops only slightly in going to a depth resolution of $0.5mm$, but begins to drop drastically at $4mm$. The pattern of results suggests that it would be interesting to determine a sensor accuracy level needed to meet a specific requirement of face recognition tasks. The accuracy requirement might vary under different conditions of subjects, such as facial muscle movement, or imaging condition changes. This initial investigation in resolution variation would bring a more explicitly decided resolution level for further experiments.

The overall quality of 3D data collected using a range camera is perhaps not as reliable as 2D intensity data. 3D sensors in the current market are not as mature as 2D sensors. Common problems with typical range finder images include missing data in eyes, cheeks, or forehead as well as several types of noise. These problems would lower the 3D recognition rate in general even though there exist ways of recovering some data in such areas.

The criteria used to decide which combination of eigenvectors to keep is the rank-one recognition rate on the gallery and probe images. So, in a way, the gallery and probe images are used in deciding what eigenvectors to use for the space, and then the results are also reported on the gallery and probe images, thereby “testing on training data”. This can be addressed by having a validation set of images to determine the set of eigenvectors to be used during the identification process so that eigenvectors to keep before the performance on the gallery and probe images are obtained.

It is generally accepted that performance estimates for face recognition will be higher when the gallery and probe images are acquired in the same acquisition session, compared to performance when the probe image is acquired after some passage of time [27]. Most envisioned applications for face recognition technology seem to occur in a scenario

in which the probe image would be acquired some time after the gallery image. In this context, it is worth noting that the dataset used here incorporates a substantial time lapse between gallery and probe image acquisition.

The dataset used in the experiments reported here will be made available to other research groups as a part of the Human ID databases. See <http://www.nd.edu/~cvrl/> for more information about the dataset and the release agreement.

6. Acknowledgments

This work was supported by the DARPA HumanID program through Office of Naval Research N000140210410 and by the National Science Foundation under grant EIA 01-30839.

References

- [1] L. Lee, B. Moghaddam, H. Pfister, and R. Machiraju, "Silhouette-based 3D face shape recovery," *Graphics Interface*, June 2003.
- [2] A. Bronstein, M. Bronstein, and R. Kimmel, "Expression-invariant 3D face recognition," *Audio and Video based Biometric Person Authentication*, pp. 62–69, 2003.
- [3] J. Huang, B. Heisele, and V. Blanz, "Component-based face recognition with 3D morphable models," *Audio and Video based Biometric Person Authentication*, pp. 27–34, 2003.
- [4] C. Chua, F. Han, and Y. Ho, "3D human face recognition using point signature," *Int'l Conf. on Automatic Face and Gesture Recognition*, pp. 233–238, 2000.
- [5] B. Achermann, X. Jiang, and H. Bunke, "Face recognition using range images," in *Proceedings Int'l Conf. on Virtual Systems and MultiMedia '97, Geneva, Switzerland*, Sept. 1997, pp. 129–136.
- [6] K. Chang, K. Bowyer, and P. Flynn, "Multi-modal 2D and 3D biometrics for face recognition," *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pp. 187–194, October 2003.
- [7] Y. Wang, T. Tan, and A. K. Jain, "Combining face and iris biometrics for identity verification," *4th Audio and Video based Person Authentication - Guildford, UK*, 2003.
- [8] K. Chang, K. Bowyer, V. Barnabas, and S. Sarkar, "Comparison and combination of ear and face images in appearance-based biometrics," *IEEE Trans. Pattern Anal. and Mach. Intel.*, vol. 25, pp. 1160–1165, 2003.
- [9] G. Shakhnarovich and T. Darrell, "On Probabilistic Combination of Face and Gait Cues for Identification," *Int'l Conf. on Automatic Face and Gesture Recognition*, pp. 169–174, 2002.
- [10] A. Ross and A. Jain, "Information fusion in biometrics," *Audio and Video based Person Authentication*, pp. 354–359, 2001.
- [11] R. Frischholz and U. Dieckmann, "BioID: A multimodal biometric identification system," *IEEE Computer*, vol. 33, no. 2, pp. 64–68, Feb. 2000.
- [12] S. Ben-Yacoub, "Multi-modal data fusion for person authentication using SVM," *Audio and Video based Person Authentication*, pp. 25–30, 1999.
- [13] L. Hong and A. Jain, "Integrating faces and fingerprints for personal identification," *IEEE Trans. Pattern Anal. and Mach. Intel.*, vol. 20, no. 12, pp. 1295–1307, 1998.
- [14] E. Bigun, J. Bigun, B. Duc, and S. Fischer, "Expert conciliation for multi modal person authentication systems by bayesian statistics," *Audio and Video based Person Authentication*, pp. 311–318, 1997.
- [15] J. Kittler, G. Matas, K. Jonsson, and M. Sanchez, "Combining evidence in personal identity verification systems," *Pattern Recognition Letters*, , no. 9, pp. 845–852, 1997.
- [16] R. Brunelli and D. Falavigna, "Person identification using multiple cues," *IEEE Trans. Pattern Anal. and Mach. Intel.*, vol. 17, no. 10, pp. 955–966, 1995.
- [17] 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.
- [18] C. Beumier and M. Acheroy, "Automatic face verification from 3D and grey level clues," *11th Portuguese Conf. on Pattern Recognition*, Sept. 2000.
- [19] B. Achermann and H. Bunke, "Combination of face classifiers for person identification," *Int'l Conf. on Pattern Recognition*, pp. 416–420, 1996.
- [20] M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [21] R. Beveridge and B. Draper, "Evaluation of face recognition algorithms (release version 4.0)," URL : <http://www.cs.colostate.edu/evalfacerec/index.html>.
- [22] J. Phillips, H. Moon, S. Rizvi, and P. Rauss, "The FERET Evaluation Methodology for Face-Recognition Algorithms," *IEEE Trans. Pattern Anal. and Mach. Intel.*, vol. 22, no. 10, pp. 1090–1104, October 2000.
- [23] W. Yambor, B. Draper, and J. Beveridge, "Analyzing PCA-based face recognition algorithms: Eigenvector selection and distance measures," *2nd Workshop on Empirical Evaluation in Computer Vision, Dublin, Ireland*, July 1 2000.
- [24] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. and Mach. Intel.*, vol. 20, no. 3, pp. 226–239, 1998.
- [25] G. Medioni and R. Waupotitsch, "Face modeling and recognition in 3-D," *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pp. 232–233, 2003.
- [26] D. Hall, *Mathematical Techniques in Multisensor Data Fusion*, Artech House, Norwood, MA., 1992.
- [27] J. Phillips, P. Grother, R. Micheals, D. Blackburn, E. Tabassi, and M. Bone, "Facial recognition vendor test 2002: Evaluation report," *NISTIR 6965*, URL: <http://www.frvt2002.org/FRVT2002/documents.htm>.