Available online at www.sciencedirect.com
SCIENCE DIRECT®
Computer Vision and Image Understanding 100 (2005) 357–384
Computer Vision and Image Understanding
www.elsevier.com/locate/cviu

Finger surface as a biometric identifier

Damon L. Woodard, Patrick J. Flynn *

Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46556, USA

Received 3 June 2004; accepted 8 June 2005
Available online 19 August 2005

Abstract

We present a novel approach for personal identification and identity verification which utilizes 3D finger surface features as a biometric identifier. Using 3D range images of the hand, a surface representation for the index, middle, and ring finger is calculated and used for comparison to determine subject similarity. We use the curvature based shape index to represent the fingers' surface. Gallery and probe shape index signatures are compared using the normalized correlation coefficient to compute a match score. A large unique database of hand images supports the research. We use data sets obtained over time to examine the performance of each individual finger surface as a biometric identifier as well as the performance obtained when combining them. Both identification and verification experiments are conducted. In addition, probe and gallery sets sizes are increased to further improve recognition performance in our experiments. Our approach yields good results for a first-of-its-kind biometric technique, indicating that this approach warrants further research.

Keywords: Biometrics; Hand shape; Performance evaluation; Surface curvature

1. Introduction

Personal identification continues to be a problem of interest to many researchers. Biometrics, which are physiological or behavioral characteristics used to distinguish

* Corresponding author. Fax: +1 574 631 9260.
E-mail addresses: dwoodard@nd.edu (D.L. Woodard), flynn@nd.edu (P.J. Flynn).

1077-3142/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2005.06.003

between individuals, are used as a solution to this problem. Researchers have investigated the use of a number of biometric identifiers including face, iris, and voice for various applications. Another source for biometric identifiers is the hand. Hand based systems are currently used in many applications.

Researchers have determined that the hand contains features that can be used for personal identification. These features include finger thickness and length, and palm thickness. A number of commercial systems which make use of these features are based on ideas that have been patented, and hence details of employed techniques are not available [5,8,18,23]. Despite this, researchers have been successful in developing systems capable of performance similar to that of commercial systems. A limited of amount of open literature related to hand based biometric research exists and is summarized in Table 1.

These efforts have investigated the use of finger characteristics as biometric features. Commonly used finger characteristics include finger length and width. A number of research efforts have examined the effectiveness of using these as biometric features. Jain et al. [10] developed a system that used measurements of the fingers and hand to establish identity. Sanchez-Reillo et al. [22] used a similar approach. Another characteristic used in prior research is that of finger shape. Jain and Duta [9] investigated the use of hand and finger shape extracted from the hand's silhouette as a biometric identifier. Very little work has been performed in 3D hand biometrics. Lay [16] used a grating pattern projected on the back surface of the hand and its distortion by the hand's shape as a biometric identifier. Our work represents the first to use the fine finger surface features such as skin folds and crease patterns extracted from dense range data as a biometric identifier. A curvature based representation is extracted from the registered finger images and used to generate a feature template. This template is compared to stored templates to determine similarity.

This paper, which is adapted from [25], is organized as follows. The paper begins with the details of the data collection and preprocessing procedures. A discussion of techniques used for matching score calculation as well as the biometric fusion rules implemented is provided. The results of identification and verification experiments are presented, to demonstrate the performance of our techniques. The paper concludes with a summary of the results and suggestions for future research.

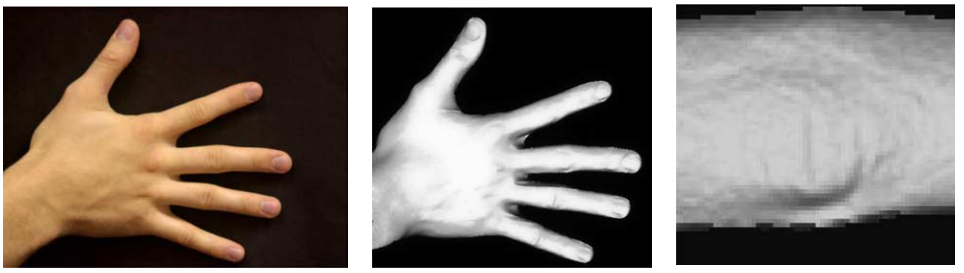
Table 1
Prior published research efforts

Author	Features	Matching Technique	Images	Subject	FAR (%)	FRR (%)
Jain et al. [10]	Hand/finger measurements	Various distances	360	—	1	10
Sanchez-Reillo et al. [22]	Hand/finger measurements	Various distances, GMMs, RBF	200	20	6.6	9
Jain and Duta [9]	Points on hand's contour	Minimization of mean alignment error	353	53	2	1.5
Lay [16]	Distorted projected grating pattern	Root mean square error	100	—	0.48	0.48

2. Data collection

Our hand data collections were part of a large multimodal database assembly effort which has been underway since early 2002. At the beginning of our research, a database of 3D hand data did not exist. To perform our research, a large database of 3D hand images would have to be constructed. For hand data collection the Minolta 900/910 sensor was used [14]. This sensor captures both a 640×480 range image and a registered 640×480 24-bit color intensity image nearly simultaneously. The sensor dimensions are 213 mm $\times$ 413 mm $\times$ 271 mm and it weighs about 11 kg. The sensor's cost, size, and weight limit the use of this sensor in a commercial biometric system but allows for the construction of our research database which will be made publicly available. See http://www.nd.edu/cvrl/ for details on access to this data. During data collection, the sensor is positioned approximately 1.3 m from a flat wall which has been covered with a black piece of cloth. Black cloth was chosen as the background to simplify the hand data segmentation task discussed later. Prior to data collection, the subject was instructed to remove all jewelry. The presence of jewelry during range image capture causes the emitted light from the sensor to scatter when contact is made with its reflective surface. The result is missing or inaccurate range image data near and at that location. The subject was instructed to place his or her right hand flat against the wall with the fingers naturally spread as the image is captured. Between image captures, the subject is instructed to remove his or her hand from the wall and then return it to approximately the same position.

Our database of collected data was obtained from male and female subjects between the ages of 18 and 70 from various ethnic groups. The majority of our data was collected from adults between the ages of 18 and 24. Data collection was performed on three separate weeks. During the first week, two images from 132 subjects were collected. Three images were collected a week later from the same 132 subjects. The third week of data collection took place approximately 16 weeks later. During the third week, three images were collected from 177 subjects of which 86 had participated in data collections during the prior 2 weeks. Hence, our efforts yielded a total of 1191 hand range images.



(A) Intensity image sample. (B) Range image sample. (C) Range image detail at finger knuckle.

Fig. 1. Sample intensity/range images and range image detail.

Fig. 1A shows a sample 640×480 color image of a hand. Fig. 1B is a pseudointensity of the same hand rendered using the 640×480 range image as a polygonal mesh. Fig. 1C depicts the surface detail detected near a knuckle. Our only requirement for hand placement is that the fingers are placed such that there is space between two adjacent fingers. By contrast, finger guide pegs were required for precise hand placement in past efforts [9,10,22].

3. Preprocessing

A number of preprocessing tasks are required prior to performing our experiments. All of the source code required for preprocessing was written in the MATLAB 6.5 programming language for easy prototyping [17]. The four required tasks were data re-sampling, hand segmentation, finger extraction, and feature template generation.

3.1. Data re-sampling

Due to slight variations in sensor position from week to week, the pixel spacing in x and y between adjacent range image pixels varied. The sampling interval values tended to cluster around 0.425 mm. We re-sampled the range images on a 0.4 mm grid (in both the x and y directions) to obtain a consistent sampling interval for all of the range images.

3.2. Hand segmentation

To work with only the range image pixels lying on the surface of the hand, the task of hand segmentation was required. To simplify this task, the intensity image of the hand was used. There is a pixel-to-pixel correspondence between intensity and range images. Therefore, we employed a combination of edge and skin detection techniques to the intensity image to reliably segment the hand from the image, thereby allowing for segmentation in the range image. The RGB color space skin detection rules specified in Kovač et al. [15] along with an implementation of a Canny edge detector comprise the segmentation module. The pseudocode of this module is as follows:

Algorithm Hand_Detect(I)
Input: I (m × n RGB image of the hand)
Output: I_{edge} (m × n binary image of edges detected with Canny edge detector)
Output: O (m × n binary image of the hand)

begin module

I_{edge} = Canny_Edge_Detector(I);
for i = 1 to m **do**

for j = 1 to n **do**
  if $I(i,j).R > 95$ AND $I(i,j).G > 40$ AND $I(i,j).B > 20$ AND
    $max\{I(i,j).R, I(i,j).G, I(i,j).B\} - min\{I(i,j).R, I(i,j).G, I(i,j).B\} > 15$ AND
    $|I(i,j).R - I(i,j).G| > 15$ AND $I(i,j).R > I(i,j).G$ AND $I(i,j).R > I(i,j).B$
  then
    $I_{skin}(i,j) = 1$; (Pixel is valid mask pixel)
  else
    $I_{skin}(i,j) = 0$; (Pixel is background pixel)
  end if
 end for
end for
$O = I_{skin} \cup I_{edge}$ (The final output is the union image of I_{skin} and I_{edge})
end module

Other color spaces such as the LUV color space could have been used for skin detection but given the high contrast between the skin complexion of the majority of the subjects and the background this was not necessary. Experimentation with RGB thresholds presented in [15] indicated that they would be suitable for our application. The RGB color space detection rules are applied to the intensity image to determine if a pixel lies on the surface of the hand. The result of this application is a binary image denoted as I_{skin}. Due to variances in illumination and subject skin complexion, there are skin pixels which are misclassified as background pixels. A Canny edge detector is applied to the intensity image to locate the outer edges of the subject's hand within the image. The result is a binary image denoted as I_{edge}. The images I_{skin} and I_{edge} are combined using the logical OR operation and holes within the resulting image filled using a combination of the open and close binary morphological operations to obtain the binary image depicted in Fig. 2A.

The binary image is traversed to extract the pixels lying on the hand silhouette's contour. Afterwards the contour of the hand's silhouette is computed. The index, middle, and ring fingers are denoted as α, β, γ, respectively, and depicted in Fig. 2B. The surfaces of α, β, and γ will be used in our experiments.

3.3. Finger extraction

The convex hull of the contour of the hand's silhouette is used to locate the valleys between the fingers represented as circles in Fig. 3A. The valley positions are used as segment boundaries allowing for α, β, and γ to be extracted and processed individually. The subjects are instructed to place their hand in relatively the same position so it is assumed that the relative positions of α, β, and γ are consistent among the collected images. Once α, β, and γ are extracted, we connected the pixels determined to be segment boundaries. We fill in this closed curve, producing a binary finger mask, as depicted in Fig. 3C. The shaded areas of Fig. 3B represent the extracted finger pixels.

We reduce the effects of noisy range data at the edges of the fingers by removing a two pixel wide portion of the finger mask perimeter. To address finger pose

(A) Segmented hand image. (B) Extracted hand silhouette.

Fig. 2. Segmented hand image and extracted contour.

(A) Finger valley locations. (B) Extracted finger pixels. (C) Finger mask example.

Fig. 3. Finger valley locations, extracted finger pixels, and finger mask example.

variations, finger masks along with their corresponding range pixels are rotated so that the major axis of the finger mask is coincident with the horizontal axis in an output finger range image. Following rotation, the finger mask pixels along with the corresponding range data is placed in a 80×240 finger image in which the finger mask is centered vertically and positioned five pixels from the right in the output image. The output finger images are used to generate the feature templates, which are used for comparisons.

3.4. Feature template generation

For each valid pixel of the finger mask in the output image, a surface curvature estimate is computed with the corresponding range data. Valid pixels of the finger mask are those in which the data at the corresponding pixel location in the range image lies on a finger surface and has been marked as valid by the sensor in the original range image. The linear regression technique, summarized in Flynn and Jain [6], is employed for this task. At each finger surface point of interest $\boldsymbol{p}$, we obtain a set of points S_p which neighbor it. We estimate the surface normal

and two orthogonal vectors which span the tangent plane centered at $\boldsymbol{p}$. A bi-cubic Monge surface

$$z = f(x,y) = ax^3 + bx^2y + cxy^2 + dy^3 + ex^2 + qxy + gy^2 + hx + iy + j$$

is then fit to S_p using linear regression. From the result, we calculate analytically the partial derivatives f_x, f_y, f_{xy}, f_{xx}, and f_{yy} to obtain the principal curvature values, $\kappa_{\min}$ and $\kappa_{\max}$ using the formula

$$\kappa_{\min,\max} = \frac{f_{xx} + f_{yy} + f_{xx}f_y^2 + f_{xx}f_x^2 - 2f_xf_yf_{xy}}{2(1 + f_x^2 + f_y^2)^{\frac{3}{2}}} \pm \sqrt{\left(\frac{f_{xx} + f_{yy} + f_{xx}f_y^2 + f_{yy}f_x^2 - 2f_xf_yf_{xy}}{2(1 + f_x^2 + f_y^2)^{\frac{3}{2}}}\right)^2 - \frac{f_{xx}f_{yy} - f_{xy}^2}{(1 + f_x^2 + f_y^2)^2}}.$$

These estimates of curvature contain noise. It has been suggested that the range data be smoothed prior to curvature estimation to limit the effects of noise. It was determined that if this approach is used in our application, many of the fine finger surface features are smoothed from the data. This problem is addressed by choosing a relatively large number of points for the Monge surface fit. The window used for determining neighboring points was varied from 3×3 to 15×15 pixels corresponding to a 2D extent of 1.2 mm $\times$ 1.2 mm and 6 mm $\times$ 6 mm, respectively. The optimal window size was chosen as 9×9 or 81 points which corresponds to a 2D extent of 3.6 mm $\times$ 3.6 mm. By using a larger number of points during surface fitting, the range data is implicitly smoothed. We found that if more than 81 points are used to fit the surface, many fine surface features are smoothed out from the range data. The computed principal curvature values were then used to compute the *Shape Index* *SI* value at each pixel, given by the formula:

$$SI = \frac{1}{2} - \frac{1}{\pi}\arctan\frac{(\kappa_{\max} + \kappa_{\min})}{(\kappa_{\max} - \kappa_{\min})} \kappa_{\max} \geqslant \kappa_{\min},$$

SI is a scalar in [0,1] with values that allow shape classification. In the rare case in which the computed principal curvature values are equal, thereby forcing the shape index formula to be undefined at a particular pixel, the shape index value at that pixel is assigned the value of zero. A zero value indicates that the surface at this pixel location is planar which is consistent with the case of equal principal curvature values. Table 2 illustrates the nine surface classes associated with each shape index value interval. Shape index was first proposed by Keonderink and van Doorn [13] and has been used successfully by Dorai and Jain [3,4] for free-form surface representation as well as global object recognition.

4. Matching technique

The match score is the sample normalized correlation coefficient given by the formula:

Table 2
Shape index intervals and associated surface classes

Surface type	Interval
Spherical cup	[0, 0.125)
Trough	[0.125, 0.25)
Rut	[0.25, 0.375)
Saddle rut	[0.375, 0.5)
Saddle	[0.5, 0.625)
Saddle ridge	[0.625, 0.75)
Ridge	[0.75, 0.875)
Dome	[0.875, 1.0)
Spherical cap	1.0

$$CC(SI_P, SI_G) = \left(\sum_{\substack{(i,j) \\ valid}} (SI_{\mathrm{P}}(i,j) - \overline{SI}_{\mathrm{P}}) * (SI_{\mathrm{G}}(i,j) - \overline{SI}_{\mathrm{G}}) \right) \Bigg/ \left(\sqrt{\left(\sum_{\substack{(i,j) \\ valid}} (SI_{\mathrm{P}}(i,j) - \overline{SI}_{\mathrm{P}})^2 \right) * \left(\sum_{\substack{(i,j) \\ valid}} (SI_{\mathrm{G}}(i,j) - \overline{SI}_{\mathrm{G}})^2 \right)} \right),$$

where $SI_{\mathrm{P}}(i,j)$, $SI_{\mathrm{G}}(i,j)$ are valid shape index values and $\overline{SI}_{\mathrm{P}}$, $\overline{SI}_{\mathrm{G}}$ are the sample mean shape index values in the probe and gallery images, respectively. If the shape index finger image is unraveled, it can be represented as a vector of shape index values. This vector can be treated as a waveform. By using the normalized version of the correlation coefficient, the shape of two shape index waveforms can be matched. The correlation between the waveforms would be independent of waveform amplitude (in our case, shape index value magnitude). The shape index value at a particular pixel among images captured from the same user could be affected by variances in the sensor's position or noise present in the range data. The normalized correlation coefficient is use to address this variance. The resulting match score lies in the interval $[-1,1]$ where a larger value indicates a better match. Earlier experiments involved the use of a count based shape class matcher [2,26]. The normalized correlation coefficient matcher outperformed the shape class matcher consistently during the current experiments and we use it exclusively henceforth.

Accurate match score calculation for each technique is dependent on proper alignment of the finger images. During preprocessing, care was taken to automatically align and center the finger mask in each output image. During each matching attempt, the number of overlapping pixels in the gallery and probe image is computed for three vertical offsets (+1 pixel, no offset, and −1 pixel) and the offset that maximizes the number of pixels in the overlap is employed during match score computation. On average, the set of overlapping pixels consists of approximately 18,500 pixels. We also experimented with horizontal shifting of images but found that it did not improve matching performance and hence was not required.

5. Score-level fusion rules

In addition to examining each individual finger's performance as a biometric, biometric fusion at the score level is implemented as described in Ross and Jain [21] and Hong et al. [7]. The matching score for each finger is treated as an output from a separate biometric system. The multiple scores are then fused into one overall match score using fusion rules proposed by Kittler et al. [12]. The first of three score fusion rules implemented is the *average fusion rule* defined as

$$FS_{\text{avg}} = \frac{1}{N}\left(\sum_{i=1}^{n}\alpha_i + \sum_{i=1}^{n}\beta_i + \sum_{i=1}^{n}\gamma_i\right), \tag{1}$$

where $\alpha_{1,\ldots,n}$, $\beta_{1,\ldots,n}$, and $\gamma_{1,\ldots,n}$ are the match scores calculated for each finger and $N = 3n$, which is the total number of match scores calculated during a single matching attempt. We perform experiments involving the use of multiple probe and gallery samples. Therefore, n is equal to the product of the number of probe and gallery samples used during a single verification attempt. The second fusion rule implemented is the *median fusion rule* defined as

$$FS_{\text{Median}} = \text{Median}\{\alpha_1, \alpha_2, \ldots, \alpha_n, \beta_1, \beta_2, \ldots, \beta_n, \gamma_1, \gamma_2, \ldots, \gamma_n\}, \tag{2}$$

where $\alpha_{1,\ldots,n}$, $\beta_{1,\ldots,n}$, and $\gamma_{1,\ldots,n}$ are the match scores calculated for each finger and FS_{Median} is the median valued match score of all match score calculated during a single verification attempt. The final fusion rule implemented is the *maximum fusion rule* defined as

$$FS_{\text{Max}} = \text{Max}\{\alpha_1, \alpha_2, \ldots, \alpha_n, \beta_1, \beta_2, \ldots, \beta_n, \gamma_1, \gamma_2, \ldots, \gamma_n\}, \tag{3}$$

where $\alpha_{1,\ldots,n}$, $\beta_{1,\ldots,n}$, and $\gamma_{1,\ldots,n}$ are the match scores calculated for each finger and FS_{Max} is the maximum valued match score calculated during a single verification attempt.

6. Identification experiments

For identification experiments, a closed-universe model (as described by Phillips et al. [20]) was used. In this model, every subject in the probe set is also in the gallery set; this allows for identification performance evaluation. All templates in the gallery set are assumed to have an identity known at enrollment. The probe image set represents a template calculated during a matching attempt. The identification experiments involved using 86 subjects in both the probe and gallery sets. Eight images were captured from each subject yielding a total of 688 images. Three fingers were extracted from each of the images producing 2064 finger images for use in identification experiments. Table 3 illustrates the various experiment configurations used. For convenience we label these experiments "Group A"–"Group I." Except for the "Group A" experiments, which did not use time-lapsed data, gallery image samples were chosen to be images collected prior to those chosen as probe images. Following this rule, only images collected during the second week can serve as probe and

Table 3
Identification experiment configurations

Experiment group	Finger	Probe	Gallery	Time lapse	No. of experiments
A	Single	Single	Single	None	42
	Fusion	Single	Single	None	42
B	Single	Single	Single	1 week	18
	Fusion	Single	Single	1 week	18
C	Single	Single	Single	16 week	45
	Fusion	Single	Single	16 week	45
D	Single	Single	Multiple	1 week	9
	Fusion	Single	Multiple	1 week	9
E	Single	Single	Multiple	16 week	18
	Fusion	Single	Multiple	16 week	18
F	Single	Multiple	Single	1 week	6
	Fusion	Multiple	Single	1 week	6
G	Single	Multiple	Single	16 week	15
	Fusion	Multiple	Single	16 week	15
H	Single	Multiple	Multiple	1 week	3
	Fusion	Multiple	Multiple	1 week	3
I	Single	Multiple	Multiple	16 week	6
	Fusion	Multiple	Multiple	16 week	6

gallery samples in the experiments. A total of 324 identification experiments were completed. During an identification attempt, only comparisons between images of the same finger type are performed. The pseudocode for the identification experiments was as follows:

Identification Experiment (P, G)
Output: GEN_SCORES (N genuine match scores)
Output: IMP_SCORES ($N(N-1)$ impostor match scores)
Output: CMC (List variable which stores N rank recognition rates)

begin module
GEN_SCORES $= \emptyset$;
IMP_SCORES $= \emptyset$;
CMC $= [0]$; (All entries in list set to zero)
RANK $= 0$;

foreach $U_P \in P$ (All subject images in probe set)
  ALL_SCORES $= \emptyset$; (List variable for all match scores computed for U_P)
  foreach $U_G \in G$ (All subject images in gallery set)
    MSC = Computed match score between U_P and U_G;
      if ($U_P.ID = U_G.ID$)
        GEN_SCORE = MSC;

```
        GEN_SCORES = GEN_SCORES ∪ MSC;
      else (U_P.ID ≠ U_G.ID)
          IMP_SCORE = IMP_SCORE ∪ MSC;
      end if
      ALL_SCORE = ALL_SCORE ∪ MSC;
  end for
  RANK = the rank of GEN_SCORE in ALL_SCORES list;
  Increment CMC(RANK) to CMC(END) by one;
end for
CMC = CMC/N;
end module
```

6.1. Experiment groups A–C: single probe/single gallery

6.1.1. Group A experiments

A set of experiments using images collected during the same day as probe and gallery set was performed to establish a baseline for identification performance. Two images were collected during the first week and three images during each of the two following weeks. This results in 14 different combinations involving images collected during the same session. The experiments were performed utilizing a single surface type and the three score level fusion rules. Fig. 4 depicts the recognition rate achieved for this set of experiments. Each curve represents the average performance achieved from 14 experiments utilizing data collected during the same session. Average rank one recognition rates between 94.5 and 98% were obtained. The α finger type experiments attained the lowest performance. Our baseline identification results suggest there is variation in the data collected during a single session. This may be due to cancellation and precision errors during shape index computation. The recognition performance obtained during these experiments was encouraging. However, this particular situation does not model a real world application. A more realistic scenario involves the comparison of probe images which are collected some period of time after gallery images. For the remainder of the experiments, we model this situation utilizing data with single and 16 weeks of time lapse.

6.1.2. Group B experiments

Fig. 5 presents the results of utilizing data collected with a single week of time lapse between gallery and probe acquisition. Each curve represents the average performance of six experiments. The average rank one recognition rate decreases between 58 and 85%. This decrease suggests that significant changes to the fingers' surface may have occurred between the collection of probe and gallery image samples. For the sake of comparison, a face recognition experiment utilizing face intensity images of the same set of subjects was performed. The experiment used the same configuration as the group B experiments, a single probe and gallery image sample was used during comparisons. The subjects displayed a neutral facial

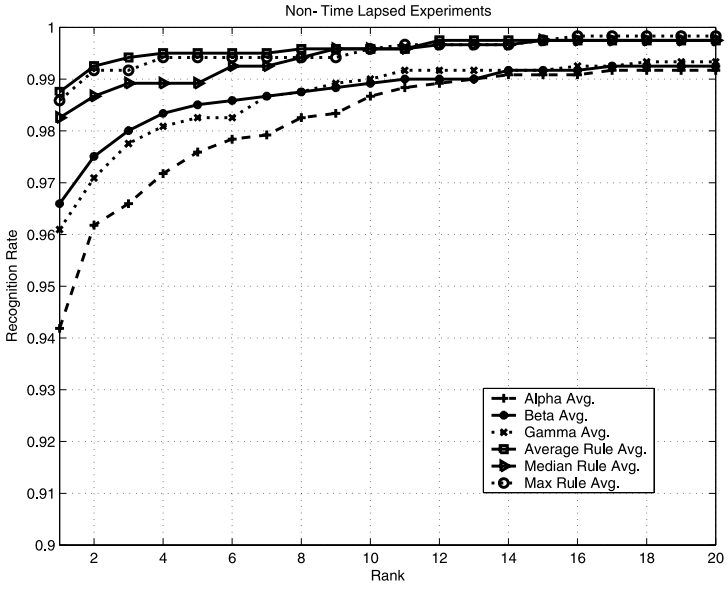


Fig. 4. Recognition performance: group A experiments.

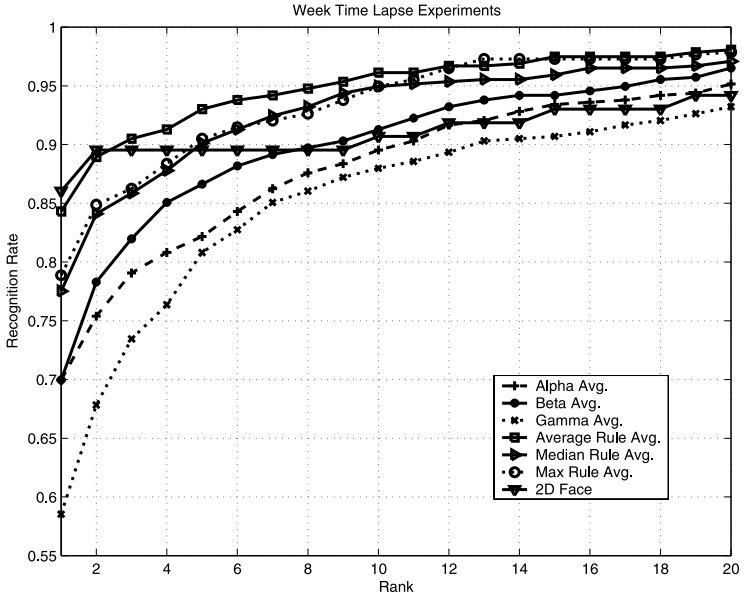


Fig. 5. Recognition performance: group B experiments.

expression under controlled lighting conditions during data collection. The face recognition experiments were conducted using the commercially available FACEIT software [24]. We can see from Fig. 5 that the performance obtained is only

slightly better, 86% as opposed to 85%, which was obtained using the average fusion rule.

6.1.3. Group C experiments

It would be reasonable to hypothesize that as the time lapse between images increases, the recognition rate should decrease significantly. Results from our group C experiments, as seen in Fig. 6, suggest that the relationship is not straightforward. Each curve represents the average of 15 experiments utilizing data with 16 weeks of time lapse between gallery and probe acquisition. There was a slight decrease in the average rank one recognition rate as compared to the group B experiments.

Even with a time lapse of 16 weeks, a average rank one recognition rate of about 84.7% was achieved when using the average fusion rule. In the majority of experiments, the fusion rule experiments performed better than experiments utilizing any single finger type. As with the single week time lapse experiments, we performed a face recognition experiment using data possessing a 16 week time lapse. The results are presented in Fig. 6. A rank one recognition rate of about 84.7% was obtained during the face recognition experiments, which is equal to rate obtained using the maximum fusion rule.

6.2. Experiment groups D–E: single probe/multiple gallery

Min et al. [19] have shown that increases in probe and gallery set sizes result in higher recognition rates. In the following sections we attempt to reduce the effects

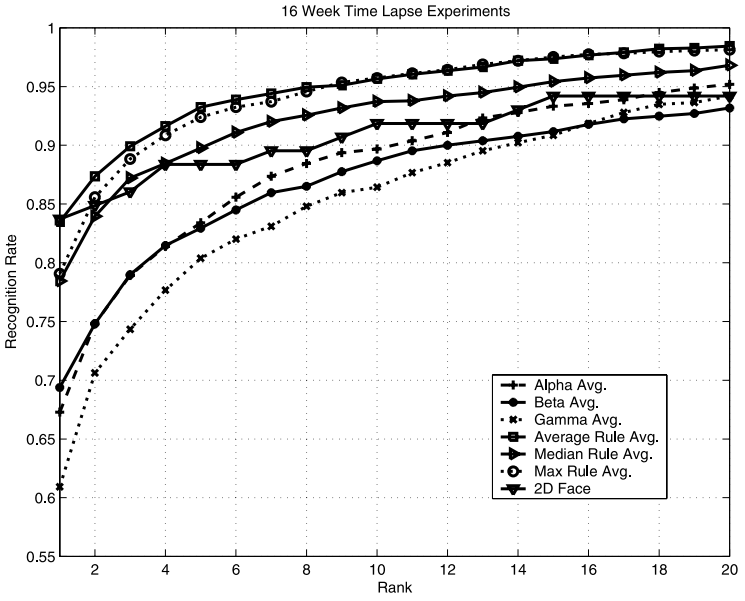


Fig. 6. Recognition performance: group C experiments.

of intra-session variation by increasing the number of image samples in the probe and gallery sets. To avoid biasing our results, the probe and gallery sets contain only images which were collected during the same acquisition session. Many biometric systems require more than a single image sample to generate the stored biometric feature template. Some hand geometry systems require three images to be captured during enrollment [27]. The features are extracted from the three image samples and averaged to generate the stored template. The next group of experiments models this situation.

6.2.1. Group D experiments

Fig. 7 depicts the results of experiments using single week time lapse data. Each curve represents the average of three experiments. In these experiments, a single probe image sample is compared to a gallery of two images during each attempt. During fusion rule experiments, this process is repeated for each finger type. A significant improvement in average rank one recognition performance is obtained. The recognition rate improves from about 84.7 to 88%. This improvement suggests that some of the variability between images may be caused by the presence of noise rather than by changes in the fingers' surface.

6.2.2. Group E experiments

The experiment configuration is repeated with data possessing a 16 week time lapse to determine if recognition rates would decline with a larger time lapse. Fig. 8 depicts the results of these experiments. Each curve of the graph represents

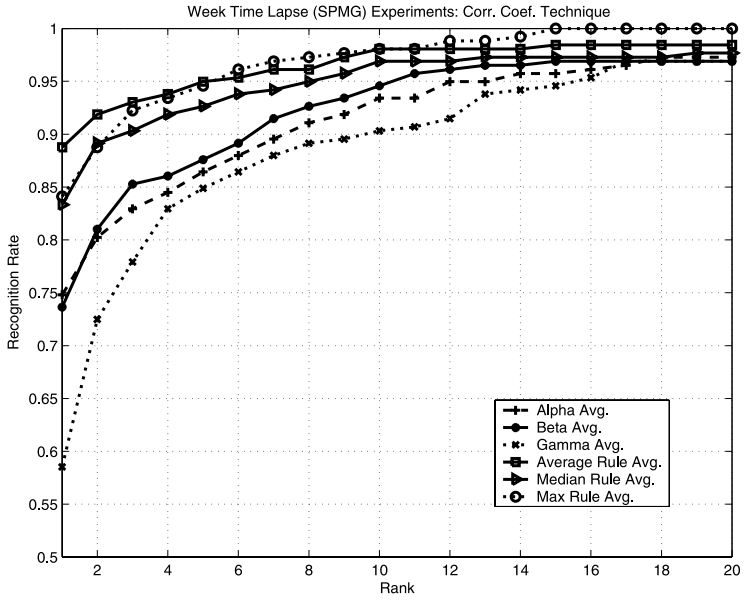


Fig. 7. Recognition performance: group D experiments.

the average of six experiments. There was a slight performance decrease compared to what was attained during single week time lapse experiments. Also, the experiments with the best recognition performance were those utilizing the maximum fusion rule.

6.3. Experiment groups F–G: multiple probe/single gallery

As suggested, intra-session variation exists among images collected during the same session. The previous experiments attempted to limit the effects of the variation in gallery images. This set of experiments attempts to do the same for probe images. For these experiments three probe image samples are compared to a single gallery image sample during each attempt. This process is repeated for each finger type during fusion rule experiments.

6.3.1. Group F experiments

Fig. 9 represents the results using single week time lapse data. Each curve of the graph represents the average performance of two experiments. There was a small decrease in identification performance as compared to the single probe/multiple gallery (SPMG) experiments. The best average rank one recognition rates decreased to about 87%. As in previous experiments, the average and maximum fusion rules performed best.

6.3.2. Group G experiments

The multiple probe/single gallery configuration is applied to data collected with a 16 week time lapse between gallery and probe acquisition. The results are depicted in

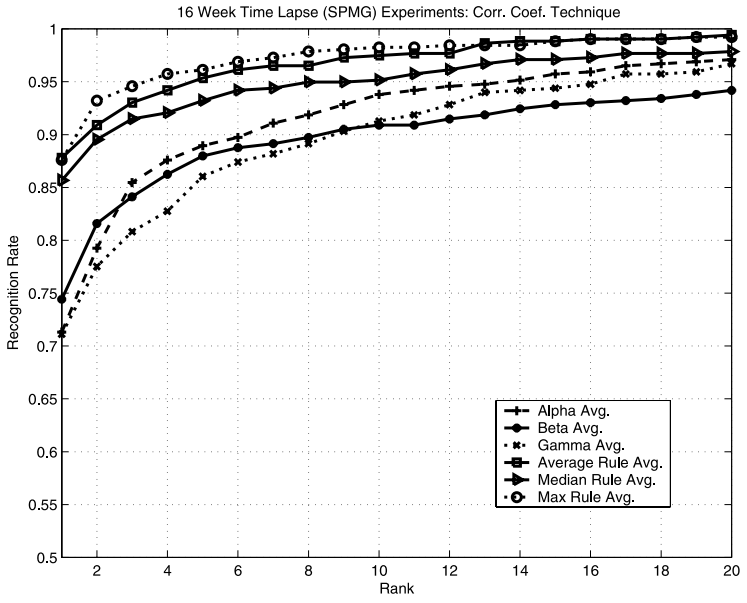


Fig. 8. Recognition performance: group E experiments.

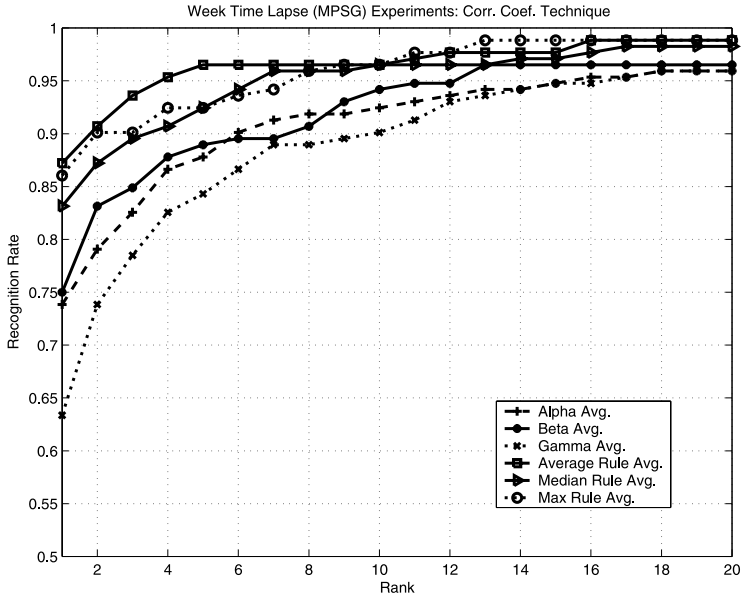


Fig. 9. Recognition performance: group F experiments.

Fig. 10. Each curve of the graph represents the average of five experiments. Each experiment involved comparing three probe images to a single gallery image during each attempt. As expected, there was a decrease in recognition rate when using data possessing a larger time lapse. Despite this, the highest recognition rates were still higher than those attained during the single probe/single gallery (SPSG) experiments.

6.4. Experiment groups H–I: multiple probe/multiple gallery

The final experiment configuration involved utilizing all images collected during a single session in the probe and gallery sets. This is believed to greatly minimize the effects of intra-session variation within the data.

6.4.1. Group H experiments

The results of single week time lapse experiments are presented in Fig. 11. Each curve is the result of a single experiment. An increase in performance was attained when using this experiment configuration. The average rank one recognition rate improved to 91%. The best performing fusion rules were the average and maximum rules. These results further support our assertion that much of the image variability is a result of measurement noise.

6.4.2. Group I experiments

Experiments using this configuration were performed using data which possessed a 16 week time lapse between gallery and probe acquisition. The results

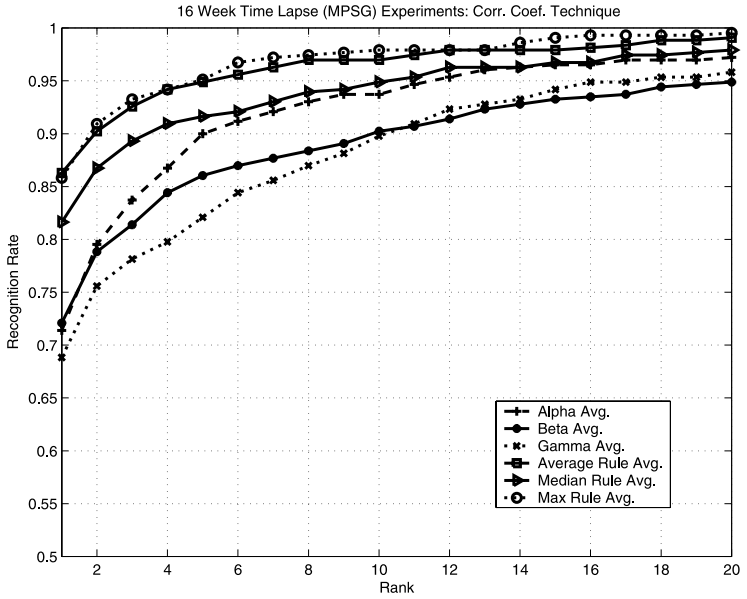


Fig. 10. Recognition performance: group G experiments.

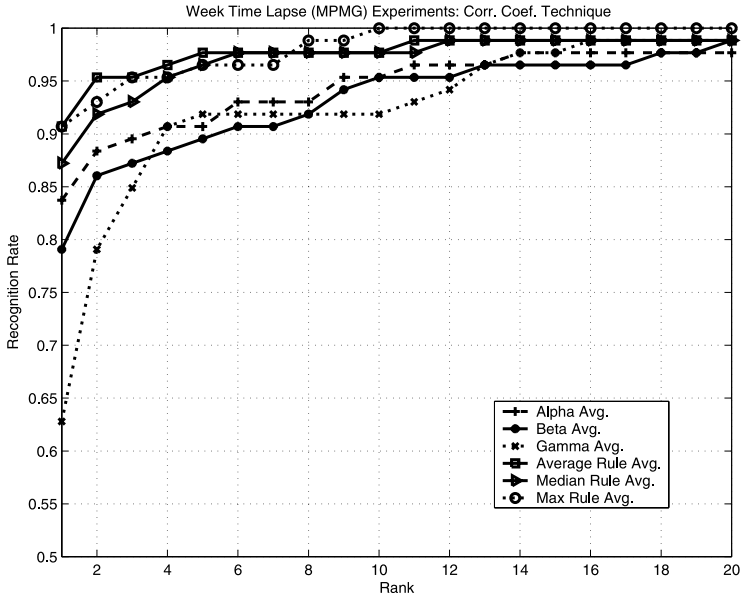


Fig. 11. Recognition performance: group H experiments.

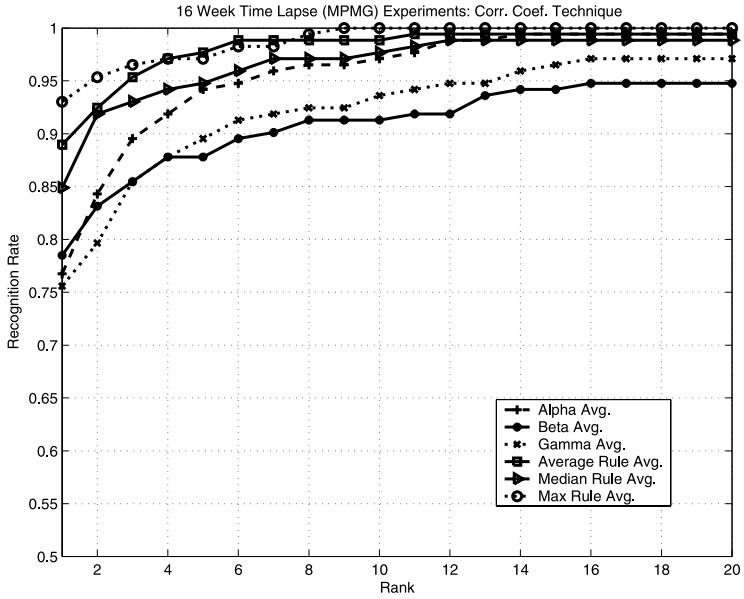


Fig. 12. Recognition performance: group I experiments.

are presented in Fig. 12. Each curve represents the average of the two such experiments. Interestingly, the highest average rank one recognition rate actually improved to approximately 94%. We believe that this may be due to a lesser sum amount of noise in the images used as compared to the previous experiments. The results from the identification experiments support the hypothesis that the finger surface may contain information which can be used to distinguish between individuals. In the following sections we present the results of verification experiments.

7. Verification experiments

Verification experiments involved the use of an open universe model, as described by Phillips et al. [20]. In this model, a subject in the probe set may or may not be present in the gallery set. The experiments used a probe set of 177 subjects and a gallery of 132 subjects. Of the subjects used, 86 are present in both the probe and gallery sets. Therefore, a total of 223 unique subjects were used for these experiments, many more than in previous related work. A total of 168 verification experiments were performed. Table 4 illustrates the various experiment configurations used.

Unlike the identification experiments, the probe and gallery images of all verification experiments possessed a 16 week time lapse between their collection. This time lapse represents the longest used in current literature. During a verification attempt, only comparisons between images of the same finger type are performed. The pseudocode for the verification experiments is as follows:

Verification Experiment (P,G)
Output: FAR (List variable which stores False Acceptance Rates for 101 threshold values)
Output: FRR (List variable which stores False Rejection Rates for 101 threshold values)

begin module
FAR = [0]; (All entries in list set to zero)
FRR = [0]; (All entries in list set to zero)

foreach $U_P \in P$ (All subject images in probe set)
  foreacn $U_G \in G$ (All subject images in gallery set)
    MSC = Computed match score between U_P and U_G;
    LOC = $\lfloor MSC/0.01 \rfloor$; (Finds the applicable threshold value)
    if ($U_P.ID = U_G.ID$)
      if LOC $\geqslant 1$
        Increment FRR(LOC + 1) to FRR(END) by one;
      else
        Increment FRR(1) to FRR(END) by one;
      end if
    end if
    if ($U_P \neq U_G$)
      if LOC $\geqslant 1$
        Increment FAR(1) to FAR(LOC + 1) by one;
      else
        Increment FAR(1) by one;
      end if
    end if
  end for
end for
end module

The following sections present the results of using each of the matching techniques for verification as Receiver Operating Characteristic (ROC). The match threshold was varied from 0 to 1 in increments of 0.01 for curve generation. At each threshold value the False Acceptance Rate (FAR) and False Rejection Rate (FRR) are calculated. For each experiment configuration we use the Equal Error Rate (EER) to quantify verification performance.

Table 4
Verification experiment configurations

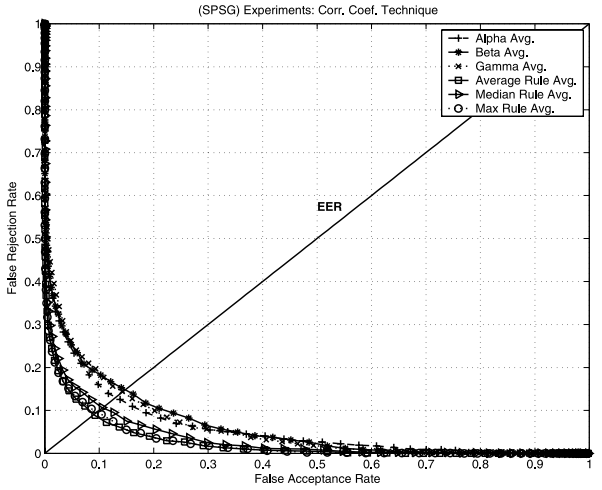
Experiment group	Finger	Probe	Gallery	Time lapse	No. of experiments
J	Single	Single	Single	16 week	45
	α, β, and γ	Single	Single	16 week	45
K	Single	Single	Multiple	16 week	18
	α, β, and γ	Single	Multiple	16 week	18
L	Single	Multiple	Single	16 week	15
	α, β, and γ	Multiple	Single	16 week	15
M	Single	Multiple	Multiple	16 week	6
	α, β, and γ	Multiple	Multiple	16 week	6

7.1. Group J experiments

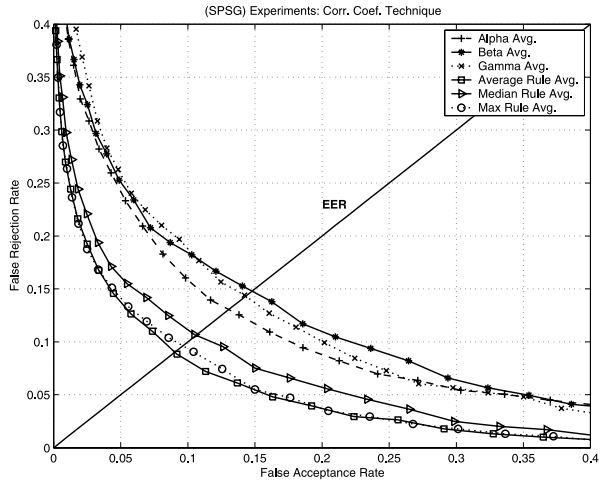
For this configuration, a single probe image is compared to a single gallery image during each verification attempt. For each experiment, templates from 132 probe subjects are compared to templates of 177 gallery subjects which result in a total of 23,364 performed verification attempts. Of these attempts, 86 are genuine and 23,278 are impostor. The false acceptance rates (FAR) and false rejection rates (FRR) are computed for each threshold value and plotted as a ROC curve in Fig. 13. Each curve represents the average of 15 experiments. As expected, the fusion rule experiments exhibit better performance than the single finger experiments. Fig. 13B provides a close view of the graph region where the equal error rates lie. The lowest equal error rate obtained is approximately 9% and achieved by the average fusion rule. The set of fusion rules experiments exhibited better performance than any of the single finger surface experiments.

7.2. Group K experiments

During this set of experiments, the number of gallery samples used during match score calculation was increased in an effort to reduce the equal error rate. During the single finger surface experiments, the overall match score is computed as the average of the matching scores computed during that verification attempt. In contrast, during the fusion rule experiments the overall match score is computed according to fusion rule. This experiment configuration involved comparing a single probe image to either two or three gallery images during each verification attempt. The curves of the ROC plot in Fig. 14 represent the average of six experiments. There was not a significant difference in performance when using either two or three gallery samples which may indicate that there was not a significant amount of variation present between the gallery sample images. The Group K experiments obtained a lowest equal error rate of about 7% as compared to 9% obtained during Group A experiments, as shown in Fig. 14B. The fusion rules continued to outperform the single finger type experiments.



(A) ROC curves.

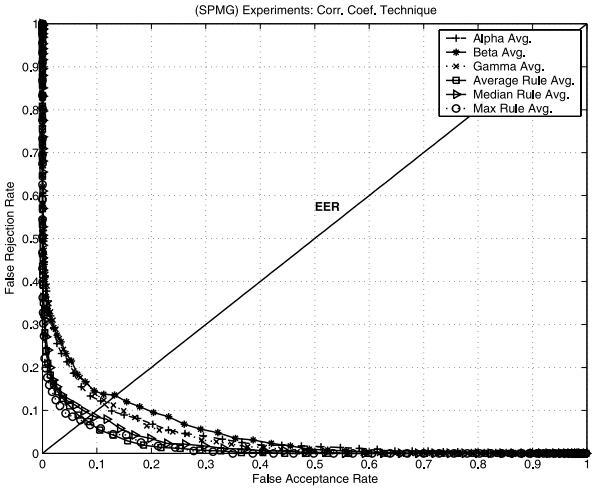


(B) Equal error rates.

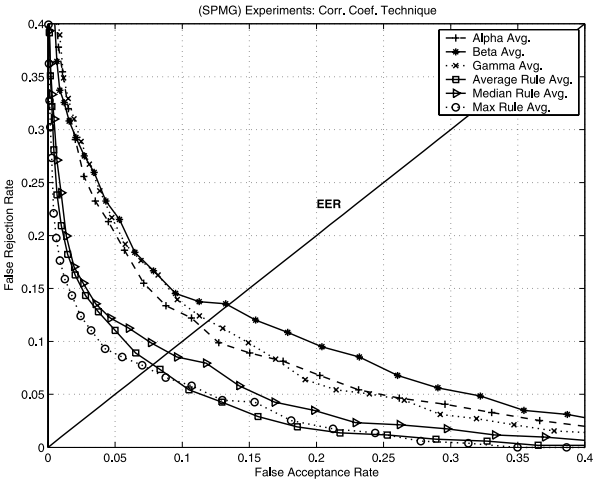
Fig. 13. Verification performance: group J experiments.

7.3. Group L experiments

Three probe image samples were compared to a single gallery image sample during this set of experiments. The curves of the ROC plot in Fig. 15 represent the average of five experiments. A slight decrease in equal error rate occurred using this configuration as compared to the previous experiment group. The lowest equal error rate was attained using the average fusion rule and was approxi-



(A) ROC curves.



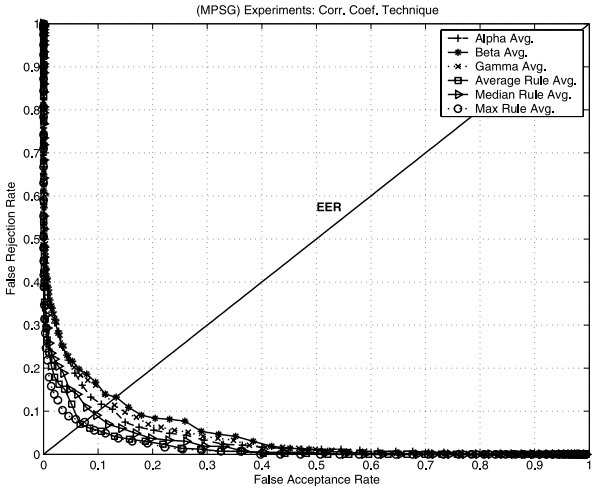
(B) Equal error rates.

Fig. 14. Verification performance: group K experiments.

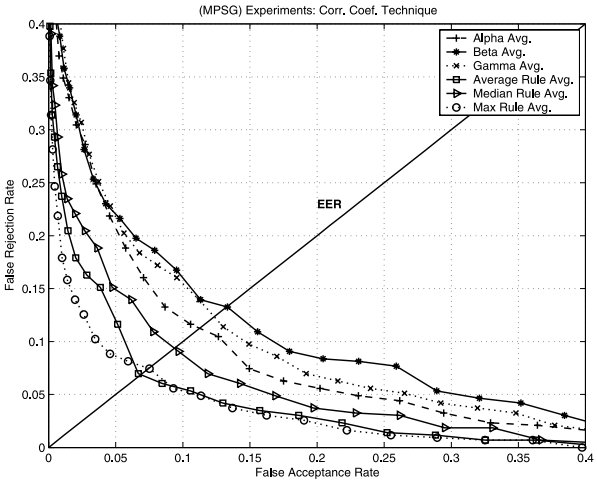
mately 6%. This performance improvement over the Group J experiments would suggest the existence of variation between probe image samples collected from the same subject.

7.4. Group M experiments

The final experiment configuration involved utilizing all images collected during a single session in the probe and gallery sets. Three probe image samples are



(A) ROC curves.



(B) Equal error rates.

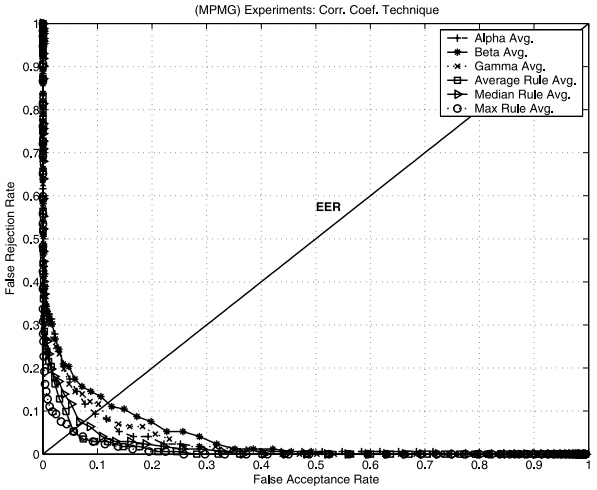
Fig. 15. Verification performance: group L experiments.

compared to all five of the gallery image samples collected 16 weeks prior. The result of a single experiment is presented in Fig. 16. Using this experiment configuration, we achieved an equal error rate of 5.5% using both the maximum and average fusion rule, as shown in Fig. 16B. This slight performance improvement in performance over the experiments performed in the pervious section suggests that there is exists a significant amount of variation between images in the gallery set.

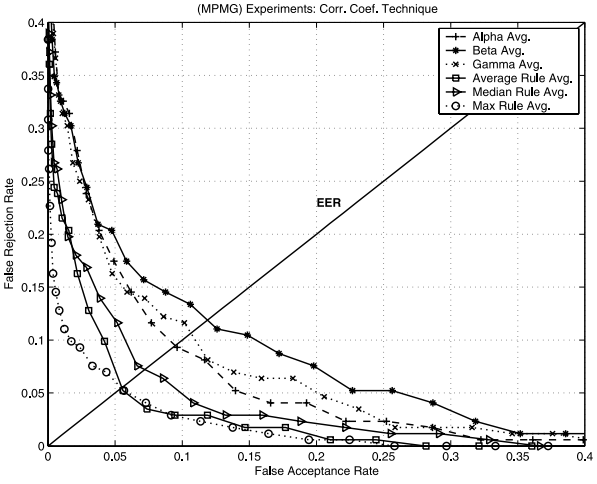
8. Conclusions

The results of an exhaustive set of identification and verification experiments were presented. The key research findings are listed as the following:

- In each of our experiments, the fusion rules outperformed each of the single finger types. As the time lapse between gallery and probe increased, this performance difference was more apparent. During identification experiments involving data



(A) ROC curves.



(B) Equal error rates.

Fig. 16. Verification performance: group M experiments.

possessing a single week time lapse between gallery and probe acquisitions, the average fusion rule achieved the best performance. When the time lapse was increased to 16 weeks, the maximum fusion rule performed best indicating that the match scores decrease as the time between gallery and probe images increases.
- There as no single best performing finger type. Different finger types performed better during particular experiment types (identification, verification) and configurations (single/multiple probe/gallery).
- Our results suggest that the matching performance is relatively stable over time. The degradation in performance when utilizing gallery and probe data possessing a large time lapse was small.
- Finger surface achieves performance similar to that of a much more researched biometric identifier, 2D face. When using face images collected from the same subjects and the same experiment configuration there was a 1% difference in the performance obtained when using finger surface as there was when using facial feature as a biometric identifier. This is encouraging considering the fact that face recognition technology is well explored and finger surface shape, by contrast, is a new technique.

9. Future work

One result of the research effort is that a number of areas for future work have been generated. These areas can be categorized as data collection, feature template refinement, matching techniques, and biometric fusion.

9.1. Data collection

The lack of a specialized apparatus to collect range images of the hand played a part in limiting the size of our database. The range sensor obtained range images relatively quickly but many more could be obtained with a faster method of range image generation. One method under consideration would involve using stereopsis to generate range data. This approach would use two digital cameras to collect an intensity image of the hand simultaneously and use the two images to reconstruct the 3D co-ordinates of the scene. The use of intensity images would not only result in faster image collection but could allow for much higher range image resolution than the 640×480 images used in our experiments. The higher resolution range data resulting from this approach could yield more biometric features for use in the feature templates and perhaps add more distinctiveness to the data collected from from subjects. The development of this application specific sensor would be less expensive than the sensor currently employed. Because 3D data collection would rely on stereopsis, a passive technique of measuring range, the effects of variance in illumination on recognition performance would require study. The reduction in time required to collect hand images would help to facilitate the creation of a much larger hand image database. A larger database of hand images is required to help answer the question

of whether a biometric system based on finger surface is scalable. A larger database of images which possesses longer time lapse in between collections could also allow for the investigation of how changes in finger shape due to swelling and subject weight loss/gain could effect recognition performance.

9.2. Feature template refinement

Many 2D face recognition algorithms are based upon principal component analysis (PCA) [11]. This statistical technique can be used to reduce the number of features considered during comparison. Only the areas of the finger surface exhibiting the greatest amount of variance among subjects, would be considered during template comparisons. The matching techniques employed in our research utilized on average 18,500 pixels or features from each finger during matching score calculation. A reduction in the number of features used could mean a significant decrease in time required for matching score calculation as well as allow for the use of more sophisticated matching algorithms. A reduction in the number of features would also result in a smaller storage requirement for templates.

9.3. Matching techniques

We achieved the best verification rates when using score-level fusion rules. The use of these rules allow for more than one finger surface to be considered in the calculation of the overall match score. However, much of the distinctiveness of the combination of finger surfaces was lost when the mean, median, or maximum score is chosen as the overall score. Employing fusion at the feature extraction level as described by Jain and Ross [21] could allow more distinctiveness to be retained. Therefore, the effectiveness of this approach for our application should investigated.

As an alternative to the matching techniques used in our work, the development of a PCA based matching technique which uses "eigenfingers" could be used. In addition to the benefits related to feature template refinement mentioned in the previous section, a possible recognition performance improvement could be obtained using this technique.

9.4. Biometric fusion

Finger surface alone has performed well as biometric identifier, but other research has demonstrated that a biometric system which utilizes multiple modalities can achieve better performance [1]. Because our data collection efforts involve obtaining both range and intensity images of the hand from a single sensor, it would be logical that perhaps the information obtained from these two modalities should be combined into a feature template. Finger surface shape information would be extracted from the range images and information such as finger measurements, color, texture, and crease patterns could be extracted from intensity images. The combination of these characteristics could result in higher identification and verification rates. Additionally, work involving the combining of finger surface data with other biometric

identifiers such as 2D/3D face data, 3D ear data, fingerprints, and iris patterns could result in higher verification rates.

Acknowledgments

The research presented in this paper was supported by the Defense Advanced Research projects Agency and the Office of Naval Research under Grant N00014-02-1-0410, and by the National Science Foundation under grant EIA-0130839. Damon L. Woodard was supported by a GEM fellowship and is currently supported by a DCI Postdoctoral Fellowship.

References

[1] K.I. Chang, P.J. Flynn, X. Chen, K.W. Bowyer, Multi-biometrics using facial appearance, shape and temperature, in: FGR, 2004, pp. 43–48.
[2] K. Chang, D.L. Woodard, P.J. Flynn, K.W. Bowyer, Three dimensional face and finger biometrics, in: 12th European Signal Processing Conference (EUSIPCO), Vienna, Austria, Sept. 2004.
[3] C. Dorai, A.K. Jain, COSMOS-A representation scheme for free-form surfaces, in: Internat. Conf. on Computer Vision, June 1995, pp. 1024–1029.
[4] C. Dorai, A.K. Jain, COSMOS-A representation scheme for 3D free-form objects, IEEE Trans. Pattern Anal. Mach. Intell. 19 (10) (1997) 1115–1130.
[5] R.H. Ernst, Hand ID System. U.S. Patent No. 3576537, 1971.
[6] P.J. Flynn, A.K. Jain, On reliable curvature estimation, in: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, vol. 89, 1989, pp. 110–116.
[7] L. Hong, A.K. Jain, S. Pankanti, Can Multibiometrics Improve Performance. Technical Report MSU-CSE-99-39, Department of Computer Science, Michigan State University, East Lansing, Michigan, 1999.
[8] I.H. Jacoby, A.J. Giordano, W.H. Fioretti, Personal Identification Apparatus. U.S. Patent No. 3648240, 1972.
[9] A.K. Jain, N. Duta, Deformable matching of hand shapes for verification, in: Proc. Internat. Conf. on Image Processing, Oct. 1999, pp. 857–861.
[10] A.K. Jain, A. Ross, S. Pankanti, A prototype hand geometry-based verification system, in: Proc. of 2nd Internat. Conf. on Audio- and Video-based Biometric Person Authentication (AVBPA), Washington DC, Mar. 1999, pp. 166–171.
[11] I.T. Jolliffe, Principal Component Analysis, Springer, Berlin, 1986.
[12] J. Kittler, M. Hatel, R.P.W. Duin, J. Matasm, On combining classifiers, IEEE Trans. Pattern Anal. Mach. Intell. 20 (3) (1998) 226–239.
[13] J.J. Koenderink, A.J. van Doorn, Surface shape and curvature scales, Image Vision Comput. 10 (8) (1992) 557–564.
[14] Konica Minolta Website. http://konicaminolta.com.
[15] J. Kovač, P. Peer, F. Solina, Human skin colour clustering for face detection, in: B. Zajc (Ed.) EUROCON 2003—Internat. Conf. on Computer as a Tool, Ljubljana, Slovenia, Sept. 2003.
[16] Y.L. Lay, Hand shape recognition, Opt. Laser Technol. 32 (1) (2000) 1–5.
[17] Mathworks Website. http://www.mathworks.com.
[18] R.P. Miller, Finger Dimension Comparison Identification System. U.S. Patent No. 3576538, 1971.
[19] J. Min, P.J. Flynn, K.W. Bowyer, Using Multiple Gallery and Probe Images per Person to Improve Performance of Face Recognition. Technical Report TR03-07, Department of Computer Science and Engineering, The University of Notre Dame, Notre Dame, Indiana, 2003.

[20] P.J. Phillips, H. Moon, S.A. Rizvi, P.J. Rauss, The FERET evaluation methodology for face-recognition algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 22 (10) (2000) 1090–1104.
[21] A. Ross, A.K. Jain, Information fusion in biometrics, Pattern Recogn. Lett. 24 (13) (2003) 2115–2125.
[22] R. Sanchez-Reillo, C. Sanchez-Avila, A. Gonzalez-Marcos, Biometric identification through hand geometry measurements, IEEE Trans. Pattern Anal. Mach. Intell. 22 (10) (2000) 1168–1171.
[23] D. Sidlauskas, 3D Hand Profile Identification Apparatus. U.S. Patent No. 4736203, 1988.
[24] Sundial Time Systems Website. http://www.identix.com/.
[25] D.L. Woodard, Exploiting Finger Surface as a Biometric Identifier. PhD thesis, The University of Notre Dame, Notre Dame, IN 46556, USA, 2004.
[26] D.L. Woodard, P.J. Flynn, 3D finger biometrics, in: D. Maltoni, A.K. Jain (Ed.), 2004 Eur. Conf. on Computer Vision, Lecture Notes in Computer Science, Prague, Czech Republic, 2004, pp. 238–247.
[27] R.L. Zunkel, Hand geometry based verification, in: A.K. Jain, R. Bolle, S. Pankanti (Eds.), BIOMETRICS Personal Identification in a Networked Society, Kluwer Academic Publishers, 1998, pp. 87–101.