A Comparative Study of Hand Recognition Systems

Gholamreza Amayeh[1], George Bebis[1,2] and Muhammad Hussain[2]
[1] Computer Science and Engineering Department, University of Nevada, Reno, NV 89557
[2] Computer Science Department, King Saud University, Riyadh, Saudi Arabia 11543
(amayeh,bebis)@cse.unr.edu & mhussain@ccis.edu.sa

***Abstract*— Hand-based recognition represents a key biometric technology with a wide range of potential applications both in industry and government. By far, many different hand-based recognition algorithms have been developed. This paper presents a comparative study to evaluate the performance of three state of the art hand-based recognition methods. Using the *University of Nevada at Reno* (UNR) and the *University of Notre Dame* (UND) hand databases, we compare a geometric-based method, a component-based approach using Zernike moments, and an algorithm employing 3D finger surface features. Both recognition and authentication experiments have been conducted to investigate the performance and robustness of the three methods. Our experimental results show that Zernike descriptors yield features that are more robust and accurate compared to hand geometric features and 3D finger surface features.**

Key words: Biometrics, hand, performance evaluation.

I. Introduction

Hand-based biometrics is among the oldest live biometrics-based recognition modalities. The existence of several hand-based authentication commercial systems and patents indicate the effectiveness of this type of biometric. With increasing interest in hand-based recognition, researchers have proposed a variety of approaches [4], [5], [15], [16]. The majority of hand-based biometric systems employ geometric measurements since the geometry of the hand contains relatively invariant features of an individual. Sanchez-Reillo et al. in [12], [11] introduced a richer set of geometric features which can be divided into four different categories: *Width, Height, Deviation and Angles*. In their system, users were asked to place their hand on a flat surface and align it, with the help of some guidance pegs. The alignment operation simplifies feature extraction and allows for high processing speeds.

Removal of pegs, to improve convenience, and use of more powerful feature extraction techniques to capture the shape of the hand more accurately has attracted much attention. More recent studies have concentrated on the design of peg-free systems [10], [8], [3], [2]. Amayeh et al. [2] proposed a new component-based approach to hand-based recognition and authentication which improves both accuracy and robustness as well as the ease of use due to avoiding pegs. Their approach accounts for hand and finger motion by decomposing the hand silhouette in different regions corresponding to the back of the palm and the fingers. They have implicitly used high-order Zernike moments, instead of hand geometric features, to represent the shape of each part of the hand.

Recently, researchers have focused on the use of three dimensional data as a source of distinguishing features for personal recognition and authentication. For example, Woodard and Flynn [13], [14] proposed using the surface of the index, middle, and ring fingers for biometric identification. A local surface descriptor, the shape index, is used to represent finger surfaces and matching scores are calculated using the normalized correlation coefficient.

Since there is no standard hand acquisition method and no benchmark hand database, most of the studies in the literature have only reported qualitative comparisons with existing methods. In this paper, first we identify three state of the art hand-based recognition methods, and then carry out a comparative study to evaluate the performance and robustness of these methods. Our intention is to reveal state of the art performance and identify the limitations of these methods. Both identification and verification experiments were conducted to determine which of these methods provides the most accurate and robust features for establishing identity.

The rest of the paper is organized as follows. Section 2 presents the details of the data sets used in this study. Section 3 describes the implementation of three methods. Section 4 provides our experimental results and discusses the performance and limitations of each method. Finally, Section 5 provides our conclusions.

II. Data Sets

The UND and UNR Hand Databases were used to carry out a comprehensive comparison on the performance of three state of the art hand recognition methods. Compared with UNR Hand Database, which includes binarized 2D silhouettes of the hand images, the UND Hand Database contains intensity and range images of the hand. In the following subsections, we describe details of each data collection.

A. UNR Data Set

A single CCD camera and a lighting table, which yields an almost binary 2D silhouette of the hand, was used to collect the data set. The acquired images are 640×480 pixels color photographs. The actual effective region for the hand placement is about 225×225 in the center of each image. The hand images were collected from 101 people of various age, sex and ethnicity. For each subject, 10 images of his/her right hand were collected during the same session. Therefore the database has a total of 1010 hand images.

Figure 1(a) shows a sample image in this database. Since the hand images in this database are almost free of shadows and noise, binarization can be performed using a fix threshold ($Th = 128$) for all the images, as mentioned in [2].

B. UND Data Set

A Minolta Vivid 910 range scanner was used to collect the data set [1]. This sensor captures both a 640×480 range image and a 640×480 color intensity image nearly simultaneously. The 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 the data was collected from adults between the ages of 18 and 24. Data collection was performed in 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, the database has a total of 1191 hand intensity and range images. Figures 1(b) and 1(c) show a sample intensity and range image from this database.

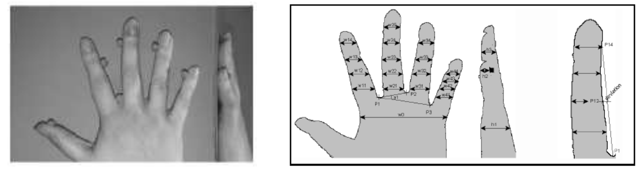
III. Hand-based Recognition Algorithms

In this study, we compare a geometric-based method [12], a component-based approach using high order Zernike moments [2], and the method reported in [13] which utilizes 3D finger surface features.

A. Geometric-Based Method

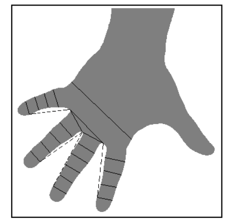
The majority of existing systems employ hand geometric features for recognition or authentication. It has been reported in the literature that these features work well and can be computed efficiency. The geometric features used in our experiments is a subset of the features proposed by Sanchez-Reillo et al. in [12]. Figure 2(a) shows a sample image taken by the image acquisition system in [12]. A total of 31 features were used (see Figure 2(b)): width of four fingers and palm in different locations (18 features), height of middle and little fingers and palm (3 features), distances between the three inter-finger points (3 features) and angles between the inter-finger points and horizontal line (3 features), distances between a middle point of the finger and the middle point of the straight line between the inter-finger point and the last height where the finger width is measured (4 features). However, the image acquisition system in [12] uses a mirror to capture a side view of the hand in addition to a top view of the hand as shown in Figure 2(a). Since the 2D images in the UNR and UND databases have a top view of the hand only, we cannot extract the height of the little and middle fingers as well as the palm (3 features). Therefore, we have used only 28 features in our experimental comparisons. Figure 3 shows the main distances measured from a binarized hand image in the UNR database.

Systems employing pegs to fix the position of the hand, such as [12], use predestined axes to facilitate feature extraction. In the case of peg-free systems, several landmarks on



(a) (b)

Fig. 2. (a) A sample hand image taken using the image acquisition system in [12], (b) Location of measurement points for feature extraction in [12].



(a)

Fig. 3. The main measured distances according [12] on a sample hand image from the UNR database.

the hand, such as fingertips and valleys, must be extracted in order to define the same or similar axes [15], [8], [16]. Here, we compute the curvature of the hand boundary to extract the fingertip and valley locations by detecting curvature minima and maxima.

To account for noise, curvature is typically smoothed using a Gaussian function [9]. Before computing the curvature, the hand boundary is re-sampled at 2^m, equal-distant, points [9]. Figure 4(b) shows the curvature of the hand contour shown in Figure 4(a). Choosing the value of σ is critical to ensure both good detection and localization. In general, smaller σ values lead to better localization, however, noise could give rise to false positives. On the other hand, larger σ values reduce false positives but good localization is difficult. To address that issue in this study, multi-resolution schemes have been employed (i.e., curvature scale-space [9]), however, time requirements are higher.

B. Component-Based Approach using Zernike Moments

Amayeh et. al. [2] proposed a component-based approach to hand-based authentication which improves both accuracy and robustness as well as ease of use due to avoiding pegs. There are two key ideas behind this approach. First, decomposing the hand image in different regions corresponding to the palm and fingers. Second, fusing information from different parts of the hand to improve accuracy and robustness. An important characteristic of this approach is that it does not require the extraction of any landmark points. Figure 5 shows the main stages of this approach.

The preprocessing stage includes the hand-forearm segmentation and the palm-fingers segmentation [2]. To separate the forearm from the hand, first the palm is detected by finding the largest circle that can be prescribed inside the hand-arm silhouette. Then the forearm is segmented by detecting its intersection with the circle and the boundary of the image. Figure 6(a) shows the resulting silhouette after

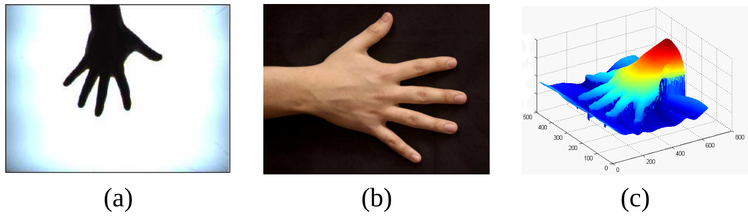


Fig. 1. (a) An image sample from the UNR database, (b) an intensity image sample from the UND database, and (c) a range image sample from the UND database.

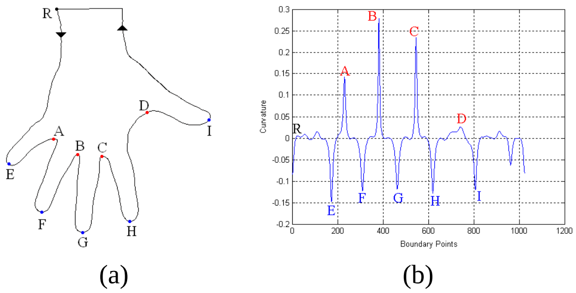


Fig. 4. (a) A sample hand contour from UNR database, (b) the curvature of the hand contour. R is a reference point which is used for identifying the landmarks. The hand boundary has been re-sampled at 1024, equal-distant, points.

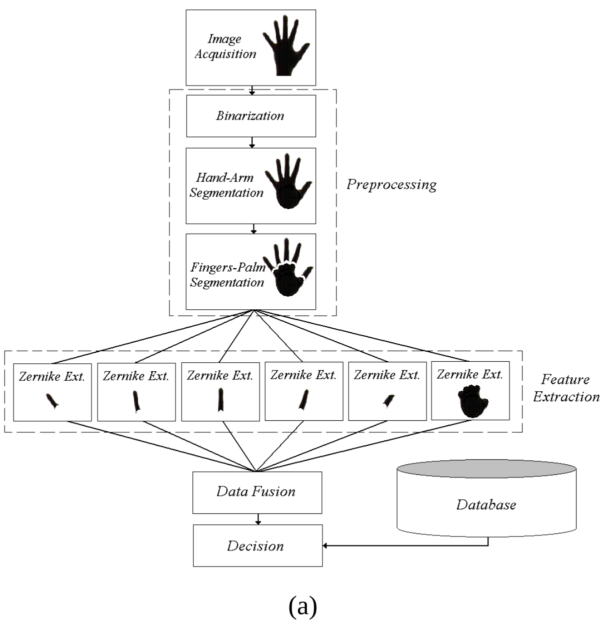


Fig. 5. Main stages of the system in [2].

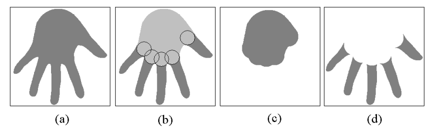


Fig. 6. (a) Hand silhouette after discarding forearm, (b) morphological closing using a circular structure element, (c) the result of closing and (d) the result of subtracting the palm from the hand silhouette.

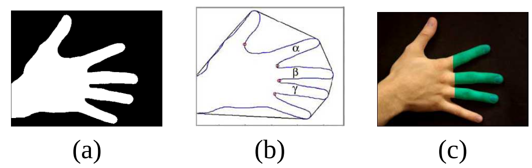


Fig. 7. (a) Segmented intensity hand image, (b) finger valley location, and (c) extracted finger pixels.

discarding the forearm region. To segment the fingers from the palm, first the fingers are filtered out using morphological closing; then, the palm is subtracted from the hand silhouette to segment the fingers [2]. The processing steps of the finger segmentation module are shown in Figure 6.

Feature extraction is performed by computing the Zernike moments of each part of the hand independently [2]. A crucial parameter here is determining the maximum Zernike moment order to represent the geometry of different parts of the hand. In [2] the maximum order is set to 20 (121 features) for each finger and to 30 (256 features) for the palm by analyzing the reconstruction error.

After extracting Zernike moments, information from different parts of the hand is fused to improve authentication accuracy and robustness. Various fusion techniques at different levels (i.e. feature level, score level and decision level) are investigated in [2]. Using UNR database the best result has been reported for authentication using majority voting rule at decision level.

C. 3D Finger Surface Method

Woodard and Flynn in [13] presented a novel approach for personal identification which utilizes finger surface features as a biometric identifier. This approach includes three main steps: *hand segmentation, finger extraction and template generation*. In order to work with only the range image pixels lying on the surface of the hand, the task of hand segmentation is required. To simplify this task in [13], the intensity image of the hand is used instead of its range image, since there is a pixel to pixel correspondence between intensity and range images. Therefore a combination of edge and skin detection techniques are employed to the intensity image to reliably segment the hand from the image, as shown in Figure 7(a), thereby allowing for segmentation in the range image.

After obtaining the hand silhouette from the intensity hand image, the convex hull of the contour of the hand silhouette is used to locate the valleys between the fingers represented as circles in Figure 7(b). The valley positions are used for segmenting the index, middle, and ring fingers [13]. The shaded areas in Figure 7(c) represent the extracted finger pixels. To address finger pose variations, each finger mask along with its corresponding range pixels is rotated and centered in a 80×240 output finger range image in which the major axis of the finger mask is coincident with the horizontal axis.

For each valid pixel of the finger mask in the output image, a surface curvature estimate is computed with the corresponding range data. The principal curvature values, k_{min} and k_{max}, are calculated for each finger surface point $\mathbf{p}$ [13].

It was suggested in [13] that the range data be smoothed prior to curvature estimation in order to limit the effects of noise. The computed principal curvature values are then used to compute the Shape Index, *SI*, value at each pixel, given by the following formula [13]:

$$SI = \begin{cases} \frac{1}{2} - \frac{1}{\pi} arctan\left(\frac{k_{max}+k_{min}}{k_{max}-k_{min}}\right) & k_{max} \geq k_{min} \\ not\ valid & k_{max} < k_{min} \end{cases} \tag{1}$$

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. The match score is the sample normalized correlation coefficient. In [13], different fusion technique such as average, median and maximum fusion rules employed to combine information of index, middle and ring fingers. Using the UND database, the best results have been reported for authentication and verification using the maximum and average rules at score level.

IV. Experimental Results and Discussion

In this section, we discuss and analyze our experimental results based on three state of the art hand recognition methods using the UNR and UND databases. First, using the UNR database, we compare the geometric-based method with the component-based approach using Zernike descriptors. Then, using the UND database, we illustrate the effectiveness of range data, as a source of distinguishing features for personal identification, by comparing the last method with the other two methods that make use of features extracted from intensity images.

A. Verification/Identification Using UNR Database

Since the UNR database contains 2D hand images only, we compare the geometric-based method with the component-based approach using Zernike descriptors. In these experiments, five templates per person were employed for enrollment while the rest of them were used for testing. We report average performance by repeating each experiment 30 times. The minimum-distance classifier was used to compute the matching distance between the query hand and an enrolled individual. Note that all the images were collected in the same session.

As it was mentioned in section III-A, a total of 28 geometric features are extracted from the hand. Before computing

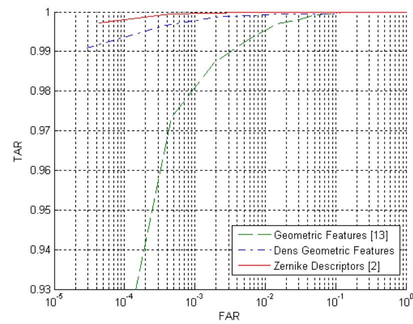


Fig. 8. Verification results using the method of [2] based on Zernike descriptors and the method of [12] based on geometric features; In the first method, majority voting was employed to fuse different components of the hand.

the curvature of the hand, all hand boundaries were resampled at 1024, equal-distant, points. Due to the high quality of the images, we have found that a σ value equal to 20 yields good detection and localization results.

Using the method summarized in section III-B, a total of 861 Zernike features were extracted. In our experiments, fusion of different parts of the hand was performed using the weighted sum rule for identification and the majority voting rule for verification [2].

Figures 8 and 9 show that the performance of the method reported in [2] (red curve) based on Zernike descriptors is much higher than the method reported in [12] (green curve) based on geometric features, both in verification and identification. One reason is the more accurate representation of hand geometry using Zernike features [2] than geometric features [12]. To investigate the effect of the number of features, we increased the number of geometric features in [12] by measuring the width of the little, ring, middle and point fingers in more places. Specifically, the resolution of the hand images allowed us to measure the width of each finger in 25 different places with no overlapping. Therefore, we increased the total number of geometric features from 28 to 111. As Figures 8 and 9 illustrate (i.e., blue curve), more features improve performance significantly such that, in the case of identification, geometric features are comparable to Zernike features. Therefore, one can conclude that the number of features effects system performance. The main advantage of using geometric features compared to Zernike descriptors is low computational cost; the drawback is the limitation in the number of features that can be extracted due to the resolution of the hand images. For example, if the width of fingers is measured at more than 25 places, in some cases, the measurements will overlap which means that no new information can be introduced in the feature set. However, in the case of Zernike descriptors, despite of the high computation cost, there is no limitation in the number of features extracted and, in theory, can be calculated to an arbitrary order.

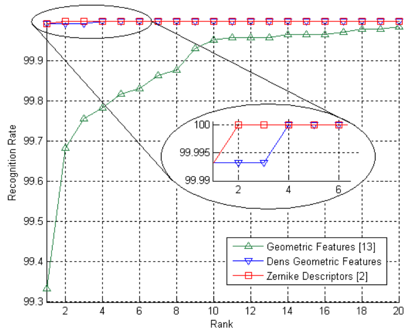


Fig. 9. Identification results using the method of [2] based on Zernike descriptors and the method of [12] based on geometric features; In the first method, a weighted sum was employed to fuse different components of the hand.

B. Verification/Identification Using UND Database

Additional experiments were performed using the UND database. To extract the hand silhouette, we used the same algorithm described in [13]. The RGB color space skin detection rules specified in [7] along with an implementation of a Canny edge detector comprise the hand segmentation module. For consistency reasons, we adopted the same setup as in [13] to form the gallery (i.e., enrollment) and probe (i.e., test) image sets. That is, gallery images were chosen to be images collected prior to those chosen as probe images [13]. Following this rule, only images collected during the second week could serve both as probe and gallery images. For each time lapse, we performed two experiments by switching the enrollment samples with the test samples. Since the number of samples was not equal in all sessions (e.g., two samples per person in the first session and three samples per person in the second and third sessions), we report average performance for each time lapse.

To extract geometric features, each hand boundary was re-sampled at 2048, equal-distant, points and the value of σ was set to 50. The geometric features were extracted based on the method of [12] (magnet curve); we have also experimented with dense geometric features (green curve) as described in the previous subsection.

Since lighting is not uniform in all images, some areas of the palm have low contrast. As a result, the hand silhouette was affected many times and we were not able to segment the palm from forearm, as described in [2], satisfactorily. Therefore, we decided to use only information from the fingers in the component-based approach using Zernike descriptors [2]. As before, fusion in verification was performed using the majority voting rule and in identification using the weighted-sum rule. In the method of [13], fusion in verification and identification were performed using the average rule.

Figures 10 and 11 show the identification results obtained for one and 16 week time lapse respectively. As Figures 10 and 11 illustrate , the performance of geometric features drops significantly over the 16 week time lapse. Due to illumination and environment changes in the samples over period of time, the geometrical measurements of the hand were affected. As a result, the performance of system (magnet vs green curves) dropped off significantly.

Figures 10 and 11 illustrate that Zernike features (red curves) have better performance over large period of time (16 week time lapse). The reason is that Zernike descriptors are quite robust to noise [6]. Moreover Zernike features can represent the geometry of the hand more accurately and without redundancy (i.e., in theory, Zernike features are not redundant due to orthogonality of the Zernike basis functions [6]). So, the information in a particular order can not be found in other orders. Interestingly enough, the recognition rate based on 16 week time lapse shown in Figure 11 is higher than the 1 week time lapse shown in Figure 10; however, this is probably due to the unequal size of the data sets.

As it can be observed in Figures 10 and 11, the recognition rate using Zernike features is higher than the recognition rate using finger surface features. The reason is that 2D hand silhouette (obtained from intensity images) is more robust than hand surface (obtained from range data). Comparing the identification results using dense geometric features and finger surface features (i.e., Figure 10) supports the claim that 2D hand silhouette is more robust than hand surface. A question that arises from the results is why the recognition rate using dense geometric features drops significantly compared to finger surface features (i.e., Figure 11). By carefully examining the UND data obtained during the third session, we noted that some of the subjects had removed/put long fingernails compared to the first two sessions. Obviously, this affects the geometric measurements (i.e. length of fingers). Moreover, variations in the contrast of intensity images affects the extraction of the hand boundary; as a result, errors are introduced in the computation of the geometric measurements (i.e. width of fingers). Note that Zernike descriptors operate on regions, so errors on extracting the hand boundary do not affect them significantly.

The verification results are shown in Figure 12 for 16 weeks time lapse. Table I reports the Equal Error Rate (EER) and True Acceptance Rate (TAR) for each method when the False Acceptance Rate (FAR) is equal to 5% assuming a 16 week time lapse as reported in [13]. Again, as it can be observed, system performance using Zernike descriptors is superior to using geometric and finger surface features. Therefore, Zernike descriptors seem to be more powerful compared to geometric and finger surface features.

TABLE I

TIME LAPSE VERIFICATION PERFORMANCE COMPARISON AMONG DIFFERENT METHODS USING THE UND DATABASE.

Time lapse	16 Week		
Method	Geometric-based [12]	Zernike-based [2]	Finger Surface-based [13]
EER (%)	10.2	1.72	5.5
TAR(FAR=5%)	76.7	99.3	94.0

Finally, comparing the identification results using geometric features (i.e., Figures 9 (blue and green curves) and 10 (magnet and green curves)) shows that the robustness of

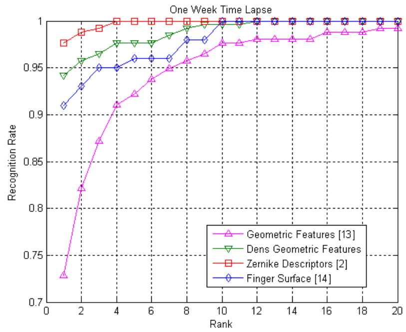


Fig. 10. Identification results of geometric-based [12], Zernike-based [2] and 3D finger surface [13] methods; Fusion was performed by weighted sum in the method of [2] and average rule in the method of [13].

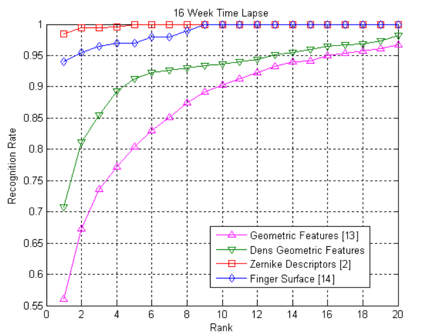


Fig. 11. Identification results of geometric-based [12], Zernike-based [2] and 3D finger surface [13] methods; Fusion was performed by weighted sum in the method of [2] and average rule in the method of [13].

hand silhouette can have an important impact on system performance. On the other hand, robust image acquisition is vital in a hand biometric system.

V. Conclusion

In this paper, we carried out a comparative study to evaluate the performance and robustness of three state of the art hand-based recognition methods. Using the UNR and UND hand databases, we compared, discussed and analyzed a geometric-based method, a component-based approach using Zernike descriptors and a method using 3D finger surface. Our experimental results showed that Zernike descriptors are superior to geometric and 3D finger surface features both in for verification and identification. The main disadvantage of Zernike features is their high computation cost.

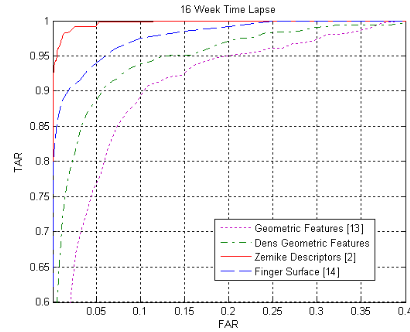


Fig. 12. Verification results of geometric-based [12], Zernike-based [2] and 3D finger surface [13] methods; Fusion was performed by majority voting in the method of [2] and average rule in the method of [13].

References

[1] . *Konica Minolta Website.* http://konicaminolta.com.

[2] G. Amayeh, G. Bebis, A. Erol, and M. Nicolescu. A component-based approach to hand verification. *IEEE Workshop on Biometrics,(in conjunction with CVPR 2007)*, Minneapolis, MN, June 18, 2007.

[3] G. Amayeh, G. Bebis, A. Erol, and M. Nicolescu. Peg-free hand shape verification using high order zernike moments. *IEEE Workshop on Multi-modal Biometrics, (in conjunction with CVPR 2006)*, New York City, NY, June 17-18, 2006.

[4] A. K. Jain, R. Bolle, and S. Pankanti. *Biometrics: personal identification in networked society*. Kluwer Academic Publishers, 1999.

[5] A. K. Jain, A. Ross, and S. Pankanti. A prototype hand geometry-based verification system. *Proc. 2nd Int. Conf. on Audio- and video-based personal authentication (AVBPA), Washington, USA*, pages 166–171, March 1999.

[6] A. Khotanzad and Y. H. Hong. Invariant image recognition by zernike moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:489 – 498, 1990.

[7] J. Kovac, P. Peer, and F. Solina. Human skin colour clustering for face detection. *International Conference on Computer as a Tool*, Sept. 2003.

[8] Y. L. Ma, F. Pollick, and W. Hewitt. Using b-spline curves for hand recognition. *Proc.of the 17th International Conference on Pattern Recognition (ICPR'04)*, 3:274–277, Aug. 2004.

[9] F. Mokhtarian and A. Mackworth. A theory of multiscale curvature-based shape representation for planar curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(8):789–805, 1992.

[10] S. Ribaric, D. Ribaric, and N. Pavesic. Multimodal biometric user-identification system for network-based applications. *IEE Proceedings on Vision, Image and Signal Processing*, 150, Issue 6:409–416, 15 Dec. 2003.

[11] R. Sanchez-Reillo. Hand geometry pattern recognition through gaussian mixture modelling. *15th International Conference on Pattern Recognition (ICPR'00)*, 2:937–940, 2000.

[12] R. Sanchez-Reillo, C. S. Avila, and A. G. Marcos. Biometric identification through hand geometry measurements. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1168–1171, October 2000.

[13] D. L. Woodard and P. J. Flynn. Finger surface as a biometric identifier. *Computer Vision and Image Understanding*, 100(3):357–384, Des. 2005.

[14] D. L. Woodard and P. J. Flynn. Personal identification utilizing finger surface features. *IEEE Conf. on Computer Vision and Pattern Recognition*, San Diego, CA, 2005.

[15] W. Xiong, C. Xu, and S. H. Ong. Peg-free human hand shape analysis and recognition. *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, 2:77–80, March 18-23 2005.

[16] E. Yoruk, E. Konukoglu, B. Sankur, and J. Darbon. Shape-based hand recognition. *IEEE Transactions on Image Processing*, 15(7):1803–1815, July 2006.