Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Pattern Recognition 38 (2005) 2270–2285
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/patcog

Score normalization in multimodal biometric systems☆

Anil Jain[a], Karthik Nandakumar[a], Arun Ross[b,*]

[a]Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824, USA
[b]Department of Computer Science and Engineering, West Virginia University, Morgantown, WV 26506, USA

Received 23 December 2003; received in revised form 18 January 2005; accepted 18 January 2005

Abstract

Multimodal biometric systems consolidate the evidence presented by multiple biometric sources and typically provide better recognition performance compared to systems based on a single biometric modality. Although information fusion in a multimodal system can be performed at various levels, integration at the matching score level is the most common approach due to the ease in accessing and combining the scores generated by different matchers. Since the matching scores output by the various modalities are heterogeneous, score normalization is needed to transform these scores into a common domain, prior to combining them. In this paper, we have studied the performance of different normalization techniques and fusion rules in the context of a multimodal biometric system based on the face, fingerprint and hand-geometry traits of a user. Experiments conducted on a database of 100 users indicate that the application of min–max, *z*-score, and tanh normalization schemes followed by a simple sum of scores fusion method results in better recognition performance compared to other methods. However, experiments also reveal that the min–max and *z*-score normalization techniques are sensitive to outliers in the data, highlighting the need for a robust and efficient normalization procedure like the tanh normalization. It was also observed that multimodal systems utilizing user-specific weights perform better compared to systems that assign the same set of weights to the multiple biometric traits of all users.

Keywords: Biometrics; Multibiometrics; Score normalization; Face; Fingerprint; Hand-geometry; Tanh; Min–max; *z*-score; Sigmoid; Parzen; User-specific weights

1. Introduction

Biometric systems make use of the physiological and/or behavioral traits of individuals, for recognition purposes [1]. These traits include fingerprints, hand-geometry, face, voice, iris, retina, gait, signature, palm-print, ear, etc. Biometric systems that use a single trait for recognition (i.e., unimodal biometric systems) are often affected by several practical problems like noisy sensor data, non-universality and/or lack of distinctiveness of the biometric trait, unacceptable error rates, and spoof attacks [2]. Multimodal biometric systems overcome some of these problems by consolidating the evidence obtained from different sources [3]. These sources may be multiple sensors for the same biometric (e.g., optical and solid-state fingerprint sensors), multiple instances of the same biometric (e.g., fingerprints from different fingers of a person), multiple snapshots of the same biometric (e.g., four impressions of a user's right index finger), multiple representations and matching algorithms for the same biometric (e.g., multiple face matchers like PCA and LDA), or multiple biometric traits (e.g., face and fingerprint).

☆ This research was supported by the Center for Identification Technology Research (CITeR), a NSF/IUCRC program.

* Corresponding author. Tel.: +1 304 293 0405; fax: +1 304 293 8602.

E-mail addresses: jain@cse.msu.edu (A. Jain), nandakum@cse.msu.edu (K. Nandakumar), ross@csee.wvu.edu (A. Ross).

0031-3203/$30.00 © 2005 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2005.01.012

The use of multiple sensors addresses the problem of noisy sensor data, but all other potential problems associated with unimodal biometric systems remain. A recognition system that works on multiple instances of the same biometric can ensure the presence of a live user by asking the user to provide a random subset of biometric measurements (e.g., left index finger followed by right middle finger). Multiple snapshots of the same biometric, or multiple representations and matching algorithms for the same biometric may also be used to improve the recognition performance of the system. However, all these methods still suffer from many of the problems faced by unimodal systems. A multimodal biometric system based on different traits is expected to be more robust to noise, address the problem of non-universality, improve the matching accuracy, and provide reasonable protection against spoof attacks. Hence, the development of biometric systems based on multiple biometric traits has received considerable attention from researchers.

In a multimodal biometric system that uses different biometric traits, various levels of fusion are possible: fusion at the feature extraction level, matching score level or decision level (as explained in Section 2). It is difficult to consolidate information at the feature level because the feature sets used by different biometric modalities may either be inaccessible or incompatible. Fusion at the decision level is too rigid since only a limited amount of information is available at this level. Therefore, integration at the matching score level is generally preferred due to the ease in accessing and combining matching scores.

In the context of verification, fusion at the matching score level can be approached in two distinct ways. In the first approach the fusion is viewed as a classification problem, while in the second approach it is viewed as a combination problem. In the classification approach, a feature vector is constructed using the matching scores output by the individual matchers; this feature vector is then classified into one of two classes: "Accept" (genuine user) or "Reject" (impostor). In the combination approach, the individual matching scores are combined to generate a single scalar score which is then used to make the final decision. Both these approaches have been widely studied in the literature. Ross and Jain [4] have shown that the combination approach performs better than some classification methods like decision tree and linear discriminant analysis. However, no single classification or combination scheme works well under all circumstances. In this paper, we use the combination approach to fusion and address some of the issues involved in computing a single matching score given the scores of different modalities. Since the matching scores generated by the different modalities are heterogeneous, normalization is required to transform these scores into a common domain before combining them. While several normalization techniques have been proposed, there has been no detailed study of these techniques. In this work, we have systematically studied the effects of different normalization schemes on the performance of a multimodal biometric system based on the face, fingerprint and hand-geometry modalities.

The rest of the paper is organized as follows: Section 2 presents a brief overview of the various approaches used for information fusion in multimodal biometrics and motivates the need for score normalization prior to matching score fusion. Section 3 describes different techniques that can be used for the normalization of scores obtained from different matchers. The experimental results are presented in Section 4 and we have outlined our conclusions in Section 5.

2. Fusion in multimodal biometrics

A biometric system has four important modules. The sensor module acquires the biometric data from a user; the feature extraction module processes the acquired biometric data and extracts a feature set to represent it; the matching module compares the extracted feature set with the stored templates using a classifier or matching algorithm in order to generate matching scores; in the decision module the matching scores are used either to identify an enrolled user or verify a user's identity. Sanderson and Paliwal [5] have classified information fusion in biometric systems into two broad categories: pre-classification fusion and post-classification fusion (see Fig. 1).

Pre-classification fusion refers to combining information prior to the application of any classifier or matching algorithm. In post-classification fusion, the information is combined *after* the decisions of the classifiers have been obtained.

2.1. Pre-classification fusion

Prior to classification/matching, integration of information can take place either at the sensor level or at the feature level. The raw data from the sensors are combined in *sensor level fusion*. For example, the face images obtained from several cameras can be combined to form a single face image [6]. In sensor level fusion, the data obtained from the different sensors must be compatible, and this may not always be possible (e.g., it may not be possible to fuse face images obtained from cameras with different resolution). *Feature level fusion* refers to combining different feature vectors that are obtained by either using multiple sensors or employing multiple feature extraction algorithms on the same sensor data. When the feature vectors are homogeneous (e.g., multiple fingerprint impressions of a user's finger), a single resultant feature vector can be calculated as a weighted average of the individual feature vectors. When the feature vectors are non-homogeneous (e.g., feature vectors obtained using different feature extraction techniques, or feature vectors of different biometric modalities like face and hand geometry), we can concatenate them to form a single feature vector. Concate-

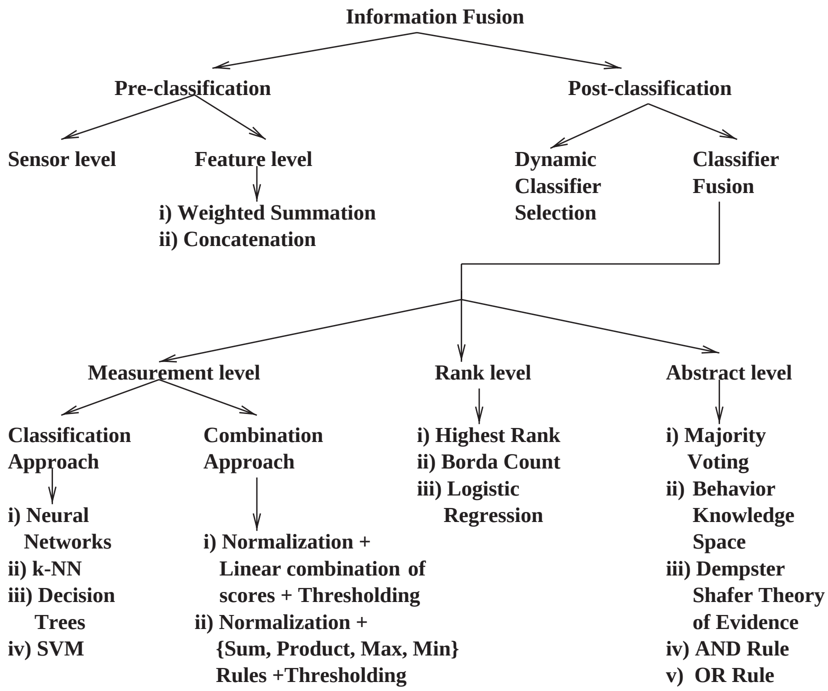

Fig. 1. Approaches to information fusion.

nation is not possible when the feature sets are incompatible (e.g., fingerprint minutiae and eigenface coefficients).

Biometric systems that integrate information at an early stage of processing are believed to be more effective than those systems which perform integration at a later stage. Since the features contain richer information about the input biometric data than the matching score or the output decision of a classifier/matcher, integration at the feature level should provide better recognition results than other levels of integration. However, integration at the feature level is difficult to achieve in practice because of the following reasons: (i) The relationship between the feature spaces of different biometric systems may not be known. In the case where the relationship is known in advance, care needs to be taken to discard those features that are highly correlated. This requires the application of feature selection algorithms prior to classification. (ii) Concatenating two feature vectors may result in a feature vector with very large dimensionality leading to the 'curse of dimensionality' problem [7]. Although, this is a general problem in most pattern recognition applications, it is more severe in biometric applications because of the time, effort and cost involved in collecting large amounts of biometric data. (iii) Most commercial biometric systems do not provide access to the feature vectors which they use in their products. Hence, very few researchers have studied integration at the feature level and most of them generally prefer post-classification fusion schemes.

2.2. Post-classification fusion

Schemes for integration of information after the classification/matcher stage can be divided into four categories: dynamic classifier selection, fusion at the abstract level, fusion at the rank level and fusion at the matching score level. A *dynamic classifier selection* scheme chooses the results of that classifier which is most likely to give the correct decision for the specific input pattern [8]. This is also known as the winner-take-all approach and the device that performs this selection is known as an associative switch [9].

Integration of information at the *abstract or decision level* can take place when each biometric matcher individually decides on the best match based on the input presented to it. Methods like majority voting [10], behavior knowledge space [11], weighted voting based on the Dempster–Shafer theory of evidence [12], AND rule and OR rule [13], etc. can be used to arrive at the final decision.

When the output of each biometric matcher is a subset of possible matches sorted in decreasing order of confidence, the fusion can be done at the *rank level*. Ho et al. [14] describe three methods to combine the ranks assigned by the different matchers. In the highest rank method, each possible match is assigned the highest (minimum) rank as computed by different matchers. Ties are broken randomly to arrive at a strict ranking order and the final decision is made based on the combined ranks. The Borda count method uses the sum of the ranks assigned by the individual matchers to calculate the combined ranks. The logistic regression method is a generalization of the Borda count method where the weighted sum of the individual ranks is calculated and the weights are determined by logistic regression.

When the biometric matchers output a set of possible matches along with the quality of each match (matching score), integration can be done at the *matching score level*. This is also known as fusion at the *measurement level* or *confidence level*. Next to the feature vectors, the matching scores output by the matchers contain the richest information about the input pattern. Also, it is relatively easy to access and combine the scores generated by the different matchers. Consequently, integration of information at the matching score level is the most common approach in multimodal biometric systems.

In the context of verification, there are two approaches for consolidating the scores obtained from different matchers. One approach is to formulate it as a classification problem, while the other approach is to treat it as a combination problem. In the classification approach, a feature vector is constructed using the matching scores output by the individual matchers; this feature vector is then classified into one of two classes: "Accept" (genuine user) or "Reject" (impostor). Generally, the classifier used for this purpose is capable of learning the decision boundary irrespective of how the feature vector is generated. Hence, the output scores of the different modalities can be non-homogeneous (distance or similarity metric, different numerical ranges, etc.) and no processing is required prior to feeding them into the classifier. In the combination approach, the individual matching scores are combined to generate a single scalar score which is then used to make the final decision. To ensure a meaningful combination of the scores from the different modalities, the scores must be first transformed to a common domain.

2.2.1. Classification approach to measurement level fusion

Several classifiers have been used to consolidate the matching scores and arrive at a decision. Wang et al. [15] consider the matching scores at the output of face and iris recognition modules as a two-dimensional feature vector, and use Fisher's discriminant analysis and a neural network classifier with radial basis function for classification. Verlinde and Chollet [16] combine the scores from two face recognition experts and one speaker recognition expert using three classifiers: *k*-NN classifier using vector quantization, decision-tree based classifier and a classifier based on a logistic regression model. Chatzis et al. [17] use fuzzy *k*-means and fuzzy vector quantization, along with a median radial basis function neural network classifier for the fusion of scores obtained from biometric systems based on visual (facial) and acoustic features. Sanderson and Paliwal [5] use a support vector machine classifier to combine the scores of face and speech experts. They show that the performance of such a classifier deteriorates under noisy input conditions. To overcome this problem, they implement structurally noise-resistant classifiers like a piece-wise linear classifier and a modified Bayesian classifier. Ross and Jain [4] use decision tree and linear discriminant classifiers for combining the scores of face, fingerprint, and hand-geometry modalities.

2.2.2. Combination approach to measurement level fusion

Kittler et al. [18] have developed a theoretical framework for consolidating the evidence obtained from multiple classifiers using schemes like the sum rule, product rule, max rule, min rule, median rule and majority voting. In order to employ these schemes, the matching scores are converted into posteriori probabilities conforming to a genuine user and an impostor. They consider the problem of classifying an input pattern Z into one of m possible classes (in a verification system, $m = 2$) based on the evidence provided by R different classifiers. Let $\vec{x}_i$ be the feature vector (derived from the input pattern Z) presented to the ith classifier. Let the outputs of the individual classifiers be $P(\omega_j|\vec{x}_i)$, i.e., the posterior probability of the pattern Z belonging to class ω_j given the feature vector $\vec{x}_i$. Let $c\varepsilon\{1, 2, \ldots, m\}$ be the class to which the input pattern Z is finally assigned. The following rules can be used to determine c:

Product Rule: This rule is based on the assumption of statistical independence of the representations $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_R$. The input pattern is assigned to class c such that

$$c = \operatorname*{argmax}_j \prod_{i=1}^{R} P(\omega_j|\vec{x}_i).$$

In general, different biometric traits of an individual (e.g., face, fingerprint and hand-geometry) are mutually independent. This allows us to make use of the product rule in a multimodal biometric system based on the independence assumption.

Sum Rule: Apart from the assumption of statistical independence of the multiple representations used in the product rule, the sum rule also assumes that the posteriori probabilities computed by the individual classifiers do not deviate much from the prior probabilities. This rule is applicable when there is a high level of noise leading to ambiguity in the classification problem. The sum rule assigns the input pattern to class c such that

$$c = \operatorname*{argmax}_j \sum_{i=1}^{R} P(\omega_j|\vec{x}_i).$$

Max Rule: The max rule approximates the mean of the posteriori probabilities by the maximum value. In this case, we assign the input pattern to class c such that

$$c = \operatorname*{argmax}_j \max_i P(\omega_j|\vec{x}_i).$$

Min Rule: The min rule is derived by bounding the product of posteriori probabilities. Here, the input pattern is assigned to class c such that

$$c = \underset{j}{\operatorname{argmax}} \min_{i} P(\omega_j|\vec{x}_i).$$

Prabhakar and Jain [19] argue that the assumption of statistical independence of the feature sets may not be true in a multimodal biometric system that uses different feature representations and different matching algorithms on the same biometric trait. They propose a scheme based on non-parametric density estimation for combining the scores obtained from four fingerprint matching algorithms and use the likelihood ratio test to make the final decision. They show that their scheme is optimal in the Neyman–Pearson decision sense, when sufficient training data is available to estimate the joint densities.

The use of Bayesian statistics in combining the scores of different biometric matchers was demonstrated by Bigun et al. [20]. They proposed a new algorithm for the fusion module of a multimodal biometric system that takes into account the estimated accuracy of the individual classifiers during the fusion process. They showed that their multimodal system using image and speech data provided better recognition results than the individual modalities.

The combined matching score can also be computed as a weighted sum of the matching scores of the individual matchers [4,15]. In this paper, we use the weighted sum of the scores of the individual matchers (assuming equal weights) and compare it to a threshold in order to decide if the user is a genuine user or an impostor. We refer to this method as the simple sum of scores method. We also use two other simple fusion methods called the max-score and min-score methods [21]. In the max-score method, the combined score is merely the maximum value among the scores of the individual matchers. The minimum value among the scores of the individual matchers is assigned as the combined score in min-score fusion.

Jain and Ross [22] have proposed the use of user-specific weights for computing the weighted sum of scores from the different modalities. The motivation behind this idea is that some biometric traits cannot be reliably obtained from a small segment of the population. For example, we cannot obtain good quality fingerprints from users with dry fingers. For such users, assigning a lower weight to the fingerprint score and a higher weight to the scores of the other modalities reduces their probability of being falsely rejected. This method requires learning of user-specific weights from the training scores available for each user. Consider a multimodal biometric system with three modalities. Let $w_{1,i}$, $w_{2,i}$, and $w_{3,i}$ be the weights assigned to the three modalities for user i. The process of learning these weights for user i can be described as follows:

(1) The weights $w_{1,i}$, $w_{2,i}$, and $w_{3,i}$ are varied over the range $[0, 1]$ in steps of 0.02, such that the constraint $w_{1,i} + w_{2,i} + w_{3,i} = 1$ is satisfied.
(2) The weighted sum of scores (s_w) for user i is computed as $s_w = w_{1,i}s_1 + w_{2,i}s_2 + w_{3,i}s_3$, where s_1, s_2, and s_3 are the scores provided by the three biometric matchers for user i.
(3) The set of weights that minimizes the total error rate (sum of the false accept and false reject rates) at some specified threshold (θ) is chosen. If more than one set of weights minimize the total error rate, then the set of weights that assigns almost equal weights to all the modalities is chosen. The threshold θ is set to a common value for all users. In [22], user-specific thresholds had been suggested.

3. Score normalization

Consider a multimodal biometric verification system that utilizes the combination approach to fusion at the match score level. The theoretical framework developed by Kittler et al. in [18] can be applied to this system only if the output of each modality is of the form $P(genuine|Z)$ i.e., the posteriori probability of user being "genuine" given the input biometric sample Z. In practice, most biometric systems output a matching score s, and Verlinde et al. [23] have proposed that the matching score s is related to $P(genuine|Z)$ as follows:

$$s = f\{P(genuine|Z)\} + \eta(Z), \tag{1}$$

where f is a monotonic function and η is the error made by the biometric system that depends on the input biometric sample Z. This error could be due to the noise introduced by the sensor during the acquisition of the biometric signal and the errors made by the feature extraction and matching processes. If we assume that η is zero, it is reasonable to approximate $P(genuine|Z)$ by $P(genuine|s)$. In this case, the problem reduces to computing $P(genuine|s)$ and this requires estimating the conditional densities $P(s|genuine)$ and $P(s|impostor)$. Snelick et al. [21] assumed a normal distribution for the conditional densities of the matching scores ($p(s|genuine)\tilde{}N(\mu_g, \sigma_g)$ and $p(s|impostor)\tilde{}N(\mu_i, \sigma_i)$), and used the training data to estimate the parameters μ_g, σ_g, μ_i, and σ_i. The posteriori probability of the score being that of a genuine user was then computed as,

$$P(genuine|s) = \frac{p(s|genuine)}{p(s|genuine) + p(s|impostor)}.$$

The above approach has two main drawbacks. The assumption of a normal distribution for the scores may not be true in many cases. For example, the scores of the fingerprint and hand-geometry matchers used in our experiments do not follow a normal distribution. Secondly, the approach does not make use of the prior probabilities of the genuine and impostor users that may be available to the system. Due to these reasons, we have proposed the use of a non-parametric technique, viz., Parzen window density estimation method

[7], to estimate the actual conditional density of the genuine and impostor scores. After estimating the conditional densities, the Bayes formula can be applied to calculate the posteriori probability of the score being that of a genuine user. Thus,

$$P(genuine|s) = \frac{p(s|genuine) * P(genuine)}{p(s)},$$

where $p(s) = \{p(s|genuine) * P(genuine) + p(s|impostor) * P(impostor)\}$ and $P(genuine)$ and $P(impostor)$ are the prior probabilities of a genuine user and an impostor, respectively.

Although the Parzen window density estimation technique significantly reduces the error in the estimation of $P(genuine|s)$ (especially when the conditional densities are not Gaussian), the error is still non-zero due to the finite training set and the problems in choosing the optimum window width during the density estimation process. Further, the assumption that the value of η in Eq. (1) is zero is not valid in most practical biometric systems. Since η depends on the input biometric sample Z, it is possible to estimate η only if the biometric system outputs a confidence measure (that takes into account the nature of the input Z) on the matching score along with the matching score itself. In the absence of this confidence measure, the calculated value of $P(genuine|s)$ is a poor estimate of $P(genuine|Z)$ and this can lead to poor recognition performance of the multimodal system (see Fig. 10). Hence, when the outputs of individual modalities are just matching scores without any measures quantifying the confidence on those scores, it would be better to combine the matching scores directly using an appropriate method without converting them into probabilities.

The following issues need to be considered prior to combining the scores of the matchers into a single score. The matching scores at the output of the individual matchers *may not be homogeneous*. For example, one matcher may output a distance (dissimilarity) measure while another may output a proximity (similarity) measure. Furthermore, the outputs of the individual matchers *need not be on the same numerical scale* (range). Finally, the matching scores at the output of the matchers *may follow different statistical distributions*. Due to these reasons, score normalization is essential to transform the scores of the individual matchers into a common domain prior to combining them. Score normalization is a critical part in the design of a combination scheme for matching score level fusion. Fig. 2 shows the conditional distributions of the face, fingerprint and hand-geometry matching scores used in our experiments. The scores obtained from the face and hand-geometry matchers are distance scores and those obtained from the fingerprint matcher are similarity scores. One can easily observe the non-homogeneity in these scores and the need for normalization prior to any meaningful combination.

Score normalization refers to changing the location and scale parameters of the matching score distributions at the outputs of the individual matchers, so that the matching scores of different matchers are transformed into a common domain. When the parameters used for normalization are determined using a fixed training set, it is referred to as *fixed score normalization* [24]. In such a case, the matching score distribution of the training set is examined and a suitable model is chosen to fit the distribution. Based on the model, the normalization parameters are determined. In *adaptive score normalization*, the normalization parameters are estimated based on the current feature vector. This approach has the ability to adapt to variations in the input data such as the change in the length of the speech signal in speaker recognition systems.

The problem of score normalization in multimodal biometric systems is identical to the problem of score normalization in metasearch. Metasearch is a technique for combining the relevance scores of documents produced by different search engines, in order to improve the performance of document retrieval systems [25]. Min–max and z-score normalization are some of the popular techniques used for relevance score normalization in metasearch. In metasearch literature [26], the distribution of scores of relevant documents is generally approximated as a Gaussian distribution with a large standard deviation while that of non-relevant documents is approximated as an exponential

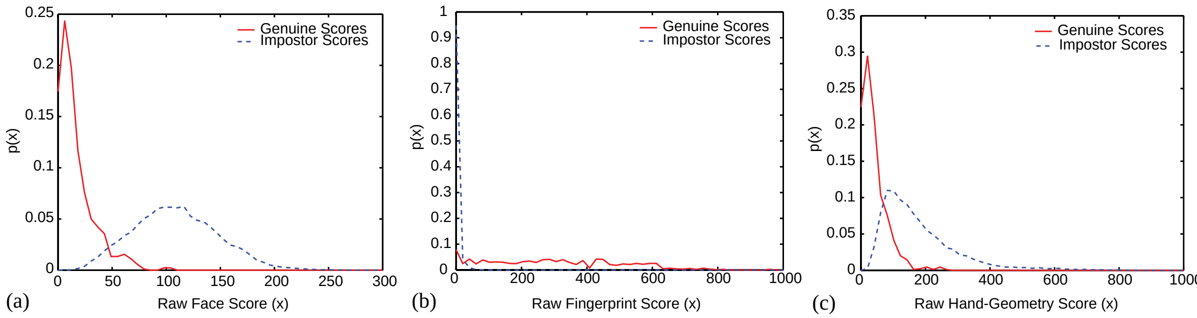

Fig. 2. Conditional distribution of genuine and impostor scores: (a) face (distance score); (b) fingerprint (similarity score); and (c) hand-geometry (distance score).

distribution. In our experiments, the distributions of the genuine and impostor fingerprint scores closely follow the distributions of relevant and non-relevant documents in metasearch. However, the face and hand-geometry scores do not exhibit this behavior.

For a good normalization scheme, the estimates of the location and scale parameters of the matching score distribution must be *robust* and *efficient*. Robustness refers to insensitivity to the presence of outliers. Efficiency refers to the proximity of the obtained estimate to the optimal estimate when the distribution of the data is known. Huber [27] explains the concepts of robustness and efficiency of statistical procedures. He also explains the need for statistical procedures that have both these desirable characteristics. Although many techniques can be used for score normalization, the challenge lies in identifying a technique that is both robust and efficient.

3.1. Normalization techniques

The simplest normalization technique is the *Min–max normalization* (Table 1). Min–max normalization is best suited for the case where the bounds (maximum and minimum values) of the scores produced by a matcher are known. In this case, we can easily shift the minimum and maximum scores to 0 and 1, respectively. However, even if the matching scores are not bounded, we can estimate the minimum and maximum values for a set of matching scores and then apply the min–max normalization. Given a set of matching scores $\{s_k\}$, $k=1,2,\ldots,n$, the normalized scores are given by

$$s'_k = \frac{s_k - \min}{\max - \min}.$$

When the minimum and maximum values are estimated from the given set of matching scores, this method is not robust (i.e., the method is highly sensitive to outliers in the data used for estimation). Min–max normalization retains the original distribution of scores except for a scaling factor and transforms all the scores into a common range $[0,1]$. Distance scores can be transformed into similarity scores by subtracting the min–max normalized score from 1. Fig. 3 shows the distribution of face, fingerprint and hand-geometry scores after min–max normalization.

Decimal scaling can be applied when the scores of different matchers are on a logarithmic scale. For example, if one matcher has scores in the range $[0,1]$ and the other has scores in the range $[0,1000]$, the following normalization could be applied.

$$s'_k = \frac{s_k}{10^n},$$

where $n = \log_{10} \max(s_i)$. The problems with this approach are lack of robustness and the assumption that the scores of different matchers vary by a logarithmic factor. In our experiments, the matching scores of the three modalities are not distributed on a logarithmic scale and, hence, this normalization technique cannot be applied.

The most commonly used score normalization technique is the *z-score* that is calculated using the arithmetic mean and standard deviation of the given data. This scheme can be expected to perform well if prior knowledge about the average score and the score variations of the matcher is available. If we do not have any prior knowledge about the nature of the matching algorithm, then we need to estimate the mean and standard deviation of the scores from

Table 1
Summary of normalization techniques

Normalization technique	Robustness	Efficiency
Min–max	No	N/A
Decimal scaling	No	N/A
z-score	No	High (optimal for Gaussian data)
Median and MAD	Yes	Moderate
Double sigmoid	Yes	High
tanh-estimators	Yes	High
Biweight estimators	Yes	High

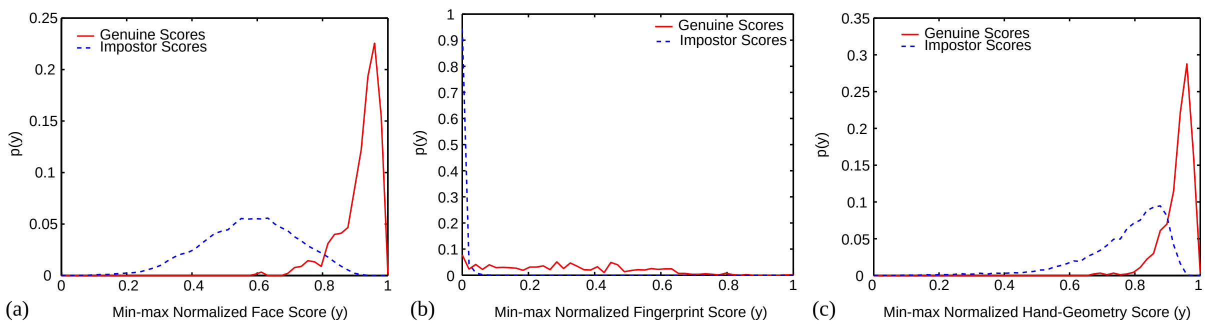

Fig. 3. Distribution of genuine and impostor scores after min–max normalization: (a) face; (b) fingerprint; and (c) hand-geometry.

a given set of matching scores. The normalized scores are given by

$$s'_k = \frac{s_k - \mu}{\sigma},$$

where μ is the arithmetic mean and σ is the standard deviation of the given data. However, both mean and standard deviation are sensitive to outliers and, hence, this method is not robust. *Z*-score normalization does not guarantee a common numerical range for the normalized scores of the different matchers. If the input scores are not Gaussian distributed, *z*-score normalization does not retain the input distribution at the output. This is due to the fact that mean and standard deviation are the optimal location and scale parameters only for a Gaussian distribution. For an arbitrary distribution, mean and standard deviation are reasonable estimates of location and scale, respectively, but are not optimal.

The distributions of the matching scores of the three modalities after *z*-score normalization are shown in Fig. 4. The face and hand-geometry scores are converted into similarity scores by subtracting from a large number (300 for face and 1000 for hand-geometry in our experiments) before applying the *z*-score transformation. It is clearly seen that *z*-score normalization fails to transform the scores of the different modalities into a common numerical range and also does not retain the original distribution of scores in the case of fingerprint modality.

The *median* and *median absolute deviation* (MAD) are insensitive to outliers and the points in the extreme tails of the distribution. Hence, a normalization scheme using median and MAD would be robust and is given by

$$s'_k = \frac{s_k - median}{MAD},$$

where MAD = median($|s_k - median|$). However, the median and the MAD estimators have a low efficiency compared to the mean and the standard deviation estimators, i.e., when the score distribution is not Gaussian, median and MAD are poor estimates of the location and scale parameters. Therefore, this normalization technique does not retain the input distribution and does not transform the scores into a common numerical range. This is illustrated by the distributions of the normalized face, fingerprint, and hand-geometry scores in Fig. 5.

Cappelli et al. [28] have used a *double sigmoid function* for score normalization in a multimodal biometric system that combines different fingerprint classifiers. The

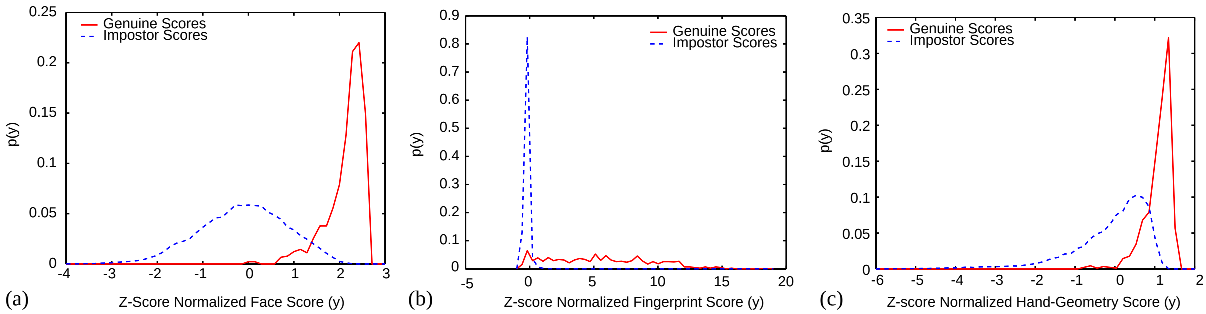

Fig. 4. Distribution of genuine and impostor scores after *z*-score normalization: (a) face; (b) fingerprint; and (c) hand-geometry.

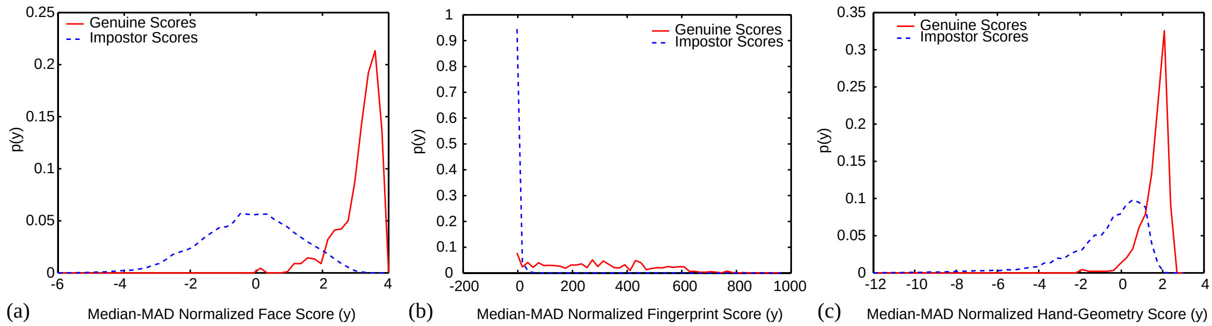

Fig. 5. Distribution of genuine and impostor scores after median-MAD normalization: (a) face; (b) fingerprint; and (c) hand-geometry.

normalized score is given by

$$s'_k = \begin{cases} \dfrac{1}{1+\exp(-2((s_k - t)/r_1))} & \text{if } s_k < t, \\ \dfrac{1}{1+\exp(-2((s_k - t)/r_2))} & \text{otherwise,} \end{cases}$$

where t is the reference operating point and r_1 and r_2 denote the left and right edges of the region in which the function is linear, i.e., the double sigmoid function exhibits linear characteristics in the interval $(t - r_1, t - r_2)$. Fig. 6 shows an example of the double sigmoid normalization, where the scores in the $[0, 300]$ range are mapped to the $[0, 1]$ range using $t = 200$, $r_1 = 20$ and $r_2 = 30$.

This scheme transforms the scores into the $[0, 1]$ interval. But, it requires careful tuning of the parameters t, r_1, r_2 to obtain good efficiency. Generally, t is chosen to be some value falling in the region of overlap between the genuine and impostor score distribution, and r_1 and r_2 are made equal to the extent of overlap between the two distributions toward the left and right of t, respectively. This normalization scheme provides a linear transformation of the scores in the region of overlap, while the scores outside this region are transformed non-linearly. The double sigmoid normalization is very similar to the min–max normalization followed by the application of two-quadrics (QQ) or logistic (LG) function as suggested by Snelick et al. [29]. When r_1 and r_2 are large, the double sigmoid normalization closely resembles the QQ-min–max normalization. On the other hand, we can make the double sigmoid normalization tend toward LG-min–max normalization by assigning small values to r_1 and r_2.

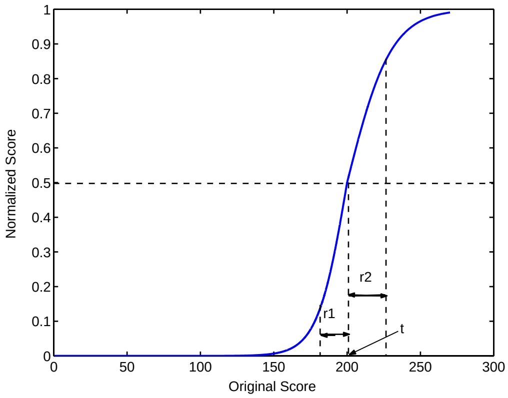

Fig. 6. Double sigmoid normalization (t=200, r_1=20, and r_2=30).

Fig. 7 shows the face, fingerprint and hand-geometry score distributions after double sigmoid normalization. The face and hand-geometry scores are converted into similarity scores by subtracting the normalized scores from 1. The parameters of the double sigmoid normalization were chosen as follows: t is chosen to be the center of the overlapping region between the genuine and impostor score distribution, and r_1 and r_2 are made equal to the extent of overlap between the two distributions toward the left and right of the center, respectively. A matching score that is equally likely to be from a genuine user and an impostor is chosen as the center (t) of the region of overlap. Then r_1 is the difference between t and the minimum of genuine scores, while r_2 is the difference between the maximum of impostor scores and t. In order to make this normalization robust, approximately 2% of the scores at the extreme tails of the genuine and impostor distributions were omitted when calculating r_1 and r_2. It must be noted that this scheme cannot be applied as described here if there are multiple intervals of overlap between genuine and impostor distributions. Although this normalization scheme transforms all the scores to a common numerical range $[0, 1]$, it does not retain the shape of the original distribution of the fingerprint scores.

The *tanh-estimators* introduced by Hampel et al. [30] are robust and highly efficient. The normalization is given by

$$s'_k = \frac{1}{2}\left\{\tanh\left(0.01\left(\frac{s_k - \mu_{GH}}{\sigma_{GH}}\right)\right) + 1\right\},$$

where μ_{GH} and σ_{GH} are the mean and standard deviation estimates, respectively, of the genuine score distribution as

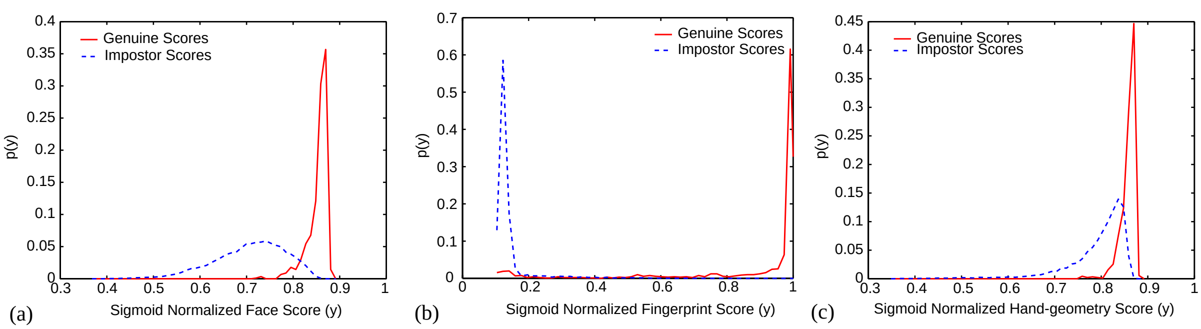

Fig. 7. Distribution of genuine and impostor scores after double sigmoid normalization: (a) face; (b) fingerprint; and (c) hand-geometry.

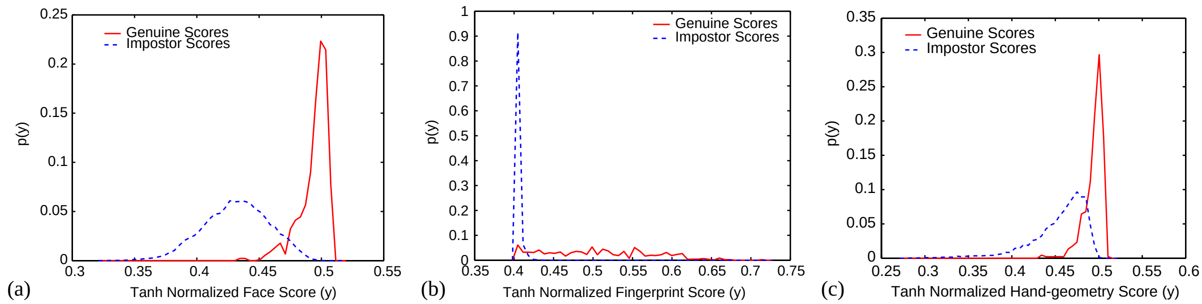

Fig. 8. Distribution of genuine and impostor scores after tanh normalization: (a) face; (b) fingerprint; and (c) hand-geometry.

given by Hampel estimators.[1] Hampel estimators are based on the following influence (ψ)-function:

$$\psi(u) = \begin{cases} u & 0 \leqslant |u| < a, \\ a \operatorname{sign}(u) & a \leqslant |u| < b, \\ a \operatorname{sign}(u)\left(\dfrac{c-|u|}{c-b}\right) & b \leqslant |u| < c, \\ 0 & |u| \geqslant c. \end{cases}$$

The Hampel influence function reduces the influence of the points at the tails of the distribution (identified by a, b, and c) during the estimation of the location and scale parameters. Hence, this method is not sensitive to outliers. If the influence of a large number of tail-points is reduced, the estimate is more robust but not efficient (optimal). On the other hand, if many tail-points influence the estimate, the estimate is not robust but the efficiency increases. Therefore, the parameters a, b, and c must be carefully chosen depending on the amount of robustness required which in turn depends on the estimate of the amount of noise in the available training data.

In our experiments, the values of a, b and c were chosen such that 70% of the scores were in the interval $(m-a, m+a)$, 85% of the scores were in the interval $(m-b, m+b)$, and 95% of the scores were in the interval $(m-c, m+c)$, where 'm' is the median score. The distributions of the scores of the three modalities after tanh normalization are shown in Fig. 8. The distance to similarity transformation is achieved by subtracting the normalized scores from 1. The nature of the tanh distribution is such that the genuine score distribution in the transformed domain has a mean of 0.5 and a standard deviation of approximately 0.01. The constant 0.01 in the expression for tanh normalization determines the spread of the normalized genuine scores. In our experiments, the standard deviation of genuine scores of face, fingerprint and hand-geometry modalities are 16.7, 202.1, and 38.9, respectively. We observe that the genuine fingerprint scores have a standard deviation that is approximately 10 times the standard deviation of the genuine face and hand-geometry scores. Hence, using the same constant, 0.01, for the fingerprint modality is entirely inappropriate. To avoid this problem, the constant factor in the tanh normalization for fingerprint modality was set to 0.1. Therefore, the standard deviation of the tanh normalized genuine fingerprint scores is roughly 0.1, which is about 10 times that of the face and hand-geometry modalities. This modification retains the information contained in the fingerprint scores even after the normalization, resulting in better performance.

Mosteller and Tukey [31] introduced the biweight location and scale estimators that are robust and efficient. But, the *biweight estimators* are iterative in nature (an initial estimate of the biweight location and scale parameters is chosen, and this estimate is updated based on the training scores), and are applicable only for Gaussian data. The biweight location and scale estimates of the data used in our experiments were very close to the mean and standard deviation. Hence, the results of this scheme were quite similar to those produced by the z-score normalization. Therefore, we have not included the results for biweight normalization in this paper.

Snelick et al. have developed a general testing framework [21] that allows system designers to evaluate multimodal biometric systems by varying different factors like the biometric traits, matching algorithms, normalization schemes, fusion methods and sample databases. To illustrate this testing methodology, they evaluated the performance of a multimodal biometric system that used face and fingerprint classifiers. Normalization techniques like min–max, z-score, median and MAD, and tanh estimators were used to transform the scores into a common domain. The transformed scores were then combined using fusion methods like simple sum of scores, maximum score, minimum score, sum of posteriori probabilities (sum rule), and product of posteriori probabilities (product rule). Their experiments conducted on a database of more than 1000 users showed that the min–max normalization followed by the sum of scores fusion method generally provided better recognition

[1] In [21,24], the mean and standard deviation of all the training scores (both genuine and impostor) were used for tanh normalization. However, we observed that considering the mean and standard deviation of only the genuine scores results in a better recognition performance.

performance than other schemes. The results of Snelick et al. [29] also show a similar trend. However, the reasons for such a behavior have not been presented by these authors. In this paper, we have tried to analyze the reasons for the differences in the performance of the different normalization schemes. We have tried to systematically study the different normalization techniques to ascertain their role in the performance of a multimodal biometric system consisting of face, fingerprint and hand-geometry modalities. In addition to the four normalization techniques employed in [21,29], we have also analyzed the double sigmoid method of normalization.

4. Experimental results

The multimodal database used in our experiments was constructed by merging two separate databases (of 50 users each) collected using different sensors and over different time periods. The first database (described in [4]) was constructed as follows: Five face images and five fingerprint impressions (of the same finger) were obtained from a set of 50 users. Face images were acquired using a Panasonic CCD camera (640 × 480) and fingerprint impressions were obtained using a Digital Biometrics sensor (500 dpi, 640×480). Five hand-geometry images were obtained from a different set of 50 users (some users were present in both the sets) and captured using a Pulnix TMC-7EX camera. The mutual independence assumption of the biometric traits allows us to randomly pair the users from the two sets. In this way, a database consisting of 50 users was constructed, each user having five biometric templates for each modality. The biometric data captured from every user is compared with that of all the users in the database leading to one genuine score vector and 49 impostor score vectors for each distinct input. Thus, 500 (50 × 10) genuine score vectors and 24,500 (50 × 10 × 49) impostor score vectors were obtained from this database. The second database also consisted of 50 users whose face images were captured using a Sony video camera (256 × 384) and fingerprint images were acquired using an Identix sensor (500 dpi, 255 × 256). The Pulnix TMC-7EX camera was used to obtain hand-geometry images. This database also gave rise to 500 genuine and 24,500 impostor score vectors. Merging the scores from the two databases resulted in a database of 100 users with 1000 genuine score vectors and 49,000 impostor score vectors. A score vector is a 3-tuple $\langle s_1, s_2, s_3 \rangle$, where s_1, s_2, and s_3 correspond to the matching scores obtained from the face, fingerprint and hand-geometry matchers, respectively. Of the 10 genuine and 10 × 49 impostor score vectors available for each user, 6 genuine and 6 impostor score vectors were randomly selected and used for training (for calculating the parameters of each normalization technique or for density estimation by the Parzen window method). The remaining 4 genuine and 4 × 49 impostor score vectors of each user were used for testing the performance of the system. Again assuming the independence of the three modalities, we create 4^3 "virtual" users from each real user, by considering all possible combinations of scores of the three modalities. Thus, we have 64,000 (100×4^3) genuine score vectors and 313,600 ($100 \times 4^3 \times 49$) impostor score vectors to analyze the system performance. This separation of the database into training and test sets was repeated 40 times and we have reported the average performance results. Fingerprint matching was done using the minutiae features [32] and the output of the fingerprint matcher was a similarity score. Eigenface coefficients were used to represent features of the face image [33]. The Euclidean distance between the eigenface coefficients of the template and that of the input face was used as the matching score. The hand-geometry images were represented by a 14-dimensional feature vector [34] and the matching score was computed as the Euclidean distance between the input feature vector and the template feature vector.

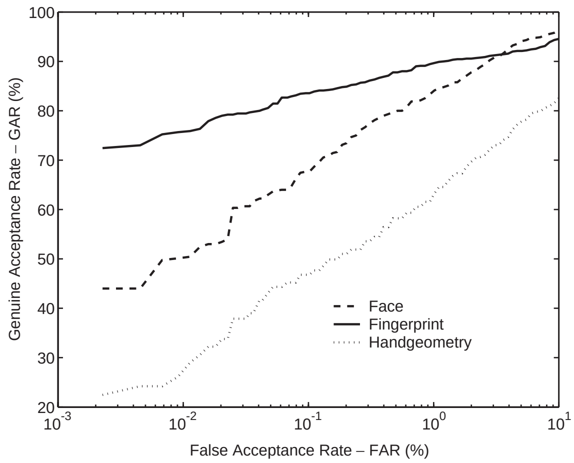

Fig. 9. ROC curves for individual modalities.

4.1. Performance results

The recognition performance of the face, fingerprint, and hand-geometry systems when operated as unimodal systems is shown in Fig. 9. From Fig. 2(a), we observe that there is a significant overlap between the genuine and impostor distributions of the raw face scores, and this explains the poor recognition performance of the face module. Fig. 2(b) shows that most of the impostor fingerprint scores are close to zero and that the genuine fingerprint scores are spread over a wide range of values. Moreover, the overlap between the two conditional densities is small and, hence, the fingerprint system performs better than the face and hand-geometry modules. The overlap between the genuine and impostor distributions of the hand-geometry system is the highest among all the three modalities as shown in Fig. 2(c). Hence, the hand geometry based recognition performance is low compared to the fingerprint and face matchers.

Table 2
Genuine acceptance rate (GAR) (%) of different normalization and fusion techniques at the 0.1% false acceptance rate (FAR)

Normalization techniques	Fusion techniques		
	Sum of scores	Max-score	Min-score
STrans	98.3 (0.4)	46.7 (2.3)	83.9 (1.6)
Min–max	97.8 (0.6)	67.0 (2.5)	83.9 (1.6)
z-score	98.6 (0.4)	92.1 (1.1)	84.8 (1.6)
Median	84.5 (1.3)	83.7 (1.6)	68.8 (2.2)
Sigmoid	96.5 (1.3)	83.7 (1.6)	83.1 (1.8)
Tanh	98.5 (0.4)	86.9 (1.8)	85.6 (1.5)
Parzen	95.7 (0.9)	93.6 (2.0)	83.9 (1.9)

Note that the values in the table represent average GAR, and the values indicated in parentheses correspond to the standard deviation of GAR.

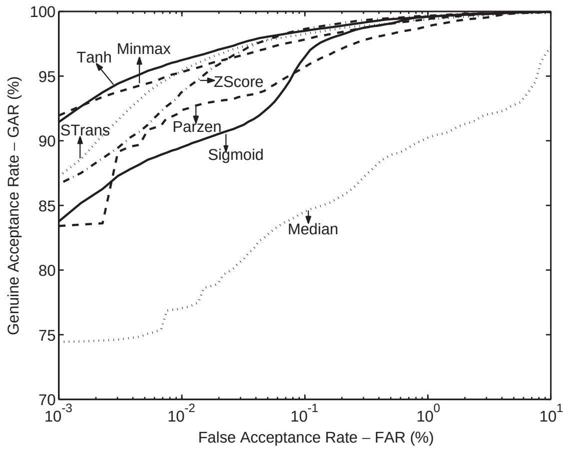

Fig. 10. ROC curves for sum of scores fusion method.

The performance of the multimodal biometric system has been studied under different normalization and fusion techniques. The simple sum of scores, the max-score, and the min-score fusion methods were applied on the normalized scores. The normalized scores were obtained by using one of the following techniques: simple distance-to-similarity transformation with no change in scale (STrans), min–max normalization (Minmax), *z*-score normalization (ZScore), median-MAD normalization (Median), double sigmoid normalization (Sigmoid), tanh normalization (Tanh), and Parzen normalization (Parzen).[2] Table 2 summarizes the average genuine acceptance rate (GAR) of the multimodal system along with the standard deviation of the GAR (shown in parentheses) for different normalization and fusion schemes, at a false acceptance rate (FAR) of 0.1%.

Fig. 10 shows the recognition performance of the system when the scores are combined using the sum of scores method. We observe that a multimodal system employing the sum of scores method provides better performance than the best unimodal system (fingerprint in this case) for all normalization techniques except median-MAD normalization. For example, at a FAR of 0.1%, the GAR of the fingerprint module is about 83.6%, while that of the multimodal system is high as 98.6% when *z*-score normalization is used. This improvement in performance is significant and it underscores the benefit of multimodal systems.

Among the various normalization techniques, we observe that the tanh and min–max normalization techniques outperform other techniques at low FARs. At higher FARs, *z*-score normalization provides slightly better performance than tanh and min–max normalization. In a multimodal system using the sum of scores fusion method, the combined score (s) is just a linear transformation of the score vector $\langle s_1, s_2, s_3 \rangle$, i.e., $s=(a_1 s_1 - b_1)+(a_2 s_2 - b_2)+(a_3 s_3 - b_3)$, where s_1, s_2, and s_3 correspond to the matching scores obtained from the face, fingerprint and hand-geometry matchers, respectively. The effect of different normalization techniques is to determine the weights a_1, a_2, and a_3, and the biases b_1, b_2, and b_3. Since the MAD of the fingerprint scores is very small compared to that of face and hand-geometry scores, the median-MAD normalization assigns a much larger weight to the fingerprint score ($a_2 \gg a_1, a_3$). This is a direct consequence of the moderate efficiency of the median-MAD estimator. The distribution of the fingerprint scores deviates drastically from the Gaussian assumption and, hence, median and MAD are not the right measures of location and scale, respectively. In this case, the combined score is approximately equal to the fingerprint score and the performance of the multimodal system is close to that of the fingerprint module. On the other hand, min–max normalization, *z*-score normalization, tanh and distance-to-similarity transformation assign nearly optimal weights to the three scores. Therefore, the recognition performance of the multimodal system when using one of these techniques along with the sum of scores fusion method is significantly better than that of the fingerprint matcher. The difference in performance between the min–max, *z*-score, tanh and distance-to-similarity transformation is relatively small. However, it should be noted that the raw scores of the three modalities used in our experiments are comparable and, hence, a simple distance-to-similarity conversion works reasonably well. If the scores of the three modalities are significantly different, then this method will not work.

The performance of the multimodal system using max-score fusion is shown in Fig. 11. Here, *z*-score and Parzen normalization provide better recognition performance compared to that of the fingerprint matcher. In a max-score fusion method that uses only the distance-to-similarity transforma-

[2] Conversion of matching scores into posteriori probabilities by the Parzen window method is really not a normalization technique. However, for the sake of convenience we refer to this method as Parzen normalization. In this case, the simple sum of scores, max score, and min score fusion schemes, respectively, reduce to the sum rule, max rule, and min rule described in [18].

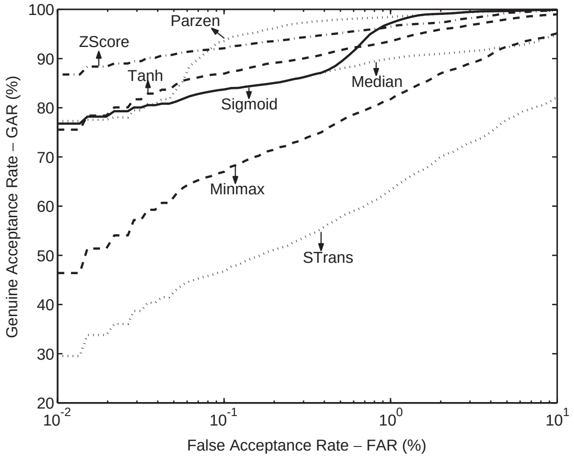

Fig. 11. ROC curves for max-score fusion method.

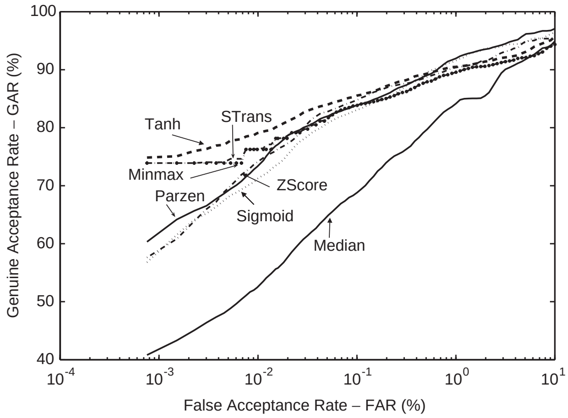

Fig. 12. ROC curves for min-score fusion method.

tion, the hand-geometry scores begin to dominate. Therefore, the performance is only slightly better than the hand-geometry module. For min–max normalization, the face and hand-geometry scores are comparable and they dominate the fingerprint score. This explains why the performance of the multimodal system is close to that of the face recognition system. When median-MAD, tanh, and double sigmoid normalization are used, the fingerprint scores are much higher compared to the face and hand-geometry scores. This limits the performance of the system close to that of the fingerprint module. In z-score normalization, the scores of the three modules are comparable and, hence, the combined score depends on all the three scores and not just the score of one modality. This improves the relative performance of the max-score fusion method compared to other normalization methods. Finally, Parzen normalization followed by max-score fusion accepts the user even if one of the three modalities produces a high estimate of the posteriori probability and rejects only if all the three modalities make errors in the probability estimation process. Hence, this method has a high genuine acceptance rate.

Fig. 12 shows the performance of the multimodal system when min-score fusion method is employed. For median-MAD normalization, most of the face scores have smaller values compared to the fingerprint and hand-geometry scores. Therefore, for median-MAD normalization the performance of the min-score method is close to the performance of the face-recognition system. On the other hand, the fingerprint scores have smaller values for all other normalization schemes and, hence, their performance is very close to that of the fingerprint matcher.

For sum of scores fusion, we see that the performance of a robust normalization technique like tanh is almost the same as that of the non-robust techniques like min–max and z-score normalization. However, the performance of such non-robust techniques is highly dependent on the accuracy of the estimates of the location and scale parameters. The scores produced by the matchers in our experiments are unbounded and, hence, can theoretically produce any value. Also, the statistics of the scores (like average or deviation from the average) produced by these matchers is not known. Therefore, parameters like the minimum and maximum scores in min–max normalization, and the average and standard deviation of scores in z-score normalization have to be estimated from the available data. The data used in our experiments does not contain any outliers and, hence, the performance of the non-robust normalization techniques were not affected. In order to demonstrate the sensitivity of the min–max and z-score normalization techniques to the presence of outliers, we artificially introduced outliers in the fingerprint scores.

For min–max normalization, a single large score whose value is 125% or 150% or 175% or 200% of the original maximum score is introduced into the fingerprint data. Fig. 13 shows the recognition performance of the multimodal system after the introduction of the outlier. We can clearly see that the performance is highly sensitive to the maximum score. A single large score that is twice the original maximum score can reduce the recognition rate by 3–5% depending on the operating point of the system. The performance degradation is more severe at lower values of FAR.

In the case of z-score normalization, a few large scores were introduced in the fingerprint data so that the standard deviation of the fingerprint score is increased by 125% or 150% or 175% or 200% of the original standard deviation. In one trial, some large scores were reduced to decrease the standard deviation to 75% of the original value. In the case of increase in standard deviation, the performance improves after the introduction of outliers as indicated in Fig. 14. Since the initial standard deviation was small, fingerprint scores were assigned a higher weight compared to the other modalities. As the standard deviation is increased, the domination of the fingerprint scores was reduced and this resulted in improved recognition rates. However, the goal of this

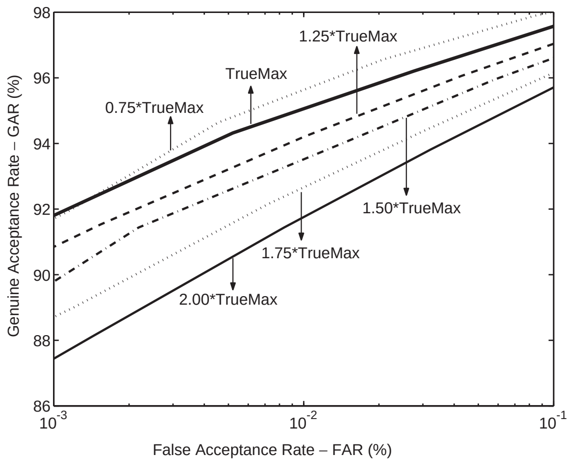

Fig. 13. Robustness analysis of min–max normalization.

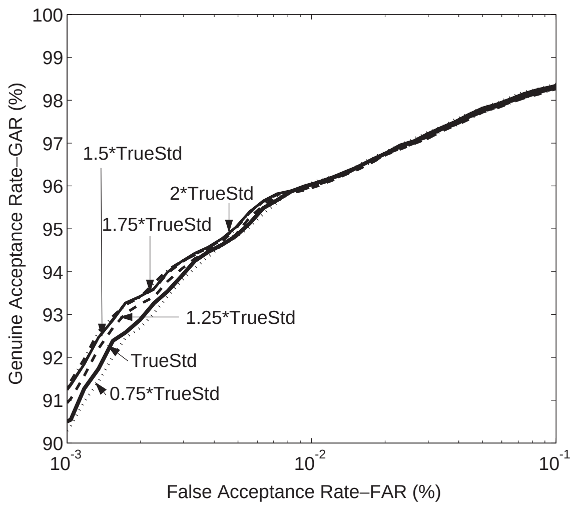

Fig. 15. Robustness analysis of tanh normalization.

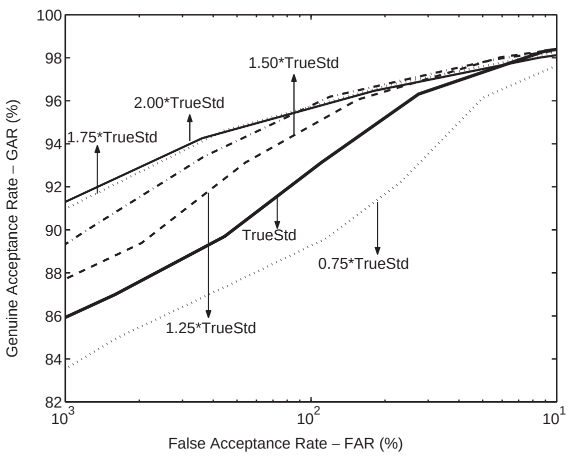

Fig. 14. Robustness analysis of *z*-score normalization.

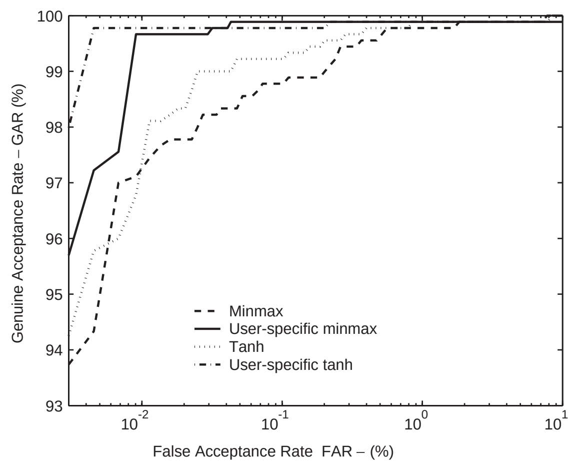

Fig. 16. Improvement in recognition rates using user-specific weights.

experiment is to show the sensitivity of the system to those estimated parameters that can be easily affected by outliers. A similar experiment was done for tanh-normalization technique and, as shown in Fig. 15, there is no significant variation in the performance after the introduction of outliers. This result highlights the robustness of the tanh normalization method.

Finally, the weighting of the three modalities during the computation of the combined score was varied for each user. The weights for each user were computed based on the training scores as explained in Section 2.2.2. The scores were normalized using min–max and tanh normalization before calculating weighted combined score. We find that the introduction of user-specific weights result in a significant improvement in the recognition performance. The ROC curves summarizing this improvement are shown in Fig. 16.

5. Conclusion and future work

This paper examines the effect of different score normalization techniques on the performance of a multimodal biometric system. We have demonstrated that the normalization of scores prior to combining them improves the recognition performance of a multimodal biometric system that uses the face, fingerprint and hand-geometry traits for user authentication. Min–max, *z*-score, and tanh normalization techniques followed by a simple sum of scores fusion method result in a superior GAR than all the other normalization and fusion techniques. We have shown that both min–max and *z*-score methods are sensitive to outliers. On the other hand, tanh normalization method is both robust and efficient. If the location and scale parameters of the matching scores (minimum and maximum values for min–max, or mean and standard deviation for z-score) of the individual modalities are known in advance, then simple normalization techniques like min–max and z-score would suffice. If these parameters are to be estimated using some training scores, and if the training scores are noisy, then one should choose a robust normalization technique like the tanh normalization. We have also explored the use of non-parametric approaches like the Parzen window density estimation method to convert the matching scores into posteriori probabilities and combining the probabilities using fusion rules. The advantage of this method is that it avoids the need for any knowledge about the distribution of matching scores. However, the performance of this method is dependent on the type and width of the kernel used for the estimation of density. Computation of the weighted sum of scores based on user-specific weights results in further improvement in the recognition performance.

Some normalization schemes work well if the scores follow a specific distribution. For example, z-score normalization is optimal if the scores of all the modalities follow a Gaussian distribution. Therefore, we need to develop rules that would allow a practitioner to choose a normalization scheme after analyzing the genuine and impostor score distributions of the individual matchers. The possibility of applying different normalization techniques to the scores of different modalities must be explored. Guidelines for choosing the design parameters of some normalization techniques (e.g., the values of the constants a, b, and c in tanh normalization, the type and width of the kernel function to be used in Parzen normalization) need to be developed. Alternative normalization and fusion techniques such as those based on Bayesian statistics also need to be explored.

References

[1] A.K. Jain, A. Ross, S. Prabhakar, An introduction to biometric recognition, IEEE Trans. Circuits Systems Video Technol. 14 (1) (2004) 4–20 (special issue on image- and video-based biometrics).

[2] A.K. Jain, A. Ross, Multibiometric systems, Commun. ACM 47 (1) (2004) 34–40 (special issue on multimodal interfaces).

[3] L. Hong, A.K. Jain, S. Pankanti, Can multibiometrics improve performance? in: Proceedings of IEEE Workshop on Automatic Identification Advanced Technologies, NJ, USA, 1999, pp. 59–64.

[4] A. Ross, A.K. Jain, Information fusion in biometrics, Pattern Recogn. Lett. 24 (13) (2003) 2115–2125 (special issue on multimodal biometrics).

[5] C. Sanderson, K.K. Paliwal, Information fusion and person verification using speech and face information, Research Paper IDIAP-RR 02-33, IDIAP, September 2002.

[6] S.S. Iyengar, L. Prasad, H. Min, Advances in Distributed Sensor Technology, Prentice-Hall, Englewood Cliffs, NJ, 1995.

[7] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, Wiley, New York, 2001.

[8] K. Woods, K. Bowyer, W.P. Kegelmeyer, Combination of multiple classifiers using local accuracy estimates, IEEE Trans. Pattern Anal. Mach. Intell. 19 (4) (1997) 405–410.

[9] K. Chen, L. Wang, H. Chi, Methods of combining multiple classifiers with different features and their applications to text-independent speaker identification, Int. J. Pattern Recogn. Artif. Intell. 11 (3) (1997) 417–445.

[10] L. Lam, C.Y. Suen, Application of majority voting to pattern recognition: an analysis of its behavior and performance, IEEE Trans. Systems Man Cybernet. Part A: Systems Humans 27 (5) (1997) 553–568.

[11] L. Lam, C.Y. Suen, Optimal combination of pattern classifiers, Pattern Recogn. Lett. 16 (1995) 945–954.

[12] L. Xu, A. Krzyzak, C.Y. Suen, Methods for combining multiple classifiers and their applications to handwriting recognition, IEEE Trans. Systems Man Cybernet. 22 (3) (1992) 418–435.

[13] J. Daugman, Combining Multiple Biometrics, Available at `http://www.cl.cam.ac.uk/users/jgd1000/combine/combine.html`.

[14] T.K. Ho, J.J. Hull, S.N. Srihari, Decision combination in multiple classifier systems, IEEE Trans. Pattern Anal. Mach. Intell. 16 (1) (1994) 66–75.

[15] Y. Wang, T. Tan, A.K. Jain, Combining face and iris biometrics for identity verification, in: Proceedings of Fourth International Conference on AVBPA, Guildford, UK, 2003, pp. 805–813.

[16] P. Verlinde, G. Cholet, Comparing decision fusion paradigms using k-NN based classifiers, decision trees and logistic regression in a multi-modal identity verification application, in: Proceedings of Second International Conference on AVBPA, Washington, DC, USA, 1999, pp. 188–193.

[17] V. Chatzis, A.G. Bors, I. Pitas, Multimodal decision-level fusion for person authentication, IEEE Trans. Systems Man Cybernet. Part A: Systems Humans 29 (6) (1999) 674–681.

[18] J. Kittler, M. Hatef, R.P. Duin, J.G. Matas, On combining classifiers, IEEE Trans. Pattern Anal. Mach. Intell. 20 (3) (1998) 226–239.

[19] S. Prabhakar, A.K. Jain, Decision-level fusion in fingerprint verification, Pattern Recogn. 35 (4) (2002) 861–874.

[20] E.S. Bigun, J. Bigun, B. Duc, S. Fischer, Expert conciliation for multimodal person authentication systems using Bayesian statistics, in: Proceedings of First International Conference on AVBPA, Crans-Montana, Switzerland, 1997, pp. 291–300.

[21] R. Snelick, M. Indovina, J. Yen, A. Mink, Multimodal biometrics: issues in design and testing, in: Proceedings of Fifth International Conference on Multimodal Interfaces, Vancouver, Canada, 2003, pp. 68–72.

[22] A.K. Jain, A. Ross, Learning user-specific parameters in a multibiometric system, in: Proceedings of International Conference on Image Processing, New York, USA, 2002, pp. 57–60.

[23] P. Verlinde, P. Druyts, G. Cholet, M. Acheroy, Applying Bayes based classifiers for decision fusion in a multi-modal identity verification system, in: Proceedings of International Symposium on Pattern Recognition “In Memoriam Pierre Devijver”, Brussels, Belgium, 1999.

[24] R. Brunelli, D. Falavigna, Person identification using multiple cues, IEEE Trans. Pattern Anal. Mach. Intell. 12 (10) (1995) 955–966.

[25] M. Montague, J.A. Aslam, Relevance score normalization for metasearch, in: Proceedings of Tenth International Conference

on Information and Knowledge Management, Atlanta, USA, 2001, pp. 427–433.
[26] R. Manmatha, T. Rath, F. Feng, Modeling score distributions for combining the outputs of search engines, in: Proceedings of 24th International ACM SIGIR Conference on Research and Development in Information Retrieval, New Orleans, USA, 2001, pp. 267–275.
[27] P.J. Huber, Robust Statistics, Wiley, New York, 1981.
[28] R. Cappelli, D. Maio, D. Maltoni, Combining fingerprint classifiers, in: Proceedings of First International Workshop on Multiple Classifier Systems, 2000, pp. 351–361.
[29] R. Snelick, U. Uludag, A. Mink, M. Indovina, A.K. Jain, Large scale evaluation of multimodal biometric authentication using state-of-the-art systems, IEEE Trans. Pattern Anal. Mach. Intell. (to appear).
[30] F.R. Hampel, P.J. Rousseeuw, E.M. Ronchetti, W.A. Stahel, Robust Statistics: The Approach Based on Influence Functions, Wiley, New York, 1986.
[31] F. Mosteller, J.W. Tukey, Data Analysis and Regression: A Second Course in Statistics, Addison-Wesley, Reading, MA, 1977.
[32] A.K. Jain, L. Hong, S. Pankanti, R. Bolle, An identity authentication system using fingerprints, Proc. IEEE 85 (9) (1997) 1365–1388.
[33] M. Turk, A. Pentland, Eigenfaces for recognition, J. Cognitive Neurosci. 3 (1) (1991) 71–86.
[34] A.K. Jain, A. Ross, S. Pankanti, A prototype hand geometry-based verification system, in: Proceedings of Second International Conference on AVBPA, Washington, DC, USA, 1999, pp. 166–171.

About the Author—ANIL JAIN is a University Distinguished Professor in the Departments of Computer Science and Engineering and Electrical and Computer Engineering at Michigan State University. He was the Department Chair between 1995–99. His research interests include statistical pattern recognition, exploratory pattern analysis, texture analysis, document image analysis and biometric authentication. Several of his papers have been reprinted in edited volumes on image processing and pattern recognition. He received the best paper awards in 1987 and 1991, and received certificates for outstanding contributions in 1976, 1979, 1992, 1997 and 1998 from the Pattern Recognition Society. He also received the 1996 IEEE Transactions on Neural Networks Outstanding Paper Award. He was the Editor-in-Chief of the IEEE Transactions on Pattern Analysis and Machine Intelligence between 1991–1994. He is a fellow of the IEEE, ACM, and International Association of Pattern Recognition (IAPR). He has received a Fulbright Research Award, a Guggenheim fellowship and the Alexander von Humboldt Research Award. He delivered the 2002 Pierre Devijver lecture sponsored by the International Association of Pattern Recognition (IAPR). He holds six patents in the area of fingerprint matching.

About the Author—KARTHIK NANDAKUMAR received his B.E. degree in Electronics and Communication Engineering from Anna University, Chennai, India in 2002. He is now a Ph.D. student in the Department of Computer Science and Engineering, Michigan State University. His research interests include statistical pattern recognition, biometric authentication, computer vision and machine learning.

About the Author—ARUN ROSS is an Assistant Professor in the Lane Department of Computer Science and Electrical Engineering at West Virginia University. Ross received his B.E. (Hons.) degree in Computer Science from the Birla Institute of Technology and Science, Pilani (India), in 1996. He obtained his M.S. and Ph.D. degrees in Computer Science and Engineering from Michigan State University in 1999 and 2003, respectively. Between July 1996 and December 1997, he worked with the Design and Development group of Tata Elxsi (India) Ltd., Bangalore. He also spent three summers (2000–2002) with the Imaging and Visualization group at Siemens Corporate Research, Inc., Princeton, working on fingerprint recognition algorithms. His research interests include statistical pattern recognition, image processing, computer vision and biometrics.