14

Introduction to Multibiometrics

Arun Ross[1], Karthik Nandakumar[2], and Anil K. Jain[2]

[1] Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV 26506, USA
arun.ross@mail.wvu.edu

[2] Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824
nandakum@cse.msu.edu,jain@cse.msu.edu

14.1 Introduction

Most biometric systems that are presently in use, typically use a single biometric trait to establish identity (i.e., they are unibiometric systems). With the proliferation of biometric-based solutions in civilian and law enforcement applications, it is important that the vulnerabilities and limitations of these systems are clearly understood. Some of the challenges commonly encountered by biometric systems are listed below.

1. Noise in sensed data: The biometric data being presented to the system may be contaminated by noise due to imperfect acquisition conditions or subtle variations in the biometric itself. For example, a scar can change a subject's fingerprint while the common cold can alter the voice characteristics of a speaker. Similarly, unfavorable illumination conditions may significantly affect the face and iris images acquired from an individual. Noisy data can result in an individual being incorrectly labeled as an impostor thereby increasing the False Reject Rate (FRR) of the system.

2. Non-universality: The biometric system may not be able to acquire meaningful biometric data from a subset of individuals resulting in a failure-to-enroll (FTE) error. For example, a fingerprint system may fail to image the friction ridge structure of some individuals due to the poor quality of their fingerprints. Similarly, an iris recognition system may be unable to obtain the iris information of a subject with long eyelashes, drooping eyelids or certain pathological conditions of the eye. Exception processing will be necessary in order to accommodate such users into the authentication system.

3. Upper bound on identification accuracy: The matching performance of a unibiometric system cannot be continuously improved by tuning the feature extraction and matching modules. There is an implicit upper bound on the number of distinguishable patterns (i.e., the number of distinct biometric feature sets) that can be represented using a template. The capacity of

a template is constrained by the variations observed in the feature set of each subject (i.e., *intra*-class variations) and the variations between feature sets of different subjects (i.e., *inter*-class variations). Table 1.2 lists the error rates associated with four biometric modalities - fingerprints, face, voice, iris - as suggested by recent public tests. These statistics suggest that there is a tremendous scope for performance improvement especially in the context of large-scale authentication systems.

4. Spoof attacks: Behavioral traits such as voice [15] and signature [16] are vulnerable to spoof attacks by an impostor attempting to mimic the traits corresponding to legitimately enrolled subjects. Physical traits such as fingerprints can also be spoofed by inscribing ridge-like structures on synthetic material such as gelatine and play-doh [38, 47]. Targeted spoof attacks can undermine the security afforded by the biometric system and, consequently, mitigate its benefits [48].

Some of the limitations of a unibiometric system can be addressed by designing a system that consolidates *multiple* sources of biometric information. This can be accomplished by fusing, for example, multiple traits of an individual, or multiple feature extraction and matching algorithms operating on the same biometric. Such systems, known as multibiometric systems [53, 25, 19], can improve the matching accuracy of a biometric system while increasing population coverage and deterring spoof attacks. In this chapter, the various sources of biometric information that can be fused as well as the different levels of fusion that are possible are discussed.

14.2 Taxonomy of Multibiometric Systems

In the realm of biometrics, the consolidation of evidence presented by multiple biometric sources is an effective way of enhancing the recognition accuracy of an authentication system. For example, the Integrated Automated Fingerprint Identification System (IAFIS) maintained by the FBI integrates the information presented by multiple fingers to determine a match in the master file. Some of the earliest *multimodal* biometric systems reported in the literature combined the face (image/video) and voice (audio) traits of individuals [9, 4].

A multibiometric system relies on the evidence presented by multiple sources of biometric information. Based on the nature of these sources, a multibiometric system can be classified into one of the following six categories [53]: multi-sensor, multi-algorithm, multi-instance, multi-sample, multimodal and hybrid.

1. Multi-sensor systems: Multi-sensor systems employ multiple sensors to capture a single biometric trait of an individual. For example, a face recognition system may deploy multiple 2D cameras to acquire the face image of a subject [35]; an infrared sensor may be used in conjunction with a visible-light sensor to acquire the subsurface information of a person's face [29, 7, 57]; a multispectral camera may be used to acquire images of the iris, face or finger

[54, 43]; or an optical as well as a capacitive sensor may be used to image the fingerprint of a subject [37]. The use of multiple sensors, in some instances, can result in the acquisition of complementary information that can enhance the recognition ability of the system. For example, based on the nature of illumination due to ambient lighting, the infrared and visible-light images of a person's face can present different levels of information resulting in enhanced matching accuracy. Similarly, the performance of a 2D face matching system can be improved by utilizing the shape information presented by 3D range images.

2. Multi-algorithm systems: In some cases, invoking multiple feature extraction and/or matching algorithms on the same biometric data can result in improved matching performance. Multi-algorithm systems consolidate the output of multiple feature extraction algorithms, or that of multiple matchers operating on the same feature set. These systems do not necessitate the deployment of new sensors and, hence, are cost-effective compared to other types of multibiometric systems. But on the other hand, the introduction of new feature extraction and matching modules can increase the computational complexity of these systems. Ross et al. [52] describe a fingerprint recognition system that utilizes minutiae as well as texture information to represent and match fingerprint images. The inclusion of the texture-based algorithm introduces additional processing time associated with the application of Gabor filters on the input fingerprint image. However, the performance of the hybrid matcher is shown to exceed that of the individual matchers. Lu et al. [36] discuss a face recognition system that combines three different feature extraction schemes (Principal Component Analysis (PCA), Independent Component Analysis (ICA) and Linear Discriminant Analysis (LDA)). The authors postulate that the use of different feature sets makes the system robust to a variety of intra-class variations normally associated with the face biometric. Experimental results indicate that combining multiple face classifiers can enhance the identification rate of the biometric system.

3. Multi-instance systems: These systems use multiple instances of the same body trait and have also been referred to as multi-unit systems in the literature. For example, the left and right index fingers, or the left and right irises of an individual, may be used to verify an individual's identity [45, 27]. The US-VISIT border security program presently uses the left- and right-index fingers of visitors to validate their travel documents at the port of entry. FBI's IAFIS combines the evidence of all ten fingers to determine a matching identity in the database. These systems can be cost-effective if a single sensor is used to acquire the multi-unit data in a sequential fashion (e.g., US-VISIT). However, in some instances, it may be desirable to obtain the multi-unit data simultaneously (e.g., IAFIS) thereby demanding the design of an effective (and possibly more expensive) acquisition device.

4. Multi-sample systems: A single sensor may be used to acquire multiple samples of the same biometric trait in order to account for the variations that can occur in the trait, or to obtain a more complete representation of

the underlying trait. A face system, for example, may capture (and store) the frontal profile of a person's face along with the left and right profiles in order to account for variations in the facial pose. Similarly, a fingerprint system equipped with a small size sensor may acquire multiple dab prints of an individual's finger in order to obtain images of various regions of the fingerprint. A mosaicing scheme may then be used to stitch the multiple impressions and create a composite image. One of the key issues in a multi-sample system is determining the *number* of samples that have to be acquired from an individual. It is important that the procured samples represent the *variability* as well as the *typicality* of the individual's biometric data. To this end, the desired relationship between the samples has to be established before-hand in order to optimize the benefits of the integration strategy. For example, a face recognition system utilizing both the frontal- and side-profile images of an individual may stipulate that the side-profile image should be a three-quarter view of the face [17, 42]. Alternately, given a set of biometric samples, the system should be able to automatically select the "optimal" subset that would best represent the individual's variability. Uludag et al. [58] discuss two such schemes in the context of fingerprint recognition. The first method, called `DEND`, employs a clustering strategy to choose a template set that best represents the intra-class variations, while the second method, called `MDIST`, selects templates that exhibit maximum similarity with the rest of the impressions.

5. Multimodal systems: Multimodal systems establish identity based on the evidence of multiple biometric traits. For example, some of the earliest multimodal biometric systems utilized face and voice features to establish the identity of an individual [4, 10, 3]. Physically uncorrelated traits (e.g., fingerprint and iris) are expected to result in better *improvement* in performance than correlated traits (e.g., voice and lip movement). The cost of deploying these systems is substantially more due to the requirement of new sensors and, consequently, the development of appropriate user interfaces. The identification accuracy can be significantly improved by utilizing an increasing number of traits although the *curse-of-dimensionality* phenomenon would impose a bound on this number. The curse-of-dimensionality limits the number of attributes (or features) used in a pattern classification system when only a small number of training samples is available [14]. The number of traits used in a specific application will also be restricted by practical considerations such as the cost of deployment, enrollment time, throughput time, expected error rate, user habituation issues, etc.

6. Hybrid systems: Chang et al. [5] use the term *hybrid* to describe systems that integrate a subset of the five scenarios discussed above. For example, Brunelli et al. [4] discuss an arrangement in which two speaker recognition algorithms are combined with three face recognition algorithms at the match score and rank levels via a HyperBF network. Thus, the system is multi-algorithmic as well as multimodal in its design. Similarly, the NIST BSSR1 dataset [40] has match scores pertaining to two different face matchers operating on the frontal face image of an individual (multi-algorithm), and a

fingerprint matcher operating on the left- and right-index fingers of the same individual (multi-instance).

Another category of multibiometric systems combine primary biometric identifiers (such as face and fingerprint) with soft biometric attributes (such as gender, height, weight, eye color, etc.). Soft biometric traits cannot be used to distinguish individuals reliably since the same attribute is likely to be shared by several different people in the target population. However, when used in conjunction with primary biometric traits, the performance of the authentication system can be significantly enhanced [23]. Soft biometric attributes also help in filtering (or indexing) large biometric databases by limiting the number of entries to be searched in the database. For example, if it is determined (automatically or manually) that the subject is an "Asian Male", then the system can constrain its search to only those identities in the database labeled with these attributes. Alternately, soft biometric traits can be used in surveillance applications to decide if at all primary biometric information has to be acquired from a certain individual. Automated techniques to estimate soft biometric characteristics is an ongoing area of research and is likely to benefit law enforcement and border control biometric applications.

14.3 Levels of fusion

Based on the type of information available in a certain module, different levels of fusion can be defined. Sanderson and Paliwal [55] categorize the various levels of fusion into two broad categories: pre-classification or fusion *before* matching and post-classification or fusion *after* matching (see Figure 14.1). Such a categorization is necessary since the amount of information available for fusion reduces drastically once the matcher has been invoked. Pre-classification fusion schemes typically require the development of new matching techniques (since the matchers used by the individual sources may no longer be relevant) thereby introducing additional challenges. Pre-classification schemes include fusion at the sensor (or raw data) and the feature levels while post-classification schemes include fusion at the match score, rank and decision levels. A brief description of each of these fusion levels is presented in this section.

14.3.1 Sensor-level fusion

The raw biometric data (e.g., a face image) acquired from an individual represents the richest source of information although it is expected to be contaminated by noise (e.g., non-uniform illumination, background clutter, etc.). Sensor-level fusion refers to the consolidation of (a) raw data obtained using multiple sensors or (b) multiple snapshots of a biometric using a single sensor. Mosaicing multiple impressions of the same finger is a good example of fusion

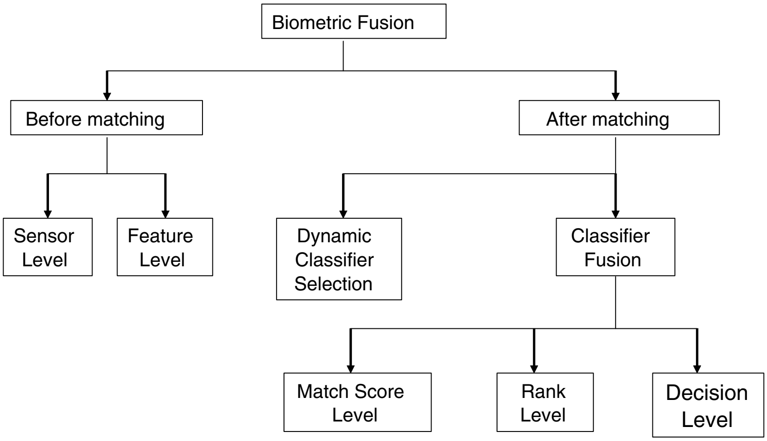

Fig. 14.1. Fusion can be accomplished at various levels in a biometric system. Most multibiometric systems fuse information at the match score level or the decision level. More recently researchers have begun to fuse information at the sensor and feature levels. In biometric systems operating in the identification mode, fusion can be done at the rank level.

at this level. Jain and Ross [24] discuss a mosaicing scheme that creates a composite fingerprint image from the evidence presented by multiple dab prints. The algorithm uses the minutiae points to first approximately register the two images using a simple affine transformation. The Iterative Closest Point (ICP) algorithm is then used to register the ridge information corresponding to the two images after applying a low-pass filter to the individual images and normalizing their histograms. The normalization ensures that the pixel intensities of the individual dab prints are comparable. Blending is accomplished by merely concatenating the two registered images. The performance using the mosaiced image templates was shown to exceed that of the individual dab print templates.

14.3.2 Feature-level fusion

In feature-level fusion, the feature sets originating from multiple biometric algorithms are consolidated into a single feature set by the application of appropriate feature normalization, transformation and reduction schemes. The primary benefit of feature-level fusion is the detection of correlated feature values generated by different biometric algorithms and, in the process, identifying a salient set of features that can improve recognition accuracy. Eliciting this feature set typically requires the use of dimensionality reduction methods [22, 46] and, therefore, feature-level fusion assumes the availability of a large number of training data. Also, the feature sets being fused are typically expected to reside in commensurate vector space in order to permit the application of a suitable matching technique upon consolidating the feature sets.

Feature-level fusion is challenging for the following reasons:

1. The relationship between the feature spaces of different biometric systems may not be known.
2. The feature sets of multiple modalities may be incompatible. For example, the minutiae set of fingerprints and the eigen-coefficients of face are irreconcilable. One is a variable length feature set (i.e., it varies across images) whose individual values parameterize a minutia point; the other is a fixed length feature set (i.e., all images are represented by a fixed number of eigen-coefficients) whose individual values are scalar entities.
3. If the two feature sets are fixed length feature vectors, then one could consider concatenating them to generate a new feature set. However, concatenating two feature vectors might lead to the curse-of-dimensionality problem ([21]) where increasing the number of features might actually degrade the system performance especially in the presence of small number of training samples. Although the curse-of-dimensionality is a well known problem in pattern recognition, it is particularly pronounced in biometric applications because of the time, effort and cost required to collect large amounts of biometric (training) data.
4. Most commercial biometric systems do not provide access to the feature sets used in their products. Hence, very few biometric researchers have focused on integration at the feature level and most of them generally prefer fusion schemes that use match scores or decision labels.

If the length of each of the two feature vectors to be consolidated is fixed across all users, then a feature concatenation scheme followed by a dimensionality reduction procedure may be adopted. Let $\mathbf{X} = \{x_1, x_2, \ldots, x_m\}$ and $\mathbf{Y} = \{y_1, y_2, \ldots, y_n\}$ denote two feature vectors ($\mathbf{X} \in R^m$ and $\mathbf{Y} \in R^n$) representing the information extracted from two different biometric sources. The objective is to fuse these two feature sets in order to yield a new feature vector, $\mathbf{Z}$, that would better represent an individual. The vector $\mathbf{Z}$ of dimensionality k, $k < (m+n)$, can be generated by first concatenating vectors $\mathbf{X}$ and $\mathbf{Y}$, and then performing feature selection or feature transformation on the resultant feature vector in order to reduce its dimensionality. The key stages of such an approach are described below.

Feature Normalization

The individual feature values of vectors $\mathbf{X} = \{x_1, x_2, \ldots, x_m\}$ and $\mathbf{Y} = \{y_1, y_2, \ldots,$
$y_n\}$ may exhibit significant differences in their range as well as form (i.e., distribution). Concatenating such diverse feature values will not be appropriate in many cases. For example, if the x_i's are in the range $[0, 100]$ while

the y_i's are in the range $[0, 1]$, then the distance between two concatenated feature vectors will be more sensitive to the x_i's than the y_i's. The goal of feature normalization is to modify the location (mean) and scale (variance) of the features values via a transformation function in order to map them into a common domain. Adopting an appropriate normalization scheme also helps address the problem of outliers in feature values. While a variety of normalization schemes can be used, two simple schemes are discussed here: the min-max and median normalization schemes.

Let x and x' denote a feature value before and after normalization, respectively. The min-max technique computes x' as

$$x' = \frac{x - \min(F_x)}{\max(F_x) - \min(F_x)}, \tag{14.1}$$

where F_x is the function which generates x, and $\min(F_x)$ and $\max(F_x)$ represent the minimum and maximum of all possible x values that will be observed, respectively. The min-max technique is effective when the minimum and the maximum values of the component feature values are known beforehand. In cases where such information is not available, an estimate of these parameters has to be obtained from the available set of training data. The estimate may be affected by the presence of outliers in the training data and this makes min-max normalization sensitive to outliers. The median normalization scheme, on the other hand, is relatively robust to the presence of noise in the training data. In this case, x' is computed as

$$x' = \frac{x - median(F_x)}{median(\mid (x - median(F_x)) \mid)}. \tag{14.2}$$

The denominator is known as the Median Absolute Deviation (MAD) and is an estimate of the scale parameter of the feature value. Although, this normalization scheme is relatively insensitive to outliers, it has a low efficiency compared to the mean and standard deviation estimators. Normalizing the feature values via any of these techniques results in modified feature vectors $\mathbf{X}' = \{x'_1, x'_2, \ldots x'_m\}$ and $\mathbf{Y}' = \{y'_1, y'_2, \ldots y'_n\}$. Feature normalization may not be necessary in cases where the feature values pertaining to multiple sources are already comparable.

Feature Selection or Transformation

Concatenating the two feature vectors, $\mathbf{X}'$ and $\mathbf{Y}'$, results in a new feature vector, $\mathbf{Z}' = \{x'_1, x'_2, \ldots x'_m\ , y'_1, y'_2, \ldots y'_n\}$, $\mathbf{Z}' \in R^{m+n}$. The curse-of-dimensionality dictates that the new vector of dimensionality $(m + n)$ need not necessarily result in an improved matching performance compared to that obtained by $\mathbf{X}'$ and $\mathbf{Y}'$ alone. The feature selection process is a dimensionality reduction scheme that entails choosing a minimal feature set of size k, $k < (m + n)$, such that a criterion (objective) function applied to the training set of feature vectors is optimized. There are several feature selection algorithms in the literature, and any one of these could be used to reduce the dimensionality of the feature set $\mathbf{Z}'$. Examples include sequential forward selection (SFS), sequential backward selection (SBS), sequential forward floating search (SFFS), sequential backward floating search (SBFS), "plus l take away r" and branch-and-bound search (see [46] and [26] for details). Feature selection techniques rely on an appropriately formulated criterion function to elicit the optimal subset of features from a larger feature set. In the case of a biometric system, this criterion function could be the Equal Error Rate (EER); the d-prime measure; the area of overlap between genuine and impostor training scores; the average GAR at pre-determined FAR values in the ROC/DET curves corresponding to the training set; or the area under the ROC curve (AUC).

Dimensionality reduction may also be accomplished using feature *transformation* methods where the vector $\mathbf{Z}'$ is subjected to a linear or a non-linear mapping that projects it to a lower dimensional subspace. Examples of such transformations include the use of principal component analysis (PCA), independent component analysis (ICA), multidimensional scaling (MDS), Kohonen Maps and neural networks ([22]). The application of a feature selection or feature transformation procedure results in a new feature vector $\mathbf{Z} = \{z_1, z_2, \ldots z_k\}$ which can now be used to represent the identity of an individual.

Ross and Govindarajan [50] apply feature-level fusion to three different scenarios: (a) multi-algorithm, where two different face recognition algorithms based on Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are combined; (b) multi-sensor, where three different color channels of a face image are independently subjected to LDA and then combined; and (c) multimodal, where the face and hand geometry feature vectors are combined.

14.3.3 Score-level fusion

A match score represents the result of comparing two feature sets extracted using the same feature extractor. A *similarity* score denotes how "similar" the two feature sets are, while a *distance* score denotes how "different" they are[3].

In score-level fusion the match scores output by multiple biometric matchers are combined to generate a new match score (a scalar) that can be subsequently used by the verification or identification modules for rendering an identity decision. Fusion at this level is the most commonly discussed approach in the biometric literature primarily due to the ease of accessing and processing match scores (compared to the raw biometric data or the feature set extracted from the data). Fusion methods at this level can be broadly classified into three categories [53]: density-based schemes, transformation-based schemes and classifier-based schemes.

[3] Consequently, a high similarity score between a pair of feature sets indicates a good match whereas a high distance score indicates a poor match.

Density-based fusion schemes

Let $\mathbf{s} = [s_1, s_2, \ldots, s_R]$ denote the scores emitted by multiple matchers, with s_j representing the match score of the j^{th} matcher, $j = 1, \ldots, R$. Further, let the labels ω_0 and ω_1 denote the genuine and impostor classes, respectively. Then, by Bayes decision theory [14], the probability of error can be minimized by adopting the following decision rule[4].

$$\text{Assign } \mathbf{s} \rightarrow \omega_i \text{ if}$$

$$P(\omega_i|\mathbf{s}) > P(\omega_j|\mathbf{s}), i \neq j, \quad and \quad i, j = 0, 1. \tag{14.3}$$

Here, the *a posteriori* probability $P(\omega_i|\mathbf{s})$, $i = 0, 1$, can be derived from the class-conditional density function $p(\mathbf{s}|\omega_i)$ using the Bayes formula, i.e.,

$$P(\omega_i|\mathbf{s}) = \frac{p(\mathbf{s}|\omega_i)P(\omega_i)}{p(\mathbf{s})}, \tag{14.4}$$

where $P(\omega_i)$ is the *a priori* probability of observing class ω_i and $p(\mathbf{s})$ denotes the probability of encountering $\mathbf{s}$. Thus, equation (14.3) can be re-written as

$$\text{Assign } \mathbf{s} \rightarrow \omega_i \text{ if}$$

$$\frac{p(\mathbf{s}|\omega_i)}{p(\mathbf{s}|\omega_j)} > \tau, i \neq j, \quad and \quad i, j = 0, 1 \tag{14.5}$$

where $\frac{p(\mathbf{s}|\omega_i)}{p(\mathbf{s}|\omega_j)}$ is known as the *likelihood ratio* and $\tau = \frac{P(\omega_j)}{P(\omega_i)}$ is a pre-determined threshold. The density $p(\mathbf{s}|\omega_i)$ is typically estimated from a training set of match score vectors, using parametric or non-parametric techniques [56]. However, a large number of training samples is necessary to reliably estimate the joint-density function $p(\mathbf{s}|\omega_i)$ especially if the dimensionality of the feature vector $\mathbf{s}$ is large. In the absence of sufficient number of training samples (which is typically the case when the multibiometric system is first deployed or if its parameters are subsequently adjusted), it is commonly assumed that the scalar scores $s_i, s_2, \ldots s_R$ are generated by R independent random processes. This assumption permits the density function to be expressed as

$$p(\mathbf{s}|\omega_i) = \prod_{j=1}^{R} p(s_j|\omega_i), \tag{14.6}$$

where the joint-density function is now replaced by the product of its marginals. The marginal densities, $p(s_j|\omega_i)$, $j = 1, 2, \ldots R$, $i = 0, 1$, are estimated from a

[4] This is known as the Bayes decision rule or the minimum-error-rate classification rule under the 0-1 loss function [14]

training set of genuine and impostor scores corresponding to each of the R biometric matchers. Equation (14.6) results in the *product rule* which combines the scores generated by the R matchers as,

$$s_{prod} = \prod_{j=1}^{R} \frac{p(s_j|\omega_0)}{p(s_j|\omega_1)}. \tag{14.7}$$

Kittler et al. [28] modify the product rule by further assuming that the *a posteriori* probability $P(\omega_i|\mathbf{s})$ of class ω_i does not deviate much from its *a priori* probability $P(\omega_i)$ resulting in the *sum rule*:

$$s_{sum} = \frac{\sum_{j=1}^{R} p(s_j|\omega_0)}{\sum_{j=1}^{R} p(s_j|\omega_1)}. \tag{14.8}$$

Similar expressions can be derived for combining the match scores using the max, min and median rules [53, 28]. All the aforementioned rules implicitly assume that the match scores are *continuous* random variables. Dass et al. [11] relax this assumption and represent the univariate density functions (i.e., the marginals in Equation (14.6)) as a mixture of discrete as well as continuous components. The resulting density functions are referred to as generalized densities. The authors demonstrate that the use of generalized density estimates (as opposed to continuous density estimates) significantly enhances the matching performance of the fusion algorithm. Furthermore, they use copula functions [41, 8] to model the correlation structure between the match scores $s_1, s_2, \ldots, s_R$ and, subsequently, define a novel fusion rule known as the *copula fusion rule.*

Transformation-based fusion schemes

Density-based schemes, as stated earlier, require a large number of training samples (i.e., genuine and impostor match scores) in order to accurately estimate the density functions. This may not be possible in most multibiometric systems due to the time, effort and cost involved in acquiring labeled multibiometric data in an operational environment. In such situations, it may be necessary to *directly* combine the match scores generated by multiple matchers using simple fusion operators (such as the simple sum of scores or order statistics) without first interpreting them in a probabilistic framework. However, such an approach is meaningful only when the scores output by the matchers are comparable. To facilitate this, a score normalization process is essential to transform the multiple match scores into a common domain. The process of score normalization entails changing the location and the scale parameters of the underlying match score distributions in order to ensure compatibility between multiple score variables.

Once the match scores output by multiple matchers are transformed into a common domain they can be combined using simple fusion operators

Table 14.1. Summary of score normalization techniques.

Normalization Technique	**Robustness**	**Efficiency**
Min-max	No	High
Decimal scaling	No	High
Z-score	No	High
Median and MAD	Yes	Moderate
Double sigmoid	Yes	High
Tanh-estimators	Yes	High

such as the sum of scores, product of scores or order statistics (e.g., maximum/minimum of scores or median score).

Classifier-based fusion schemes

In the verification mode of operation, the match scores generated by the multiple matchers may be input to a trained pattern classifier, such as a neural network, in order to determine the class label (genuine or impostor). In this approach, the goal is to directly estimate the class rather than to compute an intermediate scalar value. Classifier-based fusion schemes assume the availability of a large representative number of genuine and impostor scores during the training phase of the classifier when its parameters are computed. The component scores do not have to be transformed into a common domain prior to invoking the classifier.

In the biometric literature several classifiers have been used to consolidate the match scores of multiple matchers. Brunelli and Falavigna [4] use a HyperBF network to combine matchers based on voice and face features. Verlinde and Cholet [59] compare the relative performance of three different classifiers, namely, the k-Nearest Neighbor classifier using vector quantization, the decision tree classifier, and a classifier based on the logistic regression model when fusing the match scores originating from three biometric matchers. Experiments on the M2VTS database ([44]) show that the total error rate (sum of the false accept and false reject rates) of the multimodal system is an order of magnitude less than that of the individual matchers. Chatzis et al. [6] use classical k-means clustering, fuzzy clustering and median radial basis function (MRBF) algorithms for fusion at the match score level. The proposed system combines the output of five different face and voice matchers. Each matcher provides a match score and a quality metric indicating the reliability of the match score. These values are concatenated to form a ten-dimensional vector that is input to the classifiers. Ben-Yacoub et al. [2] evaluate a number of classification schemes for fusion including support vector machine (SVM) with polynomial kernels, SVM with Gaussian kernels, C4.5 decision trees, multilayer perceptron, Fisher linear discriminant, and Bayesian classifier. Experimental evaluations on the XM2VTS database ([39]) consisting of 295 subjects suggest the benefit of score level fusion. Bigun et al. [3]

propose a novel algorithm based on the Bayesian classifier that takes into account the estimated accuracy of the individual classifiers (i.e., matchers) during the fusion process. Sanderson and Paliwal [55] use a support vector machine (SVM) to combine the scores of face and speech experts. In order to address noisy input, they design structurally noise-resistant classifiers based on a piece-wise linear classifier and a modified Bayesian classifier. Wang et al. [60] view the match scores obtained from face and iris recognition modules as a two-dimensional feature vector and use Fisher's discriminant analysis and a neural network classifier to classify this match score vector. Ross and Jain [51] use decision tree and linear discriminant classifiers for classifying the match scores pertaining to the face, fingerprint and hand geometry modalities.

14.3.4 Rank-level fusion

When a biometric system operates in the identification mode, the output of the system can be viewed as a ranking of the enrolled identities. In this case, the output indicates the set of possible matching identities sorted in decreasing order of confidence. The goal of rank level fusion schemes is to consolidate the ranks output by the individual biometric subsystems in order to derive a consensus rank for each identity. Ranks provide more insight into the decision-making process of the matcher compared to just the identity of the best match, but they reveal less information than match scores. However, unlike match scores, the rankings output by multiple biometric systems are comparable. As a result, no normalization is needed and this makes rank level fusion schemes simpler to implement compared to the score level fusion techniques.

Let us assume that there are M users enrolled in the database and let the number of matchers be R. Let $r_{j,k}$ be the rank assigned to user k by the j^{th} matcher, $j = 1, \ldots, R$ and $k = 1, \ldots, M$. Let s_k be a statistic computed for user k such that the user with the lowest value of s is assigned the highest consensus (or reordered) rank. Ho et al. [18] describe the following three methods to compute the statistic s.

1. Highest Rank Method: In the highest rank method, each user is assigned the highest rank (minimum r value) as computed by different matchers, i.e., the statistic for user k is

$$s_k = \min_{j=1}^{R} r_{j,k}. \tag{14.9}$$

   Ties are broken randomly to arrive at a strict ranking order. This method is useful only when the number of users is large compared to the number of matchers, which is typically the case in large-scale authentication systems. If this condition is not satisfied, the system will encounter several ties thereby rendering the final ranking uninformative. An advantage of the highest rank method is that it can utilize the strength of each matcher effectively. Even if only one matcher assigns a high rank to the correct identity, it is still very likely that this user will receive a high rank after reordering.

2. Borda Count Method: The Borda count method uses the sum of the ranks assigned by the individual matchers to calculate the value of s, i.e., the statistic for user k is

$$s_k = \sum_{j=1}^{R} r_{j,k}. \tag{14.10}$$

   The magnitude of the Borda count for each user is a measure of the degree of agreement among the different matchers on whether the input belongs to that user. The Borda count method assumes that the ranks assigned to the users by the matchers are statistically independent and that all the matchers perform equally well.

3. Logistic Regression Method: The logistic regression method is a generalization of the Borda count method where a weighted sum of the individual ranks is calculated, i.e., the statistic for user k is

$$s_k = \sum_{j=1}^{R} w_j r_{j,k}. \tag{14.11}$$

   The weight, w_j, to be assigned to the j^{th} matcher, $j = 1, \ldots, R$, is determined by logistic regression [1]. The logistic regression method is useful when the different biometric matchers have significant differences in their accuracies. However, this method requires a training phase to determine the weights.

14.3.5 Decision-level fusion

Many commercial off-the-shelf (COTS) biometric matchers provide access only to the final recognition decision. When such COTS matchers are used to build a multibiometric system, only decision level fusion is feasible. Methods proposed in the literature for decision level fusion include "AND" and "OR" rules [12], majority voting [34], weighted majority voting [30], Bayesian decision fusion [61], the Dempster-Shafer theory of evidence [61] and behavior knowledge space [20].

Let M denote the number of possible decisions (also known as *class labels* or simply *classes* in the pattern recognition literature; these three terms are used interchangeably in the following discussion) in a biometric system. Also, let $\omega_1, \omega_2, \ldots \omega_M$ indicate the classes associated with each of these decisions.

1. "AND" and "OR" Rules: In a multibiometric verification system, the simplest method of combining decisions output by the different matchers is to use the "AND" and "OR" rules. The output of the "AND" rule is a "match" only when all the biometric matchers agree that the input sample matches

with the template. On the contrary, the “OR” rule outputs a “match” decision as long as at least one matcher decides that the input sample matches with the template. The limitation of these two rules is their tendency to result in extreme operating points. When the “AND” rule is applied, the False Accept Rate (FAR) of the multibiometric system is extremely low (lower than the FAR of the individual matchers) while the False Reject Rate (FRR) is high (greater than the FRR of the individual matchers). Similarly, the “OR” rule leads to higher FAR and lower FRR than the individual matchers. When one biometric matcher has a substantially higher equal error rate compared to the other matcher, the combination of the two matchers using “AND” and “OR” rules may actually degrade the overall performance [12]. Due to this phenomenon, the “AND” and “OR” rules are rarely used in practical multibiometric systems.

2. Majority Voting: The most common approach for decision level fusion is majority voting where the input biometric sample is assigned to that identity on which a majority of the matchers agree. If there are R biometric matchers, the input sample is assigned an identity when at least k of the matchers agree on that identity, where

$$k = \begin{cases} \frac{R}{2} + 1 \text{ if } R \text{ is even}, \\ \\ \frac{R+1}{2} \quad \text{otherwise}. \end{cases} \tag{14.12}$$

When none of the identities is supported by k matchers, a reject decision is output by the system. Majority voting assumes that all the matchers perform equally well. The advantages of majority voting are: (i) no apriori knowledge about the matchers is needed, and (ii) no training is required to come up with the final decision. A theoretical analysis of the majority voting fusion scheme was done by [33] who established limits on the accuracy of the majority vote rule based on the number of matchers, the individual accuracy of each matcher and the pairwise dependence between the matchers.

3. Weighted Majority Voting: When the matchers used in a multibiometric system are not of similar recognition accuracy (i.e, imbalanced matchers/classifiers), it is reasonable to assign higher weights to the decisions made by the more accurate matchers. In order to facilitate this weighting, the labels output by the individual matchers are converted into degrees of support for the M classes as follows.

$$s_{j,k} = \begin{cases} 1, \text{ if output of the } j^{th} \text{ matcher is class } \omega_k, \\ 0, \text{ otherwise}, \end{cases} \tag{14.13}$$

where $j = 1, \ldots, R$ and $k = 1, \ldots, M$. The discriminant function[5] for class ω_k computed using weighted voting is

[5] The discriminant function is used to classify an input pattern. Typically, a discriminant function is defined for each pattern class and the input pattern is assigned to the class whose discriminant function gives the maximum response.

$$g_k = \sum_{j=1}^{R} w_j s_{j,k}, \tag{14.14}$$

where w_j is the weight assigned to the j^{th} matcher. A test sample is assigned to the class with the highest score (value of discriminant function).

4. Bayesian Decision Fusion: The Bayesian decision fusion scheme relies on transforming the discrete decision labels output by the individual matchers into continuous probability values. The first step in the transformation is the generation of the confusion matrix for each matcher by applying the matcher to a training set $\mathbf{D}$. Let CM^j be the $M \times M$ confusion matrix for the j^{th} matcher. The (k, r)th element of the matrix CM^j (denoted as $cm^j_{k,r}$) is the number of instances in the training data set where a pattern whose true class label is ω_k is assigned to the class ω_r by the j^{th} matcher. Let the total number of data instances in $\mathbf{D}$ be N and the number of elements that belong to class ω_k be N_k. Let c_j be the class label assigned to the test sample by the j^{th} matcher. The value $cm^j_{k,c_j}/N_k$ can be considered as an estimate of the conditional probability $P\left(c_j|\omega_k\right)$ and N_k/N can be treated as an estimate of the prior probability of class ω_k. Given the vector of decisions made by R matchers $\mathbf{c} = [c_1, \ldots, c_R]$, we are interested in calculating the posterior probability of class ω_k, i.e., $P\left(\omega_k|\mathbf{c}\right)$. According to the Bayes rule,

$$P\left(\omega_k|\mathbf{c}\right) = \frac{P\left(\mathbf{c}|\omega_k\right) P(\omega_k)}{P(\mathbf{x})}, \tag{14.15}$$

where $k = 1, \ldots, M$. The denominator in Equation 14.15 is independent of the class ω_k and can be ignored for the decision making purpose. Therefore, the discriminant function for class ω_k is

$$g_k = P\left(\mathbf{c}|\omega_k\right) P(\omega_k). \tag{14.16}$$

The Bayes decision fusion technique chooses that class which has the largest value of discriminant function calculated using equation 14.16. To simplify the computation of $P\left(\mathbf{c}|\omega_k\right)$, one can assume conditional independence between the different matchers. Under this assumption, the decision rule is known as naive Bayes rule and $P\left(\mathbf{c}|\omega_k\right)$ is computed as

$$P\left(\mathbf{c}|\omega_k\right) = P\left(c_1, \ldots, c_R|\omega_k\right) = \prod_{j=1}^{R} P\left(c_j|\omega_k\right). \tag{14.17}$$

The accuracy of the naive Bayes decision fusion rule has been found to be fairly robust even when the matchers are not independent [13].

5. Dempster-Shafer Theory of Evidence: The Dempster-Shafer theory of evidence is based on the concept of assigning degrees of belief for uncertain events. Note that the degree of belief for an event is different from the probability of the event. This subtle difference is explained in the following example.

Suppose we know that a biometric matcher has a reliability of 0.95, i.e., the output of the matcher is reliable 95% of the time and unreliable 5% of the time. Suppose that the matcher outputs a "match" decision. We can assign a 0.95 degree of belief to the "match" decision and a zero degree of belief to the "non-match" decision. The zero belief does not rule out the "non-match" decision completely, unlike a zero probability. Instead, the zero belief indicates that there is no reason to believe that the input does not match successfully against the template. Hence, we can view belief theory as a generalization of probability theory. Indeed, belief functions are more flexible than probabilities when our knowledge about the problem is incomplete.

Rogova [49] and Kuncheva et al. [31] propose the following methodology to compute the belief functions and to accumulate the belief functions according to the Dempster's rule. For a given input pattern, the decisions made by R classifiers for a M-class problem is represented using a $R \times M$ matrix known as a decision profile (DP) [31] which is given by,

$$DP = \begin{bmatrix} s_{1,1} \cdots s_{1,k} \cdots s_{1,M} \\ \cdots \\ s_{j,1} \cdots s_{j,k} \cdots s_{j,M} \\ \cdots \\ s_{R,1} \cdots s_{R,k} \cdots s_{R,M} \end{bmatrix},$$

where $s_{j,k}$ is the degree of support provided by the j^{th} matcher to the k^{th} class. At the decision level, the degree of support is expressed as

$$s_{j,k} = \begin{cases} 1, \text{ if output of the } j^{th} \text{ matcher is class } \omega_k, \\ 0, \text{ otherwise,} \end{cases} \tag{14.18}$$

where $j = 1, \ldots, R$ and $k = 1, \ldots, M$. The decision template (DT^k) of each class ω_k is the average decision profile for all the training instances that belong to the class ω_k. When the degrees of support defined in Equation 14.18 are used, one can easily see that the elements of the decision template DT^k are related to the elements of the confusion matrices of the R matchers in the following manner.

$$DT^k_{j,r} = \frac{CM^j_{k,r}}{N_k}, \tag{14.19}$$

where N_k is the number of instances in the training set $\mathbf{D}$ that belong to class ω_k, $j = 1, \ldots, R$ and $k, r = 1, \ldots, M$. For a given test pattern X^t, the decision profile DP^t is computed after the decisions of the R matchers are obtained. The similarity between DP^t and the decision templates for the various classes is calculated as follows.

$$\Phi_{j,k} = \frac{\left(1 + \left(||DT^k_j - DP^t_j||\right)^2\right)^{-1}}{\sum_{r=1}^{M}\left(\left(1 + \left(||DT^r_j - DP^t_j||\right)^2\right)^{-1}\right)}, \tag{14.20}$$

where DT_j^k represents the j^{th} row of DT^k belonging to class ω_k, DP_j^t represents the j^{th} row of DP^t belonging to the test pattern X^t, and $||.||$ denotes the matrix norm. For every class $k = 1, \ldots, M$ and for every matcher $j = 1, \ldots, R$, we can compute the degree of belief as

$$b_{j,k} = \frac{\Phi_{j,k}\left[\prod_{r=1,r\neq k}^{M}(1-\Phi_{j,r})\right]}{1-\Phi_{j,k}\left[\prod_{r=1,r\neq k}^{M}(1-\Phi_{j,r})\right]}. \tag{14.21}$$

The accumulated degree of belief for each class $k = 1, \ldots, M$ based on the outputs of R matchers is then obtained using the Dempster's rule as

$$g_k = \prod_{j=1}^{R} b_{j,k}. \tag{14.22}$$

The test pattern X^t is assigned to the class having the highest degree of belief g_k.

14.4 Summary

Multibiometric systems are expected to enhance the recognition accuracy of a personal authentication system by reconciling the evidence presented by multiple sources of information. In this chapter, the different sources of biometric information as well as the type of information that can be consolidated was presented. Different fusion strategies were also discussed. Typically, early integration strategies (e.g., feature-level) are expected to result in better performance than late integration (e.g., score-level) strategies. However, it is difficult to predict the performance gain due to each of these strategies prior to invoking the fusion methodology. While the *availability* of multiple sources of biometric information (pertaining either to a single trait or to multiple traits) may present a compelling case for fusion, the *correlation* between the sources has to be examined before determining their suitability for fusion. Combining uncorrelated or negatively correlated sources is expected to result in a better improvement in matching performance than combining positively correlated sources. This has been demonstrated by Kuncheva et al. [32] for fusion at the decision level using the majority vote scheme. Combining sources that make complementary errors is assumed to be beneficial. However, defining an appropriate diversity measure to predict fusion performance has been elusive thus far.

References

1. A. Agresti. *An Introduction to Categorical Data Analysis*. Wiley, 1996.

2. S. Ben-Yacoub, Y. Abdeljaoued, and E. Mayoraz. Fusion of Face and Speech data for Person Identity Verification. *IEEE Transactions on Neural Networks*, 10(5):1065–1075, September 1999.
3. E. S. Bigun, J. Bigun, B. Duc, and S. Fischer. Expert Conciliation for Multimodal Person Authentication Systems using Bayesian Statistics. In *First International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA)*, pages 291–300, Crans-Montana, Switzerland, March 1997.
4. R. Brunelli and D. Falavigna. Person Identification Using Multiple Cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(10):955–966, October 1995.
5. K. I. Chang, K. W. Bowyer, and P. J. Flynn. An Evaluation of Multimodal 2D+3D Face Biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):619–624, April 2005.
6. V. Chatzis, A. G. Bors, and I. Pitas. Multimodal Decision-level Fusion for Person Authentication. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 29(6):674–681, November 1999.
7. X. Chen, P. J. Flynn, and K. W. Bowyer. IR and Visible Light Face Recognition. *Computer Vision and Image Understanding*, 99(3):332–358, September 2005.
8. U. Cherubini, E. Luciano, and W. Vecchiato. *Copula Methods in Finance.* Wiley, 2004.
9. C. C. Chibelushi, F. Deravi, and J. S. Mason. Voice and Facial Image Integration for Speaker Recognition. In R. I. Damper, W. Hall, and J. W. Richards, editors, *Multimedia Technologies and Future Applications*, pages 155–161. Pentech Press, London, 1994.
10. C. C. Chibelushi, J. S. D. Mason, and F. Deravi. Feature-level Data Fusion for Bimodal Person Recognition. In *Proceedings of the Sixth International Conference on Image Processing and Its Applications*, volume 1, pages 399–403, Dublin, Ireland, July 1997.
11. S. C. Dass, K. Nandakumar, and A. K. Jain. A Principled Approach to Score Level Fusion in Multimodal Biometric Systems. In *Proceedings of Fifth International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA)*, pages 1049–1058, Rye Brook, USA, July 2005.
12. J. Daugman. Combining Multiple Biometrics. Available at `http://www.cl.cam.ac.uk/users/jgd1000/combine/combine.html`, 2000.
13. P. Domingos and M. Pazzani. On the Optimality of the Simple Bayesian Classifier under Zero-One Loss. *Machine Learning*, 29(2-3):103–130, November/December 1997.
14. R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification.* John Wiley & Sons, 2001.
15. A. Eriksson and P. Wretling. How Flexible is the Human Voice? A Case Study of Mimicry. In *Proceedings of the European Conference on Speech Technology*, pages 1043–1046, Rhodes, 1997.
16. W. R. Harrison. *Suspect Documents, their Scientific Examination.* Nelson-Hall Publishers, 1981.
17. H. Hill, P. G. Schyns, and S. Akamatsu. Information and Viewpoint Dependence in Face Recognition. *Cognition*, 62(2):201–222, February 1997.
18. T. K. Ho, J. J. Hull, and S. N. Srihari. Decision Combination in Multiple Classifier Systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(1):66–75, January 1994.

19. L. Hong, A. K. Jain, and S. Pankanti. Can Multibiometrics Improve Performance? In *Proceedings of IEEE Workshop on Automatic Identification Advanced Technologies (AutoID)*, pages 59–64, New Jersey, USA, October 1999.
20. Y. S. Huang and C. Y. Suen. Method of Combining Multiple Experts for the Recognition of Unconstrained Handwritten Numerals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(1):90–94, January 1995.
21. A. K. Jain and B. Chandrasekaran. Dimensionality and Sample Size Considerations in Pattern Recognition Practice. In P.R. Krishnaiah and L. N. Kanal, editors, *Handbook of Statistics*, volume 2, pages 835–855. North-Holland, Amsterdam, 1982.
22. A. K. Jain, R. P. W. Duin, and J. Mao. Statistical Pattern Recognition: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4–37, January 2000.
23. A. K. Jain, K. Nandakumar, X. Lu, and U. Park. Integrating Faces, Fingerprints and Soft Biometric Traits for User Recognition. In *Proceedings of ECCV International Workshop on Biometric Authentication (BioAW)*, volume LNCS 3087, pages 259–269, Prague, Czech Republic, May 2004. Springer.
24. A. K. Jain and A. Ross. Fingerprint Mosaicking. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 4, pages 4064–4067, Orlando, USA, May 2002.
25. A. K. Jain and A. Ross. Multibiometric Systems. *Communications of the ACM, Special Issue on Multimodal Interfaces*, 47(1):34–40, January 2004.
26. A. K. Jain and D. Zongker. Feature Selection: Evaluation, Application, and Small Sample Performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2):153–158, February 1997.
27. J. Jang, K. R. Park, J. Son, and Y. Lee. Multi-unit Iris Recognition System by Image Check Algorithm. In *Proceedings of International Conference on Biometric Authentication (ICBA)*, pages 450–457, Hong Kong, July 2004.
28. J. Kittler, M. Hatef, R. P. Duin, and J. G. Matas. On Combining Classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239, March 1998.
29. A. Kong, J. Heo, B. Abidi, J. Paik, and M. Abidi. Recent Advances in Visual and Infrared Face Recognition - A Review. *Computer Vision and Image Understanding*, 97(1):103–135, January 2005.
30. L. I. Kuncheva. *Combining Pattern Classifiers - Methods and Algorithms*. Wiley, 2004.
31. L. I. Kuncheva, J. C. Bezdek, and R. P. W. Duin. Decision Templates for Multiple Classifier Fusion: An Experimental Comparison. *Pattern Recognition*, 34(2):299–314, 2001.
32. L. I. Kuncheva, C. J. Whitaker, C. A. Shipp, and R. P. W. Duin. Is Independence Good for Combining Classifiers? In *Proceedings of International Conference on Pattern Recognition (ICPR)*, volume 2, pages 168–171, Barcelona, Spain, 2000.
33. L. I. Kuncheva, C. J. Whitaker, C. A. Shipp, and R. P. W. Duin. Limits on the Majority Vote Accuracy in Classifier Fusion. *Pattern Analysis and Applications*, 6(1):22–31, 2003.
34. L. Lam and C. Y. Suen. Application of Majority Voting to Pattern Recognition: An Analysis of its Behavior and Performance. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 27(5):553–568, 1997.

35. J. Lee, B. Moghaddam, H. Pfister, and R. Machiraju. Finding Optimal Views for 3D Face Shape Modeling. In *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, pages 31–36, Seoul, Korea, May 2004.
36. X. Lu, Y. Wang, and A. K. Jain. Combining Classifiers for Face Recognition. In *IEEE International Conference on Multimedia and Expo (ICME)*, volume 3, pages 13–16, Baltimore, USA, July 2003.
37. G. L. Marcialis and F. Roli. Fingerprint Verification by Fusion of Optical and Capacitive Sensors. *Pattern Recognition Letters*, 25(11):1315–1322, August 2004.
38. T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino. Impact of Artificial Gummy Fingers on Fingerprint Systems. In *Optical Security and Counterfeit Deterrence Techniques IV, Proceedings of SPIE*, volume 4677, pages 275–289, San Jose, USA, January 2002.
39. K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre. XM2VTSDB: The Extended M2VTS Database. In *Proceedings of Second International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA)*, pages 72–77, Washington D.C., USA, March 1999.
40. National Institute of Standards and Technology. NIST Biometric Scores Set. Available at `http://http://www.itl.nist.gov/iad/894.03/biometricscores`, 2004.
41. R. B. Nelsen. *An Introduction to Copulas*. Springer, 1999.
42. A. O'Toole, H. Bulthoff, N. Troje, and T. Vetter. Face Recognition across Large Viewpoint Changes. In *Proceedings of the International Workshop on Automatic Face- and Gesture-Recognition (IWAFGR)*, pages 326–331, Zurich, Switzerland, June 1995.
43. Z. Pan, G. Healey, M. Prasad, and B. Tromberg. Face Recognition in Hyperspectral Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1552–1560, December 2003.
44. S. Pigeon and L. Vandendrope. M2VTS Multimodal Face Database Release 1.00. Available at `http://www.tele.ucl.ac.be/PROJECTS/M2VTS/m2fdb.html`, 1996.
45. S. Prabhakar and A. K. Jain. Decision-level Fusion in Fingerprint Verification. Technical Report MSU-CSE-00-24, Michigan State University, October 2000.
46. P. Pudil, J. Novovicova, and J. Kittler. Floating Search Methods in Feature Selection. *Pattern Recognition Letters*, 15(11):1119–1124, November 1994.
47. T. Putte and J. Keuning. Biometrical Fingerprint Recognition: Don't Get Your Fingers Burned. In *Proceedings of IFIP TC8/WG8.8 Fourth Working Conference on Smart Card Research and Advanced Applications*, pages 289–303, 2000.
48. N. K. Ratha, J. H. Connell, and R. M. Bolle. An Analysis of Minutiae Matching Strength. In *Proceedings of Third International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA)*, pages 223–228, Halmstad, Sweden, June 2001.
49. G. Rogova. Combining the Results of Several Neural Network Classifiers. *Neural Networks*, 7(5):777–781, 1994.
50. A. Ross and R. Govindarajan. Feature Level Fusion Using Hand and Face Biometrics. In *Proceedings of SPIE Conference on Biometric Technology for Human Identification II*, volume 5779, pages 196–204, Orlando, USA, March 2005.

51. A. Ross and A. K. Jain. Information Fusion in Biometrics. *Pattern Recognition Letters*, 24(13):2115–2125, September 2003.
52. A. Ross, A. K. Jain, and J. Reisman. A Hybrid Fingerprint Matcher. *Pattern Recognition*, 36(7):1661–1673, July 2003.
53. A. Ross, K. Nandakumar, and A. K. Jain. *Handbook of Multibiometrics*. Springer, New York, USA, 1st edition, 2006.
54. R. K. Rowe and K. A. Nixon. Fingerprint Enhancement Using a Multispectral Sensor. In *Proceedings of SPIE Conference on Biometric Technology for Human Identification II*, volume 5779, pages 81–93, March 2005.
55. C. Sanderson and K. K. Paliwal. Information Fusion and Person Verification Using Speech and Face Information. Research Paper IDIAP-RR 02-33, IDIAP, September 2002.
56. D. W. Scott. *Multivariate Density Estimation: Theory, Practice and Visualization*. Wiley Series in Probability and Statistics. Wiley-Interscience, August 1992.
57. D. A. Socolinsky, A. Selinger, and J. D. Neuheisel. Face Recognition with Visible and Thermal Infrared Imagery. *Computer Vision and Image Understanding*, 91(1-2):72–114, July-August 2003.
58. U. Uludag, A. Ross, and A. K. Jain. Biometric Template Selection and Update: A Case Study in Fingerprints. *Pattern Recognition*, 37(7):1533–1542, July 2004.
59. P. Verlinde and 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 Audio- and Video-Based Biometric Person Authentication (AVBPA)*, pages 188–193, Washington D.C., USA, March 1999.
60. Y. Wang, T. Tan, and A. K. Jain. Combining Face and Iris Biometrics for Identity Verification. In *Fourth International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA)*, pages 805–813, Guildford, UK, June 2003.
61. L. Xu, A. Krzyzak, and C. Y. Suen. Methods for Combining Multiple Classifiers and their Applications to Handwriting Recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(3):418–435, 1992.