This is to certify that the
thesis entitled

INTEGRATION OF MULTIPLE CUES IN BIOMETRIC SYSTEMS

presented by

KARTHIK NANDAKUMAR

has been accepted towards fulfillment
of the requirements for the

M. S. degree in COMPUTER SCIENCE AND ENGINEERING

Anil Jain

Major Professor's Signature

May 13, 2005

Date

INTEGRATION OF MULTIPLE CUES IN BIOMETRIC SYSTEMS

By

Karthik Nandakumar

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Department of Computer Science and Engineering

2005

ABSTRACT

INTEGRATION OF MULTIPLE CUES IN BIOMETRIC SYSTEMS

By

Karthik Nandakumar

Biometrics refers to the automatic recognition of individuals based on their physiological and/or behavioral characteristics. Unimodal biometric systems perform person recognition based on a single source of biometric information and are affected by problems like noisy sensor data, non-universality and lack of individuality of the chosen biometric trait, absence of an invariant representation for the biometric trait and susceptibility to circumvention. Some of these problems can be alleviated by using multimodal biometric systems that consolidate evidence from multiple biometric sources. Integration of evidence obtained from multiple cues is a challenging problem and integration at the matching score level is the most common approach because it offers the best trade-off between the information content and the ease in fusion. In this thesis, we address two important issues related to score level fusion. 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 fusion. We have studied the performance of different normalization techniques and fusion rules using a multimodal biometric system based on face, fingerprint and hand-geometry modalities. The normalization schemes have been evaluated both in terms of their efficiency and robustness to the presence of outliers in the training data. We have also demonstrated how soft biometric attributes like gender, ethnicity, accent and height, that by themselves do not have sufficient discriminative ability to reliably recognize a person, can be used to improve the recognition accuracy of the primary biometric identifiers (e.g., fingerprint and face). We have developed a mathematical model based on Bayesian decision theory for integrating the primary and soft biometric cues at the score level.

To My Grandfather and Grandmother

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my grandfather Shri. P.S. Venkatachari and my grandmother Smt. P.S. Vanaja for all their prayers and blessings. Without their support and encouragement at crucial periods of my life, it would not have been possible for me to pursue graduate studies and aim for greater things in my career. Their hard work and positive attitude even at an age that is soon approaching 80, is my main source of inspiration. I am proud to dedicate this thesis and all the good things in my life to them.

I am grateful to my advisor Dr. Anil Jain for providing me the opportunity to work in an exciting and challenging field of research. His constant motivation, support and infectious enthusiasm has guided me towards the successful completion of this thesis. My interactions with him have been of immense help in defining my research goals and in identifying ways to achieve them. His encouraging words have often pushed me to put in my best possible efforts. I am hopeful that my Ph.D. experience with him would continue to be as memorable and fruitful as my Masters work.

I also thank my guidance committee members Dr. Bill Punch and Dr. Arun Ross for spending their time in carefully reviewing this thesis. Their valuable comments and suggestions have been very useful in enhancing the presentation of this thesis. Special thanks goes to Dr. Arun Ross for his guidance in the score normalization research. I am also indebted to Dr. Sarat Dass for the enlightening research discussions that have shown me the right path on many occasions, especially in the soft biometrics research.

The PRIP lab is an excellent place to work in and it is one place where you can always find good company, no matter what time of the day it is. My interactions with members of this lab has certainly made me a better professional. Special thanks goes to Umut Uludag for helping me acclimatize to the lab during the initial months. I would also like to thank

Xiaoguang Lu and Unsang Park for their assistance in the soft biometrics project and Martin Law who is always ready to help in case of any problems. Finally, I would like to Steve Krawczyk for having taken the time to proof-read my thesis.

I would like to thank my friends Mahesh Arumugam and Arun Prabakaran for their great company during my stay at MSU. I am also grateful to my friends Mahadevan Balakrishnan, Jayaram Venkatesan, Hariharan Rajasekaran and Balasubramanian Rathakrishnan for all their long conversations over phone that helped me feel at home.

Finally, I would like to thank my parents who have been the pillar of strength in all my endeavors. I am always deeply indebted to them for all that they have given me. I also thank the other members of my family including my brother and two sisters for their love, affection and timely help.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Biometric Systems

Identity management refers to the challenge of providing authorized users with secure and easy access to information and services across a variety of networked systems. A reliable identity management system is a critical component in several applications that render their services only to legitimate users. Examples of such applications include physical access control to a secure facility, e-commerce, access to computer networks and welfare distribution. The primary task in an identity management system is the determination of an individual's identity. Traditional methods of establishing a person's identity include knowledge-based (e.g., passwords) and token-based (e.g., ID cards) mechanisms. These surrogate representations of the identity can easily be lost, shared or stolen. Therefore, they are not sufficient for identity verification in the modern day world. Biometrics offers a natural and reliable solution to the problem of identity determination by recognizing individuals based on their physiological and/or behavioral characteristics that are inherent to the person [1].

Some of the physiological and behavioral characteristics that are being used for biometric recognition include fingerprint, hand-geometry, iris, retina, face, palmprint, ear, DNA, voice, gait, signature and keystroke dynamics (see Figure 1.1). A typical biometric system consists of four main modules. The sensor module is responsible for acquiring the biometric data from an individual. The feature extraction module processes the acquired biometric data and extracts only the salient information to form a new representation of the

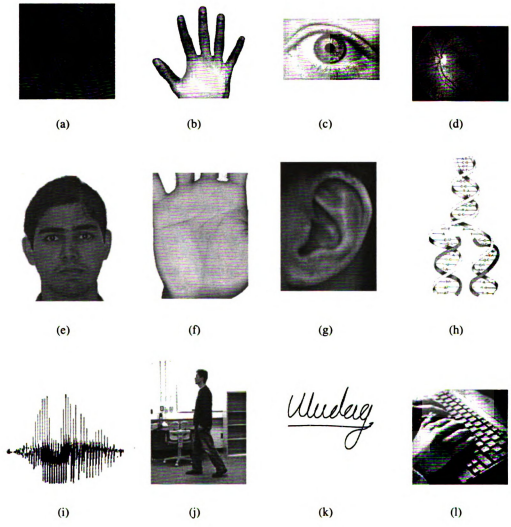


Figure 1.1: Characteristics that are being used for biometric recognition; (a) Fingerprint; (b) Hand-geometry; (c) Iris; (d) Retina; (e) Face; (f) Palmprint; (g) Ear structure; (h) DNA; (i) Voice; (j) Gait; (k) Signature and (l) Keystroke dynamics.

data. Ideally, this new representation should be unique for each person and also relatively invariant with respect to changes in the different samples of the same biometric collected from the same person. The matching module compares the extracted feature set with the templates stored in the system database and determines the degree of similarity (dissimilarity) between the two. The decision module either verifies the identity claimed by the user or determines the user's identity based on the degree of similarity between the extracted features and the stored template(s).

Biometric systems can provide three main functionalities, namely, (i) verification, (ii) identification and (iii) negative identification. Figure 1.2 shows the flow of information in verification and identification systems. In verification or authentication, the user claims an identity and the system verifies whether the claim is genuine. For example, in an ATM application the user may claim a specific identity, say John Doe, by entering his Personal Identification Number (PIN). The system acquires the biometric data from the user and compares it only with the template of John Doe. Thus, the matching is 1:1 in a verification system. If the user's input and the template of the claimed identity have a high degree of similarity, then the claim is accepted as "genuine". Otherwise, the claim is rejected and the user is considered an "impostor". In short, a biometric system operating in the verification mode answers the question "Are you who you say you are?".

In a biometric system used for identification, the user does not explicitly claim an identity. However, the implicit claim made by the user is that he is one among the persons already enrolled in the system. In identification, the user's input is compared with the templates of all the persons enrolled in the database and the identity of the person whose template has the highest degree of similarity with the user's input is output by the biometric system. Typically, if the highest similarity between the input and all the templates is less than a fixed minimum threshold, the system outputs a reject decision which implies that

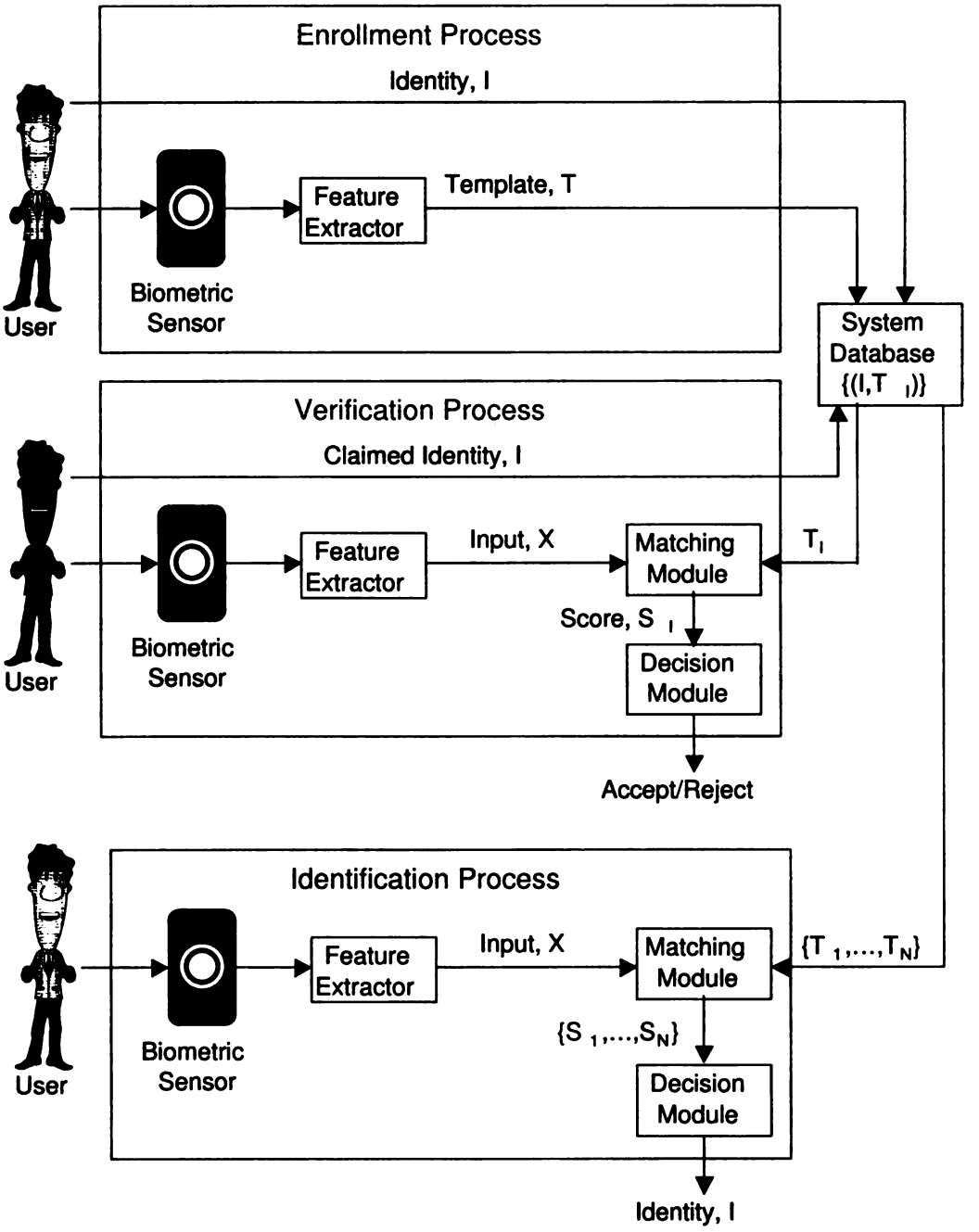


Figure 1.2: Information flow in biometric systems.

the user presenting the input is not one among the enrolled users. Therefore, the matching is 1:N in an identification system. An example of an identification system could be access control to a secure building. All users who are authorized to enter the building would be enrolled in the system. Whenever a user tries to enter the building, he presents his biometric data to the system and upon determination of the user's identity the system grants him the preset access privileges. An identification system answers the question "Are you really someone who is known to the system?".

Negative identification systems are similar to identification systems because the user does not explicitly claim an identity. The main factor that distinguishes the negative identification functionality from identification is the user's implicit claim that he is *not* a person who is already enrolled in the system. Negative identification is also known as screening. As in identification, the matching is 1:N in screening. However in screening, the system will output an identity of an enrolled person only if that person's template has the highest degree of similarity with the input among all the templates and if the corresponding similarity value is greater than a fixed threshold. Otherwise, the user's claim that he is not already known to the system is accepted. Screening is often used at airports to verify whether a passenger's identity matches with any person on a "watch-list". Screening can also be used to prevent the issue of multiple credential records (e.g., driver's licence, passport) to the same person. To summarize, negative identification answers the question "Are you who you say you are not?".

Verification functionality can be provided by traditional methods like passwords and ID cards as well as by biometrics. The negative identification functionality can be provided only by biometrics. Further, biometric characteristics are inherent to the person whose identity needs to be established. Hence, they cannot be lost, stolen, shared, or forgotten. Therefore, biometric traits provide more security than traditional knowledge-based or

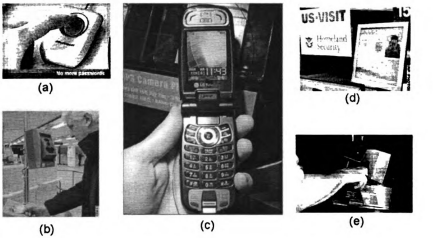


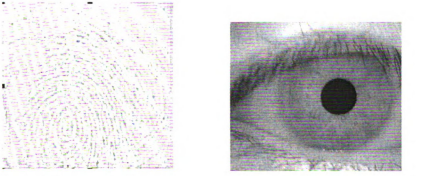
Figure 1.3: Some illustrations of deployment of biometrics in civilian applications; (a) A fingerprint verification system manufactured by Digital Persona Inc. used for computer and network login; (b) An iris-based access control system at the Umea airport in Sweden that verifies the frequent travelers and allows them access to flights; (c) A cell phone manufactured by LG Electronics that recognizes authorized users using fingerprints (sensors manufactured by Authentec Inc.) and allows them access to the phone's special functionalities such as mobile-banking; (d) The US-VISIT immigration system based on fingerprint and face recognition technologies and (e) A hand geometry system at Disney World that verifies seasonal and yearly pass-holders to allow them fast entry.

token-based identification methods. They also discourage fraud and eliminate the possibility of repudiation. Finally, they are more convenient to use because it eliminates the need for remembering multiple complex passwords and carrying identification cards. Although biometric systems have some limitations [2], they offer a number of advantages over traditional security methods and this has led to their widespread deployment in a variety of civilian applications. Figure 1.3 shows some examples of biometrics deployment in civilian applications.

1.2 Multimodal Biometric Systems

1.2.1 Why Multimodal Biometrics?

Unimodal biometric systems perform person recognition based on a single source of biometric information. Such systems are often affected by the following problems [3]:



(a) (b)

Figure 1.4: Examples of noisy biometric data; (a) A noisy fingerprint image due to smearing, residual deposits, etc.; (b) A blurred iris image due to loss of focus.

(a) (b) (c)

Figure 1.5: Three impressions of a user's finger showing the poor quality of the ridges.

- Noisy sensor data : Noise can be present in the acquired biometric data mainly due to defective or improperly maintained sensors. For example, accumulation of dirt or the residual remains on a fingerprint sensor can result in a noisy fingerprint image as shown in Figure 1.4(a). Failure to focus the camera appropriately can lead to blurring in face and iris images (see Figure 1.4(b)). The recognition accuracy of a biometric system is highly sensitive to the quality of the biometric input and noisy data can result in a significant reduction in the accuracy of the biometric system [4].

- Non-universality: If every individual in the target population is able to present the biometric trait for recognition, then the trait is said to be universal. Universality is one of the basic requirements for a biometric identifier. However, not all biometric traits are truly universal. The National Institute of Standards and Technology (NIST) has reported that it is not possible to obtain a good quality fingerprint from approximately two percent of the population (people with hand-related disabilities, manual workers with many cuts and bruises on their fingertips, and people with very oily or dry fingers) [5] (see Figure 1.5). Hence, such people cannot be enrolled in a fingerprint verification system. Similarly, persons having long eye-lashes and those

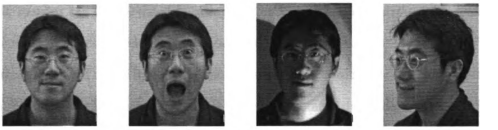
suffering from eye abnormalities or diseases like glaucoma, cataract, aniridia, and nystagmus cannot provide good quality iris images for automatic recognition [6]. Non-universality leads to Failure to Enroll (FTE) and/or Failure to Capture (FTC) errors in a biometric system.

- Lack of individuality: Features extracted from biometric characteristics of different individuals can be quite similar. For example, appearance-based facial features that are commonly used in most of the current face recognition systems are found to have limited discrimination capability [7]. A small proportion of the population can have nearly identical facial appearance due to genetic factors (e.g., father and son, identical twins, etc.). This lack of uniqueness increases the False Match Rate (FMR) of a biometric system.

- Lack of invariant representation: The biometric data acquired from a user during verification will not be identical to the data used for generating the user's template during enrollment. This is known as "intra-class variation". The variations may be due to improper interaction of the user with the sensor (e.g., changes due to rotation, translation and applied pressure when the user places his finger on a fingerprint sensor, changes in pose and expression when the user stands in front of a camera, etc.), use of different sensors during enrollment and verification, changes in the ambient environmental conditions (e.g., illumination changes in a face recognition system) and inherent changes in the biometric trait (e.g., appearance of wrinkles due to aging or presence of facial hair in face images, presence of scars in a fingerprint, etc.). Figure 1.6 shows the intra-class variations in face images caused due to expression, lighting and pose changes. Ideally, the features extracted from the biometric data must be relatively invariant to these changes. However, in most practical biometric

systems the features are not invariant and therefore complex matching algorithms are required to take these variations into account. Large intra-class variations usually increase the False Non-Match Rate (FNMR) of a biometric system.

- Susceptibility to circumvention: Although it is very difficult to steal someone's biometric traits, it is still possible for an impostor to circumvent a biometric system using spoofed traits. Studies [8,9] have shown that it is possible to construct gummy fingers using lifted fingerprint impressions and utilize them to circumvent a biometric system. Behavioral traits like signature and voice are more susceptible to such attacks than physiological traits. Other kinds of attacks can also be launched to circumvent a biometric system [10].



(a) (b) (c) (d)

Figure 1.6: Four face images of the person in (a), exhibiting variations in (b) expression, (c) illumination and (d) pose.

Due to these practical problems, the error rates associated with unimodal biometric systems are quite high which makes them unacceptable for deployment in security critical applications. The state-of-the-art error rates associated with fingerprint, face and voice

Table 1.1: State-of-the-art error rates associated with fingerprint, face and voice biometric systems. Note that the accuracy estimates of biometric systems are dependent on a number of test conditions.

	Test	Test Parameter	False Reject Rate	False Accept Rate
Fingerprint	FVC 2004 [11]	Exaggerated skin distortion, rotation	2%	2%
	FpVTE 2003 [12]	U.S. government operational data	0.1%	1%
Face	FRVT 2002 [13]	Varied lighting, outdoor/indoor	10%	1%
Voice	NIST 2004 [14]	Text independent, multi-lingual	5-10%	2-5%

biometric systems are shown in Table 1.1. Some of the problems that affect unimodal biometric systems can be alleviated by using multimodal biometric systems [15]. Systems that consolidate cues obtained from two or more biometric sources for the purpose of person recognition are called multimodal biometric systems. Multimodal biometric systems have several advantages over unimodal systems. Combining the evidence obtained from different modalities using an effective fusion scheme can significantly improve the overall accuracy of the biometric system. A multimodal biometric system can reduce the FTE/FTC rates and provide more resistance against spoofing because it is difficult to simultaneously spoof multiple biometric sources. Multimodal systems can also provide the capability to search a large database in an efficient and fast manner. This can be achieved by using a relatively simple but less accurate modality to prune the database before using the more complex and accurate modality on the remaining data to perform the final identification task. However, multimodal biometric systems also have some disadvantages. They are more expensive and require more resources for computation and storage than unimodal biometric

systems. Multimodal systems generally require more time for enrollment and verification causing some inconvenience to the user. Finally, the system accuracy can actually degrade compared to the unimodal system if a proper technique is not followed for combining the evidence provided by the different modalities. However, the advantages of multimodal systems far outweigh the limitations and hence, such systems are being increasingly deployed in security-critical applications.

1.2.2 How to Integrate Information?

The design of a multimodal biometric system is strongly dependent on the application scenario. A number of multimodal biometric systems have been proposed in literature that differ from one another in terms of their architecture, the number and choice of biometric modalities, the level at which the evidence is accumulated, and the methods used for the integration or fusion of information. Chapter 2 presents a detailed discussion on the design of a multimodal biometric system.

Four levels of information fusion are possible in a multimodal biometric system. They are fusion at the sensor level, feature extraction level, matching score level and decision level. Sensor level fusion is quite rare because fusion at this level requires that the data obtained from the different biometric sensors must be compatible, which is seldom the case with biometric sensors. Fusion at the feature level is also not always possible 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. Therefore, integration at the matching score level is generally preferred due to the presence of sufficient information content and 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 [16] have shown that the combination approach performs better than some classification methods like decision tree and linear discriminant analysis. However, it must be noted that no single classification or combination scheme works well under all circumstances.

1.3 Thesis Contributions

A review of the proposed multimodal systems indicates that the major challenge in multimodal biometrics is the problem of choosing the right methodology to integrate or fuse the information obtained from multiple sources. In this thesis, we deal with two important problems related to score level fusion.

- In the first part of the thesis, we follow the combination approach to score level 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 thesis, 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. Apart from studying the efficiency of the normalization schemes, we have also analyzed their robustness to the presence of outliers in the training data.

- The second part of the thesis proposes a solution to the problem of integrating the information obtained from the soft biometric identifiers like gender, ethnicity and height with the primary biometric information like face and fingerprint. A mathematical model based on the Bayesian decision theory has been developed to perform the integration. Experiments based on this model demonstrate that soft biometric identifiers can be used to significantly improve the recognition performance of the primary biometric system even when the soft biometric identifiers cannot be extracted with 100% accuracy.

CHAPTER 2

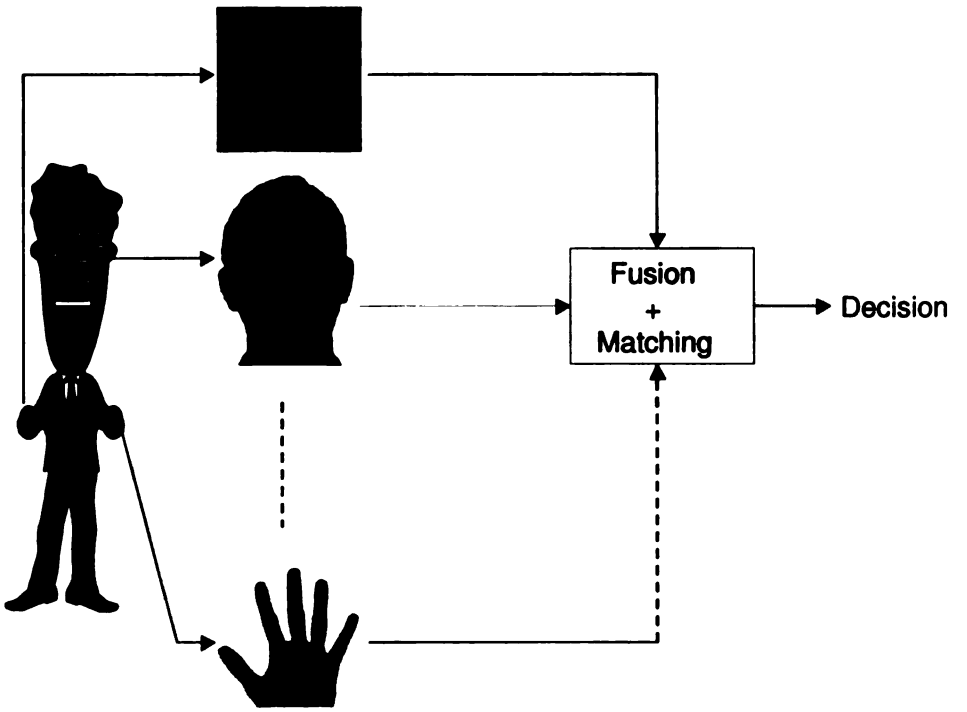
Information Fusion in Multimodal Biometrics

Multimodal biometric systems that have been proposed in literature can be classified based on four parameters, namely, (i) architecture, (ii) sources that provide multiple evidence, (iii) level of fusion and (iv) methodology used for integrating the multiple cues. Generally, these design decisions depend on the application scenario and these choices have a profound influence on the performance of a multimodal biometric system. In this chapter, we compare the existing multibiometric systems based on the above four parameters.

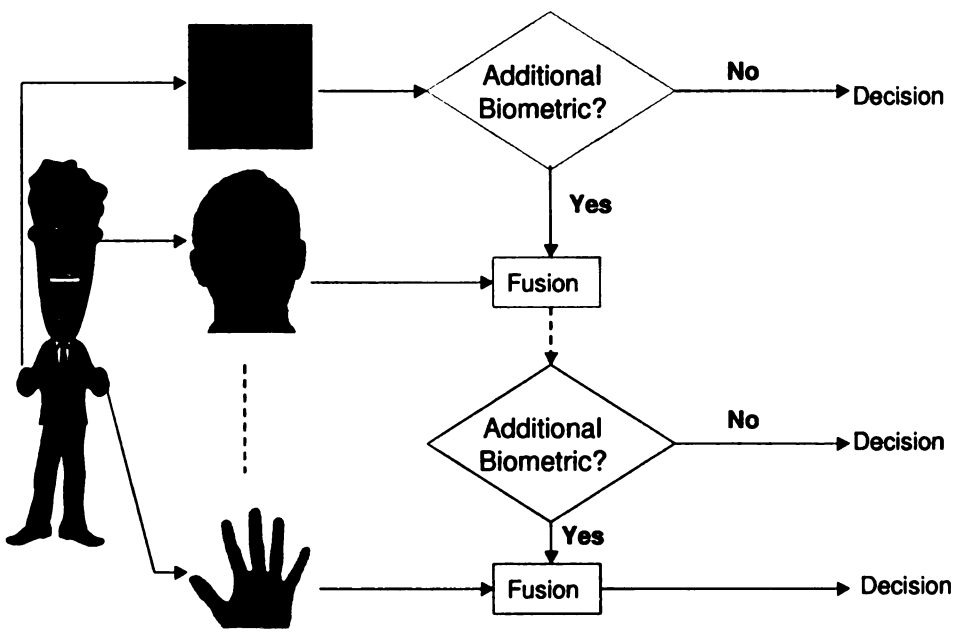
2.1 Architecture

Architecture of a multibiometric system refers to the sequence in which the multiple cues are acquired and processed. Typically, the architecture of a multimodal biometric system is either serial or parallel (see Figure 2.1). In the serial or cascade architecture, the processing of the modalities takes place sequentially and the outcome of one modality affects the processing of the subsequent modalities. In the parallel design, different modalities operate independently and their results are combined using an appropriate fusion scheme. Both these architectures have their own advantages and limitations.

The cascading scheme can improve the user convenience as well as allow fast and efficient searches in large scale identification tasks. For example, when a cascaded multimodal biometric system has sufficient confidence on the identity of the user after processing the first modality, the user may not be required to provide the other modalities. The system can also allow the user to decide which modality he/she would present first. Finally, if the



(a)



(b)

Figure 2.1: Architecture of multimodal biometric systems; (a) Serial and (b) Parallel.

system is faced with the task of identifying the user from a large database, it can utilize the outcome of each modality to successively prune the database, thereby making the search faster and more efficient. Thus, a cascaded system can be more convenient to the user and generally requires less recognition time when compared to its parallel counterpart. However, it requires robust algorithms to handle the different sequence of events. An example of a cascaded multibiometric system is the one proposed by Hong and Jain in [17]. In this system, face recognition is used to retrieve the top n matching identities and fingerprint recognition is used to verify these identities and make a final identification decision. A multimodal system designed to operate in the parallel mode generally has a higher accuracy because it utilizes more evidence about the user for recognition. Most of the proposed multibiometric systems have a parallel architecture because the primary goal of system designers has been a reduction in the error rate of biometric systems (see [16], [18] and the references therein).

The choice of the system architecture depends on the application requirements. User-friendly and less security critical applications like bank ATMs can use a cascaded multimodal biometric system. On the other hand, parallel multimodal systems are more suited for applications where security is of paramount importance (e.g., access to military installations). It is also possible to design a hierarchical (tree-like) architecture to combine the advantages of both cascade and parallel architectures. This hierarchical architecture can be made dynamic so that it is robust and can handle problems like missing and noisy biometric samples that arise in biometric systems. But the design of a hierarchical multibiometric system has not received much attention from researchers.

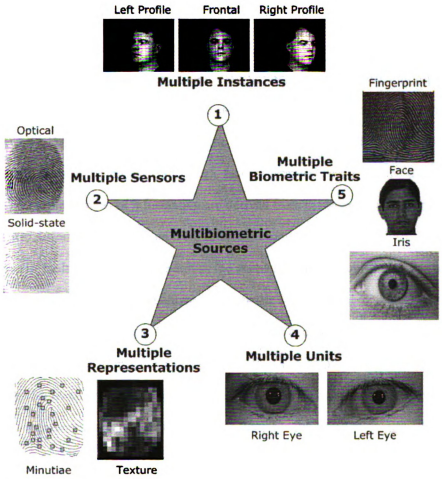


Figure 2.2: Sources of multiple evidence in multimodal biometric systems. In the first four scenarios, multiple sources of information are derived from the same biometric trait. In the fifth scenario, information is derived from different biometric traits.

2.2 Sources of Multiple Evidence

Multimodal biometric systems overcome some of the limitations of unimodal biometric systems by consolidating the evidence obtained from different sources (see Figure 2.2). These sources may be (i) multiple sensors for the same biometric (e.g., optical and solid-state fingerprint sensors), (ii) multiple instances of the same biometric (e.g., multiple face images of a person obtained under different pose/lighting conditions), (iii) multiple representations and matching algorithms for the same biometric (e.g., multiple face matchers like PCA and LDA), (iv) multiple units of the same biometric (e.g., left and right iris images), or (v) multiple biometric traits (e.g., face, fingerprint and iris). In the first four scenarios, multiple sources of information are derived from the same biometric trait. In the fifth scenario, information is derived from different biometric traits.

The use of multiple sensors can address the problem of noisy sensor data, but all other potential problems associated with unimodal biometric systems remain. A recognition system that works on multiple units 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 instances 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 some 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.

2.3 Levels of Fusion

Fusion in multimodal biometric systems can take place at four major levels, namely, sensor level, feature level, score level and decision level. Figure 2.3 shows some examples of fusion at the various levels. These four levels can be broadly categorized into fusion prior to matching and fusion after matching [19].

2.3.1 Fusion Prior to Matching

Prior to matching, integration of information can take place either at the sensor level or at the feature level. The raw data from the sensor(s) are combined in *sensor level fusion* [20]. Sensor level fusion can be done only if the multiple cues are either instances of the same biometric trait obtained from multiple compatible sensors or multiple instances of the same biometric trait obtained using a single sensor. For example, the face images obtained from several cameras can be combined to form a 3D model of the face. Another example of sensor level fusion is the mosaicking of multiple fingerprint impressions to form a more complete fingerprint image [21, 22]. In sensor level fusion, the multiple cues must be compatible and the correspondences between points in the data must be known in advance. Sensor level fusion may not be possible if the data instances are incompatible (e.g., it may not be possible to integrate face images obtained from cameras with different resolutions).

Feature level fusion refers to combining different feature vectors that are obtained from one of the following sources; multiple sensors for the same biometric trait, multiple instances of the same biometric trait, multiple units of the same biometric trait or multiple biometric traits. 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 of different biometric modalities like face and hand geometry), we

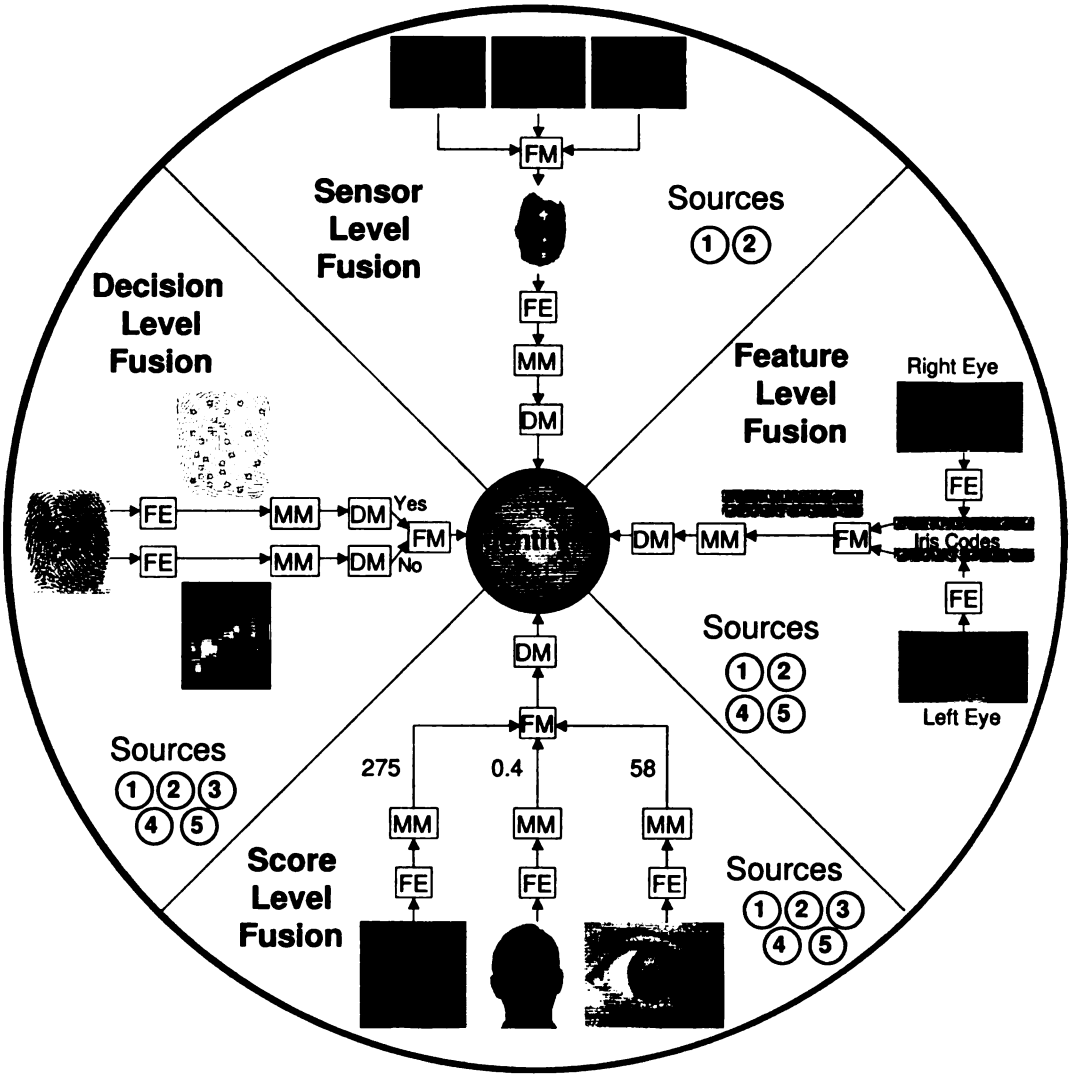


Figure 2.3: Levels of fusion in multibiometric systems.

can concatenate them to form a single feature vector. Concatenation is not possible when the feature sets are incompatible (e.g., fingerprint minutiae and eigen-face coefficients). Attempts by Kumar et al. [23] in combining palmprint and hand-geometry features and by Ross and Govindarajan [24] in combining face and hand-geometry features have met with only limited success.

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 decision of a 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 [25]. 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 fusion schemes after matching.

2.3.2 Fusion After Matching

Schemes for integration of information after the classification/matcher stage can be divided into four categories: dynamic classifier selection, fusion at the decision 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 [26]. This is also known as the winner-take-all approach and the device that performs this selection is known as an associative switch [27].

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 [28], behavior knowledge space [29], weighted voting based on the Dempster-Shafer theory of evidence [30], AND rule and OR rule [31], 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. [32] 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.

2.4 Fusion at the Matching Score Level

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.

2.4.1 Classification Approach to Score Level Fusion

Several classifiers have been used to consolidate the matching scores and arrive at a decision. Wang et al. [33] consider the matching scores resulting from face and iris recognition modules as a two-dimensional feature vector. Fisher's discriminant analysis and a neural network classifier with radial basis function are then used for classification. Verlinde and Chollet [34] 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. [35] 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 (vocal) features. Sanderson et al. [19] 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 [16] use decision tree and linear discriminant classifiers for combining the scores of face, fingerprint, and hand-geometry modalities.

2.4.2 Combination Approach to Score Level Fusion

Kittler et al. [36] 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 must be converted into posteriori probabilities conforming to a genuine user and an impostor. They consider the problem of classifying an input pattern X into one of m possible classes (in a verification system, $m = 2$) based on the evidence provided by R different classifiers or matchers. Let $\vec{x_i}$ be the feature vector (derived from the input pattern X) presented to the i^{th} matcher. Let the outputs of the individual matchers be $P(\omega_j|\vec{x_i})$, i.e., the posterior probability of the of class ω_j given the feature vector $\vec{x_i}$. Let $c \in \{\omega_1, \omega_2, \cdots, \omega_m\}$ be the class to which the input pattern X is finally assigned. The following rules can be used to estimate c:

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

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

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: The sum rule is more effective than the product rule 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

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

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

$$c = argmax_j \max_i \ P(\omega_j|\vec{x}_i).$$

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

$$c = argmax_j \min_i \ P(\omega_j|\vec{x}_i).$$

Prabhakar and Jain [37] 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. [38]. 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 [16,33]. Jain and Ross [39] 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. In [39], user-specific thresholds was also suggested.

2.5 Evaluation of Multimodal Biometric Systems

The performance metrics of a biometric system such as accuracy, throughput, and scalability can be estimated with a high degree of confidence only when the system is tested on

a large representative database. For example, face [13] and fingerprint [12] recognition systems have been evaluated on large databases (containing samples from more than 25,000 individuals) obtained from a diverse population under a variety of environmental conditions. In contrast, current multimodal systems have been tested only on small databases containing fewer than $1,000$ individuals. Further, multimodal biometric databases can be either true or virtual. In a true multimodal database (e.g., XM2VTS database [40]), different biometric cues are collected from the same individual. Virtual multimodal databases contain records which are created by consistently pairing a user from one unimodal database with a user from another database. The creation of virtual users is based on the assumption that different biometric traits of the same person are independent. This assumption of independence of the various modalities has not been explicitly investigated till date. However, Indovina et al. [41] attempted to validate the use of virtual subjects. They randomly created $1,000$ sets of virtual users and showed that the variation in performance among these sets was not statistically significant. Recently, NIST has released a true multimodal database [42] containing the face and fingerprint matching scores of 517 individuals.

2.6 Summary

We have presented a detailed discussion on the various approaches that have been proposed for integrating evidence obtained from multiple cues in a biometric system. Figure 2.4 presents a high-level summary of these information fusion techniques. Most of research work on fusion in multimodal biometric systems has focused on fusion at the matching score level. In particular, the combination approach to score level fusion has received considerable attention. However, there are still many open questions that have been left unanswered. There is no standard technique either for converting the scores into probabilities or for normalizing the scores obtained from multiple matching algorithms. A systematic

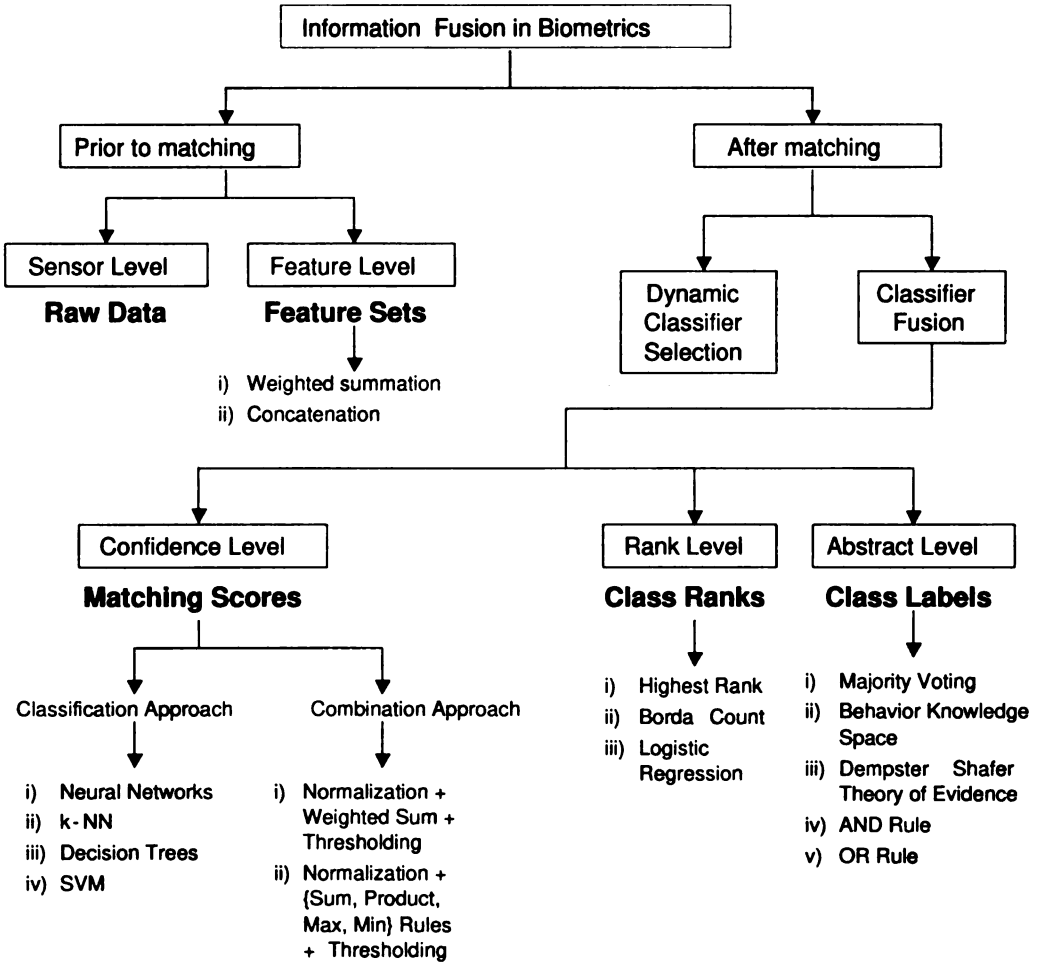


Figure 2.4: Summary of approaches to information fusion in biometric systems.

evaluation of the different normalization techniques is not available. Further, most of the score level fusion techniques can be applied only when the individual modalities can provide a reasonably good recognition performance. They cannot handle less reliable (soft) biometric identifiers that can provide some amount of discriminatory information, but are not sufficient for recognition of individuals. For example, the height of a user gives some indication on who the user could be. However, it is impossible to identify a user just based on his height. Currently, there is no mechanism to deal with such soft information.

CHAPTER 3

Score Normalization in Multimodal Biometric Systems

Consider a multimodal biometric verification system that follows the combination approach to fusion at the matching score level. The theoretical framework developed by Kittler et al. in [36] can be applied to this system only if the output of each modality is of the form $P(genuine|X)$ i.e., the posteriori probability of user being "genuine" given the input biometric sample X. In practice, most biometric systems output a matching score s. Verlinde et al. [43] have proposed that the matching score s is related to $P(genuine|X)$ as follows:

$$s = f\left(P(genuine|X)\right) + \eta(X), \tag{3.1}$$

where f is a monotonic function and η is the error made by the biometric system that depends on the input biometric sample X. 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 $P(genuine|X)$ 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. [44] assumed a normal distribution for the conditional densities of the matching scores ($p(s|genuine) \sim N(\mu_g, \sigma_g)$ and $p(s|impostor) \sim N(\mu_i, \sigma_i)$), 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 [25], 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,

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

where $p(s) = (p(s|genuine) * P(g) + p(s|impostor) * P(i))$ and $P(g)$ and $P(i)$ 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 non-Gaussian), the density estimation still has inaccuracies 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 equation (3.1) is zero is not valid in most practical biometric systems. Since η depends on the input biometric sample X, it is possible to estimate η only if the biometric system outputs a confidence measure (that takes into account the nature of the input X) 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 not a good estimate of $P(genuine|X)$ and this can lead to poor recognition performance

of the multimodal system. Hence, when the outputs of individual modalities are matching scores without any measures quantifying the confidence on those scores, it would be better to combine the matching scores directly without converting them into probabilities.

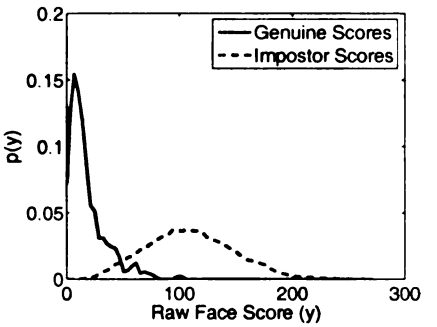
3.1 Need for Score Normalization

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. Further, 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.

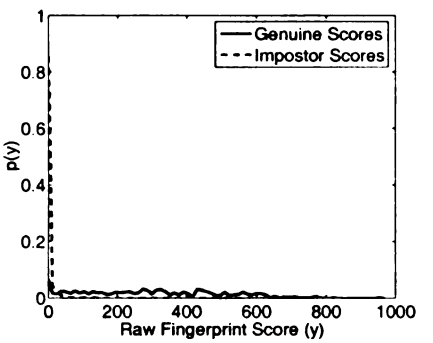
Figure 3.1 shows the conditional distributions of the face, fingerprint and hand-geometry matching scores used in our experiments. The scores obtained from the face and hand-geometry matchers are distance scores and those obtained from the fingerprint matcher are similarity scores. One can easily observe the non-homogeneity in these scores and the need for normalization prior to any meaningful combination.

3.2 Challenges in Score Normalization

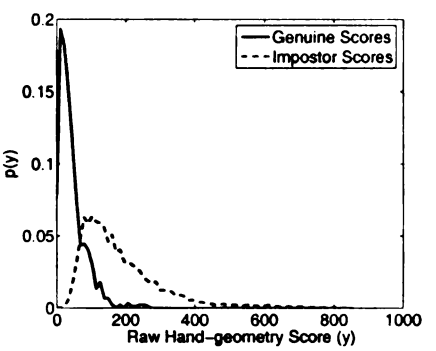
Score normalization refers to changing the location and scale parameters of the matching score distributions at the outputs of the individual matchers, so that the matching scores of different matchers are transformed into a common domain. When the parameters used for normalization are determined using a fixed training set, it is referred to as *fixed score*



(a)



(b)



(c)

Figure 3.1: Conditional distributions of genuine and impostor scores used in our experiments; (a) Face (distance score), (b) Fingerprint (similarity score) and (c) Hand-geometry (distance score).

normalization [45]. 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 [46]. Min-max normalization and z-score normalization are some of the popular techniques used for relevance score normalization in metasearch. In metasearch literature [47], 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 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 [48] 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.

3.3 Normalization Techniques

The simplest normalization technique is the *Min-max normalization*. 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. Let s_{ij} denote the j^{th} matching score output by the i^{th} modality, where $i = 1, 2, \cdots, R$ and $j = 1, 2, \cdots, M$ (R is the number of modalities and M is the number of matching scores available in the training set). The min-max normalized score for the test score s_{ik} is given by

$$s'_{ik} = \frac{s_{ik} - min\left(\{s_{i.}\}\right)}{max\left(\{s_{i.}\}\right) - min\left(\{s_{i.}\}\right)},$$

where $\{s_{i.}\} = \{s_{i1}, s_{i2}, \cdots, s_{iM}\}$. 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 normalized score from 1. Figure 3.2 shows the distributions 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 $[0, 100]$, the following normalization could be applied.

$$s'_{ik} = \frac{s_{ik}}{10^n},$$

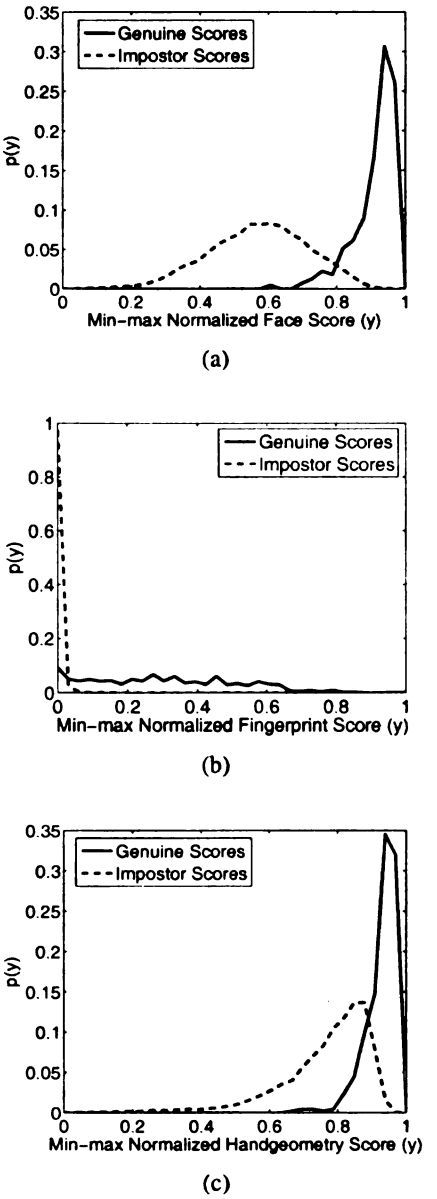


Figure 3.2: Distributions of genuine and impostor scores after min-max normalization; (a) Face, (b) Fingerprint and (c) Hand-geometry.

where n $= \log_{10} \max(\{s_{i.}\})$. The problems with this approach are the 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 uses 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 a given set of matching scores. The normalized scores are given by

$$s'_{ik} = \frac{s_{ik} - \mu}{\sigma},$$

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 distribution of the scores is not Gaussian, 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 Figure 3.3. The face and hand-geometry scores are converted into similarity scores by subtracting the scores from a large number (300 for face and 1000 for

hand-geometry in our experiments) before applying the z-score transformation. Figure 3.3 shows 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

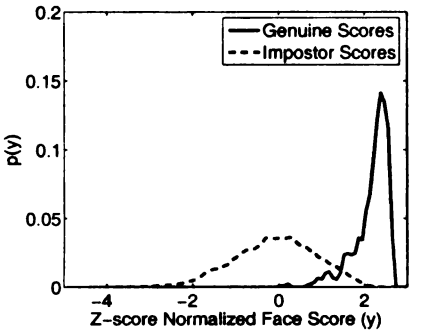
$$s'_{ik} = \frac{s_{ik} - median}{MAD},$$

where $MAD = median\left(\{|s_{i\cdot} - median\left(\{s_{i\cdot}\}\right)|\}\right)$. 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 Figure 3.4.

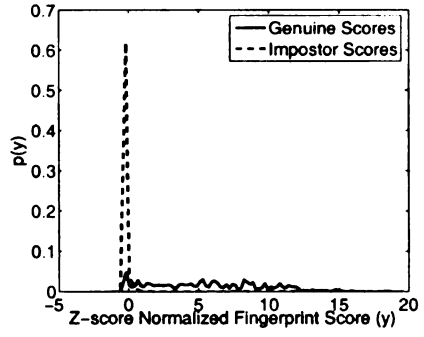
Cappelli et al. [49] have used a *double sigmoid function* for score normalization in a multimodal biometric system that combines different fingerprint matchers. The normalized score is given by

$$s'_{ik} = \begin{cases} \frac{1}{1+\exp\left(-2\left(\frac{s_{ik}-t}{r_1}\right)\right)} & \text{if } s_k < t, \\ \frac{1}{1+\exp\left(-2\left(\frac{s_{ik}-t}{r_2}\right)\right)} & \text{otherwise,} \end{cases}$$

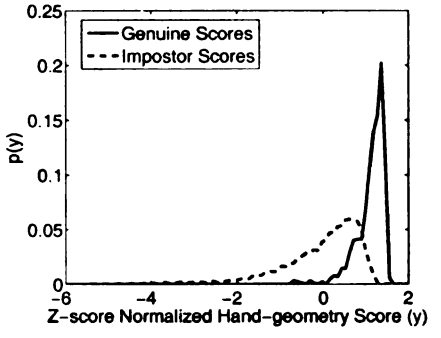
where t is the reference operating point and r_1 and r_2 denote the left and right edges of the region in which the function is linear, i.e., the double sigmoid function exhibits linear



(a)



(b)



(c)

Figure 3.3: Distributions of genuine and impostor scores after z-score normalization; (a) Face, (b) Fingerprint, and (c) Hand-geometry.

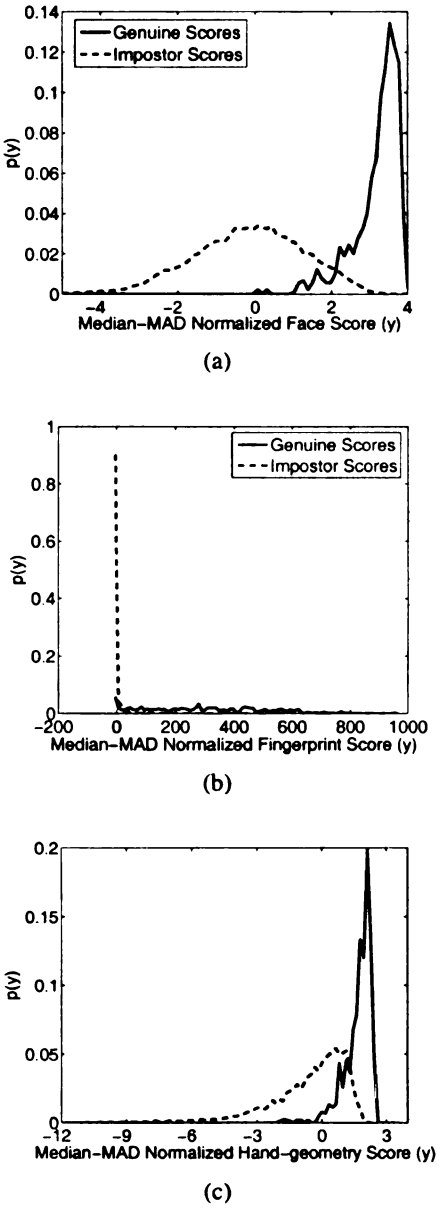


Figure 3.4: Distributions of genuine and impostor scores after median-MAD normalization; (a) Face, (b) Fingerprint and (c) Hand-geometry.

characteristics in the interval $(t - r_1,\ t - r_2)$. Figure 3.5 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 $r_2 = 30$.

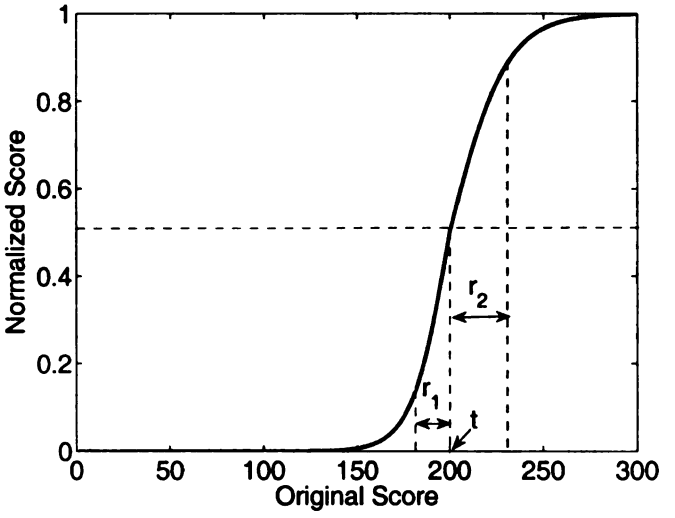


Figure 3.5: Double sigmoid normalization ($t = 200$, $r_1 = 20$, and $r_2 = 30$).

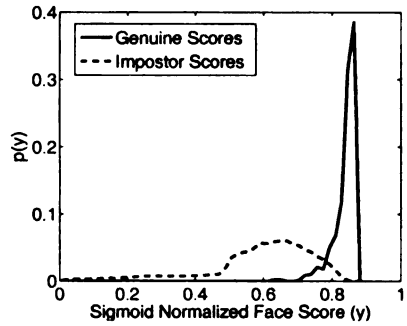
This scheme transforms the scores into the $[0, 1]$ interval. But, it requires careful tuning of the parameters t, r_1 and r_2 to obtain good efficiency. Generally, t is chosen to be some value falling in the region of overlap between the genuine and impostor score distribution, and r_1 and r_2 are set so that they correspond 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 a two-quadrics (QQ) or a logistic

(LG) function as suggested by Snelick et al. [18]. 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.

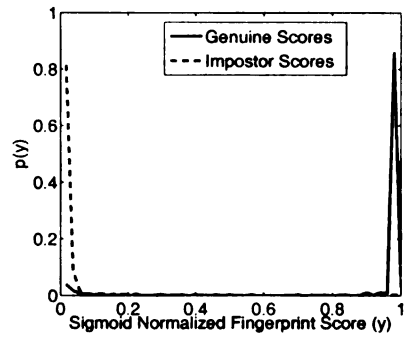
Figure 3.6 shows the face, fingerprint and hand-geometry score distributions after double sigmoid normalization. The face and hand-geometry scores are converted into similarity scores by subtracting the normalized scores from 1. The parameters of the double sigmoid normalization were chosen as follows: t is chosen to be the center of the overlapping regions between the genuine and impostor score distributions, and r_1 and r_2 are set so that they correspond to the minimum genuine similarity score and maximum impostor similarity score, 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 the genuine scores, while r_2 is the difference between the maximum of the 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. [50] are robust and highly efficient. The normalization is given by

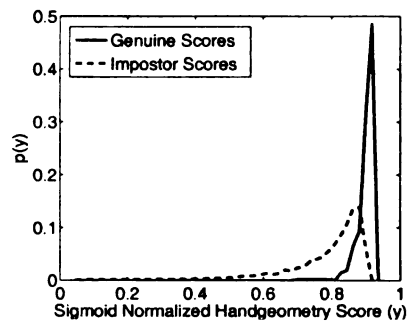
$$s'_{ik} = \frac{1}{2}\left\{\tanh\left(0.01\left(\frac{s_{ik} - \mu_{GH}}{\sigma_{GH}}\right)\right) + 1\right\},$$



(a)



(b)



(c)

Figure 3.6: Distributions of genuine and impostor scores after double sigmoid normalization; (a) Face, (b) Fingerprint and (c) Hand-geometry.

where μ_{GH} and σ_{GH} are the mean and standard deviation estimates, respectively, of the genuine score distribution as given by Hampel estimators [1]. Hampel estimators are based on the following influence (ψ)-function:

$$\psi(u) = \begin{cases} u & 0 \leq |u| < a, \\ a * sign(u) & a \leq |u| < b, \\ a * sign(u) * \left(\frac{c-|u|}{c-b}\right) & b \leq |u| < c, \\ 0 & |u| \geq c. \end{cases}$$

A plot of the Hampel influence function is shown in Figure 3.7. 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 many of the points that constitute the tail of the distributions are discarded, the estimate is robust but not efficient (optimal). On the other hand, if all the points that constitute the tail of the distributions are considered, 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 Figure 3.8. 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

[1] In [44,45], 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.

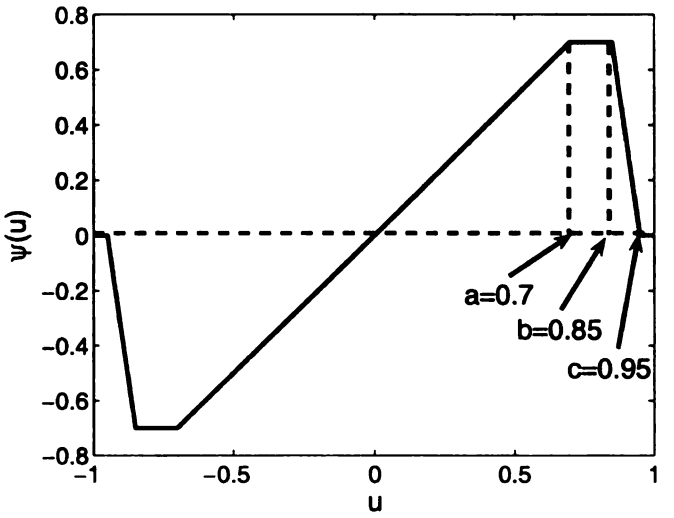
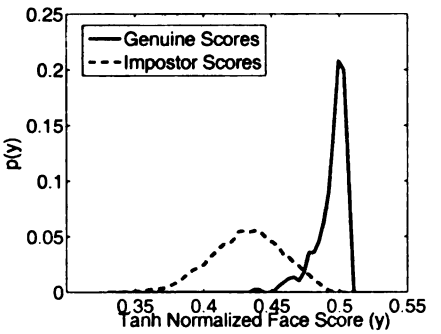


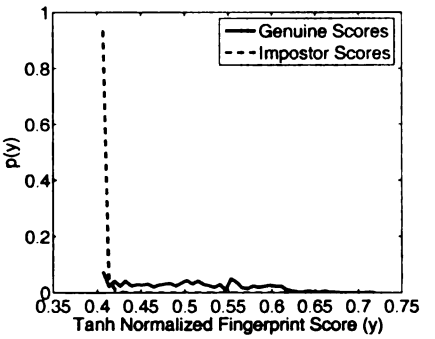
Figure 3.7: Hampel influence function ($a = 0.7$, $b = 0.85$, and $c = 0.95$).

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 not 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.

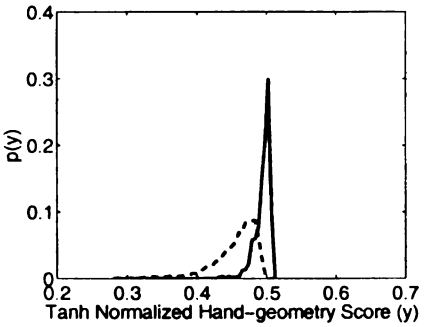
Mosteller and Tukey [51] introduced the biweight location and scale estimators that are robust and efficient. But, the *biweight estimators* are iterative in nature (an initial estimate of the biweight location and scale parameters is chosen, and this estimate is updated based



(a)



(b)



(c)

Figure 3.8: Distributions of genuine and impostor scores after tanh normalization; (a) Face, (b) Fingerprint and (c) Hand-geometry.

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 considered biweight normalization in our experiments. The characteristics of the different normalization techniques have been tabulated in Table 3.1.

Table 3.1: 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

3.4 Experimental Results

Snelick et al. [18] have developed a general testing framework 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 1,000 users showed that the min-max normalization followed by the sum of scores fusion method generally provided better recognition performance than other schemes. However, the reasons for such a behavior have not been presented by these authors. In this work, 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 [18], we have also analyzed the double sigmoid method of normalization.

3.4.1 Generation of the Multimodal Database

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 [16]) 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 $\times$ 480) and fingerprint impressions were obtained using a Digital Biometrics sensor (500 dpi, 640 $\times$ 480). Five hand-geometry images were obtained from a 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 multimodal database consisting of 50 virtual 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 $\times$ 10) genuine

score vectors and $24,500$ ($50 \times 10 \times 49$) 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 $1,000$ genuine score vectors and $49,000$ impostor score vectors. A score vector $\mathbf{s}_k$ is a 3-tuple $\{s_{1k}, s_{2k}, s_{3k}\}$, where s_{1k}, s_{2k}, and s_{3k} correspond to the k^{th} 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 [52] and the output of the fingerprint matcher was a similarity score. Eigenface coefficients were used to represent features of the face image [53]. The Euclidean distance between the eigenface coefficients of the face 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 [54] and the matching score was computed as the Euclidean distance between the input feature vector and the template feature vector.

The recognition performance of the face, fingerprint, and hand-geometry systems when operated as unimodal systems is shown in Figure 3.9. From Figure 3.1(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. Figure 3.1(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 Figure 3.1(c). Hence, the hand geometry based recognition performance is low compared to the fingerprint and face matchers.

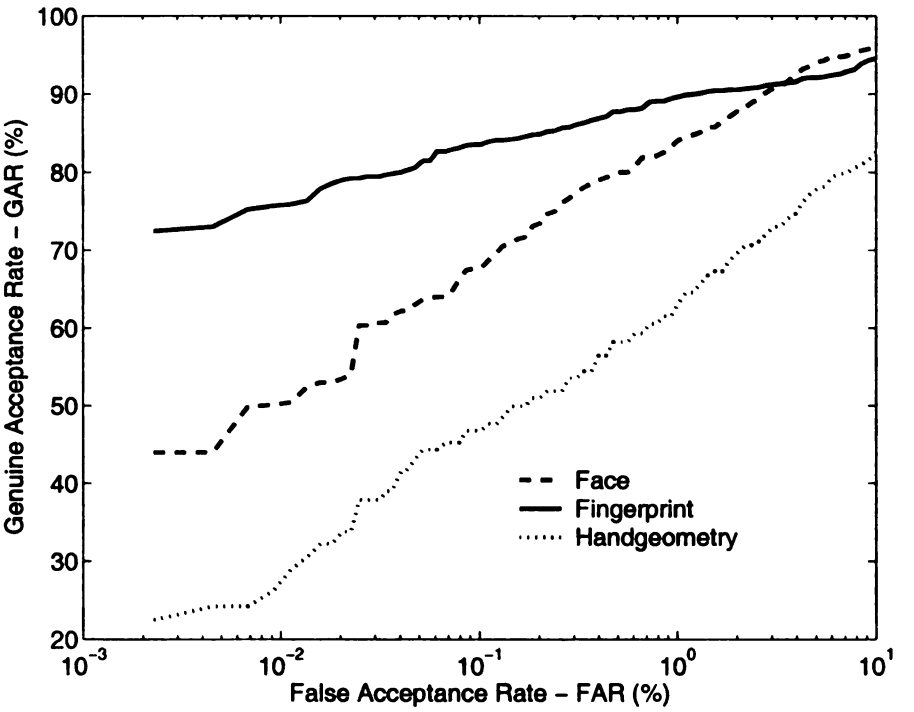


Figure 3.9: ROC curves for individual modalities.

3.4.2 Impact of Normalization on Fusion Performance

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 described in [18] 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 3.2 summarizes the average (over 40 trials) 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%.

Table 3.2: Genuine Acceptance Rate (GAR) (%) of different normalization and fusion techniques at the 0.1% False Acceptance Rate (FAR). Note that the values in the table represent average GAR, and the values indicated in parentheses correspond to the standard deviation of GAR.

Normalization Techniques		**Fusion Techniques**	
	Sum of scores	**Max-score**	**Min-score**
STrans	98.3 (0.4)	46.7 (2.3)	83.9 (1.6)
Minmax	97.8 (0.6)	67.0 (2.5)	83.9 (1.6)
Zscore	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)

[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 the case of this method, the simple sum of scores, max score, and min score fusion schemes, reduce to the sum rule, max rule, and min rule described in [36], respectively.

Figure 3.10 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.

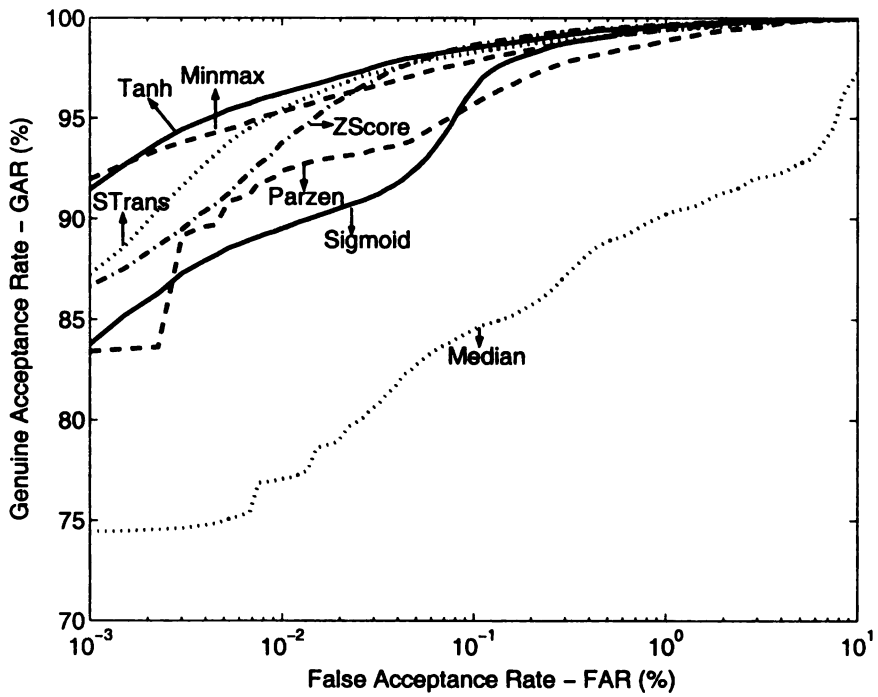


Figure 3.10: ROC curves for sum of scores fusion method under different normalization schemes.

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_k) is just a linear transformation of the score vector $\{s_{1k}, s_{2k}, s_{3k}\}$, i.e., $s_k = (a_1 s_{1k} - b_1) + (a_2 s_{2k} - b_2) + (a_3 s_{3k} - b_3)$, where s_{1k}, s_{2k}, and s_{3k} correspond to the k^{th} 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 ($a_2 >> a_1, a_3$). 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 Figure 3.11. 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 transformation, 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 the user only if all the three modalities make errors in the probability estimation process. Hence, this method has a high genuine acceptance rate.

Figure 3.12 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.

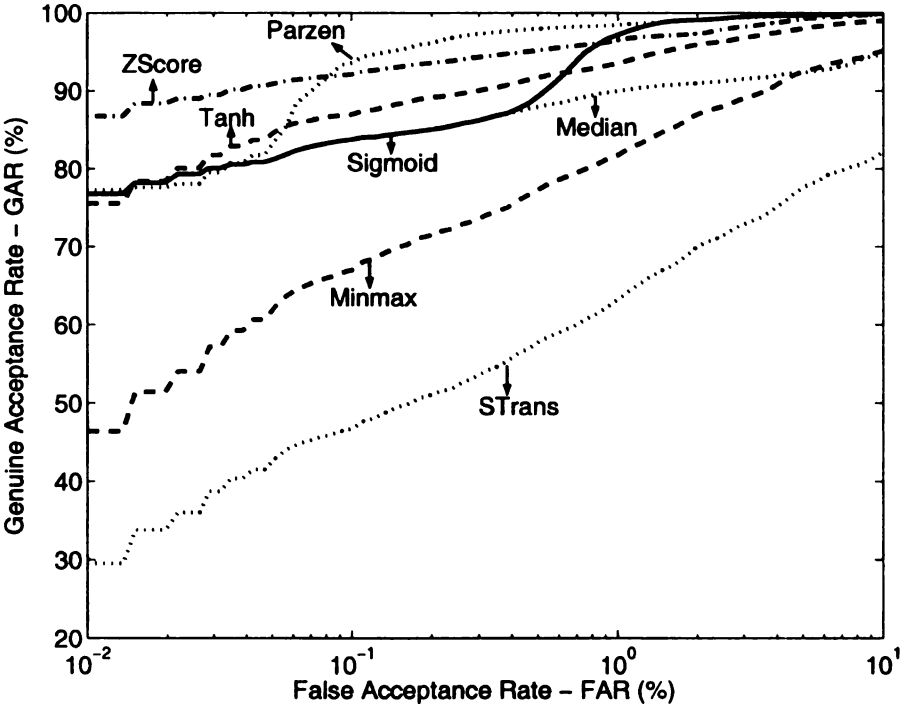


Figure 3.11: ROC curves for max-score fusion method under different normalization schemes.

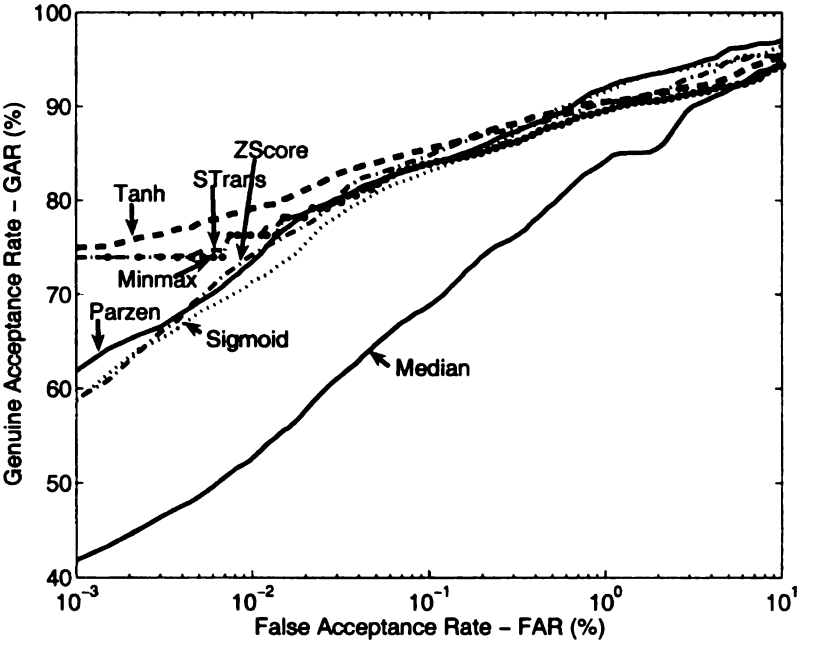


Figure 3.12: ROC curves for min-score fusion method under different normalization schemes.

3.4.3 Robustness Analysis of Normalization Schemes

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 in the presence of outliers, we artificially introduced outliers in the fingerprint scores.

For min-max normalization, a single large score whose value is 125%, 150%, 175% or 200% of the original maximum score is introduced into the fingerprint data. Figure 3.13 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%, 150%, 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

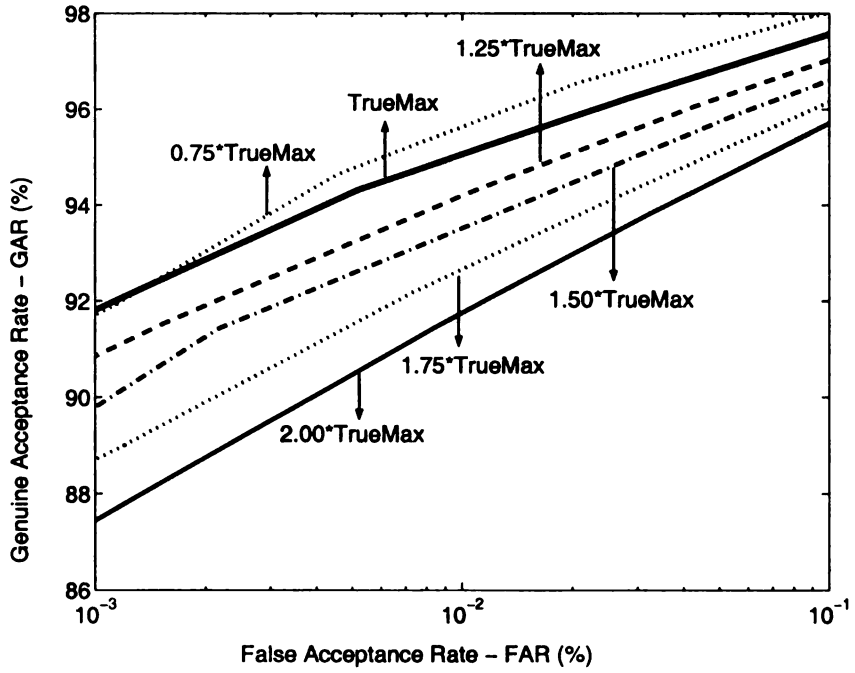


Figure 3.13: Robustness analysis of min-max normalization.

of an increase in standard deviation, the performance improves after the introduction of outliers as indicated in Figure 3.14. 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 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 Figure 3.15, there is no significant variation in the performance after the introduction of outliers. This result highlights the robustness of the tanh normalization method.

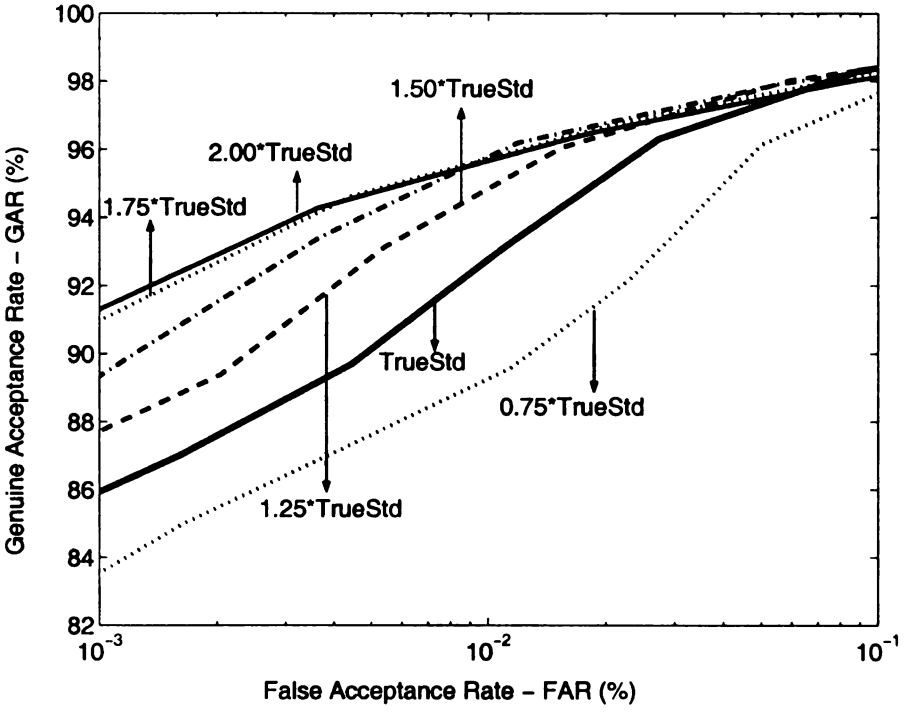


Figure 3.14: Robustness analysis of z-score normalization.

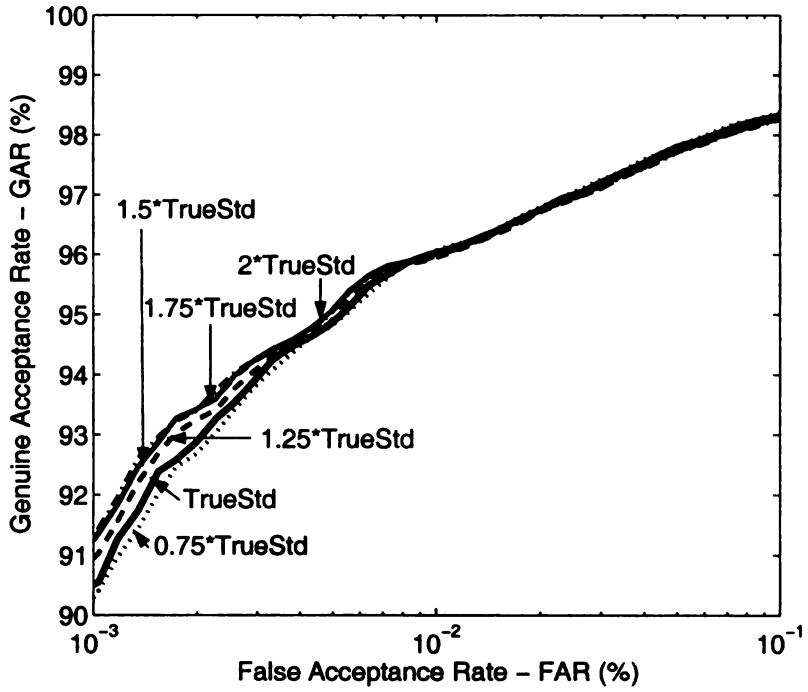


Figure 3.15: Robustness analysis of tanh normalization.

3.5 Summary

Score normalization is an important issue to be dealt with in score level fusion and it has significant impact on the performance of a multibiometric system. We have examined 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.

CHAPTER 4

Soft Biometrics

Current biometric systems are not perfect (do not have zero error rates) and problems like noise in the sensed biometric data, non-universality and lack of distinctiveness of the chosen biometric trait lead to unacceptable error rates in recognizing a person. Using a combination of biometric identifiers like face, fingerprint, hand-geometry and iris makes the resulting multibiometric system more robust to noise and can alleviate problems such as non-universality and lack of distinctiveness, thereby reducing the error rates significantly. However, using multiple traits will increase the enrollment and verification times, cause more inconvenience to the users and increase the overall cost of the system. Therefore, we propose another solution to reduce the error rates of the biometric system without causing any additional inconvenience to the user. Our solution is based on incorporating soft identifiers of human identity like gender, ethnicity, height, eye color, etc. into a (primary) biometric identification system. Figure 4.1 shows a sample application (ATM kiosk) where both primary (fingerprint) and soft (gender, ethnicity, height estimated using a camera) biometric information are utilized to verify the account holder's identity.

4.1 Motivation and Challenges

The usefulness of soft biometric traits in improving the performance of the primary biometric system can be illustrated by the following example. Consider three users A (1.8m tall, male), B (1.7m tall, female), and C (1.6m tall, male) enrolled in a fingerprint system that works in the identification mode. When user A presents his fingerprint sample X to the system, it is compared to the templates of all the three users stored in the database

Figure 4.1: An ATM kiosk equipped with a fingerprint (primary biometric) sensor and a camera to obtain soft attributes (gender, ethnicity and height).

and the posteriori matching probabilities of all the three users given the template X is calculated. Let us assume that the output of the fingerprint matcher is $P(A|X) = 0.42$, $P(B|X) = 0.43$, and $P(C|X) = 0.15$. In this case, the test user will either be rejected due to the proximity of the posteriori matching probabilities for users A and B, or be falsely identified as user B. On the other hand, let us assume that there exists a secondary system that automatically identifies the gender of the user as male and measures the user's height as 1.78m. If we have this information in addition to the posteriori matching probabilities given by the fingerprint matcher, then a proper combination of these sources of information will lead to a correct identification of the test user as user A.

The first biometric system developed by Alphonse M. Bertillon in 1883 used anthropometric features such as the length and breadth of the head and the ear, length of the middle finger and foot, height, etc. along with attributes like eye color, scars, and tatoo marks for ascertaining a person's identity [55]. Although each individual measurement in the Bertillon system may exhibit some variability, a combination of all these measurements

was sufficient to manually identify a person with reasonable accuracy. The Bertillon system was dropped in favor of the Henry's system of fingerprint identification due to three main reasons: (i) lack of persistence - the anthropometric features can vary significantly over a period of time; (ii) lack of distinctiveness - features such as skin color or eye color cannot be used for distinguishing between individuals coming from a similar background; and (iii) the time, effort, training, and cooperation required to get reliable measurements. Similar to the Bertillon system, Heckathorn et al. [56] used attributes like gender, race, eye color, height, and other visible marks like scars and tattoos to recognize individuals for the purpose of welfare distribution. More recently, Ailisto et al. [57] showed that unobtrusive user identification can be performed in non-security applications such as health clubs using a combination of "light" biometric identifiers like height, weight, and body fat percentage. However, it is obvious that the features used in the above mentioned systems provide some identity information, but are not sufficient for accurate person recognition. Hence, these attributes can be referred to as "soft biometric traits". The soft biometric information complements the identity information provided by traditional (primary) biometric identifiers such as fingerprint, iris, and voice. Therefore, utilizing soft biometric traits can improve the recognition accuracy of primary biometric systems.

Wayman [58] proposed the use of soft biometric traits like gender and age, for filtering a large biometric database. Filtering refers to limiting the number of entries in a database to be searched, based on characteristics of the interacting user. For example, if the user can somehow be identified as a middle-aged male, the search can be restricted only to the subjects with this profile enrolled in the database. This greatly improves the speed or the search efficiency of the biometric system. In addition to filtering, the soft biometric traits can also be used for tuning the parameters of the biometric system. Studies [59, 60] have shown that factors such as age, gender, race, and occupation can affect the performance of

a biometric system. For example, a young female Asian mine-worker is considered as one of the most difficult subjects for a fingerprint system [60]. This provides the motivation for tuning the system parameters like threshold on the matching score in a unimodal biometric system, and thresholds and weights of the different modalities in a multimodal biometric system to obtain the optimum performance for a particular user or a class of users. However, filtering and system parameter tuning require soft biometric feature extractors that are highly accurate.

Based on the above observations, the first challenge in utilizing soft biometrics is the automatic and reliable extraction of soft biometric information in a non-intrusive manner without causing any inconvenience to the users. It must be noted that the unreliability and inconvenience caused by the manual extraction of these features is one of the primary reasons for the failure of Bertillon-like systems. In this thesis, we have analyzed the various techniques that have been proposed for automatic extraction of characteristics like gender, ethnicity, eye color, age, and height. We also briefly describe the vision system that was implemented to accomplish the task of identifying the gender, ethnicity, height, and eye color of a person from real-time video sequences. Once the soft biometric information is extracted, the challenge is to optimally combine this information with the primary biometric identifier so that the overall recognition accuracy is enhanced. We have developed a Bayesian framework for integrating the primary and soft biometric features.

4.2 Soft Biometric Feature Extraction

Any trait that provides some information about the identity of a person, but does not provide sufficient evidence to exactly determine the identity can be referred to as soft biometric trait. Figure 4.2 shows some examples of soft biometric traits. Soft biometric traits

are available and can be extracted in a number of practical biometric applications. For example, demographic attributes like gender, ethnicity, age, eye color, skin color, and other distinguishing physical marks such as scars can be extracted from the face images used in a face recognition system. The pattern class of fingerprint images (right loop, left loop, whorl, arch, etc.) is another example of soft trait. Gender, accent, and perceptual age of the speaker can be inferred in a voice recognition system. Eye color can be easily found from iris images. However, automatic and reliable extraction of soft biometric traits is a difficult task. In this section, we present a survey of the techniques that have been proposed in the literature for extracting soft biometric information and briefly describe our system for determining height, gender, ethnicity, and eye color.

Several researchers have attempted to derive gender, ethnicity, and pose information about the users from their face images. Gutta et al. [61] proposed a mixture of experts consisting of ensembles of radial basis functions for the classification of gender, ethnic origin, and pose of human faces. Their gender classifier (male vs female) had an accuracy of 96%, while their ethnicity classifier (Caucasian, South Asian, East Asian, and African) had an accuracy of 92%. These results were reported on good quality face images from the FERET database that had very little expression or pose changes. Based on the same database, Moghaddam and Yang [62] showed that the error rate for gender classification can be reduced to 3.4% by using an appearance-based gender classifier that uses non-linear support vector machines. Shakhnarovich et al. [63] developed a demographic classification scheme that extracts faces from unconstrained video sequences and classifies them based on gender and ethnicity. The learning and feature selection modules used a variant of the AdaBoost algorithm. Even under unconstrained environments, they showed that a classification accuracy of more than 75% can be achieved for both gender and ethnicity (Asian vs non-Asian) classification. For this data, the SVM classifier of Moghaddam and Yang

Gender, Skin Color, Hair color
http://anthro.palomar.edu/adapt/adapt_4.htm
© Corel Corporation, Ottawa, Canada

Height
http://www.altonweb.com/history/wadlow/p2.html
© Alton Museum of History and Art

Weight
http://www.laurel-and-hardy.com/goodies/home6.html © CCA

Eye color
http://ology.amnh.org/genetics/longdefinition/index3.html
© American Museum of Natural History, 2001

Figure 4.2: Examples of soft biometric traits.

also had a similar performance and there was also a noticeable bias towards males in the gender classification (females had an error rate of 28%). Balci and Atalay [64] reported a classification accuracy of more than 86% for a gender classifier that uses PCA for feature extraction and a multilayer perceptron for classification. Jain and Lu [65] proposed a Linear Discriminant Analysis (LDA) based scheme to address the problem of ethnicity identification from facial images. The users were identified as either Asian or non-Asian by applying multiscale analysis to the input facial images. An ensemble framework based on the product rule was used for integrating the LDA analysis at different scales. This scheme had an accuracy of 96.3% on a database of 263 users (with approximately equal number of males and females). Hayashi et al. [66] suggested the use of features like wrinkle texture and color for estimating the age and gender of a person from the face image. However, they did not report the accuracy of their technique.

Automatic age determination is a more difficult problem than gender and ethnicity classification. Buchanan et al. [67] have studied the differences in the chemical composition

of fingerprints that could be used to distinguish children from adults. Kwon and Lobo [68] presented an algorithm for age classification from facial images based on cranio-facial changes in feature-position ratios and skin wrinkle analysis. They attempted to classify users as "babies", "young adults", or "senior adults". However, they did not provide any classification accuracy. More recently, Lanitis et al. [69] performed a quantitative evaluation of the performance of three classifiers developed for the task of automatic age estimation from face images. These classifiers used eigenfaces obtained using Principal Component Analysis (PCA) as the input features. Quadratic models, shortest distance classifier, neural network classifiers, and hierarchical age estimators were used for estimating the age. The best hierarchical age estimation algorithm had an average absolute error of 3.82 years which was comparable to the error made by humans (3.64 years) in performing the same task. Minematsu et al. [70] showed that the perceptual age of a speaker can be automatically estimated from voice samples. All these results indicate that the automatic age estimation is possible (though the current technology is not very reliable).

The weight of a user can be measured by asking him to stand on a weight sensor while providing the primary biometric. The height of a person can be estimated from a sequence of real-time images. For example, Su-Kim et al. [71] used geometric features like vanishing points and vanishing lines to compute the height of an object. With the rapid growth of technology, especially in the field of computer vision, we believe that the techniques for soft biometric feature extraction would become more reliable and commonplace in the near future.

4.2.1 A Vision System for Soft Biometric Feature Extraction

We have implemented a real-time vision system for automatic extraction of gender, ethnicity, height, and eye color. The system is designed to extract the soft biometric attributes

as the person approaches the primary biometric system to present his primary biometric identifier (face and fingerprint in our case). The soft biometric system is equipped with two Sony EVI-D30 color pan/tilt/zoom cameras. Camera I monitors the scene for any human presence based on the motion segmentation image. Once camera I detects an approaching person, it measures the height of the person and then guides camera II to focus on the person's face. More details about this system can be found in [72].

4.3 Fusion of Soft and Primary Biometric Information

4.3.1 Identification Mode

For a biometric system operating in the identification mode, the framework for integrating primary and soft biometric information is shown in Figure 4.3. The primary biometric system is based on m ($m \geq 1$) traditional biometric identifiers like fingerprint, face, iris and hand-geometry. The soft biometric system is based on n ($n \geq 1$) soft attributes like age, gender, ethnicity, eye color and height. Let $\omega_1, \omega_2, \cdots, \omega_R$ represent the R users enrolled in the database. Let $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_m]$ be the collection of primary biometric feature vectors. For example, if the primary biometric system is a multimodal system with face and fingerprint modalities ($m = 2$), then $\mathbf{x}_1$ represents the face feature vector and $\mathbf{x}_2$ represents the fingerprint feature vector. Let $p(\mathbf{x}_j|\omega_i)$ be the likelihood of observing the primary biometric feature vector $\mathbf{x}_j$ given the user is ω_i. If the output of each individual modality in the primary biometric system is a set of matching scores ($\mathbf{s} = [s_1, s_2, \cdots, s_m]$), one can approximate $p(\mathbf{x}_j|\omega_i)$ by $p(s_j|\omega_i)$, provided the genuine matching score distribution of each modality is known.

Let $\mathbf{y} = [y_1, y_2, \cdots, y_n]$ be the soft biometric feature vector, where, for example, y_1 could be gender, y_2 could be eye color, etc. We require an estimate of the posteriori

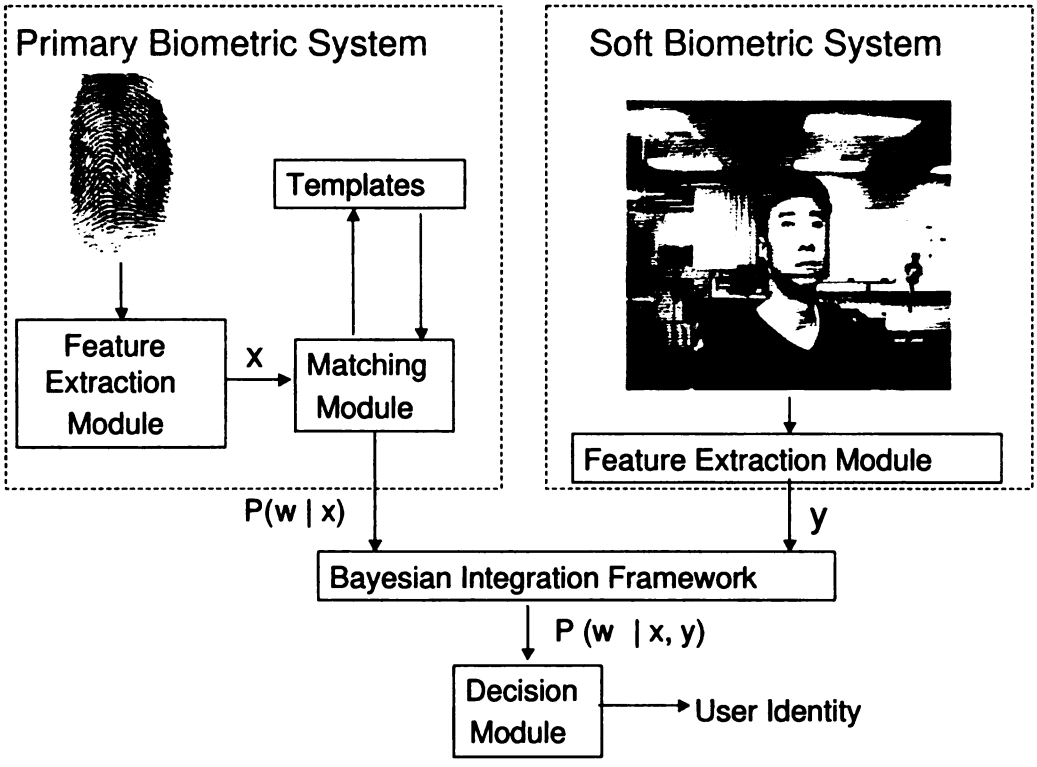


Figure 4.3: Framework for fusion of primary and soft biometric information. Here **x** is the fingerprint feature vector and **y** is the soft biometric feature vector.

probability of user ω_i given both $\mathbf{x}$ and $\mathbf{y}$. This posteriori probability can be calculated by applying the Bayes rule as follows:

$$P(\omega_i|\mathbf{x},\mathbf{y}) = \frac{p(\mathbf{x},\mathbf{y}|\omega_i)P(\omega_i)}{p(\mathbf{x},\mathbf{y})}. \tag{4.1}$$

If all the users are equally likely to access the system, then $P(\omega_i) = \frac{1}{R}$, $\forall\ i$. Further, if we assume that all the primary biometric feature vectors $(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_m)$ and all the soft biometric variables $(y_1,\ y_2,\ \cdots,\ y_n)$ are independent of each other given the user's identity ω_i, the discriminant function $g_i(\mathbf{x},\ \mathbf{y})$ for user ω_i can be written as,

$$g_i(\mathbf{x},\ \mathbf{y}) = \sum_{j=1}^{m} \log p(\mathbf{x}_j|\omega_i) + \sum_{k=1}^{n} \log p(y_k|\omega_i). \tag{4.2}$$

4.3.2 Verification Mode

A biometric system operating in the verification mode classifies each authentication attempt as either a "genuine claim" or an "impostor attempt". In the case of verification, the Bayes decision rule can be expressed as

$$\frac{P(genuine|\mathbf{x},\mathbf{y})}{P(impostor|\mathbf{x},\mathbf{y})} = \frac{p(\mathbf{x},\mathbf{y}|genuine)P(genuine)}{p(\mathbf{x},\mathbf{y}|impostor)P(impostor)} \geq \tau, \tag{4.3}$$

where τ is the threshold parameter. Increasing τ reduces the false acceptance rate and simultaneously increases the false reject rate and vice versa. If the prior probabilities of the genuine and impostor classes are equal and if we assume that all the primary biometric feature vectors and all the soft biometric attributes are independent of each other given the class, the discriminant function can be written as,

$$g(\mathbf{x},\ \mathbf{y}) = \sum_{j=1}^{m} \log\left(\frac{p(\mathbf{x}_j|genuine)}{p(\mathbf{x}_j|impostor)}\right) + \sum_{k=1}^{n} \log\left(\frac{p(y_k|genuine)}{p(y_k|impostor)}\right). \tag{4.4}$$

If the output of each individual modality in the primary biometric system is a set of matching scores ($\mathbf{s} = [s_1, s_2, \cdots, s_m]$), one can approximate $p(\mathbf{x}_j|genuine)$ and $p(\mathbf{x}_j|impostor)$ by $p(s_j|genuine)$ and $p(s_j|impostor)$, respectively, provided the genuine and impostor matching score distributions of each modality are known.

4.3.3 Computation of Soft Biometric Likelihoods

A simple method for computing the soft biometric likelihoods $p(y_k|\omega_i),\ k = 1, 2, \cdots, n$ is to estimate them based on the accuracy of the soft biometric feature extractors on a training database. For example, if the accuracy of the gender classifier on a training database is α, then

1. P(observed gender is male | true gender of the user is male) = α,
2. P(observed gender is female | true gender of the user is female) = α,
3. P(observed gender is male | true gender of the user is female) = $1 - \alpha$,
4. P(observed gender is female | true gender of the user is male) = $1 - \alpha$.

Similarly, if the average error made by the system in measuring the height of a person is μ_e and the standard deviation of the error is σ_e, then it is reasonable to assume that $p(\text{observed height}|\omega_i)$ follows a Gaussian distribution with mean $(h(\omega_i) + \mu_e)$ and standard deviation σ_e, where $h(\omega_i)$ is the true height of user ω_i. However, there is a potential problem when the likelihoods are estimated only based on the accuracy of the soft

biometric feature extractors. The discriminant function in equation (4.2) is dominated by the soft biometric terms due to the large dynamic range of the soft biometric log-likelihood values. For example, if the gender classifier is 98% accurate ($\alpha = 0.98$), the log-likelihood for the gender term in equation (3) is -0.02 if the classification is correct and -3.91 in the case of a misclassification. This large difference in the log-likelihood values is due to the large variance of the soft biometric features. To offset this phenomenon, we introduce a scaling factor β, $0 \leq \beta \leq 1$, to flatten the likelihood distribution of each soft biometric trait. If q_{ki} is an estimate of the likelihood $p(y_k|\omega_i)$ based on the accuracy of the feature extractor, the weighted likelihood $\hat{p}(y_k|\omega_i)$ is computed as,

$$\hat{p}(y_k|\omega_i) = \frac{q_{ki}^{\beta_k}}{\sum_{Y_k} q_{ki}^{\beta_k}}, \tag{4.5}$$

where Y_k is the set of all possible values of the discrete variable y_k and β_k is the weight assigned to the k^{th} soft biometric feature. If the feature y_k is continuous with standard deviation σ_k, the likelihood can be scaled by replacing σ_k with $\frac{\sigma_k}{\beta_k}$. This weighted likelihood approach is commonly used in the speech recognition community in the context of estimating the word posterior probabilities using both acoustic and language models. In this scenario, weights are generally used to scale down the probabilities obtained from the acoustic model [73].

This method of likelihood computation also has other implicit advantages. An impostor can easily circumvent the soft biometric feature extraction because it is relatively easy to modify/hide one's soft biometric attributes by applying cosmetics and wearing other accessories (like a mask, shoes with high heels, etc.). In this scenario, the scaling factor β_k can act as the measure of the reliability of the soft biometric feature and its value can be set depending on the environment in which the system operates. If the environment is

hostile (many users are likely to circumvent the system), the value of β_k must be closer to 0. Finally, the discriminant functions given in equations (4.2) and (4.4) are optimal only if the assumption of independence between all the biometric traits is true. If there is any dependence between the features, the discriminant function is sub-optimal. In this case, appropriate selection of the weights β_k during training can result in better recognition rates.

4.4 Experimental Results

Our experiments demonstrate the benefits of utilizing the gender, ethnicity, and height information of the user in addition to the face and fingerprint biometric identifiers. A subset of the "Joint Multibiometric Database" (JMD) collected at West Virginia University has been used in our experiments. The selected subset contains 4 face images and 4 impressions of the left index index finger obtained from 263 users over a period of six months. The Identix FaceIt® SDK [74] is used for face matching. Fingerprint matching is based on the algorithm in [52].

The ethnicity classifier proposed in [65] was used in our experiments. This classifier identifies the ethnicity of a test user as either Asian or non-Asian with an accuracy of 82.3%. If a "reject" option is introduced, the probability of making an incorrect classification is reduced to less than 2%, at the expense of rejecting 25% of the test images. A gender classifier was built following the same methodology used in [65] for ethnicity classification and the performance of the two classifiers were similar. The reject rate was fixed at 25% and in cases where the ethnicity or the gender classifier made a reject decision, the corresponding information is not utilized for updating the discriminant function, i.e., if the label assigned to k^{th} soft biometric trait is "reject", then the log-likelihood term corresponding to the k^{th} feature in equations (4.2) and (4.4) is set to zero. During the collection of the

WVU multimodal database, the approximate height of each user was also recorded. However, our real-time height measurement system was not applied to measure the height of the user during each biometric data acquisition. Hence, we simulated values for the measured height of user ω_i from a normal distribution with mean $h(\omega_i) + \mu_e$ and standard deviation σ_e, where $h(\omega_i)$ is the true height of user ω_i, $\mu_e = 2$ cm and $\sigma_e = 5$ cm.

4.4.1 Identification Performance

To evaluate the performance of the biometric system in the identification mode, one face image and one fingerprint impression of each user are randomly selected to be the template (gallery) images and the remaining three samples are used as probe images. The separation of the face and fingerprint databases into gallery and probe sets, is repeated 50 times and the results reported are the averages for the 50 trials. The weights (β_k) used for the soft biometric likelihood computation are obtained by exhaustive search over an independent training database. The multimodal database described in [72] is used for weight estimation. For each soft biometric trait (gender, ethnicity and height), β_k is increased in steps of 0.05 and the value of β_k that maximizes the average rank-one recognition rate of the biometric system (utilizing face, fingerprint, and the corresponding soft information) is chosen. Following this procedure, values of 0.1, 0.1 and 0.5 are obtained as the best weights for gender, ethnicity, and height, respectively.

The effects of soft biometric identifiers on the performance of three primary biometric systems, namely, fingerprint, face, and a multimodal system using face and fingerprint as the modalities are studied. Figure 4.4 shows the Cumulative Match Characteristic (CMC) of the fingerprint biometric system and the improvement in performance achieved after the utilization of soft biometric information. The use of ethnicity and gender information along with the fingerprint leads to an improvement of 1.3% in the rank one performance

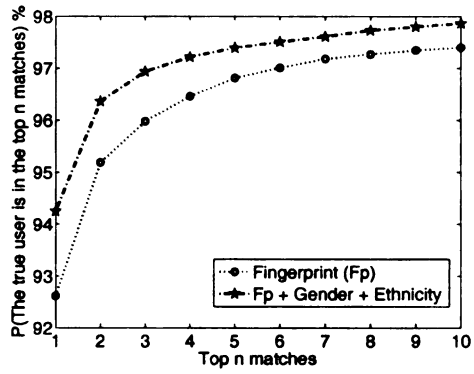
as shown in Figure 4.4(a). From Figure 4.4(b), we can observe that the height information also results in $\approx$ 1% improvement in the rank one performance. The combined use of all the three soft biometric traits results in an improvement of approximately 2.5% over the primary biometric system, as shown in Figure 4.4(c).

Ethnicity and gender information do not provide any statistically significant improvement in the performance of a face recognition system. One of the reasons for this observed effect could be that the FaceIt® algorithm already takes into account some of the features containing the gender and ethnicity information. This hypothesis is supported by the observation that more than 90% of the faces that are incorrectly matched at the rank-one level belong to either the same gender or ethnicity or both. On the other hand, we observe that the height information which is independent of the facial features, leads to an improvement of 0.5%-1% in the face recognition performance (see Figure 4.5). The failure of the ethnicity and gender information in improving the face recognition performance demonstrates that soft biometric traits would help in recognition only if the identity information provided by them is complementary to that of the primary biometric identifier.

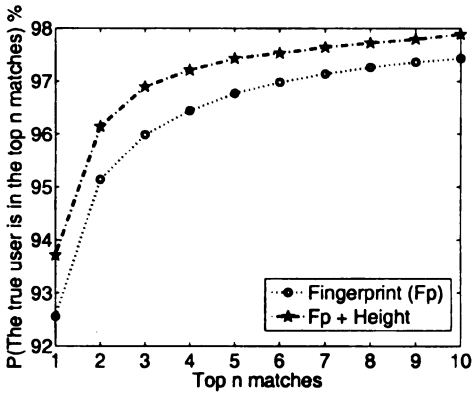
Figure 4.6 depicts the performance gain obtained when the soft biometric identifiers are used along with both face and fingerprint modalities. Although the multimodal system containing face and fingerprint modalities is highly accurate with a rank-one recognition rate of 97%, we still get a performance gain of more than 1% by the addition of soft biometric information.

4.4.2 Verification Performance

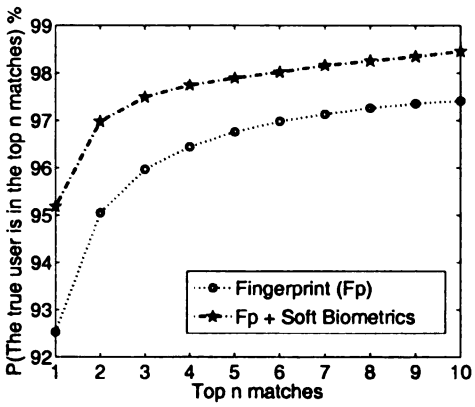
Genuine and impostor scores are obtained by computing the similarity between all distinct pairs of samples. The scores are then converted into log-likelihood ratios using the



(a)



(b)



(c)

Figure 4.4: Improvement in identification performance of a fingerprint system after utilization of soft biometric traits. a) Fingerprint with gender and ethnicity, b) Fingerprint with height, and c) Fingerprint with gender, ethnicity and height.

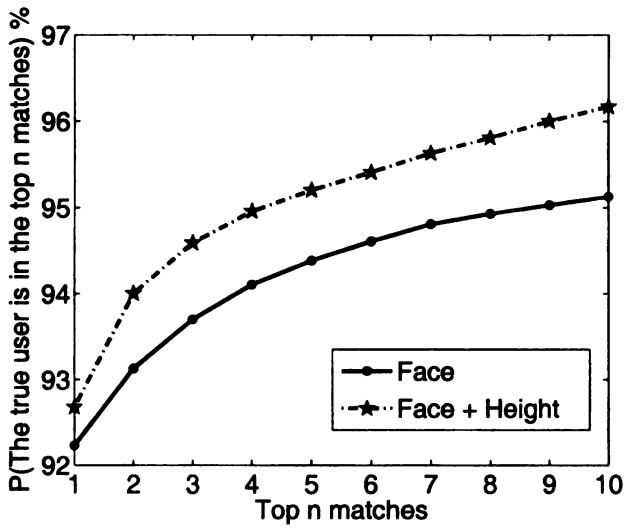


Figure 4.5: Improvement in identification performance of face recognition system after utilization of the height of the user.

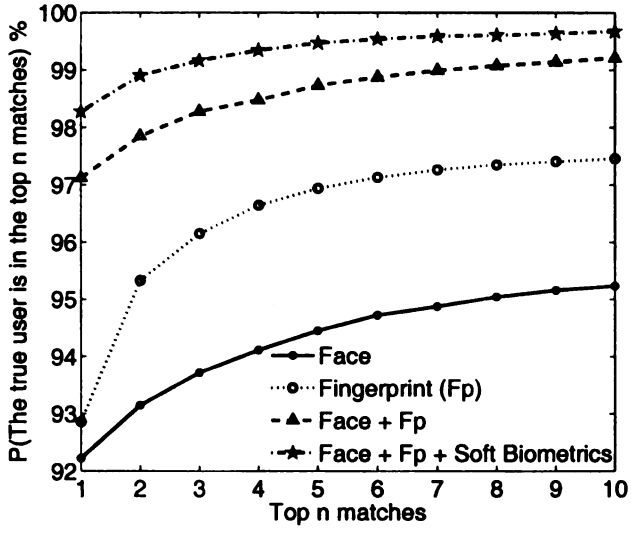


Figure 4.6: Improvement in identification performance of (face + fingerprint) multimodal system after the addition of soft biometric traits.

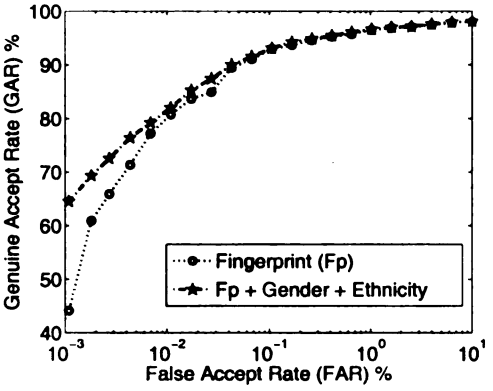
procedure outlined in [75]. This method involves the non-parametric estimation of generalized densities of the genuine and impostor matching scores of each modality. The weights (β_k) used for the soft biometric likelihood computation are estimated using a technique similar to the identification case, except that the weights are chosen to maximize the Genuine Acceptance Rate (GAR) at a False Acceptance Rate (FAR) of 0.01%. The best set of weights for the verification scenario are 0.5, 0.5 and 0.75 for gender, ethnicity, and height, respectively. This difference in the weights between the identification and verification modes can be attributed to the different discriminant functions used for fusion as given by equations (4.2) and (4.4), respectively.

Figure 4.7 shows the Receiver Operating Characteristic (ROC) curves when fingerprint is used as the primary biometric identifier. At lower FAR values, the performance of the fingerprint modality is rather poor. This is due to the large similarity scores for a few impostor fingerprint pairs that have very similar ridge structures. The addition of soft biometric information helps to alleviate this problem, resulting in a substantial improvement ($>$ 20% increase in GAR at 0.001% FAR) in the performance at lower FAR values (see Figure 4.7).

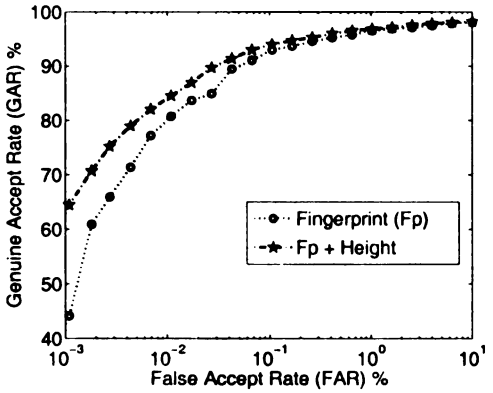
In the case of face modality and the multimodal system using both face and fingerprint, the improvement in GAR is about 2% at 0.001% FAR (see Figures 4.8 and 4.9). This improvement is still quite significant given that the GAR of the primary biometric systems at this operating point is already very high.

4.5 Summary

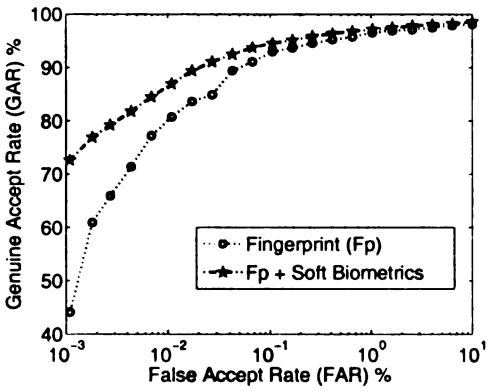
The objective of this chapter is to demonstrate that soft biometric identifiers such as gender, height, and ethnicity can be useful in person recognition even when they cannot be automatically extracted with 100% accuracy. To achieve this goal, we have developed a Bayesian framework that can combine information from the primary biometric identifiers



(a)



(b)



(c)

Figure 4.7: Improvement in verification performance of a fingerprint system after utilization of soft biometric traits. a) Fingerprint with gender and ethnicity, b) Fingerprint with height, and c) Fingerprint with gender, ethnicity and height.

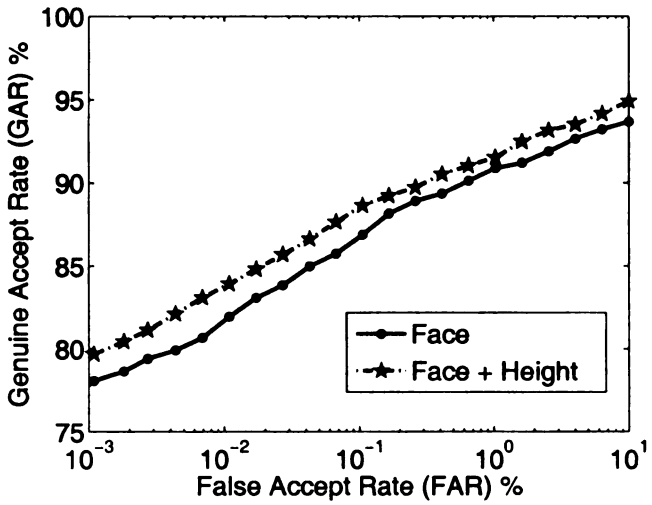


Figure 4.8: Improvement in verification performance of face recognition system after utilization of the height of the user.

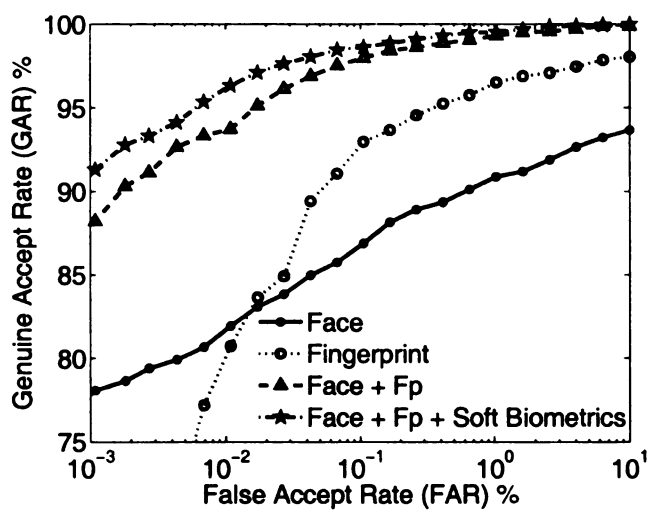


Figure 4.9: Improvement in verification performance of (face + fingerprint) multimodal system after the addition of soft biometric traits.

(face, fingerprint, etc.) and the soft biometric information such that it leads to higher accuracy in establishing the user's identity. Our experiments indicate that soft biometric traits can indeed substantially enhance the biometric system performance if they are complementary to the primary biometric traits. We have also presented a survey of the techniques that have been developed for automatically extracting soft biometric features and described our own implementation of a system that can identify soft biometric traits from real-time video sequences.

CHAPTER 5

Conclusions and Future Work

5.1 Conclusions

Although biometrics is becoming an integral part of the identity management systems, current biometric systems do not have 100% accuracy. Some of the factors that impact the accuracy of biometric systems include noisy input, non-universality, lack of invariant representation and non-distinctiveness. Further, biometric systems are also vulnerable to security attacks. A biometric system that integrates multiple cues can overcome some of these limitations and achieve better performance. Extensive research work has been done to identify better methods to combine the information obtained from multiple sources. It is difficult to perform information fusion at the early stages of processing (sensor and feature levels). In some cases, fusion at the sensor and feature levels may not even be possible. Fusion at the decision level is too simplistic due to the limited information content available at this level. Therefore, researchers have generally preferred integration at the matching score level which offers the best compromise between information content and ease in fusion. One of the problems in score level fusion is that the matching scores generated by different biometric matchers are not always comparable. These scores can have different characteristics and some normalization technique is required to make the combination of scores meaningful. Another limitation of the existing fusion techniques is their inability to handle soft biometric information, especially when the soft information is not very accurate. In this thesis, we have addressed these two important problems in a systematic manner.

We have carried out a detailed evaluation of the various score normalization techniques that have been proposed in literature in terms of their efficiency and robustness. First, we studied the impact of the different normalization schemes on the performance of the multimodal biometric system consisting of face, fingerprint, and hand-geometry modalities. Our analysis shows that min-max, z-score and tanh techniques are efficient and provide a good recognition performance. However, we also observed that the min-max and z-score normalization schemes are not robust if the training data used for estimating the normalization parameters contains outliers. The tanh method is more robust to outliers due to the application of influence functions to reduce the effect of noise. But careful selection of parameters is required for the tanh normalization scheme to work efficiently. Although, non-parametric density estimation is a theoretically sound method of normalizing the matching scores, it does not work well in practice due to limited availability of training scores and problems with choosing the appropriate kernel.

One of the major contributions of this thesis has been the development of a framework for utilizing ancillary information about the user (also known as soft biometric traits) like gender and height to improve the performance of traditional biometric systems. Although the ancillary information by itself is not sufficient for person recognition, it certainly provides some cues about the individual, which can be highly beneficial when used appropriately. We have developed a model based on Bayesian decision theory that can incorporate the soft biometric information into the traditional biometric framework. Experiments on a reasonably large multimodal database validate our claim about the utility of soft biometric identifiers. We have also built a prototype system that can extract soft biometric traits from individuals during the process of primary biometric acquisition.

5.2 Future Work

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) need to be developed.

We believe that existing score level fusion techniques are quite ad-hoc and do not consider the underlying mathematical/statistical rigor. A more principled approach would be the computation of likelihood ratios based on the estimates of genuine and impostor score distributions. Automatic bandwidth selection techniques can be utilized to find the width of the kernels to be employed in non-parametric density estimation. The use of generalized likelihoods that model the score distributions as a mixture of discrete and continuous components must be explored. This method can also take into account possible correlation between the multiple sources of information. Finally, such a method has the capability to act as a black-box and can be used without any knowledge about the biometric matcher. However, the major obstacle in following the likelihood ratio approach is the scarcity of multimodal biometric data. If the method has to work effectively, then a large number of matching scores would be required to estimate the densities accurately.

With regards to our prototype soft biometric system, performance improvement can be achieved by incorporating more accurate mechanisms for soft biometric feature extraction. The method of carrying out an exhaustive search for the weights of the soft biometric identifiers is computationally inefficient and requires a large training database. Since the

weights are used mainly for reducing the dynamic range of the log-likelihood values, it is possible to develop simple heuristics for computing the weights efficiently. The Bayesian framework in its current form cannot handle time-varying soft biometric identifiers such as age and weight. We will investigate methods to incorporate such identifiers into the soft biometric framework.

BIBLIOGRAPHY

[1] A. K. Jain, A. Ross, and S. Prabhakar. An Introduction to Biometric Recognition. *IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics*, 14(1):4–20, January 2004.

[2] S. Prabhakar, S. Pankanti, and A. K. Jain. Biometric Recognition: Security and Privacy Concerns. *IEEE Security and Privacy Magazine*, 1(2):33–42, March-April 2003.

[3] A. K. Jain and A. Ross. Multibiometric Systems. *Communications of the ACM, Special Issue on Multimodal Interfaces*, 47(1):34–40, January 2004.

[4] Y. Chen, S. C. Dass, and A. K. Jain. Fingerprint Quality Indices for Predicting Authentication Performance. In *Proceedings of Fifth International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA) (To appear)*, New York, U.S.A., July 2005.

[5] NIST report to the United States Congress. Summary of NIST Standards for Biometric Accuracy, Tamper Resistance, and Interoperability. Available at `ftp://sequoyah.nist.gov/pub/nist_internal_reports/NISTAPP_Nov02.pdf`, November 2002.

[6] BBC News. Long lashes thwart ID scan trial. Available at `http://news.bbc.co.uk/2/hi/uk_news/politics/3693375.stm`, May 2004.

[7] M. Golfarelli, D. Maio, and D. Maltoni. On the Error-Reject Tradeoff in Biometric Verification Systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):786–796, July 1997.

[8] 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, January 2002.

[9] 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.

[10] 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, Sweden, June 2001.

[11] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain. FVC2004: Third Fingerprint Verification Competition. In *Proceedings of International Conference on Biometric Authentication*, pages 1–7, Hong Kong, China, July 2004.

[12] C. Wilson, A. R. Hicklin, H. Korves, B. Ulery, M. Zoepfl, M. Bone, P. Grother, R. J. Micheals, S. Otto, and C. Watson. Fingerprint Vendor Technology Evaluation 2003: Summary of Results and Analysis Report. NIST Internal Report 7123; available at `http://fpvte.nist.gov/report/ir_7123_summary.pdf`, June 2004.

[13] P. J. Philips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and J. M. Bone. FRVT2002: Overview and Summary. Available at `http://www.frvt.org/FRVT2002/documents.htm`.

[14] D. A. Reynolds, W. Campbell, T. Gleason, C. Quillen, D. Sturim, P. Torres-Carrasquillo, and A. Adami. The 2004 MIT Lincoln Laboratory Speaker Recognition System. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Philadelphia, PA, March 2005.

[15] L. Hong, A. K. Jain, and S. Pankanti. Can Multibiometrics Improve Performance? In *Proceedings of IEEE Workshop on Automatic Identification Advanced Technologies*, pages 59–64, New Jersey, U.S.A., October 1999.

[16] A. Ross and A. K. Jain. Information Fusion in Biometrics. *Pattern Recognition Letters, Special Issue on Multimodal Biometrics*, 24(13):2115–2125, 2003.

[17] L. Hong and A. K. Jain. Integrating Faces and Fingerprints for Personal Identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12):1295–1307, December 1998.

[18] R. Snelick, U. Uludag, A. Mink, M. Indovina, and A. K. Jain. Large Scale Evaluation of Multimodal Biometric Authentication Using State-of-the-Art Systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):450–455, March 2005.

[19] 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.

[20] S. S. Iyengar, L. Prasad, and H. Min. *Advances in Distributed Sensor Technology*. Prentice Hall, 1995.

[21] A. Ross and A. K. Jain. Fingerprint Mosaicking. In *Proceedings of International Conference on Acoustic Speech and Signal Processing (ICASSP)*, pages 4064–4067, Florida, U.S.A., May 2002.

[22] Y. S. Moon, H. W. Yeung, K. C. Chan, and S. O. Chan. Template Synthesis and Image Mosaicking for Fingerprint Registration:An Experimental Study. In *Proceedings of*

International Conference on Acoustic Speech and Signal Processing (ICASSP), volume 5, pages 409–412, Quebec, Canada, May 2004.

[23] A. Kumar, D. C. M. Wong, H. C. Shen, and A. K. Jain. Personal Verification Using Palmprint and Hand Geometry Biometric. In *Proceedings of Fourth International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA)*, pages 668–678, Guildford, U.K., June 2003.

[24] A. Ross and R. Govindarajan. Feature Level Fusion Using Hand and Face Biometrics. In *Proceedings of of SPIE Conference on Biometric Technology for Human Identification*, volume 5779, pages 196–204, Florida, U.S.A., March 2005.

[25] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley & Sons, 2001.

[26] K. Woods, K. Bowyer, and W. P. Kegelmeyer. Combination of Multiple Classifiers using Local Accuracy Estimates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(4):405–410, April 1997.

[27] K. Chen, L. Wang, and H. Chi. Methods of Combining Multiple Classifiers with Different Features and Their Applications to Text-Independent Speaker Identification. *International Journal of Pattern Recognition and Artificial Intelligence*, 11(3):417–445, 1997.

[28] 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.

[29] L. Lam and C. Y. Suen. Optimal Combination of Pattern Classifiers. *Pattern Recognition Letters*, 16:945–954, 1995.

[30] 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.

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

[32] 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.

[33] Y. Wang, T. Tan, and A. K. Jain. Combining Face and Iris Biometrics for Identity Verification. In *Proceedings of Fourth International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA)*, pages 805–813, Guildford, U.K., June 2003.

[34] 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., U.S.A., March 1999.

[35] 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.

[36] 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.

[37] S. Prabhakar and A. K. Jain. Decision-level Fusion in Fingerprint Verification. *Pattern Recognition*, 35(4):861–874, 2002.

[38] E. S. Bigun, J. Bigun, B. Duc, and S. Fischer. Expert Conciliation for Multimodal Person Authentication Systems using Bayesian Statistics. In *Proceedings of First International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA)*, pages 291–300, Crans-Montana, Switzerland, March 1997.

[39] A. K. Jain and A. Ross. Learning User-specific Parameters in a Multibiometric System. In *Proceedings of International Conference on Image Processing*, pages 57–60, New York, USA, September 2002.

[40] 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., U.S.A., March 1999.

[41] M. Indovina, U. Uludag, R. Snelick, A. Mink, and A. K. Jain. Multimodal Biometric Authentication Methods: A COTS Approach. In *Proceedings of Workshop on Multimodal User Authentication*, pages 99–106, Santa Barbara, USA, December 2003.

[42] National Institute of Standards The Image Group of the Information Access Division and Technology. Biometric Scores Set - Release 1. Available at `http://www.itl.nist.gov/iad/894.03/biometricscores`, September 2004.

[43] P. Verlinde, P. Druyts, G. Cholet, and 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, February 1999.

[44] R. Snelick, M. Indovina, J. Yen, and A. Mink. Multimodal Biometrics: Issues in Design and Testing. In *Proceedings of Fifth International Conference on Multimodal Interfaces*, pages 68–72, Vancouver, Canada, November 2003.

[45] R. Brunelli and D. Falavigna. Person Identification Using Multiple Cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(10):955–966, October 1995.

[46] M. Montague and J. A. Aslam. Relevance Score Normalization for Metasearch. In *Proceedings of Tenth International Conference on Information and Knowledge Management*, pages 427–433, Atlanta, USA, November 2001.

[47] R. Manmatha, T. Rath, and F. Feng. Modeling Score Distributions for Combining the Outputs of Search Engines. In *Proceedings of Twenty-Fourth International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 267–275, New Orleans, USA, 2001.

[48] P. J. Huber. *Robust Statistics*. John Wiley & Sons, 1981.

[49] R. Cappelli, D. Maio, and D. Maltoni. Combining Fingerprint Classifiers. In *Proceedings of First International Workshop on Multiple Classifier Systems*, pages 351–361, June 2000.

[50] F. R. Hampel, P. J. Rousseeuw, E. M. Ronchetti, and W. A. Stahel. *Robust Statistics: The Approach Based on Influence Functions*. John Wiley & Sons, 1986.

[51] F. Mosteller and J. W. Tukey. *Data Analysis and Regression: A Second Course in Statistics*. Addison-Wesley, 1977.

[52] A. K. Jain, L. Hong, S. Pankanti, and R. Bolle. An Identity Authentication System Using Fingerprints. *Proceedings of the IEEE*, 85(9):1365–1388, 1997.

[53] M. Turk and A. Pentland. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[54] A. K. Jain, A. Ross, and S. Pankanti. A Prototype Hand Geometry-based Verification System. In *Proceedings of Second International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA)*, pages 166–171, Washington D.C., USA, March 1999.

[55] A. Bertillon. *Signaletic Instructions including the theory and practice of Anthropometrical Identification, R.W. McClaughry Translation*. The Werner Company, 1896.

[56] D. D. Heckathorn, R. S. Broadhead, and B. Sergeyev. A Methodology for Reducing Respondent Duplication and Impersonation in Samples of Hidden Populations. In *Annual Meeting of the American Sociological Association*, Toronto, Canada, August 1997.

[57] H. Aillisto, M. Lindholm, S. M. Makela, and E. Vildjiounaite. Unobtrusive User Identification with Light Biometrics. In *Proceedings of the Third Nordic Conference on Human-Computer Interaction*, pages 327–330, Tampere, Finland, October 2004.

[58] J. L. Wayman. Large-scale Civilian Biometric Systems - Issues and Feasibility. In *Proceedings of Card Tech / Secur Tech ID*, 1997.

[59] G. Givens, J. R. Beveridge, B. A. Draper, and D. Bolme. A Statistical Assessment of Subject Factors in the PCA Recognition of Human Subjects. In *Proceedings of CVPR Workshop: Statistical Analysis in Computer Vision*, June 2003.

[60] E. Newham. The Biometrics Report. SJB Services, 1995.

[61] S. Gutta, J. R. J. Huang, P. Jonathon, and H. Wechsler. Mixture of Experts for Classification of Gender, Ethnic Origin, and Pose of Human Faces. *IEEE Transactions on Neural Networks*, 11(4):948–960, July 2000.

[62] B. Moghaddam and M. H. Yang. Learning Gender with Support Faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):707–711, May 2002.

[63] G. Shakhnarovich, P. Viola, and B Moghaddam. A Unified Learning Framework for Real Time Face Detection and Classification. In *Proceedings of International Conference on Automatic Face and Gesture Recognition*, Washington D.C., USA, May 2002.

[64] K. Balci and V. Atalay. PCA for Gender Estimation: Which Eigenvectors Contribute? In *Proceedings of Sixteenth International Conference on Pattern Recognition*, volume 3, pages 363–366, Quebec City, Canada, August 2002.

[65] X. Lu and A. K. Jain. Ethnicity Identification from Face Images. In *Proceedings of SPIE Conference on Biometric Technology for Human Identification*, volume 5404, pages 114–123, April 2004.

[66] J. Hayashi, M. Yasumoto, H. Ito, and H. Koshimizu. Age and Gender Estimation based on Wrinkle Texture and Color of Facial Images. In *Proceedings of the Sixteenth International Conference on Pattern Recognition*, pages 405–408, Quebec City, Canada, August 2002.

[67] M. V. Buchanan, K. Asano, and A. Bohanon. Chemical Characterization of Fingerprints from Adults and Children. In *Proceedings of SPIE Photonics East Conference*, volume 2941, pages 89–95, November 1996.

[68] Y. H. Kwon and N. V. Lobo. Age Classification from Facial Images. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 762–767, April 1994.

[69] A. Lanitis, C. Draganova, and C. Christodoulou. Comparing Different Classifiers for Automatic Age Estimation. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 34(1):621–628, February 2004.

[70] N. Minematsu, K. Yamauchi, and K. Hirose. Automatic Estimation of Perceptual Age using Speaker Modeling Techniques. In *Proceedings of the Eighth European Conference on Speech Communication and Technology*, pages 3005–3008, Geneva, Switzerland, September 2003.

[71] J. S. Kim et al. Object Extraction for Superimposition and Height Measurement. In *Proceedings of Eighth Korea-Japan Joint Workshop on Frontiers of Computer Vision*, January 2002.

[72] A. K. Jain, K. Nandakumar, X. Lu, and U. Park. Integrating Faces, Fingerprints and Soft Biometric Traits for User Recognition. In *Proceedings of Biometric Authentication Workshop, LNCS 3087*, pages 259–269, Prague, Czech Republic, May 2004.

[73] F. Wessel, R. Schluter, K. Macherey, and H. Ney. Confidence Measures for Large Vocabulary Continuous Speech Recognition. *IEEE Transactions on Speech and Audio Processing*, 9(3):288–298, March 2001.

[74] Identix Inc. FaceIt Identification Software Developer Kit. Available at `http://www.identix.com/products/pro_sdks_id.html`.

[75] 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) (To appear)*, New York, U.S.A., July 2005.