Can soft biometric traits assist user recognition?

Anil K. Jain[a], Sarat C. Dass[b] and Karthik Nandakumar[a]

[a] Department of Computer Science and Engineering;
[b] Department of Statistics and Probability
Michigan State University, East Lansing, MI, USA 48824-1226

ABSTRACT

Biometrics is rapidly gaining acceptance as the technology that can meet the ever increasing need for security in critical applications. Biometric systems automatically recognize individuals based on their physiological and behavioral characteristics. Hence, the fundamental requirement of any biometric recognition system is a human trait having several desirable properties like universality, distinctiveness, permanence, collectability, acceptability, and resistance to circumvention. However, a human characteristic that possesses all these properties has not yet been identified. As a result, none of the existing biometric systems provide perfect recognition and there is a scope for improving the performance of these systems. Although characteristics like gender, ethnicity, age, height, weight and eye color are not unique and reliable, they provide some information about the user. We refer to these characteristics as "soft" biometric traits and argue that these traits can complement the identity information provided by the primary biometric identifiers like fingerprint and face. This paper presents the motivation for utilizing soft biometric information and analyzes how the soft biometric traits can be automatically extracted and incorporated in the decision making process of the primary biometric system. Preliminary experiments were conducted on a fingerprint database of 160 users by synthetically generating soft biometric traits like gender, ethnicity, and height based on known statistics. The results show that the use of additional soft biometric user information significantly improves ($\approx$ 6%) the recognition performance of the fingerprint biometric system.

Keywords: primary biometrics, soft biometrics, Bayes rule, distinctiveness, permanence

1. INTRODUCTION

A wide variety of biometric systems have been developed for automatic recognition of individuals based on their physiological/behavioral characteristics.[1] These systems make use of a single or a combination of traits like fingerprint, face, hand-geometry, iris, retina, palm-print, ear, voice, gait, signature, keystroke dynamics, etc., for recognizing a person. Biometric recognition systems have been widely deployed in forensic, government and commercial applications. Table 1 summarizes some of the major applications that use biometric technologies. In addition to these biometric traits, newer technologies based on diverse characteristics like vein scan, facial and hand thermography, ear shape, gait, keystroke dynamics and palm-print are under various stages of development.

Biometric systems that use a single trait for recognition, called unimodal biometric systems, are often affected by several practical problems like noisy sensor data, non-universality and/or lack of distinctiveness of the chosen biometric trait, unacceptable error rates, and spoof attacks. Universality is one of the basic requirements of any biometric trait that is selected for use in a biometric recognition system. However, no biometric trait is truly universal. For example, 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 and hence such people cannot be enrolled

Further author information: (Send correspondence to Karthik Nandakumar)
Anil K. Jain: E-mail: jain@cse.msu.edu, Telephone: 1 517 355 9282
Sarat C. Dass: E-mail: sdass@stt.msu.edu, Telephone: 1 517 432 5412
Karthik Nandakumar: E-mail: nandakum@cse.msu.edu, Telephone: 1 517 355 9319

Table 1. Summary of Major Biometric Applications

Biometric Trait	Major Applications
Fingerprint	Canadian Passenger Accelerated Service System (CANPASS)[2] Spanish National Social Security Identification Card (TASS)[3] FBI's Integrated Automated Fingerprint Identification System (IAFIS)[4]
Hand-geometry	Immigration and Naturalization Service PASS(INSPASS)[5] Security and Immigration System at Ben Gurion Airport in Tel Aviv[6] Member Verification System at Colombian Legislature[3]
Iris	Border Passage System at Heathrow Airport in London[7]
Voice	PORTPASS at US-Canadian Vehicle Border Crossing[8]
Face	FacePASS System for Physical Access Control Applications[9] FaceIT for Surveillance Applications[10]
Keystroke dynamics	BioPassword for Network Access Control[11]
Signature	CyberSIGN Electronic Signature Verification System for Adobe Acrobat[12]

in a fingerprint biometric system.[13] Every biometric trait has a theoretical upper bound in terms of its ability to distinguish two individuals. Golfarelli et al.[14] have shown that the most commonly used representations of hand-geometry and face biometrics, have a limited *information content* (number of distinguishable patterns) of 10^5 and 10^3, respectively. Although fingerprint[15] and iris possess better discrimination capability, existing automatic recognition systems based on these biometric identifiers are not able to deal with poor quality images and hence they do not meet the high accuracy requirements of critical applications. The state-of-the-art error rates associated with fingerprint, face and voice biometric systems[16] shown in Table 2 further highlight the unacceptable performance of unimodal biometric systems for large-scale deployment and for security critical applications. It is also possible for an impostor to circumvent a biometric system using spoof attacks. Behavioral traits like signature and voice and physical traits like fingerprint, face, hand-geometry, etc., are susceptible to such attacks.[17, 18]

Table 2. 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 2002[19]	Users mostly in the age group 20-39	0.2%	0.2%
Face	FRVT 2002[20]	Enrollment and test images were collected in indoor environment and could be on different days	10%	1%
Voice	NIST 2000[21]	Text dependent	10-20%	2-5%

Some of the problems associated with unimodal biometric systems can be overcome by the use of multimodal biometric systems that combine the evidence obtained from multiple sources.[22] These sources include multiple sensors for the same biometric (e.g., optical and solid-state fingerprint sensors), multiple instances of the same biometric (e.g., fingerprints from different fingers of a person), multiple snapshots of the same biometric (e.g., two impressions of a user's right index finger), multiple representations and matching algorithms for the same biometric (e.g., combining multiple face matchers like PCA and LDA), or multiple biometric traits (e.g., face and fingerprint). The use

of multiple sensors addresses the problem of noisy sensor data, but all other potential problems associated with unimodal biometric systems remain. A recognition system that works on multiple instances of the same biometric can ensure the presence of a live user by asking the user to provide a random subset of biometric measurements (e.g., left index finger followed by right middle finger). Multiple snapshots of the same biometric or multiple representations and matching algorithms for the same biometric may be used to improve the recognition performance of the system. However, all these methods still suffer from many of the problems faced by unimodal systems. A multimodal biometric system based on different biometric identifiers can be expected to be more robust to noise, address the problem of non-universality, improve the matching accuracy, and provide reasonable protection against spoof attacks. Thus, a multimodal biometric system that uses a large number of biometric traits (face, fingerprint, hand-geometry, iris, etc.) simultaneously, can be very effective. However, such a system will have two limitations. Firstly, the overall cost involved in building the multimodal system can be prohibitively high due to the need for multiple high quality sensors and increased storage and computational requirements. Secondly, the system will require a longer verification time thereby causing inconvenience to the users. Due to these limitations, the number of identifiers (modalities) in a multimodal biometric system is usually restricted to two or three.

A possible solution to the problem of designing a reliable and user-friendly biometric system is to use ancillary information about the user like height, weight, age, gender, ethnicity, and eye color to improve the performance of the primary biometric system. Most practical biometric systems collect such ancillary information about the users during enrollment. This information is stored either in the database or in the smart cards possessed by the user. However, this information is not currently utilized during the automatic identification/verification phase. This motivates us to utilize every available bit of information about the user to improve the performance of a biometric recognition system. Further, biometric systems used in access control applications generally have a human supervisor who oversees the operations of the system. When a genuine user is falsely rejected by the system, the human operator steps in to verify the identity of this user. This manual verification is usually done by comparing the facial appearance of the user with the facial image appearing on the user's identification card and by verifying other information on the ID card like age, gender, height, and other visible identification marks. If the soft biometric characteristics can be automatically extracted and used during the decision making process, the overall performance of the system will improve and the need for manual intervention will be reduced. The ancillary information by itself is not sufficient to establish the identity of a person and this is the reason for which characteristics providing such information are referred to as soft biometric traits. The contributions of this paper are two-fold. Firstly, we study the feasibility of automatically extracting soft biometric information from the user. Secondly, we analyze how the extracted soft information can be used in addition to the primary biometric identifiers for enhanced user recognition. The rest of the paper is organized as follows. Section 2 defines the concept of soft biometric traits and explores their utility in user recognition. This section also highlights the methods for automatic extraction of soft biometric information. In section 3 we propose a strategy for integration of soft biometrics with the main biometric system. The experimental results are presented in Section 4 and our conclusions are outlined in Section 5.

2. SOFT BIOMETRICS

The first personal identification system developed by Alphonse Bertillon[23] for identification of criminals was based on three sets of features: (i) body measurements (anthropometry) like height and length of the arm, (ii) morphological description of the appearance and shape of the body like eye color and anomalies of the fingers, and (iii) peculiar marks observed on the body like moles, scars, and tattoos. Although the Bertillon system was very useful in tracking criminals, it had an unacceptably high rate of false identification. This was due to two reasons. Firstly, several individuals can have the same set of values for these measurements. Secondly, for the same individual, these values can change over time. In other words, these characteristics do not have the distinctiveness and permanence to uniquely identify an individual over a period of time and hence we refer them as soft biometric traits. ***Soft biometric traits are those characteristics that provide some information about the individual, but lack the distinctiveness and permanence to sufficiently differentiate any two individuals***(see Figure 1 for examples of soft biometric traits). The soft biometric

traits can either be continuous or discrete. Traits such as gender, eye color, ethnicity, etc. are discrete in nature. On the other hand, traits like height and weight are continuous variables. A system that is completely based on soft biometric traits cannot provide the required accuracy in the recognition of individuals. In fact, the Bertillon system had a relatively short life since immediately after it was introduced, the Henry system for fingerprint classification[24] was adopted by the Scotland Yard for keeping track of criminals. However, soft biometric traits can be used to improve the performance of a traditional biometric system (e.g., fingerprint, hand-geometry) in many ways.

Figure 1. Examples of soft biometric traits

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 template X to the system, it is compared to the templates of all the three users stored in the database 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 biometric 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. Heckathorn et al.[25] have shown that a combination of personal attributes like gender, race, eye color, height, and other visible marks like scars, tattoos, etc. can be used to identify an individual with a fair degree of accuracy.

Wayman[26] 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. Filtering reduces the probability of obtaining a wrong match, but this is offset by the fact that the errors in filtering also reduce the probability of obtaining a correct match. Hence, in general, filtering drastically reduces the time required for identification but can degrade the recognition performance.

In addition to filtering, the soft biometric traits can also be used for tuning the parameters of the biometric system. Studies[27, 28] 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 seen as the most difficult subject for a fingerprint system.[28] This provides the motivation for tuning the system parameters like threshold on the matching score in a unimodal biometric system, and thresholds and weighting of the different modalities in a multimodal biometric system to obtain the optimum performance for a particular user or a class of users. Filtering and system parameters tuning require an accurate classification of a user into a particular class or bin (e.g., male or female, blue or brown eyes, Caucasian or Asian or African). This requires a pre-identification module that can accurately perform this classification.

2.1. Automatic Extraction of Soft Biometric Characteristics

In order to utilize soft biometrics, there must be a mechanism to automatically extract these features from the user during the recognition phase. As the user interacts with the primary biometric system, the system should be able to automatically measure the soft biometric characteristics like height, weight, age, gender, and ethnicity in a non-obtrusive manner without any interaction with the user. This can be achieved using a special system of sensors. For example, a bundle of infra-red beams could be used to measure the height. Another method for measuring the height of a person is to estimate it from a sequence of real-time images as described by Su-Kim et al..[29] The weight sensor could be installed at the place where the users stand while providing the primary biometric. A camera could be used for obtaining the facial image of the user, from which information like age, gender, and ethnicity could be derived. These observed soft biometric information could then be used to supplement the identity information provided by the user's primary biometric identifier.

Extensive studies have been made to identify the gender, ethnicity, and pose of the users from their facial images. Gutta et al.[30] proposed a mixture of experts consisting of ensembles of radial basis functions for the classification of gender, ethnic origin, and pose of human faces. They also used a SVM classifier with RBF kernel for gating the inputs. Their gender classifier classified users as either male or female with an average accuracy rate of 96%, while their ethnicity classifier classified users into Caucasian, South Asian, East Asian, and African with 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[31] 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.[32] developed a demographic classification scheme that extracts faces from unconstrained video sequences and classifies them based on gender and ethnicity. Their demographic classifier was a Perceptron constructed from binary rectangle features. The learning and feature selection modules used a variant of the AdaBoost algorithm. Their ethnicity classifier classified users as either Asian or non-Asian. Even under unconstrained environments, they showed that a classification accuracy of more than 75% can be achieved for both gender and ethnicity classification. For this data, the SVM classifier of Moghaddam and Yang had an error rate of

24.5% and there was also a notable bias towards males in the classification (females had an error rate of 28%). Balci and Atalay[33] reported a classification accuracy of more than 86% for a gender classifier that uses PCA for feature extraction and Multi-Layer Perceptron for classification.

Age determination is a more difficult problem due to the very limited physiological or behavioral changes in the human body as the person grows from one age group to another. There are currently no reliable biometric indicators for age determination.[34] Buchanan et al.[35] have been studying the differences in the chemical composition of child and adult fingerprints that could be used to distinguish children from adults. Kwon and Lobo[36] present 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 into "babies", "young adults", or "senior adults". However, they not provide any accuracy estimates for their classification scheme. It is our expectation that age determination systems providing a reasonable estimate of the age of a person would be developed in the near future.

3. GENERAL FRAMEWORK FOR INTEGRATION OF SOFT BIOMETRICS

In our framework, the biometric recognition system is divided into two subsystems. One subsystem is called the primary biometric system and it is based on traditional biometric identifiers like fingerprint, face and hand-geometry. The second subsystem, referred to as the secondary biometric system, is based on soft biometric traits like age, gender, and height. Figure 3 shows the architecture of a personal identification system that makes use of both fingerprint and soft biometric measurements. Let $\omega_1, \omega_2, \cdots, \omega_n$ represent the n users enrolled in the database. Let **x** be the feature vector corresponding to the primary biometric. Without loss of generality, let us assume that the output of the primary biometric system is of the form $P(\omega_i \mid \mathbf{x}),\ i = 1, 2, \cdots, n$, where $P(\omega_i \mid \mathbf{x})$ is the probability that the test user is ω_i given the feature vector **x**. If the output of the primary biometric system is a matching score, it is converted into posteriori probability using an appropriate transformation. For the secondary biometric system, we can consider $P(\omega_i \mid \mathbf{x})$ as the prior probability of the test user being user ω_i.

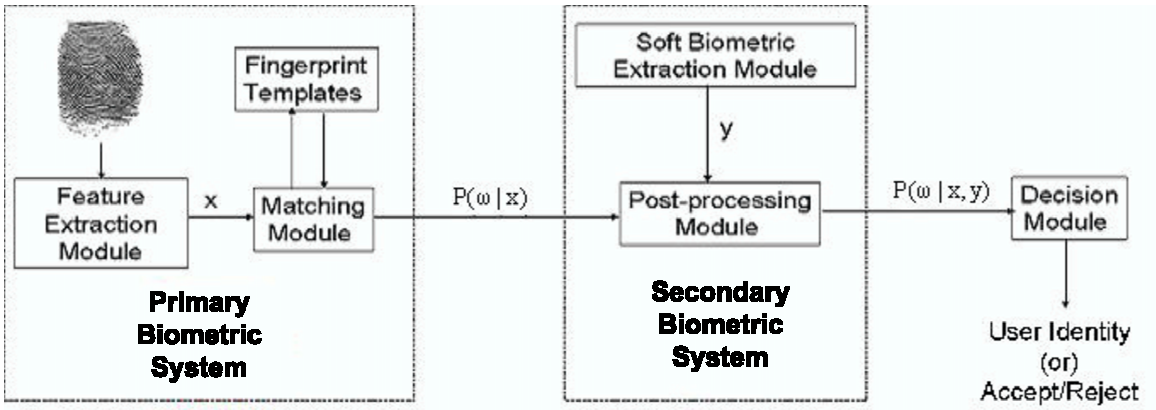

Figure 2. Integration of Soft Biometric Traits with a Fingerprint Biometric System.
(x is the fingerprint feature vector, y is the soft biometric feature vector)

Let $\mathbf{y} = [y_1,\ y_2,\ \cdots,\ y_k,\ y_{k+1},\ y_{k+2},\ \cdots,\ y_m]$ be the soft biometric feature vector, where y_1 through y_k are continuous variables and y_{k+1} through y_m are discrete variables. The final matching probability of user ω_i, given the primary biometric feature vector **x** and the soft biometric feature vector **y**, i.e., $P(\omega_i \mid \mathbf{x},\ \mathbf{y})$ can be calculated using the Bayes' rule as

$$P(\omega_i|\mathbf{x},\ \mathbf{y}) = \frac{p(\mathbf{y}|\omega_i)\ P(\omega_i|\mathbf{x})}{\sum_{i=1}^{n} p(\mathbf{y}|\omega_i)\ P(\omega_i|\mathbf{x})} \quad . \tag{1}$$

If we assume that the soft biometric variables are independent, equation (1) can be rewritten as

$$P(\omega_i|\mathbf{x},\ \mathbf{y}) = \frac{p(y_1|\omega_i)\ \cdots\ p(y_k|\omega_i)\ P(y_{k+1}|\omega_i)\ \cdots\ P(y_m|\omega_i)\ P(\omega_i|\mathbf{x})}{\sum_{i=1}^{n} p(y_1|\omega_i)\ \cdots\ p(y_k|\omega_i)\ P(y_{k+1}|\omega_i)\ \cdots\ P(y_m|\omega_i)\ P(\omega_i|\mathbf{x})} \quad . \tag{2}$$

In equation (2), $p(y_j|\omega_i),\ j = 1, 2, \cdots, k$ represents the conditional probability of the continuous variable y_j given user ω_i. This can be evaluated from the conditional density of the variable j for user ω_i. On the other hand, discrete probabilities $P(y_j|\omega_i), j = k+1, k+2, \cdots, m$ represents the probability that user ω_i is assigned to the class y_j. This is a measure of the accuracy of the classification module in assigning user ω_i to one of the distinct classes based on biometric indicator y_j. In order to simplify the problem, let us assume that the classification module performs equally well on all the users and therefore the accuracy of the module is independent of the user. The need for this simplification is illustrated by the following example. Let y_j represent the gender of the user. In general, we need to estimate probabilities like $P(user\ \omega_i\ is\ classified\ as\ a\ male\ \mid\ true\ gender\ of\ user\ \omega_i\ is\ male)$ for each user independently. This requires many training samples for each user. If we assume that the accuracy of the gender classifier is independent of the user, then we only need numerical values for the following four parameters:

1. $p_1 = P(user\ is\ classified\ as\ a\ male \mid true\ gender\ of\ the\ user\ is\ male)$
2. $p_2 = P(user\ is\ classified\ as\ a\ male \mid true\ gender\ of\ the\ user\ is\ female)$
3. $p_3 = P(user\ is\ classified\ as\ a\ female \mid true\ gender\ of\ the\ user\ is\ male)$
4. $p_4 = P(user\ is\ classified\ as\ a\ female \mid true\ gender\ of\ the\ user\ is\ female)$.

Note that $p_1 + p_3 = 1$ and $p_2 + p_4 = 1$. Therefore, there are only two unknown parameters to be determined.

Let

$$p(\mathbf{y}) = \sum_{i=1}^{n} p(y_1|\omega_i)\ \cdots\ p(y_k|\omega_i)\ P(y_{k+1}|\omega_i)\ \cdots\ P(y_m|\omega_i)\ P(\omega_i|\mathbf{x})\ .$$

The logarithm of $P(\omega_i|\mathbf{x},\ \mathbf{y})$ in equation (2) can be expressed as

$$\log P(\omega_i|\mathbf{x},\ \mathbf{y}) = \log p(y_1|\omega_i) + \cdots + \log p(y_k|\omega_i) + \log P(y_{k+1}|\omega_i) + \cdots + \log P(y_m|\omega_i) + \log P(\omega_i|\mathbf{x}) - \log p(\mathbf{y})\ . \tag{3}$$

This formulation has two main drawbacks. The first problem is that all the m soft biometric variables have been weighed equally. In practice, some soft biometric variables may contain more information than the others. For example, the ethnicity of a person may give more information about the person, than gender. Therefore, we must introduce a weighting scheme for the soft biometric traits based on an index of distinctiveness and permanence, i.e., traits that have smaller variability and larger distinguishing capability will be given more weight in the computation of the final matching probabilities. Another potential pitfall is that any impostor can easily spoof the system because the soft characteristics have an equal say in the decision as the primary biometric trait. It is relatively easy to modify/hide one's soft biometric attributes by applying cosmetics and wearing other accessories (like mask, shoes with high heels, etc.). To avoid this problem, we assign smaller weights compared to those assigned to the primary biometric traits. This differential weighting also has another implicit advantage. Even if a soft biometric trait of an user is measured

incorrectly (e.g., a male user is identified as a female), there is only a small reduction in that user's posteriori probability and the user is not immediately rejected. In this case, if the primary biometric produces a good match, the user may still be accepted. Only if several soft biometric traits do not match, there is significant reduction in the posteriori probability and the user could be possibly rejected. If the devices that measure the soft biometric traits are reasonably accurate such a situation has very low probability of occurrence. The introduction of the weighting scheme results in the following discriminant function for user ω_i:

$$g_i(\mathbf{x},\ \mathbf{y}) = a_0 \log P(\omega_i|\mathbf{x})\ +\ a_1 \log p(y_1|\omega_i) +\ \cdots\ +\ a_k \log p(y_k|\omega_i) + a_{k+1} \log P(y_{k+1}|\omega_i)\ +\ \cdots\ +\ a_m \log P(y_m|\omega_i), \quad (4)$$

where $\sum_{i=0}^{m} a_i = 1$ and $a_0 >> a_i,\ i = 1, 2, \cdots, m$. Note that a_i's, $i = 1, 2, \cdots, m$ are the weights assigned to the soft biometric traits and a_0 is the weight assigned to the primary biometric identifier.

4. EXPERIMENTAL RESULTS

Preliminary experimental results show significant improvement in recognition performance due to the utilization of soft biometric information. We used fingerprint as the primary biometric identifier and gender, ethnicity, and height as the soft biometric variables. Our database consisted of fingerprint impressions of 160 users obtained using a Veridicom sensor. Each user provided four impressions of each of the four fingers, namely, the left index finger, the left middle finger, the right index finger, and the right middle finger. The results reported in this paper are based only the four impressions of the left index finger of each user. Fingerprint matching was done using the minutiae features[37] and the output of the fingerprint matcher was a similarity score 's'. This similarity score is then converted into probabilities using a non-parametric technique, viz., Parzen window density estimation method.[38] Since our system operates in the verification mode, estimates of the conditional densities of the genuine ($p(s|genuine)$) and impostor ($p(s|impostor)$) scores were obtained using a Gaussian window function of width 1. Two-thirds of the fingerprint scores of genuine and impostor users were used for density estimation and the remaining scores were used for testing. This training-test data separation was done 20 times and each trial was performed independently i.e., for the verification attempts in different trials, the soft biometric feature vector were generated independently as described later in this section. After estimating the conditional densities, the Bayes formula is applied to calculate the posteriori probability of the score being that of a genuine user as

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

where $p(s) = p(s|genuine) * P(genuine) + p(s|impostor) * P(impostor)$,
and $P(genuine)$ and $P(impostor)$ are the prior probabilities of a genuine user and an impostor, respectively. In our experiments, both classes were assumed to be equally likely.

Our database also had additional information about the users including their age group, gender and ethnicity. However, there were no additional biometric indicators (like facial images) that could be used for the automatic estimation of these soft biometric variables during each recognition attempt. Also, the height and weight information of the users were not available. Therefore, we make the following assumptions in order to demonstrate the usefulness of the soft biometric characteristics:

1. There exists a gender classification system that has a classification accuracy of 90%. Although, the best published technique for gender classification has an accuracy of 97%,[31] this high accuracy rate has been obtained only for good quality and perfectly aligned face images. Therefore, we assume a more conservative value for the accuracy of the gender classification system. For the sake of simplicity, we assume that the system has no bias towards male or female users. As a result, the gender classification system has the following four characteristics:

(a) $P(user\ is\ classified\ as\ a\ male \mid true\ gender\ of\ the\ user\ is\ male) = 0.90$

(b) $P(user\ is\ classified\ as\ a\ male \mid true\ gender\ of\ the\ user\ is\ female) = 0.10$

(c) $P(user\ is\ classified\ as\ a\ female \mid true\ gender\ of\ the\ user\ is\ male) = 0.10$

(d) $P(user\ is\ classified\ as\ a\ female \mid true\ gender\ of\ the\ user\ is\ female) = 0.90$

2. There exists an ethnicity classification system that classifies users as either Asian or non-Asian with an accuracy of 90%. Again, we assume that the system has no bias and therefore,

   (a) $P(user\ is\ classified\ as\ Asian \mid user\ is\ Asian) = 0.90$

   (b) $P(user\ is\ classified\ as\ Asian \mid user\ is\ non\text{-}Asian) = 0.10$

   (c) $P(user\ is\ classified\ as\ non\text{-}Asian \mid user\ is\ Asian) = 0.10$

   (d) $P(user\ is\ classified\ as\ non\text{-}Asian \mid user\ is\ non\text{-}Asian) = 0.90$

3. Since we did not have the height information about the users in the database, we randomly assigned a height 'H_i' to user ω_i, where the H_i's are drawn from a Gaussian distribution with mean 165 cm and a standard deviation of 15 cm. During the recognition phase, we assume that the height of user ω_i can be measured up to an accuracy of ± 4 cm. Therefore, it is reasonable to assume that the measured height H_i^* will follow a Gaussian distribution with a mean H_i cm and a standard deviation of 4 cm.

Let $P(genuine|s)$ be the posterior probability that the test user is a genuine user given the fingerprint score 's' of the test user. Let $y_i = (G_i,\ E_i,\ H_i)$ be the soft biometric feature vector corresponding to the identity claimed by the test user, where G_i, E_i, and H_i are the true values of gender, ethnicity, and height of the claimed identity. Let $y^* = (G^*,\ E^*,\ H^*)$ be the observed soft biometric feature vector of the test user, where G^* is the observed gender, E^* is the observed ethnicity, and H^* is the observed height. The observed feature vector is generated such that $G^* = G_i$ with a probability of 0.90 and $E^* = E_i$ with a probability of 0.90. These probability values were derived from the assumptions about the accuracy of the corresponding classifiers. H^* was derived from a Gaussian distribution with mean H_i and standard deviation 4. Now the final score after considering the observed soft biometric characteristics is computed as

$$g_i(s,\ y^*) = a_0 \log P(genuine|s) + a_1 \log p(H^*|H_i) + a_2 \log P(G^*|G_i) + a_3 \log P(E^*|E_i)\ .$$

The optimum value of the weights were found to be $a_0 = 0.8$, $a_1 = 0.1$, $a_2 = 0.05$, and $a_3 = 0.05$. The final scores for all the users were obtained and the Receiver Operating Characteristics (ROC) curves were plotted by applying different threshold values. The performance results shown in Figure 3 indicate that there is a significant improvement in the recognition performance after the inclusion of the soft biometric information. At a False Acceptance Rate (FAR) of 0.1%, the Genuine Acceptance Rate (GAR) of the biometric system using only the fingerprint information is about 74%. On the other hand, the system using both the fingerprint and soft biometric information has an average GAR of 80.1% (with a standard deviation of 0.4%) at the same value of FAR. This experimental result demonstrates the benefits of using soft biometric traits along with the primary biometric identifier.

5. SUMMARY AND FUTURE DIRECTIONS

We have motivated the utilization of ancillary user information (also called soft biometrics) like gender, height, weight, age, ethnicity, and color of the eye/skin/hair to complement the identity information provided by the traditional (primary) biometric identifiers like fingerprint and face. Although these soft biometric characteristics are not as permanent and reliable as the traditional biometric identifiers like fingerprint, they provide some information about the identity of the user that can lead to higher accuracy in establishing the user identity. In order to exploit the soft biometric information, an automatic extraction mechanism is needed and there have been several attempts made in this direction.

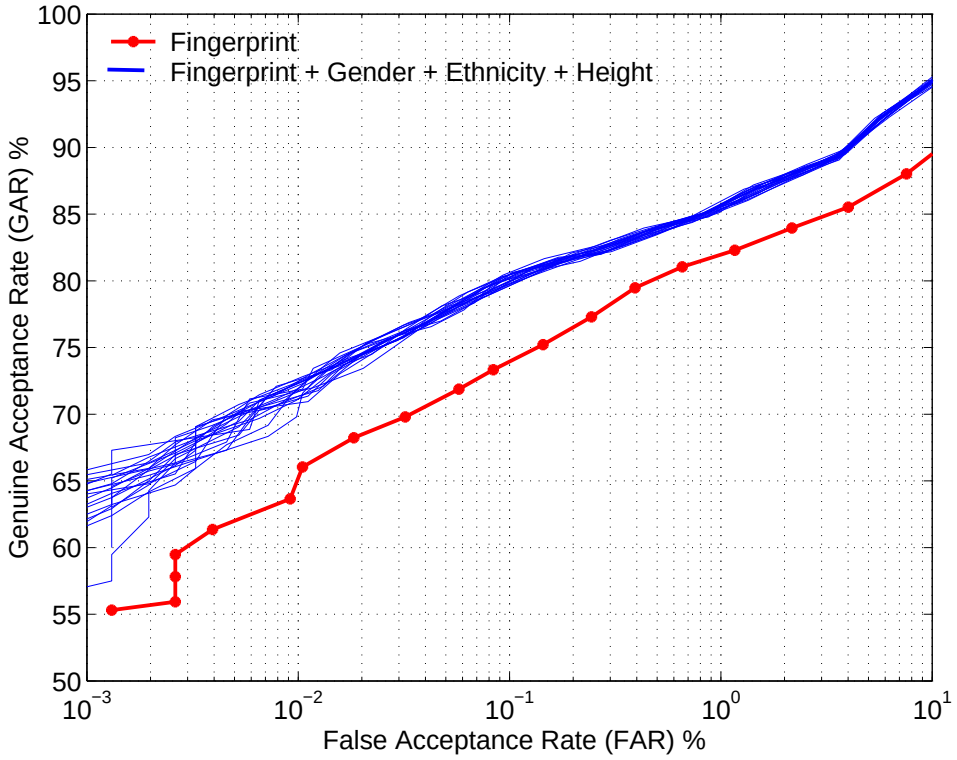

Figure 3. Improvement in authentication performance after utilization of soft biometric traits.

In this paper, we have proposed a framework for integrating the soft biometric information with the output of the primary biometric system. Our initial experiments on a fingerprint biometric system that uses soft biometric information, namely, gender, ethnicity, and height show very promising results and hence provides us the motivation to analyze this approach in greater depth.

Our future research work in this direction will involve the development of a prototype system that automatically extracts soft biometric information like gender, ethnicity, skin color, eye color, height, and weight along with a primary biometric characteristics. A rational procedure will be developed to determine the optimal set of weights for the soft characteristics based on their distinctiveness and permanence. The performance of an adaptive system of weights based on subpopulation of users will be studied. Methods to incorporate time-varying soft biometric information such as age and weight into the soft biometric framework will be explored. Finally, a more comprehensive evaluation of the system performance will be done to establish the advantages of utilizing the soft biometric traits.

REFERENCES

1. A. K. Jain, R. Bolle, and S. Pankanti, *Biometrics: Personal Identification in Networked Security*, Kluwer Academic Publishers, 1999.
2. "CANPASS – Remote Area Border Crossing (RABC) Permit." Available at `http://www.cic.gc.ca/english/visit/rabc.html`.
3. B. Miller, "PIN's Top 10 Biometric Applications." The 1997 Advanced Card and Identification Technology Sourcebook, 1997.

4. "Federal Bureau of Investigation Criminal Justice Information Services Division Homepage." Available at `http://www.fbi.gov/hq/cjisd/iafis.htm`.
5. "INSPASS Information." Available at `http://www.panynj.gov/aviation/inspassmain.htm`.
6. "Streamlined airport services take flight – Case Study." Available at `http://www.eds.com/case_studies/bgaa.pdf`.
7. "Airport tests passenger eye IDs." Available at `http://news.bbc.co.uk/2/hi/uk_news/1808187.stm`.
8. R. Zunkel, "Biometrics and Border Control." Security Technology and Design, May 1997.
9. "Facepass – Physical Access / Keyless Entry." Available at `http://www.viisage.com/facepass.htm`.
10. "Facial Surveillance – FaceIT ARGUS." Available at `http://www.identix.com/products/pro_security_bnp_argus.html`.
11. "BioPassword – Restoring and strengthening password integrity." Available at `http://www.biopassword.com`.
12. "Cyber-SIGN – Biometric Signature Verification." Available at `http://www.cybersign.com`.
13. 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.
14. 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**, pp. 786–796, July 1997.
15. S. Pankanti, S. Prabhakar, and A. K. Jain, "On the Individuality of Fingerprints," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(8), pp. 1010–1025, 2002.
16. L. O'Gorman, "Seven Issues with Human Authentication Technologies," in *Proceedings of IEEE Workshop on Automatic Identification Advanced Technologies*, pp. 185–186, (Tarrytown, U.S.A.), March 2002.
17. 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*, **4677**, pp. 275–289, January 2002.
18. 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*, pp. 289–303, 2000.
19. D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain, "FVC2002: Fingerprint Verification Competition," in *Proceedings of International Conference on Pattern Recognition*, pp. 744–747, (Quebec City, Canada), August 2002.
20. 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`.
21. N. Institute of Standards and Technology, "The 2000 NIST Speaker Recognition Evaluation." Available at `http://www.nist.gov/speech/tests/spk/2000/`, 2000.
22. L. Hong, A. K. Jain, and S. Pankanti, "Can Multibiometrics Improve Performance?," in *Proceedings of IEEE Workshop on Automatic Identification Advanced Technologies*, pp. 59–64, (New Jersey, U.S.A.), October 1999.
23. A. Bertillon, *Signaletic Instructions including the theory and practice of Anthropometrical Identification, R.W. McClaughry Translation*, The Werner Company, 1896.
24. I. Biometric Group, "The Henry Classification System." Available at `http://www.biometricgroup.com/Henry%20Fingerprint%20Classification.pdf`, 2003.
25. 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.
26. J. L. Wayman, "Large-scale Civilian Biometric Systems - Issues and Feasibility," in *Proceedings of Card Tech / Secur Tech ID*, 1997.

27. 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.
28. E. Newham, "The Biometrics Report." SJB Services, 1995.
29. 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.
30. 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**, pp. 948–960, July 2000.
31. B. Moghaddam and M. H. Yang, "Learning Gender with Support Faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**, pp. 707–711, May 2002.
32. 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.
33. K. Balci and V. Atalay, "PCA for Gender Estimation: Which Eigenvectors Contribute?," in *Proceedings of Sixteenth International Conference on Pattern Recognition*, (Quebec City, Canada), August 2002.
34. J. D. Woodward, "Testimony to the Commission on Online Child Protection on Age Verification Technologies." available at `http://www.copacommission.org/meetings/hearing1/woodward.test.pdf`, 2000.
35. M. V. Buchanan, K. Asano, and A. Bohanon, "Chemical Characterization of Fingerprints from Adults and Children," in *Proceedings of SPIE Photonics East Conference*, **2941**, pp. 89–95, November 1996.
36. Y. H. Kwon and N. V. Lobo, "Age Classification from Facial Images," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 762–767, April 1994.
37. A. K. Jain, L. Hong, S. Pankanti, and R. Bolle, "An identity authentication system using fingerprints," *Proceedings of the IEEE* **85**(9), pp. 1365–1388, 1997.
38. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, 2001.