2010

An Optimal Score Fusion Strategy For a Multimodal Biometric Authentication System for Mobile Device

Md. Saifur Rahim
University of Windsor

AN OPTIMAL SCORE FUSION STRATEGY FOR A MULTIMODAL BIOMETRIC AUTHENTICATION SYSTEM FOR MOBILE DEVICE

by
Md. Saifur Rahim

A Thesis
Submitted to the Faculty of Graduate Studies
through Computer Science
in Partial Fulfillment of the Requirements for
the Degree of Master of Science at the
University of Windsor

Windsor, Ontario, Canada
2010

AN OPTIMAL SCORE FUSION STRATEGY FOR A MULTIMODAL BIOMETRIC AUTHENTICATION SYSTEM FOR MOBILE DEVICE

by

Md. Saifur Rahim

APPROVED BY:

Dr. Ahmed Azab
Industrial & Manufacturing Systems Engineering

Dr. Alioune Ngom
Computer Science

Dr. Xiaobu Yuan, Advisor
Computer Science

Dr. Jianguo Lu, Chair of Defense
Computer Science

Sep 15, 2010

Author's Declaration of Originality

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

Abstract

For its unique advantages of preventing the loss of user identification, biometrics authentication is being increasingly used on mobile devices to meet the demand of access control and electronic transactions. Biometric community has been working on different approaches to improve reliability of security systems, multimodal authentication has attracted a lot of attention for its advantages over uni-modal biometric matchers. Nevertheless, errors caused by noises existing in real-world circumstances have become a major fact that slows down its acceptance in mobile computing.

Aimed at improving the reliability of biometric authentication, current practice uses score-level fusion to combine normalized outputs of multiple classifiers. By investigating the performance of different score-level fusion methods with normalization techniques in different noise conditions, this work develops an algorithm to analyze the individual biometric matching scores in different noise conditions and dynamically select the combinations of normalization and fusion methods that are adequate for different working environments.

Dedication

To my parents.

Acknowledgements

First and foremost, I am grateful to my advisor Dr. Xiaobu Yuan 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 also thank my guidance committee members Dr. Alioune Ngom, School of Computer Science and Dr. Ahmed Azab, Department Industrial & Manufacturing Systems Engineering for spending their time in carefully reviewing this thesis and Dr. Jianguo Lu, School of Computer Science for serving as the chair of the defense. Their valuable comments and suggestions have been very useful in enhancing the presentation of this thesis.

I would like to thank my friends Amanda Chernawski, Nathan Frenette, Sean Wendl, Michelle Gajewski and Rajkumar Vijayarangan for their great company and support during my stay at Windsor. I am also grateful to my aunts Farzana Zaman, Maggie Kozlowski and Mary Chernawski for all their long conversations over phone and messenger 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 for their love, affection and timely help.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

New technologies are designed to make our life easier and safer; however, they are often not secure enough. One of the most common identification methods is a Personal Identification Number (PIN), which can be easily eavesdropped by other people or forgotten by the user. This is why researchers have started to work on biometric methods as a way to identify people. Biometrics offers a natural and reliable solution to certain aspects of identity management by utilizing fully automated or semi-automated schemes to recognize individuals based on their inherent physical and/or behavioral characteristics [24]. A biometric system is a pattern recognition system that requires biometric data from an individual, such as a fingerprint, iris, face, hand, or voice etc. Identification systems that are based on these human characteristics have many advantages over the traditional authentication techniques based on what one knows or what one possesses [9, 24].

A wide variety of systems require person identification in order to confirm or determine the identity of an individual requesting to use said services. The purpose of such a system is to ensure that only legitimate users can access the service. One such system that needs security access is a mobile device (e.g., cellular phone). Nowadays, the new technology of

mobile devices allows us to connect to the internet, do banking transactions, maintain online address books, online shopping, send personal information, and do many other things.

Performance is a big issue in multi-biometric authentication for mobile devices when the authentication takes place under various environmental scenarios (e.g., on a dark, noisy street). Score-level fusion with different normalization techniques is a popular practice to increase the reliability of biometric authentication systems by combining the outputs of multiple classifiers. This thesis presents a new approach to the robust design of multimodal biometric systems on mobile devices. This approach uses the technique of Design of Experiments to systematically evaluate the performance of multimodal biometric systems under the influence of noise [46]. By examining the performance of different combinations of fusion methods and normalization techniques with a range of errors corresponding to real environmental noises, this thesis develops an algorithm to dynamically select the most suitable combination for optimized performance of user authentication on mobile devices in a given condition of usage.

Our goal is to build a biometric authentication system that works on face and voice samples taken by a camera and microphone already built into a mobile device. Images that are taken by a mobile phone are usually poor quality because most cameras in mobile devices are not equipped with a flash light. Images can be taken in different environments: outdoors, in an office with daylight, fluorescent, or incandescent light, or in a dark environment. In the same way, voice recordings are taken place in different noisy environments: in quiet offices, in mildly noisy hallways or at a busy street intersection with passing vehicles. We show that our multimodal biometric system selects the optimal normalization and fusion method combinations in different illumination and acoustic noise scenarios and obtains the higher accuracy compared to other authentication methods proposed so far.

In this thesis, we systematically analyzed the performance of multimodel biometric authentication systems for mobile devices under different noise conditions to select the combination of normalization techniques and fusion methods that produce the optimal performance than other combinations under every noise scenarios. The two main contributions of this thesis are 1) introducing the idea of applying different normalization and fusion methods in different environmental scenarios in multimodal biometric authentication to ensure the legitimacy of the user accessing various services over the internet from a mobile device, and 2) demonstrating the feasibility and performance of the method by means of experimentation and comparison with other approaches. Thorough investigation of face recognition and speaker recognition methods is beyond the scope of this thesis.

The remainder of this thesis has been organized into the following chapters. Chapter 2 provides an overview of the field of biometrics and multibiometrics. It describes the architecture, as well as biometric applications. Chapter 3 first provides a literature review of the existing approaches of biometrics based authentication for mobile computing, describes the method of systematic performance evaluation, and then presents the new system for the dynamic selection of biometric techniques according to different working environments for optimized user authentication. Data description, experimental results and analysis are then presented in Chapter 4. Finally, Chapter 5 gives conclusions and directions for future work.

Chapter 2

Overview of Biometric System

The etymology of the word biometrics comes from the ancient Greek words: bios life and metros measure. It is well-known the humans use inputs such as face, voice or gait to recognize each other. Recognition of people based their characteristics is important in many emerging technologies. These days, biometrics is used in a wide variety of applications that require the identification or verification schemes to confirm the identity of an individual. In this chapter, we present an overview of biometric methods and its applications.

2.1 Biometric Systems

Biometrics is a constantly evolving field that is becoming more widespread in the industry. The term biometrics is described as an automatic personal recognition system based on physiological or behavioral characteristics [9, 25, 41]. Biometrics use biological properties of a human, such as fingerprints, iris, voice recognition, face recognition, and hand geometry to identify individuals. Figure 2.1 shows examples of the biological properties used for biometrics [24].

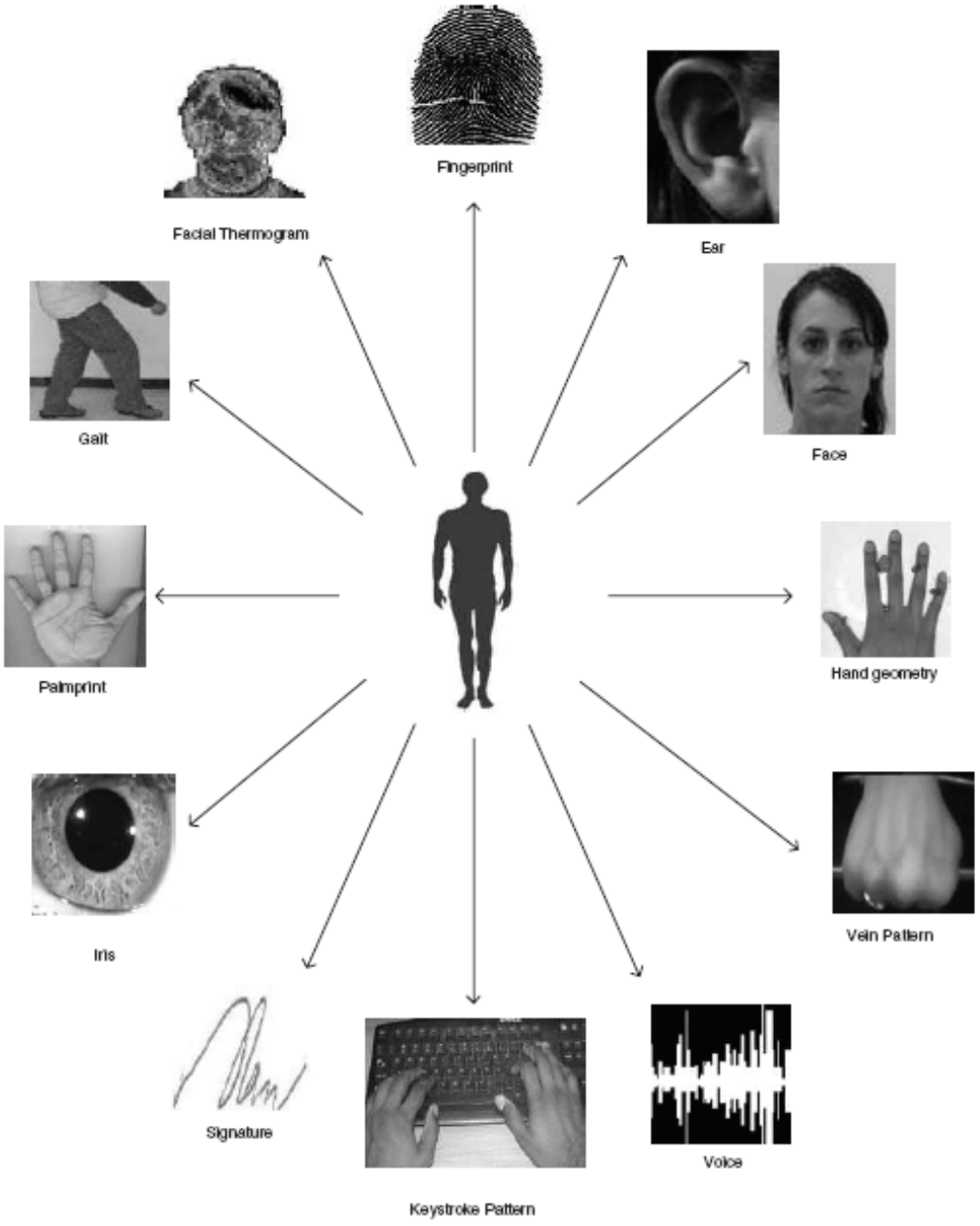

Figure 2.1: Biological properties used for biometrics

Nowadays, biometrics is no longer confined to criminal law enforcement. In addition, more businesses use biometrics to regulate access to buildings and information. Governments are considering including biometric identifiers in passports, drivers licenses, and possibly in a future national ID card. Also, digital video surveillance has already been spread in private and public places.

A biometric system is basically a pattern recognition system that can recognize a person based on specific physiological or behavioral features. A biometric system usually runs in two modes:

(1) *Verification mode*, where the system validates a persons identity by comparing the captured biometric characteristic with individuals biometric template, which is stored in the system database. Identity verification is typically used for positive recognition, where the aim is to prevent different people from using the same identity [41, 57].

(2) *Identification mode*, the system recognizes an individual by searching the entire template database for a match. The system conducts a one-to-many comparison to establish an individuals identity. Identification is a critical component of negative recognition applications, in which the system establishes whether the person is who he/she claims to be [41, 57]. The purpose of negative recognition is to prevent a single person from using multiple identities.

However, each biometric system regardless if it operates on verification or identification mode contains the following parts:

- A sensor unit that represents the interface between the user and the machine. This is the point where the biometric trait is acquired.

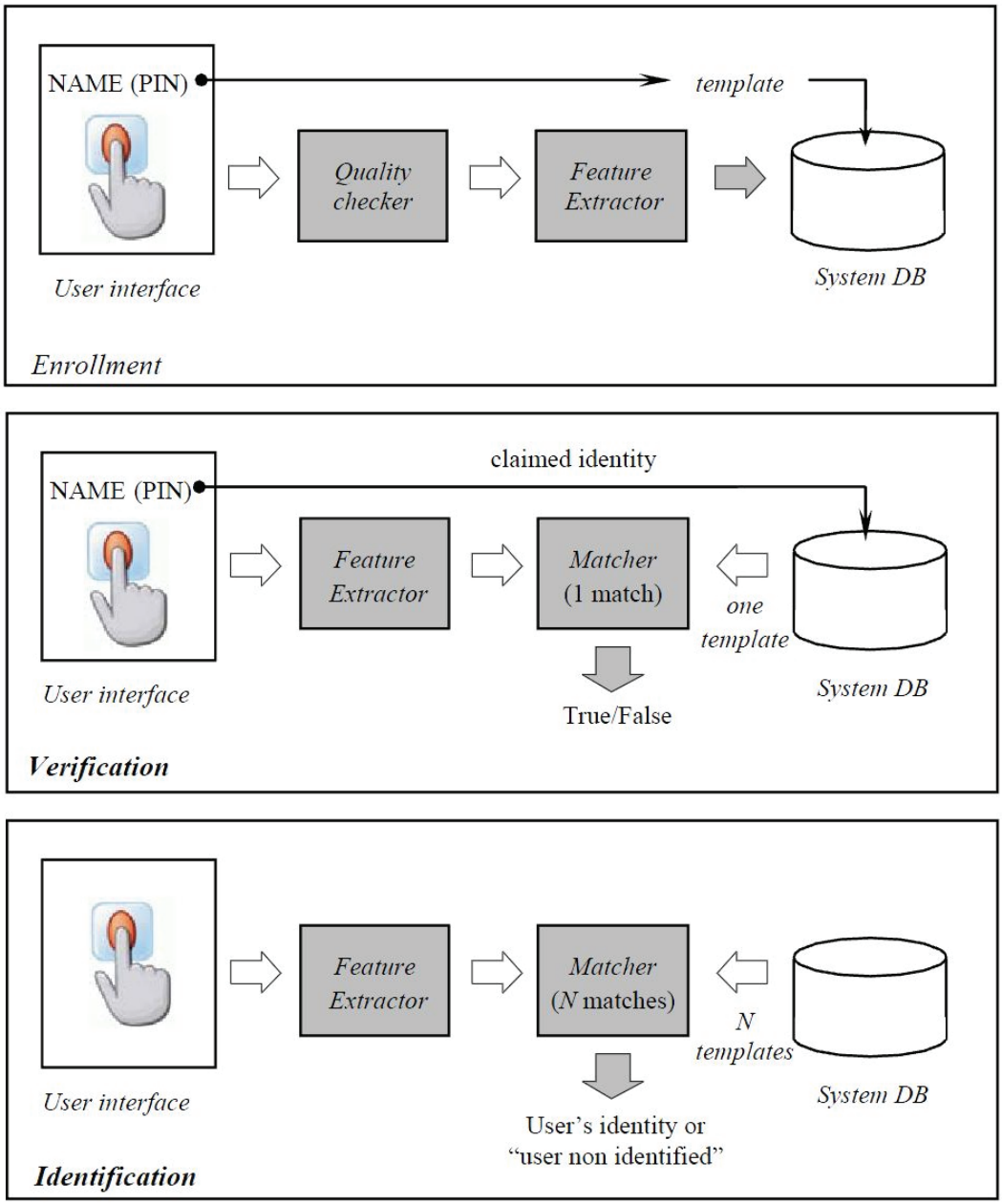

Figure 2.2: Diagrams identifying system modules within various application context modes [24]

- A processing unit where the acquired biometric is sampled segmented and features are being extracted. It also includes quality assurance to determine if the quality of the biometric is good enough to be used further in the process. If the quality of the acquired biometric is poor, the user may be asked to present the biometric again.

- A database unit where all the enrolled biometric templates are being stored and where the templates are being retrieved from in the authentication process.

- A matching unit that compares the newly acquired biometric template with the template stored in the database and based on decision rules determines either if the presented biometric is a genuine/impostor or if the user is identified or not.

Figure 2.2 shows the enrolment and recognition process flow in a biometric system.

Physiological biometrics is based on measurements and data derived from direct measurement of a part of human body. Here we list some example of those biometrics:

- **DNA:** Deoxyribonucleic acid: Except for the identical twins, it is the one-dimensional ultimate unique code for one individual. DNA is currently used mostly in the forensic applications for person recognition.

- **Ear recognition:** It is based on matching the distance of salient points on the pinna from a landmark location on the ear. The evidence from two studies [22, 25] supports the hypothesis that the ear contains unique physiological features, since in both studies all examined ears were found to be unique though identical twins were found to have similar, but not identical, ear structures especially in the Concha and lobe areas. Having shown uniqueness, it remains to ascertain if the ear provides biometrics which are comparable over time.

- **Facial, hand, and hand vein infrared thermogram:** The pattern of heat radiated by the human body is a characteristic of an individual and can be captured by an infrared camera in an unobtrusive way much like a regular (visible spectrum) photograph [42].

- **Fingerprint:** This is one of the most commonly used features that have been used to identify humans. Prior to the advent of biometric tools, fingerprints (captured on paper using ink marks) have been used extensively in forensics for the identification and verification of criminals. Provided the advent of new technologies, fingerprints are now captured using optical, capacitive or ultrasonic sensors, that measure the ridges, valleys and islands in a fingerprint [33].

- **Palmprint:** The palms of human hands contain patterns of ridges and valleys much like the fingerprints. The area of the palm is much larger than the area of a finger and as a result, palmprints are expected to be even more distinctive than the fingerprints. Some of palmprint techniques distinguish between identical twins [29].

- **Retina recognition:** The retinal vasculature is rich in structure and is supposed to be a characteristic of each individual and each eye. It is claimed to be the most secure biometric since it is not easy to change or replicate the retinal vasculature. Retina is unique for each individual, even for identical twins [25].

- **Hand and finger geometry:** Hand geometry recognition systems are based on a number of measurements taken from the human hand, including its shape, size of palm, and lengths and widths of the fingers. The geometry of a hand and fingers it is not very distinctive, and cannot be used for systems requiring identification of an individual from a large population [25].

- **Iris recognition:** Iris recognition systems scan the surface of an iris in order to com-

pare patterns. Each iris is distinctive and, like fingerprints and retinas, even the irises of identical twins are different [15].

- **Face recognition:** Humans are conditioned to recognize each other based on facial features. Consequently, facial features can be considered an "inherent" modality since it is widely used for recognition amongst humans. Captured usually as an image, facial features are normally used for identification or verification in a multimodal biometric system. Commonly used algorithms that support this process include measuring the distance between the facial features. Another approach employs scalar comparison between parts of the face using the sample image and the template set. Face recognition involves computer recognition of personal identity based on geometric or statistical features derived from face images [7, 27, 39, 52, 58].

Behavioral characteristics are based on an action taken by a person. On the other hand, behavioral biometrics are based on measurements and data derived from an action and indirectly measure characteristics of the human body. The following are the examples of biometric techniques based on behavioral characteristics:

- **Gait:** Gait is the way one walks and is a complex spatio-temporal biometric. Gait is not supposed to be very distinctive, but is sufficiently discriminatory to allow verification in some low-security applications [25].

- **Signature recognition:** The way a person signs his/her name is known to be a characteristic of that individual. Signatures change over a period of time and are influenced by physical and emotional conditions of the signatories [35].

- **Voice recognition:** Voice recognition systems use the characteristics of the voice in order to recognize a person. The behavioral part of the speech of a person changes

over time due to age, medical conditions (such as common cold), emotional state, etc; therefore, voice is not very distinctive and may not be appropriate for large-scale identification [11].

- **Keystroke:** It is hypothesized that each person types on a keyboard in a characteristic way. It is not unique to each individual but it offers sufficient discriminatory information to permit identity verification [34].

There are seven factors defined by Jain, Bolle, and Pankanti [25] that determine the suitability of a physical or a behavioral trait to be used in a biometric application.

1. **Universality:** each person accessing the application should posses the trait.
2. **Uniqueness:** the given trait should be sufficiently different across individuals comprising the population.
3. **Permanence:** the characteristic should be sufficiently invariant with respect to the matching criterion over a period of time.
4. **Collectability:** the characteristic should be measured quantitatively.
5. **Performance:** the recognition accuracy and the resources required to achieve that accuracy should meet the constraints imposed by the application.
6. **Acceptability:** individuals in the target population that will utilize the application should be willing to present their biometric trait to the system.
7. **Circumvention:** this reflects how easily the system can be fooled using fraudulent methods.

Table 2.1 presents a brief comparison of the physiological and behavioral biometric techniques based on these seven factors described above.

Biometrics	Universality	Uniqueness	Permanence	Collectability	Performance	Acceptability	Circumvention
Face	HIGH	LOW	MEDIUM	HIGH	LOW	HIGH	LOW
Finger print	MEDIUM	HIGH	HIGH	MEDIUM	HIGH	MEDIUM	HIGH
Hand geometry	MEDIUM	MEDIUM	MEDIUM	HIGH	MEDIUM	MEDIUM	MEDIUM
Keystrokes	LOW	LOW	LOW	MEDIUM	LOW	MEDIUM	MEDIUM
Hand veins	MEDIUM	MEDIUM	MEDIUM	MEDIUM	MEDIUM	MEDIUM	HIGH
Iris	HIGH	HIGH	HIGH	MEDIUM	HIGH	LOW	HIGH
Retinal scan	HIGH	HIGH	MEDIUM	LOW	HIGH	LOW	HIGH
Signature	LOW	LOW	LOW	HIGH	LOW	HIGH	LOW
Voice	MEDIUM	LOW	LOW	MEDIUM	LOW	HIGH	LOW
Facial thermograph	HIGH	HIGH	LOW	HIGH	MEDIUM	HIGH	HIGH
Odor	HIGH	HIGH	HIGH	LOW	LOW	MEDIUM	LOW
DNA	HIGH	HIGH	HIGH	LOW	HIGH	LOW	LOW
Gait	MEDIUM	LOW	LOW	HIGH	LOW	HIGH	MEDIUM
Ear Canal	MEDIUM	MEDIUM	LOW	MEDIUM	MEDIUM	HIGH	MEDIUM

Table 2.1: Comparison of biometric technologies [24]

2.1.1 Biometric Applications

Several biometric characteristics are in use in various applications. These applications of biometrics can be divided into three main groups [41, 48]:

1. Commercial applications, such as computer network login, electronic data security, ecommerce, internet access, ATM, credit card, physical access control, cellular phone.

2. Government applications, such as national ID card, correctional facility, drivers license, social security, welfare-disbursement, border control, passport control, etc.

3. Forensic applications, such as corpse identification, criminal investigation, missing children, etc.

Traditionally, commercial applications have used knowledge-based systems, such as PIN and passwords; however, these methods are not secure enough because passwords or PINs are easy to crack and easy to forget. In addition, the password can be shared by the user with his/her colleagues and then there is no way for the system to know who the actual user is. Government applications have used token-based systems, such as ID cards and badges. These systems have their problems as well. Firstly, the token or ID card can be stolen, shared, duplicated or lost; whereas, biometrics cannot be stolen, lost or forgotten. Only forensic applications have relied on human experts to match biometric features. These days, biometric methods are also used more often for security purposes. The first airport that applied biometrics for passengers verification was the Schipol airport in Amsterdam (Netherlands). This airport is equipped with iris scans that validate passengers (Privium) [5]. The border passage identifies a passenger using iris recognition. It is safe and considerably faster compared with manual passport control. Such a system also exists in Canada (CANPASS) [6] at the airports in the following cities: Edmonton, Winnipeg, Calgary, Halifax, Ottawa, Montreal, Toronto and Vancouver.

2.1.2 Biometrics and Mobile Device

Biometric systems can be integrated with mobile devices such as cell phones in two ways [40]: As a biometric collecting device or as a stand-alone system in order to protect unauthorized use of the mobile device such as cell phone. In the first, case Mobile devices are used as collecting the biometric and then they are passing it via internet a remote location (e.g., server) where it is processed and matched. This proves the usefulness for remote transactions when the identity of the user has to be proven. As an example, the user log in to the his bank account through the mobile web browser or through a banking software to

make a transaction; he is going to introduce himself as Saifur Rahim and in order to verify his identity he is asked to recite a pass phrase. The voice recording is done at the mobile device and then sends to the server to be processed and compared with the sample that was collected when the user enrolled in the system. Face, signature or key stroke are other biometric traits that todays mobile devices have the capabilities to collect and transfer them to remote location.

The other implementation of biometric system on mobile devices is that the entire biometric authentication system resides on the mobile device and it serves the purpose of preventing unauthorized access to the mobile devices functions and data.

Todays implementations of biometric systems on mobile devices include face recognitions, voice recognitions, gait recognitions, signature recognitions and keystroke recognitions for unimodal or multimodal authentication [40].

2.1.3 Limitation of Unimodal Biometric Systems

In real world applications there are several problems with unimodal biometric systems which operate on a single biometric modality. The limitations of unimodal biometric systems are as follows [24]:

- **Noise in sensed 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. Failure to focus the camera appropriately can lead to blurring in face and iris images.

- **Intra-class variations:** Biometric data acquired from an individual during an authentication session may be different from the data that was used to generate the template

during enrollment. 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.).

- **Distinctiveness:** While a biometric trait is expected to vary significantly across individuals, there may be large inter-class similarities in the feature sets used to represent these traits. This limitation restricts the discrimination power provided by the biometric trait. Inter-user similarity refers to the overlap of the biometric samples from two different individuals in the feature space.

- **Non-universality:** While every user is expected to possess the biometric trait being acquired, in reality it is possible that some users do not possess that particular biometric characteristic. 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) [3]. 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 [4].

- **Spoof attacks:** An individual may attempt to fake the biometric trait. It is easy for behavioral characteristics, such as when signature and voice are used as an identifier. Some of the limitations imposed by unimodal biometric systems can be overcome by using multiple biometric modalities [8, 10, 28]. Multibiometric systems are described in the next section.

2.2 Multimodal Biometric Systems

As the name suggests, multimodal biometric systems combine biometric information from multiple sources to establish the authenticity of a person. As identified in [45], multimodal biometric systems resolve, to a degree, the issue posed by non-universality. This is done by taking into account multiple biometric traits that can better identify a person when used in conjunction as opposed to a single modality. Multimodal biometric systems also act as deterrent to spoof attacks by making it more difficult to replicate the information since any illegitimate use will require the subject to imitate multiple features. More details have been provided in the following sub-section.

2.2.1 Necessity of Multimodal Biometric Systems

In section 2.1.3, some limitations of biometric systems relying on a single trait or modality have been identified. Multimodal biometric systems counter these limitations and present an improvement in the authentication performance. These improvements have been listed below:

- The noise present in the data due to factors such as defective equipment, alteration in the biometric trait or limitations in the physical environment have a lesser probability

of affecting multiple hardware and multiple traits. Hence, a multimodal biometric system ensures improved performance.

- Intra-class variations are mitigated provided any degree of difference in user's interaction with a particular component of a multimodal system is distributed over the entire system during the authentication process, therefore, lessening its effects. The probability of change in hardware throughout the system is also less compared to a single modality biometric system.

- Inter-class variations are also mitigated provided the commonality in physical or psychological traits within individuals is of much lesser probability than a single trait.

- Non-universality is addressed in multimodal biometric systems due to the increased size of the biometric traits' set. The probability of finding a biometric trait to authenticate a user increases with an increase in the number of modalities.

- Spoof attacks are also limited in multimodal biometric systems, simply owing to the number of biometric traits that must be imitated to carry out such an attack.

Multimodal biometric systems, consequently, provide an improved performance over unimodal systems in their ability to authenticate a user in presence of various limiting factors discussed above. In addition, multimodal biometric systems also provide improved security within the systems themselves.

2.2.2 Multimodal Biometric Systems−Schemes

As described in previous sections, a multimodal biometric system is created by combining various unimodal systems. The information retrieved in these individual systems is

combined to create a multimodal system. According to Figure 2.3, in such systems, the information can be combined through [36]:

1. **Multi-sensor**, in this system a single biometric trait is imaged using multiple sensors in order to extract information from registered images. For instance, the face images of an individual obtained using a thermal infrared camera and a visible light camera [14].

2. **Multi-modal**, these systems combine the evidence presented by different body traits for establishing identity. The cost of these systems is high since new sensors must be added [13, 20, 37, 43].

3. **Multi-instance**, these systems use multiple instances of the same body trait. For example, the left and right index fingers. These systems are cost-effective, because they require neither new sensors nor new algorithms for feature extraction [24].

4. **Multi-algorithm**, in this system the same biometric data is processed using multiple algorithms. This system does not require the use of new sensors and therefore is cost-effective. For example, a texture-based algorithm and minutiae based algorithm can operate on the same fingerprint image [24].

5. **Multi-sample**, in these systems a single sensor is used to obtain multiple samples of the same biometric trait. For example, face pictures, frontal profile, left and right profiles.

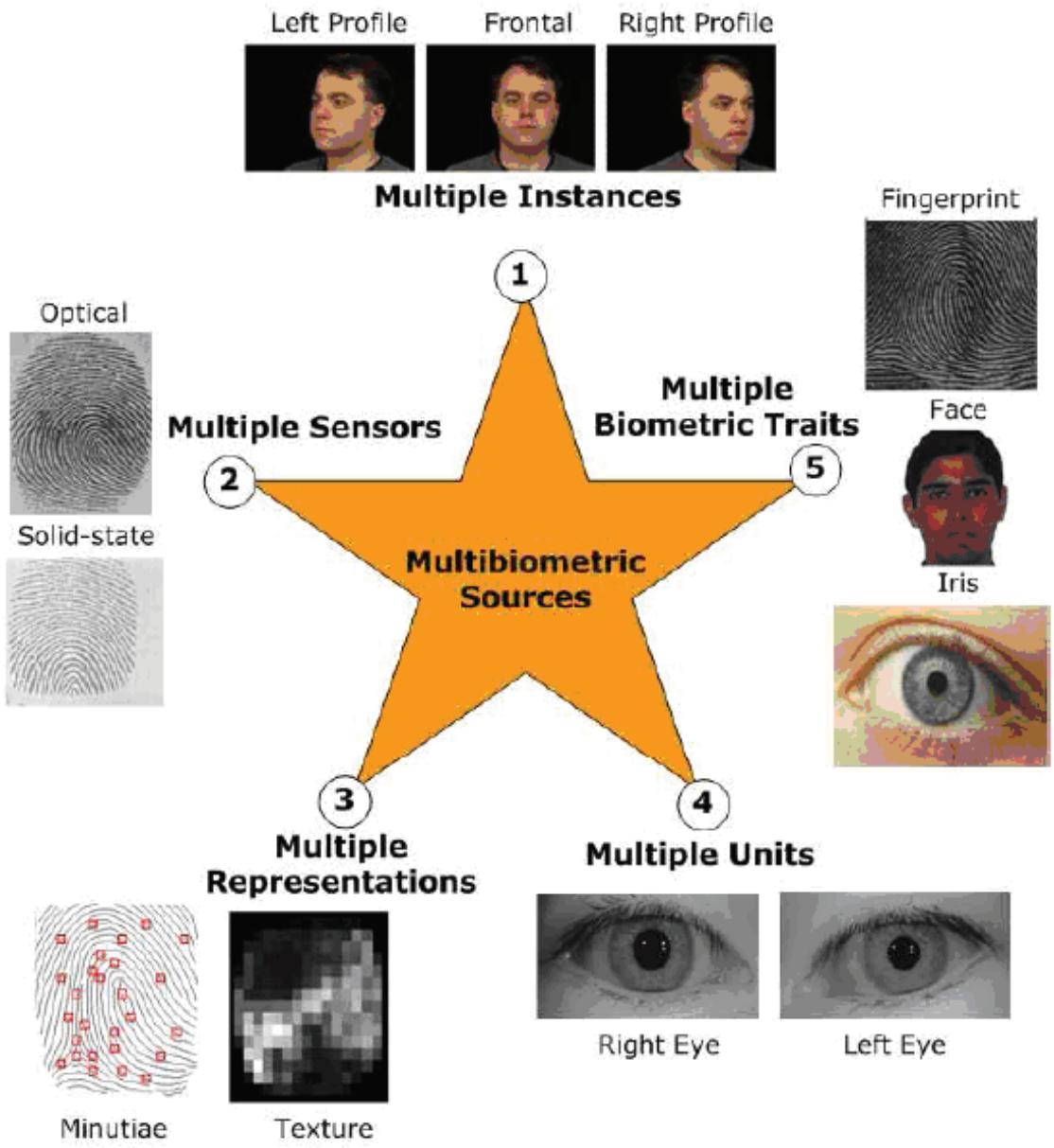

Figure 2.3: Sources of multiple evidence in multimodal biometric systems [36]. In the first four scenarios, multiple sources of information are derived from the same biometric trait. In the fifth scenario, information is derived from different biometric traits

2.2.3 Fusion in Biometrics

Combining information within multimodal biometric systems is referred to as the process of fusing information. The information captured from various sources following the schemes mentioned in the previous section can be fused at any of the following levels [16]:

Feature extraction level: Fusion at the feature extraction level stands for immediate data integration at the beginning of the processing chain. The information extracted from the different sensors is encoded into a joint feature vector, which is then compared to an enrollment template (which itself is a joint feature vector stored in a database) and is assigned a matching score as in a single biometric system (see Figure 2.4).

Decision level: In this level, a separate authentication decision is made for each biometric trait. These decisions are then combined into a final vote, as shown in Figure 2.5. Methods proposed in the literature for decision level fusion include AND and OR rules [2], majority voting [31], weighted majority voting [30], Bayesian decision fusion [60], the Dempster-Shafer theory of evidence [60] and behavior knowledge space [29].

Matching score level: This level is also known as confidence level or measurement level. Fusion at this level is much more effective than fusion at the decision level. Matching score is a measure of similarity between features derived from a presented sample and a

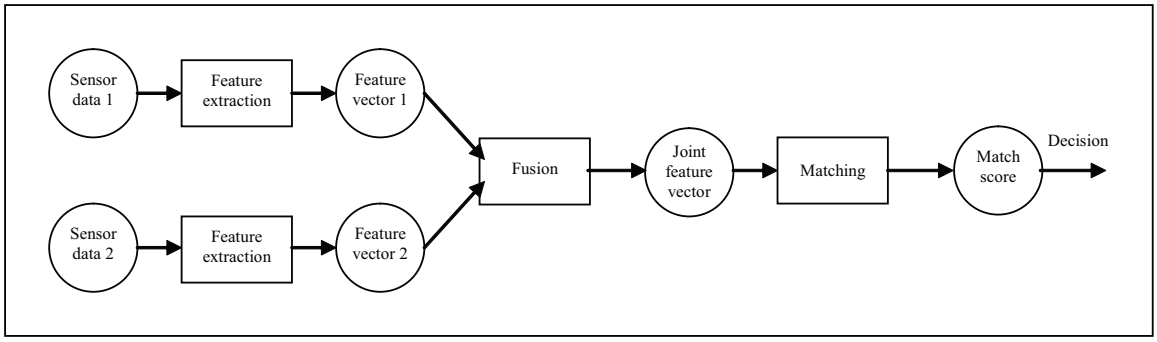

Figure 2.4: Block diagram of fusion at feature extraction level

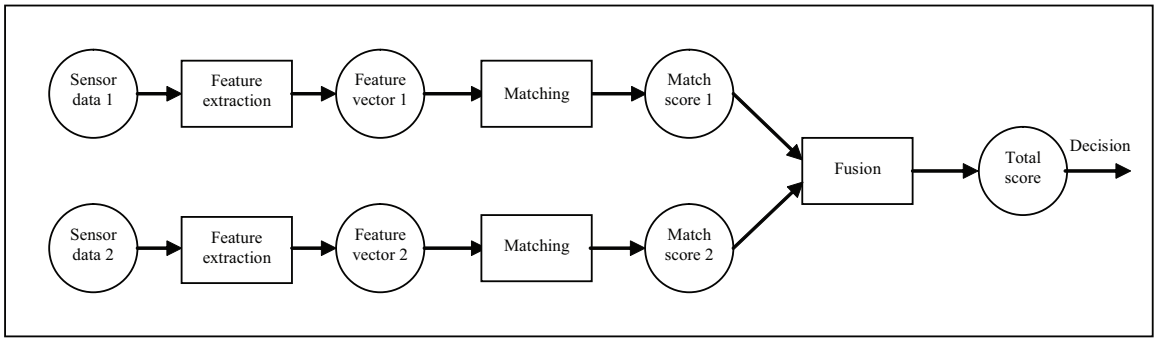

Figure 2.5: Block diagram of the fusion at the decision level

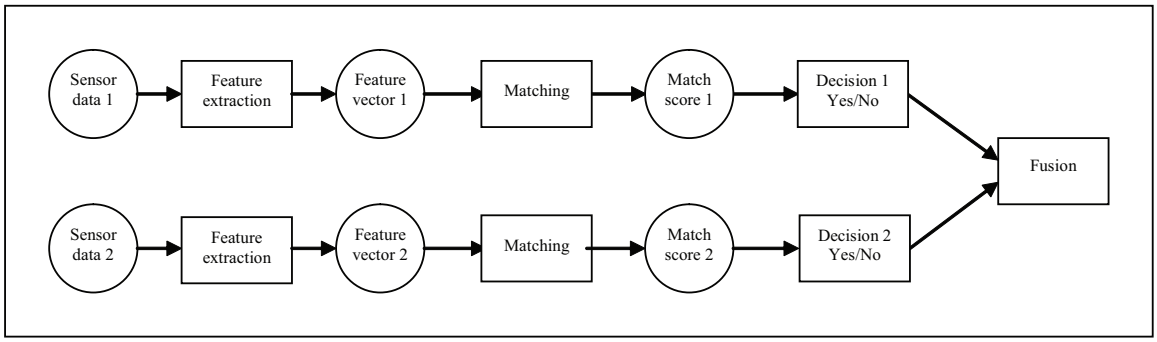

Figure 2.6: Block diagram of fusion at the matching score level

stored template. Each unimodal biometric system measures and calculates its own matching score and these matching score are fused to reach a final match/non match decision based on a certain decision threshold.

There are two approaches for consolidating the scores obtained from different matchers. One approach is to formulate it as a classification problem where each biometric modality a feature vector is constructed using the matching scores. This feature vector is then classified into one of two classes: Accept (genuine user) or Reject (impostor user). In general the classifier used in this scenario has the ability to learn the decision boundary irrespective of generation of feature vector. The output scores of the different modalities can be non-homogeneous (distance or similarity metric, different numerical ranges, etc). They are not

required to be processed before being fed into the classifier.

The second approach combines the individual matching scores to generate a single scalar score, which is then used to make the final decision. Since the matching scores from the different modalities, normalization is required to transform the scores into a common domain. Snelick et el. [51] analyzed the advantages of fusion at matching score level in several aspects. Firstly, matching score fusion does not affect the existing proprietary biometric systems, allowing for a common middleware layer to handle the multimodal application but with a small amount of common information. These existing and proprietary unimodal biometric systems can be easily combined into a multimodal biometric system given some basic information provided. Secondly, the data from prior evaluations of single-modal biometric systems can be reused. This avoids live testing or re-running individual biometric algorithms.

Another advantage is that the matching scores output by the matchers contain the second richest information about the input pattern next to the feature vectors; however it is much easier to access and to combine the scores generated by the different matchers compares to fusion at the feature extraction level.

Consequently, integration of information at the matching score level is the most common approach in multimodal biometric systems nowadays.

2.3 Combination Approach to Score Level Fusion

When comparing two approaches for score level fusion, experiments indicate that the combination approach performs better than the classification approach [44]; we will therefore discuss more about combination approach to score level fusion.

Prior combining score of different matchers into a single score, several issues need to be considered. First of all, the match scores generated by the individual matchers may not be compatible. For example, one matcher may output a distance (dissimilarity) measure while another may output a similarity measure. Furthermore, the outputs of the individual matchers may have different numerical scales (range). For example, one matcher may output the interval within (0, 1) while another output the interval within (0, 100). Finally, the match scores may follow different probability distributions. Normalization technique is then used to address these problems.

2.3.1 Normalization methods

To address the problem of incomparable classifier output scores in different combination classification systems, normalization methods are used to change the location and scale parameters of the matching score distributions at the outputs of the individual matchers. In such a way, various matching scores of different matchers are converted into a common domain and can be combined later on [26].

It is highly desirable that the normalization of the location and scale parameters of the matching score distribution must be robust and efficient. Huber [21] defines robustness as insensitivity to the optimal estimate when the distribution of the data is known. Huber also argues even though many techniques can be used for score normalization, the challenging work is to identify a technique that can be both robust and efficient.

Below we discussed a list of normalization methods that are commonly used and their robustness and efficiency have been examined [61]. We donate a raw matching score set $\{S_k\}$ of all the scores for a matcher, and the corresponding normalized score set as $\{S_k^{'}\}$.

1. Min-max normalization: Min-max is the simple normalization technique. It is best suited for the case where the bounds (maximum and minimum values) of the matcher are known. In this case, we can easily shift the minimum scores to 0 and 1, respectively. Min-max normalization keeps the original distribution of score except for a scaling factor and transforms all the scores into a common range [0, 1].

   The normalized score is give by

$$S_k' = \frac{S_k - min}{max - min},$$

   We can estimate the minimum and maximum values for a set of matching score from the training set even if the matching scores are not bounded. But the method is not robust in the case as it is highly sensitive to outliers in the training set used for estimation.

2. Decimal scaling 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.

$$S_k' = \frac{S_k}{10^n},$$

   Where n=$\log_{10}\max(s_i)$.

   The problem with this approach is lack of robustness and the assumption that the scores of different matchers vary by a logarithmic factor [26]. If the matching scores of the modalities are not distributed on a logarithmic scale, then this normalization

technique cannot be applied.

3. Z-score normalization: Z-score is one of the most commonly used score normalization technique. The normalized score is calculated using the arithmetic mean and standard deviation of the given data. If we have known the nature of the matching algorithm, it will work well by using the scheme, otherwise we have to estimate the average score and score variations of the matcher from a given set of matching scores. The normalized scores are given by

$$S_k^{'} = \frac{S_k - \mu}{\sigma},$$

Where μ is the arithmetic mean and σ is the standard deviation of the given data.

4. Median and median absolute deviation (MAD) normalization: The median and median absolute deviation (MAD) is insensitive to outliers and the points in the extreme trails of the distribution. Hence, median and MAD method are robust and is given by

$$S_k^{'} = \frac{S_k - median}{MAD},$$

Where MAD = median ($|S_k$ - median$|$).

5. Tanh-estimators normalization: The tanh-estimators introduced by Hampel et al. [17] are robust and highly efficient.

The normalization is given by

$$S_k^{'} = \frac{1}{2}\left\{\tanh\left(0.01\left(\frac{S_k - \mu_{GH}}{\sigma_{GH}}\right)\right) + 1\right\}$$

Where μ_{GH} and σ_{GH} are the mean and standard deviation estimates, respectively, of the genuine score distribution as given by Hampel estimators.

Hampel estimators are used to reduce the influence of outliers in the distribution based on the influence (ψ) - function below

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

The Hampel influence function can reduce 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, this method is therefore insensitive to outliers. However, tradeoff between the robustness and efficiency of this method should be decided cautiously, if too many points from the tail of the distributions are removed, estimation becomes robust but not efficient. Otherwise efficiency increases and robustness goes down when points from the tail are kept as many as possible. Practically parameters (a, b, and c) are chosen depending on the amount of noise in the training data set because it decides the extent of robustness the system requires.

2.3.2 Fusion Methods

In the famous theoretical framework [28] for consolidating the evidence obtained from multiple classifiers, Kittler et al. offer a number of fusion schemes including Min rule, Max rule, Sum rule and Product rule. These techniques can be applied to the system only if due output of each modality is in the form of P(genune|X),where X is the input pattern. That is,

what to be fused in the system is the posteriori probability of user being genuine given the input biometric sample X However, practically most biometric systems output a matching score s.

One solution is approximating P(genuine|X) by P(genuine|s) which can be calculated from the matching scores. But Jain et al. [24] argue that without corresponding confidence measure, the calculated value of P(genuine|s) is not a good estimate of P(genuine|X) and this can result in poor recognition performance. Hence, when consolidating the matching scores of individual modalities which dont offer confidence measure, it would be better to combine the matching scores directly using an appropriate method without converting them into probabilities.

In [61], the author has discussed the following commonly used fusion methodologies to combine multiple modalities at the matching scores level.
If S_i, is the matching score from i^{th} modality, S represents the resulting fused score.

1. **The *Simple Product Rule*** combines the scores by multiplying all of the individual scores,

$$S = S_1 * S_2 * \ldots * S_n$$

2. **The *Simple Sum Rule*** combines the scores as a linear transformation.

$$S = (a_1 s_1 - b_1) + \ldots + (a_n s_n - b_n)$$

a_i and b_i represents the weights and biases, respectively, which can be specified by the user.

3. **The *Simple Max Rule*** is the maximum score from the different modalities.

$$S = Max(S_1, S_2, \ldots, S_n)$$

4. **The *Simple Min Rule*** is the minimum score from the different modalities.

$$S = Min(S_1, S_2, \ldots, S_n)$$

5. **Biometric Gain against Impostor (BGI)/Likelihood Ratio of Genuine to Impostor (LRGI)**

   The BGI is a very useful concept. It is a measurement about how many times more likely we believe it that the claimant is an impostor, after having made biometric measurements, than we believed it beforehand. Its mathematical definition is the ratio of the a posteriori to the a priori probabilities of the claimant being an impostor [48].

$$BGI = \frac{\text{Probability of being an imposter, given the biometric evidence too}}{\text{Probability of being an impostor, given only prior knowledge}}$$

   The modified BGI as the Likelihood Ratio of Genuine to Impostor (LRGI) is a very good approximation to the BGI during most of the time.

$$BGI \approx LRGI = \frac{\text{Probability of seeing evidence from an impostor}}{\text{Probability of seeing it from the expected genuine subject}}$$

   Every score that comes out of the biometric devices is transformed to the LRGI scale. This is a score normalization process. Then the various scores are combined by mul-

tiplication or by addition of log likelihood ratios. This characteristic of BGI/LRGI fusion method exempts itself from score normalization in the sense it can normalize and fuse the matching scores together and no normalization is needed when using this fusion method.

Due to the fact some biometric traits can not be reliably obtained in some cases (e.g., good quality faces can not be obtained from users with dry faces), Jain and Ross [23] have proposed the use of user specific weights for computing the weighted sum of scores from the different modalities. For the example of dry face users, a lower weight can be assigned to the face score while raising the weight to the scores of the other modalities. The same scheme can be applied to threshold. Jain [24] has shown that the use of user specific weights and thresholds can improve the performance by approximately 3% and 2%, respectively. However, this method requires learning of user-specific weights from the training scores available for each user.

2.4 Disadvantages of Multimodal Biometric Systems

Multibiometric systems also have a few disadvantages when compared to unibiometric systems. They are more expensive and require more resources for computation and storage than unibiometric systems. Multibiometric systems generally require additional time for user enrollment, causing some inconvenience to the user; however multibiometric systems achieve better accuracies.

Chapter 3

Optimal Score Fusion Strategy

This chapter describes the approach that was implemented for this thesis research, including systematic performance analysis, matching performance metrics and dynamic technique selection in multimodal biometrics. Section 3.1 deals with the performance evaluation of multimodal biometric systems with performance measurable. The next section provides a literature review of the existing approaches of biometrics based authentication for mobile computing. Section 3.3 first describes the method of systematic performance evaluation, and then presents the new system for dynamic selection of biometric techniques according to different noise conditions for optimized user authentication.

3.1 Performance Evaluation of Biometric Systems

Biometrics-based authentication has started to find its way into mobile computing for its unique ability to prevent the theft, loss, and forgotting of user identification. The usage of mobile devices, however, usually involves situations in which there is no control over conditions such as lighting (e.g., indoor or outdoor, in sunny or cloudy days) and levels of

Biometrics	EER (%)	FAR (%)	FRR (%)	Subjects	Comments	Reference
Face	n/a	1	10	37437	Varied lighting, indoor/outdoor	FRVT (2002)
Fingerprint	n/a	1	0.1	25000	US Government operational data	FpVTE (2003)
Fingerprint	2	2	2	100	Rotation and exaggerated skin distortion	FVC (2004)
Hand geometry	1	2	0.1	129	With rings and improper placement	(2005)
Iris	<1	0.94	0.99	1224	Indoor environment	ITIRT (2005)
Iris	0.01	0.0001	0.2	132	Best conditions	NIST (2005)
Keystrokes	1.8	7	0.1	15	During months period	(2005)
Voice	6	2	10	310	Text independent, multilingual	NIST (2004)

Table 3.1: Error Rates of Biometric Systems using Standard Measurable

noise (e.g., at a bus stop or in a hotel lobby). The uncontrolled environment of usage results in unstable performance, and real-world circumstances have become major factors in slowing down the acceptance of biometrics-based authentication in mobile computing. The performance of biometric systems is an important issue not only in high security applications in government organizations, like forensics in crime analysis, but also in commercial applications like mobile computing [38].

Table 3.1 is a comparison table that lists the error rates based upon different modalities using different reference databases [53]. *(EER, FAR and FRR are measurable units for biometric systems explained in the following sections)*

3.1.1 FAR (False Accept Rate)

FAR represents the frequency with which a given biometric system identifies an impostor as a genuine subject. Mathematically, the FAR is the ratio of successful fraudulent attempts and the total number of fraudulent attempts. This is denoted by

$$FAR(n) = \frac{\text{successful fraudulent attempts made for identity n}}{\text{all fraudulent attempts made for identity n}}$$

where n is a unique identity.

The overall FAR of a biometric system can be calculated as an average using the formula

$$FAR(n) = \frac{1}{N} \sum_{n=1}^{N} FAR(n).$$

where N represents all identities being evaluated by the system. The FAR represents a statistical value, and therefore is dependent on the size N of the identities against which the biometric system is tested as well as the number of fraudulent attempts made.

3.1.2 FRR (False Reject Rate)

The FRR represents the frequency with which a biometric system rejects a genuine user, failing to correctly match the provided biometric input with the stored template. Essentially, the FRR is the ratio of the number of failed authentication attempts for genuine users and the total number of authentication attempts made by genuine users. The formula for the FRR is denoted by

$$FAR(n) = \frac{\text{rejected genuine attempts made for identity n}}{\text{all genuine attempts made for identity n}}$$

where n is a unique identity in the system. The overall FRR of a biometric system can be calculated by finding the average through the formula,

$$FRR(n) = \frac{1}{N} \sum_{n=1}^{N} FRR(n).$$

where N represents all identities within the biometric system.

Similar to the FAR, the FRR represents a statistical value dependent on the size N

of the identities against which the biometric system is tested as well as the number of authentication attempts made.

3.1.3 GAR (Genuine Accept Rate)

The GAR represents the frequency in which a biometric system accepts genuine users as authentic. The GAR is related to the FRR through the formula

$$GAR = 1 - FRR$$

To measure the performance of a biometric system, the FAR is usually mapped against the GAR in an ROC curve.

3.1.4 EER (Equal Error Rate)

The FAR and the FRR are both performance measures that rely on the chosen threshold values. On the other hand the Equal Error Rate, EER, on the other hand is independent of the threshold. In general, the EER is the value on the ROC curve where the FAR and FRR are equal. A low value of EER is considered to represent a biometric system with highly accurate performance. A trade-off between FAR and FRR is achieved by varying the acceptance threshold, so that as errors of one type decreases, errors of the other type increases. Thus, EER is a common way of evaluating the performance of a biometric system. Figure 3.1 is a representative ROC curve identifying the EER.

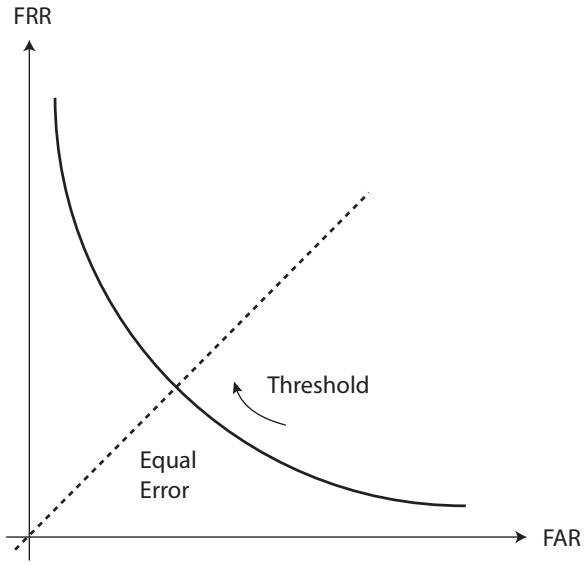

Figure 3.1: ROC curve indicating the Equal Error Rate (EER), where EER = FRR = FAR

3.2 Prior Work and Motivation

At the current stage of development, several multimodal biometric systems have been proposed for mobile and hand held devices, but most of them perform poorly in the presence of various noises. To improve the performance of multimodal biometric authentication system on handheld and mobile devices, H. Shaber proposed a simple mobile phone security system using biometrics for voice and finger print recognition [49]. Unfortunately, he did not present any experiments that could show the effectiveness of the proposed system.

In the approach presented in [40] by Pocovnicu, several biometric traits were proposed for mobile phone authentication, including finger prints, facial features, and handwritten signatures. This system has limited use in mobile computing as it requires the use of a fingerprint scanner, which is only available on very few mobile devices.

Vildjiounaite et al. proposed a method in [54] for the authentication of frequent mobile device users, and tested it in offline experiments on a database of 31 persons. The database

contained voice clips recorded at different noise levels and gait data with three placements of accelerometer module - in hand, in the breast pocket, and in the hip pocket. Experiment results showed that performance was significantly improved in comparison to performances using individual modalities. However, gait recognition works only when the user is in motion, and therefore has a limited use in practice.

Using face and voice, Hazen et al. developed a user verification system on an iPAQ handheld computer [18, 19]. This system showed good performance in experiments with a database of 35 persons in a strictly controlled low-noise environment. The system also conducted experiments with faces under different lighting conditions but with front images only. The authors acknowledged that the rotation of faces presents additional challenges compared to the recognition of frontal images, but they expect that users will cooperate with the system during the identification process and will generally be looking at the handheld computer screen while using it.

In addition, mobile devices are often used under noisy conditions in which voice recognition does not produce reliable results. For example, Lee et al. reported in [32] that the measured SNR inside a car can vary between plus 15 and minus 10 dB. Similar situations may occur in places where there is heavy traffic, e.g. at intersections, or other machinery in the vicinity. The performance of face recognition also depends on lighting/illumination conditions.

Though the performance of recognition with voice under noisy conditions and with face under changed illumination is not satisfying, a combined approach may produce better results. However, actual results depend on the chosen metric techniques and on the performance of each modality. If one of the modalities is significantly worse than the others, the performance of the multimodal system can even be poorer than that of the best modality

[45].

Despite continuous efforts made by different researchers to combine different biometrics techniques, there is a clear shortage of systematic studies of the outcome from combinations of relevant techniques, especially in real life conditions. The next section fills in the blanks, and presents a new system for the dynamic selection of fusion methods and normalization techniques based upon joint performance scores in different noise scenarios.

3.3 Proposed Approach

This section first presents a method for the systematic evaluation of performance in combinations of normalization and fusion techniques under the influence of noises. An authentication system is then proposed to dynamically select optimal combinations for different noisy environments.

3.3.1 Systematic Performance Evaluation

In an authentication system for handheld or mobile devices, the selection of fusion and normalization techniques is controlled by the system, but the condition of noise is uncontrollable as it depends on the usage of the device. For the purpose of systematic performance analysis and dynamic selection of biometric techniques, an evaluation matrix can be constructed with both controllable technical factors and uncontrollable noise factors [61]. As shown in Table 3.2, this matrix consists of three regions. The left region contains u controllable technical factors (**cf**), n combinations of the controllable factors (**cfc**), and an $n \times u$ array of their combination values (cfv). Similarly, the upper-right region contains v uncontrollable noise factors (**nf**), m combinations of the uncontrollable factors (**nfc**), and a

				$\mathbf{ncf}_1$	…	$\mathbf{ncf}_m$	
				$nfv_{1,1}$	…	$nfv_{1,m}$	$\mathbf{nf}_1$
				…	…	…	…
				$nfv_{l,1}$		$nfv_{l,m}$	$\mathbf{nf}_l$
				…	…	…	…
	$\mathbf{cf}_1$	…	$\mathbf{cf}_u$	$nfv_{v,1}$	…	$nfv_{v,m}$	$\mathbf{nf}_v$
$\mathbf{cfc}_1$	$cfv_{1,1}$	…	$cfv_{1,u}$	$\mathrm{r}_{1,1}$	…	$\mathrm{r}_{1,m}$	
…	…	…	…	…	…	…	
$\mathbf{cfc}_i$	$cfv_{i,1}$	…	$cfv_{i,u}$	$\mathrm{r}_{i,1}$	…	$\mathrm{r}_{i,m}$	
…	…	…		…	…	…	
$\mathbf{cfc}_n$	$cfv_{n,1}$	…	$cfv_{n,u}$	$\mathrm{r}_{n,1}$	…	$\mathrm{r}_{n,m}$	

Table 3.2: Evaluation matrix for multimodal biometric systems

$v \times m$ array of their combination values (nfv).

The lower right region of the matrix is the array R whose elements $r_{i,j}$, $1 \leq i \leq n$ and $1 \leq j \leq m$, record experiment results. The impact of operational precision is studied by conducting experiments on the configurations that associate all combinations of the controllable technical selections with all combinations of the uncontrollable noise influence. This procedure simulates the variation in performance of multimodal biometric systems in practical situations, and the results fill the evaluation matrix in its two arrays $[cfv_{i,l}]$ and $[nfv_{k,j}]$, where $1 \leq i \leq n$, $1 \leq l \leq v$, $1 \leq k \leq u$, and $1 \leq j \leq m$. When the combination of different noise conditions is too large for the evaluation matrix to handle, statistical analysis offers the mechanism to help reduce the size of arrays with orthogonal arrays [46].

Using the combinations of controllable and uncontrollable factors in the evaluation ma-

trix, experiments can be conducted to produce results according to pre-determined performance metrics for typical Genuine Accept Rate (GAR) or False Accept Rate (FAR). Experiment results are then filled into the elements of array R. As each column of R relates to a specific scenario of noise condition, experiment results based upon the chosen metrics then provide the information about which combination of normalization and fusion methods performs the best in a particular environment of usage for the mobile device, thus allowing the proposed authentication system to make the best choice for optimal performance in practice.

3.3.2 Dynamic Selection of Optimal Combinations

Figure 3.2 illustrates the proposed algorithm that uses simple arithmetic to determine different noise scenarios for the individual matcher from trained data for individual modalities, such as face and voice. Environmental noise conditions can be determined by the matching score generated by the matchers under that noise scenario. Each biometric classifier generates a distinctly separable matching score in different noise scenarios. The mean value of matching score indicates the particular noise condition. There are four actions for the system at the training stage:

1. Enroll users with biometric traits acquired in noise free condition.
2. Authenticate users in different noise conditions separately.
3. Calculate the average (mean) of genuine scores in each noise condition.
4. Follow 1-3 for each modality.

The mean values distinguish noise scenarios by the similarity of the matching score with the mean value in a noise condition.

Algorithm Biometric Technique Selector

Input:
n: Number of matchers
S: Array containing matching scores from biometric matchers $\{s_1, s_2, ... s_n\}$
A: Array containing mean values in different noise conditions of every matcher
$\{M_1, M_2, M_3, ..., M_n\}$
Where, M_i = Array of mean values in different noise conditions for the i^{th} matcher

HN-F: Hash table of normalization-fusion combination

Output: *Selected Normalization-Fusion Combination*

```
1    for i = 1 to n
2        min_value = max_possible {Contains minimum value for matcher i}
3        for j = 1 to Mi.length
4            min_value = min(min_value, CalculateDistance(Si, Mi[j])
5        end for
6        NoiseCondition[i] = FindNoiseCondition(Mi, min_value)
7    end for
8    optimal-combination = SelectNormFusion(HN-F, noiseCondition)
9    return optimal-combination
```

End Algorithm

Figure 3.2: Algorithm for dynamic selection of biometric techniques

The inputs of the algorithm are the matching scores S from each of the classifiers, calculated mean values $A = \{M_1, M_2, \ldots, M_n\}$ (where, M_i contains the mean values for i^{th} matcher) during the training stage of every noise condition for each modality. Hash table *HN-F* contains robust fusion methods with normalization techniques that were selected after analyzing the evaluation matrix result. At the time of authentication: for each modality, the matching score is compared with every mean values in M_i of the corresponding modality i in order to find the minimum distance and stores in *min_value*. Function *FindNoiseCondition* uses this minimum distance *min_value* with the mean value array M_i (for i^{th} matcher) and identifies the noise condition for that modality. The same steps

are repeated for every modality and noise conditions are stored in array *NoiseCondition*. The *SelectNormFusion* function then uses this array *NoiseCondition* containing all the determined noise conditions to select the normalization and fusion combination from the hashtable $HN-F$.

Time Complexity

For every modality, the algorithm iterates through all corresponding mean values in M_i. There are n modalities and let, $m = max(length(M_1), length(M_2), ...length(M_n))$. Thus the running time is $T(n) : \mathrm{O}(n \times m)$.

Shown in Figure 3.3 is the block diagram of the proposed system. In the diagram, multimodal biometric traits are acquired by the mobile device, and sent over the network to the remote server where feature vectors $F_1, F_2, \ldots, F_N$ are extracted. At the server, these feature vectors are employed to generate matching scores $s_1, s_2, \ldots, s_n$ from the corresponding templates acquired during the registration. The proposed system then follows the algorithm in Figure 3.2 to analyze the individual matching score and to identify the noise condition of the mobile device that is acquiring the biometric trait. The dynamic selection of normalization and fusion techniques finally takes place in correspondence with their performance of authentication during training for different noise conditions in real-life scenarios.

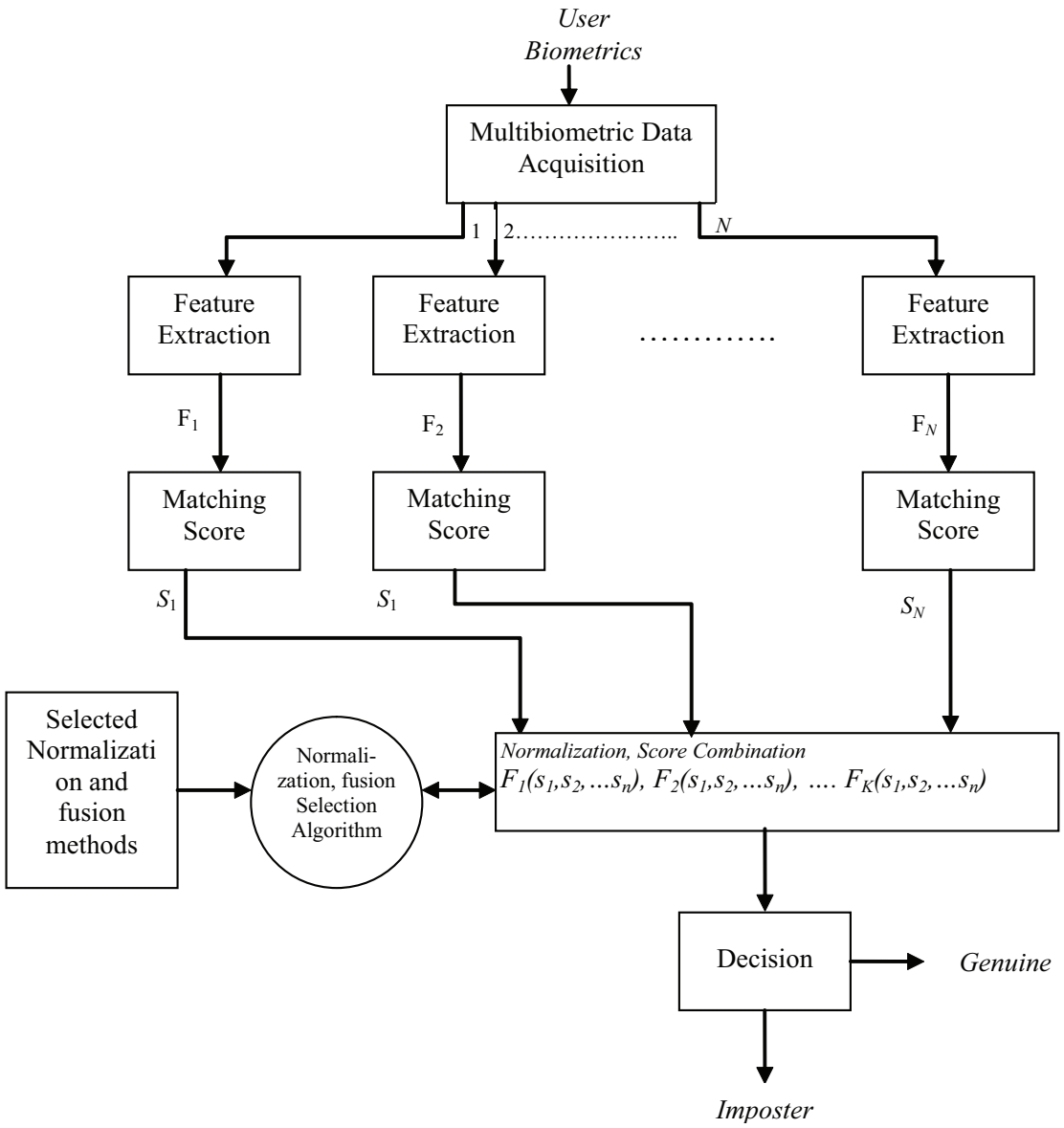

Figure 3.3: Proposed system: Dynamic selection of score-level fusion with normalization

Chapter 4

Experiments and Discussions

To verify the presented concepts, we constructed a multimodal face and speaker identification system and conducted experiments on an offline database in two parts. The first part performs a systematic analysis and identifies the best combinations of fusion methods and normalization techniques in different noise scenarios. The second part of the experiment evaluates the performance of the proposed system of multimodal biometric authentication in different illumination and acoustic noise scenarios and compares the results with existing techniques using the Equal Error Rate (EER).

4.1 Database Selection and Description

For our experiment's purpose, the following two databases were selected:

(a) Pose Illumination Expression (PIE) database: The CMU Pose, Illumination, and Expression (PIE) database [50] contains facial images (640x486 pixels) of 68 people acquired across different poses, under different illuminations/lightings, and with different facial expressions. There are 384 images of each person in different illumina-

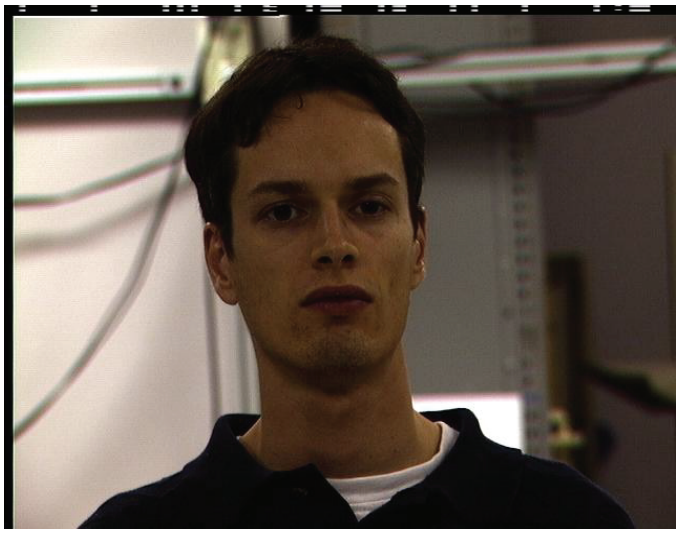

Figure 4.1: A sample image from CMU PIE database

tions/lighting conditions. The main reason for the selection of the CMU PIE database was that it has been extensively used to analyze facial images under different illuminations and in different poses and to benchmark the development of facial recognition algorithms to handle such distortions. Figure 4.1 is a sample image from the CMU PIE database.

(b) The MIT Mobile Device Speaker Verification (MIT MDSVC) Corpus: The MIT Mobile Device Speaker Verification Corpus [59] contains two sessions. Each session consists of the same 48 speakers with 54 speech samples per user. Data was collected in three different locations (a quiet office, a mildly noisy lobby, and a busy street intersection) as well as with two different type of microphones (the built-in internal microphone of the handheld device and an external earpiece headset) leading to six distinct test conditions with nine speech samples in every condition. Table 4.1 shows a sample list of the phrases spoken in each session during the time of data collection.

Office	Lobby	Intersection
alex park	alex park	alex park
peppermint stick	pralines and cream	chunky monkey
ken steele	ken steele	ken steele
peppermint stick	pralines and cream	chunky monkey
thomas cronin	thomas cronin	thomas cronin
peppermint stick	pralines and cream	chunky monkey
sai prasad	sai prasad	sai prasad
peppermint stick	pralines and cream	chunky monkey
trenton young	trenton young	trenton young

Table 4.1: A sample list of phrases spoken in each session

4.1.1 Generation of the Multimodal Database

Two separate multimodal databases were constructed for our experiments by merging the two databases (of 48 users each) described above. The databases were constructed as follows: 15 facial images (of 5 for each illumination/lighting condition: normal, dark and bright) and 15 voice samples recorded by an internal microphone (of 5 for each location: a quiet office, a mildly noisy lobby and a busy street intersection). In our databases, facial images were not confined only to frontal images in [18, 19] or to the same expression. The mutual independence assumption [44] of biometric traits allowed us to randomly pair the users from the two sets. In this way, a multimodal database consisting of 48 virtual users was constructed with each user having 15 (5 for each noise condition) biometric templates for each modality. The biometric data captured from each user is compared with that of all other users in the database under the same noise condition (e.g., mildly lobby) leading to one genuine score vector and 47 impostor scores for each distinct input. Thus, 240 (48 x 5) genuine score vectors and 11280 (48 x 5 x 47) impostor score vectors were obtained from this database under each noise condition.

The first database was used to systematically evaluate multimodal biometric authenti-

cation systems in order to identify the best combination of normalization techniques and fusion methods and also to train our proposed authentication system. The later database was used for evaluating the performance of our proposed system.

4.2 Experimental Setup

All of the experimental results were collected using the same hardware configuration for both the systematic performance evaluation and the proposed system for all experiments: a laptop with an AMD Turion^{TM}X2 processor running at a speed of 2.00GHZ and containing 3 GB of RAM. The developed application is in C#, Java and Matlab R2009b.

4.2.1 MUBI Off-line- Analysis Tool

MUBI is an offline multimodal biometric analyzer tool that can be downloaded from the Center for Identification Technology Research (CITeR) web site run by the department of Computer Science and Electrical Engineering at West Virginia University

`http://Www.citer.wvu.edu/downloads/software.php`

MUBI was developed as an independent multimodal biometrics system analysis tool in a great effort to empower biometric system designers to evaluate different normalization and fusion methods and to choose the the best integration techniques in the context of their application [47].

The inputs of the MUBI tool are the genuine and the impostor scores for each modality. Several modalities (for our experiment: face matching scores and voice matching scores) can be added to make up a multimodal biometrics system so as to evaluate the performance of this hypothetical system. After the modalities have been added to the system, the densi-

ties of genuine and impostor scores for each modality can be plotted, the data partitioning of a chosen method can be created and a number of normalization and fusion methods can then be employed. A ROC curve will eventually be plotted for the system designer to study the performance of the selected combination of techniques using the GAR and the FAR.

4.2.2 Selection of Noise Factors

For the selection of the combination of normalization and fusion method in different noise scenarios, we selected the following illumination and acoustic noise scenarios to mimic the real world scenarios when the authentication will take place on the mobile device:

(a) For the face database:

  (i) Normal lighting condition.

  (ii) Bright lighting condition (Low contrast).

  (iii) Dark (High contrast).

(b) For the voice database:

  (i) Office (Quiet).

  (ii) Lobby (Mild noise).

  (iii) Busy Street Intersection (Loud noise).

Therefore, in total we had the following nine different combinations of illumination and acoustic noise scenarios:

1. Normal lighting condition, Office (Quiet).

2. Normal lighting condition, Lobby (Mild noise).

3. Normal lighting condition, Busy Street Intersection.
4. Bright lighting condition (Low contrast), Office (Quiet).
5. Bright lighting condition (Low contrast), Lobby (Mild noise).
6. Bright lighting condition (Low contrast), Busy Street Intersection.
7. Dark (High contrast), Office (Quiet).
8. Dark (High contrast), Lobby (Mild noise).
9. Dark (High contrast), Busy Street Intersection.

4.2.3 Selection of Controllable Factors

The following normalization and fusion methods were selected as controllable factors [61] for the systematic performance evaluation:

A. Normalization:
    1. Min-Max.
    2. Decimal Scaling.
    3. Z-Score.
    4. MAD.
    5. Tanh.

B. Fusion:
    1. Simple Sum.

2. Simple Product.
3. Simple Maximum.
4. Simple Minimum.
5. BGI.

4.2.4 Evaluation Matrix

There are a total of 21 different controllable factors: 20 combinations created by pairing the first four fusion methods with all of the different normalization techniques and BGI, which uses its own normalization. A hold-out partition scheme [47] was selected for cross validation in the MUBI tool. Table 4.2 was the evaluation matrix we used to identify the best combination of normalization and fusion methods under our selected noise conditions.

4.2.5 Face Detection & Recognition

Identifying people from images of their faces is a widely studied problem. A thorough review of the literature on this topic is available in [62]. In this section, we discuss only the technologies used in our experiments.

Before face recognition techniques are applied, the face must first be detected and located within a given image. The Viola-Jones face detection algorithm (which is based on a boosted cascade of feature classifiers) is a commonly used approach which we have used as our face detection algorithm [55].

We used 40 components based Eigenface coefficients to represent features of the facial image [52] for face identification. The Euclidean distance between the Eigenface coefficients of the face template and that of the inputted face was used as the matching score.

			1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**No**
			1	1	1	2	2	2	3	3	3	a
No	A	B	1	2	3	1	2	3	1	2	3	b
1	1	1										
2	1	2										
3	1	3										
4	1	4										
5	2	1										
6	2	2										
7	2	3										
8	2	4										
9	3	1										
10	3	2										
11	3	3										
12	3	4										
13	4	1										
14	4	2										
15	4	3										
16	4	4										
17	5	1										
18	5	2										
19	5	3										
20	5	4										
21	*	5										

Table 4.2: Evaluation matrix for face and voice multimodal biometric systems

4.2.6 Speaker Recognition

Our speaker recognition was text-independent and was performed using the widely known MASV (Munich Automatic Speaker Verification) environment [1]. MASV uses a Gaussian Mixture Models (GMM) classifier and allows changes to be made in many of the input parameters, including the number of GMM components and the feature set based on Mel Frequency Cepstrum Coefficients (MFCC). The GMM in our verifier used 32 components and the feature vector contained 39 components. The world model was generated from a small subset of the training samples.

4.2.7 Training

The face and speaker recognition systems were trained on the enrollment data for 48 users from the first database and the second database was used for evaluation. During the training phase, a hold out partition (2:3) was used for enrollment and threshold traiting on the development dataset. The threshold for multi-biometric authentication was selected with respect to 1% FAR. For evaluation purpose, face identification scores from the testing dataset were pairwise combined with speaker identification scores under different lighting and acoustic noise scenarios to imitate the practical working environments.

4.3 Results and Analysis

Our experiments demonstrate the benefits of dynamically selecting and applying different combinations of normalization and fusion methods in different noise scenarios. After 181 experiments were carried out using the MUBI tool [47], the results were recorded in the evaluation matrix as seen in Table 4.3.

			1	2	3	4	5	6	7	8	9	No
			1	1	1	2	2	2	3	3	3	a
No	A	B	1	2	3	1	2	3	1	2	3	b
1	1	1	99.842	97.842	96.514	98.571	96.28	96.012	97.224	96.885	94.221	
2	1	2	95.101	95.891	94.783	95.01	94.451	94.236	95.891	94. 257	93.942	
3	1	3	86.15	87.551	85.12	85.447	84.521	83.265	85.447	84.4 49	83.221	
4	1	4	95.121	93.379	93.987	93.678	93.454	93.334	94.234	94.965	93.798	
5	2	1	88.02	87.01	86.453	83.464	87.464	84.454	83.654	86.6 53	83.024	
6	2	2	95.541	94.579	86.454	82.756	84.453	85.234	83.456	86.234	83.024	
7	2	3	90.234	90.01	84.453	83.454	85.464	78.443	84.464	78. 675	78.012	
8	2	4	87.354	87.031	83.454	84.343	83.245	84.264	87.235	86.454	83.157	
9	3	1	99.421	97.225	96.456	96.343	96.343	95.342	96.645	95 .454	94.789	
10	3	2	94.247	94.278	94.342	94.454	94.343	94.353	94.324	93 .245	89.386	
11	3	3	84.915	84.783	85.342	84.645	84.343	83.343	84.342	83 .234	82.541	
12	3	4	99.051	96.865	93.234	94.234	94.243	93.234	95.676	96 .754	89.241	
13	4	1	99.354	97.012	97.531	96.354	96.968	96.024	96.552	97.025	95.01	
14	4	2	98.022	96.893	96.243	96.23	96.01	96.01	96.23	92.78	89.521	
15	4	3	98.334	96.037	96.464	96.575	96.234	95.365	96.454	96 .236	94.637	
16	4	4	97.983	95.769	95.354	95.465	95.745	95.33	94.343	96. 345	95.842	
17	5	1	99.542	96.022	94.157	95.236	96.327	97.152	92.841	95.573	96.421	
18	5	2	99.01	94.541	94.533	96.256	95.243	94.453	92.75	96.5 6	90.02	
19	5	3	99.212	93.247	95.865	95.354	95.345	95.675	96.364	96 .365	90.24	
20	5	4	99.234	93.247	93.543	93.567	93.676	93.867	93.686	93 .454	92.341	
21	*	5	98.101	97.542	97.021	97.051	97.741	95.542	98.112	97.237	95.015	

A1--A5: Min-Max/ Decimal Scaling/ Z-Score/ Median a nd MAD/ Tanh-Estimators
B1--B5: Simple Sum/ Simple Product/ Simple Minimum/ Simple Maximum/ BGI
a1--a3: Normal lighting / Bright lighting (Low contrast)/ Dark(High contrast).
b1--b3: Office (Quiet)/ Lobby (Mild noise)/ Busy Street Intersection.

Table 4.3: Experiment results of performance evaluation (GAR (%) at 1% FAR)

No.	Noise Scenarios	Normalization-Fusion
1	normal lighting, office (quiet)	Min-Max, Simple Sum
2	normal lighting, lobby (mild noise)	Min-Max, Simple Sum
3	normal lighting, busy street intersection	Median and MAD, Simple Sum
4	bright lighting (low contrast), office (quiet)	Min-Max, Simple Sum
5	bright lighting (low contrast), lobby (mild noise)	BGI
6	bright lighting (low contrast), busy street intersection	Tanh, Simple Sum
7	dark (high contrast), office (quiet)	BGI
8	dark (high contrast), lobby (mild noise)	Median and MAD, Simple Sum
9	dark (high contrast), busy street intersection	Tanh, Simple Sum

Table 4.4: Selected normalization techniques and fusion methods from evaluation matrix

In the table, the fusion-normalization combination that produces the best performance under each of the nine noise conditions is highlighted. The combinations are Min-Max, Simple Sum for the first, second, and forth; MAD, Simple Sum for the third and eighth; Tanh, Simple Sum for the sixth and ninth; and BGI for the fifth and seventh noise conditions.Table 4.4 lists these selected combinations from the evaluation matrix.

The second part of the experiment evaluates the performance of the proposed system of multimodal biometric authentication in different illumination and acoustic noise scenarios, and compares the results with existing techniques using EER.

Since our face and speaker recognition systems were trained with a fairly small amount of data, their performances were not as good as the top results achieved in face recognition [56] or in speaker recognition [12], although a comparison with state-of-the-art experiments is difficult because the databases are different. In our experiments the EER for face recognition under normal lighting was 2.23% and the performance under bright/dark conditions is shown in Table 4.5. The EER for clean speech in an office environment was 8.635% and the performance under noisy conditions is shown in Table 4.6.

Authentication Location	Normal Lighting	Bright Lighting	Dark Lighting
GAR	97.04%	94.13%	92.4%
FAR	1.8%	1.87%	2.10%
EER	2.23%	4.12%	4.85%

Table 4.5: GAR/FAR/EER for face recognition under different lighting conditions

Authentication Location	Office	Lobby	Intersection
GAR	85.98%	78.98%	64.57%
FAR	3.25%	4.28%	6.25%
EER	8.635%	12.65%	17.84%

Table 4.6: GAR/FAR/EER for speaker recognition under different noise conditions

Table 4.7 provides the results for combined face and voice recognition in which the combination of fusion methods and normalization techniques has been dynamically selected as the one with the best performance in training. Figures 4.2-4.4 further put the two approaches together graphically for comparison.

No.	Lighting	Recording	GAR (%)	FAR (%)	EER (%)
1	normal lighting	office	98.684	0.924	1.12
2	normal lighting	lobby	98.564	0.944	1.19
3	normal lighting	street	97.57	1.025	1.73
4	bright lighting	office	96.465	0.925	2.23
5	bright lighting	lobby	95.612	0.952	2.67
6	bright lighting	street	94.47	1.13	3.83
7	dark	office	94.274	0.974	3.35
8	dark	lobby	92.833	0.993	4.08
9	dark	street	92.705	1.245	4.27

Table 4.7: Combined Performance of Proposed system.

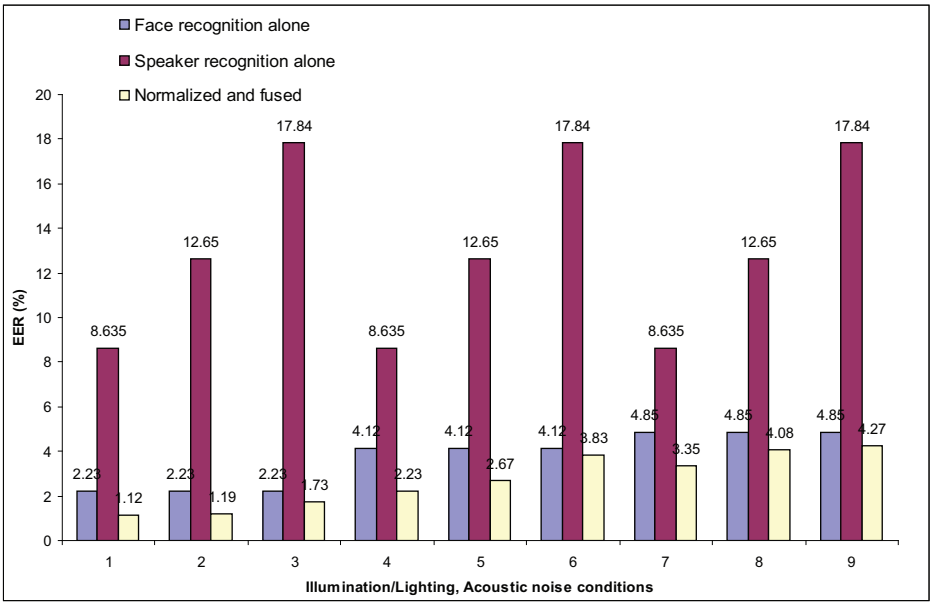

Figure 4.2: Performance in EER (%) of the combined face and speaker recognition classifiers at different lighting and acoustic noises by comparison with face and speaker recognition alone

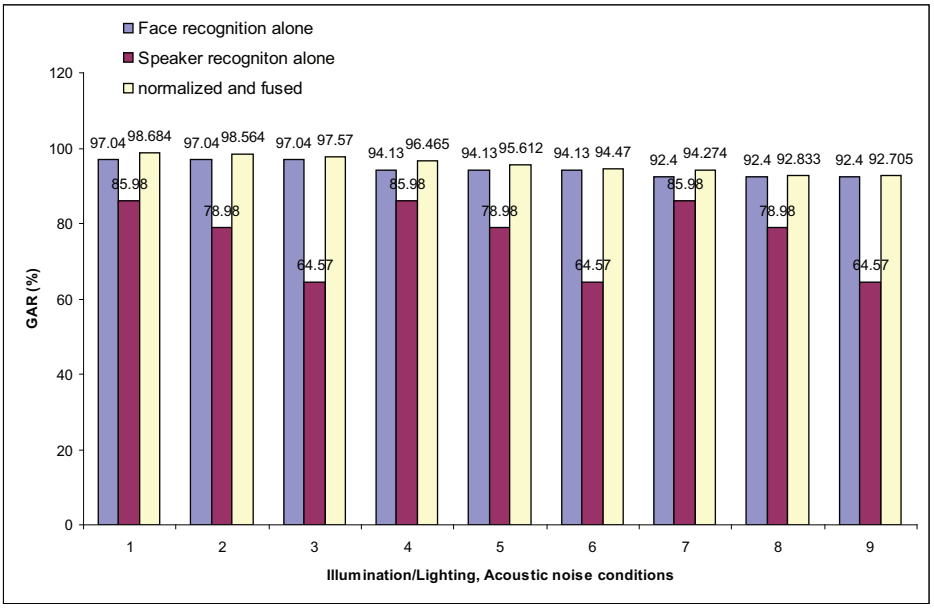

Figure 4.3: Performance in GAR (%) of the combined face and speaker recognition classifiers at different lighting and acoustic noises by comparison with face and speaker recognition alone

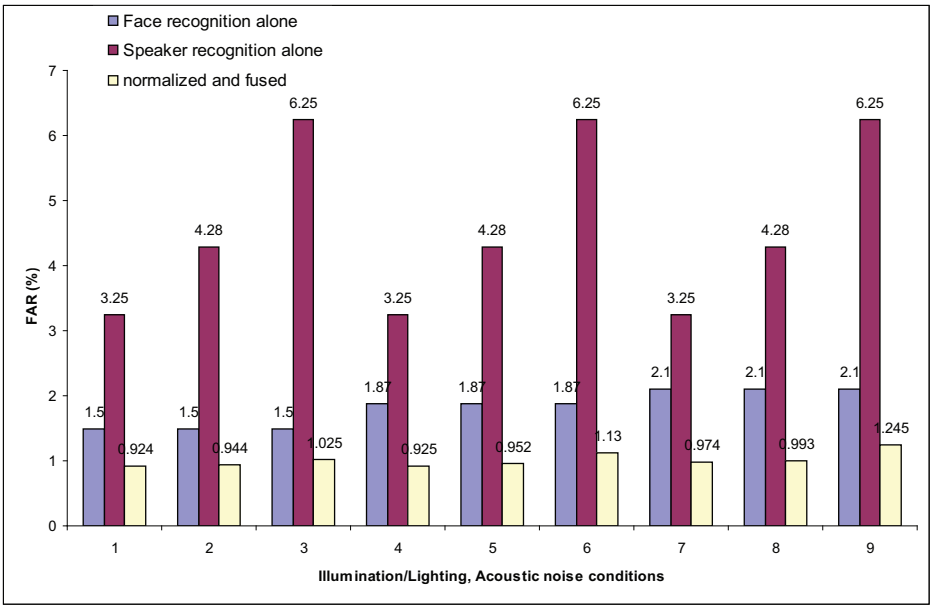

Figure 4.4: Performance in FAR (%) of the combined face and speaker recognition classifiers at different lighting and acoustic noises by comparison with face and speaker recognition alone

No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
A	1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4	4	4	4	4	*
B	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	5
GAR(%)	98.684	94.23	87.15	94.01	87.02	94.841	88.134	87.354	97.21	92.547	83.715	97.051	97.54	96.23	96.34	95.83	97.42	98.01	97.12	97.34	97.10
FAR(%)	0.924	0.987	0.997	1.025	1.035	1.055	1.022	1.10	1.07	1.08	1.08	1.012	1.023	1.011	10.57	1.027	0.987	1.024	1.013	1.026	1.025
EER(%)	1.12	3.379	6.924	3.508	7.008	3.107	6.444	6.873	1.93	4.267	8.683	1.981	4.403	2.391	7.115	2.599	1.784	1.507	1.947	1.843	1.963

Table 4.8: Comparison of dynamically selected combination with all other combinations in noise condition No.1 (normal, office)

No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
A	1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4	4	4	4	4	*
B	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	5
GAR(%)	98.564	94.891	86.351	93.379	86.50	92.879	88.23	86.31	96.125	93.378	85.783	96.265	95.512	94.593	95.137	94.269	94.82	92.541	92.147	92.847	96.142
FAR(%)	0.944	0.985	1.025	1.005	0.954	1.025	1.214	1.045	1.034	1.024	1.025	1.011	1.018	1.12	1.15	1.02	0.97	1.035	1.14	1.009	0.987
EER(%)	1.19	3.047	7.337	3.813	7.227	4.073	6.492	7.368	2.455	3.823	7.621	2.373	2.753	3.264	3.007	3.376	3.075	4.247	4.497	4.081	2.423

Table 4.9: Comparison of dynamically selected combination with all other combinations in noise condition No.2 (normal, lobby)

No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
A	1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4	4	4	4	4	*
B	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	5
GAR(%)	94.814	93.783	84.12	93.687	86.253	85.494	83.137	82.154	94.056	92.992	83.157	92.034	97.57	95.43	96.464	95.354	92.757	93.102	93.165	91.743	95.157
FAR(%)	1.18	1.24	1.257	1.34	.957	1.189	1.201	1.199	1.15	1.211	1.098	1.23	1.025	1.028	1.088	1.15	1.134	1.213	1.34	1.125	1.095
EER(%)	3.183	3.729	8.569	3.827	7.352	7.848	9.032	9.523	3.547	4.11	8.971	4.598	1.73	2.799	2.312	2.898	4.189	4.056	4.09	4.691	2.969

Table 4.10: Comparison of dynamically selected combination with all other combinations in noise condition No.3 (normal, street)

No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
A	1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4	4	4	4	4	*
B	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	5
GAR(%)	96.465	94.51	85.147	93.278	83.164	82.556	83.205	84.443	94.147	93.54	84.45	92.134	94.54	94.231	95.175	93.865	93.36	94.156	93.254	91.567	94.517
FAR(%)	0.925	1.021	1.01	1.023	1.012	0.981	0.987	1.23	1.05	1.15	1.01	1.068	1.067	1.054	1.064	1.001	0.97	0.987	1.09	1.101	1.023
EER(%)	2.23	3.256	7.932	3.873	8.924	9.213	8.891	8.394	3.452	3.805	8.28	4.467	3.264	3.412	2.945	3.568	3.805	3.416	3.918	4.767	3.253

Table 4.11: Comparison of dynamically selected combination with all other combinations in noise condition No.4 (bright, office)

No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
A	1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4	4	4	4	4	*
B	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	5
GAR(%)	94.28	94.151	84.321	93.054	87.02	84.453	85.412	81.245	94.843	94.31	84.143	94.043	94.168	93.18	94.34	93.45	94.327	93.437	94.005	93.676	95.612
FAR(%)	1.012	1.025	1.011	1.056	1.24	1.08	1.025	1.187	1.029	0.987	0.947	1.028	1.145	1.234	1.024	1.213	1.024	1.347	1.014	1.15	0.952
EER(%)	3.366	3.437	8.345	4.001	7.11	8.314	7.807	9.971	3.093	3.339	8.402	3.493	3.489	4.027	3.342	3.882	3.349	3.955	3.505	3.737	2.67

Table 4.12: Comparison of dynamically selected combination with all other combinations in noise condition No.5 (bright, lobby)

No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
A	1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4	4	4	4	4	*
B	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	5
GAR(%)	86.152	84.237	83.265	89.347	84.454	85.234	78.443	84.264	91.342	90.537	83.143	90.346	89.247	88.017	85.657	85.247	94.47	90.122	91.012	90.67	89.421
FAR(%)	1.141	1.18	1.16	1.24	1.247	1.348	1.247	1.189	1.248	1.312	1.257	1.214	1.234	1.147	1.247	1.29	1.13	1.23	1.21	1.24	1.15
EER(%)	7.495	8.472	8.948	5.947	8.397	8.057	11.402	8.463	4.953	5.388	9.057	5.434	5.994	6.565	7.795	8.022	3.83	5.554	5.099	5.285	5.865

Table 4.13: Comparison of dynamically selected combination with all other combinations in noise condition No.6 (bright, street)

No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
A	1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4	4	4	4	4	*
B	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	5
GAR(%)	90.24	89.91	85.147	90.347	83.547	81.56	80.467	86.357	86.95	91.24	84.342	90.67	90.152	90.237	91.454	91.347	93.412	92.35	90.347	90.15	94.274
FAR(%)	1.123	1.234	1.157	1.005	1.245	1.125	1.025	1.045	1.002	0.945	1.023	1.023	1.16	1.085	1.245	1.001	1.025	1.019	1.241	0.917	0.974
EER(%)	5.442	5.662	8.005	5.329	8.849	9.783	10.279	7.344	7.026	4.853	8.341	5.177	5.504	5.424	4.896	4.827	3.807	4.335	5.447	5.384	3.35

Table 4.14: Comparison of dynamically selected combination with all other combinations in noise condition No.7 (dark, office)

No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
A	1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4	4	4	4	4	*
B	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	5
GAR(%)	86.885	90.257	84.149	88.96	85.53	85.234	78.75	85.54	84.54	89.457	83.24	86.54	92.833	84.781	86.36	86.451	85.73	86.156	86.365	90.454	90.378
FAR(%)	1.025	1.351	1.023	1.037	1.037	1.012	1.321	1.212	1.247	1.147	1.255	1.264	0.993	1.24	1.12	1.18	1.023	1.295	1.024	1.045	1.23
EER(%)	7.07	5.547	8.437	6.039	7.754	7.889	11.286	7.836	8.354	5.845	9.008	7.368	4.08	8.23	7.38	7.365	7.647	7.57	7.33	5.296	5.426

Table 4.15: Comparison of dynamically selected combination with all other combinations in noise condition No.8 (dark, lobby)

No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
A	1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4	4	4	4	4	*
B	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	5
GAR(%)	84.21	83.442	83.221	88.798	83.024	83.24	78.012	81.157	85.789	79.386	82.541	84.41	88.01	78.521	84.637	85.842	92.705	84.02	78.24	84.341	89.015
FAR(%)	1.265	1.378	1.547	1.471	1.247	1.227	1.87	1.57	1.345	1.873	1.341	1.387	1.474	1.247	1.285	1.297	1.245	1.678	1.387	1.354	1.287
EER(%)	8.528	8.968	9.163	6.337	9.112	8.994	11.929	10.207	7.778	11.244	9.4	8.489	6.732	11.363	8.324	7.728	54.27	8.829	11.574	8.507	6.136

Table 4.16: Comparison of dynamically selected combination with all other combinations in noise condition No.9 (dark, street)

In examining the results, first of all, experiment results in Table 4.3 provide clear evidence of the need for the dynamic selection of normalization-fusion combinations, as no single combination works the best in all noise conditions. The results in Figure 4.2 demonstrate that the proposed system produces significant performance improvement over individual modality. The best EER achieved in a normal lighting, quiet office environment was 1.12%, but under dark, lobby and dark, noisy street environment the EER rose up to 4.08% and 4.27% respectively.This degradation in normalized and fused performance is due to a higher EER for face and voice modality in noisy environment. This result is consistent with the observation made in Chapter 2 about poor combined performance due to the poor performance of a modality. However, combined performance is always better than individual recognition performance alone. The performance in GAR (%) is shown in Figure 4.3. The combined approach always outperforms the individual modality alone.

Figure 4.4 also shows the significant improvement of FAR compared to face and speaker recognition alone. FAR is an important metric for a multimodal biometric authentication system performance as lower FAR reduces the chance of impostors being identified as legitimate users by the system.

Figures 4.8-4.16 show the comparison of the dynamically selected combinations that were selected during training with the other combinations in every noise condition. From these tables, we can observe that the the dynamically selected combination always out performs other combinations as it is the best in performance among all.

Tables 4.17 and 4.18 show the performance of the proposed biometric technique selection algorithm in terms of the number of correct noise condition identification.

Also, our proposed authentication scheme will use limited processing power in a mobile device. Facial image of and voice sample from the claimant will be acquired in the mobile

Lighting condition	Correct identification
normal	98.47%
bright	95.878%
dark	97%

Table 4.17: Lighting condition selection performance of the algorithm

Acoustic noise condition	Correct identification
office (quiet)	99.5%
lobby (mild noise)	98.6%
busy street intersection	98.2%

Table 4.18: Acoustic noise condition selection performance of the algorithm

device. The size of voice sample is on average 50KB. The face detection will be done in the mobile device, that significantly reduces the size of a face image, on average 3KB for our image database. These facial image and voice recording will be then transmitted over the internet to the authentication server where the preprocessing of data, feature extraction and actual identification will take place. The identification result will then be sent back to the mobile device. Due to the small size of the image and the recorded voice, our authentication system can be implemented for those mobile devices that use not only WIFI and 3G but also a low speed EDGE network.

Though it is difficult to make comparison between different multimodal biometric authentication techniques based on different modalities, algorithms and databases, here we just compared the performance achieved by a gait and speaker based authentication system [54]. In a clean speech environment, this system achieved an EER of 2.19% at best when the accelerometer device was in the breast pocket while our system achieved an EER of 1.12% under normal lighting and quiet office environment. When the authentication took place in a bright lighting, quiet office environment or in a dark, quiet office environment, the EER

of our system rose up to 2.23% and 3.35% respectively. However, a frequently encountered urban noise type is mild city noise and busy street noise (e.g., car noise). In both cases, our system achieved better performance than the multibiometric gait and speaker authentication system where the best EER for those two conditions based on the accelerometer device position were 4.91% and 8.44% respectively.

Chapter 5

Conclusions and Future Works

Multimodal biometrics refers to an automatic recognition of a person based on more than one of his/her behavioral and/or physiological characteristics. Many businesses/government organizations have already applied multimodal biometric methods in practice, mostly for identification or authentication purposes; however, more of work is left to improve the accuracy rates in various environmental noise scenarios. Multimodal biometrics has been adopted in a variety of large scale identification applications, such as border control, criminal investigations and security.

In this thesis, we proposed a new approach of user authentication for use on handheld and mobile devices. In this approach, a systematical method played an important role in performance evaluation of different combinations of fusion methods and normalization techniques in different noise scenarios. The systematic performance evaluation made possible for the development of an algorithm and the construction of an authentication system to dynamically select combinations of biometrics techniques that produces optimal performance in environments where user authentication takes place in practice. The key contributions of the thesis can be summarized as follows:

- We extended the previous research that uses the technique of Design of Experiments to show the importance of investigating and systematically analyze the performance of different score-level fusion methods with normalization techniques in different noise scenarios for the purpose of the designing a robust multimodal biometric authentication system for mobile or handheld devices that will be operated in different noise scenarios.

- We developed an algorithm to dynamically select the most suitable normalization and fusion method combination in different noise scenarios in multimodal biometric authentication for ensuring the legitimacy of the user accessing various services over the internet from a mobile or handheld device, and demonstrated the feasibility and performance of the method by means of experiments and comparison with other methods.

The system presented here is highly customizable in terms of selecting noise scenarios where the biometric traits (e.g. facial image, key stroke, voice) will be collected by the mobile device, combination of normalization techniques and fusion methods and the decision threshold. Hence, the system can be improved significantly by selecting robust recognition algorithms that show better performance in different noise scenarios. As our face and voice database were relatively small, we could not use leave-one-out cross-validation that divides the dataset to n-1 training samples and 1 testing sample, for *N* different times. The *N* results from the folds then can be averaged (or otherwise combined) to produce a single performance estimation. Including this partitioning method for validation and other newly developed biometric techniques, recognition algorithms in performance evaluation, use of other databases for more experiments and development of a fully automatic realtime user authentication system for mobile or handheld devices is left for future research.

Bibliography

[1] Masv, http://www.bas.uni-muenchen.de/bas/sv/. Availabe online.

[2] Combining multiple biometrics, http://www.cl.cam.ac.uk/ jgd1000/combine/combine.html, 2000. Availabe online.

[3] Nist report to the united states congress. summary of nist standards for biometric accuracy, tamper resistance, and interoperability. ftp://sequoyah.nist.gov/pub/nist_internal_reports/nistapp_nov02.pdf, 2002. Available online.

[4] Bbc news. long lashes thwart id scan trial. http://news.bbc.co.uk/2/hi/uk_news/politics/3693375.stm, 2004. Available online.

[5] Privium - fast border passage with iris scan, http://www.cbsa-asfc.gc.ca/travel/canpass/menu-e.html, 2007. Available online.

[6] Canpass: Streamlines customs clearance for frequent travellers, http://www.cbsa-asfc.gc.ca/travel/canpass/menu-e.html, 2007. Available online.

[7] P. Belhumer, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[8] E. Bigun, J. Bigun, and S. Fisher. Expert conciliation for multimodal person authentication systems using baysian statistics. In *Proceedings of the International Conference on Audio and Video-Based Biometric Person Authentication*, volume 1206, pages 291–300. Springer Berlin, 1997.

[9] R. M. Bolle, J. H. Connell, S. Pankanti, N. K. Ratha, and A. W. Senior. *Guide to Biometrics*. Springer-Verlag, New York, 2004.

[10] R. Brunelli and D. Falavigna. Person identification using multiple cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(10):955–966, 1995.

[11] J. Campbell. Speaker recognition: a tutorial. *Proceedings of the IEEE*, 85(9):1437–1462, Sep 1997.

[12] J. P. Campbell, D. A. Reynolds, and R. B. Dunn. Fusing high- and low-level features for speaker recognition. In *Proceeding of the 8th European Conference on Speech Communication and Technology*, pages 2665–2668, Geneva, Switzerland, 2003.

[13] C. H. Chen and C. T. Chu. Fusion of face and iris features for multimodal biometrics. In *ICB*, pages 571–580, 2006.

[14] X. Chen, P. J. Flynn, and K. W. Bowyer. Ir and visible light face recognition. *Comput. Vis. Image Underst.*, 99(3):332–358, 2005.

[15] J. Daugman. How iris recogntion works? *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):21–30, 2004.

[16] M. Faundez-Zanuy. Data fusion in biometrics. *IEEE Aerospace and Electronic Systems Magazine*, 20:34–38, 2005.

[17] 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.

[18] T. Hazen, E. Weinstein, R. Kabir, A. Park, and B. Heisele. Multimodal face and speaker identification on a handheld device. In *Proceedings of the Workshop on Multimodal User Authentication*, pages 113–120, Dec 2003.

[19] T. Hazen, E. Weinstein, and A. Park. Towards robust person recognition on handheld devices using face and speaker identification technologies. In *Proceedings of the 5th international conference on Multimodal interfaces*, pages 289–292, Nov 2003.

[20] L. Hong and A. K. Jain. Intergrating faces and fingerprints for personal identification. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 20:1295–1307, 1998.

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

[22] A. Iannarelli. *Forensic Identification Series*. Paramont Publishing Company, Fremont, California, 1989.

[23] A. Jain and A. Ross. Learning user-specific parameters in a multibiometric system. In *Proceedings of International Conference on Image Processing*, volume 1, pages 57–60, 2002.

[24] A. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):4–20, Jan 2004.

[25] A. K. Jain, R. Bolle, and S. Pankanti. *Biometrics, Personal Identification in Networked Society: Personal Identification in Networked Society*. Kluwer Academic Publishers, Norwell, MA, USA, 1998.

[26] A. K. Jain, K. Nandakumar, and A. Ross. Score normalization in multimodal biometric systems. *Pattern Recognition*, 38(12):2270–2285, 2005.

[27] T. Kanade. *Picture processing by Computer Complex and Recognition of Human-Faces*. PhD thesis, Kyoto University, 1973.

[28] J. Kittler, M. Hatef, R. Duin, and J. Matas. On combining classifiers. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(3):226–239, 1998.

[29] A. Kong, D. Zhang, and G. Lu. A study of identical twins' palmprints for personal verification. *Pattern Recognition*, 39:2149–2156, 2006.

[30] L. Kuncheva. *Combining Pattern Classifiers - Methods and Algorithms*. Wiley, 2004.

[31] L. Lam and S. Suen. Application of majority voting to pattern recognition: an analysis of its behavior and performance. *IEEE Transactions on Man and Cybernetics Systems, Part A: Systems and Humans*, 27(5):553–568, 1997.

[32] B. Lee, M. Hasegawa-johnson, C. Goudeseune, S. Kamdar, S. Borys, M. Liu, and T. Huang. Avicar: Audio-visual speech corpus in a car environment. In *Proceedings of Conference on Spoken Language*, pages 2489–2492, 2004.

[33] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar. *Handbook of Fingerprint Recognition*. Springer-Verlag, 2003.

[34] F. Monrose and A. Rubin. Authentication via keystroke dynamics. In *Proceedings of the 4th ACM conference on Computer and communications security*, pages 48–56. ACM, 1997.

[35] V. Nalwa. Automatic on-line signature verification. *Proceedings of the IEEE*, 85(2): 215–239, 1997.

[36] K. Nandakumar. *Integration of multiple Cues in Biometric Systems*. PhD thesis, Michigan State University, 2005.

[37] C. Park, T. Choi, Y. Kim, S. Kim, J. Namkung, and J. Paik. Multi-modal human verification using face and speech. In *Proceedings of the Fourth IEEE International Conference on Computer Vision Systems*, pages 54–60, 2006.

[38] H.-A. Park, J. W. Hong, J. H. Park, J. Zhan, and D. H. Lee. Combined authentication-based multilevel access control in mobile application for dailylifeservice. *IEEE Transactions on Mobile Computing*, 9(6):824–837, 2010.

[39] P. Penev and J. Atick. Local feature analysis: A general statistical theory for object representation. *Network: Comput. Neural Syst*, 7(3):47–500, 1996.

[40] A. Pocovnicu. Biometric security for cell phones. *Informatica Economica Journal*, 13(1):57–63, 2009.

[41] S. Prabhakar, S. Pankanti, and A. Jain. Biometric recognition: security and privacy concerns. *Security Privacy, IEEE*, 1(2):33–42, Mar-Apr 2003.

[42] F. Prokoski. Disguise detection and identification using infrared imagery. In *Proceedings of SPIE, Optics, and Images in Law Enforcement II*, pages 27–31, 1982.

[43] J. Rokita, A. Krzyzak, and C. Y. Suen. Cell phones personal authentication systems using multimodal biometrics. In *ICIAR '08: Proceedings of the 5th international conference on Image Analysis and Recognition*, pages 1013–1022, Portugal, 2008.

[44] A. Ross and A.K.Jain. Information fusion in biometrics. *Pattern Recognition Letter*, 24(13):2115–2125, 2003.

[45] A. Ross and A. Jain. Multimodal biometrics: An overview. In *Proceedings of 12th European Signal Processing Conference*, pages 1221–1224, 2004.

[46] P. Ross. *Taguchi Techniques for Quality Engineering*. McGraw-Hill, 1995.

[47] N. Samoska. *Evaluation and performance prediction of multimodal biometric systems*. PhD thesis, West Virgina University, 2006.

[48] N. Sedgwich. *The Need for Standardization of Multi-Modal Biometric Combination*. Business/technical presentation, Cambridge Algorithmica Limited, New York, 2004.

[49] H. Shabeer and P. Suganthi. Mobile phones security using biometrics. In *International Conference on Computational Intelligence and Multimedia Applications*, volume 4, pages 270 –274, 2007.

[50] T. Sim, S. Baker, and M.Bsat. The CMU pose, illumination, and expression (PIE) database. In *Proceedings of the 5th International Conference on Automatic Face and Gesture Recognition*, 2002.

[51] 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. National Institute of Standards and Technology, 2003.

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

[53] A. Ulery, A. Hicklin, C. Watson, W. Fellner, and P. Hallinan. Studies of biometric fusion. Technical Report 7346, National Institute of Standards and Technology, 2006.

[54] E. Vildjiounaite, S.-M. Makela, M. Lindholm, R. Riihimaki, V. Kyllonen, J. Mantyjarvi, and H. Ailisto. Unobtrusive multimodal biometrics for ensuring privacy and information security with personal devices. In *Pervasive*, pages 187–201, 2006.

[55] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 511–518, 2001.

[56] V. Štruc and N. Pavešić. Gabor-based kernel partial-least-squares discrimination features for face recognition. *Informatica (Vilnius)*, 20(1):115138, 2009.

[57] J. Wayman, A. Jain, D. Maltoni, and D. Maio. *Biometric Systems: technology, Design and Performance Evaluation*. Springer-Verlag, London, 2005.

[58] L. Wiskott, J. Fellous, N. Kruger, and C. Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:775–779, 1997.

[59] R. Woo, A. Park, and T. Hazen. The mit mobile device speaker verification corpus: Data collection and preliminary experiments. In *Proceedings of Odyssey, The Speaker & Language Recognition Workshop*, pages 1–6, 28-30 2006.

[60] L. Xu, A. Krzyzak, and C. Suen. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Man and Cybernetics Systems*, 22(3):418–435, 1992.

[61] X. Yuan and W. Gan. A statistical approach towards performance analysis of multimodal biometric systems. In *Proceedings of IEEE International Conference on Robotics and biometrics*, pages 877–882, 2008.

[62] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computer Survey*, 35(4):399–458, 2003.

Vita Auctoris

Md. Saifur Rahim was born in 1983 in Mymensingh, Bangladesh. He received his Bachelors degree in Computer Science and Engineering from University of Dhaka, Dhaka, Bangladesh in 2008. His research interests include computer security, human computer interaction and multimodal biometrics.