**NTNU – Trondheim**
Norwegian University of
Science and Technology

# Multibiometric Systems


## Pushpa Dhamala


Master of Telematics - Communication Networks and Networked
Submission date: June 2012
Supervisor: Danilo Gligoroski, ITEM
Co-supervisor: Yanling Chen, ITEM


Norwegian University of Science and Technology
Department of Telematics

# Thesis Description

Reliable and efficient personal recognition is a critical concern in today's widely interconnected society. As a newly emerging technique, biometric recognition systems are being increasingly used by government, business and forensic applications. In this thesis, multibiometric systems are of interest due to their advantages in improving the matching accuracy, increasing population coverage, detering spoofing attacks and imparting fault tolerance to biometric applications.

Multibiometric systems are biometric systems that consolidate multiple sources of biometric evidences. The integration of evidences is known as fusion. In biometrics, various levels of fusion can be categorized into two broad categories: preclassification (fusion before matching) and postclassification (fusion after matching). In this thesis, a survey of different levels of fusion will be conducted. In particular, fusion at match score level will be examined in detail, since it is the dominant level of fusion in biometric systems.

In a multibiometric system, multiple sources of biometric information are used. Various sources that can be fused will be studied in this thesis. Besides, depending on the nature of these sources, multibiometric systems can be classified into different categories, for instance, multi-sensor systems, multi-algorithm systems, multi-instance systems, multi-sample systems, multimodal systems and hybrid systems. In this thesis, an overview of these systems will be provided. Special attention will be devoted to multimodal systems since multimodal systems consolidate the evidence presented by different body traits and the use of multiple body traits improves the identification accuracy significantly.

Assignment Given: February 2012

Professor: Danilo Gligoroski

Supervisor: Yanling Chen

# Abstract

Reliable user authentication has become very important with rapid advancements in networking and mobility coupled with increased concerns about security. Biometric systems perform recognition based on specific physiological or behavioral characteristics(s) possessed by a user. Biometrics establishes identity based on who you are rather than what you possess (e.g, tokens) or what you remember (e.g, passwords). Biometric systems have now been deployed in various commercial, civilian, and forensic applications for reliable individual recognition. Unibiometric systems rely on the evidence of a single source of information whereas multibiometric systems consolidate multiple sources of biometric evidences. Multibiometric systems, if designed properly, are able to enhance the matching performance. Moreover, they are expected to increase population coverage, deter spoofing attacks and provide fault tolerance to biometric systems. In this thesis, we perform a survey of various categories multibiometric systems based on the levels of fusion and sources of evidences being consolidated. Based on the type of information being consolidated, we discuss the fusion at sensor level, feature level, score level, rank level and decision level with examples from literature. Based on the sources of evidences being consolidated, we discuss the multi-sensor, multi-algorithm, multi-instance, multi-sample, multimodal and hybrid systems with examples from literature.

# Acknowledgement

I would like to express my cordial thanks to my professor Danilo Gligoroski for his guidance and valuable advices. I wish to express my sincere gratitude to my supervisor Yanling Chen. Her interests and continuous feedbacks on my work throughout the year proved to be very fruitful in reaching my goals. I am very grateful to her for motivating me throughout the period and always finding sufficient time for me from her schedule.

I would like to thank the Department of Telematics for providing all facilities and good environment. I am also very thankful to the Norwegian Government for providing me the Quota Scheme Scholarship for my master studies.

Finally, I would like to express my heartfelt thanks to my beloved parents and sister for their blessings, wishes and support.

# Table of Contents

# List of Tables

# List of Figures

# List of Acronyms

| | |
|---|---|
| COTS | Commercial Off-the-Shelf |
| CMC | Cumulative match characteristic |
| DWT | Discrete Wavelet Transform |
| EER | Equal Error Rate |
| FTA | Failure to Acquire |
| FTC | Failure to Capture |
| FTE | Failure to enroll |
| FAR | False Acceptance Rate |

| | |
|---|---|
| FMR | False Match Rate |
| FNMR | False Non-match Rate |
| FRR | False Rejection Rate |
| FBI | Federal Bureau of Investigation |
| GA | Genetic Algorithm |
| GAR | Genuine Accept Rate |
| IAFIS | Integrated Automated Fingerprint Identification System |
| JFV | Joint Feature Vector |
| k-NN | k-nearest-neighbor |
| k-NN+VQ | k-NN classifier with vector quantization |
| LDA | Linear Discriminant Analysis |
| MAD | Median Absolute Deviation |
| NIST-BSSR1 | National Institute of Standards and Technology Biometric Score Set-Release 1 |
| PDA | Personal Digital Assistant |
| PIN | Personal Identity Number |
| PCA | Principle Component Analysis |
| ROC | Receiver Operating Characteristics |
| RJFV | Reduced Joint Feature Vector |
| TAD | Threshold Absolute Distance |
| TER | Total Error Rate |

# Chapter 1

# Introduction

## 1.1 Thesis Motivation

Reliable identity establishment/conformance is becoming critical in a variety of applications. Some examples of such applications are sharing networked computer resources, performing remote financial transactions, border security control, and forensic applications. Traditional methods of establishing identity are either knowledge based (e.g., passwords) or possession based (e.g., ID cards). Individuals have certain distinct physiological and behavioral traits that are used by biometric systems for reliable authentication. Biometric systems provide better security and greater convenience than the traditional systems.

Our motivation for working on this project comes from the fact that in near future biometric systems will be supplementing or replacing the traditional systems in many applications. Most of the biometric systems presently being used are unibiometric systems typically making use of a single biometric trait for recognition purpose. There are several limitations of unibiometric systems and some of these can be addressed by designing multibiometric systems that consolidate multiple sources of biometric information. Multibiometric systems can improve the matching accuracy of a biometric system [54]. They also address challenges such as non-universality, noise, susceptibility to spoof attacks and large intra-class variations.

## 1.2 Related Work

As multibiometric systems can be one of the important solutions for various applications in near future, there has been a long list of articles addressing this topic. Ross et al. [54] provide a very good survey of multibiometric systems. The authors focus on the survey of various levels of fusion and go into the details of score level fusion. The ISO/IEC Technical Report [25] contains descriptions and analysis on current practices on various multibiometric fusion. It also discusses requirements and possible routes of standardization to support multibiometric systems.

There are numerous research papers on the various levels of fusion in multibiometric systems. Ratha et al. [49] propose a mosaicking scheme which constructs a composite fingerprint image fingerprint by integrating multiple partial fingerprints as the user rolls finger on the sensor surface. Singh et al. [56] propose a face recognition system combining visible and thermal Infrared (IR) images at sensor level. Kong et al. [18] also discuss a face recognition system performing fusion of visual and thermal infrared images with eyeglass removal at sensor level. Son et al. [60] perform the feature level fusion of face and iris. Ross

et al. [50] perform feature level fusion of hand and face biometrics and perform experiments in three different scenarios. Kittler et al. [34] develop a common theoretical framework for combining classifiers and discuss the various classifier combination strategies. Verlinde et al. [68] compare the performance of score level fusion using three different classifiers based on the k-nearest-neighbor (k-NN) classifier, decision trees and logistic regression. Jain et al. [53] use the classifiers decision trees and linear discriminant function for fusion of match scores. Jain et al. [26] study the performance of different normalization techniques and fusion methods in a fusion scenario involving face, fingerprint and hand geometry modalities. Ho et al. [19] describe the three methods: the highest rank method, the borda count method, and the logistic regression method, to combine the ranks assigned by different matchers. Decision level fusion process is categorized into simple decision level fusion and advanced decision level fusion and discussed in [25].

Many works discuss on the various categories of multibiometric systems based on the sources of information being consolidated. Marcialis et al. [39] discuss a multi-sensor fingerprint system employing optical and capacitive sensors. Jain et al. [31] propose a multi-algorithm system which integrates the evidence obtained from three different minutiae based fingerprint matchers. Han and Bhanu [17] propose a multi-algorithm gait recognition system which probabilistically combines different gait classifiers based on different environmental contexts. Wang et al. [69] discuss a multi-instance iris recognition system where the left and right irises of an individual are combined. Jain et al. [29] describe a multi-sample system which constructs a composite fingerprint template from multiple impressions of the same finger using mosaicking scheme. Bowyer et al. [3] evaluate the performances of face recognition system using both the multi-sensor and multi-sample approaches. Jain et al. [27] investigate a multimodal biometric identification system combining face, fingerprint and voice modalities. Yao et al. [70] propose a multimodal biometric system combining face and palmprint features. Thian et al. [43] propose a hybrid multibiometric system where fusion of multiple samples obtained from multiple modalities is performed at score level.

## 1.3 Thesis Outline

**Chapter 1. Introduction** outlines the motivations for working on this thesis, the related works and the thesis outline.

**Chapter 2. Background** presents general information about biometric system operations and functional processes, biometric system errors, biometric characteristics and their desirable properties, challenges in biometric systems and advantages of multibiometric systems over unibiometric systems, etc.

**Chapter 3. Levels of Fusion in Biometrics** provides an overview on various levels of fusion. It describes sensor level fusion, feature level fusion, score level fusion, rank level fusion and

decision level fusion with corresponding examples from the literature. The score level fusion is dealt in more details.

**Chapter 4 Sources of Evidences** provides an overview of six categories of biometric systems depending on the nature of the sources of information being fused. Multi-sensor, multi-algorithm, multi-instance, multi-sample, multimodal and hybrid systems are described with examples from the literature. More examples are studied for multimodal systems.

**Chapter 5 Conclusion** summarizes the main findings and concludes the thesis.

# Chapter 2

# Background

## 2.1 Biometric Systems

Biometric systems perform recognition of individuals on the basis of their physical and/or behavioral traits. Some commonly used traits are fingerprint, face, iris, retina, palmprint, voice pattern, signature, gait, etc. Most biometric systems will serve one of the two purposes: identification or verification/authentication. Biometric systems provide several advantages over the traditional methods. Unlike passwords and tokens, biometric traits cannot be lost, forgotten or manipulated. Biometric traits cannot be easily copied, shared, distributed or forged. Biometric systems also add to user convenience by alleviating the need to design and remember passwords. Moreover, use of biometrics can provide negative recognition and non-repudiation which is not possible through traditional methods. Negative recognition is a process by which an individual is found to be enrolled in a system despite his unwillingness to be identified. Non-repudiation is a way to guarantee an individual accessing a certain facility cannot later deny having used it. Multibiometric systems consolidate multiple sources of biometric evidences. The integration of evidences is known as fusion. Multibiometric systems combine the information from multiple sensors, samples or traits of an individual, matching algorithms operating on the same biometric.

## 2.2 Biometric System Functional Processes

A biometric system involves the following three main functional processes:

*Enrollment Process*:

In enrollment process, a subject presents his/her biometric characteristics to the sensor along with his/her non-biometric information. Non-biometric information related to subjects could be name, social security number, driver license's number, etc. Biometric features extracted from the captured sample and the non-biometric information are enrolled in the database.

*Verification Process:*

In a verification process, the question being answered is "Is this person who he claims to be?". Subject who desires to be recognized claims his identity which could be a Personal Identity Number (PIN), a username or a smartcard and presents his biometric characteristic(s). The system then compares the extracted template (from the captured

sample) with the enrolled template linked to the claimed identity and determines whether the claim is true or false. Identity verification is used in positive recognition applications where a subject is willing to be recognized.

*Identification Process*:

In an identification process, the question being answered is "Who is this person?". In this process, an individual is recognized by searching the templates of all users in an enrollment database against the captured and extracted biometric features for a match. Identification is a critical component in negative recognition applications where the user tries to avoid being found out who he is [45]. Some examples of negative recognition applications are background checks, forensic criminal identification or preventing terrorists from entering certain areas. Though traditional recognition methods such as passwords, PIN, tokens work for positive recognition; the only viable approach for negative recognition is biometric identification [45].

## 2.3 Biometric System Operations

The overall conceptual structure of a biometric system as given in [24] is shown in Figure 2.1. The biometric system usually consists of five subsystems enumerated below [24].

*Biometric data capture subsystem*

This subsystem comprises of suitable capture devices or sensors. A sensor is required to collect signals from a biometric trait and convert the captured signals into a biometric sample such as a fingerprint image, iris image or voice recording.

*Signal processing subsystem*

This subsystem is responsible for extracting a set of salient discriminatory features from a biometric sample. The extracted feature set represents the underlying trait. The biometric features are suitable for comparing with those extracted from other biometric samples. The biometric feature set extracted in the enrollment process is stored in the data storage subsystem which serves as a biometric reference during recognition process.

*Data storage subsystem*

During enrollment phase, the feature sets extracted are stored in a data storage subsystem. The feature sets are possibly stored along with other non-biometric information related to subject such as name, PIN, social security number, etc. In practice, biometric templates and non-biometric information are often stored in different databases which are logically or physically separated for security and privacy concerns.

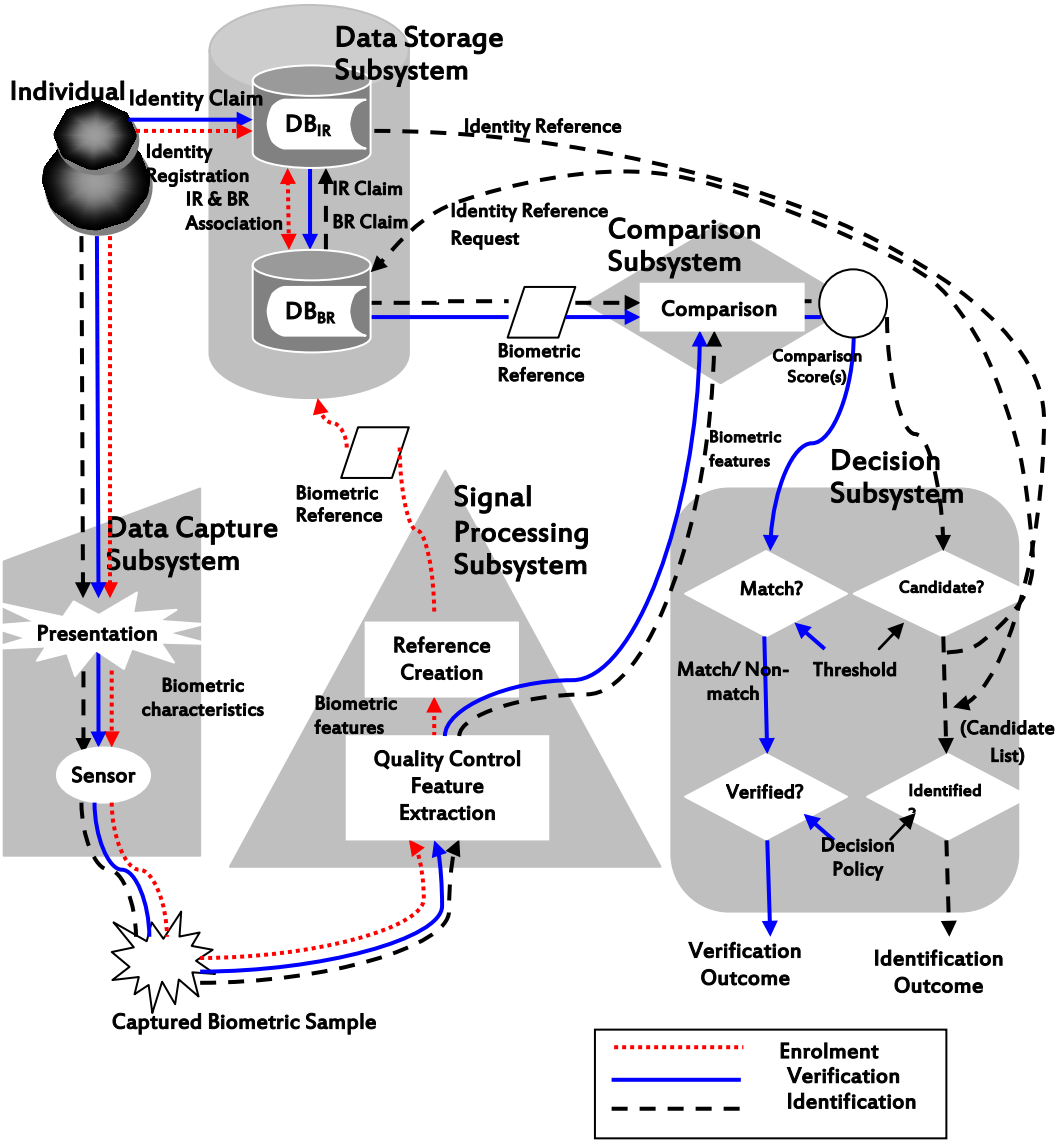



Figure 2.1: Conceptual structure of a biometric system [24].

*Comparison or Matching subsystem*

In a comparison subsystem, the similarity (or difference) between the extracted features (from input sample) and the enrolled biometric templates is determined. In case of verification process, the captured biometric template is compared with the corresponding enrolled biometric template to produce a comparison score. In the identification process, an extracted feature set of a subject is compared against a set of enrolled biometric templates of more than one subject to return a set of comparison scores.

*Decision subsystem*

Based on the comparison score(s) and decision policy, a decision subsystem determines if the captured biometric sample and enrolled template are derived from the same subject. In case of verification process, decision made based on a comparison score is either the acceptance or rejection of the subject. In case of identification, a ranking of enrolled identities that meet the decision policy is presented in order to identify an individual.

## 2.4 Desirable Properties of Biometric Characteristics

Some desirable properties of biometric characteristics for good subject discrimination and reliable recognition performance are described below [54]:

*Universality*: Every individual should possess the characteristic.

*Uniqueness*: The characteristics should be sufficiently distinguishable across individuals comprising the population.

*Permanence*: The biometric characteristics should be sufficiently invariant over a period of time.

*Measurability*: It should be possible to acquire the characteristics without causing undue inconvenience. The acquired raw data should be suitable for further processing.

From an application point of view, following properties should also be taken into account.

*Performance*: The required recognition accuracy in an application should be achievable using the characteristics.

*Acceptability*: Acceptability refers to the willingness by the subject to present his biometric characteristics.

*Spoof Resistance*: This refers to how difficult it is to use artifacts (for example, fake fingers) in case of physiological characteristics and mimicry in case of behavioral characteristics.

## 2.5 Biometric Characteristics

There are various physiological and behavioral biometric characteristics that can be used during recognition. The choice of a biometric characteristic to be used in a specific application is made depending upon the nature and requirements of applications and the properties of the biometric characteristics [32]. Physiological biometric traits include but are not limited to fingerprint, face, iris, hand geometry, hand/finger vein, retina, DNA and palm print. Behavioral characteristics include but are not limited to signature and gait. We introduce some of the commonly used biometric characteristics discussed in [32].

*Face:*

Face recognition is a non-intrusive method and also requires minimum cooperation from the subject. The dimensions, proportions and physical attributes of a person's face are unique. In some application scenario like crowd surveillance, face recognition probably is the only feasible modality to be used. Face recognition can be in a static controlled environment or a dynamic uncontrolled environment. One popular approach to face recognition is based on the location, dimensions and proportions of facial attributes such as eyes, eyebrows, nose, lips, and chin and their spatial relationships. Another approach being widely used is based on the overall analysis of the face image that represents face as a weighted combination of a number of canonical faces.

Face recognition involves two major tasks: i) face location and ii) face recognition. Face location is determining the location of face in the input image. For recognizing the located face, the eigenface approach is one of the very popular methods. The eigenface-based recognition method consists of two stages: i) training stage and ii)operational stage. In the training stage, training set of face images are acquired. The acquired face images are projected into lower dimensional subspace using Principle Component Analaysis (PCA) [63]. A set of images that best describe the distribution of training images in a lower dimensional facespace (the eigenspace) is computed. Then the training facial images are projected into this eigenspace to generate representation of the training images in the eigenspace. In the operational stage, the input face image is projected into the same eigenspace that the training samples were projected into. Then, recognition can be performed by a classifier operating in the eigenspace.

*Fingerprints:*

Fingerprints are unique and consistent over time and hence being used since a long time. A fingerprint is a pattern of ridges and valleys on the surface of a fingertip. Ridges are the upper skin layer segments of the finger and valleys are the lower segments. The various kinds of discontinuities in ridges (minutiae) have sufficient discriminatory information to recognize fingerprints. Ridge bifurcation (where the ridge splits) and ridge ending (where the ridge ends) are the important minutiae points. A minutiae-based fingerprint recognition usually represents fingerprint by these two ridge characteristics called as minutiae.

The uniqueness of a fingerprint can be determined by the pattern of ridges and furrows as well as the minutiae points. Availability of multiple fingerprints of a person makes fingerprint recognition suitable for use in large-scale identification involving millions of identities. However, the problem with the large scale fingerprint recognition system is the requirement of huge amount of computational resources, especially in the identification mode.

*Hand geometry*:

Hand geometry recognition systems are based on the different measurements such as shape of the hand, size of palm, lengths and widths of the fingers. Hand features are not very distinctive. They are suitable for verification but not for identification [52]. In certain situations such as immigration and border control, biometrics such as fingerprints may not be suitable because they infringe on privacy. In such situations hand geometry can be used for verification as hand geometry is not very distinctive. Hand geometry features may not be invariant during the growth period of children. The size of such recognition systems is large and hence it is difficult to embed the systems in other devices such as laptops.

*Palmprint*:

Human palms also contain pattern of ridges and valleys like human fingerprints. Palmprint based recognition is based on the principle lines, wrinkles and ridges on the surface of the palm. Palmprint is distinct for each person. Palmprint scanners are bulkier and more expensive than fingerprint sensors. Features such as principal lines and wrinkles can be captured even with a low resolution scanner. When a high-resolution scanner is used, all features such as geometry, ridges and valley features, principle lines and wrinkles can be combined to achieve higher accuracy. Kumar et al. [36] use both palmprint and hand geometry features for personal verification. Both features are simultaneously acquired from a single hand image.

*Iris*:

Iris is the annular region of the eye regulating the size of the pupil. It is bounded by pupil and sclera (white of the eye) on either side. Iris develops during prenatal period and stabilizes during the first two years of life. The complex iris texture carries very distinctive information useful for personal recognition. Irises of twins are different as well. Iris based recognition systems provide promising speed and accuracy and support large scale identification operations as well. Contact lenses printed with fake iris [11] can be detected. The hippus movement of the eye can also be used for liveness detection.

*Signature*:

Signature is a behavioral biometrics. Electronic signature, for example taken at a POS terminal is compared to the signature on our driving license (or another type of ID) for verification. This is not signature recognition but can be called as ‘simple signature comparison’ [9]. Signature recognition involves a process known as ‘dynamic signature recognition’ where the focus is not only on the ‘look’ of the signature, but on the behavioral patterns inherent to the process of signing. This includes changes in timing, pressure, and speed. It is easy for an imposter to duplicate the visual appearance of signature. However, it is difficult to mimic the behavioral characteristics. Signature recognition is particularly suitable for high-value transactions. Signature recognition is also non-invasive. However, the

system is prone to high error rates when the behavioral characteristics of signatures are not consistent.

*Voice*:

Voice recognition is a combination of both physical characteristics and behavioral biometric characteristics. Voice recognition uses the acoustic features of speech that vary among individuals to discriminate among users. The variations in these acoustics properties arise because of the anatomical differences naturally occurring in individuals and the differences in learned speaking habits. The physical characteristics remain constant whereas the behavioral characteristics of voice could change over time because of age, medical conditions, emotional state, etc. Voice is not distinctive enough to be used for large scale identification. A voice recognition system could be either text-dependent or text-independent. In a text based system any subject needs to utter a specific phrase whereas a text-independent system recognizes subject independent of what he speaks. Text-independent systems are more difficult to design and also more robust against frauds.

*Gait*:

Image-based recognition methods employing fingerprint, face or iris modalities require co-operation from subject, physical contact or close proximity with capture devices. Gait recognition is based on recognizing individuals on the basis of the way they walk. This technique can be appropriate in many practical cases where the environmental condition is changing; subject is not cooperating and is at a distance from the capture device. Gait recognition has several challenges. Gait can be affected by clothing, injuries or other environmental context. There can be large variation in gait characteristics of an individual (both intentionally and unintentionally) making it less unique compared to iris or fingerprint. However, it is still useful in many visual surveillance applications.

## 2.6 Biometric System Errors

Two samples of a single user's biometric trait are rarely read exactly the same. This occurs due to various reasons such as imperfect sensing condition, alterations in user's biometric characteristics, changes in ambient conditions and user's interaction with the sensor. Therefore, the output of a biometric matching system is a similarity score(s) that quantifies similarities between the enrolled and input templates. The system decision depends on the set threshold $t$. Pairs of biometric samples generating a similarity score $s$ greater than $t$ are inferred as mate pairs belonging to the same person. Pairs of samples with similarity score less than $t$ as inferred as non-mate pairs belonging to different persons. The distribution of match scores generated from pairs of samples from different persons is called an imposter distribution and the distribution of match scores generated from pairs of samples of the same person is called a genuine distribution (see Figure 2.2).

False Acceptance Rate (FAR) (or, the False Match Rate (FMR)) of a biometric system is the rate at which the non-authorized persons are falsely recognized during matching process. False Rejection Rate (FRR) (or, the False Non-match Rate (FNMR)) of a biometric system is the rate at which authorized people are falsely not recognized during matching process. Total Error Rate (TER) is obtained by combining these two errors. TER = (Number of False Accepts + Number of False Rejects) / (Total Number of Accesses).

Regulating the threshold $t$ changes both FAR and FRR. If the threshold $t$ is increased in order to attain higher system security, FRR increases. If the threshold is decreased in order make the system more tolerant to input variations and noise, and reduce annoyance, FAR increases. Therefore, a biometric system needs to make a tradeoff between FAR and FRR. The system performance at all operating points (thresholds $t$) can be depicted by Receiver Operating Characteristics (ROC) Curve. ROC curve represents the FAR as a function of FRR (see Figure 2.3). In many cases ROC curve plots the (1-FRR) (instead of FRR) against the FAR. The Equal Error Rate (EER), which is the FAR and FRR when they are equal, is often used as a performance measure. However, EER is not a robust measure for system performance [4]. This is because most practical biometric systems do not have threshold adjusted for FAR=FRR. ROC curves of various systems could be very different and thus two systems with the same EER could differ by decades for other ROC points.

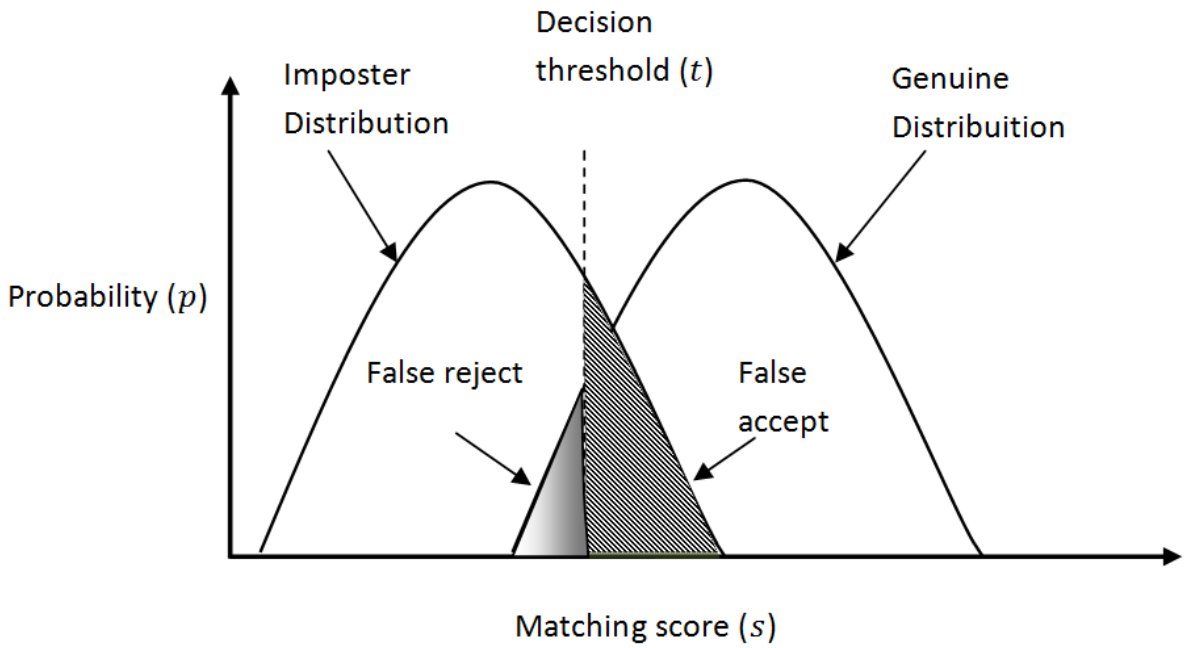



Figure 2.2: Biometric system error rates: The curves show FAR and FRR for a given threshold $t$ over the genuine and impostor score distributions. FAR is the percentage of the non-mate pairs whose matching scores are greater than or equal to $t$, and FRR is the percentage of the mate pairs whose matching scores are less than $t$ [45].

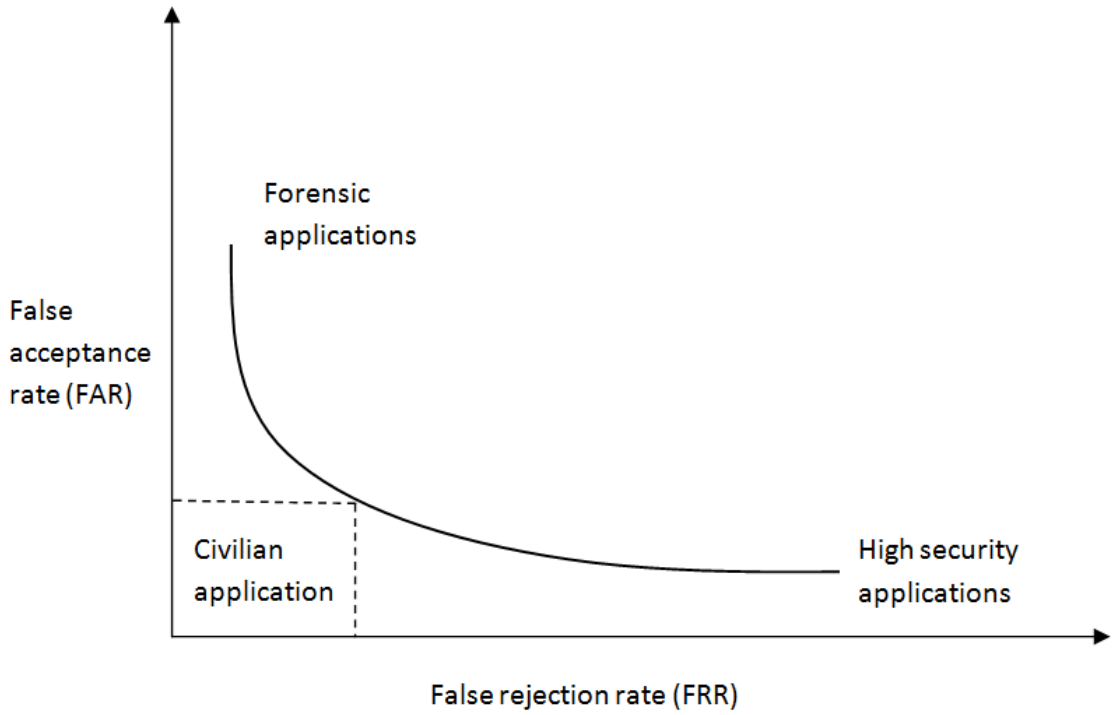



Figure 2.3: Receiver operating characteristic curve: Different biometric application types make different trade-offs between the FAR and FRR. [45].

Besides the two error rates FAR and FRR discussed above, some other error rates are also used to characterize biometric system's accuracy. The Failure to Acquire (FTA) (also known as Failure to Capture (FTC)) rate shows the rate at which biometric device fails to automatically capture a sample when presented with a biometric characteristic. This usually occurs because of low quality of inputs (for example, extremely faint fingerprint) and also sensor wear and tear. Failure to enroll (FTE) rate is the proportion of users who cannot be successfully enrolled in a biometric system. FTE rate usually occurs when the system rejects poor quality templates during enrollment.

We now discuss Cumulative Match Characteristic (CMC) curves. These are the most common graphs for evaluating closed-set identification performances of a system. In closed-set identification, every probe sample has a corresponding match in the database. However, such systems usually exist in laboratories and there are very few real-world applications operating under closed-set identification task. CMC curve is relevant to a recognition scenario, in which a probe sample is matched against each of a set of gallery samples. The gallery sample exhibiting best match (for example best similarity score) with the probe sample represents the identity of the probe. If the trial is repeated for all probe samples, it is possible to know how often the top match selected by the system represents a correct identity. The rank one identification rate is the percentage of the probes for which the closest match (top similarity score) in the gallery represents the correct identity. The percentage of probes for which either the closest or second-closest match (top or second ranked score) in the gallery represents the correct identity is the rank two identification rate.

Thus rank forty identification rate gives the probability that the correct identity lies somewhere in the top forty similarity scores. A CMC curve shows the probability of identification at numerous ranks. CMC curves are largely dependent on gallery size. The probability of correct identification at a certain rank is higher for smaller databases. For instance, for identification at rank 10, the probability of correct identification is much higher when the database size is 100 than when the database size is 10,000. Therefore, it is important to state the size of database for any CMC curve.

## 2.7 Social Acceptance and Privacy Issues

The ease and comfort in interaction with the biometric systems largely contribute to acceptability. Acceptability may also be influenced by religious, cultural and ethnic factors. For example, use of contactless biometric features such as face, iris, or voice may be considered as more user-friendly and hygienic [33]. Likewise, systems requiring lesser co-operation from user may be considered as more convenient to users. However, biometric characteristics which can be captured without user participation might be captured without the knowledge of user which may be perceived as threat to individual privacy.

"Privacy is the ability to lead life free of intrusions, to remain autonomous and to control access to one's personal information"[45]. The use of biometric data raises several privacy concerns which need to be addressed. We mention some threats to privacy discussed in [24]. Biometric data could be misused for applications other than originally intended. Biometric data might be misused to retrieve or analyze some other information that is not required for recognition purpose. For example, the subject's health status or ethnic background could be determined from biometric traits. Biometric data could also be used in linking information of a subject across different databases or systems. Public need to be ensured that their biometric data is used only for the intended purpose and the biometric information remains private by the companies and agencies operating biometric systems. Appropriate legislation is necessary to ensure that the biometric information is not abused and the misuse is punished. Biometric applications with highly decentralized recognition capabilities are the most acceptable [45]. This can be done by storing the biometric information in decentralized encrypted databases over which a subject has his control. For example, a system can issue user a smart card with his fingerprint template stored on it in an encrypted format.

## 2.8 Challenges in Biometric Systems

Most biometric systems presently in use employ a single biometric trait for recognizing individuals. Even though these unibiometric systems have offered a reliable solution for identification and verification applications, it is important to consider the vulnerabilities and

limitations of these systems. Some of the challenges encountered by the unibiometric systems are described below [51]:

*Noise in sensed data*:

The biometric data is contaminated by noise mainly due to slight variations in the biometric trait itself or imperfect acquisition conditions. For example, a fingerprint image with a scar or a voice sample altered by cold is noisy data. Inappropriate ambient conditions like poor illumination of user's face in face recognition or imperfectly maintained sensors like a fingerprint sensor with dirt on its surface lead to noisy data. Noisy data can result in rejection of a genuine user.

*Non-universality*:

Biometric system may not be able to acquire meaningful biometric data from a subset of individuals. This results in a failure-to-enroll (FTE) error. For example, an iris recognition system may not be able to obtain the iris information of users with long eyelashes, drooping eyelids or certain pathological conditions of eyes.

*Spoof attacks*:

Spoofing attack is more relevant in cases where behavioral traits such as signature and voice are used. In such cases, an imposter tries to mimic the traits corresponding to the enrolled user. However, physical traits such as fingerprints and iris are also vulnerable to spoof attacks by creating biometric artifacts. Matsumoto et al. [41] report that the gummy fingers, made using cheap and easily obtainable tools and materials, were accepted with high rates by the 11 different fingerprint systems they used. Those fingerprint systems employed optical or capacitive sensors. Gelatin, a readily available and cheap soft plastic material was used to make gummies. Not only gummy fingers made using impression taken from live fingers but also the gummy fingers made from residual fingerprints were readily accepted by their systems. Targeted spoof attacks can seriously undermine the security of biometric systems. Different ways have been suggested to protect the system from spoofing attacks. For example, in the case of fingerprint and iris, liveness detection can be used. There are two complementary approaches to liveness detection [65]. One is detection of the known artifacts (e.g. silicon and gelatin fingerprints, photograph of a face etc.). The other approach is to look for evidences of liveness in the presented biometric (for example temperature, pulse, humidity etc). In the case of behavioral traits such as voice, a challenge response mechanism could be used (for example system prompts "Please say 5-3-4-8").

*Intra-class variations*:

Changes in biometric characteristics of a person with the passage of time (for example, change in hand geometry) or user interactions with the sensor in a wrong manner (for example, incorrect facial pose) are the main factors resulting in intra-class variations

between the enrolled and input template of an individual. Some ways to address the intra-class variations could be storing multiple templates for every user during enrollment and updating these templates at certain intervals of time [65]. Intra-class variations are more serious concerns in biometric systems using behavioral traits since the variations in psychological makeup of an individual might result in very different behavioral traits at different times. For example, the voice of a person can vary depending on stress levels, health conditions. Similarly, gait can be affected by clothing, injuries, inebriation and other environmental context.

*Inter-class similarities*:

Inter-class similarity refers to overlapping of feature spaces corresponding to multiple classes or individuals. Inter-class similarity is prominent in an identification system comprising a large population of enrolled users resulting in an increased false match rate. Therefore, there is an upper bound on the number of individuals that can be discriminated effectively which determines the capacity of an identification system.

## 2.9 Advantages of Multibiometric Systems over Unibiometric Systems

We discuss some of the advantages multibiometric systems offer over unibiometric systems in the following paragraphs [51].

Multibiometric systems address the issue of non-universality i.e., limited population coverage. For example, if a person's poor quality of fingerprints prevents him from enrolling in the system; then the use of other biometric traits such as iris, face, voice etc. will help the system acquire meaningful biometric data and enroll the user.

It is extremely difficult to spoof multiple biometric traits of a legitimate user. If each subsystem determines the probability of the particular trait being a spoof, it is possible to find out the probability of the user being an imposter by using an appropriate fusion technology. Moreover, a challenge response mechanism can be included that asks user to present the random subset of traits (in a particular order) at the point of acquisition. This would ensure that the system is interacting with a live user.

Multibiometric systems effectively address the problem arising because of noisy data. When the information acquired from one biometric trait is corrupted by noise, it is possible to use information acquired from the other biometric trait. Some systems also take into considerations the quality of acquired input biometric signals during the fusion process. Estimating the quality of acquired biometric data is in itself a challenging problem. However, if done appropriately, multibiometric systems gain significant benefits.

A multibiometric system acts as a fault tolerant system by continuing to operate even when information from certain biometric sources becomes unreliable because of sensor or software malfunctions or intentional user manipulation. Fault tolerance is usually desirable in authentication systems involving large number of subjects (for example, in border control applications).

Consolidation of evidences from multiple sources can offer substantial improvement in the accuracy of biometric systems. Use of proper sources of information and the right fusion methodology determines the improvement in matching accuracy. The availability of multiple sources also increases the feature space thereby increasing the number of individuals that can be discriminated reliably. Therefore, the capacity (i.e., the number of users that can be enrolled) of an identification system can be increased.

## 2.10 Levels of Fusion in Multibiometric Systems

Fusion in multibiometric systems can be performed utilizing information available in any of the modules (data capture module to decision module). Fusion can take place at these levels: i) sensor level ii) feature level iii) score level iv) rank level and v) decision level. In sensor level fusion raw data captured by the sensor(s) are combined. In feature level fusion features originating from each individual biometric process are combined to form a single feature set or vector. In score level fusion, match scores provided by different matchers indicating degree of similarity (differences) between the input and enrolled templates, are consolidated to reach the final decision. In rank level fusion each biometric sub-system assigns a rank to each enrolled identity and the ranks from the subsystems are combined to obtain a new rank for each identity. In decision level fusion the final Boolean result from every biometric subsystem are combined to obtain final recognition decision. We provide a more detailed description of fusion at various levels in Chapter 3.

# 2.11 Sources of Evidences in Multibiometric Systems

Various sources of biometric information can be used in a multibiometric system. Based on these sources, multibiometric systems can be classified into six different categories [54]: multi-sensor, multi-algorithm, multi-instance, multi-sample, multimodal and hybrid. Multi-sensor systems employ multiple sensors to capture a single biometric trait in order to extract diverse information. In multi-algorithm systems, multiple algorithms are applied to the same biometric data. Multi-instance systems use multiple instances of the same body trait (for example, left and right irises or left and right index fingers). In multi-sample system, multiple samples of the same biometric trait are acquired using the same sensor in order to obtain a more complete representation of the underlying trait. Multimodal systems combine evidences obtained from different (two or more) biometric traits. In [54] hybrid is used to refer to those systems integrating two or more of the scenarios mentioned earlier. We

conduct a detailed survey of multibiometric systems based on the sources of information in Chapter 4.

## 2.12 Application of Biometric Systems

Biometric applications can be categorized into three main groups [45]:

1) Commercial applications such as computer network login, e-commerce, Internet access, ATMs or credit cards, physical access control, mobile phones, Personal Digital Assistant (PDA)s, medical records management, distance learning, etc.
2) Government applications such as national ID card, driver's license, social security, border control, passport control, welfare-disbursement, etc.
3) Forensic applications such as corpse identification, criminal investigation, terrorist identification, parenthood determination, etc.

# Chapter 3

# Levels of Fusion in Biometrics

It is important to determine the type of information that should be consolidated during fusion process. The amount of information available decreases after each level of processing in different modules of a biometric system. The raw data represents the richest source of information whereas the final decision just contains an abstract level of information. The various levels of fusion are categorized as (i) preclassification or fusion before matching and (ii) postclassification or fusion after matching [55]. This categorization is based on the fact that the amount of information available for fusion is drastically reduced once the matcher is invoked. Fusion before matching can take place either at the sensor level or at the feature level. Fusion at score level, rank level and decision level occur after matching module is invoked (postclassification). In this chapter we discuss the various levels of fusion in multibiometric systems.

## 3.1 Sensor Level Fusion

The raw biometric data represents the richest source of information. However, it is highly probable that raw data is contaminated by noise (for example, non-uniform illumination, background clutter, etc.). Sensor level fusion refers to the consolidation of raw data obtained using multiple compatible sensors or multiple snapshots of a biometric using a single sensor [51].

***Example 3.1 Mosaicking multiple fingerprint impressions to construct rolled fingerprint***

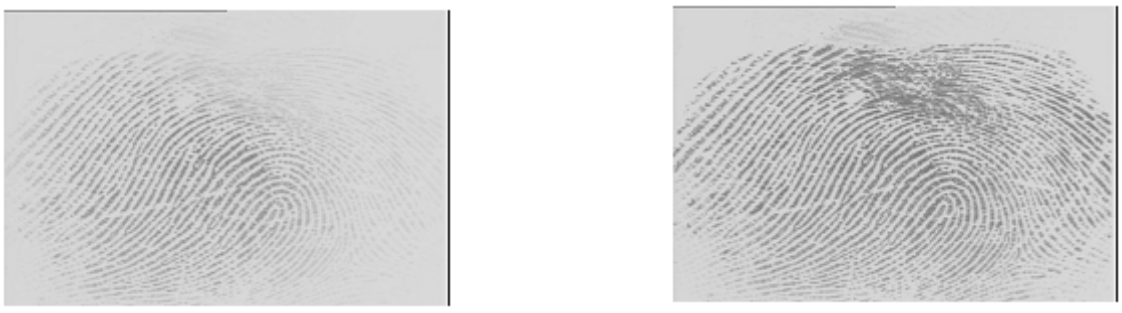
Image mosaicking refers to aligning of two or more images into a new aggregate image without distortion in the overlapping areas. Mosaicking multiple fingerprint impressions to construct a composite image is an example of sensor level fusion. Ratha et al. [49] describe a mosaicking scheme which constructs a rolled fingerprint by integrating multiple partial fingerprints as the user rolls finger on the sensor surface. A rolled fingerprint is preferable over plain touch impression known as dab during enrollment of a person in database. A sample rolled fingerprint and dab are shown in Figure 3.1. This rolled fingerprint covers larger area of the finger, thereby including larger number of feature points. Therefore, the overlap is higher when the partial fingerprint impression (query impression) is matched to rolled fingerprint template than when it is matched to another partial fingerprint.

(a) (b)

Figure 3.1: Images (a) rolled fingerprint (b) dab [49].

The first step to fingerprint mosaicking algorithm is to segment each frame into foreground, the fingerprint area and background, the non-fingerprint area. The second step is to construct a rolled fingerprint mosaic from the set of frames of partial fingerprint impression. For this purpose, the frames stacked are visualized as image planes. If it is assumed that there was no slipping when user rolled his finger on sensor, the resultant fingerprint should be the aggregate of the individual image components. To determine the aggregate, a pixel in all the frames is considered and the resultant pixel is computed as a mathematical function of the pixels. Authors describe five schemes for constructing rolled fingerprint image. The results with different composing schemes are shown in Figure 3.2.

The simplest approach is naïve averaging over the whole image. The second approach ignores the foreground masks and only takes the minimum of the intensity value at each pixel. The third approach does averaging only in the region where fingerprint is detected. The fourth approach is similar but it uses a mask that tapers from zero at the edges of foreground to one at the central region. The last approach shrinks the foreground mask and only uses the central portion of each fingerprint image. The final step is to compute the confidence level at each pixel in order to evaluate the reconstructed image.



a)Naive b) Minimum

c) Foreground

d) Smoothed

e) Center

Figure 3.2: Result with different composing schemes [49].

***Example 3.2 Fusion of infrared (IR) and visible face Images for face recognition***

Fusion of visible and thermal face images at sensor level is discussed in several literatures. Singh et al. [56] describe a face recognition system by fusion of visible and thermal infrared images at sensor level. Face recognition is not sufficiently accurate in uncontrolled environments even when efficient approach to face recognition like the eigenface approach is implemented. Using IR images can be a good alternative to using visible images for face recognition applications under changing illuminations as the IR images are relatively insensitive to illumination changes. However, IR image has other limitations. It is opaque to glass and is sensitive to surrounding temperature changes and variations in the heat patterns of the face. On the other hand, visible image is more robust to the mentioned factors but very sensitive to illumination changes. In [56], authors concentrate on the sensitivity of IR images to eyeglasses. Eye glasses act as temperature screen and hide the parts located behind them degrading the recognition performance significantly. The experiments in [56] show that face recognition performance in IR spectrum is significantly degraded when eyeglasses are present in the probe image but not in the gallery image and vice versa. In order to address the serious problem arising from the sensitivity of IR image to facial occlusion due to eyeglasses, [56] proposes fusion of information from both IR and visible spectra. Genetic Algorithm (GA) (see [56] for more on GA) is employed for feature selection and fusion where group of wavelet features (see [56] for review on wavelets) from

visible and thermal face images are selected and fused to form a fused image. Experiments are performed using the Equinox face dataset [21]. The eigenface approach to face recognition is used. The experimental results show substantial improvements in recognition performance suggesting the potentials of fusing IR with visible images.

***Example 3.3 Fusion of visible and infrared images with eyeglass removal for face recognition***

Another example of fusion of visual and thermal infrared images at sensor level is by Kong et al. [18]. By integrating visual and thermal face images, a new face image is obtained that is invariant to illumination conditions and also robust under low lighting conditions. In the fusion process, eyeglasses which block thermal energy are detected from thermal images with an ellipse fitting method (see [18] for more). The detected eyeglass regions are replaced with template eye pattern in order to retain information for face recognition. A commercial face recognition software FaceIt® is used as an individual recognition module. From the experiments performed under conditions of varying illumination and facial expressions, it is observed that sensor-fusion based face recognition outperforms individual visual and infrared face recognitions.

## 3.2 Feature Level Fusion

In feature level fusion, feature sets originating from multiple information sources are integrated into a new feature set. For homogeneous feature sets (for example, multiple measurements of a person's hand geometry), fusion can be achieved by calculating the weighted average of the individual feature vectors [54]. For non-homogeneous feature sets (for example, features of different modalities like face and hand geometry), a single feature set can be obtained by concatenation. However, for incompatible feature sets (for example, fingerprint minutiae and eigenface coefficients) concatenation is not possible. Dimensionality reduction scheme like feature selection/transformation is applied to obtain a minimal feature set. The key benefit of this fusion scheme is that it enables detection/removal of correlated feature values improving recognition accuracy. Fusion at match score level and decision level are extensively studied in literatures. Fusion at feature level is relatively less studied.

Feature level fusion is challenging for the following reasons [58]:

1) The feature vectors of multiple modalities might be incompatible. For example, the minutiae set of fingerprints and eigen-coefficients of face.
2) The relationship between the feature spaces of different biometric systems may not be known.
3) Concatenation of two feature vectors might result in a feature vector with very large dimensionality leading to the curse-of-dimensionality problem. In such cases, when

sufficiently large numbers of training samples are not available, increasing number of features might degrade system performance.

4) Most commercial biometric system vendors do not provide access to the feature sets.
5) More complex matchers might be required to operate on concatenated feature vectors.

***Example 3.4 Feature level fusion of face and iris***

Son et al. [60] perform feature level fusion of face and iris (see Figure 3.3). They apply multilevel two-dimensional Discrete Wavelet Transform (DWT) to extract feature vectors from the iris and face images. For fusion, concatenation is done between the iris and face feature vectors to form a Joint Feature Vector (JFV). The feature dimensionality is further reduced by applying Direct Linear Discriminant Analysis (DLDA) in order to extract Reduced Joint Feature Vector (RJFV). RJFV has a lower dimensionality and a higher discriminating power than the JFV. Their experiments show that the multimodal authentication system using RJFV exhibits considerably better performance than unimodal system.

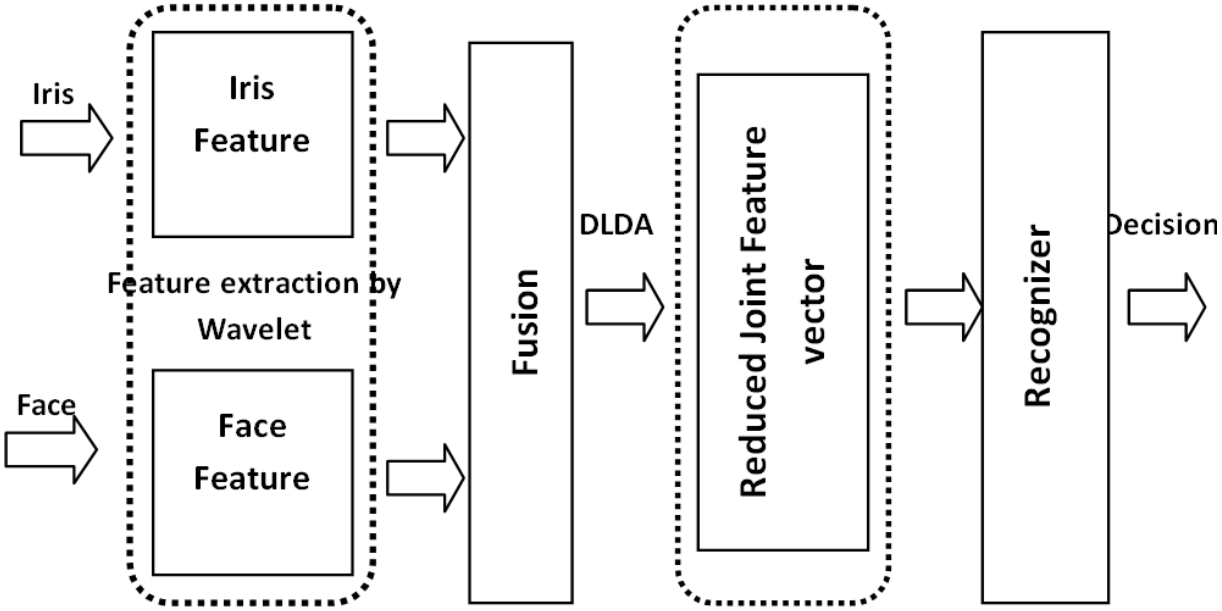



Figure 3.3: Bimodal biometric system using iris and face [59].

***Example 3.5 Feature level fusion of hand and face biometrics***

In this section we summarize the feature level fusion suggested by Ross et al. [50].

Let $X = \{x_1, x_2, \dots, x_m\}$ and $Y = \{y_1, y_2, \dots, y_n\}$ denote the feature vectors ($X \in R^m$ and $Y \in R^n$) representing information extracted from two different sources. In order to yield the new feature vector $Z$, vectors $X$ and $Y$ are augmented and then feature selection is performed on the resultant vector in order to reduce its dimensionality. The different stages adopted in [50] are:

*Feature Normalization:* The individual feature values of the vectors $X$ and $Y$ may be significantly different in terms of their range and distribution. For example, the values of $x_i$'s may be in the range [0,100] while $y_i$'s values may be in the range [0,1]. Therefore, feature normalization is performed to modify the mean and variance of the feature values in order to ensure the contribution of each feature vector is comparable [30]. Ross et al. test two normalization techniques: the simple min-max and median normalization (see [23] for details on these techniques). In their experiments they use the median normalization scheme because of its robustness to presence of outliers in the training data. An outlier is an observation that is numerically distant from the rest of the data. After normalization the modified feature vectors are represented as $X' = \{x'_1, x'_2, \dots, x'_m\}$ and $Y' = \{y'_1, y'_2, \dots, y'_n\}$.

*Feature Selection*: When two feature vectors $X'$ and $Y'$ are augumented, a new feature vector $Z' = \{x'_1, x'_2, \dots, x'_m, y'_1, y'_2, \dots, y'_n\}$ ($Z' \in R^{m+n}$) is obtained. The curse of dimensionality dictates that the augmented vector might not result in an improved performance [62]. Feature selection process is a dimensionality reduction scheme. Some feature values maybe noisy compared to others. In the feature selection process, a minimal feature set of size $k < (m + n)$ is chosen such that classification performance on a training set of feature vectors is improved. The feature selection algorithm employed here is sequential forward floating selection technique (see [47] for more on this technique). A new feature vector $Z = \{z_1, z_2, \dots, z_k\}$ is obtained when the feature selection algorithm is applied.

*Match Score Generation:* Let $(X_i, Y_i)$ and $(X_j, Y_j)$ be the feature vectors obtained at the two different time instances $i$ and $j$ where $X$ and $Y$ represent the feature vectors derived from two different information sources. Let $(Z_i, Z_j)$ denote the corresponding fused feature vectors.

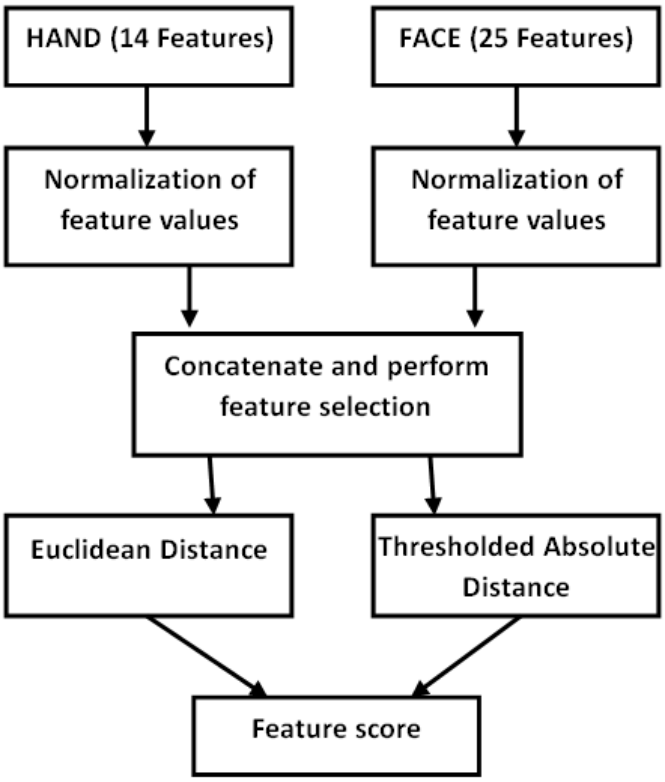



Figure 3.4: Procedure adopted in [50] to perform feature level fusion.

Let $(s_x, s_y)$ denote the normalized match score generated by comparing $X_i$ with $X_j$ and $Y_i$ with $Y_j$ respectively. Let $s_{match} = (s_x + s_y)/2$ represent the fused match score obtained using simple sum rule.

To compare the fused vectors $Z_i$ and $Z_j$, two different distance measures are used. They are:

Euclidean distance ($s_{euc}$) = $\sum_{r=1}^{k}(z_{i,r} - z_{j,r})^2$

Threshold Absolute Distance or TAD ($s_{tad}$) = $\sum_{r=1}^{k} I(\left|z_{i,r} - z_{j,r}\right|, t)$

Here, $I(y, t) = 1$, if $y > t$ (and 0, otherwise) and $t$ is a pre-specified threshold. Thus, we see that TAD measure determines the number of normalized feature values that differ by a magnitude greater than the set threshold $t$. One feature level score $s_{feat}$ is obtained from $s_{euc}$ and $s_{tad}$ using simple sum rule (Figure 3.4). Finally, the information at match score level $s_{match}$ and the feature level $s_{feat}$ are combined using simple sum rule to obtain the final score $s_{tot}$ (Figure 3.5).

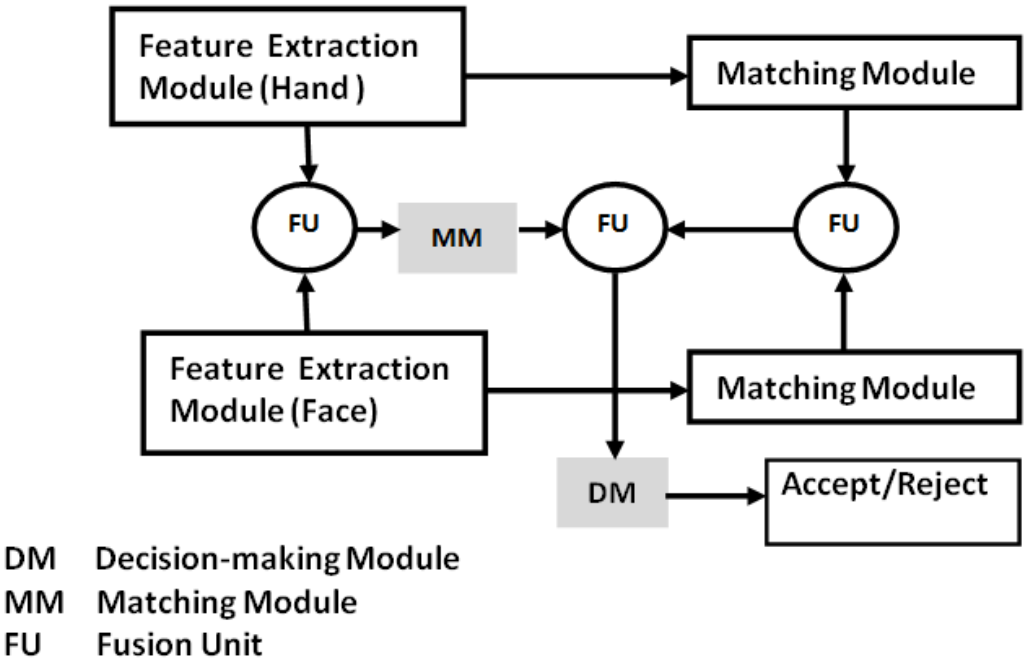



Figure 3.5: Information flow when data from the feature level and match score level are combined [50].

Ross et al. [50] carry out experiments in three different scenarios:

a) *Fusion of Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA) coefficients of face:* Two different face recognition algorithms based on PCA and LDA are combined at feature level (see [1] for details on these methods). It is observed that performance of LDA-based matcher is higher than the performance of PCA-based matcher. In this situation applying match score level fusion is found to degrade matching performance. The proposed fusion involving the combination of feature level and match score level fusion neither degrades nor improves matching performance. Authors mention that using fusion rules other than simple sum rule could have however improved performance.

b) *Fusion of R, G, B channels:* Three different feature sets are generated for a face image by subjecting each color channel to LDA separately. These feature sets are then combined at both feature and match score levels. It is observed that the scheme combining feature level and match score level information performs significantly better than match score level fusion.

c) *Fusion of Hand and Face Biometrics:* Face and hand feature sets are combined performing multimodal fusion. The matching performance of the scheme combining feature level and match score level fusion is slightly inferior compared to match score level fusion. However, when the same experiment is conducted with different dataset, the performance of the proposed scheme is found to be superior compared to match score level fusion.

## 3.3 Score Level Fusion

In score level fusion, different biometric matchers provide match scores indicating the degree of similarity between the input and template vectors. These match scores are consolidated to reach the final recognition decision. After the sensor level and feature level information, match scores contain the richest information about the input biometric sample. Fusion at score level provides the best tradeoff between the available information content and convenience of fusion. Therefore, this scheme is extensively studied in literature. This is also known as fusion at measurement level or confidence level.

From theoretical point of view the performance obtained by combining match scores from any number of matchers is guaranteed (on average) to be no worse than the best of the individual biometric matcher [25]. The key is to identify the appropriate method which combines the matching scores reliably and maximize the matching performance. Two guidelines for good combination of scores are mentioned in [25]. Firstly, each biometric matcher must provide a match score to the combiner. Secondly, in advance of operational use, each biometric matcher must make available to the combiner, its technical performance (such as score distributions).

Match scores generated by individual matchers might not be homogenous. For example, one matcher may produce a similarity score where a high value indicates better match whereas the other matcher may produce a dissimilarity score where a smaller value indicates better match. The match scores generated from different matchers may not be in the same range and may have different probability distributions. Because of these reasons, scores are normally normalized prior to fusion. However, some fusion methods use probability density functions (PDFs) directly and do not require normalization methods. The general flow of information in a match score level fusion taking normalization into account is shown in Figure 3.6.

Fusion methods at score level can be broadly classified into three categories [54]: density-based schemes, transformation-based schemes, and classifier-based schemes.

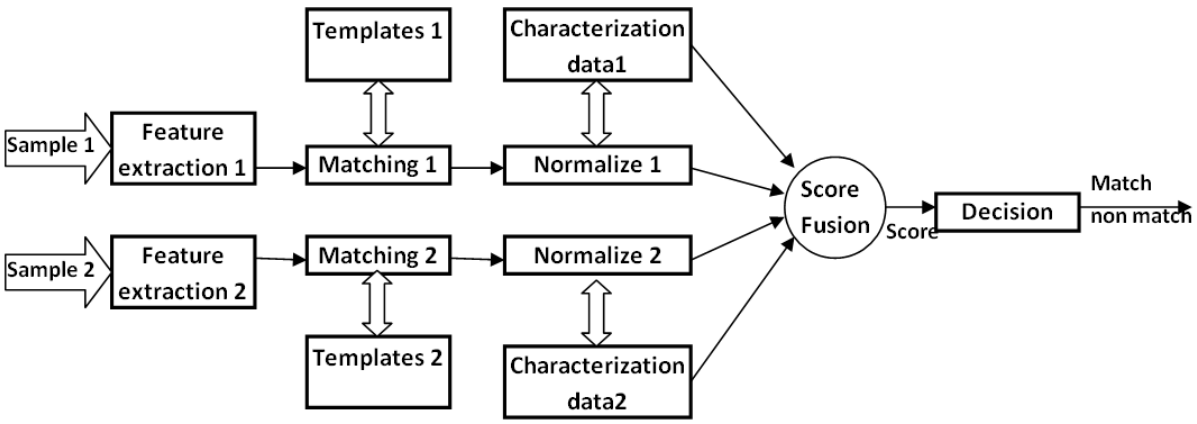



Figure 3.6: Match score level fusion [25].

### 3.3.1 Classifier Combination Rules

Kittler et al. [34] developed a common theoretical framework for consolidating the evidences obtained from different classifiers. They consider a pattern recognition problem where pattern $X$ is to be assigned into one of the $M$ possible classes $(\omega_1, \omega_2, \dots, \omega_M)$ based on the evidence provided by $R$ classifiers. Each of the $R$ classifiers represents the given pattern by a distinct feature vector. Let $x_j$ denote the feature vector derived from input pattern $X$ and presented to the $j^{\text{th}}$ classifier. In the feature space each class $\omega_k$ is modeled by the probability density function $p(x_j|\omega_k)$ and its prior probability of occurrence is denoted by $P(\omega_k)$.

According to Bayesian theory [13], given the feature vectors $x_j,\ j = 1, \dots, R$, the pattern $X$ should be assigned to class $\omega_r$ that maximizes the posterior probability, i.e.

$$\text{assign } X \to \omega_r \text{ if}$$

$$P(\omega_r\,|x_1, \dots, x_R) = \max_k P(\omega_k|x_1, \dots, x_R) \qquad (3.1)$$

where $k = 1, \dots, M$. The Bayesian decision rule in equation (3.1) is known as the minimum error-rate classification rule in pattern recognition. Using Bayes theorem, the posteriori probability $P(\omega_k|x_1, \dots, x_R)$ can be rewritten as

$$P(\omega_k|x_1, \dots, x_R) = \frac{p(x_1, \dots, x_R|\omega_k)P(\omega_k)}{p(x_1, \dots, x_R)} \qquad (3.2)$$

where $p(x_1, \dots, x_R|\omega_k)$ is the conditional joint probability density of the feature vectors. The unconditional joint probability density $p(x_1, \dots, x_R)$ can be expressed as

$$p(x_1, \dots, x_R) = \sum_{l=1}^{M} p(x_1, \dots, x_R|\omega_l)P(\omega_l) \qquad (3.3)$$

[34] suggests approximations to simplify equation (3.2) which lead to five classifier combination rules used in practice. All the five combination rules are based on the assumption that the $R$ feature representations $x_1, \dots, x_R$ used are statistically independent. With this assumption the conditional joint probability density $p(x_1, \dots, x_R|\omega_k)$ can be expressed as

$$p(x_1, \dots, x_R|\omega_k) = \prod_{j=1}^{R} p\left(x_j\middle|\omega_k\right) \qquad (3.4)$$

where $k = 1, \dots, M$.

The assumptions in equation (3.4) is reasonable for the multimodal systems using features from different biometric traits (for example face, fingerprint and hand geometry) that are mutually independent [54]. However, for some systems like the multi-sample systems (for

example, two representations of the same finger), using the same representation scheme (for example, using minutiae) the assumption may be unrealistic.

*Product Rule*

Using equation (3.2) and equation (3.4) which assume the statistical independence of feature representation, the product decision rule given below is obtained.

$$\text{Assign } X \to \omega_r \text{ if}$$

$$P^{-(R-1)}(\omega_r)\prod_{j=1}^{R} P(\omega_r|x_j) = \max_{k=1,\dots,M} P^{-(R-1)}(\omega_k)\prod_{j=1}^{R} P(\omega_k|x_j) \quad (3.5)$$

Even if a single classifier output is close to zero, the product of $R$ posterior probability becomes very small and leads to wrong decision. Therefore, this scheme is very sensitive to errors in estimation of posteriori probabilities.

*Sum Rule*

In sum rule, it is further assumed that posteriori probabilities computed by the classifiers do not deviate much from the prior probabilities, i.e,

$$P(\omega_k|x_j) = P(\omega_k)(1 + \partial_{kj}) \quad (3.6)$$

Though this is a strong assumption, it may be readily satisfied when the input is noisy, leading to errors in the estimation of posteriori probabilities. Some simplifications using equation (3.5) and (3.6) leads to the sum decision rule given below.

$$\text{Assign } X \to \omega_r \text{ if}$$

$$(1-R)P(\omega_r) + \sum_{j=1}^{R} P(\omega_r|x_j) = \max_{k=1,\dots,M}\left[(1-R)P(\omega_k) + \sum_{j=1}^{R} P(\omega_k|x_j)\right] \quad (3.7)$$

The feature vectors $x_1, \dots, x_R$ contain significant discriminatory information about the pattern class. Therefore, the assumption that posteriori probabilities $P(\omega_k|x_j)$ do not deviate much from prior probabilities $P(\omega_k)$ is unrealistic in most cases. However, [34] showed that the sum rule is relatively insensitive to the errors in estimation of posteriori probabilities. Therefore, the rule works well and is routinely used in practice.

Starting from the decision rules (3.5) and (3.7) and introducing other assumptions, some other classifier combination strategies are developed in [34]. These combination schemes are the max rule, min rule, median rule and majority vote rule. All the schemes and their relationships are illustrated by Figure 3.7.

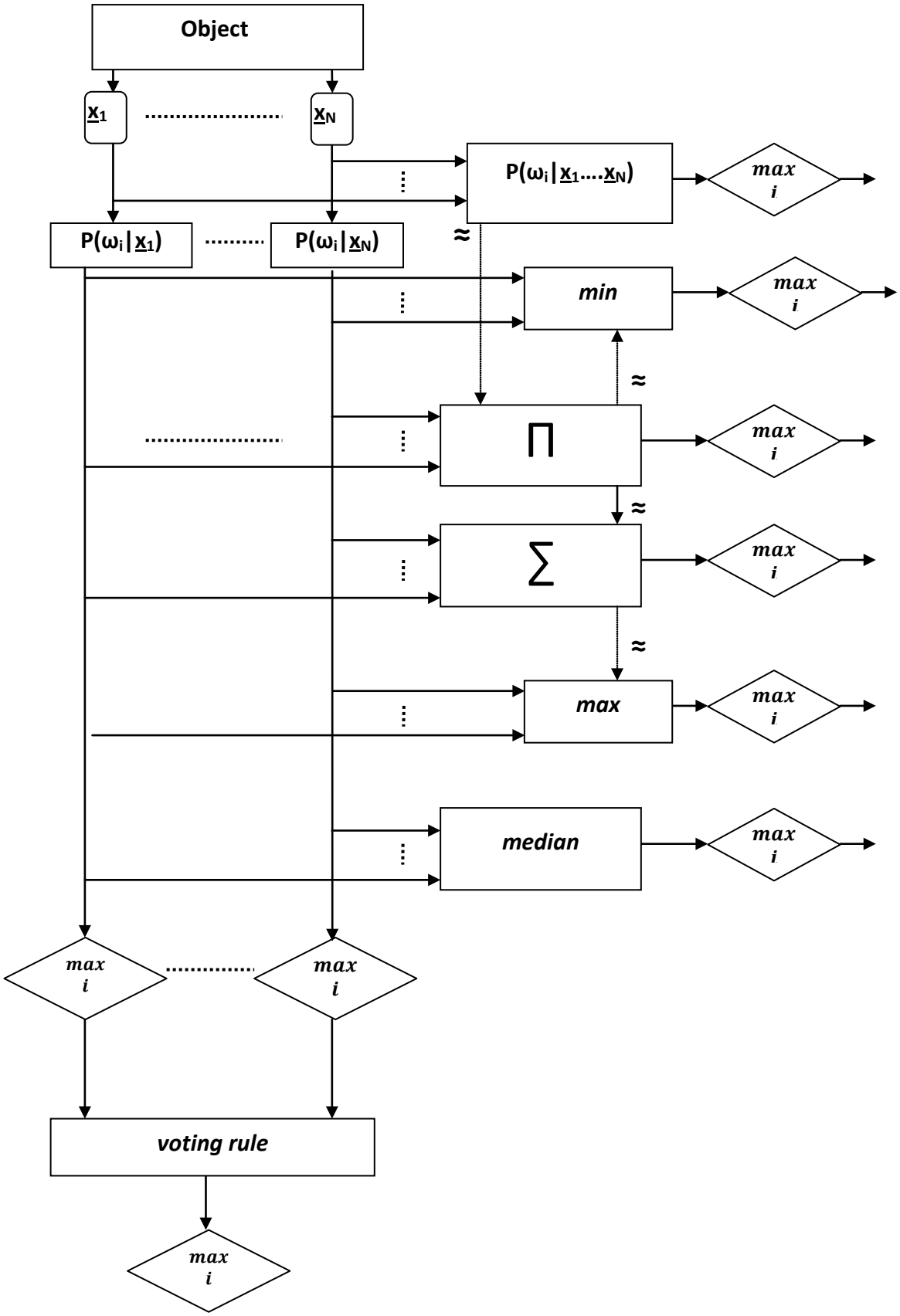



Figure 3.7: Classifier combination schemes and their relationships [34].

### 3.3.2 Score Fusion Techniques

If we consider a multibiometric system operating in the verification mode, the output of each biometric matcher is a match score (the formulation can be extended to identification scenario as well).The number of classes $(M)$ is reduced to two as our interest is to determine whether the input biometric sample $X$ belongs to a "genuine" user or an "imposter". All types of misclassifying errors are assumed to have equal costs in the minimum error-rate decision rule in equation (3.1). However, most of the practical verification systems assign different costs to FAR and FRR which would require modified Bayesian decision rule. If $\mu$ be the ratio of the cost values associated with false accept and false reject errors, the modified Bayesian decision rule becomes,

Assign $X \rightarrow genuine$ if

$$\frac{P(genuine|x_1, \dots, x_R)}{P(imposter|x_1, \dots, x_R)} \geq \mu \qquad (3.8)$$

Let $s_j$ represent the match score given by the $jth$ matcher for $j = 1,2, \dots, R$ (the subscript $k$ is dropped since the class is fixed as either genuine or imposter). As it is assumed that the feature representations of the $R$ biometric matchers $x_1, x_2, \dots, x_R$ are not available in score level fusion, the posteriori probabilities $P(genuine|x_1, x_2, \dots, x_R)$ and $P(imposter|x_1, x_2, \dots, x_R)$ need to be estimated from $s$. Classifer based score fusion, transformation based score fusion and the density-based score fusion are the three different approaches proposed for estimating these probabilities.

### 3.3.2.1 Density-based Score Fusion

We summarize the density based score fusion based on the survey in [54].

The density based score fusion approximates the posteriori probabilities $P(genuine|x_1, x_2, \dots, x_R)$ and $P(imposter|\, x_1, x_2, \dots, x_R)$ by $P(genuine|s = [s_1, s_2, \dots, s_R])$ and $P(imposter|s = [x_1, x_2, \dots, x_R])$, respectively. The conversion of vector scores $s$ into probabilities $P(genuine\ |s)$ and $P(imposter\ |s)$ requires the estimation of corresponding conditional densities $p(s \mid genuine)$ and $p(s \mid imposter)$. Finally, Bayesian decision rule in equation (3.8) can be applied to make a decision.

Suppose that $S_{gen}$ and $S_{imp}$ be the random variables representing the genuine and imposter match scores respectively. Let $F_{gen}(s)$ and $f_{gen}(s)$ be the distribution function and density function of $S_{gen}$ respectively. Then,

$$P(S_{gen} \leq s) = F_{gen}(s) = \int_{-\infty}^{s} f_{gen}(v)dv.$$

Similarly, let $F_{imp}(s)$ and $f_{imp}(s)$ be the distribution function and density function of $S_{imp}$ respectively. Then,

$$P(S_{imp} \leq s) = F_{imp}(s) = \int_{-\infty}^{s} f_{imp}(v)dv.$$

The density functions $f_{gen}(s)$ and $f_{imp}(s)$ are known as class conditional densities because they represent the probability density functions of the match scores given that score comes from genuine or imposter class $(p(s|genuine))$ and $(p(s|imposter))$, respectively. The density functions $f_{gen}(s)$ and $f_{imp}(s)$ are generally not known and have to be estimated from the set of training scores from the genuine and imposter classes respectively. Density estimation can be done either by parametric or non-parametric methods [13]. In the former, the training data is used for estimation of the parameters of density function with known form of the density function. In the later, i.e., non-parametric density estimation method, there are not any standard form for the density function and they are essentially data driven. A good selection of a specific parametric method for the density of genuine and imposter scores is a difficult task. The lack to the access of large amount of training data (especially genuine scores) for reliable estimation of genuine and imposter densities is another problem. Because of this limited availability of training scores, especially the genuine scores, the selection of density estimation method must be done carefully.

Snelick et al. [57] estimate the conditional densities of the match scores by parametric method. They assume a normal distribution for the conditional densities of the match scores. However, the assumption of normal distribution is generally not true for the biometric match scores. This method assumes that prior probabilities of the genuine and imposter classes are equal and the matchers used are statistically independent in the estimation process.

Jain et al. [30] estimate the conditional density of the genuine and imposter scores by the use of the Parzen window dependent non-parametric method. This method is suitable for estimating the conditional densities especially in cases when the densities are non-Gaussian. However, the output can be inaccurate because of the finite training data set and problems in the selection of optimum window width during density estimation process.

The approaches discussed above by Snelick et al. [57] and Jain et al. [30] estimate only the marginal densities of individual matchers. The combination of marginal densities is done under the assumption of statistical independence of the feature vectors or the biometric matchers based on framework by Kittler et al. [34]. Prabhakar et al. [44] state that the assumptions made by Kittler et al. [34] of statistical independence of matchers may not be true in a multi-algorithm biometric system that uses different feature representations and different matching algorithms on the same biometric trait. A non-parametric estimation of the joint multivariate density is proposed in [44]. In this approach the $R$ variate densities

$p(s_1, \ldots, s_R|genuine)$ and $p(s_1, \ldots, s_R|imposter)$ are directly estimated using the genuine and imposter test match scores. However, estimation of joint multivariate densities require larger number of training samples than estimating marginal (univariate) densities.

It is important to note that the biometric matching algorithms set certain thresholds at different stages in the matching process. This results in discrete components in the distribution of match scores which cannot be accurately modeled by using a continuous density function. To address this problem, discrete and continuous components of the density should be separately modeled in order to avoid large errors in estimating $f_{gen}(s)$ For such situations, [10] proposes a scheme for combining the match scores from multiple matchers based on generalized densities estimation.

### 3.3.2.2 Classifier-based Score Fusion

In classifier-based score fusion, a trained pattern classifier is used to learn the relationship between the vector of match scores $[\mathrm{s}_1, \mathrm{s}_2, \ldots, \mathrm{s}_\mathrm{R}]$ (from R matchers) and the posteriori probabilities of the genuine and imposter classes, namely, $\mathrm{P}(\mathrm{genuine}|\mathrm{s}_1, \mathrm{s}_2, \ldots, \mathrm{s}_\mathrm{R})$ and $\mathrm{P}(\mathrm{imposter}|\mathrm{s}_1, \mathrm{s}_2, \ldots, \mathrm{s}_\mathrm{R})$ [54]. The vector of match scores generated by multiple matchers is input to the trained classifier which classifies the vector into one of the two classes, genuine or imposter. The classifier defines two different decision regions in the feature space for the genuine and imposter classes. The decision regions are separated by a decision boundary. The decision boundaries can be simple as a line in a linear discriminant function or more complex such as multilayer neural networks depending on the complexity and nature of distributions of the two classes [25]. In general the classifier is capable of learning the decision boundary irrespective of how the feature vector is constructed. Therefore, the output scores from different matchers do not need to be transformed into a common domain prior to invoking the classifier. This fusion scheme requires a large number of training scores of both genuine and imposter classes during the training of classifier. A limitation of the classifier-based score fusion approach is that it is not easy to fix one type of error (say FAR) and then compute the other type of error(say FRR) at that specified FAR.

***Example 3.6 Score level fusion using classifiers based on the k-nearest-neighbor (k-NN) classifier, decision trees and logistic regression.***

Verlinde et al. [68] perform experiments and compare the performance of fusion using three different classifiers based on the k-nearest-neighbor (k-NN) classifier, decision trees and logistic regression. The three monomodal systems based on profile face image, frontal face image and voice provide match scores in parallel as input to the classifier module which has to take the decision either accept or reject. All the experiments are carried out using the multimodal M2VTS databases. All verification results are given in terms of FRR, FAR, and TER. For each error the 95% level confidence intervals given between square brackets.

The performances achieved by three monomodal identity verification systems based on profile face image, frontal face image and voice expert are given in Table 3.1.

| Expert | FRR (%) [37 tests] | FAR (%) [1332 tests] | TER (%) [1369 tests] |
|---|---|---|---|
| **Profile** | 21.6 [11.4, 37.2] | 8.5 [7.1, 10.1]] | 8.9 [7.5, 10.5] |
| **Face** | 21.6 [11.4, 37.2] | 8.3 [6.9, 9.9] | 8.7 [7.3, 10.3] |
| **Vocal** | 5.4 [1.5, 17.7] | 3.6 [2.7, 4.7] | 3.7 [2.8, 4.8] |

Table 3.1`: Verification results for single modalities [68].

A k-NN classifier is a very simple classifier that needs no specific training. It needs reference data points for both the imposter and client classes. The Eucledian distance between the test point and all the reference points is calculated and k-nearest neighbors corresponding to k-smallest Eucledian distances are sorted out. The test point is attributed the same class label as the class label of the majority of its k-nearest neighbors. Exhaustive distance calculation results in higher computing time which is the major drawback of this scheme. Since the experiments have large number of imposter (1332) and small number of client (37) reference points, large FRR and small FAR is observed resulting in rejection of lot of clients. Clustering technique is chosen as a solution. The clustering is performed by the k-means algorithm which allows to fix a priori the number of prototypes P. This algorithm uses the Eucledian distance measure to group the imposter references into P clusters. Each cluster is then replaced by the centroid of its samples. From the experiments performed after clustering, it is observed that for small P, FRR is very low and for large P, FAR is very low. The optimal number of imposter prototypes P depends on the cost-function specified by the application. The advantage with k-NN classifier with vector quantization (k-NN+VQ) is the considerable decrease in the number of calculations with the reduction in the number of imposter reference points. Though this method gives good results, the computing time is still high.

A decision tree is a tree-structured classifier representing the learned function using training data. Some tree classifiers are CART, C4.5, QUEST [38]. Some tree classifiers generate binary trees and some of them generate multi-branch trees. In [68], the C4.5 algorithm has been chosen. Decision trees classify unknown instances by sorting them from root to the leaf node. The topmost node is called root node. Leaf nodes are the terminal nodes represented by rectangles and are tagged with class labels. At each node in the tree some attribute of the instance is tested and each branch descending from that node corresponds to one of the possible values for this attribute. In [68] the attributes are match scores of the instance obtained for the different modalities. Split at the node is done in order to attain as homogeneous a set of labels as possible in each partition [38]. The best attribute is selected

for testing at the root. In C4.5 algorithm, the information gain determines the best attribute. Information gain is defined as the reduction in entropy caused by splitting the instances according to this attribute. When a full tree is grown it needs to be pruned before using it for classifying unknown instances. Pruning is done to avoid over-fitting problems that arise because of the training data that cannot represent the test data. Pruning is implemented by cutting back the tree branches moving from the bottom to the top. The process starts form the leaf node and moving upwards sub-trees are removed in intermediate nodes wherever required. Such intermediate node then becomes a leaf-node. The pruning criteria that is used in the C4.5 is reduced-error pruning which specifies that nodes are to be removed only if the resulting pruned tree performs no worse than the original one.

The influence of reducing number of imposter data points is studied in decision tree classifier as well by using k-means clustering algorithm as in the case with the k-NN based classifier. For lower value of P, the decision tree based classifier shows lower performance than the k-NN based classifier. However, decision tree requires lower computation time than the k-NN based classifier.

For a two-class problem, classification method based on the principles of logistic regression can be used. In this method, through statistical analysis of the training data, discrimination function which is the logistic distribution function, is implemented. The function implemented is:

$$E(Y/x) = \pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}}$$

In the expression, $E(Y/x)$ is the conditional probability for the binary output variable $Y$ given the input vector $x$. $g(x) = \beta_0 + \beta_1 x_1 + \cdots + \beta_d x_d$ where $x = (x_1, x_2, \ldots, x_d)$ is the $d$-dimensional input vector. Thus the function $\pi(x)$ gives the probability for the input vector $x$ of belonging to the class of clients $(Y = 1)$. Likewise, the probability $1 - \pi(x)$ for the input vector $x$ belonging to the class of imposters $(Y = 0)$ is also known.

The logistic regression parameters $\beta_i$s are obtained using the maximum likelihood principle in order to maximize the probability of finding the observed training data. Since each $\beta_i$ $(i \neq 0)$ multiplies one of the $d$-modalities, its value depends on the importance of that particular modality in the fusion process. After $\beta_i$s are estimated on training data, $\pi(x)$ is calculated for test pattern. This calculated value is then compared to the theoretical optimal threshold, which is the EER threshold calculated on training data. It is observed that the results in this experiment are the best among the three experiments conducted. Moreover, the computing time required in this method is less than the other two methods.

Table 3.2 gives summary of best verification results obtained for all 3 classifiers.

| Method | FRR (%) [37 tests] | FAR (%) [1332 tests] | TER (%) [1369 tests] |
|---|---|---|---|
| **k-NN** | 8.0 [2.7, 21.2] | 0.0 [0.0, 0.3] | 0.2 [0.1, 0.6] |
| **k-NN+VQ** | 0.0 [0.0, 9.4] | 0.5 [0.2, 1.0] | 0.5 [0.2, 1.0] |
| **Dec. Tree** | 7.7 [0.5, 13.8] | 0.3 [0.1, 0.8] | 0.5 [0.2, 0.1] |
| **Log. Reg.** | 2.7 [0.5, 13.8] | 0.0 [0.0,0.3] | 0.1 [0.0,0.5] |

Table 3.2: Summary table of verification results [68].

***Example 3.7 Decision trees and linear discriminant function classifiers***

Jain et al. [53] use the classifiers decision trees and linear discriminant function for fusion of match scores obtained from three modalities- face, fingerprint and hand geometry. The match scores from all three modalities are mapped to the range [0,100]. As the face and hand scores are distance scores, they are converted to similarity scores by subtracting them from 100.

*Decision trees:* The C5.0 program Quinlan (1992) is used to generate a tree from the training set of genuine and imposter score vectors. Both the training set and test set consist of 11,125 imposter score vectors and 250 genuine score vectors.

The confusion matrices (Table 3.3) given below illustrate the performance of the C5.0 decision tree.

| | Genuine | Imposter |
|---|---|---|
| **Genuine Class** | 239 | 11 |
| **Imposter Class** | 2 | 11,123 |

Evaluation on training data

| | Genuine | Imposter |
|---|---|---|
| **Genuine Class** | 226 | 24 |
| **Imposter Class** | 4 | 11,121 |

Evaluation on test data

Table 3.3: Confusion matrices indicating performance of C5.0 decision tree [53].

*Linear discriminant function*: Through linear discriminant analysis of training set, the three dimensional score vectors are transformed into a new feature space that maximizes the separation between the two-classes. The centroids of both classes in the new feature space are calculated. A test vector is classified by measuring the Mahalanobis distance from the vector to centroid of both classes, and assigning the vector to the class for which the Mahalanobis distance is minimum. The confusion matrices in Table 3.4 show the performance of the linear discriminant classifier on three different trials.

| | | Genuine | Imposter |
|---|---|---|---|
| **Trial 1:** | | | |
| | **Genuine Class** | 250 | 0 |
| | **Imposter Class** | 54 | 11,071 |
| **Trial 2:** | | | |
| | **Genuine Class** | 250 | 0 |
| | **Imposter Class** | 50 | 11,075 |
| **Trial 3:** | | | |
| | **Genuine Class** | 250 | 0 |
| | **Imposter Class** | 72 | 11,053 |

Table 3.4: Performance of linear discriminant classifier on three different trials [53].

Besides the two classifiers discussed, authors also use the sum rule which combines three scores corresponding to the three modalities. Weighted average of the scores from the multiple modalities is calculated. This is done for all possible combinations of the three modalities. Equal weight is assigned to each modality. Figure 3.8 and Figure 3.9 show the improvement in performance when the scores are combined using the sum rule.

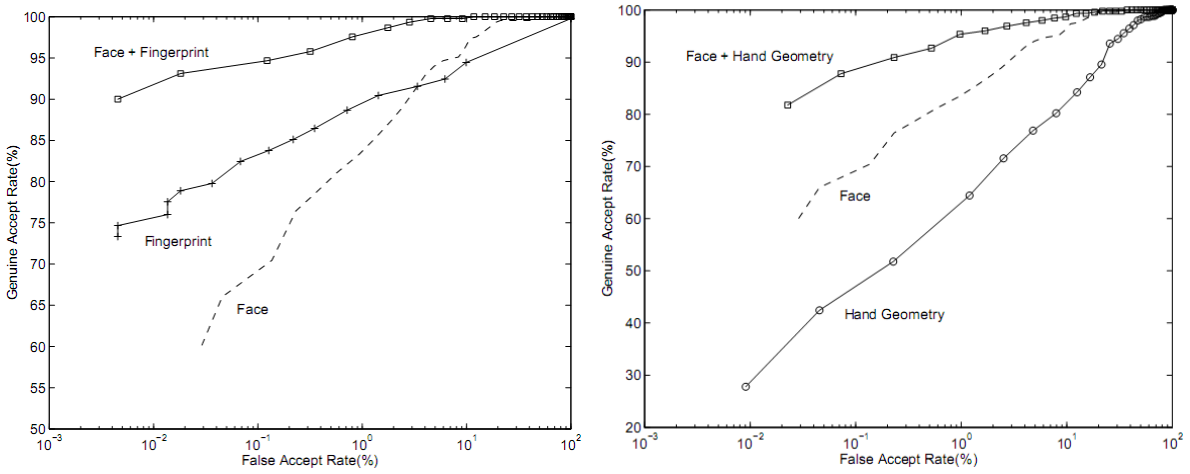



Figure 3.8: ROC curves when the scores are combined using the sum rule: (a) combining face and fingerprint scores (b) combining face and hand geometry scores [53].

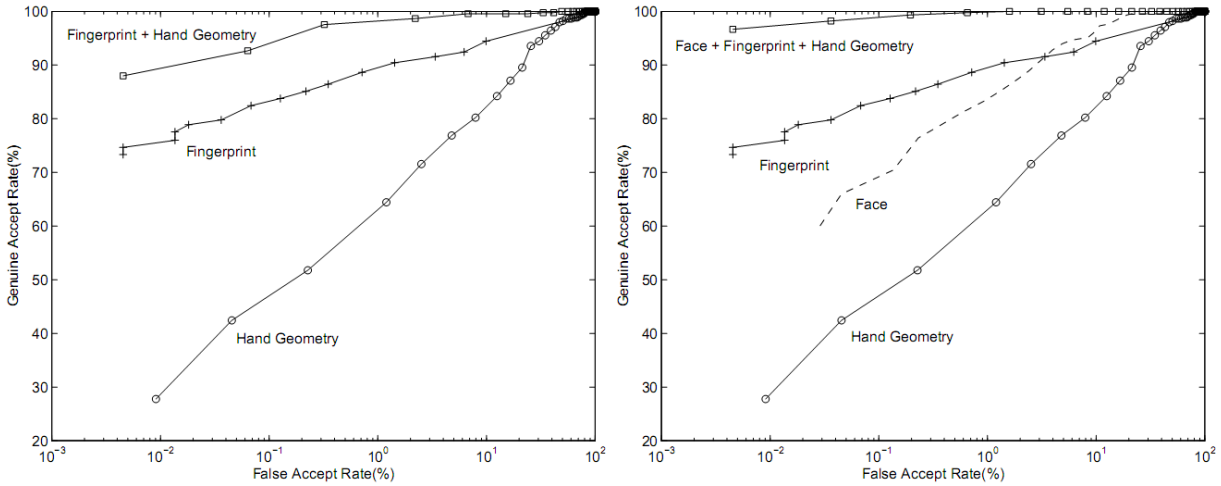



Figure 3.9: ROC curves when the scores are combined using the sum rule: (a) combining fingerprint and hand geometry scores and (b) combining face, fingerprint and hand geometry scores [53].

In the experiments, analysis of the FAR and FRR rates of all three schemes clearly suggested that the sum rule showed better performance than the other two classifiers.

### 3.3.2.3 Transformation-based Score Fusion

Density-based schemes require a large number of training samples (genuine and imposter match score) in order to accurately estimate the joint conditional densities $p(s = [s_1, s_2, \dots, s_R])|genuine)$ and $p(s = [s_1, s_2, \dots, s_R])|imposter)$ [54]. However, the availability of training data is limited due to the time, cost and efforts required in collecting data. In such a scenario, it is appropriate to combine the match scores from different matchers without converting them into posteriori probabilities. The match scores require to be compatible to achieve meaningful combination. A process known as score normalization is applied in order to transform the match scores from different matchers into a common domain. When the normalized match scores are available, different methods like the sum, max and min classifier combination rules can be applied to combine these scores.

#### Score Normalization

Score normalization refers to changing the location and scale parameters of the match score distributions so as to transform them into a common domain. A scale parameter determines the statistical dispersion of the probability distribution. A larger scale parameter implies a more spread out distribution and a smaller scale parameter implies a more concentrated distribution. The location parameter determines where the origin will be located and can be either positive or negative. The location parameter is used to shift a distribution in one direction or another. For a good normalization scheme, the estimates of the location and scale parameters of the match score distribution must be robust and efficient [20].

Robustness refers to insensitivity to the presence of outliers and efficiency refers to proximity of the obtained estimate to the optimal estimate when the distribution of data is known. Figure 3.10 shows the conditional distributions of the face, fingerprint and hand-geometry matching scores used in experiments by Jain et al. [26]. It is apparent from the distributions that the scores from different modalities are non-homogeneous and require normalization before they can be combined.

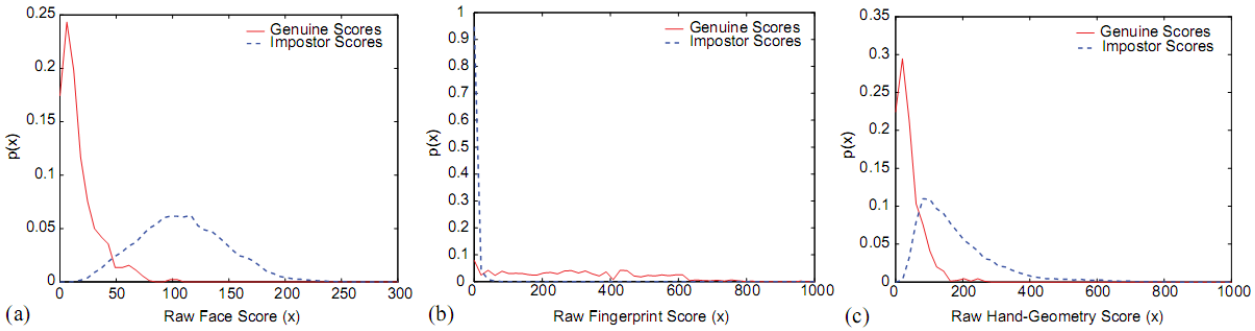



Figure 3.10: Conditional distribution of genuine and imposter scores: (a) face (distance score); (b) fingerprint (similarity score); and (c) hand-geometry (distance score) [26].

In the following paragraphs, we summarize some normalization techniques described in [26].

*Min-max normalization*: The simplest normalization technique is the min-max normalization. This technique is best suited for the case where the bounds (maximum and minimum values) of the scores produced by a matcher are known. In such case, the minimum and maximum scores can be easily shifted to 0 and 1 respectively. When the match scores are not bounded, minimum and maximum values can be estimated for the given set of training match scores and then min-max normalization can be applied. Let $\{s_k\}, k = 1, 2, \ldots, n$ denote a set of matching scores, then the normalized scores are given by:

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

In the case where minimum and maximum values are estimated from the given set of scores, the method is sensitive to the presence of outliers in the given data and hence is not robust. After min-max normalization, the original distribution of scores is maintained but now all the scores lie in the range [0, 1]. After normalization, distance scores can be converted to similarity score easily by subtracting the distance score from 1.

Figure 3.11 shows the distribution of the face, fingerprint and hand-geometry scores after min-max normalization.

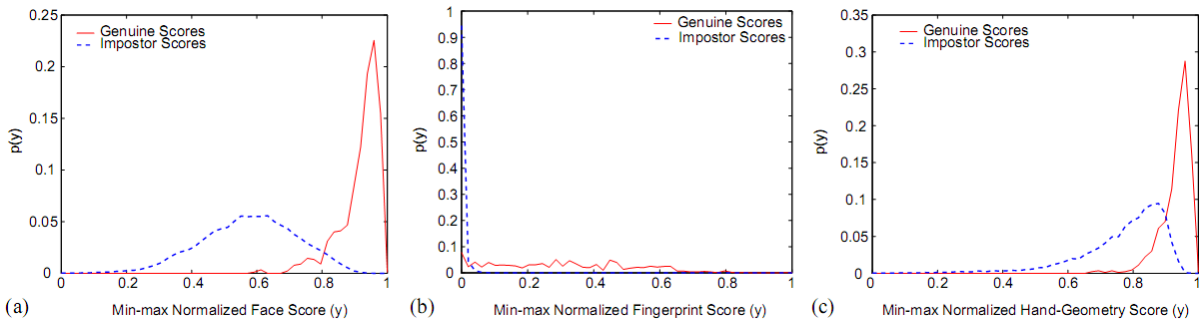



Figure 3.11: Distribution of genuine and imposter scores after min-max normalization: (a) face; (b) fingerprint ; and (c) hand-geometry [26].

*Decimal scaling normalization:* Decimal scaling can be applied when the scores of different matchers are on a logarithmic scale. The following normalization can be applied.

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

where $n = log_{10} max(s_i)$. For example, if two matchers had scores in the range [0, 10] and [0, 1000] this normalization technique can be applied and the scores can be transformed to a common range [0, 1]. The values of n in this case would be 1 and 3. The decimal scaling normalization is not robust. The other problem is the implicit assumption that the scores of different matchers vary by a logarithmic factor.

*Z-score normalization:* The most commonly used normalization technique is z-score normalization. It uses the arithmetic mean and standard deviation of training data. This scheme performs well when the average and variance of score distribution of matchers are known. If this prior knowledge is not available, the mean and standard deviation of the scores need to be estimated from given training data. The normalized scores are given by

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

where $\mu$ is the arithmetic mean and $\sigma$ is the standard deviation. Since both mean and standard deviation are sensitive to outliers, this method is not robust. This method does not guarantee a common numerical range for the normalized scores. If the input scores are not Gaussian distributed, this method does not retain the input distribution at the output. This is because the mean and standard deviation are the optimal location and scale parameters only for a Gaussian distribution. For arbitrary distributions, mean and standard deviation are the reasonable estimates of location and scale but are not the optimal.

Figure 3.12 shows the distribution of the face, fingerprint and hand-geometry scores after z-score normalization. It can be observed that the scores are not transformed into a common numerical range and for the fingerprint modality the original distribution of scores is not retained.

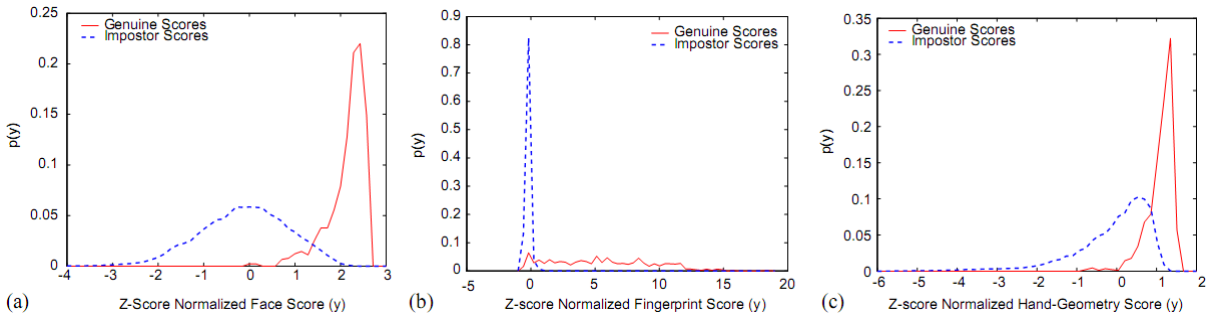



Figure 3.12: Distribution of genuine and imposter scores after z-score normalization: (a) face; (b) fingerprint ; and (c) hand-geometry [26].

*Median-MAD normalization:* The median and median absolute deviation (MAD) statistics are less sensitive to the outliers and the points in the extreme tails of the distribution. Therefore, a normalization scheme using median and MAD is robust. The normalized scores using this scheme are given by

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

where $MAD = median(|s_k - median|)$. For score distributions other than Gaussian, median and MAD are poor estimates of the location and scale parameters. Therefore, this technique does not preserve the input distribution and does not transform the scores into a common numerical range. Figure 3.13 shows the distribution of the face, fingerprint and hand-geometry scores after median-MAD normalization.

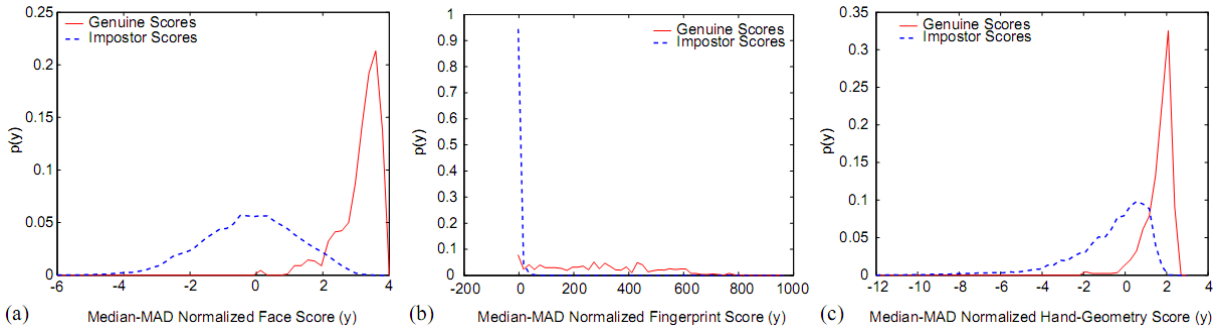



Figure 3.13: Distribution of genuine and imposter scores after median-MAD normalization: (a) face; (b) fingerprint ; and (c) hand-geometry [26].

*Double sigmoid normalization:* The normalized scores obtained using double sigmoid function are given by

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

where $t$ is the reference operating point and $r_1$ and $r_2$ denote the left and right edges of the linear region of the function. Figure 3.14 shows an example of the double sigmoid normalization, where the scores in the range [0,300] are mapped to the [0, 1] range using $t = 200, r_1 = 20$ and $r_2 = 30$.

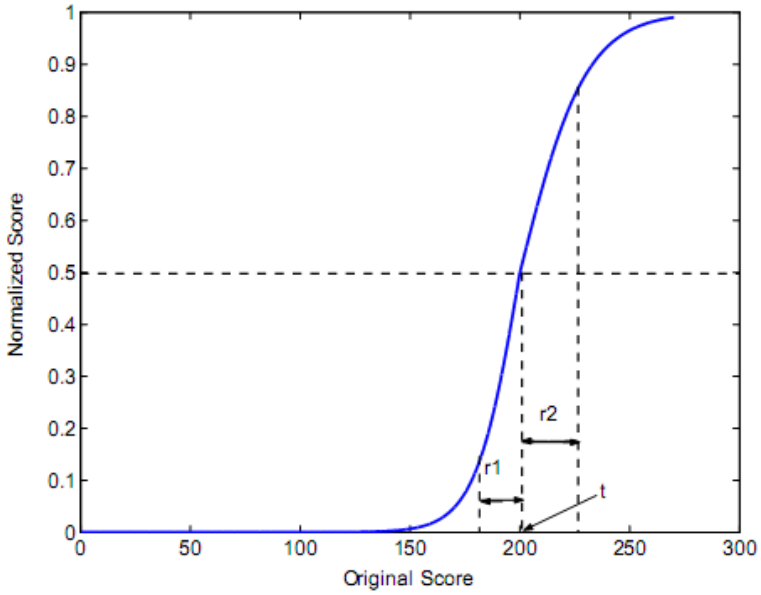



Figure 3.14: Double sigmoid normalization ($t = 200, r_1 = 20, and\ r_2 = 30$) [26]

This scheme transforms the scores into the [0, 1] interval. To obtain a good efficiency, appropriate tuning of the parameters $t, r_1, r_2$ is important. Generally, $t$ is chosen such that it falls in the region of overlap between genuine and imposter score distributions and $r_1$ and $r_2$ are set so that they correspond to the extent of overlap between the two distributions towards the left and right of $t$, respectively.

Figure 3.15 shows the distribution of the face, fingerprint and hand-geometry scores after double sigmoid normalization. It can be seen that all the scores are transformed to a common numerical range [0, 1], but the shape of the original fingerprint distribution scores is not retained.

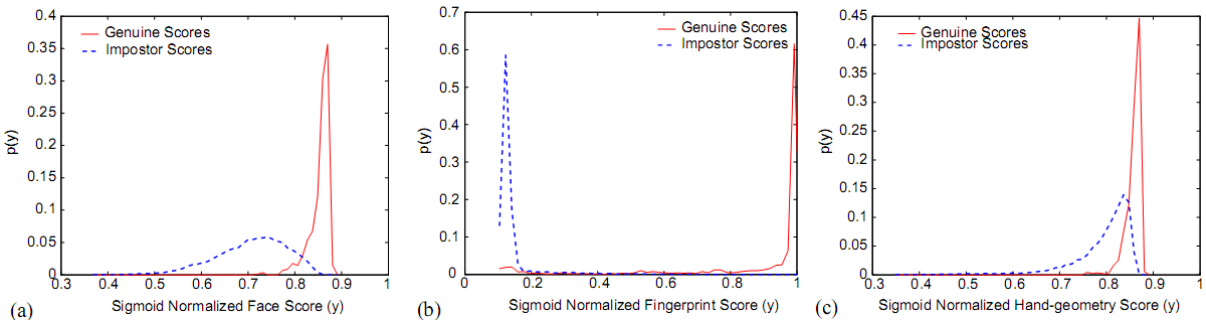



Figure 3.15: Distribution of genuine and imposter scores after double sigmoid normalization: (a) face; (b) fingerprint ; and (c) hand-geometry [26].

*Tanh normalization:* Hampel et al. [16] introduced the tanh-estimators which 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 $u_{GH}$ and $\sigma_{GH}$ are the mean and standard deviation estimates, respectively, of the genuine score distribution as given by Hampel estimators. Hampel estimators are based on the following influence ($\varphi$)-function:

$$\varphi(u) = \begin{cases} u & 0 \le |u| < a, \\ a\ sign(u) & a \le |u| < b, \\ a\ sign(\mu)\left(\dfrac{c - |u|}{c - b}\right) & b \le |u| < c, \\ 0 & |u| \ge c. \end{cases}$$

The Hampel influence function reduces the influence of the scores at the tails of the distribution (identified by a, b, c) in estimating the location and scale parameters. This normalization technique is therefore not sensitive to outliers. Discarding many points at the tails of the distribution results in an estimate that is more robust but not efficient (optimal). On the other hand, including many points at the tails of the distribution results in an estimate that is more efficient but not robust. Therefore, the parameters a, b, c must be carefully chosen depending on the required level of robustness which depends on the amount of noise present in the training data. Figure 3.16 shows the distribution of the face, fingerprint and hand-geometry scores after tanh normalization.

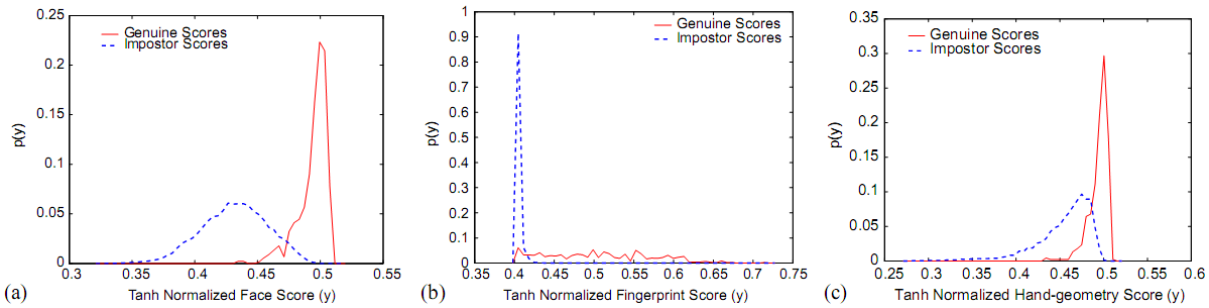



Figure 3.16: Distribution of genuine and imposter scores after tanh normalization: (a) face; (b) fingerprint ; and c) hand-geometry [26].

A summary of the characteristics of different score normalization techniques is given in Table 3.5.

| Normalization technique | Robustness | Efficiency |
|---|---|---|
| **Min-max** | NO | N/A |
| **Decimal Scaling** | NO | N/A |
| **Z-Score** | NO | High<br>(Optimal for Gaussian Data) |
| **Median-MAD** | YES | Moderate |
| **Double Sigmoid** | YES | High |
| **Tanh** | Yes | High |

Table 3.5: Summary of normalization techniques [26].

### *Example 3.8 Performance of various normalization techniques*

Choice of the normalization scheme giving the best performance depends on the fusion problem. It is recommended that a number of normalization schemes need to be evaluated to determine the optimal scheme for the given problem. We discuss the main results from the experiments performed by Jain et al. [26]. They studied the performance of a multimodal system where score-level fusion of face, fingerprint and hand geometry modalities was performed using different normalization and fusion techniques.

The recognition performance of the three unimodal systems is shown in Figure 3.17. From Figure 3.10 (b) and Figure 3.10 (c) it can be seen that the overlap between the conditional densities of genuine and imposter scores is highest for hand-geometry system and smallest for fingerprint system. This explains the worst performance of hand geometry based system and the best performance of fingerprint based system.

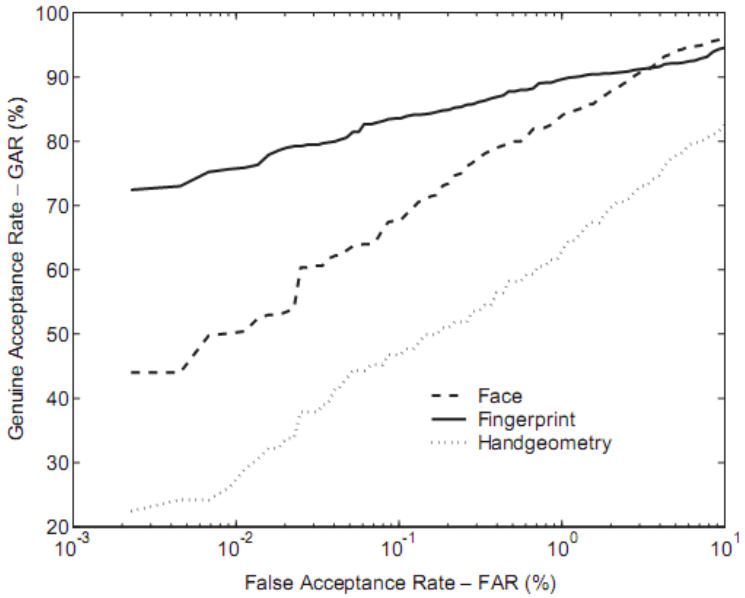



Figure 3.17: ROC curves for unimodal systems [26]

Three fusion techniques: the simple sum of scores, the max-score, and the min-score fusion are applied on normalized scores in [26]. Normalized scores are obtained by using the techniques: simple distance-to-similarity transformation with no change in scale (STrans), min-max normalization (Minmax), z-score normalization (ZScore), median-MAD normalization (Median), double sigmoid normalization (Sigmoid), tanh normalization (Tanh), and Parzen normalization (Parzen). We have mentioned in Section 3.3.2.1 that Parzen window density estimation method is used to covert match scores into posteriori probabilities which is not a normalization technique. "However, [26] treats the ratio of the posteriori probabilities of the genuine and imposter classes as a normalized match score and refer to this technique as Parzen normalization" [54]. The match scores output by different matchers may not be homogenous. One matcher may output a distance (dissimilarity) measure while other may output a proximity (similarity) measure. In order to obtain a set of homogenous match scores, simple distance to similarity conversion is performed in STrans transformation.

Table 3.6 summarizes the average Genuine Accept Rate (GAR=1-FRR) at a False Accept Rate (FAR) of 0.1% of the multimodal systems and the standard deviation of the GAR (shown in parentheses) for various normalization and fusion schemes. From the table it becomes clear that the sum of scores fusion gave better performance than the max-score and min-score techniques.

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

Table 3.6: GAR (%) of different normalization and fusion techniques at 0.1% FAR [26].

Figure 3.18 shows the recognition performance of the multimodal system when the scores normalized using various methods are combined by the sum of scores method. It is clear from the graphs that the sum of scores method gives better result than the fingerprint based system (the best unimodal system in this case) for all normalization techniques except the median-MAD normalization. At lower values of FARs, the tanh and min-max normalization techniques provide higher performance than other techniques. At higher FARs, z-score normalization provides slightly better performance than the tanh and min-max normalization. The min-max, z-score, tanh and distance-to-similarity transformation show similar performances. The raw scores of the three modalities used in the experiment are comparable. It is to be noted that when the raw scores are significantly different, distance-to-similarity transformation method does not work.

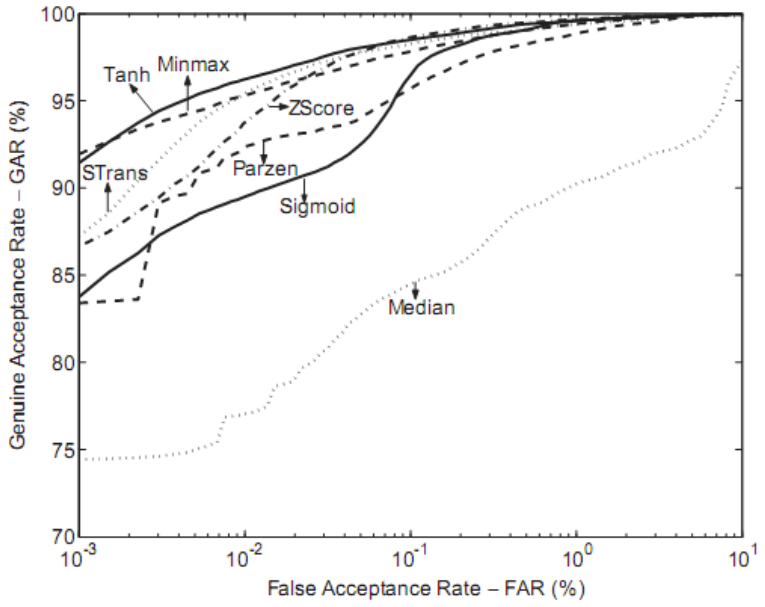



Figure 3.18: ROC curves for sum of score fusion method [26].

Jain et al. [26] also perform experiments which reveal that both min-max and z-score methods are sensitive to outliers whereas tanh method is both robust and efficient. When the minimum and maximum values of scores produced by a matcher are known, min-max technique works well. The z-score normalization works well when prior knowledge about the average score and score variations of the matcher is available. However, if these values are to be estimated from training scores and the available training scores are noisy, then a robust normalization technique like tanh normalization is the appropriate choice.

Authors also compute weighted sum of scores based on user-specific weights. The weights for each user are computed based on the training scores [26]. The scores are first normalized using min-max and tanh normalization and then the weighted combined score is calculated. It is observed from the experiments that the use of user-specific weights results in a significant improvement in recognition performance.

## 3.4 Rank Level Fusion

Rank level fusion is appropriate in the multibiometric systems operating in the identification mode where each component biometric system associates a rank with every enrolled identity (a higher rank indicating a better match) [54]. Rank level fusion entails consolidating the multiple ranks output by individual biometric subsystem in order to determine a new rank for each identity. The final decision is established based on these new ranks for all identities. Ranks reveal more information compared to just the identity of the best match and less information compared to the match scores.

Ho et al. [19] describe three methods to combine the ranks assigned by different matchers. Those are the highest rank method, the borda count method, and the logistic regression method.

### Highest Rank Method

In the highest rank method each identity is assigned the highest (minimum value) rank of all the ranks computed by different matchers. Ties are broken randomly to achieve a strict ranking order. This method requires the number of matchers to be small relative to the number of identities which is usually the scenario in multibiometric identification systems. If this is not the case, many identities could have ties and the final ranking is not reliable. The advantage of this method is its ability to utilize the strength of each matcher. As long as at least one matcher assigns a high rank to the correct identity, it is highly probable that the true identity will get high rank after reordering.

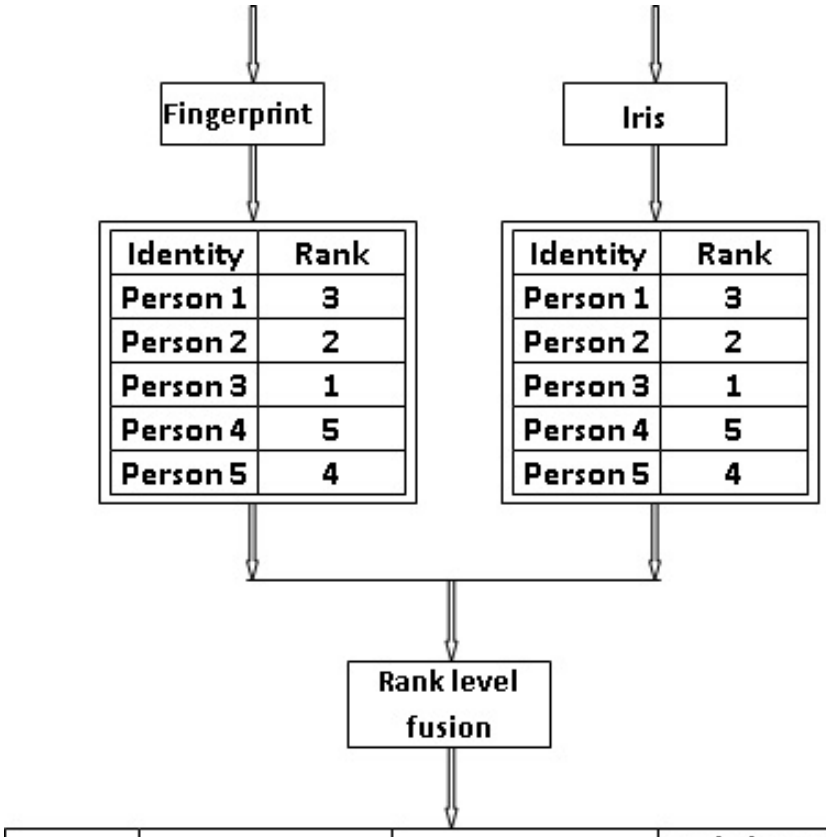
### Borda Count Method

In this method, the final rank for an identity is calculated as the sum of the ranks assigned to the identity by individual matchers. This method assumes that the ranks assigned to a given

identity by different matchers are independent. It also assumes similar accuracy of all matchers.

### Logistic Regression Model

Logistic regression model is a modification of borda count method where the final rank is calculated as the weighted sum of the individual ranks. Therefore, logistic regression model takes into consideration the differences in performances of different matchers. The weights are calculated during the training phase using logistic regression method.




| | Highest rank | | Borda Count | | Logistic Regression(Finger Print=0.3, iris=0.7) | |
|---|---|---|---|---|---|---|
| User Identity | Fused Score | Reordered rank | Fused Score | Reordered rank | Fused Score | Reordered rank |
| Person 1 | 2 | 3 | 5/2=2.5 | 2 | 2.3/2=1.15 | 2 |
| Person 2 | 1 | 1 | 3/2=1.5 | 1 | 1.3/2=0.65 | 1 |
| Person 3 | 1 | 2 | 6/2=3.0 | 3 | 3.8/2=1.9 | 4 |
| Person 4 | 4 | 5 | 9/2=4.5 | 5 | 4.3/2=2.15 | 5 |
| Person 5 | 3 | 4 | 7/2=3.5 | 4 | 3.3/2=1.65 | 3 |

Figure 3.19: Example of rank level fusion (adopted from [48] )

Figure 3.19 illustrates a simple example of rank level fusion using the highest rank, borda count and logistic regression methods. Ranks assigned to Person 1 by the fingerprint and iris matchers are 3 and 2 respectively. For the highest rank method, the fused score is the higher of the two ranks i.e, 2 for Person 1. After fused scores for all identities are deterimined using highest rank method, we can see that there is tie between person 2 and person 3 for rank 1. This tie is broken randomly and the final reordered rank is obtained. Using borda count method for Person 1, the fused score is obatined by adding the two ranks 3 and 2 and dividing the sum by 2 (the number of matchers) which results in 2.5. The highest rank and borda count methods both assume equal performances of both matchers. Therefore, the reordered ranks by these methods are influenced equally by the ranks assigned individually by both matchers. For logistic regression method, weights 0.7 and 0.3 are assigned for the iris and fingerprint matchers. A lower value of weight is assigned to the more accurate matcher (the fingerprint matcher being more accurate here).Therefore, the result here can be expected to be more influenced by the ranks assigned by the fingerprint matcher which is the case.

## 3.5 Decision Level Fusion

Decision level fusion is performed using the decisions output by the biometric matcher components. Many Commercial Off-the-Shelf (COTS) biometric matchers provide only the final recognition result which is the match or non-match decision. When such matchers are used in multibiometric systems, fusion is only possible at decision level.

Decision level fusion process is broadly categorized into a) Simple decision-level fusion and b) Advanced decision level fusion in [25].

### Simple Decision Level Fusion

When the multibiometric systems comprise of few components, AND and OR rules can be used to conveniently combine decisions output by different matchers. The output of fusion using AND rule is a “match” only when outputs of each component matcher is a “match”. The output of fusion using OR rule is a “match” if at least one of the outputs of component matcher is a “match”. Though these rules are convenient, the drawback is that they result in extreme operating points. The AND rule results in a lower FAR and a higher FRR than that of individual matchers. Likewise, the OR rule results in a higher FAR and a lower FRR than that of individual matchers.

Majority voting scheme is a commonly used approach in decision level fusion where the input biometric sample is assigned to that identity on which the majority of matchers agree [54]. When none of the identities is agreed by majority of matchers then “reject” decision is output by the system. This scheme assumes that performances of all matchers are similar. However, when the matchers used do not have similar recognition accuracy, it is reasonable

to assign higher weights to more accurate matchers. This is done in weighted majority voting scheme.

### Advanced Decision Level Fusion

Advanced decision level fusion into two subgroups 1) layered and 2) cascaded by [25]. Decision level fusion for both sub-groups are shown in Figure 3.20.

*Layered System*

A layered system uses the result of threshold test of a biometric score to determine the pass/fail threshold for the subsequent test of the next biometric score. Figure 3.20 shows an example of a layered decision level fusion comprising of three modalities P1, P2 and P3. The match score of P1 first enters the system. It is checked against the system threshold for the modality P1. The result (pass/fail) of this check determines the adjustment required for the next threshold for the modality P2. When the threshold for modality P2 is reset, match score of P2 enters the system. The same process is repeated for P2 and then for P3. After the process for P3 is completed, the final accept/reject decision is made.

*Cascaded System*

In cascaded systems the results of threshold test and strength test for a biometric sample determine whether additional biometric samples from other modalities are required in order to reach the final system decision. A simple model of a cascaded system shown in Figure 3.20. The match score of P1 enters the system which is checked against the threshold for P1. If the match score for P1 meets the threshold requirement, next decision is made on the strength of the result. If the strength is sufficient, then the final accept decision is made. If the strength is not sufficient or the match score fails the initial threshold test, the same process has to be repeated for the match score P2. If the match score P2 also does not pass both tests, the process has to be repeated for match score P3 as well. However, when one match score passes both tests, the system does not require samples of other modalities to be captured.

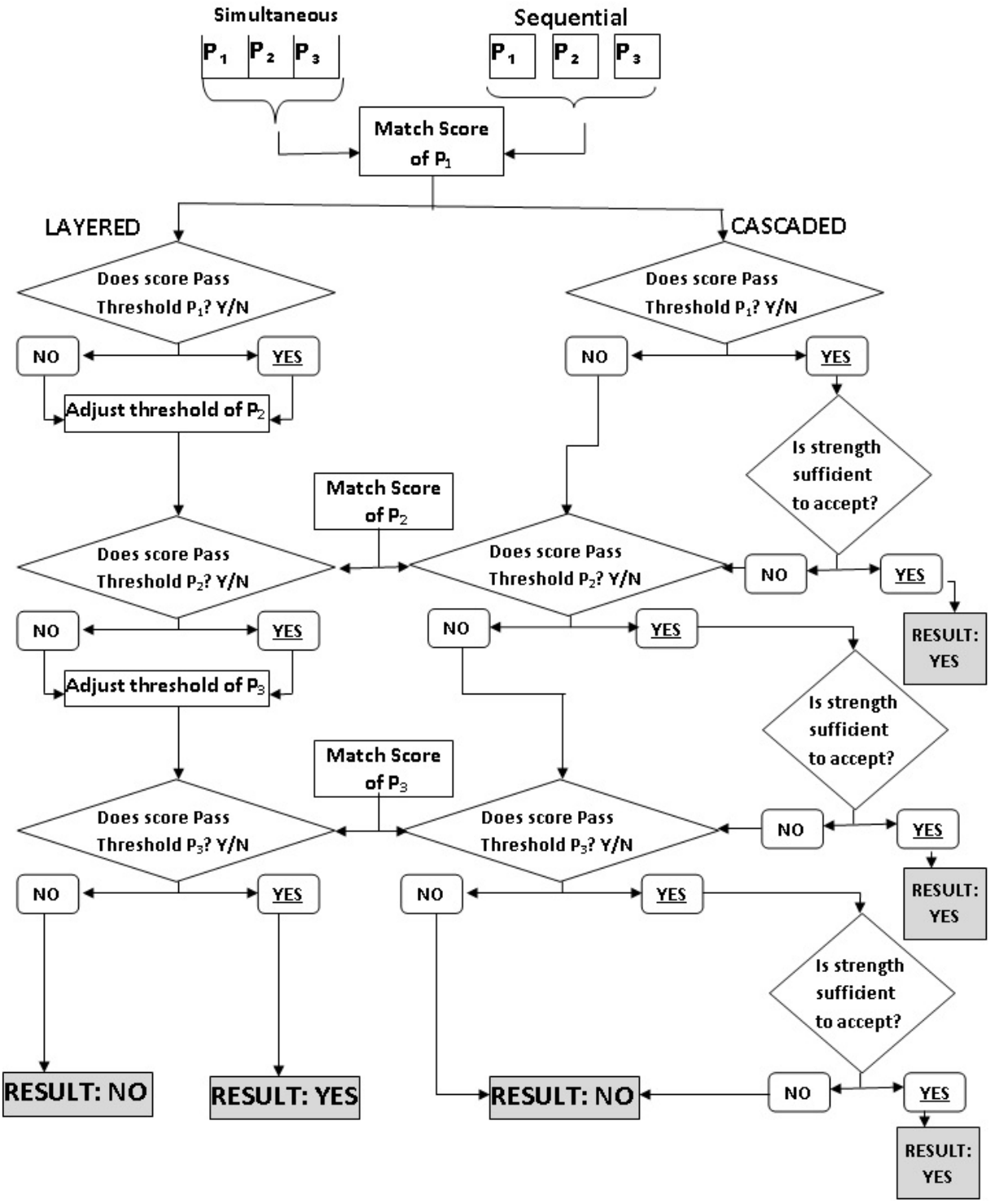



Figure 3.20: Advanced decision level fusion [25]

# Chapter 4

# Sources of Evidences

A multibiometric system performs recognition based on the evidences obtained from multiple sources of biometric information. Depending on the nature of sources, multi-biometric systems can be classified into six categories [54]: multi-sensor, multi-algorithm, multi-instance, multi-sample, multimodal, and hybrid. In this chapter we discuss all six scenarios. Table 4.1 below illustrates the five multibiometric categories by the simple case of using 2 of something.

| Category | Modality | Algorithm | Biometric trait (e.g., fingerprint, iris etc) | Sensor |
|---|---|---|---|---|
| Multi-sensor | 1(always) | 1(usually)[a] | 1(always, and same instance) | 2(always) |
| Multi-algorithm | 1(always) | 2(always) | 1(always) | 1(always) |
| Multi-instance | 1(always) | 1(always) | 2 instances(subtypes) of 1 body trait (e. g, left and right index finger) | 1(usually)[b] |
| Multi-sample | 1(always) | 1(always) | 2 samples of 1 biometric trait (e. g, 2 fingerprints of the same finger) | 1(always) |
| Multimodal | 2(always) | 2(always) | 2(always) | 2(usually)[c] |

Table 4.1: Comparison between the different multibiometric systems (categorized on the basis of sources of evidence) [58].

[a]Exception: It is possible that two samples from separate sensors are processed by using separate "feature extraction" algorithms, and then through a common comparison algorithm, making this "1.5 algorithms", or two completely different algorithms.

[b]Exception case may be using two individual sensors each capturing one instance.

[c]Exception: a multimodal system with a single sensor used to capture two different modalities(e. g, a high resolution image used to extract face and iris).

## 4.1 Multi-sensor Systems

In multi-sensor systems a single biometric trait is captured using multiple sensors in order to extract diverse information. For example, Chen et al. [8] investigate multi sensor face recognition system employing visible light camera and infrared camera. The PCA-based

recognition using visible light images showed better matching performance than PCA-based recognition using Infrared images. They further demonstrate that the integration of evidences provided by these two images substantially outperform the systems using either of these images.

***Example 4.1 Multi-sensor fingerprint system employing optical and capacitive fingerprint sensors***

We discuss the multi-sensor fingerprint system by Marcialis et al. [39] employing optical and capacitive fingerprint sensors. Fusion is performed at the match score level where the match scores obtained separately from two sensors are combined by score transformation fusion rules. Implementing multi-sensor fingerprint verification system demands increase in system cost and user co-operation. Therefore, acceptability of such systems depends on the improvement in verification accuracy achieved compared to that of a single-sensor system. In addition to improvement in performance accuracy, multi-sensor fingerprint verification system has other advantages too. A single sensor is not equally suited for all type of fingerprints. Hence, use of more than one sensor also increases the coverage of user population. Moreover, the difficulty of presenting fake fingers increases when multiple sensors are used as multiple sensors might require different fake fingers. This helps in prevention of fraudulent attempts.

Figure 4.1 shows the architecture of multi-sensor fingerprint verification system proposed in [39]. In the first step the fingerprint of user is acquired by both the optical and capacitive sensors. The acquired images from both the optical and capacitive sensors are processed to extract feature sets in terms of minutiae points. A minutiae based matching algorithm is separately applied to both the feature sets to obtain two match scores.

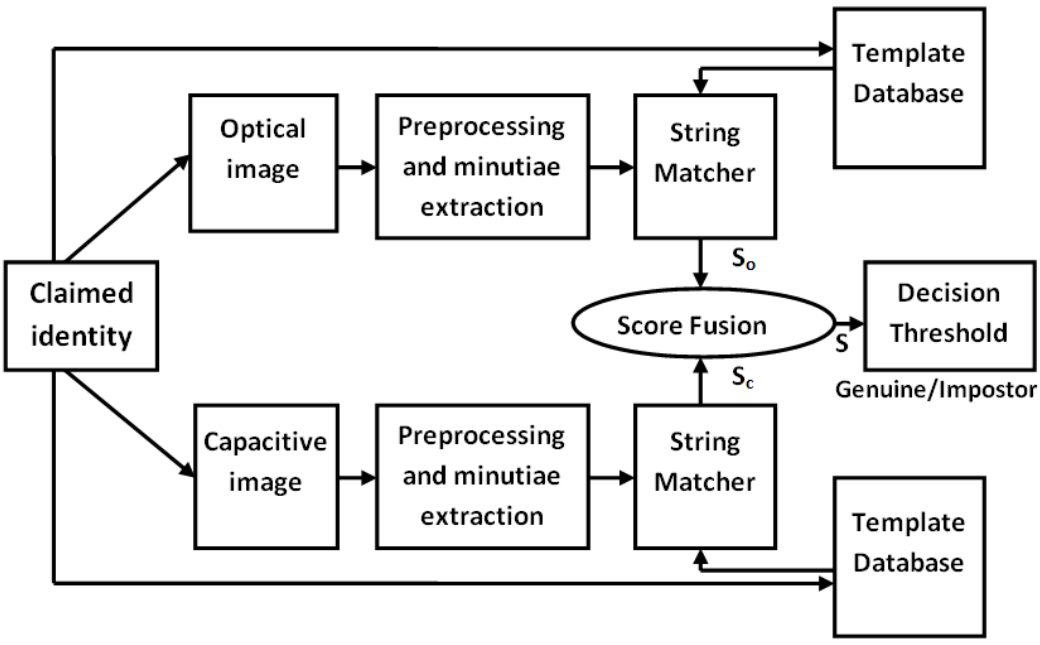



Figure 4.1: Architecture of the proposed multi-sensor fingerprint verification system [39].

Two different match scores are obtained after applying matching algorithm to the images acquired from optical and capacitive sensors. Authors investigate two kinds of score transformations for fusion of the obtained scores. In the first fusion rule, the fused score is determined as the mean of the two match scores. The second fusion rule, the logistic transformation, requires training phase as well to determine certain parameters in order to compute the fused score. If the fused score exceeds the threshold set, the identity is classified as genuine else it is classified as imposter.

The result from the experiment performed in [39] is summarized in Table 4.2. Firstly EER is computed on training set. Then, FAR and FRR are computed on test set using the EER threshold estimated from training set.

| | EER | FAR | FRR |
|---|---|---|---|
| Optical | 3.4 | 3.2 | 3.6 |
| Capacitive | 18.5 | 18.2 | 18.8 |
| Fusion by Mean | 2.9 | 3.1 | 2.8 |
| Fusion by Logistic | 2.3 | 2.4 | 2.3 |

Table 4.2: Errors of single and multi-sensor fingerprint verification systems [39].

The capacitive sensor performs very worse compared to optical sensor as expected. This is because the sensing surface is reduced when a capacitive sensor is used and this results in reduced number of minutiae extracted. The results indicate performance improvement after fusion. In particular, when the logistic fusion is implemented, the EER reduces from 3.4% (for the optical sensor) to 2.3%. Moreover, the results show that the deviation between the training set performance and test set performance is reduced by fusion, particularly when logistic fusion rule is used , improving the system's robustness. Despite the difference in performances between the optical and capacitive sensors, fusion provides better performance than the best individual sensor. This result suggests that the optical and capacitive sensors are strongly complementary. Complementarity refers to the capability of the sensors of recovering patterns misclassified by the other sensor.

## 4.2 Multi-algorithm Systems

Multi-algorithm systems process the same biometric sample using multiple algorithms. They can use multiple feature sets (i.e., multiple representations) extracted from the same biometric sample or multiple matching schemes operating on a single feature set [54]. These systems employ single sensor and hence reduce the cost as well as avoid the need for users to interact with multiple sensors. However, multi-algorithm systems require additional feature extractor modules or matching modules.

***Example 4.2 Fingerprint verification system combining three minutiae based fingerprint matchers***

Jain et al. [31] integrate the evidence obtained from three different minutiae based fingerprint matchers in order to improve performance of the proposed fingerprint verification system. Different fingerprint matching algorithms are usually based on different representations for a fingerprint and thus provide complementary information. In [31] output scores from three different minutiae matchers are integrated using logistic transform. The input and enrolled features (minutiae extracted by using minutiae extraction algorithm) are matched using one of the three matching algorithms or using a combination of these algorithms to obtain a matching score. The three matching algorithms used are Hough transform based matching, string distance based matching and dynamic programming based matching (see [31] for more on the algorithms). Results from experiments conducted using a large fingerprint database reveal the potential of integrating information from multiple matchers for performance improvement. The performance improvement achieved by integrating all three algorithms was the same as achieved by integrating string distance based matching algorithm and dynamic programming based matching algorithm. This is because the Hough transform based matching algorithm is substantially inferior compared to the other two algorithms and does not provide complementary information. Authors point out that integrating two matching algorithms do not guarantee improved performance. A poor matching algorithm may not add to overall performance improvement with its integration. Factors such as correlation between the matching algorithms used, disparity in the efficiencies of those algorithms and the fusion technique employed impact the performance that can be achieved with fusion [54].

***Example 4.3 Gait recognition system combining three gait classifiers based on environmental contexts***

Han and Bhanu [17] propose a context-based gait recognition system by probabilistically combining different gait classifiers based on different environmental contexts. This system is an example of multi-algorithm system combing different matching modules (classifiers). During enrollment phase templates representing individuals are stored (in the gallery set). These templates can be acquired under similar environmental condition for biometrics such as fingerprint and iris with sufficient discriminating features. However, this approach is not appropriate for gait recognition as changes in environmental context (for example, walking surface, temperature, carrying objects etc.) can lead to large appearance changes in detected human silhouette. Large gait variation occurring with changes in environmental contexts necessitates acquisition of more gallery images from all possible environmental contexts for each individual. However, this requirement is unrealistic in practical situations. In practice, limited number of gait gallery images are obtained under one or more environmental contexts for each individual. A single classifier is not able to reliably recognize an individual under different environmental contexts when the gallery images from those

contexts are not available. One classifier could be insensitive to changes in one context whereas another classifier could be insensitive to changes in other context. To improve recognition performance, it is possible to combine different classifiers for which the environmental context of the given probe sample (to be recognized) needs to be detected.

The basic idea of the proposed context-based recognition system is illustrated in Figure 4.2.

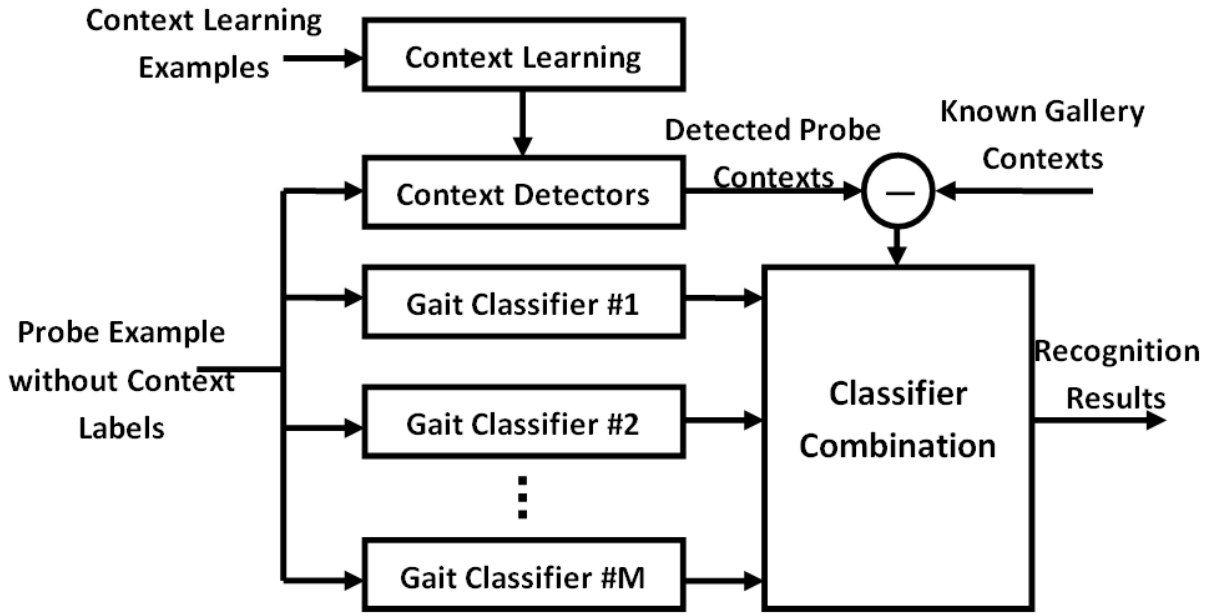



Figure 4.2: Gait recognition by combining context-based classifiers [17].

In the initial stage, the acquired context training samples are used in learning of context properties in order to construct context detectors. These context detectors are used to find out the context of a given probe gait sample. The context changes between the probe sample and gallery samples are determined under the assumption that all gallery samples are obtained under similar environmental contexts. One classifier is robust to change in one context whereas other classifier is robust to change in some other context. This makes different classifiers suitable for recognition under different context changes. Therefore, these classifiers are probabilistically combined based on the detected context changes for better recognition of a given probe gait.

Han and Bhanu [17] present a context-based gait recognition system where walking surface is the context type (see Figure 4.3). They design a real gate classifier for recognizing probe samples having no surface type changes with respect to gallery templates and synthetic gait classifier for recognizing probe samples with surface type changes. Experiments are carried out in three scenarios: i) using real classifier, ii) using synthetic classifier and iii) using the context-based combined classifier. In the third approach, the two gait classifiers (real and synthetic classifiers) are combined based on walking surface. Probabilistic approach is used in combining classifiers. In experiments where the surface type of probe samples is different from that of gallery samples, the performance of synthetic classifier is significantly higher

than that of the real classifier. In other experiments where the surface type of probe and gallery samples is the same, the performance of real classifiers is better. These results demonstrate the suitability of the designed real and synthetic classifiers for their desired contexts. The combined classifier based on the surface context shows better performance than the two individual classifiers in most experiments indicating the advantages of using context information in biometric fusion. When using the combined classifier, the context information is detected and the classifier takes advantage of the merits in individual classifiers. Only one walking surface type is detected and used as context information in the proposed system. Authors suggest that further performance improvement can be achieved if some more context information such as carrying objects, clothing etc. are detected, corresponding classifiers designed and incorporated into the recognition system.

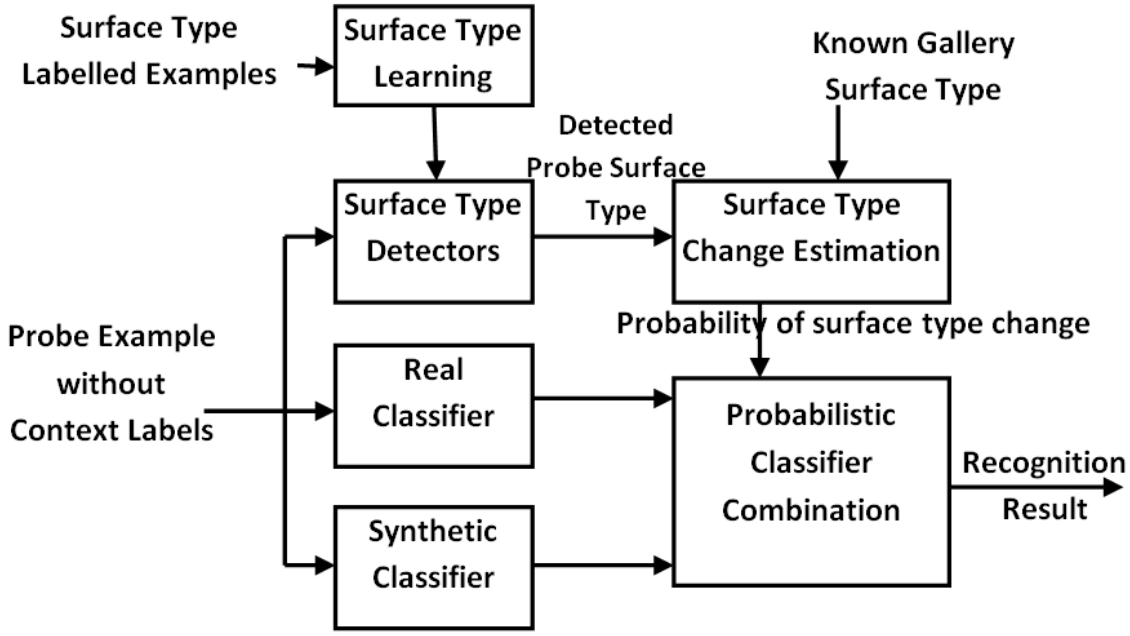



Figure 4.3: Gait recognition by combining context-based classifiers. The context investigated in the system is walking surface type [17].

## 4.3 Multi-instance Systems

Multi-instance systems involve fusion of information from multiple instances within the same biometric modality. For example, evidence from the left and right irises or the left and right index fingers can be combined for the recognition of an individual. Multi-instance systems are particularly useful for the individuals whose biometric traits cannot be reliably captured due to inherent problems. For example, it might not be possible to acquire sufficient features when the skin is very dry. In such cases, combining information obtained from fingerprints of multiple fingers of an individual provides with better discriminatory information required for recognition. Multi-instance systems generally do not require additional sensors and also do not necessitate new feature extraction and matching

algorithms. However, in some applications a new sensor arrangement might be required to capture various instances simultaneously [54]. Multi-instance systems are necessary in applications where the size of database is very huge. Integrated Automated Fingerprint Identification System (IAFIS) is a national fingerprint and criminal history system maintained by the Federal Bureau of Investigation (FBI) with a huge database [22]. FBI's IAFIS combines evidence from all ten fingers to determine a match in the database. The fingerprints from multiple fingers are obtained simultaneously in IAFIS.

***Example 4.4 Iris recognition system combining left and right irises***

Wang et al. [69] discuss a multi-instance iris recognition system where the left and right irises of an individual are combined. Figure 4.4 shows the block diagram of the proposed multi-instance iris recognition system. Iris recognition involves preprocessing, feature extraction, matching and decision making. During verification, the left and right irises go through the pre-processing and feature extraction steps individually. The features extracted for both the left and right irises are matched with their corresponding enrolled templates. Now, a score vector $(x_1, x_2)$ can be constructed, where $x_1$ and $x_2$ are the match scores obtained after matching the left and right irises with their respective templates. The next step is fusion at matching score level. Among the classification and combination approach to fusion, combination approach is preferred here. In combination approach scores are fused to generate a single scalar score which is then compared to the decision threshold for final decision. The decision threshold can be adjusted in order to meet requirements under different circumstances. Therefore, combination approach is chosen for the higher flexibility it provides. Then two matching scores from the two irises are fused using a fusion strategy based on minimax probability machine [37] to generate a fused score. When the fused score is obtained, decision on whether the individual is genuine or imposter is made based on the predefined threshold.

Experiments are performed on CASIA [6] and UBIRIS [46] iris databases in order to evaluate the performance of the proposed multi-instance fusion scheme. Experiments using single instance (either left or right iris) are also carried out. The experimental results for both the CASIA and UBIRIS databases show that the performance of multi-instance system is significantly better compared to the single-instance system. The EER for the left-iris, right-iris and multi-instance iris recognition systems on the CASIA database are 0.47%, 0.53% and 0.13% respectively. The EER for the left-iris, right-iris and multi-instance iris recognition systems on the UBIRIS database are 0.63%, 0.71% and 0.18% respectively.

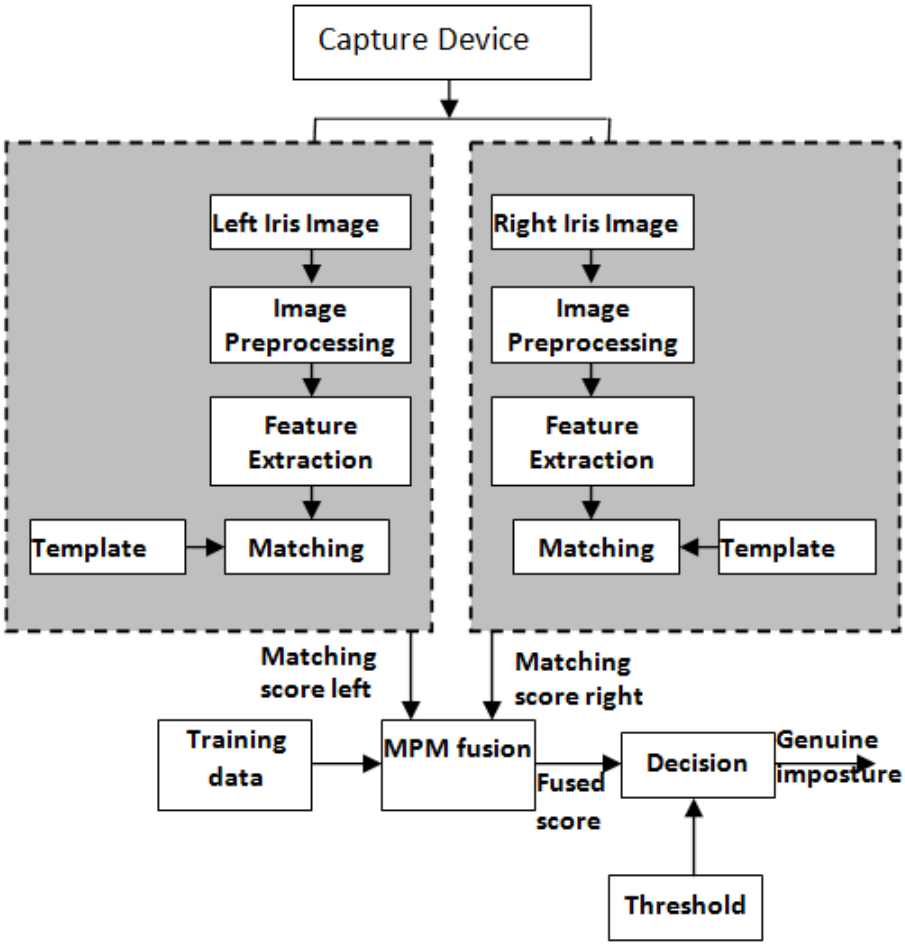



Figure 4.4: Multi-instance fusion block diagram [69].

***Example 4.5* Fingerprint *and eigenfinger-based multi-instance recognition system***

Uhl et al. [64] present a multi-instance fingerprint and eigenfinger-based biometric system. Acquisition of multiple instances in serial order results in additional transaction time costs. This can be avoided when multiple instances of a single biometrics are acquired from a single input source simultaneously. The proposed system in [64] exploits this advantage and the multiple features (minutiae and eigenfinger features) are extracted from a high-resolution scan of the entire palm. For fusion, transformation-based approaches at the confidence and rank levels are employed. The combination schemes max, median, min, product and sum are used for verification mode and score sum, borda count are used for identification mode. From the experiments conducted, it is verified that the multi-instance fusion increases performance significantly. In the case of combining solely eigenfinger or minutiae scores (multi-instance intra-feature fusion), sum rule showed the best performance. The total EER for combination involving minutiae and eigenfinger are observed to be 0.21% and 1.45% respectively when the sum rule is used. They also evaluate the case where cross feature combination of all minutiae and eigenfinger scores is performed. This case is the heterogeneous multiple-matcher and multiple instance combination of minutiae and eigenfinger scores. In this total combination scenario, product rule showed the best

performance with the lowest error rates of 0.08% EER, 0.55% zero false match rate (the lowest FNMR for zero false matches) and 0.14% zero false non-match rate (the lowest FMR for zero false non-matches). Zero false match rate and zero false non-match rate are given for the assessment of high-security and high-convenience scenarios, respectively.

## 4.4 Multi-sample Systems

Multiple samples of the same biometric trait can be acquired by using a single sensor to account for the variations that can occur in a biometric trait or to obtain a more complete representation of the underlying trait [53]. For example, different profiles such as the frontal profile, left profile and right profile of a face can be fused to address challenges arising from variations in facial pose [53].

***Example 4.6 A multi-sample system using multiple impressions of the same finger***

Jain and Ross [29] describe a mosaicking scheme which constructs a composite fingerprint template using multiple impressions of the same finger. This is a multi-sample system where fusion is performed at sensor level. Because of the limited area of contact offered by solid-state fingerprint sensors, they are not capable of providing sufficient information for highly reliable user verification. Moreover, the limited overlap between template and query impressions results in fewer corresponding points (Figure 4.5). In order to address these problems [29] use image mosaicking technique. [29] discusses some advantages of the composite image constructed using the image mosaicking technique. When the templates used are the multiple individual impressions (of the same finger), the query image is compared to each of the individual template. Because of the small size of individual template impression, it is probable that the amount of overlap between the template impression and query image is small resulting in a false reject. Use of composite image can reduce the false rejects arising from this problem. Moreover, when a composite image is used as the template, only a single comparison is required. This reduces the matching time considerably. The composite image also avoids the need to weight the individual templates during the matching process.

a)Template Image b)Query image

Figure 4.5: Limited overlap between the two impressions of the same finger [29].

***Example 4.7 Comparative performances of the multi-sample and multi-sensor face recognition systems.***

Bowyer et al. [3] compare the performances of face recognition systems obtained using two approaches, a multi-sensor approach and a multi-sample approach. The multi-sample recognition system combines normal intensity images with three-dimensional or infrared images. The multi-sample recognition system combines multiple normal intensity images.

The variations in pose, lightening, facial expressions create challenges for face recognition systems. Therefore, sufficient accuracy might not be achieved for demanding applications by using a single intensity image provided by standard camera. Besides normal intensity images, the use of 3-D shape information and infrared images for face recognition have been investigated in literature. Some advantages with using 3-D image are that it allows for better pose correction and the shape is defined independent of lightening. The major advantage with using IR images is that they are relatively unaffected by changes in lightening conditions.

Experiments were performed on an acquired image dataset containing images of same persons for each of the three image types. The intensity, IR and 3-D images of a subject for a session were all acquired within a period of a few minutes. This ensured that the images of each type were comparable. A total of 191 subjects participated in one or more image acquisition sessions held at weekly intervals over a period of several months. The eigenface was used as the recognition algorithm with each image type.

In the beginning, eigenface recognition was performed using each of the three image types individually. The experimental results are presented in two formats: the CMC curve and the ROC curve. In case of both the ROC curve and the CMC curve, it was observed that the performance using 2-D images was slightly better than using 3-D images. The performance was significantly lower while using IR images compared to using 2-D or 3-D images. However, the images were acquired in controlled indoor lightening environment which is particularly well-suited to normal intensity images. If the images were acquired in outdoor environments with highly varying lighting conditions, then IR images could be expected to provide better performances than normal intensity images.

Secondly, fusion of different image types (a multi-sensor approach) was performed at score level. Using eigenface approach with the cosine of the Mahalanobis angle [67] as the distance metric, the obtained score for a match between two points in the face space range between -1 and +1. The scores from each face space are normalized linearly to the same range (the range chosen was 0 to 100). The scores are then combined using the weighted sum combination method. The weight for the score from a given face space is based on the distribution of the top three ranks in that space.

Weight ($w$) for each face space is computed as follows:

$$w = \frac{score_2 - score_1}{score_3 - score_1} \quad \ldots \quad (4.1)$$

$score_k$ is the $kth$ closest distance from a gallery point to the probe point. It is clear from equation $(4.1)$ that greater difference between the first and second ranked matches implies higher chances of the top ranked match being correct. The weighted sum rule sums the weighted scores for each gallery subject for all three face spaces and selects the gallery subject with the smallest sum.

Experiments were performed to evaluate the performances of multi-sensor systems. It was observed from the results that each of the multi-sensor recognition system showed improved performance over the single-sensor systems. The results from combining all three image-types were slightly better compared to combining any two of image-types. However, the difference between performances of two-image type based and three image-type based systems was not statistically significant. Authors suggest that using a more challenging dataset could possibly detect the performance differences if any.

The third categories of experiments performed were to evaluate the performances of multi-sample systems. In the same image acquisition sessions outlined earlier, four different intensity images were acquired for each person. The four images were taken with variations in lighting conditions and facial expressions. Each person was asked to make to make two facial expressions, smile and neutral expression, in each lightening condition. The two lighting conditions used were referred to as “LM” and “LS” and the two facial expressions as “FA” and “FB”. Therefore, the four image conditions were FALM, FALF, FBLM, and FBLF. The same eigenface algorithm that was used for the intensity images in earlier experiments was used in this case too. In this experiment two, three or all four images of a person were combined.

In order to understand how the overall score is computed, let us consider the first case when a person is represented by two images for the gallery and also by two images for the probe. In order to compare one probe person to one gallery person, each of the two probe images need to be compared to each of the two gallery images resulting in four match scores. The sum of these four match scores is then the overall match score obtained for matching this probe person to this gallery person. Likewise, when a person is represented by three images for the gallery and also by 3 images for the probe, the overall match score is obtained as the sum of nine match scores. When a person is represented by all four images then the overall score is a sum of 16 match scores.

When a person was represented by two images (the FALM and the FALF conditions), the rank one recognition rate was 96.1%. When three images (the FALM, FALF, and FBLM conditions) were used, the recognition rate was 98.4%. Using all four image conditions, the recognition rate reached 100%.

We present some interesting issues discussed in [3] while comparing multi-sample versus multi-sensor recognition performance. It was observed from the experimental results that using four intensity images acquired at various lightening and facial expression conditions achieves the same recognition performance as achieved using three different image-types. In both the multi-sensor and multi-sample recognitions the same eigenface algorithm and score level fusion were used. In this case, it appears that multi-sample recognition is a better choice as it is cheaper and more practical to acquire several intensity images than to acquire multiple image-types. However, using multiple images adds improvement in performance only if there is some variation between the individual images of the subject.

***Example 4.8 Template selection in multi-sample systems: a case study in fingerprints***

An important issue with multi-sample systems is determining the correct number of samples to be acquired from an individual. The samples acquired should represent the variability and typicality of an individual. In [66], two methods are proposed to automatically select prototype fingerprint templates for a finger from multiple templates stored. The template selection problem authors describe involves selecting K templates that best represent the variability as well as typicality observed in the given N fingerprint images of a single finger (K<N). It is assumed that the value of K is known.

In the first method proposed in [66], called DEND, K clusters are formed from N fingerprint impressions in such a way that impressions within a cluster possess more similarity compared to impressions in all other clusters. From each cluster, a representative impression that best represents the typicality of the impressions within the cluster is chosen resulting in K prototype impressions. It is clear that the template set selected by this technique represents the variability observed in the fingerprint impressions.

The second method, called the MDIST, sorts the fingerprint impressions based on their average distance from the other impressions, and then selects K impressions with the smallest average distances. Therefore, this method selects templates that exhibit maximum similarity with other impressions representing typicality observed in the fingerprint impressions.

The experimental result in [66] demonstrates that a systematic template selection procedure results in better performance than random template selection. It was also observed that the MDIST method results in better performance than the DEND method. MDIST selects a template set that represents typicality in the fingerprint impressions. Hence, the probability of the selected templates being matched correctly with the same fingerprint impression is high. The DEND method selects templates representing variability and can possibly select outliers as well, increasing the probability of false rejects. However, combination of both methods is desirable as they select template sets of complementary natures.

## 4.5 Multimodal Systems

Multimodal systems combine two or more different biometric modalities (body traits) for establishing identity. Multimodal systems have several advantages. Better recognition rates can be achieved combining different modalities. Higher performance improvement can be expected by using physically uncorrelated traits (e.g., fingerprint and iris) than using correlated traits (e.g., voice and lip movement) [54]. They provide very high protection against spoofing as it is quite difficult for an imposter to spoof more than one biometric trait simultaneously. Multimodal systems also address the problems of noisy data. Even if one input is very noisy, input from other biometric trait might aid in recognition process.

It can be expected that increasing the number of traits could improve recognition performance. However, the curse of dimensionality phenomenon dictates that there is bound to the number of attributes that can be used in a pattern classification system without degrading the performance. This upper bound arises because of the limited availability of training samples. Several practical considerations such as the cost of deployment, enrollment time, throughput time, ease of use etc. restrict the number of traits that can be used [54]. There are some problems in deploying multimodal systems. Cost and complexity of added sensors and the appropriate user interfaces are increased. It is also more difficult to control the acquisition environment simultaneously for several traits.

We have discussed (in Section 3.2) an example of multimodal system by Son et al. [59] where face and iris biometrics are fused at feature level. The experimental results reveal that this multimodal authentication system performs significantly better than unimodal systems. We have also discussed the multimodal system proposed by Jain et al. [30] (in Section 3.3.2.3) where face, fingerprint and hand geometry modalities are fused at score-level using various normalization and fusion techniques. In the following sections, we introduce few more examples of multimodal fusion from literature. All the multimodal biometric system examples described in this section vary in many terms such as the level of fusion, types of modalities being used, and the fusion strategy employed.

***Example 4.9 Multimodal system using fingerprint, face and speech***

Jain et al. [27] investigate a multimodal biometric identification system integrating face recognition, fingerprint verification and speaker verification. Preliminary results show that identity established by multimodal system is more reliable than the identity established by individual systems. The proposed system is targeted for verification applications where identity claimed by the user is to be authenticated. There are four main blocks in the proposed system: i) acquisition module ii) template database iii) enrollment module and iv) verification module. Acquiring of fingerprint images, face images and voice signal of users is done by acquisition module. Template database contains all template records of enrolled users. Enrollment module performs tasks as user enrollment, user deletion, user update, training, etc. The verification module is responsible for authenticating the identity claimed

by user. The verification process consists of fingerprint verification, face recognition, speaker verification and finally fusion. The fusion process integrates match scores obtained as outputs from the fingerprint verification, face recognition and speaker verification systems. Minutiae-based fingerprint verification is employed which consists of two steps: i) minutiae extraction and ii) minutiae matching [28]. In the scenario of personal identification, face recognition refers to static controlled full frontal portrait recognition [7]. Static implies the used facial portraits are still images (intensity or range). Controlled here means the type of background, lightening conditions, the distance between the acquisition devices and faces, etc. are fixed for image acquisition. Face recognition is performed using the eigenface approach. In [27] a text dependent speaker recognition system is implemented which uses left-to-right hidden Markov model to make a verification [5]. The input to the system is a visual of a random combination of four digits (1, 2, 7 and 9) on a video monitor which is to be spoken by the user for verification.

*Fusion*

If the outputs of individual systems are similarity (dissimilarity) scores, then fusion can be performed by accumulating the confidence associated with each individual decision. Authors propose fusion at match score level.

Let $X_1, X_2$ and $X_3$ be the random variables indicating similarity(dissimilarity) between the input and template for fingerprint verification, face recognition and speaker recognition respectively. Let $p_j(X_j|\omega_i)$ ( $j = 1, 2, 3$ and $i = 1, 2$) be the class-conditional probability density functions of $X_1$, $X_2$ and $X_3$. When $X_1$, $X_2$ and $X_3$ are statistically independent, the joint class-conditional probability density function of $X_1$, $X_2$ and $X_3$ can be expressed as:

$$p(X_1, X_2, X_3|\omega_i) = \prod_{j=1}^{3} p_j(X_j|\omega_i), \quad i = 1,2.$$

In order to classify the observation based on evidence provided by different sub-systems, any one of the various statistical decision theory frameworks can be chosen depending on the desired level of accuracy. In [27], the fusion scheme needs to determine a decision boundary which satisfies the FAR specification and at the same time minimizes the FRR. Neyman-pearson rule [42] is used to establish the decision boundary. Let $R^3$ be the three dimensional space spanned by$(X_1, X_2, X_3)$; $R_1^3$ and $R_2^3$ denote the $\omega_1$-region and $\omega_2$-region respectively ($R^3 = R_1^3 + R_2^3$) and $\in_0$ denote the FAR set in advance. According to Neyman-Pearson rule, a given observation $X^0 = (X_1^0, X_2^0, X_3^0)$, is classified as:

$$(X_1^0, X_2^0, X_3^0) \in \begin{cases} \omega_1, & if \dfrac{p_1(X_1^0, X_2^0, X_3^0|\omega_1)}{p_2(X_1^0, X_2^0, X_3^0|\omega_2)} > \lambda \\ \omega_2, & otherwise \end{cases}$$

where λ is the minimum value that satisfies the following

$$\lambda = \frac{p_1(X_1, X_2, X_3|\omega_1)}{p_2(X_1, X_2, X_3|\omega_2)} \quad and$$

$$\epsilon_0 = \int_{R_1} p_2(X_1, X_2, X_3|\omega_2) dX_1 dX_2\, dX_3.$$

Performance of the proposed multimodal system has been evaluated on a small set of data acquired in the laboratory. The genuine and imposter distributions are estimated from the training data. After the distributions are estimated, the decision boundary satisfying a pre-specified FAR is derived using the Neyman-Pearson rule. The ROC curves of the individual systems and the multimodal system are shown in Figure 4.6 in which the authentic acceptance rate (the percentage of genuine individuals being accepted, i.e., 1-FRR) is plotted against FAR. These graphs clearly demonstrate that the multimodal system showed better verification performance than the individual systems.

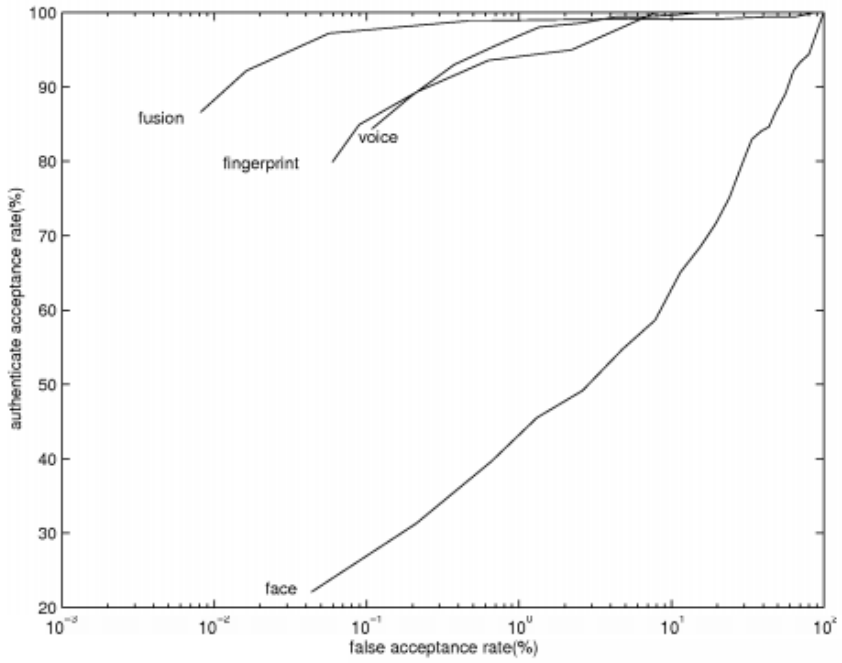



Figure 4.6: Receiver operating curves using Neyman-Pearson rule [27].

***Example 4.10 Multimodal system employing feature level fusion of face and palmprint traits***

Yao et al. [70] present a multimodal biometric system combining face and palmprint features. They present a novel approach to single sample biometrics recognition problem where discriminant features are extracted using Gabor based image processing and PCA techniques and then feature level fusion is performed designing a distance based separability weighting strategy. Single sample biometrics recognition problem, an extreme case of small sample size problem, is a challenge in Biometric Authentication. It might lead to bad recognition performance. Authors try to address this problem by implementing fusion strategy taking into account two important considerations. First consideration is selecting

appropriate biometrics having supplementary properties. In [70], face trait, a representative of contactless biometrics and palmprint trait, a representative of contact biometrics, are chosen for fusion. The second consideration is designing appropriate fusion method. In the case of single sample biometrics recognition, availability of rich information for fusion is very important. Therefore, fusion is performed at feature level since the feature vectors contain very rich information about the input pattern.

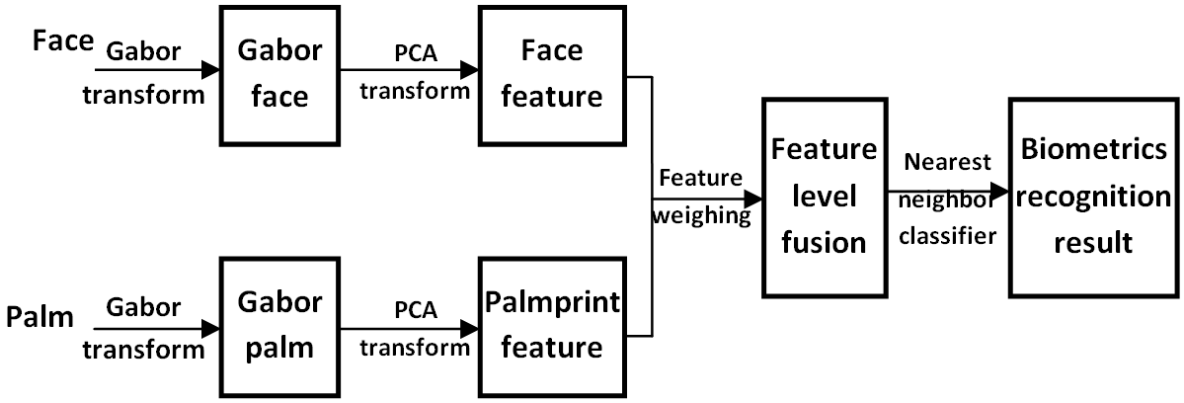



Figure 4.7: The single sample biometrics recognition procedure [70].

Figure 4.7 shows the entire recognition procedure adopted in [70]. Both Gabor and PCA transforms are used to extract the discriminant features. Gabor transform is useful for analyzing gradually changing data such as face, palmprint and iris images. Gabor filters can provide accurate time-frequency location and robustness against varying brightness and contrast of images [35]. In [70], circular Gabor filter is used. Let $X_{face}$ and $X_{palm}$ represent the face and palmrint image sample sets respectively. Gabor transform is performed on each sample in $X_{face}$ and $X_{palm}$ to obtain the transformed sample sets $X_{gabor-face}$ and $X_{gabor-palm}$. In the next step PCA transform is used to extract discriminant features from $X_{gabor-face}$ and $X_{gabor-palm}$ to obtain the corresponding face and palmprint discriminant feature sets $Y_{face}$ and $Y_{palm}$.

Before fusion, feature vector normalization is performed and . Let $y_{face}$ represent one sample of $Y_{face}$. Then, $y_{face}$ is normalized to obtain the normalized face feature vector $y_{nor-face}$ as follows:

$$y_{nor-face} = (y_{face} - \mu_{face})/\sigma_{face},$$

where $\mu_{face}$ and $\sigma_{face}$ represent the mean value and variance value of training sample set of $Y_{face}$. Similarly, normalized palmprint feature vector $y_{nor-palm}(of\ Y_{palm})$ is obtained. Weight value is computed using the weighing strategy that is directly linked to the Nearest Neighbour (NN) Classifier. NN-classifier finds the class with the least distance to the testing sample and assigns the same class label to the testing sample. It is assumed that there are $M$ testing samples in $Y_{face}$ and $Y_{palm}$. For a testing sample of $Y_{face}$, say the $j$th testing sample $(j = 1,2, \dots, M)$, distance vector $d$ is obtained using NN classifier where $d = [d_1, d_2, \dots, d_c]$,

$c$ is the class number and $d_1 < d_2 < \cdots < d_c$ . If $d'$ denotes the mean of $d$, then distance-based separability value of face features, $s_{face}$, for the particular testing sample is obtained as $s_{face} = d'/d_1$. Similarly, for the corresponding testing sample of $Y_{palm}$ , the separability value $s_{palm}$ of palmprint features is calculated. Let $w_j$ denote the ratio of separability values of palmprint and face feature vectors, that is, $w_j = s_{palm}/s_{face}$ . Weighing values $[w_1, w_2, \ldots, w_M]$ are computed for all $M$ testing samples and the average weighing value $w'$ is obtained. Weight w′ is assigned to all palmprint feature vectors and weight 1 is assigned to all face feature vectors. Now, the face feature vector $y_{nor-face}$ and its corresponding palmprint feature vector $y_{nor-palm}$ are serially combined as:

$$y_{fuse} = \left[y_{nor-face},\ w'.y_{nor-palm}\right].$$

Finally, a fused sample set $Y_{fuse}$ is obtained. Then, the NN-classifier is used to classify $Y_{fuse}$ . The distance $d(.)$ between a training sample $y_1$ and a test sample $y_2$ is obtained as,

$d(y_1, y_2) = \|y_1 - y_2\|_2$ , where $\|\ .\|_2$ denotes the Eucledian distance.

A large public face image database, the AR database [40] and a palmprint database provided by the Hong Kong Polytechnic University [35] were used to perform experiments.

| **Compared methods** | | | **Average recognition rates (%)** |
|---|---|---|---|
| **Undo Gabor transform** | **Single modal recognition** | **AR-PCA** | **43.25** |
| | | **Palm-PCA** | **56.4** |
| | **Multimodal fusion(weighing or not)** | **ARPalm-PCA-directfusion** | **77.07** |
| | | **ARPalm-PCA-weightfusion** | **80.49** |
| **Do Gabor transform** | **Single modal recognition** | **AR-GaborPCA** | **52.57** |
| | | **Palm-GaborPCA** | **62.72** |
| | **Multimodal fusion(weighing or not)** | **ARPalm-GaborPCA-directfusion** | **87.84** |
| | | **ARPalm-GaborPCA-weightfusion** | **90.73** |

Table 4.3: Average recognition rates using AR and palmprint databases [70].

Table 4.3 shows the average recognition rates obtained from the experiments using all the approaches. From the experimental results it is obvious that multimodal fusion of face and palmprint biometrics outperformed single mode recognition indicating suitability of face and palmprint fusion. Discriminant feature extraction based on Gabor and PCA transforms lead to significant performance improvement compared to using only PCA approach. Implementation of the proposed distance-based separability weighing strategy further improved the recognition performance.

***Example 4.11 BioID: A multimodal biometric identification system***

Implementing Biometric identification can ensure much higher security in systems controlling access to banking transactions, computer networks, and secured locations. For instance, biometric features such as fingerprint, face can be stored on a microchip in a secured card like a credit card. In systems depending on passwords or PIN numbers, imposter can steal these information and misuse the system without being detected. In a system implementing biometric identification even if the card in stolen, the system will reject access as the imposter's features will not match the features stored in the card.

We describe BioID which is a multimodal identification system combining three biometric traits: face, voice and lip movement [14]. BioID, developed by Dialog Communication Systems (DCS AG), is commercially available since 1998. This system employing three modalities achieves much higher accuracy than a unimodal system. Use of lip movement, which is a dynamic feature, makes BioID more secure against frauds than those systems which use only static features such as fingerprint.

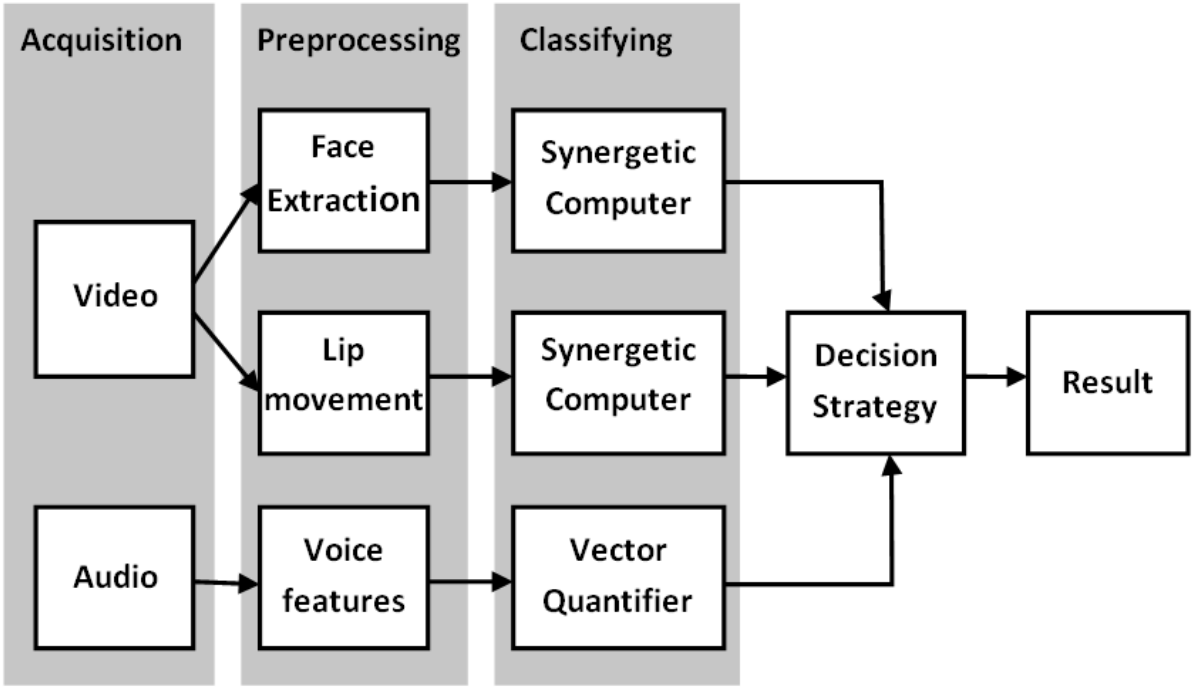



Figure 4.8: BioID's main functional units [14].

Figure 4.8 shows the four major functional units of BioIDs: Acquisition, Preprocessing, Classifying, and Decision Making units. The input acquired by the system is a recorded sample of a person speaking a word. The one-second sample consists of a video-sequence and an audio signal. From the video sequence, the preprocessing module extracts optical features for face and lip movement. In order to extract those features, the preprocessing module needs to determine the exact face location first. After the face boundaries are

detected, the eyes are located and then further processing for feature extraction takes place. Acoustic preprocessing module is responsible for extracting audio feature vector.

We discuss the vector quantifier [2] used in BioID as a classifier. Let us assume there are $d$ (here $d = 3$) features in which the classes distinguish themselves most. The first step to classification using vector classifier involves crossplotting the training data in a $d$-dimensional feature space (resulting in black points in Figure 4.9). This feature space is then segmented (dashed lines in Figure 4.9) in such a way that the resulting point clouds from different classes are best separated from each other. Each segment thus formed is assigned an alphabet symbol label (A, B and C in Figure 4.9). The next step is vector quantization in which a point is crossplotted in the $d$-dimensional feature set. The point is assigned the alphabet corresponding to the nearest segment based on calculation of Euclidean distance.

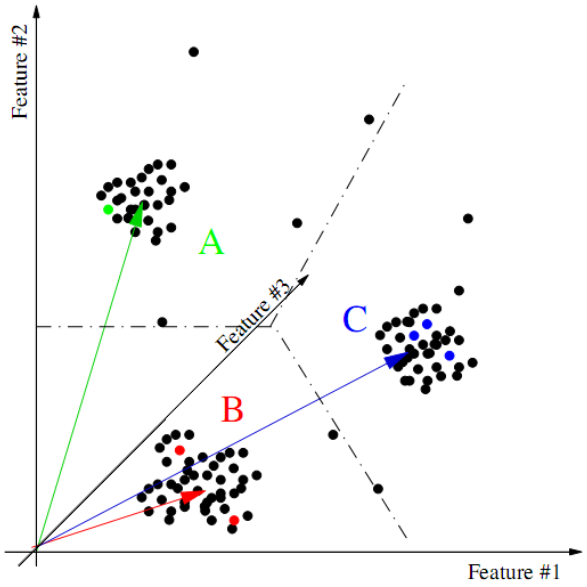



Figure 4.9: Sketch of a vector quantifier. A sequence of green, blue, blue, blue, red and red points is vector quantified by assigning alphabet corresponding to nearest segment resulting in the symbol sequence A, C, C, C, B and B [2].

During enrollment phase, biometric templates are generated for each feature for each person and stored. Classification process involves comparing the recorded input pattern with the corresponding stored templates. Synergetic computer is used to classify optical features and a vector quantifier is used to classify audio features. The synergetic computer is a set of algorithms that simulate synergetic phenomena in theoretical physics [15].

The classification results are combined into a final result selecting an appropriate strategy depending on the desired level of security. Figure 4.10 shows the available options for fusion. For normal operations, a two-out-of-three strategy is chosen in which classification results from two out of three traits need to agree to an enrolled class without falling below the threshold values set in advance. When higher security is desired three-out-of three strategy can be used which requires agreement of all three traits. This strategy helps achieve

very low FAR but FRR increases as well. The other option is finding fused result through weighted summation of all three classification results. It is possible to assign different weights to individual traits.

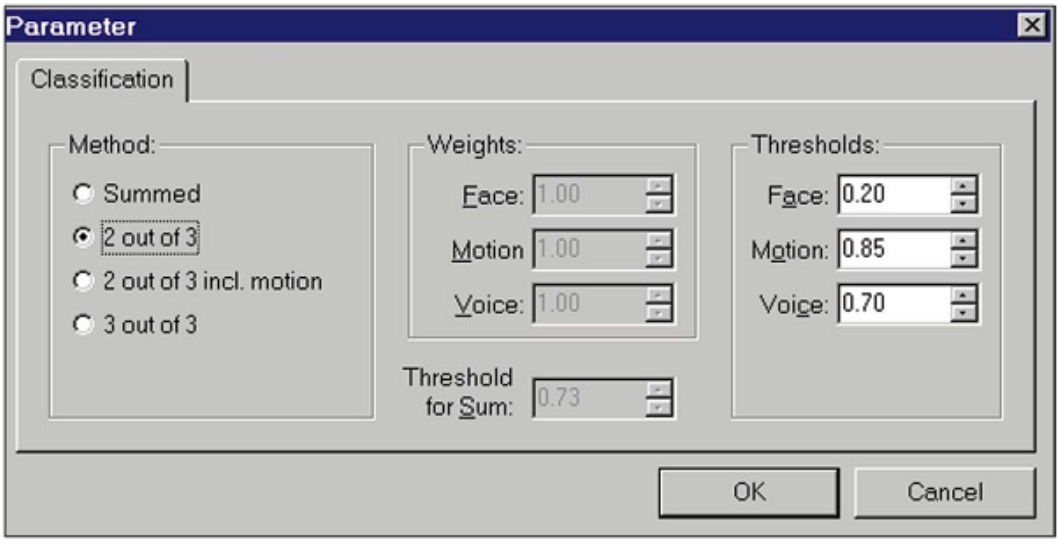



Figure 4.10: Sensor fusion options [14].

BioID functions in identification or verification mode depending on requirement. In identification mode, the system identifies a person through his biometric traits. In verification mode, a person gives his name or a number, which is then verified by the system using biometric traits. A test was conducted involving 150 persons for three months. The results showed that with BioID implementation, FAR were reduced significantly below 1 percent, depending on the desired security level. For higher security requirements, very low FAR has to be achieved and in doing so FRR increases. Therefore, an appropriate FRR needs to be determined without causing the FAR to increase undesirably.

## 4.6 Hybrid Systems

Hybrid term is used to refer to systems that integrate one or more of the five scenarios discussed above. Hybrid systems, if properly designed, can achieve the highest recognition performance in biometric systems. However, they are also the most complex systems, requiring higher storage, processing and architectural complexity.

The National Institute of Standards and Technology - Biometric Score Set Release 1(NIST-BSSR1) comprises of a set of output similarity scores from two different face recognition matchers operating on the frontal faces (multi-algorithm) and a fingerprint matcher operating on left and right index live-scan fingerprints (multi-instance) [61]. The release includes similarity scores from comparisons of faces and fingerprints of the same individuals. The release is intended to facilitate interested parties to investigate problems in the fields of biometrics. The data is particularly suitable for study of score level fusion in different

scenarios such as multimodal, multi-algorithm, multi-sample or a combination of two or more of these scenarios resulting in a hybrid scenario.

***Example 4.12 A multi-sample and multimodal (hybrid) biometric system***

We discuss a hybrid multibiometric system proposed by Thian et al. [43] where fusion of multiple samples obtained from multiple modalities is performed at score level. This system is both multimodal and multi-sample in its design. The basic idea of the proposed system is depicted in Figure 4.11.

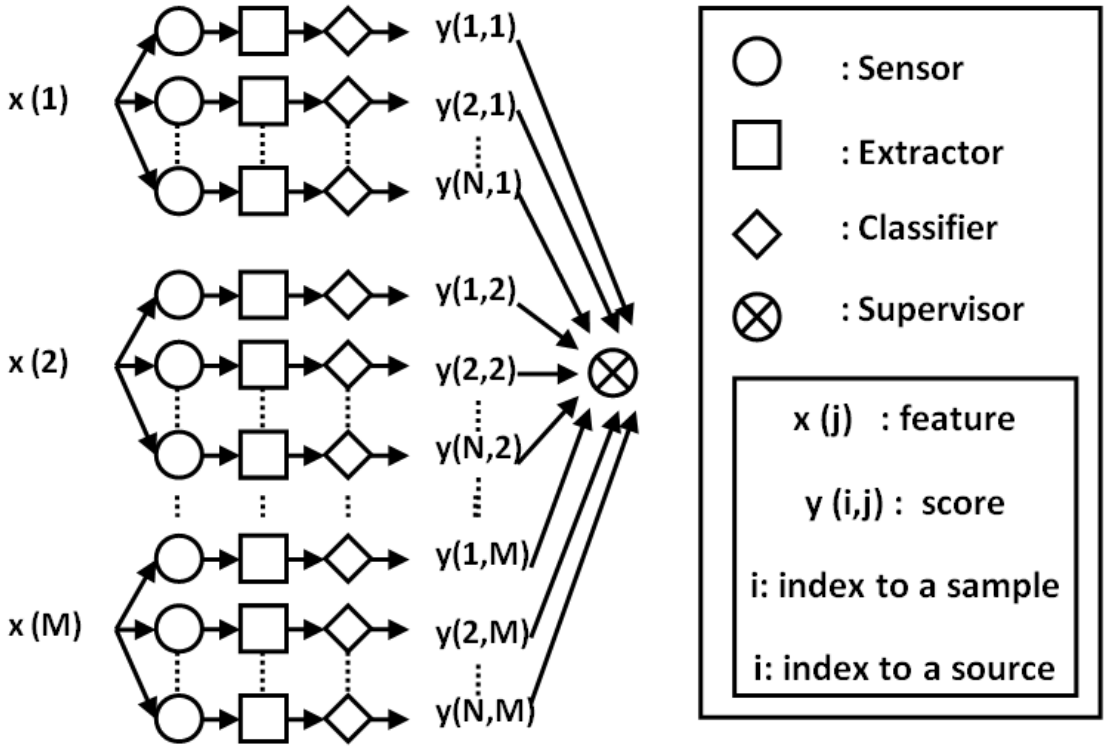



Figure 4.11: Multi-sample and multimodal (hybrid) biometric model [43].

In Figure 4.11, $x(j)$ is the $jth$ biometric modality and $j = 1,2, \dots, M$. $y(i,j)$ is the match score obtained from the classifier for the $ith$ sample of the $jth$ biometric modality and $i = 1,2, \dots, N$. The supervisor here combines scores from multiple classifiers and provides the final match score for decision making.

For conducting experiments, a publicly available LSIIT database [43] comprising of two modalities voice and face was chosen. The database consisted of 30 persons. Even with the small database of 30 persons perfect verification was achieved. With the increase in number of probe samples the accuracy improved faster for face modality than for speech modality. For experiments, 1 to 5 samples of each modality were taken. Improvement in accuracy was faster for samples of different modalities than for multiple samples of the same modality.

# Chapter 5

# Conclusion

Reliable and efficient identity management system has become very important in this highly interconnected world with increased concerns of identity fraud and national security. Biometric systems provide a greater degree of security and user convenience than the traditional authentication methods. Moreover, biometric systems also provide negative recognition and non-repudiation that traditional systems don't. Multibiometric systems, if properly designed, are able to increase the matching accuracy of a recognition system, increase population coverage and deter spoofing attacks.

Various types of information can be combined. In this thesis, we have discussed the various levels of fusion in multibiometric systems. Sensor level fusion combines the information at raw level. Although raw data is the richest in information, it is highly probable that raw data is contaminated by noise. Feature level fusion involves augmenting the feature sets originating from multiple information sources (from multiple feature extractors). Compared to the raw data, noise is suppressed in feature-level representation. Moreover, feature transformation algorithms can be applied to the augmented feature sets which enable the detection/removal of correlated feature values improving recognition accuracy. Match scores contain the richest information after the raw data and feature sets obtained from raw data. Moreover, it is easy to access and combine the match scores from different biometric matchers. Therefore, fusion at score level is the most common approach in multibiometric systems. Three main categories of score level fusion, namely, density-based, transformation-based and classifer-based schemes are studied in this thesis. Density-based schemes require a large number of training samples in order to estimate the joint conditional densities. When the available training data is limited, it is appropriate to use transformation-based schemes. Match scores generated from different matchers might not be homogeneous. We have discussed various normalization schemes which transform the match scores into a comparable domain. After the match scores are normalized, different classifier combination rules such as sum, max and min can be used for fusion. In classifier-based fusion, the vector of match scores generated by multiple matchers is input to the trained classifier. The trained classifier classifies the vector into one of the two classes, genuine or imposter. In rank level fusion each classifier associates a rank with every enrolled identity. Hence, rank level fusion is appropriate for systems operating in the identification mode. In decision level fusion, information is combined at abstract level. However, decision level fusion is the only viable approach for combining outputs from the commercial matchers which provide only the final recognition result.

Depending on how the multiple sources of evidence are obtained, multibiometric systems are categorized into: multi-sensor systems, multi-algorithm systems, multi-instance systems, multi-sample systems, multimodal systems and hybrid systems. It is very difficult to determine the best suited sources of biometric information for a specific application in order to achieve the best matching performance. During system design, factors such as cost, system speed and throughput, robustness, acceptability, ease of use, environmental flexibility, scalability, etc. have to be considered and tradeoffs must be made [54]. All these factors must be considered while selecting the sources of biometric information and a particular fusion strategy.

An important future addition to this work can be the study of fusion incorporating ancillary information. The input to fusion module consists of raw images, feature vectors, match scores, ranks or identity decisions provided by the individual biometric matchers. In addition, ancillary information may be available to some applications which can be utilized for decision making. Intrinsic ancillary information (e. g, quality of acquired biometric sample) is derived from the same sample used for verifying or establishing identity of the user. Extrinsic information such as gender, ethnicity, height or weight is not derived from the acquired biometric sample of the user. Ancillary information can be useful in different ways for recognition. The challenges associated with incorporating ancillary information in multibiometric systems can be studied.

# 6 References


1. Belhumeur, P., Hespanda, J. and Kriegman, D., “Eigenfaces versus fisherfaces: recognition using class specific linear projection,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, July 1997.

2. Beyreuther, M., “Speech recognition based automatic earthquake detection and classification”, Ludwig Maximilians University, Muenchen, Fakultaet fuer Geowissenschaften, PhD thesis, 2011.

3. Bowyer, K.W., Chang, K.I., Flynn, P.J., Chen, X., “Face recognition using 2-D, 3-D, and infrared: is multimodal better than multisample ?”, *Proceedings of the IEEE,* vol. 94, no. 11, pp. 2000-2012, November 2006.

4. Bromba, M.U.A., “Bioidentification frequently asked questions”, available at http://www.bromba.com/faq/biofaqe.htm#ROC, visited 21.05.2012.

5. Campbell, J., “Speaker recognition: a tutorial”, *Proceedings of IEEE*, 85, 9, pp. 1437-1462, September 1997.

6. CASIA iris image database, http://www.sinobiometrics.com (March, 2006).

7. Chellappa, R., Wilson, C.L., and Sirohey, S., “Human and machine recognition of faces: a survey” *Proceedings of the IEEE*, 83, 5, pp. 705-740, May 1995.

8. Chen, X., Flynn, P.J., Bowyer, K.W., “IR and visible light face recognition”, Computer Vision and Image Understanding, vol. 99. no. 3, pp. 332–358, September 2005.

9. Das, R., “Signature Recognition”, in Keesing Journal of Documents & Identity, issue 24, 2007.

10. Dass, S.C., Nandakumar, K., and Jain, A.K., “A principled approach to score level fusion in multimodal biometric systems”, In Fifth AVBPA, Rye Brook, pp. 1049–1058, July 2005.

11. Daugman, J., “Recognizing persons by their Iris patterns”, in *Biometrics: Personal Identification in a Networked Society*, A. K. Jain, R. Bolle, and S. Pankanti, Eds. Norwell, MA: Kluwer, pp. 103–121, 1999.

12. Daugman, J.G., “How iris recognition works”, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 21–30, January 2004.

13. Duda, R., Hart, P. and Stork, D., *Pattern Classification*, second ed. John Wiley & Sons, 2001.

14. Frischholz, R.W., and Dieckmann, U., “Bioid: A multimodal biometric identification system,” *IEEE Computer*, vol. 33, no. 2, pp. 64–68, 2000.

15. Frischholz, R.W., Boebel, F.G., and Spinnler, K.P., “Face recognition with the synergetic computer”, *Proc. Int’l Conf. Applied Synergetics and Synergetic Eng.*, Fraunhofer Gesellschaft für Integrierte Schaltungen, Erlangen, Germany, pp. 107-110, 1994.

16. Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A., *Robust Statistics: The Approach Based on Influence Functions*. New York: Wiley, 1986.

17. Han, J. and Bhanu, B., “Gait recognition by combining classifiers based on environmental contexts”, *Lecture Notes in Computer Science*, vol. 3546/2005, pp. 113-124, 2005.

18. Heo, J. Kong, S.G., Abidi, B.R., and Abidi, M.A., “Fusion of visual and thermal signatures with eyeglass removal for robust face recognition”, *Proc. Joint IEEE Workshop Object Tracking and Classification beyond the Visible Spectrum*, June 2004.

19. Ho, T.K., Hull, J.J., and Srihari, S.N., “Decision combination in multiple classifier systems,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 16, no. 1, pp. 66-75, January 1994.

20. Huber, P.J., *Robust Statistics*. New York: Wiley, February 1981.

21. “Human identification at a distance” available at http://www.equinoxsensors.com/products/HID.html, visited 3.6.2012.

22. “IAFIS - All about integrated automated fingerprint identification system”, available at http://www.policeone.com/police-products/investigation/articles/1802754, visited March 27, 2012.

23. Indovina, M., Uludag, U. Snelick, R., Mink, A., and Jain, A., “Multimodal biometric authentication methods: A COTS approach”, *Proc. MMUA 2003, Workshop Multimodal User Authentication*, pp. 99-106, December 2003.

24. ISO/IEC 24745:2011, “Information technology - Security techniques – biometric information protection”, 2011.

25. ISO/IEC TR 24722:2007, Information technology– Biometrics: Multimodal and other multibiometric fusion, July 2007.

26. Jain A., Nandakumar, K., Ross, A., “Score normalization in multimodal biometric systems”, *Pattern Recognition*, vol. 38, no. 12, pp. 2270–228, December 2005.

27. Jain, A., Hong, L. and Kulkarni, Y., “A multimodal biometric system using fingerprint, face and speech”, in: *Second Internat. Conf. on AVBPA*, Washington, DC, USA. pp. 182-187, 1999.

28. Jain, A., Hong, L., and Bolle, R., “On-line fingerprint verification”, *IEEE Trans. Pattern Anal. and Machine Intell*., 19, 4, pp. 302-314, 1997.

29. Jain, A.K., and Ross, A., “Fingerprint mosaicking”, in *Proc. Int’l Conf Acoustic Speech and Signal Processing*, vol. 4, pp. 4064-4067, 2002.

30. Jain, A.K., Nandakumar, K., and Ross, A., “Score normalization in multimodal biometric systems, ” *Pattern Recognition*, vol. 38, no. 12, pp. 2270-2285, December 2005.

31. Jain, A.K., Prabhakar, S., Chen, S. “Combining multiple matchers for a high security fingerprint verification system”, *Pattern Recognition Letters*, vol. 20, 11-13, pp 1371–1379, November 1999.

32. Jain, A.K., Ross, A. and Pankanti, S., “Biometrics: A tool for information security”, *IEEE Trans. Inf. Forensics Security*, vol. 1, no. 2, pp. 125–143, January 2006.

33. Jain, A.K., Ross, A., Prabhakar, S., “An introduction to biometric recognition”, in *IEEE Trans. Circuits Systems Video Technology,* vol.14, no. 4, pp. 4–20, January 2004.

34. Kittler, J., Hatef, M., Duin, R.P.W., and Mates, J. “On combining classifiers”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, 1998.

35. Kong, W.K., Zhang, D., Li, W. X., “Palmprint feature extraction using 2-D Gabor filters”, *Pattern Recognition 36*, pp. 2339-2347, 2003.

36. Kumar, A., Wong, D.C., Shen, H.C. and Jain, A.K., “Personal verification using palmprint and hand geometry biometric”, in *Proc. 4th Int.Conf. Audio-and Video-based Biometric Person Authentication*, Guildford, U.K., January 9–11, 2003.

37. Lanckriet, G.R.G., Ghaoui, L.El., Bhattacharyya, C., and Jordan, M.I., *J. Machine Learning Res.* 3, 552, 2002.

38. Ma, Y., Cukic, B., and Singh, H., “A classification approach to multibiometric score fusion,” *Proc. Fifth Int’l Conf. Audio Video-Based Biometric Person Authentication*, pp. 484-493, July 2005.

39. Marcialis, G.L., and Roli, F., “Fingerprint verification by decision-level fusion of optical and capacitive sensor”, *Lecture Notes in Computer Science*, vol. 3087/2004, pp. 307-317, 2004.

40. Martinez, A.M., Benavente, R., “The AR face database”, *CVC Technical Report No. 24*, June 1998.

41. Matsumoto, T., Matsumoto, H., Yamada, K. and Hoshino, S., “Impact of artificial gummy fingers on fingerprint systems”, *Proc. SPIE, Optical Security and Counterfeit Deterrence Techniques IV*, vol. 4677, pp. 275-289, January 2002.

42. Morrell, D., “Lecture Note 2: Neyman/Pearson decision theory”, *EEE 598C: Statistical Pattern Recognition*, September 3, 1996.

43. Poh, N., Bengio, S., and Korczak, J., “A multi-sample multi-source model for biometric authentication”, *Proc. IEEE Workshop Neural Networks for Signal Processing*, pp.375-384, 2002.

44. Prabhakar S., Jain, A.K., “Decision-level fusion in fingerprint verification”, *Pattern Recognition* 35, 4, pp. 861–874, 2002.

45. Prabhakar, S., Pankanti, S. and Jain, A.K., “Biometric recognition: security and privacy concerns,” *IEEE Security Privacy Mag.*, vol. 1, no. 2, pp. 33–42, 2003.

46. Proenca, H. and Alexandre, L.A., “UBIRIS iris image database”, http://iris.di.ubi.pt (December, 2006).

47. Pudil, P., Novovicova, J. and Kittler, J., “Floating search methods in feature selection,” *Pattern Recognition Letters*, vol. 15, no. 11, pp. 1119-1125, 1994.

48. Radha, N., and Kavitha, A., “Rank level fusion using fingerprint and iris biometrics”, *Indian Journal of Computer Science and Engineering,* vol. 2, no. 6, December 2011-January 2012.

49. Ratha, N.K., Connell, J.H., and Bolle, R.M., “Image mosaicing for rolled fingerprint construction”, in *Proc. Int. Conf. Pattern Recog.*, vol. 2, pp. 1651–1653, August 1998.

50. Ross A., and Govindarajan, R., “Feature level fusion using hand and face biometrics”, in *Proc. SPIE Conf. Biometric Technology for Human Identification II*, pp. 196–204, March 2005.

51. Ross, A. and Jain, A.K., “Fusion techniques in multibiometric systems”, in *Face Biometrics for Personal Identification*, R. Hammound, B. Abidi, and M. Abidi, Eds. Berlin, Germany: Springer, 2007.

52. Ross, A. and Jain, A.K., “Hand Geometry”, available at http://biometrics.cse.msu.edu/hand_geometry.html, visited 4.6.2012.

53. Ross, A., and Jain, A.K., “Information fusion in biometrics”, *Pattern Recognition Letters*, vol. 24, no. 13, pp. 2115-2125, 2003.

54. Ross, A., Nandakumar, K. and Jain, A.K., “Handbook of Multibiometrics”, *Springer-Science + Business Media, LLC*, 2006.

55. Sanderson, C., Paliwal, K.K., “Information fusion and person verification using speech and face information”, *Research Paper IDIAP-RR 02-33*, IDIAP, September 2002.

56. Singh, S., Gyaourova, A., Bebis, G. and Pavlidis, I., "Infrared and visible image fusion for face recognition", *SPIE Defense and Security Symposium*, pp.585-596, 2004.

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

58. Soh, J., Deravi , F . and Triglia, A., “Multibiometrics and data fusion standardization”, in *Encyclopedia of Biometrics*, S.Z. Li and A.K. Jain, New York, Heidelberg: Springer, 2009.

59. Son, B. and Lee, Y., “Biometric authentication system using reduced joint feature vector of iris and face”, in *Proc. 5th Int. Conf. Audio and Video-Based Biometric Person Authent.*, Rye Brook, NY, pp.513–522, 2005.

60. Son, B., and Lee, Y., “Biometric authentication system using reduced Joint feature vector of iris and face”, *Lecture Notes in Computer Science*, vol. 3546, pp. 261-273, 2005.

61. The National Institute of Standards and Technology (NIST), available at http://www.nist.gov/itl/iad/ig/biometricscores.cfm, visited 12.04.2012.

62. Trunk, G., “A problem of dimensionality: a simple example,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 1, no. 3, pp. 306-307, 1979.

63. Turk, M. and Pentland, A., “Eigenfaces for recognition”, *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991

64. Uhl A., and Wild, P., “Single-sensor multi-instance fingerprint and eigenfinger recognition using (weighted) score combination methods”, *International Journal of Biometrics,* vol. 1, no. 4, pp. 442-462, 2009.

65. Uludag, U. and Jain, A.K., “Attacks on biometric systems: a case study in fingerprints,” in *Proc. SPIE-EI Security, Steganography and Watermarking of Multimedia Contents VI*, San Jose, CA, pp. 622–633, January 2004.

66. Uludag, U., Ross, A., Jain, A.K., “Biometric template selection and update: a case study in fingerprints”, *Pattern Recognition*, 37, 7, pp. 1533-1542, 2004

67. Vercellis, C., “Business intelligence: data mining and optimization for decision making”, *Wiley Online Library*, 2009.

68. Verlinde, P. and Cholet, G., “Comparing decision fusion paradigms using k-NN based classifiers, decision trees and logistic regression in a multi-modal identity verification application,” in *Proc. Int. Conf. Audio and Video-Based Biometric Person Authentication (AVBPA)*, Washington, DC, pp. 188–193, March 1999.

69. Wang, F., Yao, X., and Han, J., “Improving iris recognition performance via multi-instance fusion at the score level”, *Chinese Optics Letters*, vol.6, no. 11, pp. 824-826, 2008.

70. Yao, Y.-F., Jing, X.-Y., Wong, H.-S., “Face and palmprint feature level fusion for single sample biometrics recognition”, *Neurocomputing*, vol. 70, pp. 1582–1586, March 2007.