CERIAS Tech Report 2008-22

ANALYSIS OF FINGERPRINT SENSOR INTEROPERABILITY ON SYSTEM PERFORMANCE

by Shimon K. Modi

Center for Education and Research in
Information Assurance and Security,
Purdue University, West Lafayette, IN 47907-2086

ANALYSIS OF FINGERPRINT SENSOR INTEROPERABILITY ON SYSTEM PERFORMANCE

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Shimon K. Modi

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

August 2008

Purdue University

West Lafayette, Indiana

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

AFIS - Automated Fingerprint Identification System

ANOVA - Analysis of Variance

FMR - False Match Rate

FNMR – False Non Match Rate

FTE – Failure to Enroll

H_0 – Null Hypothesis

H_1 – Alternate Hypothesis

ISO – International Organization for Standardization

M – Mathematical mean

NBIS – NIST Biometric Image Software

NFIQ – NIST Fingerprint Image Quality

NIST- National Institute of Standards and Technology

ROC – Receiver Operating Characteristic

SD – Mathematical standard deviation

WSQ – Wavelet Scalar Quantization

ABSTRACT

Modi, Shimon K. Ph.D., Purdue University, August, 2008. Analysis of fingerprint sensor interoperability on system performance. Major Professor: Stephen Elliott.

The increased use of fingerprint recognition systems has brought the issue of fingerprint sensor interoperability to the forefront. Fingerprint sensor interoperability refers to the process of matching fingerprints collected from different sensors. Variability in the fingerprint image is introduced due to the differences in acquisition technology and interaction with the sensor. The effect of sensor interoperability on performance of minutiae based matchers is examined in this dissertation. Fingerprints from 190 participants were collected on nine different fingerprint sensors which included optical, capacitive, and thermal acquisition technologies and touch, and swipe interaction types. The NBIS and VeriFinger 5.0 feature extractor and matcher were used. Along with fingerprints, characteristics like moisture content, oiliness, elasticity and temperature of the skin were also measured. A statistical analysis framework for testing interoperability was formulated for this dissertation, which included parametric and non-parametric tests. The statistical analysis framework tested similarity of minutiae count, image quality and similarity of performance between native and interoperable datasets. False non-match rate (FNMR) was used as the performance metric in this dissertation. Interoperability performance analysis was conducted on each sensor dataset and also by grouping datasets based on the acquisition technology and interaction type of the acquisition sensor. Similarity of minutiae count and image quality scores between two datasets was not an indicator of similarity of FNMR for their interoperable datasets.

Interoperable FNMR of 1.47% at fixed FMR of 0.1% was observed for the optical touch and capacitive touch groupings. The impact of removing low quality fingerprint images on the effect of interoperable FNMR was also examined. Although the absolute value of FNMR reduced for all the datasets, fewer interoperable datasets were found to be statistically similar to the native datasets. An image transformation method was also proposed to compensate for the differences in the fingerprint images between two datasets, and experiments conducted using this method showed significant reduction in interoperable FNMR using the transformed dataset.

CHAPTER 1. INTRODUCTION

1.1. Introduction

Authentication of individuals is a process that has been performed in one form or another since the beginning of recorded history. Evidence of attempts to authenticate individuals has been found dating back to 500 B.C (Ashbourn, 2000). The quest for improving authentication methodologies has intrigued humans since these early times. Establishing and maintaining the identity of individuals, and accurate automated recognition is becoming increasingly important in today's networked world. As technology advances, the complexity of these tasks has also increased. There are three main methods of authenticating an individual: 1) using something that only the authorized individual has knowledge of e.g. passwords 2) using something that only the authorized individual has possession of e.g. physical tokens 3) using physiological or behavioral characteristics that only the authorized individual can reproduce i.e. biometrics. The increasing use of information technology systems has created the concept of digital identities which can be used in any of these authentication mechanisms. Digital identities and electronic credentialing have changed the way authentication architectures are designed. Instead of stand-alone and monolithic authentication architectures of the past, today's networked world offers the advantage of distributed and federated authentication architectures. The development of distributed authentication architectures can be seen as an evolutionary step, but also raises the issue always accompanied by an attempt to mix disparate systems: Interoperability. What is the effect on performance of the authentication process if an individual establishes his/her credential on one system, and then authenticates him/her-self on a different system? This issue is of relevance to all kinds of authentication mechanisms, and of particular relevance to biometric recognition systems. The last decade has witnessed a huge increase in deployment of biometric systems, and while most of these systems have been single vendor, monolithic architectures the issue of interoperability is bound to arise as distributed architectures become more pervasive.

Fingerprint recognition systems are the most widely deployed biometric systems, and most commercially deployed fingerprint systems are single vendor systems. A recent market report published by International Biometric Group (IBG) (2007) found that fingerprint recognition systems account for approximately 68% of the biometric system deployments. The single point of interaction of a user with the fingerprint system is during the acquisition stage, which has the maximum probability of introducing inconsistencies in the biometric data. Fingerprint sensors are based on a variety of different technologies like electrical, optical, thermal etc. The physics behind these technologies introduces inconsistent distortions and variations in the feature set of the captured image, which makes the goal of interoperability even more challenging. There are obvious technical, commercial and operational advantages to understanding these issues more in-depth, in addition to making the objective of creating a globally interoperable biometric system a more realistic one.

This dissertation is organized into five chapters. The first chapter discusses the statement of the problem, the significance of the problem being addressed, and rationale behind this research. The second chapter covers background material and review of literature available in related areas. Selected works in the area of fingerprint recognition, which are relevant to this dissertation, are summarized. The third chapter outlines the methodology of research, the experimental setup and the analysis framework to be followed during this study. The fourth chapter contains the results and analysis of the procedures described in the third chapter. The fifth chapter contains conclusions from the results analysis and recommendations for future work.

1.2. Statement of Problem

The distortions and variations introduced when acquiring fingerprint images propagate from the acquisition subsystem all the way to the matching subsystem. These variations ultimately affect performance rates of the overall fingerprint recognition system. Fingerprint images captured using the same sensor technology during enrollment and recognition phases will introduce similar distortions, thus making it easier to compensate for such distortions and reducing its effect on the performance of the overall fingerprint recognition system. However, when different fingerprint sensor technologies are used during enrollment and recognition phases, an impact on performance is expected, but is unpredictable.

Financial institutions provide a relevant example of the requirement for interoperability of fingerprint sensors. Some institutions are starting to deploy Automated Teller Machines (ATM) which use fingerprint recognition for authenticating customers. Such a system can be designed to take advantage of distributed acquisition architecture and use a centralized storage and matching architecture as shown in Figure 1.1. Without proper understanding of how fingerprints captured from different sensors affect the overall recognition rates, a financial institution would be forced to deploy the same fingerprint sensor at all of their ATM’s. An extraordinary level of confidence and trust in the fingerprint sensor manufacturer would be required in order to choose just a single manufacturer. The lack of choice could also be a hurdle to mass adoption of this technology. If the sensor manufacturer was to stop supporting the particular fingerprint sensor, a financial institution would be forced to replace all the sensors and re-enroll all its clients. A financial institution would incur massive capital and labor costs, which could be a deterrent to using this technology. There is need to understand the effect of different fingerprints on recognition rates not just from an algorithm advancement perspective, but also from a technology usage perspective.

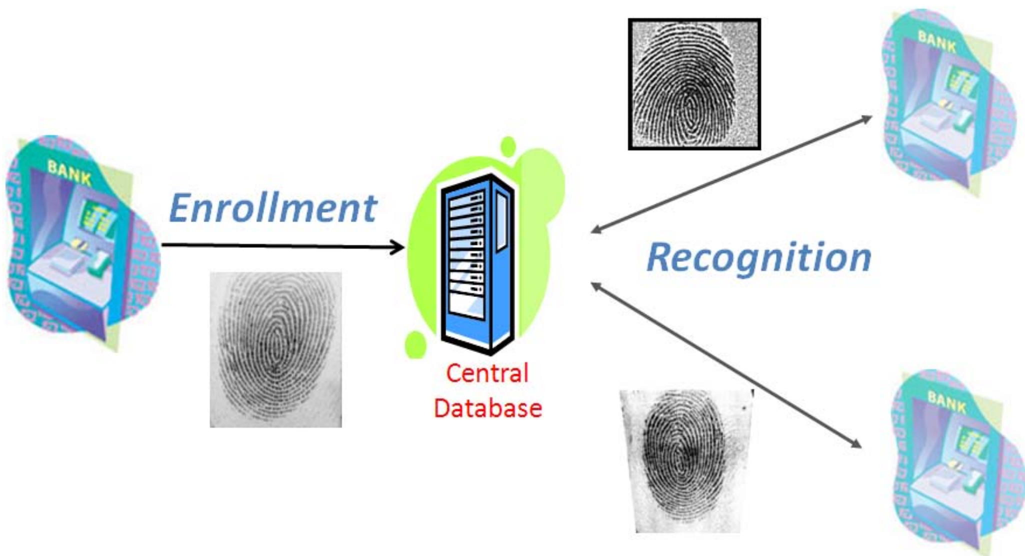

Figure 1.1 Example of an Interoperability Scenario.

1.3. Significance of Problem

One of the main components of minimizing error rates in a fingerprint recognition system is the acquisition subsystem since it is the first point of contact between the user and the system. Sensor variability results in moving the distribution of genuine scores away from the origin, thereby increasing error rates and negatively impacting the performance of the system (Wayman, 1997). Bolle and Ratha (1998) describe a list of challenges of matching two fingerprints. Existence of spurious features and missing features compared to the database template, transformation or rotation of features, and elastic deformation of features are several problems faced by fingerprint matchers. The acquisition subsystem is responsible for introducing part or all of the variability. This issue is amplified when a fingerprint recognition system is deployed in a distributed architecture because of the use of different components during acquisition, signal processing, and matching operations. A fingerprint recognition system deployed in a distributed architecture can benefit from gaining a deeper insight into interoperability of sensors and its effect on error rates.

Jain and Ross (2004) observed that Equal Error Rates (EER) for a fingerprint dataset consisting of images acquired from two different sensors was 23.10% whereas the EER for the two native datasets were 6.14% and 10.39%. In a report submitted to U.S. Department of Agriculture, Ford and Sticha (1999) outlined feasibility of biometric systems in reducing fraud in the food stamp program and other welfare programs, and paid particular attention to fingerprint recognition. In 2001 eight states in the U.S.A. required applicants for welfare in at-least some counties to submit to fingerprint recognition (Dean, Salstrom, & Wayman, 2001). These previously mentioned deployments are distributed architectures which acquire fingerprints from various locations and have a higher degree of centralization for its matching and storage functionalities. The ability of a recognition system to integrate different fingerprint acquisition technologies without a significant degradation of error rates has definite technological and financial advantages. It not only allows the owners of the deployment to pick and choose sensors that best fit its application, but also have the ability to switch a few of the sensors without overhauling the entire deployment, or re-enrolling its entire user-base. Campbell and Madden (2006) concluded that sensor effects on performance are not trivial, and need to be investigated independently. A deeper understanding of the effect of matching fingerprints acquired from different sensors is imperative for mass adoption of this technology.

IBG (2007) has provided an analysis of adoption of biometric technologies and applications, and a forecast of global revenues from 2007-2012 in their report. They project annual revenues to grow from $3,012 million USD to approximately

$7,407 million USD. Commercial fingerprint recognition systems currently make up approximately 25% of the entire biometrics market, and Automated Fingerprint Identification Systems (AFIS) and other civilian fingerprint systems make up approximately 33% of the entire biometrics market. Revenue generated from the biometric market is projected to double in the next five years, and fingerprint recognition systems are poised to generate a majority of it. Several industry support and government enforced initiatives are expected to be prime growth drivers for the biometrics industry but the issue of interoperability will have to be tackled for a realization of this growth potential. An ever increasing networked world is going to necessitate use of distributed architectures, and the ability to integrate different fingerprint sensors in a heterogeneous system without a significant impact on performance will be of critical importance.

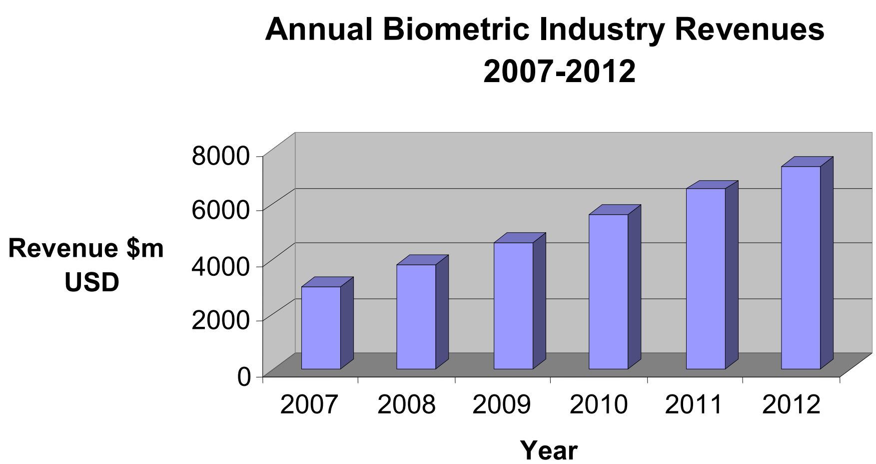

Figure 1.2 Annual Industry Revenue Projections 2007-2012 (IBG, 2007).

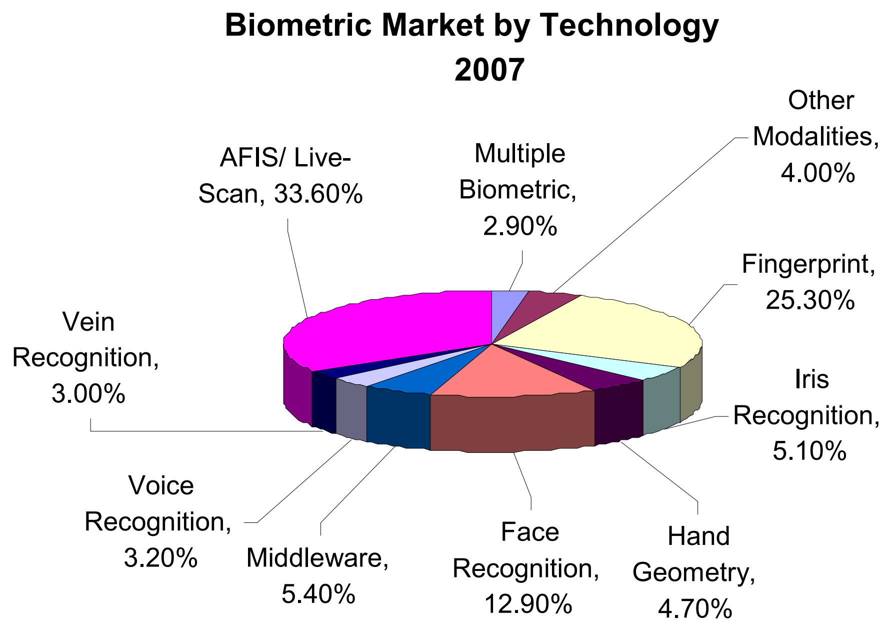

Figure 1.3 Biometric Market Share by Technology (IBG, 2007).

1.4. Purpose of Study

Fingerprint recognition systems are primarily based on two different types of matchers: minutiae based matchers, and pattern based matchers. The purpose of this study was to examine the effect of sensor dependent variations and distortions, and characteristics of the sensor on the interoperability matching error rates of minutiae based fingerprint recognition systems. This study achieved this aim by acquiring fingerprints from different acquisition technologies, and examining error rates for native and interoperable fingerprint datasets. This study did not attempt to isolate or examine specific variations and distortions in different fingerprint acquisition technologies. The ultimate effect on matching error rates of fingerprint datasets because of distortions introduced by different technologies was the focal point of analysis. The outcome of such a study will be useful in designing matching algorithms which are better at handling fingerprint images from various sensors. The problem space affecting error rates of fingerprint datasets is vast – every subsystem of a fingerprint recognition system has an influence on performance of a fingerprint dataset. This study examined an exclusive aspect of the problem space - the acquisition subsystem. The end objective of this study was to provide greater insight into the effect of a fingerprint dataset acquired from various sensors on performance measured in terms of false non match rates (FNMR) and false match rates (FMR).

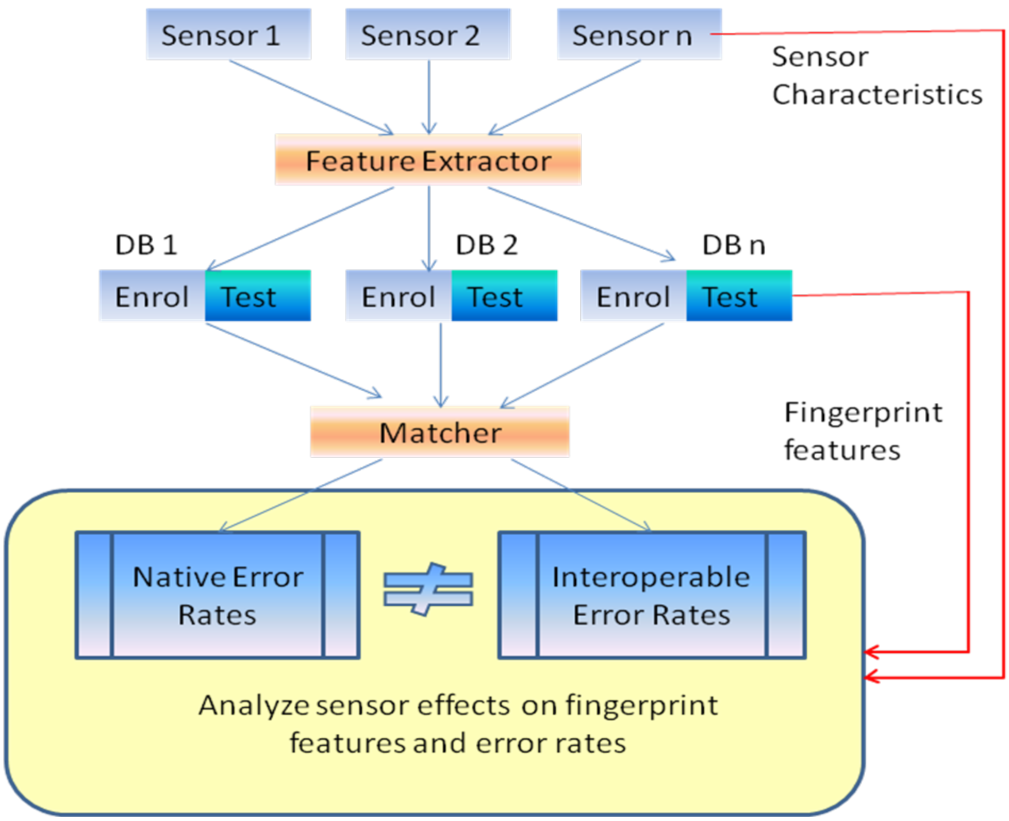

Figure 1.4 Purpose of Study.

1.5. Definitions

Biometric Algorithm represents the finite number of steps used by a biometric engine to compute whether a biometric sample and template is a match (M1, 2004).

Biometrics is defined as automated recognition of individuals based on their behavioral and biological characteristics (*ISO/IEC JTC1 SC37 SD2 - Harmonized Biometric Vocabulary*, 2006).

Biometric Sample is the raw data representing a biometric characteristic of an end-user as captured and processed by a biometric system (*ISO/IEC JTC1 SC37 SD2 - Harmonized Biometric Vocabulary*, 2006).

Biometric System is an automated system capable of the following (*ISO/IEC JTC1 SC37 SD2 - Harmonized Biometric Vocabulary*, 2006):

1. capturing a biometric sample from the end user
2. extracting biometric data from the sample
3. comparing the biometric data to one or more reference templates
4. computing how well they match
5. indicating based on decision policy if identification or verification has been achieved

Core is the turning point on the inner most ridge of a fingerprint (Amin & Yager, 2004).

Delta is the point on a ridge at or nearest to the point of divergence of two ridge lines, and located at or directly in front of point of divergence (M1, 2004).

Equal Error Rate (EER) is the operational point where FNMR=FMR

Enrollment is the process of converting a captured biometric sample and the subsequent preparation and storage of the biometric template representing the individual’s identity (M1, 2004).

Environment is defined as the physical surroundings and conditions where biometric capture occurs, and it also includes operational factors such as operator skill and enrollee cooperation level (*Biometric Sample Quality Standard - Part 1: Framework*, 2006).

False Accept Rate (FAR) is the expected number of transactions with wrongful claims of identity that are incorrectly confirmed (Mansfield & Wayman, 2002).

False Reject Rate (FRR) is the expected number of transactions with truthful claims of identity that are wrongly denied (Mansfield & Wayman, 2002).

False Match Rate (FMR) is the expected probability that a sample will be falsely declared to match a single randomly selected non genuine template (Mansfield & Wayman, 2002).

False Non Match Rate (FNMR) is the expected probability that a sample will be falsely declared not to match a template from the same user supplying the sample (Mansfield & Wayman, 2002).

Failure to Acquire (FTA) Rate is the expected rate of sample acquisitions which cannot be processed.

Failure to Enroll (FTE) Rate is the expected proportion of subjects who cannot be enrolled in the system.

Friction Ridge is the ridges present on skin of the fingers which make contact with an incident surface under normal touch (M1, 2004).

Hybrid testing describes a methodology in which samples are collected in real time and testing of the samples is performed in an off-line environment (Grother, 2006).

Identification is the one-to-many process of comparing a submitted biometric sample against all of the biometric reference templates on file to determine whether it matches any of the templates (M1, 2004).

Interoperable fingerprint dataset refers to the fingerprint dataset of 3 enrollment fingerprint images and 3 testing fingerprint images collected from two different sensors.

Live capture is the process of capturing a biometric sample by an interaction between an end user and a biometric system (IS*O/IEC JTC1 SC37 SD2 - Harmonized Biometric Vocabulary*, 2006).

Native fingerprint dataset refers to the fingerprint dataset comprised of 3 enrollment fingerprint images and 3 test fingerprint images collected from the same sensor.

Performance is measured in terms of false non match rates and false match rates which are fundamental error rates in offline testing (Grother et al., 2006).

Receiver Operating Characteristic (ROC) curves are a means of representing results of performance of diagnostic, detection and pattern matching systems (Mansfield & Wayman, 2002).

Template is the data which represents the biometric measurement of an enrollee and is used by the biometric system for comparison against subsequently submitted samples (*ISO/IEC JTC1 SC37 SD2 - Harmonized Biometric Vocabulary*, 2006).

Valley is the area surrounding a friction ridge which does not make contact with incident surface under normal touch (M1, 2004).

Verification is the process of comparing a submitted biometric sample against the biometric reference template of a single enrollee whose identity is being claimed, to determine whether is matches the enrollee's template (M1, 2004).

1.5.1. Sensor Related Distortions and Variations

Fingerprint sensors are responsible for capturing the unprocessed representation of ridge flow from the finger skin. Invariably, the capture process introduces undesirable changes which make it different from the source of the sample. These undesirable changes are called distortions. The causes of fingerprint image distortions and variations can be categorized into the following:

1. Interaction type
2. Acquisition technology type
3. Sensor characteristics

The two most popular modes of interacting with fingerprint sensors are touch and swipe. All sensors used in this study were one of these two types of interactions. Both of these interaction types have a distinct effect on the image. Swipe sensors utilize an image reconstruction technique which takes into account nonlinear distortions between two consecutive frames (Zhang, Yang, & Wu, 2006). Touch sensors are affected by elasticity of skin and amount of pressure placed on the finger when it is placed on the sensor. Swipe sensors are not affected by residue and latent prints, which can affect touch sensors. The artifacts introduced by latent prints can distort the image captured from a touch sensor.

There are several different types of acquisition technologies used in fingerprint sensors: optical, capacitive, thermal, ultrasonic, piezoelectric etc. The acquisition technology introduces distortion because of the physics behind it. The three types of acquisition technologies used in this study were optical, capacitive, and thermal.

Most optical sensors are based on the phenomenon of frustrated total internal reflection (FTIR) (O'Gorman & Xia, 2001). This technology uses a glass platen, a light source and a Charged Coupled Device (CCD) or Complementary Metal Oxide Semiconductor (S3OS) camera for constructing fingerprint images (O'Gorman & Xia, 2001). Optical sensors introduce distortions which are characteristic of its technology. The edges of fingerprint images captured using optical sensors have a tendency of getting blurred due to the setup of the lenses. Optical physics could potentially cause out of focus images which can be attributed to the curvature of the lens. Sometimes residual incident light is reflected from the ridges which can lead to a low contrast image (Secugen). A phenomenon called Trapezoidal Distortion is also noticed in fingerprint images captured from optical sensors due to the unequal optical paths between each point of the fingerprint and the image focusing lens (Igaki, Eguchi, Yamagishi, Ikeda, & Inagaki, 1992). The level of contrast in resulting fingerprint images is affected by the acquisition technology. Optical sensors tend to introduce grey areas in the image due to residual light getting scattered from the ridge and not reflecting completely.

Capacitance sensors do not produce geometric distortions, but they are prone to introduce distortions due to the electrical nature of the capture technology. Electrostatic discharge can affect the resulting image since the conductive plates are sensitive to it. Capacitance sensors can also be affected from the 60Hz power line and electrical noise from within the sensor (2006).

Thermal sensors also do not introduce geometric distortions. They work on the principle of measuring difference in heat flux, which makes the sensor susceptible to humid and warm conditions. Since this type of technology measure the difference in heat flux, the finger surface has to be swiped across the sensor and introduces distortions typical of swipe technology.

Sensor characteristics like sensing surface and resolution are also responsible for introducing distortions. The horizontal and vertical resolutions are responsible for determining the distance observed between pairs of minutiae points. Sensors with different resolutions will provide images which show a different Euclidean distance for the same pair of minutiae points. The area of the sensing surface will determine the level of overlap in the same fingerprint image captured on a different sensor. It is more likely for two different sensors with a similar sensing areas to capture the same fingerprint image with a higher level of overlap compared to sensors with different sensing area.

1.6. Assumptions

- All participants were willing participants in the study, and thus were treated as cooperative participants and provided samples in genuine and good faith.
- The integrity of fingerprint sensors was maintained throughout the study.
- The fingerprint sensors, minutiae extractors and minutiae matchers were either open source or commercially available. Their integrity and performance were assumed to be consistent with their product specifications.

1.7. Delimitations

- The participants in the study were recruited from Purdue University.
- All data collection was performed in a single visit.
- The participants in the study were allowed to adjust the height of the chair and positioning of the fingerprint sensor to suit their comfort needs.
- Mistakes in placement of finger on sensor invalidated the fingerprint sample provided by the participant.

- All fingerprint sensors were peripheral devices connected using USB port.
- An ultrasonic fingerprint sensor was not used.
- Environment conditions were controlled throughout the data collection process.
- Not all participants were familiar with fingerprint recognition technology. All participants had to complete a practice session to familiarize themselves with the process of providing fingerprint images.
- Two different fingerprint feature extractors and fingerprint feature matchers were used: VeriFinger 5.0 extractor and matcher, and MINDTCT and BOZORTH3. VeriFinger 5.0 extractor and matcher create one subsystem and MINDTCT and BOZORTH3 create another subsystem. This study did not compare interoperability of the different subsystems.

1.8. Limitations

- The experience of participants with fingerprint sensors participating in the study was different. Previous studies have observed that prior experience and habituation of users with fingerprint devices has an effect on performance rates (Thieme, 2003).

1.9. Summary

This chapter introduced the problem in global terms by giving a description of the problem space and its significance to overall advancement of fingerprint recognition. An overall frame of reference is created for the reader which will assist in highlighting the focus areas of this study.

CHAPTER 2. REVIEW OF LITERATURE

2.1. Introduction

This chapter outlined several studies and experiments related to the impact of sensor specific issues on fingerprint recognition systems. Previous studies were analyzed to identify and justify the design of experiment. The statistical analysis methodology used in this dissertation was formulated based on previous research analysis methodologies. This chapter also presented a historical and evolutionary view of fingerprint recognition systems.

2.2. History of Biometrics

Archaeological evidence has been found indicating fingerprints were used as a form of authentication dating back to the 7000 to 6000 B.C. (O'Gorman, 1999). Clay pottery from these times used a fingerprint impression to mark the identity of the potter who had made it. Bricks used in the ancient city of Jericho have been discovered with impression of fingerprints, and its most probable use was to recognize the mason. Likewise, in ancient Egypt allowance claimants were identified by checking against a record which contained the name, age, place of origin, and other relatively unique physical and behavioral characteristics (Ashbourn, 2000).

2.2.1. Studies Related to Finger Anatomy and Ridge Structure

The earliest scientific fingerprint studies aimed at studying the anatomy and function of papillary ridges on the surface of the finger. Two microscopists, Govard Bidloo and Macello Malphighi published their findings and conclusions on papillary ridges in 1685 and 1687 respectively (Cole, 2001). The main point of discussion in those publications was if papillary ridges were organs of touch or if they facilitated sweating. Anatomist J.C. Mayer (1788) made a claim in his publication that the arrangement of skin ridges could never duplicated in two persons. This was among the first recorded claim about uniqueness of fingerprints. In 1823 Czech physician Jan Purkyne classified papillary ridges on human fingerprints into nine categories in an effort to establish a connection between vision and touch (Hutchings, 1997). This was the first attempt at classifying fingerprint patterns and laid the foundation for future fingerprint identification systems. Although Purkyne was a trained physician, he was also a student of philosophy and believed that every natural object is identical to itself and thus claimed that no two fingerprint patterns are identical.

2.2.2. Early Experiments with Fingerprint Identification: Criminalistics

William Herschel and Henry Faulds were the first to apply fingerprints for verifying the claim of an individual's identity by checking it against a catalogue of recorded fingerprints. In 1858 Willliam Herschel asked a road contractor to impress his hanS4rint in ink on a deed with the intention of adding the element of non-repudiation to the contract. Upon further investigation of the details of the hanS4rint he believed that he had found a way of verifying the identity of every man, and tried to institute it as an identification method in a local prison to identify the inmates, but his ideas did not impress his superiors and his ideas were never tested. The impracticality of comparing a visual representation of a fingerprint with all the catalogued fingerprints was its primary drawback. In 1880 Henry Faulds observed that fingerprint patterns could be used to identify criminals. Faulds devised an alphabet classification scheme so that a person's set of ten prints could be represented by a word. These words could be catalogued in a dictionary which would then be used to search for fingerprints. Faulds published his observations and methodology in a letter to *Nature* in 1880 and also approached Scotland Yard with his ideas. Although he provided an improvement over previous methods with the use of a cataloguing system, Scotland Yard declined to follow up on his technique.

2.2.3. Galton’s Fingerprint Experiments

In 1880 Francis Galton followed up on Fauld’s experiment on using fingerprints to identify individuals. Galton started off by devising a classification system which built on Purkyne’s 9 pattern types, but he found that they were not discriminative enough. There were fingerprints with patterns which blurred the division between the 9 types. Eventually Galton came up with 60 different patterns which he believed to be discriminative and inclusive but then realized the impracticality of such a system. He reconsidered his efforts and reclassified all patterns into three categories: “arches”, “loops” and “whorls”. Galton recognized the key to fingerprint classification lay in grouping, not in differentiating. The empirical results from Galton’s experiments showed that the patterns were not uniformly distributed. The loop pattern was the most common pattern, so Galton divided it into inner loop pattern and outer loop pattern. According to Galton’s classification inner loops opened towards the thumb and outer loops opened towards the little finger. For purposes of classification Galton suggested classifying all ten fingers based on the four pattern types and using that combination of patterns to catalogue the individual. Galton demonstrated his classification scheme to the British Home Office in 1893 but non uniform distribution of fingerprint patterns posed a challenge of representing large collections, which dissuaded them from adopting Galton’s scheme.

In the course of his experiments he noticed features of ridges where the papillary ridge ended, split into two, or rejoined from two ridges into one. He went on to propose that two fingerprints could be matched by comparing these features that he called “minutiae”. Although theoretically this method had merit, Galton believed this method to be infeasible in practical terms. The number of minutiae that needed to match correctly in order to make a confident claim of a match was a stumbling block for Galton's minutiae matching method. This discovery would not be used to its fullest strength until advent of automated minutiae matchers in 1960's.

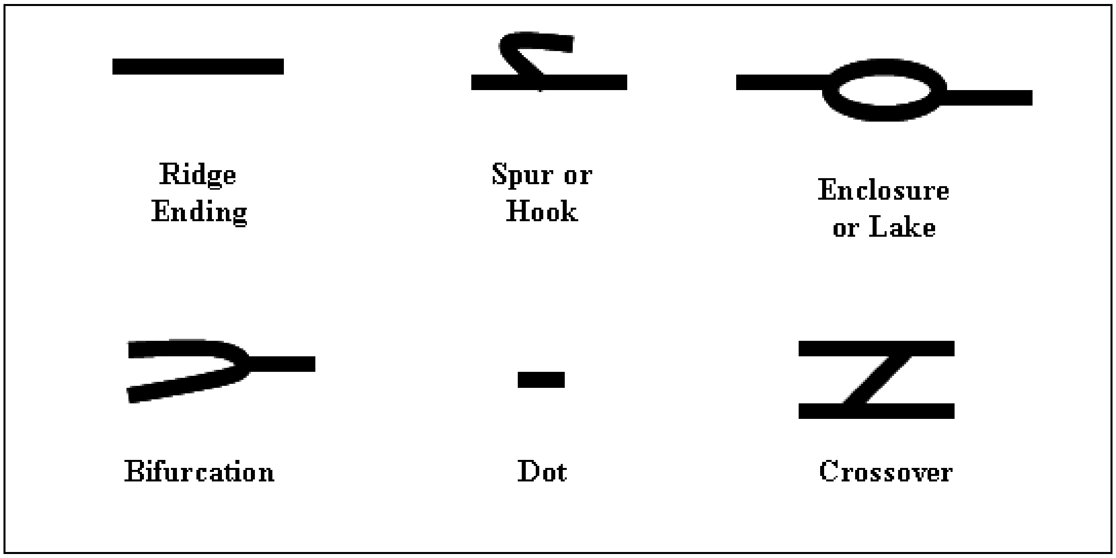

Figure 2.1 Examples of Galton features.

2.2.4. Henry's Fingerprint Experiments

Edward Henry was a colonial police officer in India interested in using fingerprints to identify criminals since he believed the Bertillon system based on anthropometrics was inadequate (Cole, 2004). Henry and his assistants set about creating a feasible and practical fingerprint classification scheme and devised a solution using ridge counting and ridge tracing. They subdivided loops by counting the number of ridges between the delta and core. They subdivided whorls by tracing the ridge from the delta and determining whether it passed inside, outside or met the core. They also added a fourth group called composites to the original three classification groups. The concept of ridge tracing and ridge counting was crucial in making it workable. Henry and his assistants, Haque and Bose, built upon the concept of ridge tracing and ridge counting and created a heuristic based fingerprint classification methodology which was capable of accommodating up to 100,000 prints (Haylock, 1979).

Henry introduced this system in his jurisdiction in 1895 and by 1897 it was adopted all over India. Henry published his system in a publication called “The Classification and Uses of Finger Prints” in 1900 and by 1902 Scotland Yard had fully adopted fingerprinting for purposes of identification.

2.2.5. Fingerprint Recognition in South America, Europe

Parallel development of fingerprint identification and verification techniques was being pursued in Argentina and France in late 19th century and early 20th century. Juan Vucetich published “Dactiloscopia Comprada” in 1904 which is one of the earliest notable works on fingerprinting (Polson, 1951). Vucetich came across Galton’s lecture on “Patterns in Thumb and Finger Marks” which motivated him to apply fingerprinting to the problem of reliable identification. Vucetich developed a 10 digit system based on Galton’s classification. He used the four basic pattern types of arch, left loop, right loop and whorl. In addition, he used a secondary classification of five sub-types: loop with plain pattern, loop with adhering ridges, internal loop approximating a central pocket, external loop approximating a central pocket, and irregular loops that did not include any of the previous types (Cole, 2004). In 1905 an autonomous finger print bureau was created in Argentina based on Vucetich’s fingerprint identification system.

In 1891 in France, Forgeot described a technique of development of latent fingerprints based on previous work on micro-chemical tests for analysis of sweat. Locard of Lyons is credited with inventing poroscopy when he demonstrated that pores of skin are relatively unique in shape, size and position in each individual (Polson, 1951). Locard applied his method in the Boudet-Simonin case which led to a conviction of burglary. Bertillon, who is considered widely to be the inventor of anthropometry, also used bloody fingerprints in case of Henri Scheffer to prove the identity of the murderer.

2.3. Automated Fingerprint Recognition

Attempts to automate the process of identification can be found dating back to 1920. IBM punch card sorters and other automated data processing technologies were employed to alleviate the problem of handling large number of cards. In 1919 the California State Bureau of Identification introduced a mechanized punch card system called the Robinson F-index to assist in storing and retrieval of information. In 1934 the Federal Bureau of Investigation (FBI) began using an IBM card sorter to search coded fingerprint classifications, and in 1937 the New York State Division of Criminal Identification started using an IBM card sorter which could sort 420 cards per minute (Cole, 2001). These approaches solved part of the problem, but human examiners were still required to inspect fingerprints. The need to automate the entire matching process was widely acknowledged, but technology constraints remained a barrier to progress. The first experiments with optical recognition of fingerprint images began in the 1960’s (Allen, Prabhakar, & Sankar, 2005). In 1963 Joseph Wegstein and Raymond Moore began work on an automated fingerprint identification system under the auspices of FBI and National Bureau of Standards (Reed, 1981). In 1972 the FBI installed a system with a fingerprint scanner built by Cornell Aeronautical Laboratory and a prototype fingerprint reader system built by North American Aviation (Cole, 2001). With considerable progress achieved on digitizing and automating the matching process, the FBI started scanning all their fingerprint records for persons born after January 1, 1929 and by 1980 they had a databank of 14.3 million records. Throughout the 1980’s various city police departments and state criminal justice bureaus in the U.S.A. started deploying Automated Fingerprint Identification Systems (AFIS) confirming the maturity of the technology.

2.4. General Biometric Recognition System

A generic biometric authentication system can be divided into five subsystems based on its functionality: data collection, transmission, signal processing, decision and data storage (A. Jain, Maltoni, Maio, & Wayman, 2005). Wayman (1997) described a general biometric model in “A Generalized Biometric Identification System Model” The data collection subsystem samples the raw biometric representation and the sensor converts it into an electronic representation. The transmission subsystem is responsible for transportation of the electronic representation of the biometric to the signal processing subsystem. Data compression and adjustment for noise are the main tasks of this subsystem. The signal processing subsystem takes the biometric signal and converts it into a feature vector. The tasks of this subsystem differ according to the modality being processed, but quality assessment and feature extraction is performed by this subsystem. The storage subsystem is responsible for storing the enrolled templates, and depending on the policy governing a particular biometric system the raw signal might also be stored. Binning and classification of templates according to predefined criteria are also employed within this subsystem in order to increase efficiency of the matching process. Inputs to the decision subsystem are measures resulting from comparison between the feature vector of the input sample and the enrolled templates. Using the system decision policy, a match or a non-match is declared.

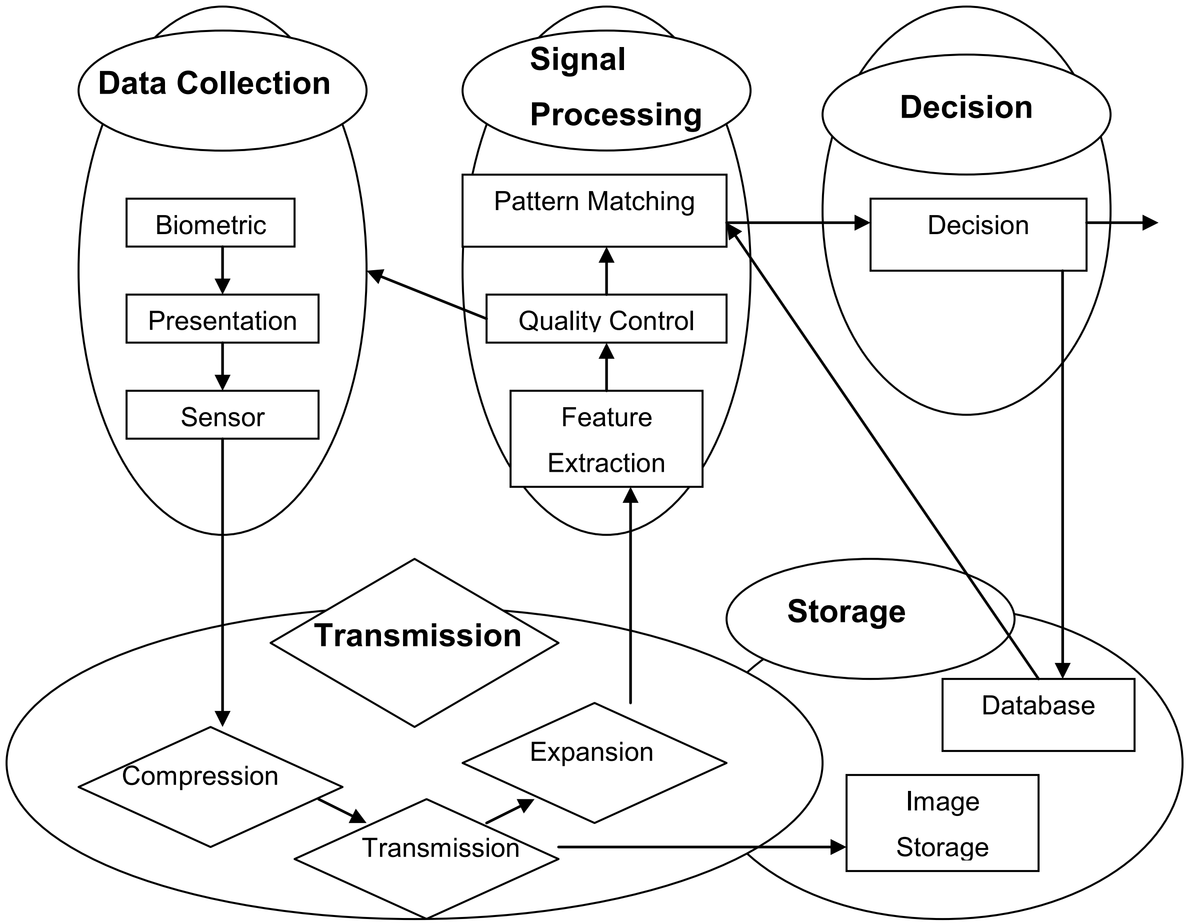

Figure 2.2 The General Biometric System (Wayman, 1997)

R. M. Bolle, Connell, Pankanti, Ratha and Senior (2004b) described a biometric authentication system as a pattern recognition system comprised of biometric readers, feature extractors, and feature matchers. They categorized the authentication system based on the processes of enrollment and authentication. In their model, the enrollment subsystem captures the biometric measurements, extracts the relevant features and then stores them in a database. Depending on system policy, ancillary information like identification number can also be stored with the biometric template. The authentication subsystem is concerned with recognizing a individual once he/she has been enrolled in the biometric system.

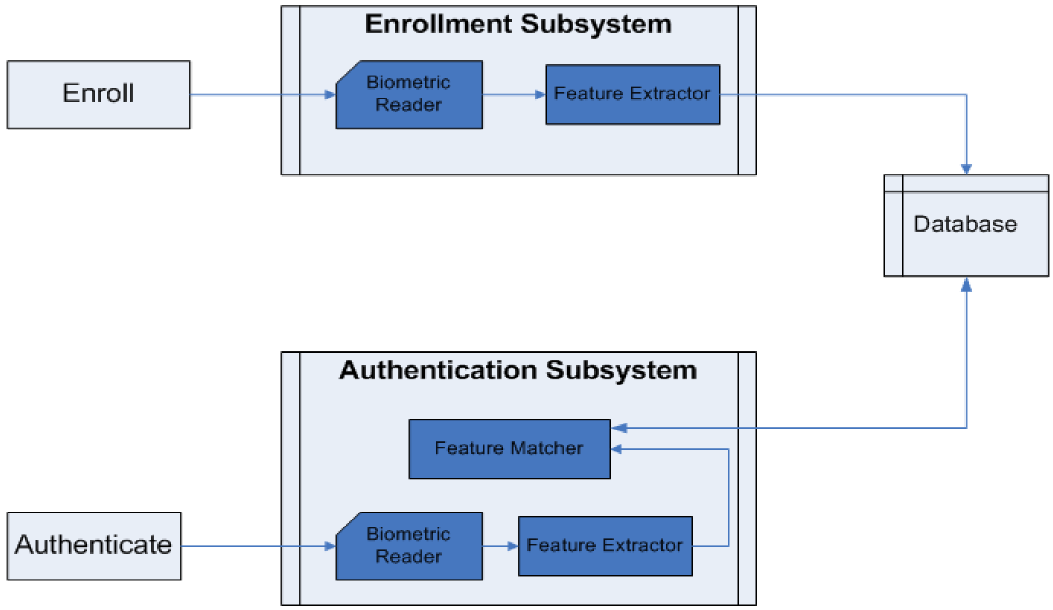

Figure 2.3 Biometric Authentication System (R. M. Bolle, Connell, Pankanti, Ratha, & Senior, 2004b)

Wayman et al. (2005) use the functionality of each subsystem, and process flow as a means of describing a general biometric model, and Bolle et al. (2004) use the different stages of the authentication process as a means of describing a general biometric model. The differences in description of the two models highlighted the possibilities for the underlying system architectures for biometric systems.

A detailed model was developed within ISO/IEC JTC1 SC37 which described the general biometric system in terms of the functional components of the system and process flow in the system. A notable deviation from the previously described biometric system model was removal of the transmission subsystem as a specific part of the biometric model. In a distributed and networked authentication system, transmission is a general function of the overall system instead of a function specific to the biometric system.

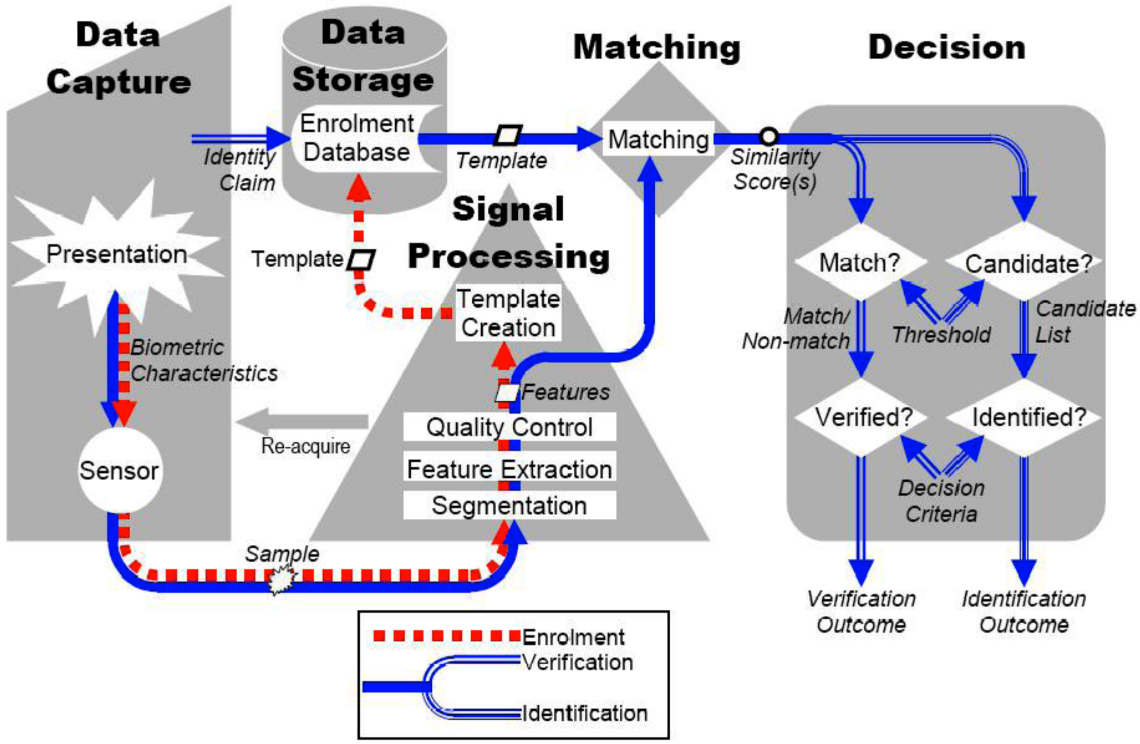

Figure 2.4 ISO/IEC JTC1 SC37 Conceptual Biometric Model (ISO)

2.4.1. Fingerprint Recognition Model

A distributed authentication architecture is an fundamental reason for focusing on the issues of interoperability. A distributed fingerprint recognition system can be designed in a multitude of ways, and a list of possible architectural configurations can be considered based on acquisition location of fingerprint sample, storage location of the fingerprint template, and matching location of input sample and fingerprint template. Table 2.1 shows the four possible processing locations.

The server is defined as a centrally located computer which is at a different physical location than the requesting client. The local workstation is where a user initiates interaction with the fingerprint recognition system. The peripheral device can also be connected with the local workstation using input/output ports, or an embedded device. Physical token refers to smartcards, PS3CIA cards and other small scale devices that could support any of the required processes. Figure 2.5 shows a list of all possible architectural configurations. There are $4^3 = 64$ configurations possible, but not all of them are feasible in a practical implementation. The purpose of depicting these possible configurations is to give an idea about the complexity of designing a distributed architecture.

Table 2.1 Possible Acquisition/Storage/Matching Location

Server
Local Workstation
Peripheral Device
Physical Token

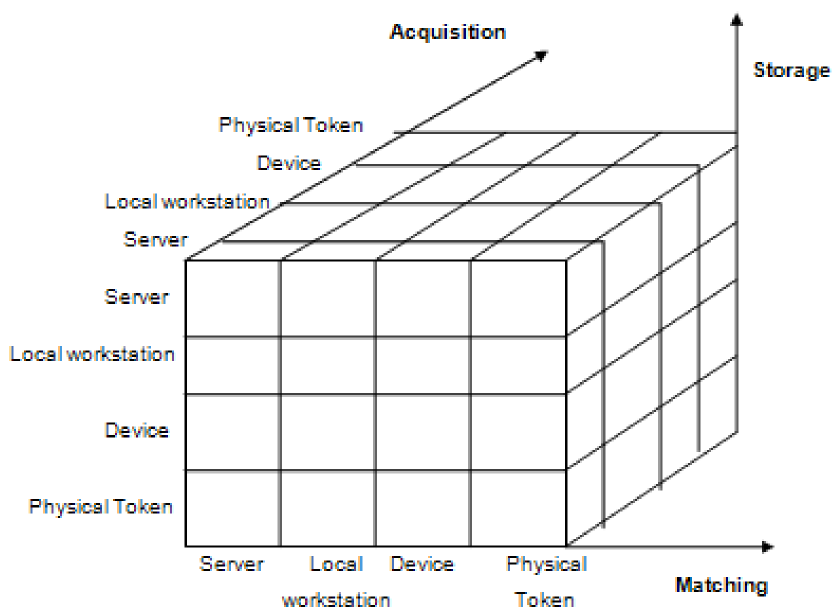

Figure 2.5 Possible Distributed Architectures

Several live implementations use one of the possible configurations as their distributed architecture design. Digital Persona U.are.U Pro for Active Directory replaces passwords with fingerprint recognition for authentication used in enterprise network applications (DigitalPersona, 2006). This product allows organizations to centralize the network authority processes and is an example of distributed acquisition and centralized storage/matching architecture. The Clear Registered Traveler program provides registered members with a smart card that stores fingerprint, face, and iris templates for card’s owner ("Clear Registered Traveler", 2006). For verification purposes the registered individual steps up to a kiosk, provides their fingerprint, face, and iris samples which are matched with the templates on the card. In this scenario, the matching happens on the machine in the kiosk, and this is an example of architecture that uses a physical token to store, and a local machine to perform the matching operation. In this architecture, the storage and matching operations are performed on separate devices. The increasing use of distributed architectures brings into focus the issues of interoperability. The following section describes the history and development of acquisition technologies used in fingerprint recognition systems.

2.4.2. Fingerprint Acquisition Technologies

The inked method of fingerprint acquisition was the preferred means of data capture from the 1870’s when fingerprints were first considered for law enforcement purposes until the 1960’s when the first generation AFIS was developed. Advances in storage of digital information and automated recognition of fingerprints led to innovations in fingerprint acquisition technologies. Rapid advances were made in this field and the following is a brief listing of the family of imaging technologies:

- Optical
- Electrical
- Thermal
- Ultrasonic
- Piezoelectric

Optical, electrical and thermal are the most popular types of fingerprint acquisition and imaging technologies which will be discussed more in-depth in the following section. The proportion of ultrasonic and piezoelectric fingerprint acquisition devices is relatively small compared to the other three types of sensors and were not considered for this research, but it still does deserved mention because it is a feasible method of capturing fingerprints.

2.4.2.1. Optical Sensors

Optical fingerprint capture devices have the longest history of all automated fingerprint image acquisition devices (O'Gorman & Xia, 2001). The earliest methods of optical capture involved using a camera and taking direct images of fingerprints. This method had several issues - the most prominent one was that fingerprint ridges and valleys were not differentiated by color and a shadow needed to be introduced to differentiate between ridges and valleys (Setlak, 2004). Using the principal of frustrated total internal reflection (FTIR) was a major step forward for optical sensors.

> When light is incident on the interface of two media, it is partly transmitted into the second medium and partly reflected into the first. If however, the index of refraction of medium 1 is greater than the index of medium 2 and the angle of incidence exceeds the critical angle, total internal reflection occurs. The incident light is reflected completely back into medium 1 (Hawley, Roy, Yu, & Zhu, 1986).

The refractive index is different for ridges and valleys when a finger touches the platen of the optical sensor. Due to this phenomenon, light incident on valleys gets totally reflected and the light incident on ridges is not reflected which results in the ridges appearing dark in the final image (O'Gorman & Xia, 2001).

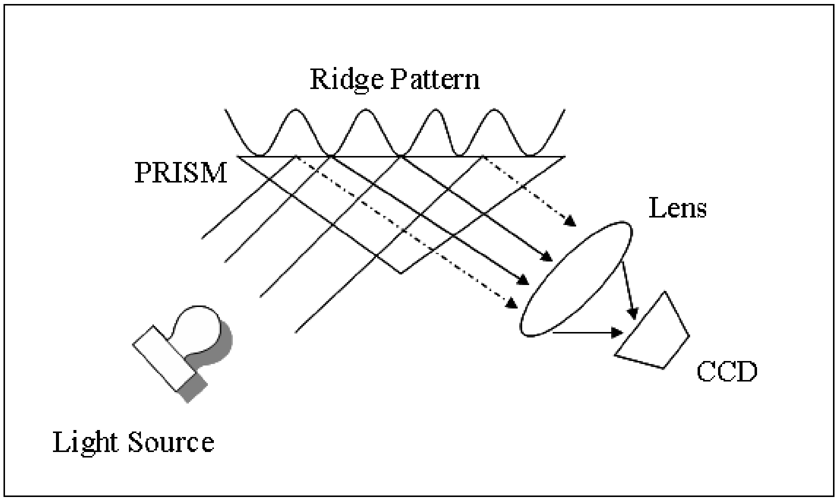

Figure 2.6 Optical Fingerprint Sensor

Figure 2.7 Fingerprint Image from Optical Fingerprint Sensor

As shown in Figure 2.6 this technology focuses light incident on the ridges and valleys onto a CCD/S3OS camera which captures the image. Older generation optical sensors used CCD cameras, but newer generation optical sensors used S3OS cameras. CCD cameras are usually analog which requires additional hardware to convert the analog signal to digital signal. S3OS cameras incorporate an analog-digital conversion process thus simplifying the cost and complexity of the system (O'Gorman & Xia, 2001).

Along with S3OS cameras, multiple mirrors and sheet prisms are also used to reduce size of sensors. The reason for introducing reflective surfaces within the optical sensor is to reduce the total optical length between the finger surface and the camera. The optical length can be modified by changing the lens focus and width of the camera array. Using multiple mirrors and sheet prisms is a method used to reduce the optical length, thereby reducing the size of the sensor.

Optical sensors are not susceptible to electrostatic discharge which can disrupt the image capture process in solid state capacitive sensors. If a direct light source is pointed at the sensor it could have an effect on the fingerprint image. Optical fingerprint sensors have a large platen area which allows it capture a large portion of the fingerprint thereby providing a larger area of the fingerprint in the image. The large size of optical fingerprint sensors is a disadvantage since it makes them impractical for use in small scale devices like cell phones. Optical fingerprint sensors are also affected by residue on the platen which could lead to an imperfect representation of the fingerprint.

2.4.2.2. Solid State Capacitive Sensors

Capacitive sensors work on premise that skin on finger surface is an equi-potential surface – constancy of potential is a requirement to obtain representative fingerprint images (Lim & Moghavvemi, 2000). In the late 1960's several inventors proposed methods of capturing fingerprint images using two dimensional conductive arrays. Rapid advances were made during the 1980's that led to new approaches to capacitive sensing technologies. But these early concepts all had a common drawback; they did not use integrated circuits that could take advantage of signal amplification and scanning. In the late 1980's researchers working with memory devices discovered that a finger place on a memory array caused data errors that resembled the spatial patterns of the fingerprint (Setlak, 2004). These memory chips used periodically refreshed charge stored on a capacitor in the memory cell and the difference in capacitance between ridges and valleys on skin of the finger caused the bits to flip. This phenomenon was refined and has been present in commercially available capacitive sensors available since mid 1990’s.

Capacitive sensors are constructed using a two-dimensional array of conductive plates. The finger is placed on a surface above the array so that the electrical capacitance of these plates is affected. The sensor plates under the ridge will have a larger capacitance than the sensor plates beneath the valley. This is because air has lower permittivity than skin, which leads to an increased capacitance in plates under the skin.

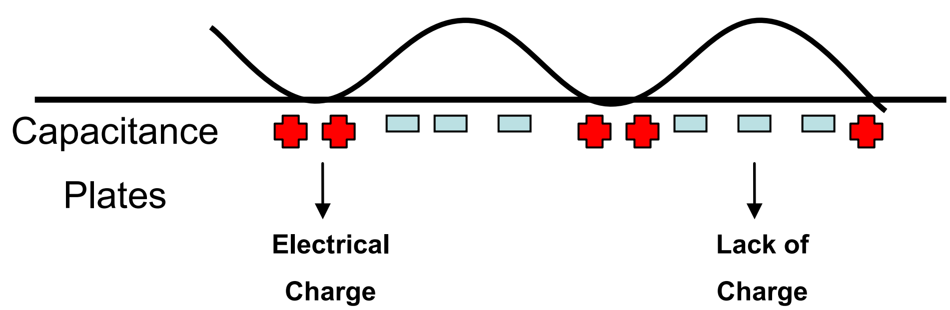

Figure 2.8 Capacitive Fingerprint Sensor

Capacitive fingerprint sensors can be classified into two classes: single plate capacitive sensors and double plate capacitive sensors. In a single plate capacitive sensor each pixel has at-least a capacitive sensor, a sensor circuit and additionally can have a logic circuit (Machida, Morimura, Okazaki, & Shigematsu, 2004). Each pixel is charged separately and measured separately in order to generate a complete fingerprint image. In double plate capacitive sensors two adjacent conductive plates correspond to one pixel. The capacitance differential between two adjacent plates is used to generate a pixel value.

Figure 2.9 Fingerprint Image from Capacitive Fingerprint Sensor

Capacitive sensors have a relatively smaller sensing area, which makes them suitable for use in small scale devices. The smaller sensing area results in a smaller image of the fingerprint which can be a disadvantage if the users are not consistent with their finger placement. Capacitive sensors are susceptible to electrostatic discharge, and to moisture content of the finger. They are not affected by light reflecting on the sensor surface.

2.4.2.3. Thermal Sensors

Sensors were designed which use thermal energy flux to capture fingerprints. When a ridge is in contact with a sensor surface of different temperature, heat flows between the ridge and sensor surface. The sensor surface is made up of an array of micro-heater elements, and a cavity is formed under between the valley of the fingerprint surface and heater element (J. Han, Kadowaki, Sato, & Shikida, 1999). Since the valley is not in contact with the sensor surface there is no heat flow between the valley and sensor surface. The heat flux is measured and converted into a digital representation of the fingerprint surface.

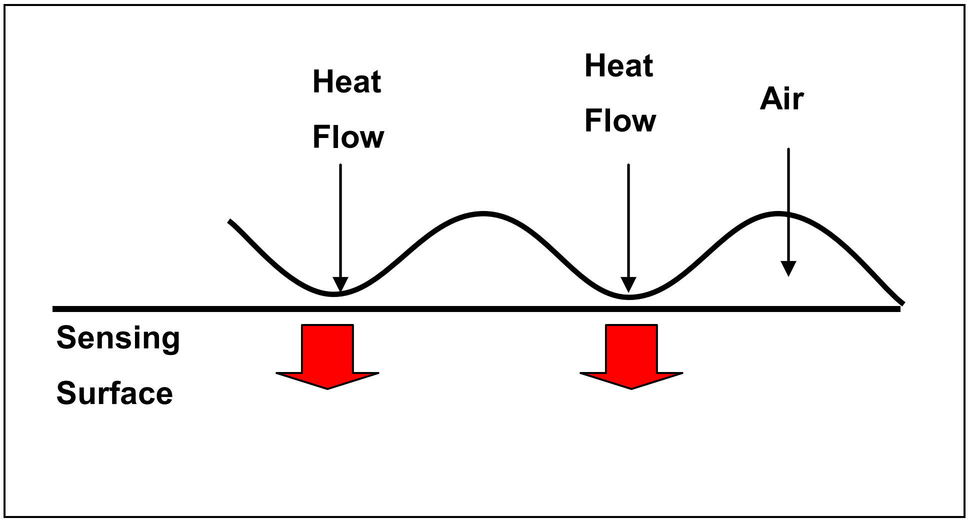

Figure 2.10 Thermal Fingerprint Sensor

When surfaces with different temperatures come in contact, they attempt to reach equilibrium as quickly as possible. Because of this phenomenon heat flux decays very rapidly when the finger is held is a static position. Thermal sensors are designed such that a finger has to be moved over the sensor surface in a constant motion in order to maintain an acceptable level of heat flux.

Figure 2.11 Fingerprint Image from Thermal Fingerprint Sensor

The thermal nature of the sensing technology makes it sensitive to environmental temperature conditions which is one of its major drawbacks. Thermal sensors are less influenced by moisture content of the finger (Adhami & Meenen, 2001). Thermal sensors are also not affected by light reflecting on the sensor surface. The small size of thermal sensors makes it a suitable choice for small scale devices.

2.4.3. Fingerprint Image Quality Assessment

Automated and consistent quality assessment of input samples is an important component of any biometric system. The ability of a system to detect and handle samples of varied quality levels is a significant contributor to performance of a biometric recognition system. The same holds true for fingerprint recognition systems. Fingerprint sample quality assessment is a topic of great interest as it can be used to ensure samples of appropriate quality are used for the matching process. The term quality is used in three different contexts as it relates to biometric sample quality (ISO, 2006):

1. Fidelity: reflects the accuracy of a sample's representation of the original source.
2. Character: reflects the expression of inherent features of the source.
3. Utility: reflects the observed or predicted positive or negative contribution of the biometric sample to overall performance of a biometric system.

Quality assessment algorithms use fidelity, character, utility or a combination of the three to compute a quality score. The following sections explain different types of fingerprint quality assessment algorithms and related research studies.

2.4.3.1. Types of Fingerprint Image Quality Assessment Algorithms

Aguilar, Fernandez and Garcia (2005) categorized existing image quality assessment algorithms into four broad categories:

1. Based on local features.
2. Based on global features.
3. Based on classifiers.
4. Hybrid algorithms based on local and global features.

They categorized algorithms which subdivided the fingerprint image into blocks, and computed quality scores for each block as local feature quality algorithms. This type of analysis takes into account specific local features. They categorized algorithms which search for abrupt changes in ridge orientation as global feature quality assessment algorithms. These algorithms tend to use 2-D discrete Fourier transform and energy concentration analysis of global structure to assess image quality of fingerprints. They third category of quality assessment algorithms were based on the premise that a quality measure should define a degree of separation between match and non-match distributions of a fingerprint. Using a relatively large dataset, classifiers can be trained using degree of separation as a response variable based on a vector of predictors, and then map the degree of separation to a quality index. They categorized hybrid algorithms as the ones which used an aggregation of local and global feature analysis to compute a quality index. The next section describes research related to the four categories of image quality assessment algorithms.

2.4.3.2. Experiments Related to Image Quality Assessment

Jain, Chen and Dass (2005) proposed a fingerprint image quality score computation which used both global and local quality scores. Energy concentration in frequency domain was used as a global feature. Their analysis as based on the premise that good quality images had a more peaked energy distribution while poor ones had a more diffused distribution and this property was used to compute the global image quality score. The local image analysis was performed by dividing the image into sub-blocks and computing the clarity of local ridge-valley orientation in each sub-block. A single quality score was computed as a weighted average of all the sub-blocks. These two methods were tested for improving fingerprint matcher performance. A decrease of 1.94% in EER was observed when images with the lowest 10% quality scores were pruned from the FVC 2002 DB3 database.

Jian, Lim and Yau (2002) described an image quality assessment method which used three different local and global features of the fingerprint. The first factor was a local feature. The strength and direction of ridge orientation information was computed since it determined contrast of the image and clarity of ridges. The second factor was a global feature. A gray-level analysis was performed to determine clarity of ridge valley structure. The third factor was an assessment of continuity of ridges which was also a global feature. The local and global quality scores were used to compute a final quality score. They also defined a quality benchmark, which was a function of minimal area of the fingerprint image required for authentication, number of falsely detected minutiae, number of correctly detected minutiae, and number of undetected minutiae. Regression analysis was performed by plotting the image quality score against the quality benchmark score and the model had a R^2 value of 0.76. The authors concluded their method could be used reliably for image quality assessment.

Abdurrachim, Kuneida, Li and Qi (2005) described a hybrid method for fingerprint image quality calculation which used seven indices. These seven indices were computed from local and global feature analysis. A Gabor feature index, smudginess and dryness index, foreground area index, central position of foreground index, minutiae count index, and singular point index were used as input parameters for a quality function. The weighting for each index was computed using an overlapping area methodology and linear regression methodology. Their results showed that their hybrid quality measure decreased EER by 12% - 34% when 10% of the images with lowest quality score were pruned from the database.

Jiang (2005) described a computationally efficient method for computing image quality using local features. A two stage averaging framework was used to take advantage of linear and normalized vector averaging. In the first stage, orientation and anisotropy estimation was performed using principal component analysis in order to determine local dominant orientation of the fingerprint image. Anisotropy is defined as the property of being directionally dependent. The modulus of the orientation vector was set to be its anisotropy estimate, whereas previous methodologies had used unity to set the modulus of the orientation vector. An improvement in EER was observed when compared to other methods of averaging the gradient orientation vectors.

Tabassi and Wilson (2005) proposed a novel methodology for assessing fingerprint image quality. They defined fingerprint image quality as a predictor of matcher performance before the matching operation is performed. They computed a normalized match score for every fingerprint in a database, and use quality scores of 11 different local features of a fingerprint as a non-linear combination to predict the normalized match score. They used a 3 layer feed-forward artificial neural network as their non-linear classification method. 5,244 images were used for training the network, and 234,700 images were used for testing the network. Their methodology divided quality into 5 levels, and matching performance of images at those 5 levels was compared. Their results showed that true accept rates (TAR) decreased and false accept rates (FAR) increased as quality of the fingerprints decreased. The [FAR,TAR] for images at quality level 1 (excellent quality) was [0.0096, 0.999] and for images at quality level 5 (poor quality) was [0.075,0.889].

Bigun, Fronthaler and Kollreider (2006) proposed a method to assess quality of an image using local features. They used the orientation tensor to draw conclusions about the quality of the image. The orientation tensor holds edge information, and this idea was used to determine if information was structured.

Their method decomposed the orientation tensor into symmetry representations and the symmetries were related to particular definition of quality. In order to test their quality assessment algorithm they used the QMCYT fingerprint database comprised of 750 unique fingers. Their proposed quality assessment algorithm was applied to all the fingerprints, and the fingerprints were split into 5 equally sized partitions of increasing quality. Each partition was tested on two different fingerprint matching algorithms, and a decrease in equal error rate was observed for each increasing quality partition. A reduction in EER by approximately 3% was observed between the worst and best quality partitions.

Alonso-Fernandez et al. (2006) performed a study to examine the performance of individual users under varying image quality conditions for a fingerprint database collected using three different fingerprint sensors. One of the main aims of their study was to investigate the effects of image quality in the performance of individual users. They proposed using a score normalization scheme adapted to the quality score of individual users. They collected 12 fingerprint images from 6 fingers of 123 participants on 3 different fingerprint sensors, which gave them a total of 26,568 fingerprint images. Image quality scores and matching scores were computed using NFIQ and BOZORTH3 respectively, both available from NIST. On performing a correlation test for verification EER threshold and quality scores they noticed a strong positive correlation between the two variables. They formulated a score normalization technique which incorporated the quality score of the fingerprint image and the verification scores in order to compute a new normalized matching score. They observed an EER reduction of 17% for one of a dataset of fingerprints collected using an optical sensor. They also observed that the score normalization technique improved EER for datasets which contained low quality fingerprints.

2.4.4. Fingerprint Feature Extraction

Most fingerprint matchers use minutiae details, ridge pattern details, or a combination of both to perform the matching operation. Minutiae based matchers extract minutiae details from the spatial domain fingerprint image and create a minutiae map of these features (Eshera & Shen, 2004). The transformation of spatial domain details to a minutiae map can be broken down into the following steps: pre-processing, direction calculation, image enhancement, singular point detection, and minutiae extraction(L. C. Jain, Erol, & Halici, 1999).

Direction calculation involves creating a direction map of the spatial domain details. A true direction map will contain orientation of every pixel of the ridge line, but this is a computationally expensive operation. In order to capture the direction details in a manner that is efficient, the fingerprint image is divided into blocks, and orientation is assigned to each block. Various algorithms have been proposed which use either overlapping blocks or completely separate blocks. Enhancement of a fingerprint image can be performed using a low-pass filter to eliminate noise. Filters have been proposed that use average ridge width and valley width as parameters. Since enhancement can have a significant impact on performance of a fingerprint matcher, a lot of work has been done in this area to explore different methodologies for enhancement. Binarization and skeletonization of fingerprint images involves eroding the ridge lines until their width is just a single pixel. The process of binarization and skeletonization reduces the amount of information needed to process for minutiae extraction (Jain et al., 1999). The core and delta are categorized as singularity points. Calculation of the rate of change of orientation on a ridge line has been used to detect the singularity points, also known as Poincare index. Core and delta points on a fingerprint image will typically have a Poincare index between -1/2 and +1/2, whereas ordinary points will have a Poincare index of 0 (Muller, 2001). Any discontinuity in the ridgeline flow is categorized as a minutiae point, and different methodologies for extracting these minutiae points have been proposed. Most of these algorithms will do some sort of ridgeline tracing and mark out all the discontinuities, and use the direction map to categorize their type. The next section discusses feature extraction methodologies in detail. The following section describes several experiments related to minutiae extraction.

2.4.4.1. Experiments Related to Minutiae Extraction

The FBI minutiae reader uses the binarization method as part of its minutiae extraction process. A composite approach based on local thresholding and a slit comparison formula was used to compare alignment of pixels along the dominant direction (Stock & Swonger, 1969).

Boashash, Deriche and Kasaei (1997) described a feature enhancement and feature extraction algorithm that used foreground/background segmentation which leads to a more precise ridge extraction process. The segmentation was based on the premise that in a given block, noisy regions had no dominant direction, which lead to foreground regions exhibiting a high variance in direction orthogonal to the orientation of the pattern, and a low variance along its dominant ridge direction (Mehtre, 1993). The ridge extraction process looked at each pixel in each foreground block with a dominant direction and determined if the pixel was part of the ridge. The obtained ridge image was then thinned to width of one pixel. Boashash et. al (1997) designed a minutiae extraction process which calculated the Crossing Number (CN) at point a point *P* which was expressed as (pp.5):

$$CN = 0.5 \sum_{i=1}^{8} | Pi - Pi + 1 | \quad P_9 = P_1 \qquad \text{Eq. 2.1}$$

where P_i was pixel value in 3 X 3 neighborhood of *P* as shown in Figure 2.12

P_4	P_3	P_2
P_5	P	P_1
P_6	P_7	P_8

Figure 2.12 3X3 matrix of P (Boashash et. al, 1997)

Depending on the calculated CN for a point, it could be categorized as an isolated point, end point, continuing point, bifurcation point or a crossing point. The x-coordinate, y-coordinate, and local ridge direction θ were extracted for every end point and bifurcation point.

Maio and Maltoni (1997) presented a direct gray scale minutiae detection approach without binarization and thinning. Their methodology followed ridge lines on a gray scale image according to the fingerprint directional image. A set of starting points was determined, and for each starting point the algorithm kept following the ridge lines until they intersected or terminated. They defined the ridge line as a set of points which were local maxima along one direction. The extraction algorithm located at each step a local maximum relative to a section orthogonal to the ridge direction. The consecutive local maxima points were connected to form an approximation of a ridge line. Once all ridge lines were detected, the minutiae points were detected by following the ridge lines along both directions of the ridge line. Their performance results showed a high proportion of errors caused by end point minutiae being detected as bifurcation minutiae points.

Foo, Ngo and Sagar (1995) described a fuzzy logic minutiae extraction which used a set of small windows to scan over the entire fingerprint image. A set of fuzzy logic based rules based on pixel intensities were used to determine if the window contains minutiae points. “Dark” and “bright” labels were used to describe gray pixel values. The membership functions mapped the gray pixel values against a degree of membership towards dark or bright categories. Their approach used a combination of six sub-windows to model the possible types and orientation of ridge endings and bifurcations.

Chan, Huang, and Liu (2000) proposed a novel methodology for extracting minutiae from gray level fingerprint images. Their methodology used the relationship between ridges and valleys to determine minutiae locations in a fingerprint. The relationship between ridges and valleys are affected by minutiae due to the discontinuities that minutiae introduce. A ridge and its two surrounding valleys ere traced, and if the two valleys met at a point then the heuristic decided a minutiae ending was detected. The heuristic decided a ridge bifurcation was detected if the distance and direction of the two valleys changed.

Amengual, Pereze, Prat, Saez and Vilar (1997) described a three phase approach for minutiae extraction. The first phase involved the processes of segmentation, directional image computation, directional image smoothing and binarization of image. The second phase involved thinning of the ridge lines in order to reduce the number of undesirable structures such as spurs and holes in the ridge lines. The minutiae extraction algorithm was based on method proposed by Rafat and Xiao (1991). They also used the concept of a CN. The CN for a pixel was computed using (Rafat et. al, 1991):

Eq. 2.2

$$CN = \sum_{k=1}^{k=8} |\Gamma(k+1) - \Gamma(k)| \quad \text{where } \Gamma(9) = \Gamma(1)$$

$\Gamma(p) = 1$ if pixel p is foreground

$\Gamma(p) = 0$

They used the following heuristic to classify the type of minutiae point:

if CN = 2 categorize as ridge ending

else if CN = 6 categorize as ridge bifurcation

The type of minutiae point, the x-coordinate, the y-coordinate, and the angle of orientation was also extracted for each minutiae point.

Engler, Frank and Leung (1990) investigated a neural network classification approach for extracting minutiae from noisy, gray scale fingerprint images. In this approach gray scale fingerprint images were first passed through a rank of six Gabor filters where each filter corresponded to a specific orientation of the minutiae to be extracted. The outputs from the filters were fed into a 3 layered back propagation neural network where the final output of the network was the determination of presence of minutiae.

The National Institute of Standards and Technology (NIST) designed a minutiae detection algorithm in the NIST Biometric Image Software (NBIS). Their minutiae detection algorithm, called MINDTCT, was designed to operate on a binary image where the black pixels represented ridges and white pixels represent valleys on surface of the finger. A pixel was assigned a binary value based on detection of ridge flow associated with the block. Detection of ridge flow in the block indicated the need to analyze the pixel intensities surrounding the pixel being examined, and a comparison with the grayscale intensities in the surrounding region indicated if the pixel was part of a ridge or valley. Once the binarization of the fingerprint image was complete, the binarized image was scanned to identify localized pixel patterns. Candidate ridge endings were detected in the binary image by scanning consecutive pairs of pixels in the fingerprint image looking for sequences which matched the pattern indicated in Figure 2.13.

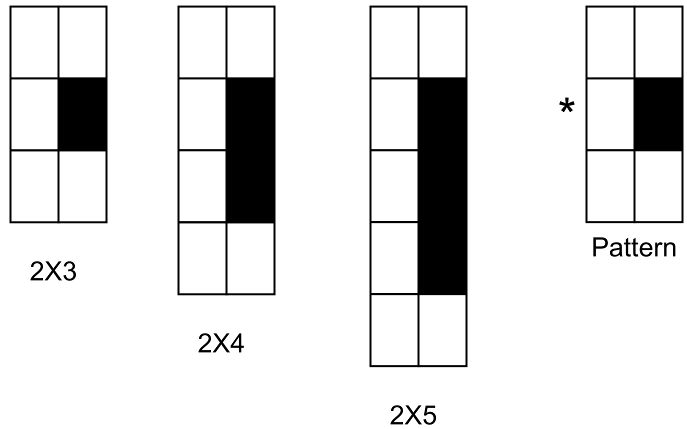

Figure 2.13 Pixel pattern used to detect ridge endings (Garris et al., 2004)

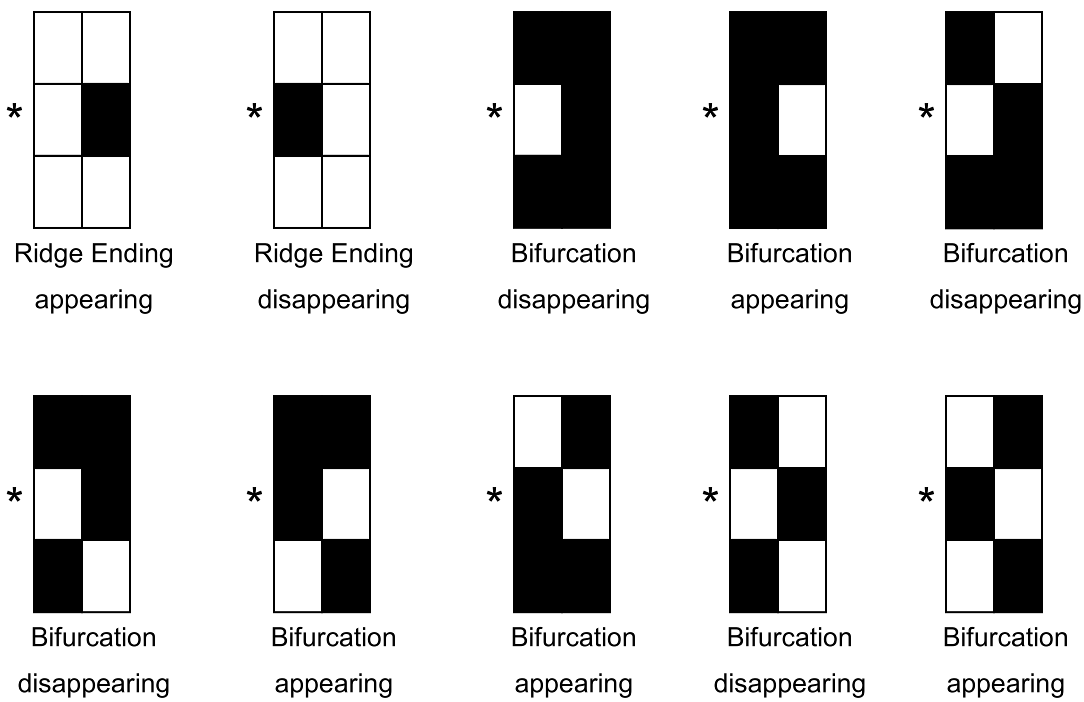

Figure 2.14 Pixel patterns for minutiae detection (Garris et al., 2004)

Once a candidate minutiae point was detected, the type of minutiae point was determined by comparing the pixel pair patterns as shown in Figure 2.14. All pixel pair sequences matching these patterns formed a list of candidate minutiae points.

The algorithm then performed post processing on the list of candidate minutiae points to ascertain if they were spurious minutiae points and removed any unnecessary artifacts in ridge lines, like islands and lakes (Figure 2.15).

Figure 2.15 Unnecessary features (Garris et al., 2004)

2.4.5. Fingerprint Matching

Fingerprint matching is the process of finding a degree of similarity between two fingerprints. There are different techniques of performing fingerprint matching, each of which examines different features of the fingerprint. Currently there are two primary methods used for fingerprint matching. The first method is a feature based approach that uses minutiae details for matching purposes. The second approach uses global ridge patterns of fingerprints for matching purposes. These matching algorithms will output a match score, i.e. the likelihood of a given fingerprint being the same as the one it is compared to. Fingerprint matching is a challenging process because of inconsistencies introduced during interaction of an individual with the sensor, temporary injuries to the surface of the finger, inconsistent contact, and a variety of other issues. The following section discusses research conducted previously in the field of minutiae based matching.

2.4.5.1. Experiments Related to Minutiae Based Matching

Chen, Jain, Karu and Ratha (1996) described a minutiae based matching algorithm based on Hough Transform algorithm. The algorithm first estimated the transformation parameters d_x, d_y, θ, and s between two fingerprint representations. d_x and d_y were translations along the x and y-axis, θ as the rotation angle, and s was the scaling factor. The second step involved aligning the two representations and counting the number of matched pairs within a margin of error. The matching scores for multiple transformations were collected in an array, and the transformation which maximized the number of matched pairs was considered to be the most accurate transformation. The final output of the matching algorithm was a set of top 10 fingerprints that matched the input fingerprint. They used a random sample of 100 fingerprints from the NIST-9 database to test the algorithm, and they achieved 80% accuracy rate at FRR of 10%. They also reported that their matching algorithm took about one second for each match, and would require parallel processing for large scale matching operations.

Bolle et al. (1997) described a string distance minutiae matching algorithm which used polar coordinates of extracted minutiae features instead of their Cartesian coordinates. This reduced the 2-D features to 1-D string by concatenating points in an increasing order of radial angle in polar coordinates. Their algorithm associated a ridge segment with each minutia which was used for aligning the two fingerprints being compared. The rotation and translation which resulted in maximum number of matched minutiae pairs as considered the most accurate rotation. All minutiae points for the two fingerprints were converted into a 1-D string at the optimal rotation angle, and the distance between the two strings was computed. Tests of their algorithm on the NIST-9 database resulted in a FAR of 0.073% and FRR of 12.4% at threshold of seven, and FAR of0 .003% and FRR of 19.5% at threshold level of 10.

A fingerprint matching algorithm is also included in NBIS. The matching algorithm is a modified a version of a fingerprint matcher developed by Allan S. Bozorth while at FBI. The matcher called BOZORTH3, uses only the x-coordinate, y-coordinate, and the orientation of the minutiae point to match the fingerprints. The main advantage of this matcher is that it is rotation and translation invariant. The main steps of the algorithm are as follows:

1. Minutiae features are extracted and an intra-fingerprint minutiae comparison table is created for the input fingerprint and the fingerprint to be matched against.
2. Use the minutiae table for the input fingerprint to the minutiae table for the comparison fingerprint to create a new inter fingerprint minutiae compatibility table.
3. Traverse the inter fingerprint minutiae compatibility table and calculate a matching score.

Tests performed by NIST using the BOZORTH3 matcher showed a matching score of 40 or higher as an acceptable threshold to consider the two fingerprints to be from the same finger. Elliott & Modi (2006) performed a study using the BOZORTH3 matcher which resulted in 1% FNMR for fingerprint dataset comprised of individuals 18-25 years old.

2.5. Large Scale Systems and Distributed Authentication Architectures

Fingerprint systems vary in size depending on its application, and they are typically characterized by the number of individuals in its database. Large scale systems are characterized by high demand storage and performance requirements, and currently large scale fingerprint systems retain fingerprint templates for approximately 40 million people and process tens of thousands of search requests everyday (Khanna, 2004).

The Registered Traveler (RT) program was authorized under the Aviation and Transportation Security Act (ATSA) as a means to establish requirements to expedite security screening of passengers (TSA, 2006). The RT program biometric system architecture consists of a Central Information Management System (CIMS), Enrollment Provider (EP), Verification Provider (VP), and Verification Station. The EP collects biographic and biometric information from applicants and transmits it to the CIMS. The CIMS aggregates the information and returns the biometric and biographic information packaged so that it can be stored on a smart card for the applicant. The EP has to use fingerprint sensors approved by the FBI, but the VP is not constrained in their choice of fingerprint sensors creating the possibility of enrolling and verifying on different fingerprint sensors. Currently there are 5 airports participating in the pilot program with a potential of more airports joining the program. Although the verification process requires matching only against the template on the smart card, the issue of interoperability still exists.

Large scale biometrics systems have to confront two different types of challenges: those common to large scale information management systems, and those specific to biometric implementations (Jarosz, Fondeur, & Dupre, 2005). The main considerations of a large scale system can be categorized into the following: architectural issues, administration issues, security issues, operational issues, and performance issues. For any biometric system transaction throughput and expected performance are of utmost importance, and the acquisition subsystem, feature extraction and template generation subsystem, storage subsystem, and matching subsystem contribute to those two factors. The acquisition subsystem is the first point of interaction between a individual and the system, and any inconsistencies introduced here will prorogate throughout the system. For a large scale system based on distributed architecture, the issues which arise at the first point of interaction become even more significant. The following section describes the experiments performed related to fingerprint sensor interoperability.

2.5.1. Fingerprint System Interoperability Experiments

Ko and Krishnan (2004) presented a methodology for measuring and monitoring quality of fingerprint database and fingerprint match performance of the Department of Homeland Security's Biometric Identification System. They proposed examining fingerprint image quality not only as a predictor of matcher performance but also at the different stages in the fingerprint recognition system which included: image quality by application, image quality by site/terminal, image quality by capture device, image quality by new participants or repeat visits, image quality by matcher enrollment, image quality by finger, and image quality trend analysis. They also pointed out the importance of understanding the impact on performance if fingerprints captured by a new fingerprint sensor were integrated into an existing identification application. Their observations and recommendations were primarily to facilitate maintenance and matcher accuracy of large scale applications.

Marcialis and Roli (2004) described a study which utilized a multi-sensor system for fingerprint verification. Along with examining interoperability error rates for fingerprint datasets collected from optical and capacitive sensors, they proposed a score fusion rule to decrease error rates for interoperable datasets. They used Biometrika FX2000 optical fingerprint sensor and Precise Biometrics MC100 capacitive sensor. Both sensors provided 500 S4i resolution fingerprint images. Using the FVC data collection process 10 fingerprint images were collected from 6 fingers of each of the 20 participants, resulting in a total of 1200 fingerprint images on each sensor. The evaluation process consisted of four stages. In the first stage the participant enrolled their index finger, middle finger, and ring finger of both hands on a capacitive and optical fingerprint sensor using the same minutiae extraction and template creation algorithm. In the second stage the participants presented the same finger on both the sensors. In the third stage the fingerprint acquired from the optical sensor in Stage 2 was matched to the template created in Stage 1 from the optical sensor and the capacitive sensor.

This resulted in two sets of match scores for each participant. In the fourth stage these two sets of match scores were fused to make a decision based on an acceptance threshold. Their results showed a False Accept Rate (FAR) and False Reject Rate (FRR) of 3.2% and 3.6% respectively for fingerprints from the optical sensor, a FAR and FRR of 18.2% and 18.8% respectively for fingerprints from the capacitive sensor, and a FAR and FRR of 0.7% and 1.3% respectively for the fusion rule.

Jain & Ross (2004) investigated the problem related to fingerprint sensor interoperability, and defined sensor interoperability as the ability of a biometric system to adapt raw data obtained from different sensors. They defined the problem of sensor interoperability as the variability introduced in the feature set by different sensors. They collected fingerprint images on an optical sensor manufactured by Digital Biometrics and a solid state capacitive sensor manufactured by Veridicom. The fingerprint images were collected on both sensors from 160 different individuals who provided four impressions each for right index, right middle, left index, and left middle finger. They used a minutiae based fingerprint matcher developed by Hong et al. (1997) to match images collected from the sensor, and then compared images collected from Digital Biometrics sensor to images collected from Veridicom sensor. Their results showed that EER of 6.14%for matching images collected from Digital Biometrics sensor and EER of 10.39% for matching images collected from Veridicom sensor .The EER for the matching images collected from Digital Biometrics sensor to Veridicom sensor was 23.13%. Their results demonstrated the impact of sensor interoperability on matching performance of a fingerprint system. Nagdir & Ross (2006) proposed a non-linear calibration scheme based on thin plate splines to facilitate sensor interoperability for fingerprints. Their calibration model was designed to be applied to the minutiae dataset and to the fingerprint image itself. They used the same fingerprint dataset used in the study conducted by Jain & Ross (2004), but used the VeriFinger minutiae based matcher and BOZORTH3 minutiae based matcher for the matching fingerprints. They applied the minutiae and image calibration schemes to fingerprints collected from Digital Biometrics sensor and Veridicom sensor and matched the calibrated images from the two sensors against each other. Their results showed an increase in Genuine Accept Rate (GAR) from approximately 30% to 70% for the VeriFinger matcher after applying the minutiae calibration model. For the BOZORTH3 matcher an increase in GAR from approximately 35% to 65% was observed.

Han et al (2006) examined the resolution different and image distortion due to differences in sensor technologies. They proposed a compensation algorithm which worked on raw images and templates. For raw images, all pixels of the image were offset along horizontal upper, horizontal center, horizontal down, vertical left, vertical center, and vertical right directions using the original resolution of the fingerprint sensor. This compensation lead to a transformation of location and angle of all minutiae points, along with a transformation of all pixels in the image. For templates transformation, only the positions and angles of the minutiae are transformed using the Unit Vector Method. Statistical analysis of their experimental results showed a reduction in the mean differences between non-compensated images and compensated images.

The International Labor Organization (ILO) conducted a biometric evaluation in an attempt to understand the causes of the lack of interoperability (Campbell & Madden, 2006). 10 products were tested, where each product consisted of a sensor paired with an algorithm capable of enrollment and verification. Data was collected from 184 participants and a total of 67,902 fingerprint images were collected on all 10 products. The test was designed to be a multi visit study, where the participant enrolled their left and right index finger and verified each finger 6 times on each product during each visit. Native and interoperable FAR and FRR were computed. Mean genuine FRR of 0.92% was observed at genuine FAR of 1%. The objectives of this test were twofold: to test conformance of the products in producing biometric information records in conformance to ILO SID-

0002, and to test which products could interoperate at levels of 1% FAR and 1% FRR. The results showed that out of the 10 products, only two products were able to interoperate at the mandated levels. The experimental design ensured extraneous factors like temperature and ordering of products would have a minimal effect on the results and outlined in detail in the report. This study was a pure assessment test, and no attempt was made to analyze the lack of interoperability.

The National Institute of Standards and Technology (NIST) performed an evaluation test to assess the feasibility of interoperability of INCITS 378 templates. The objectives of the test were three fold:

1. To determine the level of interoperability in matching two INCITS 378 templates generated by different vendors.
2. To estimate the verification accuracy of INCITS 378 templates compared to proprietary templates.
3. To compare the performance between the base INCTIS 378 template and enhanced INCITS template which includes ridge count information.

The minutiae template interoperability test was called the MINEX 2004 test. Four different datasets were used which were referred to as the DOS, DHS, POE and POEBVA. The POE and POEBVA dataset were culled from the US-VISIT dataset. All datasets had left and right index fingers using live-scan impressions. All the datasets were operational datasets gathered in on-going US government deployments. All the datasets combined consisted of fingerprint images from a quarter of a million people were used, and approximately 4.4 billion comparisons were executed in order to generate the performance evaluation reports. The POEBVA dataset was collected in separate locations, in different environments and different sensors. Also the data capture process for the authentication images was unsupervised. Fourteen different fingerprint vendors participated in the MINEX test.

The testing protocol examined three different interoperable verification scenarios:

1. The enrollment and verification templates were created using the same vendor's generator and template matcher from a different vendor.
2. The enrollment template generator and matcher were from the same vendor and the verification template generator was from a different vendor.
3. The enrollment template generator, the verification template generator and template matcher were all from different vendors.

The performance evaluation criteria was based on false non match rates (FNMR) and false match rates (FMR). FNMR was calculated at fixed FMR of 0.01% for all the matchers. Performance matrices were created which represented FNMR of native and interoperable datasets and provided a means for a quick comparison. The tests did not show a particular pattern of better performance for comparison of templates from the same generators compared to templates from different generators.

The MINEX report identified quality of the datasets as a factor which affected level of interoperability. The DOS and DHS datasets were of lower quality and did not exhibit a level of interoperability as that of POEBVA and POE databases. The MINEX test acknowledged it did not use images from sensors of different technologies like capacitive and thermal. The test team also indicated that performance for datasets which used fingerprints from capacitive, thermal and optical sensors would be different than what was observed in the MINEX test. One of the main conclusions from the MINEX test was related to the quality of the datasets used in the test. Along with the quality metric, the environment conditions and the interaction characteristics of data collection was not constant for the fingerprint collection process. Environmental conditions and interaction factors like sensor placement have an impact on the quality of fingerprints (Elliott, Kukula, & Modi, 2007). The unknown nature of the data collection process was not a part of the evaluation of the MINEX test.

Bazin & Mansfield (2007) described a study which focused on ISO/IEC 19794-2 standard for finger minutiae data exchange. The tests described in the paper encompassed the first phase of Minutiae Template Interoperability Testing (MTIT) evaluation with the aim of identifying key issues which contribute to increased error rates in interoperability scenarios. This study analyzed compact card templates along with data interchange record formats. The main difference between record and compact card templates is based on representation of ridge endings either by valley skeleton bifurcation points or by ridge skeleton end points. In record templates ridge endings can be represented only by valley skeleton bifurcation points. Four different vendors participated in this study and the three interoperability verification scenarios used in MINEX test were used for this study. Along with testing interoperability between vendors, this test also examined level of interoperability between proprietary, record and card formats. The two index fingers from 4,041 individuals were used in the final performance evaluation test. As with the MINEX test, only optical live-scan sensors were used. Also, fingerprints were also collected on 10-print scanners, and then segmented for use in the test. In order to evaluate performance, FNMR performance matrices were generated as several different fixed FMR. These matrices contained native and interoperable dataset performance rates. The results from this study showed lower FNMR for native matching comparisons compared to interoperable matching comparisons at all levels of FMR.

As with the MINEX test, this test also examined the link between interoperability and amount of data included in the standardized templates. The source of fingerprints, characteristics of the fingerprint sensors, or characteristics of the finger skin was not part of their analysis. The fingerprints were collected from different environments which could potentially lead to extraneous variables confounding results of the test. Not considering the source of the fingerprints should not be seen as a failing of this test since it was not a part of their research agenda, but the results point to the importance of studying fingerprint sensor related issues as a means of reducing interoperability error rates.

Poh, Kettler and Bourlai (2007) attempted to solve the problem of acquisition mismatch using class conditional score distributions specific to biometric devices. Their problem formulation depended on class conditional score distributions, but these probabilities were not known a-priori. They devised an approach which used observed quality measures for estimating probabilities of matching using specific devices. Their device specific score normalization procedure reduced the probability of mismatch for samples collected from different devices. They applied their approach to fingerprint and face recognition systems. They used the BioSecure database for their study. They collected data from 333 participants, of which 206 participants were considered to be genuine users. The remaining 126 participants were treated as zero-effort imposters. The fingerprint data was collected on an optical fingerprint sensor and a thermal fingerprint sensor. The BOZORTH3 matcher was used. They used the fingerprint quality indices described by Jain, Chen, & Dass (2005). Using Expected Performance Curves (EPC), which provided unbiased estimates of performance at various operating points, their results showed better performance for matching fingerprints from different devices using their framework.

2.6. Performance Metrics

Error rates are the basic units of performance assessment for matching engines. These error rates not only reflect the strength of the matching algorithm, but also the probabilistic nature of the decision making part of the algorithm. Attempting to compare performance metrics for multiple matching engines has been a challenge that researchers have attempted to address. The following sections outline different performance metrics that map functional and operational relationships between different error rates.

2.6.1. Receiver Operating Characteristic Curves

Receiver operating characteristic (ROC) curves are a means of representing results of performance of diagnostic, detection and pattern matching systems (Mansfield & Wayman, 2002). A ROC curve plots as a function of decision threshold the FMR on the x-axis and true positive rates on the y-axis.

$$ROC (T) = (FMR (T), TAR (T)) \text{ where T is the threshold} \qquad \text{Eq. 2.3}$$

For performance comparison of biometric systems, a modified ROC curve called Detection Error Tradeoff (DET) curve is used (Doddington, Kamm, Martin, Ordowski, & Przybocki, 1997). A DET curve will plot FAR on the x-axis and FRR on the y-axis as function of decision threshold. A DET curve can also be created by plotting FMR on the x-axis and FNMR on the y-axis as function of decision threshold. Comparing DET curves for different systems allows comparison of the systems at a threshold that is deemed preferable for the application of the biometric system.

$$DET (T) = (FMR (T), FNMR (T)) \text{ where T is the threshold} \qquad \text{Eq. 2.4}$$

2.6.2. Equal Error Rate

Using multiple DET curves to compare matching performances of different matchers assumes having a known operational threshold. A more convenient comparison of multiple matchers would necessitate the reduction of the DET curve to a single number (R. M. Bolle, Connell, Pankanti, Ratha, & Senior, 2004a). The EER operating point is a computation which is generally regarded as an obvious choice to judge quality of a matcher. The EER is the operational point where FNMR=FMR. This is a useful point of comparison only if the FNMR and FMR are supposed to be equal. For unequal FNMR and FMR, analysis at EER operational point will not provide any useful information.

2.6.3. Difference in Match and Non-Match Scores

Measuring the difference in match score density and non-match score density has been used as a performance assessment criteria. Daugman and Williams (1996) define the measure of separation for a matcher as

$$d' = \frac{\mu matchscores - \mu nonmatchscores}{\sqrt{(1/2)\,(\sigma^2 matchscores + \sigma^2 nonmatchscores)}} \qquad \text{Eq. 2.5}$$

$\mu_{matchscore}$ and $\sigma^2_{matchscore}$ is mean and variance of match scores of genuine users, and $\mu_{non\text{-}matchscore}$ and $\sigma^2_{non\text{-}matchscore}$ is mean and variance of non-match scores of mismatch fingerprints, respectively. *d'* can be used to compare multiple matchers, but it will be reliable only if there is a significant difference in performance. Matchers that might have similar genuine score distributions can be hard to compare, but using *d'* measure is useful because it uses non-match score distributions as part of the computation.

2.6.4. User Probabilities and Cost Functions

The previous two methodologies assume that the probability of FMR and FNMR is the same. A particular biometric system might have to be analyzed keeping in mind the probability of a user being an imposter or a genuine user, and the consequences of a false match or a false non-match. There are methodologies which use security and convenience tradeoff to determine which system is better suited for particular requirements.

In Equation 2.6 P_G is the prior probability of a user being genuine, C_{FNM} is the cost associated with a false non match, C_{FM} is the cost associated with a false match (R. M. Bolle, Connell, Pankanti, Ratha, & Senior, 2004a). This is a useful method for quantifying the cost of the system in dollars, but it is only as accurate as the input parameters.

$$\text{Cost} = C_{FNM} * \text{FNMR} * P_G + C_{FM} * \text{FMR} * (1\text{-}P_G) \qquad \text{Eq. 2.6}$$

2.6.5. Cumulative Match Curve

Cumulative Match Curve (S3C) is a statistic that measures capabilities of matching engine that returns a ranked list. The ranked list can be interpreted as a list of enrolled samples that are arranged according to the match scores against an input sample. A better matching engine will return a better ranked list. ROC curves and its derivatives are useful for analysis of matching engines that produce continuous random variables; a S3C is useful for analysis of matching engines that produce discrete random variables (R. M. Bolle, Connell, Pankanti, Ratha, & Senior, 2004b)

CHAPTER 3. METHODOLOGY

3.1. Introduction

The chapter outlines the data collection, data processing, and data analysis methodology utilized in this dissertation. The main purpose of this chapter is to ensure the experiment can be repeated in a reliable manner.

3.1.1. Research Design

This dissertation used the experimental research method. Experimental research methods require a highly controlled environment with a clear identification of independent variable/variables, dependent variables, and extraneous variables. In this study the fingerprint sensors were the independent variables and minutiae count, image quality scores, and match scores were the dependent variables. This study also required an experimental research method to ensure all subjects followed the same process of providing fingerprints on different sensors. The experimental research method is useful for controlling sources of error. For this experiment, the sources of error were controlled by randomizing assignment of sensors to subjects, maintaining repeatability of the process, and controlling the extraneous variables like room lighting, temperature, and humidity. The experimental research method also allowed for an experimental design which minimized internal and external validity issues. These are discussed in Section 3.5.

3.2. Data Collection Methodology

Each participant was required to follow the same data collection procedure. At the beginning of the experiment, the participant completed a questionnaire, the details of which are listed in Section 3.2.8. The responses were collected to provide further research opportunities of investigating their correlation with image quality and performance metrics. The detailed questionnaire can be found in Appendix B.

After completing the questionnaire, the participant proceeded with the fingerprint data collection procedure. A test administrator supervised the data collection. There were nine fingerprint sensors used in this study, and their details can be found in Section 3.2.4. The participant interacted with each fingerprint sensor while sitting in a chair. The participant provided six fingerprint images from the index finger of his/her dominant hand for each sensor. Previous studies have showed that index fingers tend to provide higher quality images compared to other fingers (Elliott & Young, 2007). If a participant was ambidextrous, the index finger from the preferred writing hand was used. The participant was given a randomized list of fingerprint sensors and started the data collection with the first sensor on the randomized list. Before interacting with each fingerprint sensor, the participant's moisture content, elasticity, oiliness and temperature was recorded from the surface of the finger skin. The fingerprint sensor was placed on a pressure measuring device. The peak pressure applied by the participant's finger on the sensor while providing each fingerprint image was recorded. A detailed description of the hardware used in this study is given in Sections 3.2.3 and 3.2.4. Data collection was complete only after the participant provided fingerprints on all sensors. A flowchart of the data collection protocol is given in Appendix A. The physical layout of the data collection area is given in Appendix D.

3.2.1. Participant Selection

Participants were selected from the population available at the West Lafayette campus of Purdue University. For the purpose of this study convenience sampling was performed to form the participant pool. 190 participants completed the study.

3.2.2. Timeline

Data collection for this study commenced on January 31st, 2008 and was completed on April 25th, 2008.

3.2.3. Data Collection Hardware & Software

Moisture content, oiliness, and elasticity finger skin were measured using Triplesense TR-3 manufactured by Moritex Corporation (Figure 3.1). This is a commercially available hand held device capable of providing all three measurements. Measurements are on a normalized scale of 1-100 based on testing performed by the manufacturer. This device came pre-calibrated from the manufacturer. No additional calibration was performed on the device.

The temperature of finger skin was measured using Raytek MiniTemp^{TM} Infrared Thermometer (Figure 3.2). This is a commercially available device which provides readings on a Fahrenheit scale. This device came pre-calibrated from the manufacturer and no additional calibration was performed.

Figure 3.1 Triplesense TR-3

Figure 3.2 Raytek MiniTemp™

A Vernier Force Plate was used to measure the force applied on the fingerprint sensor. The device was calibrated to zero N before each interaction.

3.2.4. Fingerprint Sensors

Fingerprints were collected using nine different sensors. There were three swipe and six touch interaction type sensors. There was one thermal sensor, four capacitive sensors and four optical sensors. All sensors were of approximately 500 S4i resolution. All fingerprint sensors chosen were made by commercial sensor manufacturers and were commercially available. None of the fingerprint sensors were specifically built for this dissertation. The fingerprint sensors were also chosen on basis of availability and with the intent of including sensors that were of swipe and touch interaction type and optical, thermal, and capacitive acquisition technology. There are other types of acquisition technologies available which were not selected due to their limited use in the live deployments. The specifications for each sensor are listed in Table 3.1.

Table 3.1 List of Fingerprint Sensors

Fingerprint Sensor	Type of Sensor	Action	Capture Area (mm)	Resolution S4i (approximate)
Sensor 1	[redacted]		14 X .4	500
Sensor 2			13.8 X 5	500
Sensor 3			30.5 X 30.5	500
Sensor 4			14.6 X 18.1	500
Sensor 5			12.8X15	500
Sensor 6			16 X 24	500
Sensor 7			15 X 15	500
Sensor 8			12.8 X 18	508
Sensor 9			12.4 X .2	500

3.2.5. Fingerprint Sensor Maintenance

The fingerprint sensors were calibrated the first time they were setup for the experiment. Calibration was performed using the software which accompanied each sensor. The fingerprint sensors were not recalibrated after the first setup.

At the beginning of each fingerprint collection session, every fingerprint sensor was cleaned by wiping it with a piece of cloth (Campbell & Madden, 2006). This action was performed to minimize any accumulation of residue.

3.2.6. Software or Sensor Malfunction

In case of hardware or software malfunctions, the following steps were taken:

- Restart the data collection software.
- Reconnect and reinitialize the fingerprint sensor.
- Reboot the data collection computer.

In case of a sensor failure, all fingerprints collected from that sensor were to be removed from the data analysis section. None of the fingerprint sensors used for data collection failed during the data collection period.

3.2.7. Variables Measured during Data Collection

Fingerprint images collected from all sensors were stored for the duration of the study. Information about the participant's age, gender, ethnicity, occupation, missing fingers, temperature of finger surface, moisture content of finger surface, oiliness of skin, elasticity of skin and pressure applied on the sensor was collected as performed in a previous study by Kim et al. (2003). Failure to acquire (FTA) was recorded if Verifinger 5.0 extractor determined a fingerprint image to be of insufficient quality. If three consecutive failure to acquire attempts occurred, a failure to enroll (FTE) was recorded and the participant was asked to proceed with the next fingerprint sensor on the randomized list. For the purpose of this study gender, ethnicity, and occupation were treated as categorical variables; missing fingers, temperature, moisture content, oiliness and elasticity of finger skin, and pressure applied on fingerprint sensor were treated as interval variables; age was treated as a ratio variable. This information was collected to report the demographic mix of the participants, but was not used for analysis.

3.2.8. Variables Controlled during Data Collection

Each participant had the discretion of adjusting the height of the chair during the data collection session. The lighting level of the room, room temperature, relative humidity in the room was kept within the Knoy Hall, West Lafayette, building regulations during the experiment. The tilt angle of the fingerprint sensor was controlled by laying it flat on the pressure plate surface. The surface of the pressure plate and the back of each fingerprint sensor was covered with Velcro to minimize the movement of the sensor when the participant interacted with the sensor. The surfaces of all sensors were cleaned at the start of each data collection session.

3.3. Data Processing Methodology

3.3.1. Minutiae Count and Image Quality Processing

The basic variables analyzed for this study were minutiae count, and image quality score of the fingerprint image. Minutiae count and image quality scores were generated using Aware Image Quality, and MINDTCT and NFIQ which are a part of NBIS. Minutiae count was also generated using VeriFinger 5.0 extractor. For Aware Image Quality Software, minutiae count and image quality scores were ratio variables. Minutiae count was always greater than 0, and image quality scores ranged from zero to 100. For MINDTCT, minutiae count was a ratio variable and was always greater than zero. NFIQ scores ranged from one to five and was treated as an interval variable. For VeriFinger 5.0 minutiae count was a ratio variable. The determination of the variable types was used to formulate the appropriate statistical tests.

3.3.2. Fingerprint Feature Extractor and Matcher

Fingerprint feature extractors and fingerprint matchers provided with VeriFinger 5.0 and NBIS were used. VeriFinger 5.0 is a commercially available fingerprint feature extractor and matcher. It has been a participant in Fingerprint Verification Competitions in 2000, 2002, 2004, and 2006, and in the INCITS 378 Fingerprint Template performance and interoperability test conducted by NIST. MINDTCT and BOZORTH3, which are a part of NBIS, were used for extraction and matching respectively. The genuine match scores and imposter non-match scores were also ratio variables and were always a non-negative number.

3.4. Data Analysis Methodology

Hybrid testing, which combined live acquisition and offline matching, was performed to analyze the data collected in this experiment (Grother, 2006). Genuine match scores and imposter match scores were generated offline once all 190 participants had completed their data collection sessions. A hybrid testing scenario was necessary for an experiment which incorporated multiple sensors because live users would not want to sit through combinatorial use of multiple sensors. The scenario also provided an opportunity to use the entire dataset for testing purposes.

3.4.1. Score Generation Methodology

The process outlined in this section was performed using VeriFinger 5.0 and BOZORTH3. All participants provided six fingerprint images on each sensor from the index finger of their dominant hand, or their preferred writing hand if they were ambidextrous. The resulting six fingerprint images participant were split into two groups; the first three images were placed in an enrollment dataset and the last three images were placed in a test dataset. Enrollment template datasets were created for all nine sensors, and test template datasets were created for all nine sensors using feature extractors from VeriFinger 5.0 and BOZORTH3. This methodology is graphically represented in Figure 3.3. Enrollment templates from each dataset were compared against templates from each test dataset resulting in a set of scores *S*, where

$S = \{(E_i, V_j, score_{ij})\}$
i= 1,.. ,number of enrolled templates
j = 1,.., number of test templates
$score_{ij}$ = match score between enrollment template and test template

The set of scores, which consisted of genuine match scores and imposter match scores, were placed in a matrix form (Figure 3.4). This matrix of scores had the following properties:

1. The number of rows and columns were equal, and corresponded to the number of fingerprint sensors used.
2. The diagonal represented the native datasets, and all the cells not in the diagonal represented interoperable datasets.

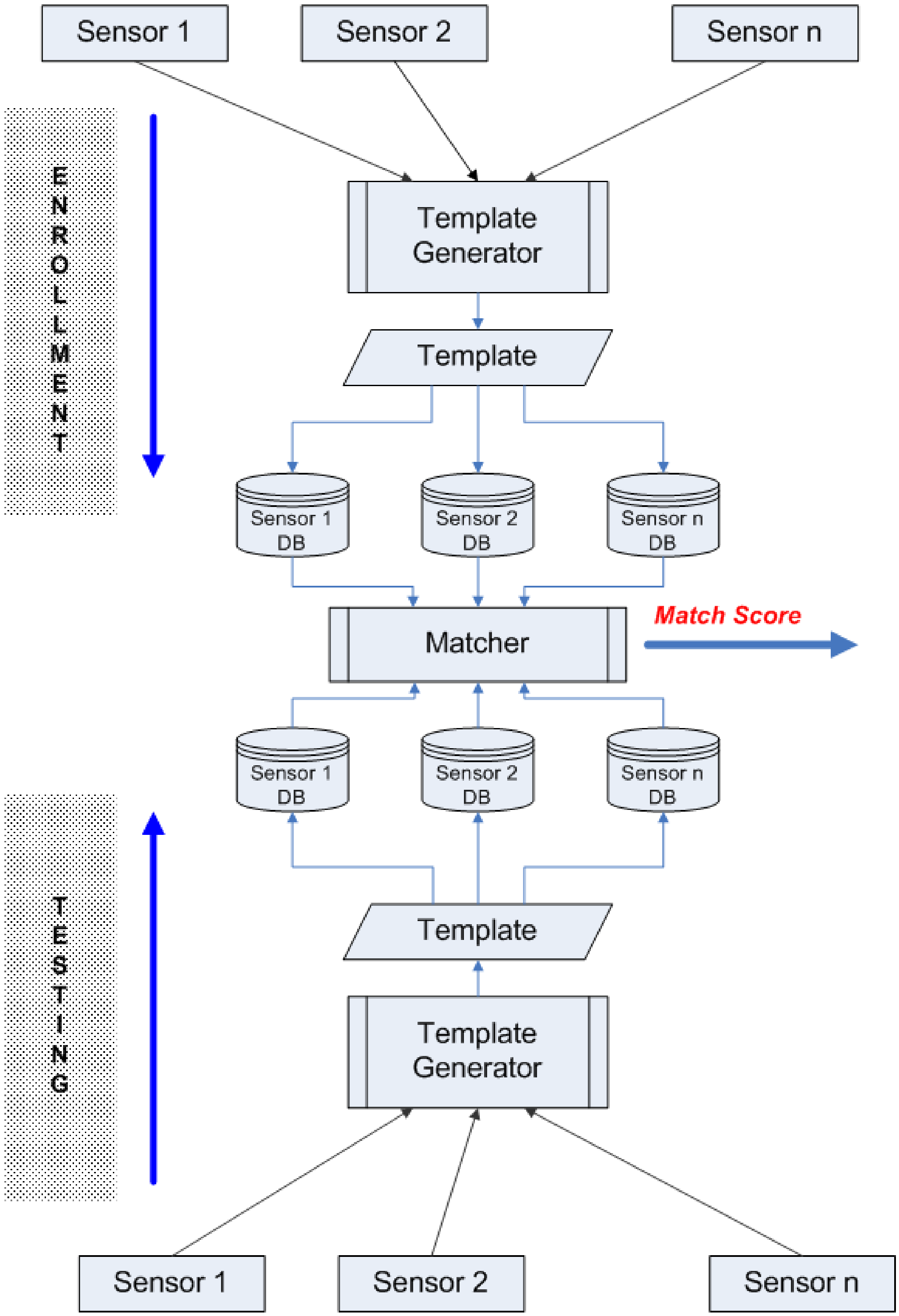

Figure 3.3 Generation of match scores

Match Score

	Sensor 1	Sensor 2	………	Sensor n	
Sensor 1	$Score_{11}$	$Score_{21}$	Score	$Score_{n1}$	Enrollment Sensor
Sensor 2	$Score_{21}$	$Score_{22}$	Score	$Score_{n2}$	
. .					
Sensor n	$Score_{n1}$	$Score_{n2}$	Score	$Score_{nn}$	
	Verification Sensor				

Figure 3.4 Scores of native and interoperable datasets

3.4.2. Analysis Techniques

The steps outlined in this section were performed on data generated from Aware Image Quality Software, VeriFinger 5.0, MINDTCT and BOZORTH3. Based on previous research conducted in fingerprint sensor interoperability experiments, the analysis methodology was categorized into the following:

1. Basic fingerprint feature analysis.
2. Match scores analysis.

3.4.2.1. Basic fingerprint feature analysis

An important factor when considering interoperability is the ability of different sensors to capture similar fingerprint features from the same fingerprint. Although human interaction with the sensor introduces its own source of variability, it is nonetheless important to statistically analyze variability introduced from different sensors. The flowchart of the fingerprint feature statistical analysis methodology performed in this dissertation is shown in Appendix C.

The analysis described in this section was performed using all six fingerprints collected from the participant. The basic fingerprint feature analysis involved examination of minutiae count and image quality scores. Minutiae count and image quality scores were calculated for each fingerprint image using Aware Image Quality, MINDTCT and NFIQ. Statistical tests were performed to test similarities in minutiae count between all fingerprint datasets. This analysis was performed separately on minutiae count generated by Aware Image Quality and MINDTCT. Statistical tests were performed to test similarities in image quality between all fingerprint datasets. This analysis was performed separately on image quality scores generated by Aware Image Quality and NFIQ. A model adequacy check for a parametric F-test was performed which involved the following:

1. Normality of residuals.
2. Constancy of variance of error terms.
3. Independence of observations.

For a parametric F-test of a between groups effect, it is more appropriate to check normality of the residuals than the raw scores (Montgomery, 1997). A parametric F-test is a ratio test of between group variance and within group variance. The within groups variance is the average square of the residuals of the group mean. A test of normality of residuals will show if any particular observation has a very strong influence on the group mean compared to other observations. This ensures that scores which are used to calculate the within-group variance are normally distributed. A violation of any of the assumptions 1) or 2) can be fixed by transformation of data (Montgomery, 1997).

Two sets of hypotheses were formed to test for similarity of minutiae count and image quality. The first set of hypothesis was used for testing similarity of minutiae count.

$H1_0$: There is no statistically significant difference among the mean minutiae counts of all fingerprint datasets.
$H1_A$: There is a statistically significant difference among the mean minutiae counts of any fingerprint datasets.

$H1_0$: $\mu_{i\ minutiae\_count} = \mu_{2\ minutiae\_count} \ldots\ldots = \mu_{n\ minutiae\_count}$ Eq. 3.1
$H1_A$: $\mu_{i\ minutiae\_count} \neq \mu_{2\ minutiae\_count} \ldots\ldots \neq \mu_{n\ minutiae\_count}$
n = number of datasets

The second set of hypothesis was used for testing similarity of image quality scores.
$H2_0$: There is no statistically significant difference among the mean image quality scores of all fingerprint datasets.
$H2_A$: There is a statistically significant difference among the mean image quality scores of any of the fingerprint datasets.

$H2_0$: $\mu_{i\ qscore} = \mu_{2\ qscore} \ldots\ldots = \mu_{n\ qscore}$ Eq. 3.2
$H2_A$: $\mu_{i\ qscore} \neq \mu_{2\ qscore} \ldots\ldots \neq \mu_{n\ qscore}$

If a statistically significant difference was observed for the test, all possible pairs of means were compared. Tukey's Honestly Significant Difference (HSD) was used to test all pairwise mean comparisons. Tukey's HSD is effective at controlling the overall error rate at significance level α, and thus preferred over other pairwise comparison methods (Montgomery, 1997). Two sets of hypotheses were tested for pair-wise differences in means, one set of hypothesis corresponding to the similarity of minutiae counts, and one set corresponding to the similarity of image quality scores.

$H3_0$: $\mu_{i\ minutiae\_count} = \mu_{j\ minutiae\_count}$ for all $i \neq j$ Eq. 3.3
$H3_A$: $\mu_{i\ minutiae\_count} \neq \mu_{j\ minutiae\_count}$ for all $i \neq j$

$H4_0$: $\mu_{i\ qscore} = \mu_{j\ qscore}$ for all $i \neq j$ Eq. 3.4

$H4_A$: $\mu_{i\ qscore} \neq \mu_{j\ qscore}$ for all $i \neq j$

3.4.2.2. Match Scores Analysis

The following analysis was performed separately for genuine match scores and imposter match scores generated by VeriFinger 5.0 and BOZORTH3. The set of genuine match scores and imposter match scores were analyzed with the data from the matrix of set of scores shown in Figure 3.5. The performance of interoperable datasets was analyzed using the following methods:

1. Performance interoperability matrix consisted of FNMR for all datasets collected in the experiment (Campbell & Madden, 2006). The VeriFinger 5.0 matcher was used to generate the set of raw scores and FNMR at fixed FMR of 0.01% and 0.1%. The VeriFinger 5.0 matcher deduces the decision threshold of FMR operational points based on internal tests of the matcher. For each fixed FMR a FNMR interoperability matrix was generated (Figure 3.5). These two matrices had exactly the same properties as the matrix of set of scores, as outlined in Section 3.4.1. For the BOZORTH3 matcher a single decision point was used to create the interoperability performance matrix. A match score of 40 or above indicated a true match as suggested by Garris et al. (2004). Since the score of 40 does not correspond to a particular FMR operational point, no operational point was assigned to performance interoperability matrix generate by NBIS matcher scores.

	Sensor 1	Sensor 2	....	Sensor n	
Sensor 1	*Score*$_{11}$	*Score*$_{21}$	*Score*	*Score*$_{n1}$	
Sensor 2	*Score*$_{21}$	*Score*$_{22}$	*Score*	*Score*$_{n2}$	
. .					*Enrollment Sensor*
Sensor n	*Score*$_{n1}$	*Score*$_{n2}$	*Score*	*Score*$_{nn}$	
	Testing Sensor				

FNMR at FMR x%

	Sensor 1	Sensor 2	....	Sensor n	
Sensor 1	*FNMR*$_{11}$	*FNMR*$_{21}$	*FNMR*	*FNMR*$_{n1}$	
Sensor 2	*FNMR*$_{21}$	*FNMR*$_{22}$	*FNMR*	*FNMR*$_{n2}$	
. .					*Enrollment Sensor*
Sensor n	*FNMR*$_{n1}$	*FNMR*$_{n2}$	*FNMR*	*FNMR*$_{nn}$	
	Testing Sensor				

Figure 3.5 Performance interoperability matrix at different operational FMR

For each interoperable dataset a simple placement consistency metric was formulated. If VeriFinger 5.0 extractor detected a core in the pair of fingerprints being matched by VeriFinger 5.0 matcher, the placement was considered to be consistent. Using this measure, a percentage of fingerprint matching pairs in which both had a core was calculated for all interoperable datasets. This analysis was performed to analyze the relation between consistency of placement and interoperable FNMR.

2. Test of FNMR homogeneity was performed on FNMR calculated in the previous section to test for equality using the chi-square distribution testing for homogeneity of proportions. The main objective of this test was to examine if the difference in FNMR among the datasets was statistically significant in their difference. This test aided in statistically testing equality of FNMR for native datasets and interoperable datasets. Since there were nine sensors, there were nine corresponding sets of hypothesis. Eq 3.5 shows a sample of the null hypothesis and alternate hypotheses for dataset S1:

$H1_0$: $p_i = p_2 = \ldots\ldots\ldots = p_n$ Eq. 3.5

$H1_A$: $p_i \neq p_2 \neq \ldots\ldots\ldots \neq p_n$

n = total number of native and interoperable datasets for sensor 1.

p= FNMR.

The chi-square test statistic is calculated as follows:

$$\chi^2 = \sum_n \frac{(fo - fc)^2}{fc}$$ Eq. 3.6

f_o = observed frequency

f_c= theoretical frequency if no false non matches occurred

The critical value of χ^2 was computed at a significance level of 0.05 and degrees of freedom (n-1), and then compared to the test statistic χ^2 from Eq. 3.6. If the test statistic exceeded the critical value, the null hypothesis was rejected. This same hypothesis test was repeated for the other eight fingerprint sensors. If the null hypothesis was rejected, further analysis was performed to examine which interoperable datasets caused the rejection. This was done using the Marascuillo procedure for a significance level of 0.05. The Marascuillo procedure simultaneously tests the differences of all pairs of proportions for all groups under investigation. The critical range was calculated as shown in Eq. 3.7 and compared to the result of Eq. 3.8. If the result from Eq. 3.8 exceeded the critical range, the difference was statistically significant.

$$r_{ij} = \sqrt{\chi^2_{(\alpha,k-1)}}\sqrt{\frac{p_i(1-p_i)}{n_i}+\frac{p_j(1-p_j)}{n_j}} \qquad \text{Eq. 3.7}$$

n_i = sample size of group i

p_i = FNMR of group i

k = total number of sensor groups

$$|\, p_i - p_j \,| \qquad \text{Eq. 3.8}$$

3. Test of Match of Scores: In order to statistically analyze the differences in genuine match scores and imposter match scores between the native dataset and interoperability dataset, either the Dunnet's comparison method (Montgomery, 1997) or the Kruskal Wallis test was used depending on model adequacy checks described in Section 3.4.2.1. Dunnet's method is a modified form of a t-test. The mean genuine match score and the mean imposter match score for the native dataset would be considered as the control. The mean genuine match score and mean imposter match score for each interoperable dataset would be tested against the control (i.e. the native dataset scores). One measure of fingerprint sensor interoperability can be described as the consistency of match scores of the matcher for native fingerprint datasets and interoperable fingerprint datasets. In statistical terms, this would be examined by testing for a significant difference of the genuine match scores between native and interoperable fingerprint datasets. The same

test would be repeated to test for a significant difference of imposter match scores between native and interoperable fingerprint datasets.

3.4.3. Impact of Image Quality on Interoperable Datasets

The impact of image quality on interoperable datasets was examined by removing low quality images and recalculating the test of FNMR homogeneity. Only fingerprint images which had NFIQ quality scores of one, two or three were used for the matching operation. The NIST Minutiae Exchange Interoperability Test report indicated that a relative reduction in interoperable dataset FNMR would be observed since poorly performing images were not used (Grother et al., 2006). This analysis was performed to examine if using higher quality images would lead to a higher degree of similarity in FNMR between native and interoperable datasets. After fingerprint images with NFIQ score of one, two or three were selected from the interoperable datasets and VeriFinger 5.0 template generator and matcher was used to perform the matching operations.

The test of homogeneity of proportions was conducted on the FNMR of native and interoperable datasets comprised of fingerprint images with NFIQ score of 1, 2 or 3. This was repeated for all possible interoperable datasets.

3.4.4. Post Hoc Analysis

The moisture content, oiliness, elasticity, and temperature of the finger skin were measured before an individual interacted with each new sensor. Basic descriptive analysis was performed on these variables to analyze their range, mean and variance. An exploratory analysis involving correlation matrices was performed on moisture content, oiliness, elasticity and temperature of the finger skin and its relation to image quality of the fingerprint provided. This analysis can be found in Appendix J.

An exploratory analysis using correlation matrices was also performed between all of the following relations: match scores of a pair of fingerprint images and the difference in quality of the pair of images, difference in ridge bifurcation count of the pair of images, difference in ridge ending count of the pair of images, and different types of acquisition technologies that were used to capture the pair of fingerprint images.

3.5. Threats to Internal and External Validity

3.5.1. Internal Validity

For the purpose of this study internal validity was defined as the confidence placed in the cause and effect relationship between the independent variable and dependent variables without the relationship being influenced by extraneous variables (Sekran, 2003). The following seven internal validity issues are discussed: history, selection, maturation, instrumentation, mortality rate, repeated testing, and experimenter bias.

History

Outside events may influence participants during the experiment or between repeated measures of the dependent variable (“Psychology 404,” 1998). Biometrics is currently receiving a lot of media coverage because of privacy issues. These events could have influenced a participant’s willingness to participate or comply with the instructions. Although these events are beyond the control of the experimenter, the participants were given an opportunity to ask questions about the experiment before proceeding with it. History effects are a more relevant threat to experiments which require repeated measures spread over multiple sessions. Since this was a single session experiment history effects were deemed to be minimal.

Selection

The threat of selection arises from selection of participants for multiple groups in an experiment. Experiments which employ a control group and different test groups are susceptible to this threat, as they utilize specific criteria for placing participants in certain groups. This experiment did not have multiple groups of participants -thus this threat was not a concern.

Maturation

For the purpose of this study, maturation was defined as the difference in measurements of the dependent variable due to passage of time (“Psychology 404”, 1998). Participants becoming more skillful, observing new details of the experiment, and acclimatizing themselves to the experimental conditions can affect the variables of interest. Multiple visit studies or studies with a long gap between subsequent visits are vulnerable to this threat. In this case participants completed the study only once, which reduced the risk of maturation. Participants waiting to complete the study could observe the previous participant, potentially causing the participants to change their interactions with the fingerprint sensors. For this reason, the data collection room was kept separate from the waiting room, and this threat did not have an impact on data collection.

Instrumentation

The reliability of instruments used to capture, measure and gauge the variables of interest is of utmost importance to an experimental setup. In order to have confidence in the results, the instruments should provide consistent readings throughout the study. All instruments used in this study are commercially available products, and remained the same for all participants. The fingerprint collection software, fingerprint feature extraction software, and fingerprint matching software are all commercially available or open source. A PC with Windows XP™ operating system was used for this study, which was not changed for the duration of data collection.

Mortality Rate

For the purpose of this study, mortality rate was defined as the dropout rate of participants from the experiment. Participants completed the study in a single session. The participants did not have to return for repeated sessions. This reduced the threat of mortality. Participants had the option of terminating the data collection session at any time during the study. Although this posed a constant threat to internal validity, the threat was mitigated by providing a positive and unambiguous experience to the participant.

Repeated Testing

Repeated testing can confound the measurement of dependent variables as the participants get repetitive in their experimental interaction and the variables of interest are not affected by manipulation of independent variables (Sekran, 2003). The experimental setup of this study required each participant to participate in a single session which reduced the threat of repeated testing. During this session, the participants were asked to provide six fingerprint images from the index finger of their natural hand on all fingerprint sensors. It was infeasible to completely control for repeated testing in such a setup. In order to minimize the order effects introduced by repeated interaction, each participant was presented with a random ordering of fingerprint sensors. This reduced the chances of random error affecting only one participant.

Experimenter bias

Expectations of an outcome by persons running the experiment could significantly influence the outcome of the experiment (“Psychology 404”, 1998). The data recording in this study was automated except for one fingerprint sensor which required the administrator to click the capture button. The ability of the administrator to influence the fingerprint placement of a participant was a constant threat, but it existed equally for all participants thereby minimizing its ability to affect specific participants. The generation of results for analysis was automated by the fingerprint feature extractor and matcher thereby reducing the threat of experimenter bias.

3.5.2. External Validity

For the purpose of this study, external validity was defined as the ability to generalize the results from a study across the majority of the experimental population (Sekaran, 2003). Threats to external validity can arise from population difference, experimental setup, and localization issues. These issues and their respective mitigation methods are discussed in detail in the following sections.

Population Difference

Population difference arises when samples selected for the experiment are not representative of the population to which the results are generalized. This threat can typically be mitigated by choosing a sample that is representative of the population. Data collection for this study was conducted at the West Lafayette campus of Purdue University. A majority of participants were students attending Purdue University, and although participants from a diverse age group were sought for this experiment in order to make the sample makeup more representative of the population, this external validity issue constrained generalizability of this study. Previous studies have shown that there is an age effect on fingerprint recognition performance, and therefore the generalizability of this study was restricted to a similar aged population (Elliott & Sickler, 2005).

Experimental Setup

Fingerprint recognition can be used for logical or physical access control. The deployment environments for these two types of access control can be vastly different. Physical access control systems are placed in a relatively uncontrolled environment and the interaction postures also differ vastly from logical access control. Logical access control systems are used with some kind of a computing resource, like a desktop, and are generally in a controlled environment. In order to control the environmental threat, the experiment was designed for logical access control applications used for network or desktop login. Participants interacted with the fingerprint devices while seated in a chair, and were allowed to adjust the seat height as to their preference. This setup was representative of a real world scenario for logical access control in an office environment. The temperature and lighting in the experimental area were controlled so that their effects were representative of an office environment.

Localization Issues

Applying the results from an experiment conducted in a laboratory to a general operational scenario can lead to localization issues. This threat is more relevant to experiments which can be affected by a change in geographical locations. This experiment was conducted in a highly controlled environment which simulated an office environment. Office environments across different geographical regions tend to have similar conditions, so the results of this study are generarlizable across office environments irrespective of their locations. The skin on fingers is affected by outdoor environmental conditions, and this can affect fingerprint recognition systems. For the purpose of this experiment this was not expected to be severe due to the ability of the human body to acclimate itself to a controlled environment. Localization effects were not anticipated to be a major threat for this study.

3.6. Evaluation Classification

Table 3.2 partitions the test along seven categories and provides a summary of the experimental evaluation.

Table 3.2 Evaluation Classification (Mansfield & Wayman, 2002).

Experimental Application Types	**Classification for this research**
Application Classification	Technology
Co-operative or Non Co-operative	Co-operative Users
Overt versus Covert	Overt
Habituated versus Non-Habituated	Both
Attended versus Non-Attended	Attended
Standard Environment	Yes
Public versus Private	N/A
Open versus Closed System	Closed

3.7. Summary

This chapter has outlined the data collection, data processing and data analysis methodology that were used in this dissertation. Both the data collection procedures and test methodology for this study were based on prior research related to interoperability of fingerprint sensors and fingerprint recognition systems. This chapter also acknowledged the different extraneous variables that potentially affected the validity of the study, and explained how the study attempted to mitigate the effects of the variables to increase validity of the study.

CHAPTER 4. DATA ANALYSIS

This chapter discusses results from surveys completed by the participants, followed by analysis of minutiae count, image quality scores and match scores of native and interoperable datasets. The data analysis methodologies and statistical tests described in Chapter 3 were performed on the datasets collected from the nine sensors. Each native dataset was named after the sensor that was used to collect it. Table 4.1 outlines the details of each sensor and the dataset name given to each sensor.

Table 4.1 Coding of Fingerprint Sensors and Datasets

Dataset Name	**Fingerprint Sensor**	**Type of Sensor**	**Action**	**Capture Area (mm)**
S1	[redacted]	[redacted]	Swipe	14 X .4
S2	[redacted]	[redacted]	Swipe	13.8 X 5
S3	[redacted]	[redacted]	Touch	30.5 X 30.5
S4	[redacted]	[redacted]	Touch	14.6 X 18.1
S5	[redacted]	[redacted]	Touch	12.8X15
S6	[redacted]	[redacted]	Touch	16 X 24
S7	[redacted]	[redacted]	Touch	15 X 15
S8	[redacted]	[redacted]	Swipe	12.4 X .2
S9	[redacted]	[redacted]	Touch	12.8 X 18

Throughout this chapter, interoperable datasets were named using the notation {ES,TS} where ES refers to the source of the enrollment images and TS refers to the source of the test images. For example, {S1, S2} refers to S1 Fingerchip being the source of enrollment images and S2 being the source of test images.

Results of the survey filled out participants at beginning of the data collection session can be found in Appendix I.

4.1. Failure to Enroll (FTE)

Data acquisition was performed using the native acquisition software provided by the sensor manufacturers. All capture software except for S3 LC 300 used an auto capture acquisition mode and had an inbuilt quality assessment algorithm which made a determination of the acceptability of the fingerprint image. Failure to Enroll (FTE) was not based on acceptability of the fingerprint image by the native software but instead by VeriFinger 5.0. FTE was recorded for an individual if three fingerprint images could not be processed using VeriFinger 5.0 extractor. Determination of FTE was conducted in offline processing mode. S3 LC 300 could not to be used in auto capture acquisition mode. The determination of capturing the fingerprint image using S3 LC 300 was made by the test administrator based on subjective assessment of the clarity of the fingerprint image. Table 4.2 shows the number of FTE for each sensor. Fingerprints collected using the S5 sensor had the highest FTE out of all the sensors.

Table 4.2 Number of participants that recorded Failure to Enroll (FTE) using VeriFinger 5.0

Sensor	***N***	**FTE Rate**
Sensor 1	0	0%
Sensor 2	3	1.5%
Sensor 3	0	0%
Sensor 4	2	1.0%
Sensor 5	5	2.6%
Sensor 6	1	0.5%
Sensor 7	1	0.5%
Sensor 8	2	1.0%
Sensor 9	3	1.5%

4.2. Basic Fingerprint Feature Analysis

This section describes results of minutiae count analysis and image quality analysis performed on fingerprint images collected from all nine fingerprint sensors. The dataset effects were treated as the main effects for all statistical tests.

4.2.1. Minutiae Count Analysis

Three different software were used to extract the minutiae count: Aware Quality Software, MINDTCT, and VeriFinger 5.0 extractor. The results of these three different software were not compared to one another.

The preliminary test of assumptions for a single factor parametric F- test was performed on minutiae count generated by Aware Quality Software. A model adequacy check was performed by examining the residual values. A visual analysis showed that the normal probability plot, independence of samples, and constancy of variance did not violate the assumptions for performing the parametric F-test. These graphs can be found in Appendix E. Table 4.4 shows the values of mean (*M*) and standard deviation (*SD*) for each of the fingerprint datasets. Figure 4.1 shows the boxplot of minutiae count for all datasets and Figure 4.2 shows the histogram of minutiae count.

Table 4.3 Descriptive statistics for Aware software minutiae count analysis

Dataset	***n***	***M***	***SD***
S1	1140	72.91	18.16
S2	1122	31.72	11.61
S3	1140	39.93	13.67
S4	1128	39.48	10.57
S5	1110	29.16	8.92
S6	1134	46.50	10.83
S7	1134	32.68	9.16
S8	1128	35.39	10.39
S9	1122	47.65	11.57

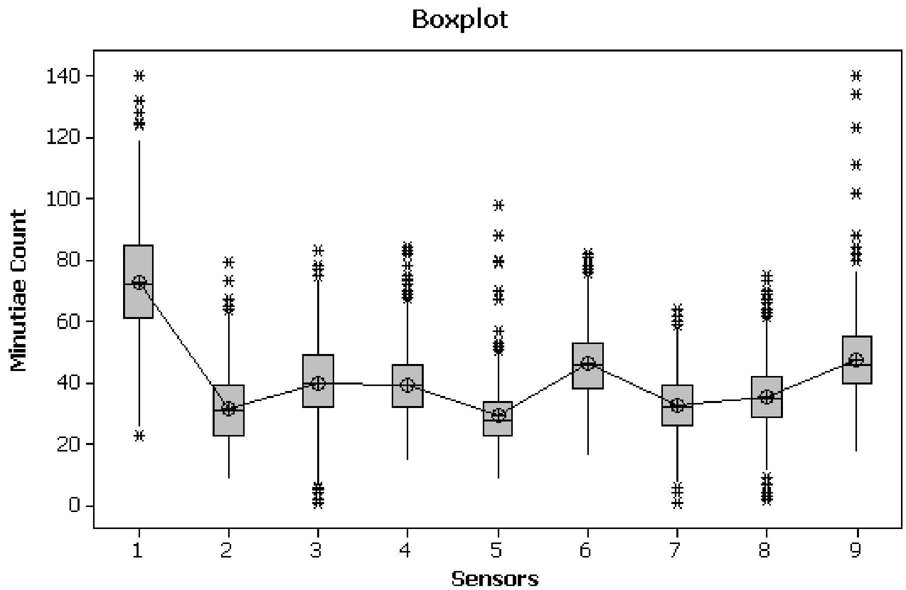

Figure 4.1 Boxplot of Aware software minutiae count

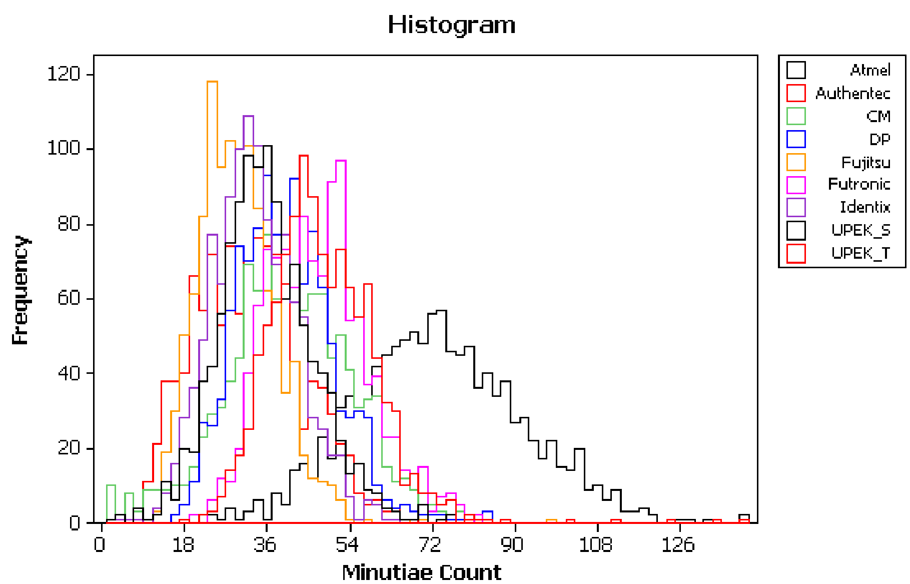

Figure 4.2 Histogram of Aware software minutiae count

The omnibus test for main effect of the sensor was statistically significant, $F(8, 10149) = 1403.79$, $p < 0.001$ at $\alpha = 0.05$. Tukey's Honestly Significant Difference (HSD) test for pairwise sensor effects was performed to determine statistical significance of difference between each possible pair in the group of datasets. For pairwise comparisons described in this section, if the p value was greater than 0.05, the comparison was not statistically significant. Table 4.4 gives the p values of each pairwise comparison. Tukey's Honestly Significant Difference (HSD) test of S2 and S7; S3 and S4; and S6 and S9 were found to be not statistically significant at $\alpha = 0.05$. The pairwise comparison of S2 and S7 was interesting because S2 was collected from a capacitive sensor of swipe interaction type, while S7 was collected from an optical sensor of touch interaction type. S6 was collected using an optical touch sensor and S9 was collected using a capacitive touch sensor. The sensor acquisition technology or interaction type did not have an impact on similarity of minutiae count. S3 and S4 were both collected using an optical touch sensor. Results of this test were symmetric due to the mechanics of the Tukey's HSD test.

Table 4.4 *p* values for Tukey's HSD test for pairwise dataset effects for Aware software minutiae count

	S2	S3	S4	S5	S6	S7	S8	S9
S1	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05
S2		<0.05	<0.05	<0.05	<0.05	>0.05	<0.05	<0.05
S3			>0.05	<0.05	<0.05	<0.05	<0.05	<0.05
S4				<0.05	<0.05	<0.05	<0.05	<0.05
S5					<0.05	<0.05	<0.05	<0.05
S6						<0.05	<0.05	>0.05
S7							<0.05	<0.05
S8								<0.05

Model adequacy checking was performed by examining residual values for minutiae count generated by MINDTCT module in NBIS. A visual analysis of the normal probability plot, independence of samples, and constancy of variance did not violate the assumptions for performing single factor parametric F- test. These graphs can be found in Appendix E. Table 4.5 shows the values of *M* and *SD* for each of the datasets. Figure 4.3 shows the boxplot of the minutiae count for all datasets.

Table 4.5 Descriptive Statistics for MINDTCT minutiae count

Dataset	***N***	***M***	***SD***
S1	1140	78.62	19.58
S2	1122	48.71	15.61
S3	1140	57.40	14.48
S4	1128	52.08	13.36
S5	1110	47.52	11.02
S6	1134	58.36	12.64
S7	1134	35.32	9.08
S8	1128	47.03	11.54
S9	1122	49.46	12.01

The omnibus test for main effect of sensor dataset was statistically significant, $F(8, 10149) = 851.0$, $p < 0.001$ at $\alpha = 0.05$. Tukey's HSD test was performed to determine the statistical significance of difference between each possible pair in the group of datasets. For pairwise comparisons described in this section, if the *p* value was greater than 0.05, the comparison was not statistically significant. Table 4.6 gives the *p* values of each pairwise comparison. Tukey's HSD test of S2 and S5; S2 and S8; S2 and S9; S3 and S6; and S5 and S8 were found to be not statistically significant at $\alpha = 0.05$. The interesting results were of the pairwise comparison of S2 and S5 and S2 and S9 because S2 was collected from a capacitive swipe sensor which S5 and S9 were collected using capacitive touch sensors. S8 which was collected using a capacitive swipe sensor showed a statistical similar count to every capacitive sensor except for S9. Results of this test were symmetric.

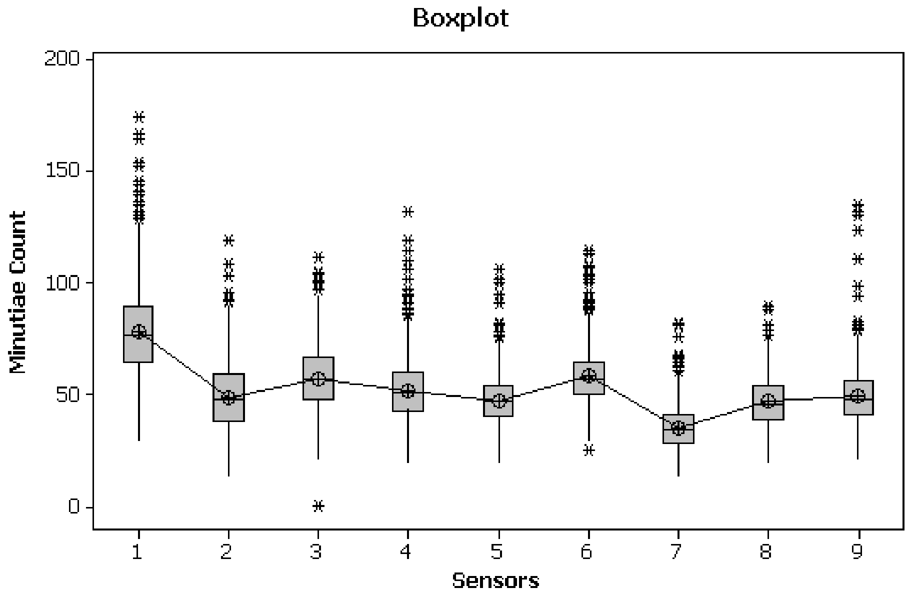

Figure 4.3 Boxplot of Minutiae Count Extracted by MINDTCT

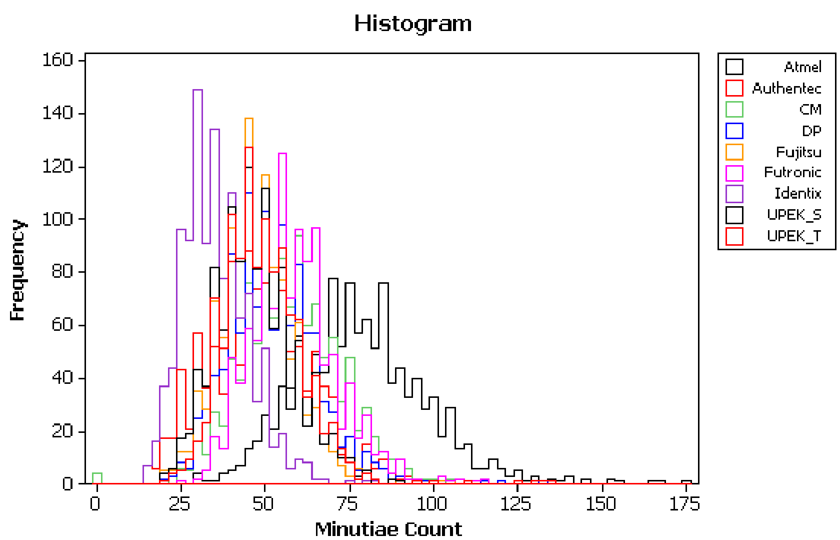

Figure 4.4 Histogram of Minutiae Count Extracted by MINDTCT

Table 4.6 *p* values for Tukey's HSD test for pairwise dataset effects for MINDTCT minutiae count

	S2	S3	S4	S5	S6	S7	S8	S9
S1	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05
S2		<0.05	<0.05	>0.05	<0.05	<0.05	>0.05	>0.05
S3			<0.05	<0.05	>0.05	<0.05	<0.05	<0.05
S4				<0.05	<0.05	<0.05	<0.05	<0.05
S5					<0.05	<0.05	>0.05	<0.05
S6						<0.05	<0.05	<0.05
S7							<0.05	<0.05
S8								<0.05

Model adequacy checking was performed by examining residual values for minutiae count generated by VeriFinger 5.0 extractor. A visual analysis of the normal probability plot, independence of samples, and constancy of variance did not violate the assumptions for performing single factor parametric F- test. Table 4.7 shows the values of *M* and *SD* for each of the datasets. Figure 4.5 shows the boxplot of minutiae count extracted by VeriFinger 5.0.

Table 4.7 Descriptive statistics for VeriFinger 5.0 minutiae count

Dataset	***n***	***M***	***SD***
S1	1140	41.72	10.57
S2	1122	28.32	10.73
S3	1140	40.25	10.12
S4	1128	30.74	8.06
S5	1110	24.38	6.87
S6	1134	38.62	9.18
S7	1134	27.53	7.69
S8	1128	26.11	6.69
S9	1122	26.15	6.73

The omnibus test for main effect of sensor dataset was statistically significant, $F(8, 10149) = 685.86$, $p < 0.001$ at $\alpha = 0.05$. Tukey's HSD test was performed to determine the statistical significance of difference between each possible pair in the group of datasets. For pairwise comparisons described in this section, if the *p* value was greater than 0.05, the comparison was not statistically significant. Table 4.8 gives the *p* values of each pairwise comparison. Tukey's HSD test of S8 and S8 was found to be not statistically significant at $\alpha = 0.05$. S8 was collected using a capacitive swipe sensor and S9 was collected using a capacitive touch sensor. Results of this test were symmetric.

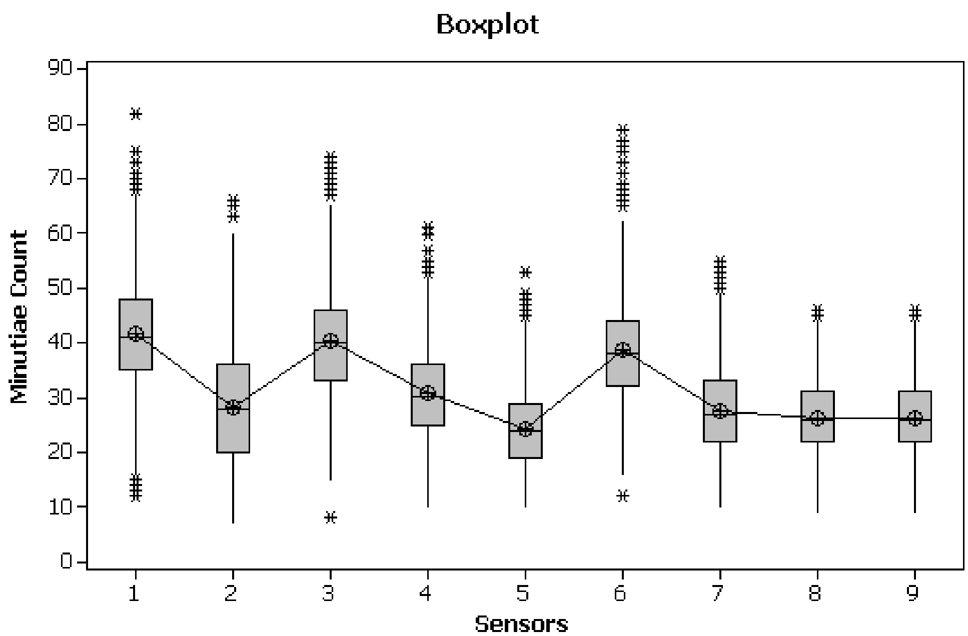

Figure 4.5 Boxplot of Minutiae Count Extracted by VeriFinger 5.0

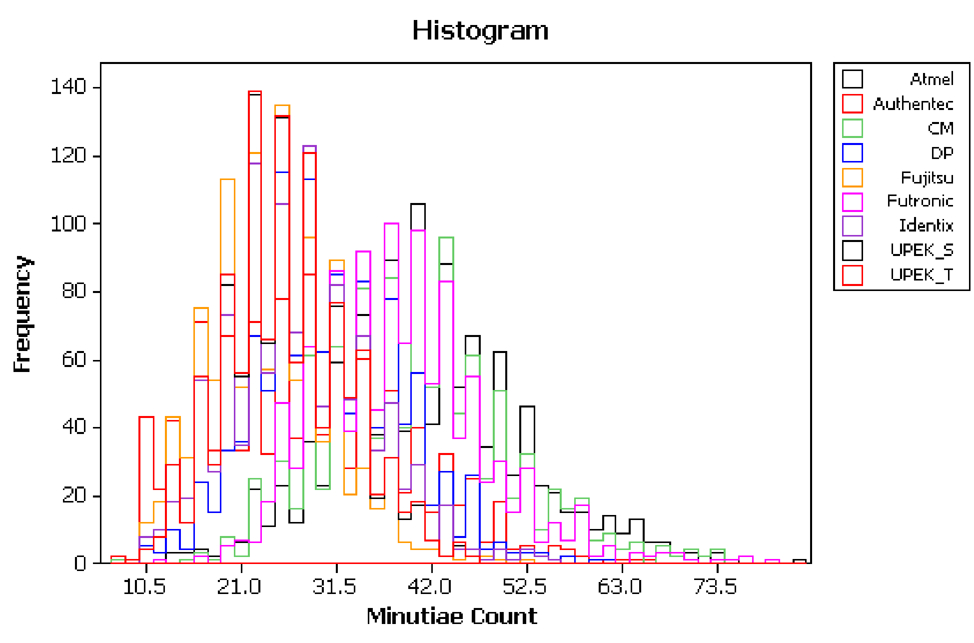

Figure 4.6 Histogram of Minutiae Count Extracted by VeriFinger 5.0

Table 4.8 *p* values for Tukey's HSD pairwise test for dataset effects

	S2	S3	S4	S5	S6	S7	S8	S9
S1	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05
S2		<0.05	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05
S3			<0.05	<0.05	<0.05	<0.05	<0.05	<0.05
S4				<0.05	<0.05	<0.05	<0.05	<0.05
S5					<0.05	<0.05	<0.05	<0.05
S6						<0.05	<0.05	<0.05
S7							<0.05	<0.05
S8								>.05

4.2.2. Image Quality Analysis

Two different software were used to extract the image quality scores: Aware Quality Software, and NFIQ. The results of these two different software were not compared to one another.

The descriptive statistics for image quality scores generated by Aware software are given in Table 4.9. Figure 4.7 shows the boxplot of quality scores generated by Aware software.

Table 4.9 Descriptive statistics for Aware software quality scores

Dataset	***N***	***M***	**Median**
S1	1140	44.49	45
S2	1122	79.52	81
S3	1140	68.76	75
S4	1128	72.64	75
S5	1110	79.92	82
S6	1134	71.02	73
S7	1134	75.88	80
S8	1128	76.58	80
S9	1122	68.43	70

The image quality scores generated by Aware software did not satisfy the assumptions for performing the parametric F-test. The model adequacy check of normality of residuals did not hold. Figure 4.9 shows the normality plot of the residuals.

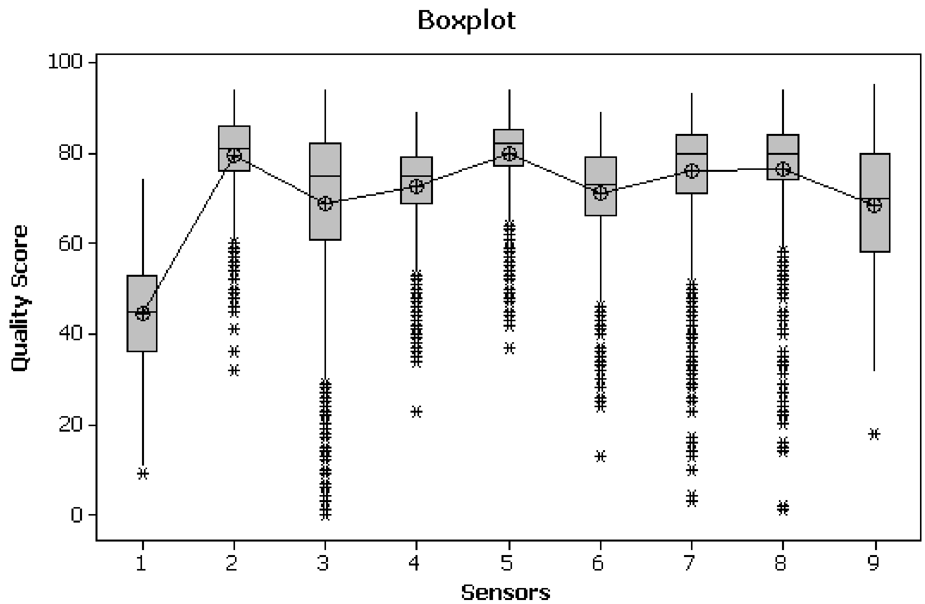

Figure 4.7 Boxplot of Image Quality Scores Computed by Aware

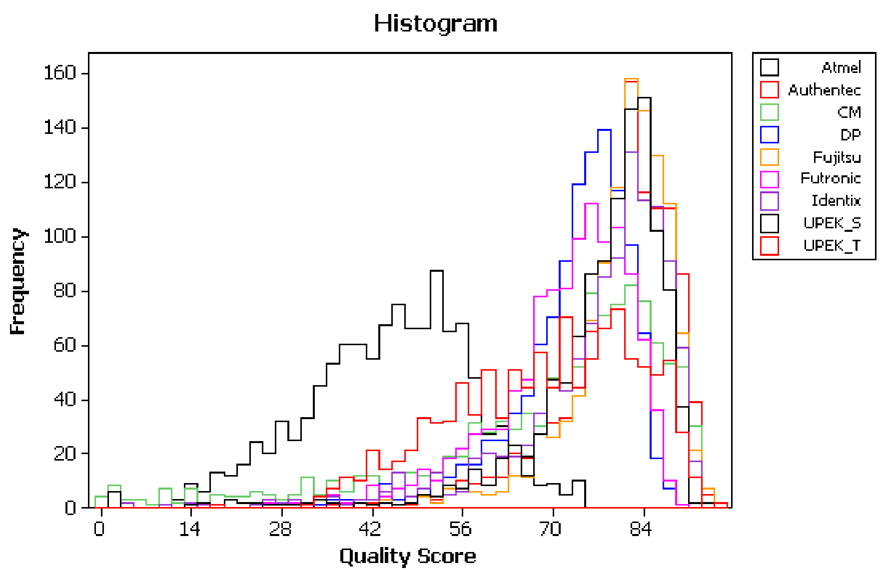

Figure 4.8 Histogram of Image Quality Scores Computed by Aware

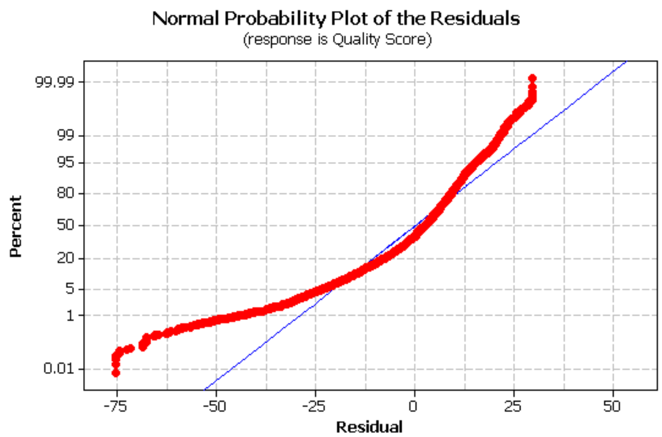

Figure 4.9 Normality plot of residuals

Instead an analysis of variance on rank of the response variable was performed using the Kruskal Wallis test (Tolley, 2006). The test for main effect of the sensor dataset was statistically significant, H (8) = 3379.77, $p < 0.001$ at $\alpha = 0.05$.

Follow up tests were performed on pairwise comparisons of datasets to determine which pairs of sensors were statistically significant in their differences. Tukey' HSD pairwise comparisons were performed on the ranks of observations for each dataset (Tolley, 2006). Table 4.11 gives the p values of each pairwise comparison. The test of pairwise comparison for S2 and S5; S3 and S4; S4 and S6; S6 and S9, and S7 and S8 were found to be not statistically significant at α = 0.05. Table 4.10 shows the different acquisition technology and interaction types of the sensors used to collect the datasets in the pairwise comparisons. The notation in the acquisition technology and interaction type columns correspond to the respective elements in the pairwise dataset column

Table 4.10 Summary of Sensor Acquisition Technologies and Interaction Types

Pairwise Dataset	**Acquisition Technology**	**Interaction Type**
(S2/S5)	(Capacitive/Capacitive)	(Swipe/ Touch)
(S3/S4)	(Optical/Optical)	(Touch/Touch)
(S4/S6)	(Optical/Optical)	(Touch/Touch)
(S6/S9)	(Optical/Capacitive)	(Touch/Touch)
(S7/S8)	(Optical/Capacitive)	(Touch/Swipe)

Each optical touch sensor dataset was found to have similar quality scores with dataset collected from another optical touch sensor or capacitive touch and capacitive swipe sensor, but none of the datasets showed similarity of scores with more than one other dataset. As a group, the optical touch sensors showed a higher level of similarity among their datasets compared to capacitive touch sensors. It should also be noted that the pairwise comparison results of minutiae count from Table 4.4 for S3 and S4, and S6 and S9 were found to be similar as well. Results of this test are symmetric.

Table 4.11 *p* values for test for difference in quality scores in every possible pairwise comparison using Aware Quality scores

	S2	S3	S4	S5	S6	S7	S8	S9
S1	<0.05	<0.05	<0.05	<0.05	<0.05	<.001	<0.05	<0.05
S2		<0.05	<0.05	>0.05	<0.05	<0.05	<0.05	<0.05
S3			>0.05	<0.05	<0.05	<0.05	<0.05	<0.05
S4				<0.05	>0.05	<0.05	<0.05	<0.05
S5					<0.05	<0.05	<0.05	<0.05
S6						<0.05	<0.05	>0.05
S7							>0.05	<0.05
S8								<0.05

Table 4.12 shows the descriptive statistics for image quality scores generated by NFIQ. A score of one indicates highest possible quality score and five indicates lowest possible quality score.

Table 4.12 Descriptive Statistics for NFIQ Image Quality Analysis

Dataset	***N***	***M***	**Median**
S1	1140	1.29	1
S2	1122	1.91	2
S3	1140	1.75	1
S4	1128	2.00	2
S5	1110	2.18	2
S6	1134	1.77	2
S7	1134	2.03	2
S8	1128	2.14	2
S9	1122	1.58	2

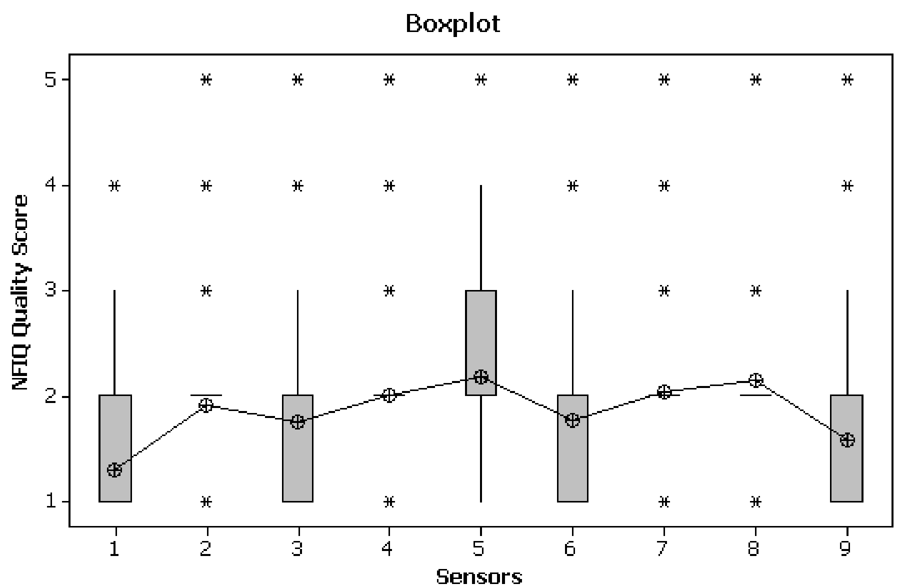

Figure 4.10 Boxplot of Image Quality Scores Computed by NFIQ

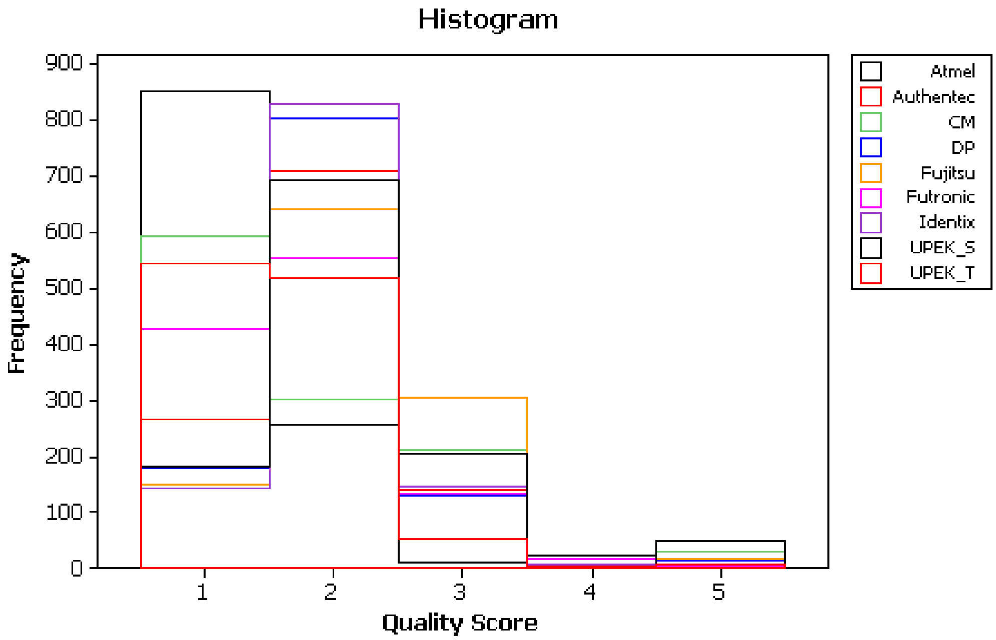

Figure 4.11 Histogram of Image Quality Scores Computed by NFIQ

NFIQ uses a 3-layer feed forward nonlinear perceptron model to predict the image quality values based on the input feature vector of the fingerprint image (Tabassi, Wilson, & Watson, 2004). Neural networks are non-parametric processors, which implies that the results produced have non-parametric characteristics (Shu-Long, Zhong-Kang, & Yan-Yan, 1991). This property of NFIQ values precludes the use of parametric based approach for detecting differences in quality scores between all fingerprint datasets. Instead an analysis of variance on rank of the response variable was performed using the Kruskal Wallis test. The test for main effect of the sensor was statistically significant, H (8) = 1646.07, $p < 0.001$ at $\alpha = 0.05$.

Follow up tests were performed on pairwise comparisons of sensor datasets to determine which pairs of sensors were statistically significant in the differences of NFIQ scores. Tukey's HSD pairwise comparisons were performed on the ranks of the observations for each dataset. If the p value was greater than 0.05 for any of the pairwise comparisons, the comparison was not statistically significant. Table 4.13 gives the *p* value of each pairwise comparison. The test of pairwise comparison for S4 and S7; S4 and S8; S5 and S9; and S7 and S9 was found to be not statistically significant at $\alpha = 0.05$. S4 and S7 datasets were both collected from optical sensors and touch interaction type sensors. S7 dataset was collected from an optical touch sensor and S8 was collected from a capacitive swipe sensor. It should be noted that S5 and S8 also showed no statistically significant difference in minutiae count extracted by MINDTCT (Table 4.6). S8 showed a similarity of quality scores with the most number of other datasets, and the other datasets were of both optical touch and capacitive touch types. As a group, neither optical touch sensors nor capacitive touch sensors showed a high level of similarity of quality scores. Results of this test are symmetric.

Table 4.13 *p* values for test for difference in quality scores in every possible pairwise comparison using NFIQ scores

	S2	S3	S4	S5	S6	S7	S8	S9
S1	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
S2		<.001	<.001	<.001	<.001	<.001	<.001	<.001
S3			<.001	<.001	<.001	<.001	<.001	<.001
S4				<.001	<.001	>0.05	>0.05	<.001
S5					<.001	<.001	>0.05	<.001
S6						<.001	<.001	<.001
S7							>0.05	<.001
S8								<.001

4.3. Match Score Analysis

This section contains the match score results obtained using VeriFinger 5.0 and BOZORTH3, and the subsequent analysis performed on the match score results.

4.3.1. VeriFinger 5.0 Interoperability Error Rates

The interoperability FNMR matrices are shown in Table 4.14, and Table 4.15. The cells along the diagonal indicate enrollment and test fingerprint images from the same fingerprint sensor. The cells off the diagonal indicate fingerprint images from different sensors. The error rates matrix generation methodology was described in Sections 3.10 and 3.12.2. The sensor dataset in the rows indicate the source of the enroll sensor and the sensor dataset in the columns indicate the source of the test sensor. All the native FNMR were found to be lower than interoperable datasets except for the interoperable dataset {S5,S9} where S5 and S9 were both collected from a capacitive touch type sensor. The interoperable FNMR for {S5, S9} = 0.55% and native FNMR for S5 = 1.02% as shown in Table 4.14. The dataset for S9 had a mean minutiae count of 26.15 while dataset for S5 had a mean minutiae count of 24.38 using VeriFinger 5.0 extractor. This result indicated the interoperable dataset {S5, S9} had a larger number of minutiae points to match compared to native dataset of S5. The capture area of the sensor used to for S9 dataset was larger compared to capture area used for S5 dataset. Also interesting was the extremely high FNMR for interoperable S8 datasets. S8 dataset was collected from a capacitive sensor of swipe interaction type and had an extremely high FNMR with datasets collected from all different types of sensors. A cross reference analysis of the FNMR matrix with similarity of minutiae count and quality scores from the previous section did not show any specific relations.

Table 4.14 FNMR at FMR 0.01% (in percentage values)

	TEST									
		S1	S2	S3	S4	S5	S6	S7	S8	S9
ENROLL	S1	0.64	8.70	7.63	6.15	7.72	44.42	15.07	100	5.19
	S2	10.61	5.58	8.43	9.30	13.81	20.62	13.42	100	7.44
	S3	11.61	6.18	0.30	1.46	1.95	4.56	2.44	100	0.90
	S4	7.29	5.40	1.31	0.23	1.70	2.13	1.95	100	1.26
	S5	9.32	9.67	1.16	1.03	1.02	6.54	3.02	100	0.55
	S6	46.14	17.92	5.33	1.54	6.89	0.17	2.18	100	2.18
	S7	18.99	11.63	2.50	2.42	3.68	3.54	1.29	100	2.56
	S8	100	99.69	100	100	100	100	100	1.30	100
	S9	4.90	3.39	0.24	1.38	0.73	2.57	1.19	100	0.11

Table 4.15 FNMR at FMR 0.1% (in percentage values)

		TEST								
		S1	S2	S3	S4	S5	S6	S7	S8	S9
ENROLL	S1	0.47	6.79	5.60	4.44	5.61	31.03	11.30	99.81	3.76
	S2	7.33	5.05	6.49	7.44	10.92	16.18	10.83	99.81	6.11
	S3	8.57	4.78	0.24	1.07	1.28	2.73	1.79	100	0.54
	S4	5.10	4.44	0.95	0	1.39	1.60	1.83	100	1.08
	S5	5.56	7.64	0.85	0.85	0.78	4.42	2.23	100	0.42
	S6	33.86	13.42	2.99	1.30	4.47	0.17	1.41	100	2.15
	S7	14.92	9.89	1.73	2.12	2.35	2.71	0.94	100	1.85
	S8	100	99.63	100	100	100	100	100	1.00	99.93
	S9	2.87	2.96	0.12	0.90	0.49	1.73	0.77	100	0.11

4.3.1.1. Core Overlap

Table 4.16 summarizes the percentage of pairs used in genuine comparisons that had cores present in both images of the pair. A higher number indicated a larger number of pairs of images in which a core was detected.

A scatter plot (Figure 4.12) was created where the x-axis represented the percentage of fingerprint image pairs with core overlap from each interoperable dataset and the y-axis represented its corresponding interoperable FNMR at fixed FMR of 0.1%. The scatter plot only contained interoperable FNMR and interoperable dataset core overlap percentage since interoperable FNMR were the data points of interest. An inverse relation between FNMR and percentage of pairs with core overlap was observed. There were two data points of interest which represent interoperable datasets {S1,S6} and {S6,S1}. The relation between percentage of images with cores and FNMR of these two interoperable datasets did not follow the trend observed in the other interoperable datasets, where a higher percentage of core overlap between images was related to a lower FNMR. S1 was collected using thermal swipe sensor and S6 was collected using optical touch sensor. These points indicate existence of other underlying factors which affected the interoperable FNMR, which are discussed in Section 4.7.

Table 4.16 Percentage of fingerprint pairs with detected core

		TEST								
		S1	S2	S3	S4	S5	S6	S7	S8	S9
ENROLL	S1	75.32	68.93	77.16	76.02	71.15	77.75	74.08	70.69	77.43
	S2	65.53	68.80	72.79	71.95	67.34	73.74	71.62	70.69	72.35
	S3	78.79	77.15	88.77	86.41	81.51	88.84	85.06	84.07	87.68
	S4	78.23	77.29	86.89	86.99	81.11	89.18	84.69	82.91	86.12
	S5	72.27	71.59	80.82	80.83	77.03	81.93	79.37	76.91	79.79
	S6	78.65	77.27	87.23	86.86	80.45	90.03	84.63	82.32	86.29
	S7	75.57	75.29	84.05	83.33	78.68	85.34	82.71	80.92	83.63
	S8	75.31	74.00	82.60	81.38	76.12	82.88	79.22	81.96	82.09
	S9	78.39	76.42	86.58	84.86	79.88	86.81	83.63	82.99	86.42

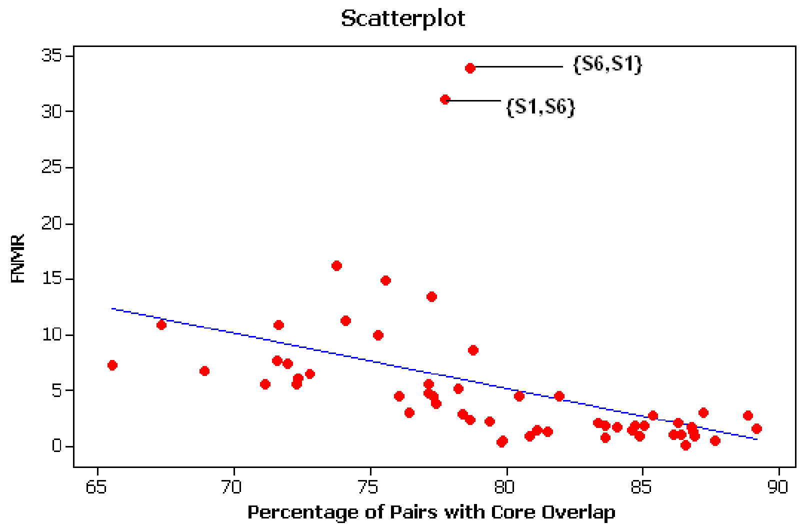

Figure 4.12 Scatter plot of Consistency of Placement vs. FNMR (FMR=0.1%)

4.3.2. NBIS Interoperability Error Rates

The interoperable FNMR for match scores generated by the NBIS matcher is shown in Table 4.17. As described in Section 3.4.2.2, a single performance interoperability matrix was generated using a decision threshold of 40.

Table 4.17 FNMR (at threshold of score of 40)

		TEST								
		S1	S2	S3	S4	S5	S6	S7	S8	S9
ENROLL	S1	1.71	21.71	10.77	11.07	23.04	20.22	21.07	60.54	9.98
	S2	27.22	19.50	23.24	20.78	29.59	31.72	27.51	63.33	20.51
	S3	12.56	18.40	2.98	5.62	11.29	10.64	9.68	60.15	4.69
	S4	11.25	16.02	4.24	1.95	9.18	7.03	6.65	59.99	2.72
	S5	24.12	25.34	10.43	8.24	11.06	18.75	13.28	62.16	8.42
	S6	20.78	26.65	8.98	6.18	18.86	0.84	11.11	60.37	5.85
	S7	23.78	24.26	10.14	8.78	15.29	13.23	8.28	62.01	7.49
	S8	60.89	61.54	60.43	60.13	61.54	60.90	61.58	4.18	59.85
	S9	9.63	14.32	3.23	2.93	8.27	5.53	5.96	60.00	1.28

Native FNMR for S5 was higher compared to interoperable FNMR for datasets {S5,S3}, {S5,S4}, and {S5,S9} where S5 was collected from capacitive touch type sensor and S3 was also collected from an optical touch type sensor, S4 was collected from an optical touch sensor, and S9 was collected from a capacitive touch sensor. Native FNMR for S7 was also higher compared to interoperable FNMR for datasets {S7,S9}. S7 was collected using an optical touch sensor and S9 was collected using a capacitive touch sensor.

Fingerprints from S8 showed a very high FNMR with interoperable datasets but its native dataset FNMR was not the highest native dataset FNMR in the matrix. A similar relation was observed in FNMR matrix generated using VeriFinger 5.0 matcher (Table 4.14 and Table 4.15).

4.3.3. Test of Proportions for VeriFinger 5.0 Match Scores

The test of homogeneity of proportions using the χ^2 distribution was performed for comparing FNMR of the native dataset to FNMR of interoperable dataset. FNMR calculated at fixed FMR 0.1% was used for this test. Significance level of 0.05 was used. This test was performed for the hypothesis stated in Eq. 3.5, which has been restated in Eq.4.1. For pairwise comparisons described in this section, if the *p* value was greater than 0.05, the comparison was not statistically significant.

$$H1_0: p_i = p_2 = \ldots\ldots = p_n \qquad \text{Eq. 4.1}$$
$$H1_A: p_i \neq p_2 \neq \ldots\ldots \neq p_n$$

For S1 dataset the FNMR was compared to all other interoperable datasets of S1, and the test was significant at $\chi^2(8, N = 15182) = 8938.87$, $p < .001$. Since the null hypothesis was rejected the Marascuillo procedure, described in Section 3.4.2.2, was used to simultaneously test differences of all interoperable datasets with S1 native dataset. The results of the Marascuillo procedure can be interpreted in the same way as Tukey's HSD pairwise test, where the results of the test indicate which interoperable datasets within the group are statistically significant in their differences compared to the native dataset. The results for pairwise comparison with S1 are shown in Table 4.19.

This overall test of homogeneity of proportions was performed one at a time for each native dataset S2, S3, S4, S5, S6, S7, S8 and S9 and their respective interoperable datasets. The results of the overall test are given in Table 4.18.

Table 4.18 Results of Overall Test of Proportions

Native Dataset	**χ^2 Statistic**	***p* value**
S2	χ^2 (8, N =14595) = 8128.5	$p < 0.001$
S3	χ^2 (8, N =14814) = 12101.87	$p < 0.001$
S4	χ^2 (8, N =15075) = 12742	$p < 0.001$
S5	χ^2 (8, N =14783) = 11880	$p < 0.001$
S6	χ^2 (8, N =14985) = 9837	$p < 0.001$
S7	χ^2 (8, N =15135) = 10919	$p < 0.001$
S8	χ^2 (8, N =14987) = 14747	$p < 0.001$
S9	χ^2 (8, N =14944) = 13486	$p < 0.001$

Table 4.18 showed that all tests were statistically significant in their differences which indicated that the FNMR calculated for the interoperable matrix in Table 4.15 were all different from their respective native dataset FNMR. Since the null hypothesis was rejected for all overall tests of proportions, the Marascuillo procedure was used to simultaneously test the differences of all pairs. The results for pairwise comparison for each native dataset are shown in Table 4.19. The first cell in the rows in Table 4.19 indicated the native dataset and the remaining cells in that row indicated its corresponding interoperable datasets. A *p* value of less than 0.05 indicated a statistically significant difference. All pairwise comparisons for the native S1 and S8 datasets were statistically different. S1 was collected using a thermal swipe sensor and S8 was collected using a capacitive swipe sensor. S1 dataset showed a statistically significant difference

in all pairwise comparisons of minutiae count and image quality scores as well. The other pairwise comparisons test showed an even distribution of similarity of FNMR with interoperable datasets without any apparent trend among acquisition technologies and interaction types of the sensors that the datasets were collected from. The test of {S2, S1} was interesting since it did not show a difference in the pairwise test of proportions but showed a difference in minutiae count and image quality scores for all software used. S2 was collected a capacitive swipe sensor and S7 was collected using an optical touch sensor. The test of {S3,S4} was interesting since it did not show a difference in pairwise test of proportions, and also did not show a difference in minutiae count and image quality scores. S3 and S4 were both collected using optical touch sensors. These two tests indicated that impact of minutiae count similarity and image quality score similarity was not consistent on the pairwise test of proportions. Interoperable datasets which contained S9 as its second dataset showed a high level of similarity to the native datasets which were used to create the interoperable dataset. This indicated that S9 did not degrade the performance of an interoperable dataset compared to the performance of native datasets.

Table 4.19 *p* values of pairwise test of proportions

Native Dataset	Pairwise Dataset for Comparison								
	S1	S2	S3	S4	S5	S6	S7	S8	S9
S1		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
S2	>.05		>.05	>.05	<.001	<.001	<.001	<.001	>.05
S3	<.001	<.001		>.05	>.05	<.001	<.001	<.001	>.05
S4	<.001	<.001	>.05		>.05	<.001	<.001	<.001	>.05
Fujtisu	<.001	<.001	>.05	>.05		<.001	>.05	<.001	>.05
S6	<.001	<.001	<.001	>.05	<.001		>.05	<.001	>.05
S7	<.001	<.001	>.05	>.05	>.05	>.05		<.001	>.05
S8	<.001	<.001	<.001	<.001	<.001	<.001	<.001		<.001
S9	<.001	<.001	>.05	>.05	>.05	<.001	>.05	<.001	

4.3.4. Test of Proportions for BOZORTH3 Match Scores

The test of homogeneity of proportions using the χ^2 distribution was performed for FNMR generated using BOZORTH3. Significance level of 0.05 was used. This test was performed for the hypothesis stated in Eq. 3.5, which was restated in Eq. 4.2. For pairwise comparisons described in this section, if the *p* value was greater than 0.05, the comparison was not statistically significant.

$$H1_0: p_i = p_2 = \ldots\ldots = p_n \qquad \text{Eq. 4.2}$$

$$H1_A: p_i \neq p_2 \neq \ldots\ldots \neq p_n$$

For S1 dataset, the test was significant at the $\chi^2(8, N = 15192) = 4874$, $p < .001$. Since the null hypothesis was rejected, the Marascuillo procedure, described in Section 3.4.2.2, was used to simultaneously test differences of all interoperable datasets with S1. The results of the Marascuillo procedure can be interpreted in the same way as Tukey's HSD pairwise test, where the results of the test indicate which specific pairs within the group are statistically significant in their differences. The results for pairwise comparison with S1 are shown in Table 4.21.

The overall test of homogeneity of proportions was performed one at a time for each native dataset S2, S3, S4, S5, S6, S7, S8 and S9 and their respective interoperable datasets. The results of the overall test are given in Table 4.20.

Table 4.20 Results of overall test of proportions

Native Dataset	χ^2 Statistic	*p* value
S2	χ^2 (8, N =15021) =2921	$p < 0.001$
S3	χ^2 (8, N =15237) = 6517	$p < 0.001$
S4	χ^2 (8, N =15086) = 7742	$p < 0.001$
S5	χ^2 (8, N =14886) = 5270	$p < 0.001$
S6	χ^2 (8, N =15165) = 5941	$p < 0.001$
S7	χ^2 (8, N =15174) = 5595	$p < 0.001$
S8	χ^2 (8, N =15102) = 12755	$p < 0.001$
S9	χ^2 (8, N =15008) = 8326	$p < 0.001$

Table 4.21 showed that all tests were statistically significant in their differences which indicated that the FNMR calculated for the interoperability matrix in Table 4.17 were all different from their respective native dataset FNMR. Since the null hypothesis was rejected for all overall tests of proportions, the Marascuillo procedure was used to simultaneously test the differences of all pairs. The results for pairwise comparison for each native dataset are shown in Table 4.21.

The first cell in the rows in Table 4.21 indicated the native dataset and the remaining cells in that row indicated its corresponding interoperable datasets. S5 and S7 exhibited the highest number of interoperable datasets with FNMR similar to their respective native datasets. S5 was collected using a capacitive touch sensor and S7 was collected using an optical touch sensor. The pairwise comparison of interoperable {S5,S3}, {S5,S4}, {S5, S9} with S5, and pairwise comparisons of interoperable {S7,S3}, {S7,S4}, {S7,S9} with S9 were not statistically different. The interoperable datasets created with S3, S4, and S9 were common to both tests. S3 and S4 were both optical touch sensors and S9 was a capacitive touch sensor.

Table 4.21 *p* values of pairwise test of proportions

Native Sensor		Pairwise Sensor for Comparison								
		S1	S2	S3	S4	S5	S6	S7	S8	S9
S1			<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
S2		<.001		<.001	<.001	<.001	<.001	<.001	<.001	<.001
S3		<.001	<.001		<.001	<.001	<.001	<.001	<.001	>.05
S4		<.001	<.001	<.001		<.001	<.001	<.001	<.001	>.05
S5		<.001	<.001	>.05	>.05		<.001	<.001	<.001	>.05
S6		<.001	<.001	<.001	<.001	<.001		<.001	<.001	<.001
S7		<.001	<.001	>.05	>.05	<.001	<.001		<.001	>.05
S8		<.001	<.001	<.001	<.001	<.001	<.001	<.001		<.001
S9		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	

4.3.5. Test of Similarity of VeriFinger 5.0 Match Scores

This section contains results for test of similarity of the genuine and imposter match scores generated by VeriFinger 5.0 for native and interoperable datasets. These match scores were the raw scores computed by VeriFinger 5.0. The test was performed by running nine different similarity tests. For each test, one native dataset was chosen as the control dataset and the rest of its interoperable datasets were compared to the control dataset.

The first test of similarity was performed by choosing S1 dataset as the native dataset and comparing its raw genuine match scores to the corresponding raw match scores from the other eight interoperable datasets. The residuals of genuine match scores generated for this test did not satisfy the model adequacy tests for performing a parametric F-test. Figure 4.13 shows the normality plot of residuals for S1 dataset genuine match scores and its corresponding interoperable dataset genuine match scores.

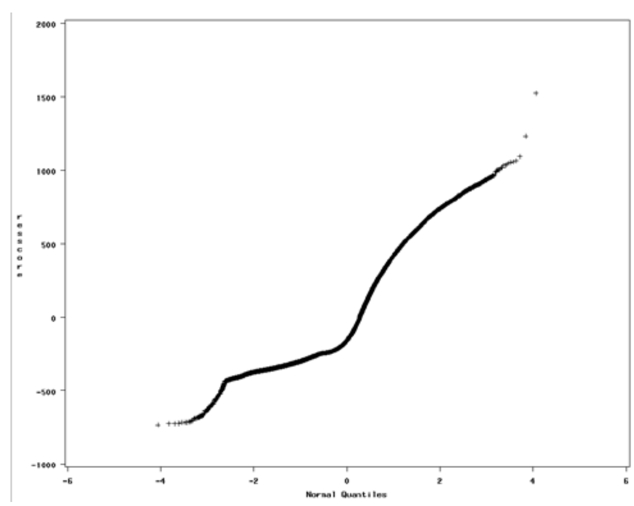

Figure 4.13 Normality plot of residuals of S1 genuine match scores

The Kruskal-Wallis test for main effect of the sensor was found to be statistically significant, H (8) = 4009, $p < 0.001$ at $\alpha = 0.05$. Follow up tests performed on pairwise comparison of every interoperable dataset with the native dataset showed that the test was significant with $p < 0.001$ at $\alpha = 0.05$. The same tests were performed one at a time for remaining native datasets. The normality plots for other datasets can be found in Appendix H. Table 4.22 shows the results of all tests. The *p* value for each test showed that the differences were statistically significant and the null hypothesis was rejected. Follow up tests were performed by comparing each pair of interoperable dataset with its corresponding native dataset. All pairwise comparisons were found to be statistically significant in their differences.

Table 4.22 Results of overall test of similarity of genuine match scores

Native Dataset	Kruskal Wallis Test	*p* value
S2	$H\ (8) = 2209$	$p < 0.001$
S3	$H\ (8) = 9716$	$p < 0.001$
S4	$H\ (8) = 9476$	$p < 0.001$
S5	$H\ (8) = 4128$	$p < 0.001$
S6	$H\ (8) = 4684$	$p < 0.001$
S7	$H\ (8) = 4111$	$p < 0.001$
S8	$H\ (8) = 2369$	$p < 0.001$
S9	$H\ (8) = 4885$	$p < 0.001$

A test of similarity was performed on raw match scores of imposter comparisons for S1 dataset and its corresponding interoperable datasets. The raw match scores generated for this test did not satisfy the model adequacy tests for performing a parametric F-test. The normality plot is shown in Figure 4.14. The Kruskal-Wallis test for main effect of the sensor was found to be statistically significant, $H\ (8) = 80685$, $p < 0.001$ at $\alpha = 0.05$. Follow up tests were performed on pairwise comparison of every S1 interoperable dataset with the S1 native dataset showed that the test was statistically significant with $p < 0.001$ at $\alpha = 0.05$. The tests did not show any similarity of raw match scores of genuine comparisons between native and interoperable datasets.

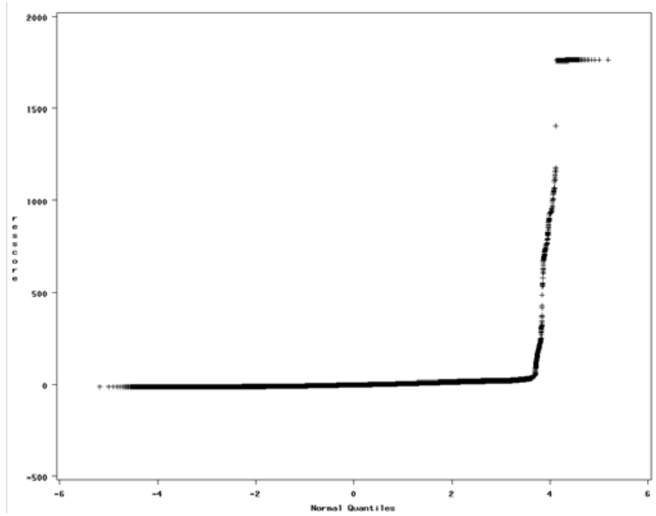

Figure 4.14 Normality plot of residuals of S1 imposter match scores

The same tests were performed one at a time for remaining native datasets. The normality plots for other datasets can be found in Appendix G. Table 4.23 shows the results of the tests. The *p* value for each test showed that the differences were statistically significant and the null hypothesis was rejected for each test. The pairwise comparison for each native dataset was also found to be statistically significant. The tests did not show a similarity of raw match scores of imposter comparisons between native and interoperable datasets.

Table 4.23 Results of overall test of similarity of imposter match scores

Native Dataset	Kruskal Wallis Test	*p* value
S2	H (8) = 71246	$p < 0.001$
S3	H (8) = 99085	$p < 0.001$
S4	H (8) = 112211	$p < 0.001$
S5	H (8) = 122373	$p < 0.001$
S6	H (8) = 163290	$p < 0.001$
S7	H (8) = 161126	$p < 0.001$
S8	H (8) = 240592	$p < 0.001$
S9	H (8) =129842	$p < 0.001$

4.4. Impact of Quality on Interoperability

This section describes the impact of removing low quality fingerprint images and recalculating the interoperable FNMR. The fingerprint images which had NFIQ quality scores of four and five were not used for the matching operation. NFIQ provides five categories of quality scores and experiments have been performed by NIST to predict the impact of NFIQ quality scores on matching operations. Image quality scores from Aware Quality Software provided scores on a scale of 1-100 and would have required an arbitrary decision threshold for removal of fingerprint images. The NIST MINEX test report indicated that a reduction in interoperability FNMR would be observed since poorly performing images were not used (Grother et al., 2006). The analysis performed in this section examined the impact of removing low quality images on variance of interoperability FNMR. VeriFinger 5.0 template generator and matcher were used to perform the matching operations. Note that dataset S8 was not included in this test since there was no impact of changing the fixed FMR operational points on the interoperable FNMR and the average quality of the fingerprint images from S8 was not the worst of all the native datasets. Section 4.7 analyzes the potential factors which affected the interoperable FNMR for S8. Table 4.24 shows the number of genuine comparisons performed on full datasets and Table 4.25 shows the number of genuine comparisons performed on datasets with fingerprint images of NFIQ scores of less than four.

Table 4.24 Number of comparisons in full datasets

	TEST								
		S1	S2	S3	S4	S5	S6	S7	S9
ENROLL	S1	1698	1677	1677	1689	1657	1695	1698	1675
	S2	1677	1683	1647	1665	1629	1668	1311	1653
	S3	1679	1650	1666	1671	1634	1533	1674	1649
	S4	1686	1665	1671	1692	1647	1683	1692	1665
	S5	1652	1623	1632	1644	1655	1650	1653	1628
	S6	1701	1668	1536	1683	1653	1701	1692	1671
	S7	1695	1668	1674	1692	1656	1692	1701	1674
	S9	1671	1650	1647	1665	1631	1668	1674	1680

Table 4.25 Number of comparisons for datasets with images of NFIQ < 4

		TEST							
		S1	S2	S3	S4	S5	S6	S7	S9
	S1	1655	1636	1652	1659	1623	1668	1668	1647
E	S2	1639	1660	1626	1635	1599	1644	1293	1635
N	S3	1628	1595	1624	1620	1593	1500	1633	1611
R	S4	1645	1616	1635	1650	1612	1653	1653	1632
O	S5	1626	1590	1620	1617	1627	1635	1635	1614
L	S6	1651	1618	1504	1647	1620	1648	1662	1641
L	S7	1654	1621	1643	1651	1626	1666	1667	1644
	S9	1667	1644	1647	1662	1626	1665	1671	1660

Table 4.26 FNMR at FMR 0.1% (in percentage values)

		TEST							
		S1	S2	S3	S4	S5	S6	S7	S9
ENROLL	S1	0.001	0.050	0.038	0.033	0.038	0.267	0.097	0.028
	S2	0.055	0.046	0.057	0.060	0.098	0.150	0.098	0.051
	S3	0.065	0.033	0	0.008	0.006	0.022	0.007	0.002
	S4	0.037	0.036	0.007	0.001	0.01	0.014	0.011	0.007
	S5	0.046	0.066	0.005	0.006	0.004	0.040	0.015	0.003
	S6	0.306	0.116	0.019	0.011	0.032	0	0.012	0.012
	S7	0.119	0.084	0.007	0.007	0.012	0.019	0.004	0.007
	S9	0.025	0.026	0	0.007	0.004	0.013	0.007	0.001

Table 4.26 shows the interoperability FNMR matrix for the reduced datasets at fixed FMR of 0.1%. Note that the interoperable FNMR for {S1,S6} increased to 0.267% compared to the full dataset interoperability FNMR of 0.25%. S1 dataset was collected using a thermal swipe sensor and S6 dataset was collected using an optical touch sensor. This is interesting since all the other FNMR reduced compared to FNMR calculated for full datasets.

The tests performed in Section 4.4.3 were repeated on interoperable fingerprint datasets which contained fingerprint images with NFIQ score of one, two or three. Table 4.27 shows the results for overall test of proportions performed on each native dataset. All the tests were found to be significant in their differences at $\alpha = 0.05$.

Table 4.27 Results of overall test of proportions

Native Dataset	**χ^2 Statistic**	***p* value**
S1	χ^2 (8, *N* =14793) = 9231.49	$p < 0.001$
S2	χ^2 (8, *N* =14039) = 8233.23	$p < 0.001$
S3	χ^2 (8, *N* =14356) = 12297.23	$p < 0.001$
S4	χ^2 (8, *N* =14666) = 12755.65	$p < 0.001$
S5	χ^2 (8, *N* =14543) = 11923.89	$p < 0.001$
S6	χ^2 (8, *N* =14569) = 9819.06	$p < 0.001$
S7	χ^2 (8, *N* =14753) = 11380.38	$p < 0.001$
S9	χ^2 (8, *N* =14847) = 13461.69	$p < 0.001$

The results from pairwise comparisons are shown in Table 4.28. The recalculated pairwise comparisons showed that the test of proportions for {S3,S7} dataset did not show any statistically significant difference while in Section 4.4.3 the test of proportion for the datasets showed a statistically significant difference. S3 and S7 both collected using optical touch sensors. This test indicated that removing the low quality images reduced the difference of variance of FNMR between these datasets.

The recalculated results for {S6,S4}, {S6,S7}, and {S6,S9} datasets showed a statistically significant difference. The results from Section 4.4.3 of the same test showed no statistically significant difference between the native and interoperable datasets. This result was interesting as it indicated that difference of variance of FNMR increased for these interoperable datasets when the lowest NFIQ score images were removed. It should also be noted that S6, S4, and S7 were all optical touch sensors. Removal of low quality images did not have a consistent effect on FNMR of interoperable datasets which were all collected using optical touch sensors.

The rows in Table 4.28 indicate the native sensor dataset and each row indicates the results of the pairwise comparison with remaining sensor datasets with the native dataset.

Table 4.28 Results of pairwise test of proportions

Native Sensor		Pairwise Sensor for Comparison							
		S1	S2	S3	S4	S5	S6	S7	S9
S1			<.001	<.001	<.001	<.001	<.001	<.001	<.001
S2		>.05		>.05	>.05	<.001	<.001	<.001	>.05
S3		<.001	<.001		>.05	>.05	<.001	>.05	>.05
S4		<.001	<.001	>.05		>.05	<.001	>.05	>.05
Fujtisu		<.001	<.001	>.05	>.05		<.001	>.05	>.05
S6		<.001	<.001	<.001	<.001	<.001		<.001	<.001
S7		<.001	<.001	>.05	>.05	>.05	>.05		>.05
S9		<.001	<.001	>.05	>.05	>.05	<.001	>.05	

4.5. Acquisition and Interaction Level Interoperability

A follow up evaluation of interoperability was performed by grouping the datasets into three categories: datasets collected using swipe interaction type sensors, datasets collected using optical touch type sensors, and datasets collected using capacitive touch type sensors. The evaluation of different groups allowed for examination of interoperability at the acquisition and interaction level, not the sensor level. S1 and S2 were placed in the Swipe group. S3, S4, S6 and S7 were placed in the Optical Touch group. S5 and S9 were placed in the Capacitive Touch group. S8 dataset was excluded from this analysis as it showed an extremely high interoperability FNMR with the rest of the datasets (Table 4.15). VeriFinger 5.0 was used to generate the interoperability FNMR matrix at fixed FMR of 0.1%.

Table 4.29 FNMR at FMR 0.1% (in percentage values)

		TEST		
		Swipe	Optical Touch	Capacitive Touch
ENROLL	Swipe	4.81	11.75	6.58
	Optical Touch	11.93	1.53	1.89
	Capacitive Touch	4.76	1.47	0.45

The {Capacitive Touch, Optical Touch} interoperable dataset showed the lowest interoperable FNMR indicating a high level of interoperability between optical touch and capacitive touch technologies. The {Capacitive Touch, Swipe} interoperable dataset had a lower FNMR than the Swipe native dataset which indicated that the interoperable dataset performed better than the native dataset. The {Swipe, Optical Touch} dataset had the highest FNMR and indicated the lowest degree of interoperability. The results showed that Capacitive Touch dataset had the lowest interoperable FNMR compared to Optical Touch and

Swipe interoperable datasets. The interoperable datasets generated with S9 dataset showed a high level of similarity of FNMR with the other native datasets. Since S9 was part of the Capacitive Touch group it showed a better interoperable FNMR with other groups. In combination with the sensor level interoperability analysis, these results show the effect of interoperability at the acquisition and interaction level.

4.6. Investigative Analysis

This section analyzed the impact of difference in quality scores, ridge bifurcation count and ridge ending count on the match scores, and the results were grouped by acquisition technology. The matching results of interoperable datasets with S8 were not used for this analysis since previous analysis did not indicate quality or minutiae count being responsible for the extremely high FNMR. A correlation matrix was generated for examining the relation between match score of a pair of fingerprint images, the difference in quality scores for the pair of fingerprint images, the difference in ridge bifurcation count of the pair of images, and the difference in ridge ending count of the pair of images (Table 4.30).

Table 4.30 Correlation matrix of match score, quality score difference, ridge bifurcation count difference, ridge ending count difference

	Match Score	Quality Score Difference	Ridge Bifurcation Count Difference
Quality Score Difference	0.16		
Ridge Bifurcation Count Difference	-0.15	0.40	
Ridge Ending Count Difference	0.11	0.52	0.32

A scatter plot of difference in Aware quality scores between the pair of images and the match scores was generated. The match scores were the raw match scores generated by VeriFinger 5.0 for genuine comparisons of native and interoperable datasets. The scatter plot is grouped by the acquisition technology used to capture the two fingerprint images that were matched. A higher difference in quality scores was related to a lower match score. The *Same Type* and *Optical-Capacitive* groupings showed a relatively high match score with high difference in match scores. The grouping of data points showed that *Thermal-Capacitive* had relatively lower match scores compared to *Optical –Capacitive* and *Same Type* groupings. The *Thermal-Capacitive* and *Optical-Thermal* groupings did not show any specific relation between difference in quality scores and match scores, which indicated that difference in quality scores and its relation to match scores was impacted by the acquisition technologies.

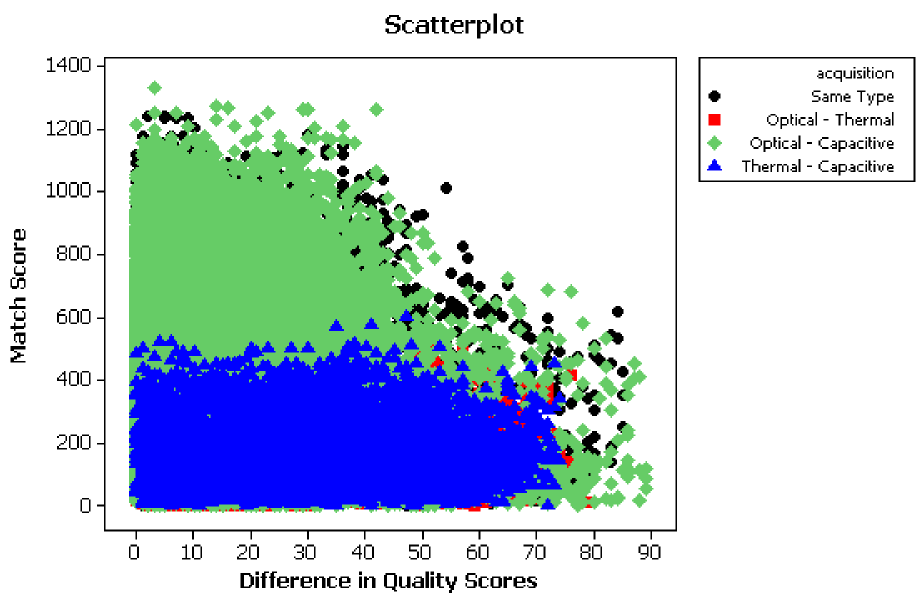

Figure 4.15 Scatter plot of VeriFinger 5.0 Match Score vs. Difference in Quality Scores

A scatter plot of match score vs. difference in ridge bifurcation count was generated and grouped by acquisition technologies used to capture the pair of fingerprints that were matched. A lower difference in ridge bifurcation count indicated that a similar number of ridge bifurcation points were detected in the two fingerprints being compared. The scatter plot showed that difference between ridge bifurcation count and the match score showed a negative relation for *Same Type* and *Optical-Capacitive* groupings. The difference in ridge bifurcation count and match score did not show a specific relation for the *Optical-Thermal* and *Thermal-Capacitive* groupings.

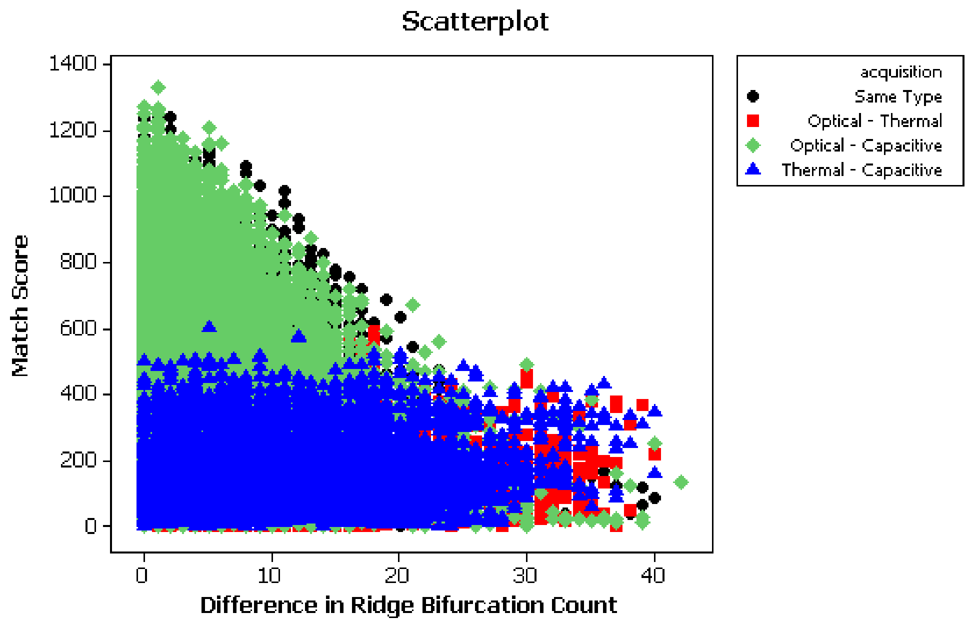

Figure 4.16 Scatter plot of VeriFinger 5.0 Match Score vs. Difference in Ridge Bifurcation Count

A scatter plot of match score vs. difference in ridge ending count was generated and grouped by acquisition technologies used to capture the pair of fingerprints that were matched. A lower difference in ridge ending count indicated that a similar number of ridge ending points were detected in the two fingerprints being compared. A similar relation to difference in ridge bifurcation count was observed in this scatter plot. The difference in ridge ending count and match score did not show a specific relation for *Thermal –Capacitive* and *Optical-Thermal* groupings. The *Same Type* and *Optical –Capacitive* groupings showed a stronger relation between the two variables.

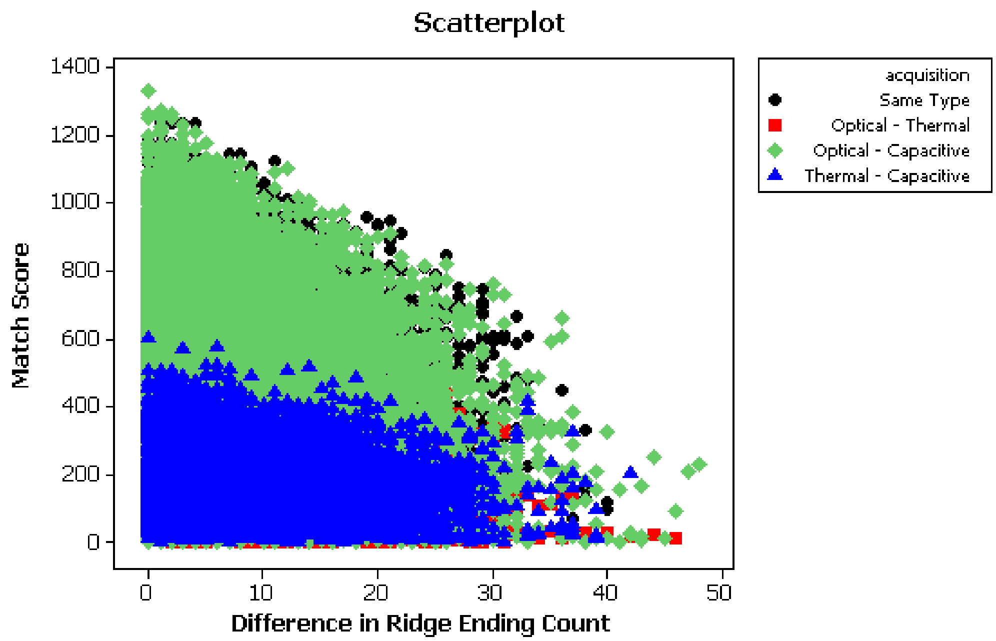

Figure 4.17 Scatter plot of VeriFinger 5.0 Match Score vs. Difference in Ridge Ending

The three scatter plots showed that images from thermal sensor when compared to optical and capacitive sensors did not follow the relation observed for the fingerprint images from the other groups. This makes prediction of interoperability between thermal and other types of sensors harder compared to interoperability between capacitive and optical sensors. These results also indicated a need to perform further analysis on interoperability match scores for fingerprint images collected from thermal sensors.

4.7. Fingerprint Image Transformation

It was observed that fingerprint images of native S8 dataset had FNMR very similar to other native datasets, but its interoperable dataset had lowest FNMR of 99.63%. Fingerprint for S8 dataset were collected using a capacitive swipe sensor. The image quality scores for S8 dataset did not show an extreme deviation from image quality scores of other datasets and the same was observed for minutiae count. The consistency of placement measure in Section 4.4.1.1 also did not show a deviation from other interoperable datasets. These factors indicated that the even though similar minutiae were being detected between images of S8 dataset and other datasets, the placement of minutiae in the fingerprint images were different enough for the matcher to incorrectly reject a genuine comparison. Further evaluation of the raw captured images showed that distance between ridge lines for S8 dataset images was shorter compared to all other fingerprint datasets. Figure 4.18 shows a skeletonized fingerprint image of the same finger from dataset S3 and S8. S3 was collected using an optical touch sensor.

Figure 4.18 Skeletonized fingerprint image from S8(left) and S3(right)

A ridge spacing profile for each image was created using the IMPROFILE image processing method in MATLAB^{TM}. The ridge spacing profile was created by traversing a distance of 100 pixels starting from the core and moving along the y-axis one pixel at a time. The IMPROFILE method computes the intensity values along a traversal line in an image. The 100 pixel distance was chosen as it would allow the IMPROFILE method to intersect sufficient ridge lines to get an understanding of the distance between successive ridge lines. Figure 4.19 shows the traversal path.

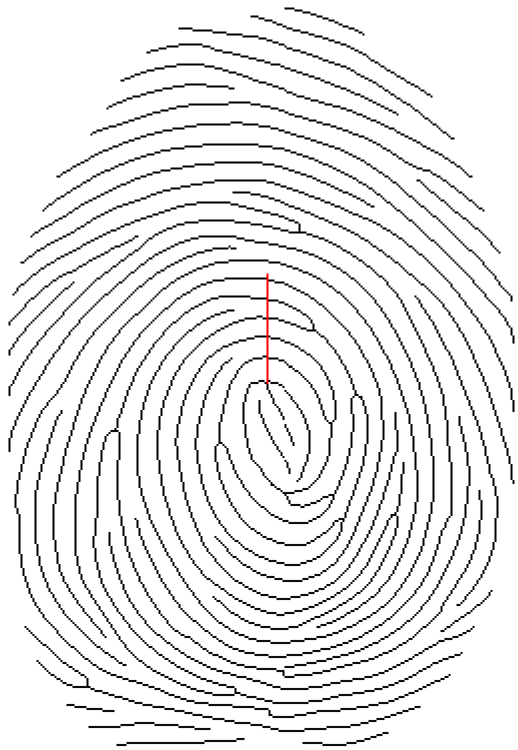

Figure 4.19 Traversal path for IMPROFILE

Figure 4.20 shows output from the IMPROFILE procedure for a single image from S8 dataset, where a downward spike in the graph indicated the intersection of the traversal route with a ridge line. Figure 4.21 shows output from the IMPROFILE procedure for a single image from S3 dataset. Both images from S3 and S8 dataset were of the same finger. Comparison of the two graphs indicated that distance between successive ridges for S3 image was significantly larger than distance between successive ridges for S8 image. In order to make the distance between successive ridges more consistent between the two images, all images for S8 dataset were transformed using 2-D affine spatial transformation process. The transformation parameters included scaling along the x-axis and y-axis. Figure 4.22 shows the image after transformation. It should be noted that the scaling parameters were approximated using only one pair of image, and then applied to all the images of S8 dataset.

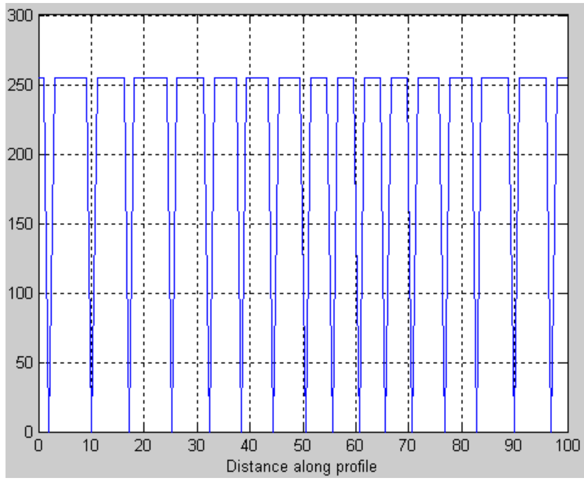

Figure 4.20 IMPROFILE output for S8 dataset image

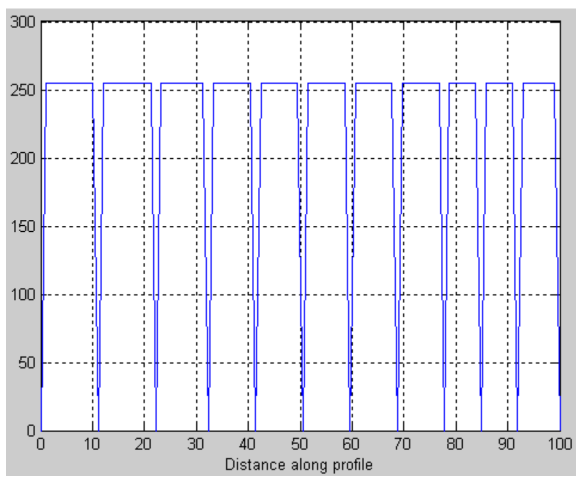

Figure 4.21 IMPROFILE output for S3 dataset image

Figure 4.22 Fingerprint image after transformation

The transformed dataset was relabeled to S8’. Using images from S8’, native and interoperable FNMR were recalculated using VeriFinger 5.0. Table 4.31 shows the FNMR at fixed FMR of 0.1%.

Table 4.31 Recalculated FNMR (FMR = 0.1%)

	S1	S2	S3	S4	S5	S6	S7	S8'	S9
S8'	11.81	8.84	6.43	6.70	12.16	21.68	9.94	0.65	6.57

Table 4.32 FNMR for S3 dataset (FMR=0.1%)

	S1	S2	S3	S4	S5	S6	S7	S9
S3	8.57	4.78	0.24	1.07	1.28	2.73	1.79	0.54

The FNMR for interoperable S8’ datasets showed a significant reduction, and also a decrease in native dataset FNMR. It is worth noting that interoperable FNMR of datasets {S8’,S4} and {S8’,S9} were comparable with FNMR of {S8’,S3}. Comparing these results with FNMR in Table 4.32, which were have been restated from Table 4.15, it was observed that interoperable FNMR for S8’ were lower for datasets which had a lower interoperable FNMR with S3, specifically {S3,S4} and {S3,S9}. This indicated that by transforming images of S8 dataset to be more similar to S3 dataset, there was an associated effect of reducing interoperable FNMR of {S8’,S4} and {S8’,S9}.

The datasets {S1,S6} and {S6,S1} also exhibited relatively high interoperable FNMR of 31.03% and 33.86% respectively, using VeriFinger 5.0 matcher. The ridge spacing profile was computed for a single fingerprint image from S1 and S6 dataset along the x-axis and the y-axis. Figure 4.23 shows the ridge spacing profile along the x-axis for S1 and S6, and Figure 4.24 shows the ridge spacing profile along the y-axis for S1 and S6.

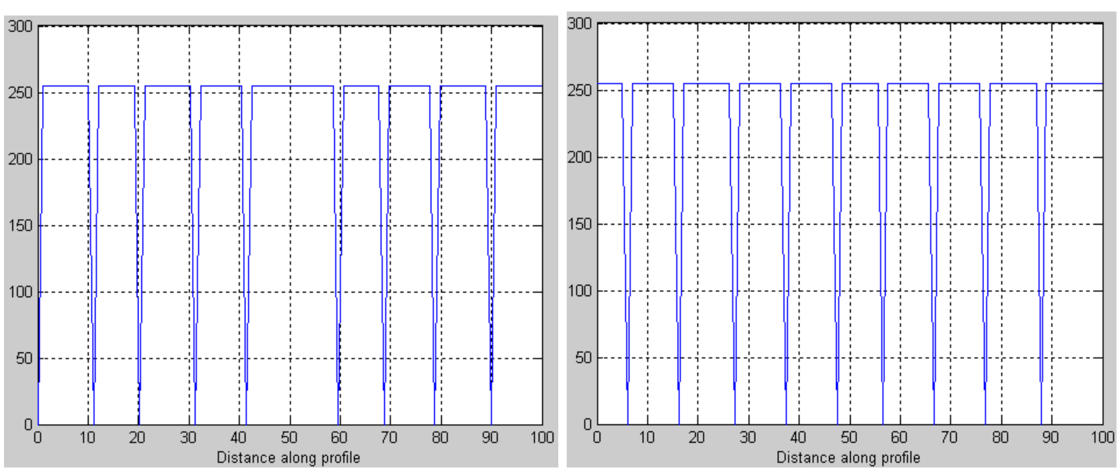

Figure 4.23 Ridge spacing profile along x-axis. S1 image on left and S6 image on right

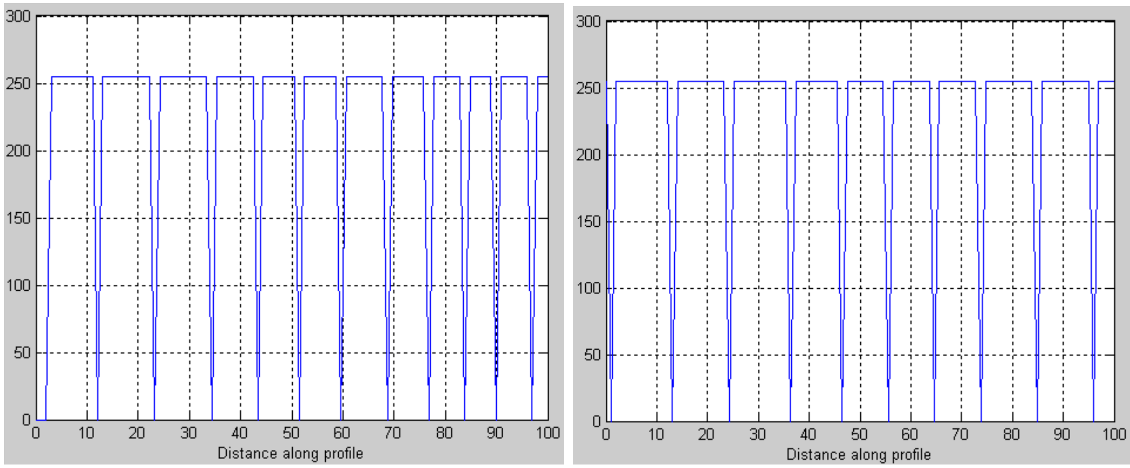

Figure 4.24 Ridge spacing profile along y-axis. S1 mage on left and S6 image on right

Figure 4.23 showed that S1 ridge spacing profile along the x-axis had a larger distance between successive ridges compared to ridge spacing profile for S6. An opposite effect was observed in the ridge spacing profile along the y-axis for the S1 and S6 dataset images. S1 was collected using a thermal swipe sensor and S6 was collected using an optical touch sensor. The difference in interaction between the two sensors was a potential factor in the difference in ridge spacing between the two images. The swipe action introduced an elastic deformation of the finger skin which lead to an increased space between ridges and contributed to a higher FNMR for {S1, S6} and {S6, S1} datasets.

Reliability of the image transformation method was entirely dependent on the ability to detect a core in both the fingerprint images. This was necessary as a common anchor point for the two fingerprint images was required as the starting point of the IMPROFILE method. This method would be efficient only if used in a 1:1 matching, or verification mode. The application of the transformation methodology described in this section was performed at a sensor level since only a pair of images was used to approximate the transformation scaling parameters for the entire dataset collected using the specific sensors. This method can be modified where a ridge spacing profile is created every time a pair of images need to be matched, and then generate the scaling parameters based on the ridge spacing profile. Such a method would make it a sensor agnostic method of transforming images and would have wider applicability. The results in this section provided justification for further development and analysis of transforming images using the ridge spacing profile.

CHAPTER 5. CONCLUSIONS AND RECOMMENDATIONS

This dissertation has described the design, formulation and application of a statistical analysis framework for testing interoperability of fingerprint sensors and its effect on system performance. This chapter reviews the work presented in this dissertation, highlights the contributions made by this work and provides recommendations for future work.

5.1. Conclusions

- Minutiae count similarity of fingerprint images collected from different sensors did not show a relation to a specific acquisition technology or interaction type. There were no common pairwise tests which were statistically similar for minutiae count datasets extracted using Aware, MINDTCT, and VeriFinger 5.0.
- Fingerprint images collected from optical touch sensors showed a higher level of similarity in quality scores with fingerprints collected from other optical touch sensors.
- Performance of minutiae based matchers was significantly affected if the pair of fingerprints being matched were captured from different sensors.
- Similarity of minutiae count and image quality scores did not have an impact on similarity of FNMR for native and interoperable datasets.
- Higher minutiae count for a particular dataset did not have an impact on FNMR of its corresponding interoperable datasets.

- Performance of interoperable datasets cannot be predicted by separately analyzing performance of native datasets.
- Test of similarity of raw match scores between native and interoperable datasets did not provide remarkable results for the raw match scores provided by VeriFinger 5.0 and BOZORTH3 matchers. This is not a deficiency of the test, but indicated a need for transforming the scores over a range which provides a normal distribution.
- Interoperable datasets which had a higher percentage of pairs of fingerprint images in which a core was detected had a positive relation with a lower FNMR, with the only exception of {S1,S6} and {S6,S1} interoperable datasets. Consistent interaction of the finger with a sensor is an important factor in improving performance, and it becomes even more important when the fingerprints are captured from different sensors. A simple metric for consistent placement is required for collecting fingerprints from different sensors, and this dissertation showed that core overlap provides a metric which has an impact on FNMR of interoperable datasets.
- Removing low quality images from interoperable datasets did not lead to a reduction in statistical variance of FNMR for interoperable datasets, although the absolute FNMR was reduced for all native and interoperable datasets.
- Consistency of ridge spacing between interoperable datasets was a very important factor in reducing FNMR of interoperable datasets.
- The Capacitive Touch sensor group showed the lowest interoperable FNMR with Swipe and Optical Touch sensor groups.
- A negative relation between match scores and ridge ending count and ridge bifurcation count was observed for fingerprints collected from the same type of sensors and optical and capacitive sensors. Matching of fingerprints collected from thermal sensor and optical sensors, and

thermal sensor and capacitive sensors did not show a relation between those variables.

5.2. Contributions

- Dataset of fingerprints from 190 individuals on nine different fingerprint sensors in a controlled environment which can be used in testing interoperability of feature extractors and feature matchers.
- Collection of finger skin characteristics and pressure placed on the sensor.
- Formulation of a framework for statistically testing interoperability of fingerprint sensors in terms of FNMR.
- Application of analysis framework to data collected in a controlled experiment.
- Enhanced enrollment procedures for reducing FNMR of interoperable datasets.
- Ability to statistically test similarity of FNMR of native dataset with interoperable dataset.
- A sensor agnostic transformation process to reduce FNMR of interoperable datasets.

5.3. Future Work

Since this dissertation limited itself to a verification scenario, performance was solely measured in terms of FNMR. The dataset collected as part of this dissertation can be used to analyze False Match Rates (FMR) if an identification scenario is of interest. The statistical analysis framework can be modified to test for FMR of interoperable datasets, and it would be interesting to understand the impact of interoperability on FMR of fingerprint datasets collected from multiple sensors.

The impact of removing low quality images from interoperable datasets did not lead to a higher level of similarity in FNMR between interoperable datasets and native datasets. The level of similarity was used as a measure of variance between interoperable and native datasets FNMR. The results indicated that further work is required to investigate the impact of quality on the variance of interoperable datasets performance. The relation between impact of quality and the types of sensors which were used to create the interoperable datasets needs to be analyzed more closely. The FBI has created image quality specifications to measure the fidelity of the fingerprint sensor. The impact of the different components of the image quality specifications on the interoperable error rates would be a relevant research.

The relationships between image quality, temperature, moisture content, oiliness and elasticity were too complex for fitting linear models. Although separate analysis of the relations between quality score and finger skin characteristics were performed, analysis which examines these interactions together would be interesting. Higher order models or neural networks would be more appropriate to understand the relation between skin characteristics and image quality. Understanding the impact of the physiological skin factors on quality is important since improving quality can reduce FNMR of interoperable datasets.

The dataset collected for this dissertation can be used for evaluating interoperability of sensors, feature extractors, and feature matchers as part of the same experiment. This dissertation did not analyze interoperability of feature extractors and feature matchers, and previous studies related to interoperability did not analyze the effect of interoperability of sensors. A combined analysis of fingerprint sensors, feature extractors and feature matchers would be beneficial.

The image transformation model described in Section 4.7 indicated the need to make the spacing between ridgelines in the pair of fingerprint images more consistent in order to reduce FNMR of interoperable datasets. The method described in this dissertation used a linear transformation to make the spacing between ridges more uniform. This is a simple model which works if distortion between the two fingerprint images is relatively consistent, but only differing in scale. This model can be extended to non linear transformation models by creating a ridge spacing profile along eight quadrants as shown in Figure 5.1. A corresponding set of points extracted along different directions of the fingerprint images can be used to generate transformation models where the distortion varies only in certain parts of the fingerprint image.

Figure 5.1 Eight quadrant ridge spacing profile

The distortion of fingerprint images is heavily affected by the type of sensor and the habituation of users with the sensor. Further work into transforming the image so that the distortion of fingerprint images would be reduced without having any a-priori knowledge about the fingerprint sensors would be of immense interest to the field of fingerprint recognition. Sensor agnostic transformation methods, even if limited in its capabilities, would significantly augment the current methodologies being investigated.

LIST OF REFERENCES

Abdurrachim, D., Kunieda, H., Li, D., & Qi, J. (2005). *A hybrid method for fingerprint image quality calculation.* Paper presented at the Automatic Identification Advanced Technologies, Buffalo, N.Y.

Adhami, R., & Meenen, P. (2001). Fingerprinting for security. *IEEE Potentials, 20*(3), 33-38.

Aguilar, J., Fernandez, F., & Garcia, J. (2005). *A Review of Schemes for Fingerprint Image Quality Computation.* Paper presented at the 3rd COST 275 Workshop Biometrics on the Internet.

Allen, R., Prabhakar, S., & Sankar, P. (2005). Fingerprint Identification Technology. In J. Wayman, A. Jain, D. Maltoni & D. Maio (Eds.), *Biometric Systems* (pp. 21-61). London: Springer Verlag.

Alonso-Fernandez, F., Veldhuis, R. N. J., Bazen, A. M., Fierrez-Aguilar, J., & Ortega-Garcia, J. (2006, May). *On the relation between biometric quality and user-dependent score distributions in fingerprint verification.* Paper presented at the Workshop on Multimodal User Authentication, Toulouse, France.

Amengual, J. C., Perez, J. C., Prat, F., Saez, S., & Vilar, J. M. (1997, 14th July). *Real Time Minutiae Extraction in Fingerprint Images.* Paper presented at the 6th International Conference on Image Processing and its Applications, Dublin.

Amin, A., & Yager, N. (2004). Fingerprint Verification based on Minutiae Features: a review. *Pattern Analysis Application, 7*, 94-113.

Ashbourn, J. (2000). *Biometrics: Advanced Identitiy Verification*. London: Springer Verlag.

Bazin, A., & Mansfield, T. (2007, 7-8 June). *An Investigation of Minutiae Template Interoperability.* Paper presented at the AutoID 2007, Alghero, Italy.

Bigun, J., Fronthaler, H., & Kollreider, K. (2006). *Automatic Image Quality Assessment with Application in Biometrics.* Paper presented at the Computer Vision and Pattern Recognition Workshop, New York.

Biometric Sample Quality Standard - Part 1: Framework. (2006).): ISO/IEC JTC1/SC37.

Boashash, B., Deriche, M., & Kasaei, S. (1997). *Fingerprint Feature Enhancement using Block-Direction on Reconstructed Images.* Paper presented at the International Conference on Information, Communications, and Signal Processing, Singapore.

Bolle, R., & Ratha, N. K. (1998, 16-20 August). *Effect of Controlled Image Acquisition on Fingerprint Matching.* Paper presented at the 14th International Conference on Pattern Recognition, Brisbane, Australia.

Bolle, R. M., Connell, J. H., Pankanti, S., Ratha, N. K., & Senior, A. W. (2004a). Basic System Errors. In *Guide to Biometrics* (pp. 61-86). New York: Springer Verlag.

Bolle, R. M., Connell, J. H., Pankanti, S., Ratha, N. K., & Senior, A. W. (2004b). Performance Testing. In *Guide to Biometrics* (pp. 105-128). New York: Springer-Verlag.

Campbell, J., & Madden, M. (2006). *ILO Seafarers' Identity Documents Biometric Interoperability Test Report* Geneva: International Labour Organization.

Chan, K., Huang, Z., & Liu, J. (2000). *Direct Minutiae Extraction from Gray-Level Fingerprint Image by Relationship Examination.* Paper presented at the International Conference on Image Processing, Vancouver, Canada.

Chen, S., Jain, A. K., Karu, K., & Ratha, N. K. (1996). A Real-Time Matching System for Large Fingerprint Database. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 18*(8), 799-813.

Clear Registered Traveler. (2006). Retrieved January 31, 2007, from HTTP://WWW.FLYCLEAR.COM/INDEX.HTML

Cole, S. A. (2001). *Suspect Identities: A History of Fingerprinting and Criminal Identification*. Cambridge, Massachusetts: Harvard University Press.

Cole, S. A. (2004). History of Fingerprint Pattern Recognition. In N. Ratha & R. Bolle (Eds.), *Automatic Fingerprint Recognition Systems* (pp. 1-25). New York: Springer.

Daugman, J., & Williams, G. O. (1996). *A proposed standard for biometric decidability.* Paper presented at the CardTech SecureTech, Atlanta, GA.

Dean, B., Salstrom, R., & Wayman, J. (2001). *Studies of Fingerprint Imaging Systems in Social Service Applications.* Paper presented at the International Conference on Management of Engineering and Technology Portland.

DigitalPersona. (2006). DigitalPersona Pro. Retrieved January 31, 2007, from HTTP://WWW.DIGITALPERSONA.COM/RESOURCES/DOWNLOADS/PRO%20V4_010-06.PDF

Doddington, G., Kamm, T., Martin, A., Ordowski, M., & Przybocki, M. (1997). *The DET curve in assessment of detection task performance.* Paper presented at the Eurospeech, Greece.

Elliott, S. J., Kukula, E., & Modi, S. (2007). Issues Involving the Human Biometric Sensor Interface. In S. Yanushkevich, P. Wang & S. Srihari (Eds.), *Image Pattern Recognition Synthesis and Analysis in Biometrics* (Vol. 67, pp. 400): World Scientific Publishers.

Elliott, S. J., & Modi, S. K. (2006). *Impact of Image Quality on Performance: Comparison of Young and Elderly Fingerprints.* Paper presented at the 6th International Conference on Recent Advances in Soft Computing (RASC), Canterbury, UK.

Elliott, S. J., & Sickler, N. C. (2005). *An evalulation of fingeprint image quality across an elderly population vis-avis an 18-25 year old population.* Paper presented at the International Carnahan Conference on Security Technology, Las Palmas, Gran Canaria.

Elliott, S. J., & Young, M. (2007). *Image Quality and Performance Based on Henry Classification and Finger Location.* Paper presented at the AutoID 2007, Alghero, Italy.

Engeler, W., Frank, P., & Leung, M. (1990, 24th September). *Fingerprint Image Processing Using Neural Network.* Paper presented at the IEEE Conference on Computer and Communication Systems, Hong Kong.

Eshera, M., & Shen, W. (2004). Feature Extraction in Fingerprint Images. In R. Bolle & N. Ratha (Eds.), *Automatic Fingerprint Recognition Systems* (pp. 145-180). New York: Springer Verlag.

Foo, K. C., Ngo, D. B. L., & Sagar, V. K. (1995, 18th October). *Fuzzy Feature Selection for Fingerprint Identification.* Paper presented at the IEEE International Carnahan Conference on Security Technology, Sanderstead, U.K.

Ford, J. P., & Sticha, P. J. (1999). *Introduction to Biometric Identification Technology: Capabilities and Applications to the Food Stamp Program*. Arlington, Virginia: R. Lewis & Company.

Garris, M. D., Janet, S., McCabe, R., Tabassi, E., Watson, C. I., & Wilson, C. L. (2004). *User's Guide to NIST Fingerprint Image Software 2 (NFIS2)*. Gaithesburg, MD: National Institute of Standards and Technology.

Grother, P. (2006, September). *An Overview of ISO/IEC 19795-4.* Paper presented at the Biometric Consotium, Baltimore.

Grother, P., McCabe, M., Watson, C., Indovina, M., Salamon, W., Flanagan, P., et al. (2006). *MINEX Performance and Interoperability of the INCITS 378 Fingerprint Template*. Gaithersburg, Maryland: National Institute of Standards and Technology.

Han, J., Kadowaki, T., Sato, K., & Shikida, M. (1999). *Thermal Analysis of Fingerprint Sensor Having a Microheater Array.* Paper presented at the International Symposium on Micromechatronics and Human Science, Nagoya, Japan.

Han, Y., Nam, J., Park, N., & Kim, H. (2006). *Resolution and Distortion Compensation based on Sensor Evaluation for Interoperable Fingerprint Recognition.* Paper presented at the 2006 International Joint Conference on Neural Networks, Vancouver, Canada.

Hawley, D., Roy, R., Yu, A. W., & Zhu, S. (1986). Frustrated Total Internal Reflection: A demonstration and review. *American Journal of Physics, 54*(7), 602-607.

Haylock, S. E. (1979). Khan Bahadur Azizul Haque. *Fingerprint World, 5*(17), 28-29.
Hong, L., Bolle, R., Jain, A. K., & Pankanti, S. (1997). An Identity Authentication System Using Fingerprints. *Proceedings of the IEEE, 85*(9), 1365-1388.
Hutchings, P. J. (1997). Modern Forensics: Photography and Other Suspects. *Cardozo Studies in Law and Literature, 9*(2), 229-243.
IBG. (2007). *Biometrics Market and Industry Report*. NY: IBG.
Igaki, S., Eguchi, S., Yamagishi, F., Ikeda, H., & Inagaki, T. (1992). Real-time fingerprint sensor using a hologram. *Applied Optics, 31*(11), 1794-1802.
ISO. *Biometric Data Interchange Formats -Part 2, Part4*. Geneva: SO/IEC JTC1 SC37 WG3.
ISO. (2006). *Text of 1st Working Draft 29794-1, Biometric Sample Quality Standard - Part 1: Framework*. Geneva: ISO/IEC JTC 1/SC37.
ISO/IEC JTC1 SC37 SD2 - Harmonized Biometric Vocabulary. (2006).). Geneva: ISO/IEC.
Jain, A., Chen, Y., & Dass, S. (2005). *Fingerprint Quality Indices for Predicting Authentication Performance.* Paper presented at the 5th International Conf. on Audio- and Video-Based Biometric Person Authentication, Rye Brook, NY.
Jain, A., Maltoni, D., Maio, D., & Wayman, J. (2005). *Biometric Systems Technology, Design and Performance Evaluation*. London: Spring Verlag.
Jain, A., & Ross, A. (Eds.). (2004). *Biometric Sensor Interoperability* (Vol. 3067). Berlin: Springer-Verlag.
Jain, L. C., Erol, A., & Halici, U. (1999). Introduction to Fingerprint Recognition. In L. C. Jain, U. Halici, I. Hayashi, S. B. Lee & S. Tsutsui (Eds.), *Intelligent Biometric Techniques in Fingerprint and Face Recognition* (pp. 1-34). Boca Raton, Florida: CRC.
Jarosz, H., Fondeur, J. C., & Dupre, X. (2005). Large Scale Identification System Design. In J. Wayman, A. Jain, D. Maltoni & D. Maio (Eds.), *Biometric Systems* (pp. 263-288). London: Springer Verlag.
Jian, X., Lim, E., & Yau, W. (2002). *Fingerprint quality and validity analysis.* Paper presented at the International Conference on Image Processing, Rochester, N.Y.
Jiang, X. (2005). On Orientation and Anisotropy Estimation for Online Fingerprint Authentication. *IEEE Transactions on Signal Processing, 53*(10), 4038-4049.
Khanna, R. (2004). Systems Engineering for Large-Scale Fingerprint Systems. In N. Ratha & R. Bolle (Eds.), *Automatic Fingerprint Recognition Systems* (pp. 283-304). New York: Springer.
Kim, H., Kang, H., Lee, B., Shin, D., & Kim, J. (2003). *A Study on Performance Evaluation of Fingerprint Sensors.* Paper presented at the AVBPA 2003, Guildford, UK.

Ko, T., & Krishnan, R. (2004). *Monitoring and Reporting of Fingerprint Image Quality and Match Accuracy for a Large User Application.* Paper presented at the Applied Imagery Pattern Recognition Workshop, Washington, D.C.

Lim, T. W., & Moghavvemi, M. (2000). *Capacitive Fingerprint Sesnsor Chip for Automatic Matching.* Paper presented at the TENCON 2000, Kuala Lumpur, Malaysia.

M1. (2004). *INCITS 378 - Finger Minutiae Format for Data Interchange*. Washington D.C.: ANSI/INCITS

Machida, K., Morimura, H., Okazaki, Y., & Shigematsu, S. (2004). *Fingerprint Sensor Technology- Sensor Structure/Chip, Sensing Scheme, and System.* Paper presented at the 7th International Conference on Solid-State and Integrated Circuits Technology, Beijing, China.

Maio, D., & Maltoni, D. (1997). Direct Gray-Scale Minutaie Detection in Fingerprints. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 19*(1), 27-40.

Mansfield, A., & Wayman, J. (2002). *Best Practices*. Middlesex: National Physics Laboratory.

Marcialis, G., & Roli, F. (2004). Fingerprint Verification by Fusion of Optical and Capacitive Sensors. *Pattern Recognition Letters, 25*(11), 1315-1322.

Mayers, J. C. (1788). *Anatomical Copper-plates with Appropriate Explanations*.

Mehtre, B. M. (1993). Fingerprint Image Analysis for Automatic Identification. *Machine Vision and Applications, 6*, 124-139.

Montgomery, D. C. (1997). *Design and Analysis of Experiments* (4th ed.). New York: John Wiley & Sons.

Muller, R. (2001). *Fingerprint Verification with Microprocessor Security Tokens.* Munich University of Technology, Munich.

Nagdir, R., & Ross, A. (2006). *A Calibration Model for Fingerprint Sensor Interoperability.* Paper presented at the SPIE Conference on Biometric Technology for Human Identificaiton III, Orlando, USA.

O'Gorman, L. (1999). *Biometrics Personal Identification in Networked Society*. Norwell: Kluwer Academic Publishers.

O'Gorman, L., & Xia, X. (2001). Innovations in fingerprint capture devices. *Pattern Recognition, 36*(1), 361-369.

Poh, N., Kittler, J., & Bourlai, T. (2007, September). *Improving Biometric Device Interoperability by Likelihood Ratio-based Quality Dependent Score Normalization.* Paper presented at the BTAS 2007, Washington, D.C.

Polson, C. J. (1951). Finger Prints and Finger Printing: An Historical Study. *American Journal of Police Science, 41*(5), 690-704.

Raafat, H., & Xiao, Q. (1991). Fingerprint Image Postprocessing: A Combined Statistical and Structural Approach. *Pattern Recognition, 24*(10), 985-992.

Reed, B. (1981). Automated Fingerprint Identification: From Will West to Minnesota Nine-Fingers and Beyond. *Journal of Police Science and Adminstration, 9*, 318-319.

Secugen. SEIR Optic Technolgoy. Retrieved May 2, 2007, from HTTP://WWW.SECUGEN.COM/DOWNLOAD/SGWP_SEIR.PDF

Sekran, U. (2003). *Research Methods for Business*. New York: John Wiley & Sons, Inc.

Setlak, D. R. (2004). Advances in Fingerprint Sensors Using RF Imaging Techniques. In N. Ratha & R. Bolle (Eds.), *Automatic Fingerprint Recognition Systems* (pp. 27-53). New York: Springer-Verlag.

Shu-Long, J., Zhong-Kang, S., & Yan-Yan, W. (1991). *The Non-Parametric Detection with Neural Network.* Paper presented at the International Conference on Circuits and Systems, China.

Stock, R. M., & Swonger, C. W. (1969). *Development and Evaluation of a Reader of Fingerprint Minutiae*. Ithica, N.Y. : Cornell Aeronautical Laboratory.

Tabassi, E., & Wilson, C. L. (2005). *A novel approach to fingerprint image quality.* Paper presented at the International Conference on Image Processing, Genoa, Italy.

Tabassi, E., Wilson, C. L., & Watson, C. I. (2004). *Fingeprint Image Quality* (NISTIR 7151). Gaithersburg, MD: NIST.

Thieme, M. (2003). *Effect of User Habituation in the Context of Biometric Performance Testing*. Washington D.C.: ANSI/INCITS.

Tolley, B. (2006). Notes for Using SAS & Homework Problems [Electronic Version], 34. Retrieved 30th March, 2008 from HTTP://WWW.UTMEM.EDU/PREVMED/PM/FILELIBRARY/SASNOTES2006_2007.PDF.

TSA. (2006). *Registered Traveler Program Business Model Version 2.1*. Washington D.C.: U.S. Department of Homeland Security.

Wayman, J. (1997). *A Generalized Biometric Identification System Model.* Paper presented at the 31st Conference on Signals, Systems and Computers, Pacific Grove, California.

Zhang, Y., Yang, J., & Wu, H. (2006). Coarse-to-fine image registration for sweep fingerprint sensors. *Optical Engineering, 45*(6), 1-3.

Appendix A.

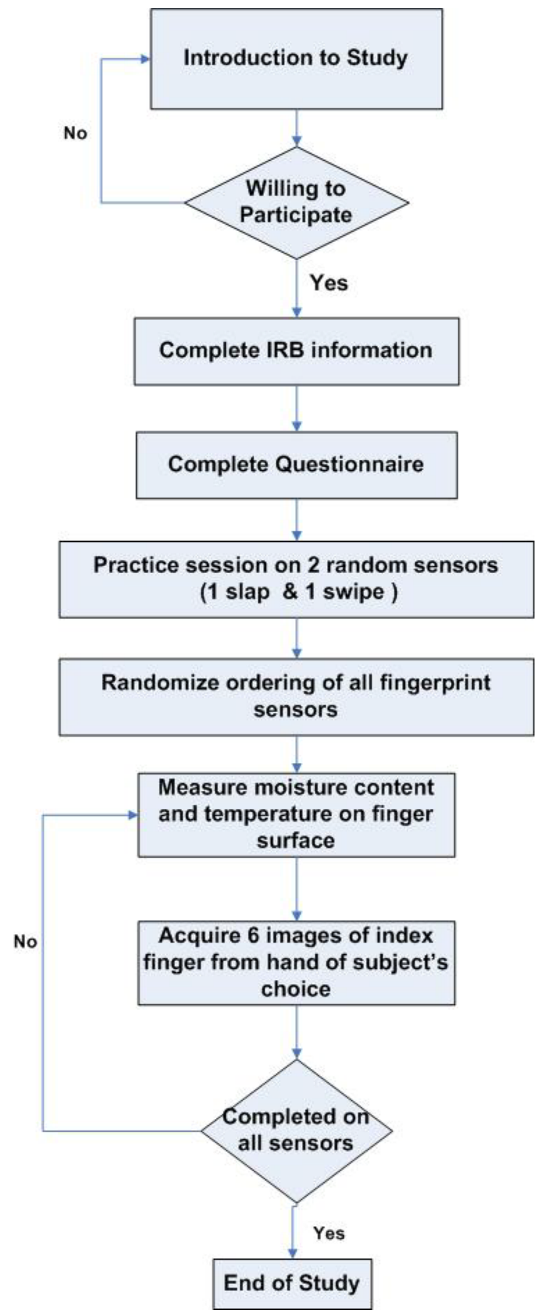

Figure A.1 Data Collection Protocol

Appendix B.

List of questions to be answered by each participant:

1. Age.
2. Gender.
3. Handedness.
4. Occupation.
5. Ethnicity.

Data Collected before first interaction with the fingerprint sensors:

1. Temperature in the lab.
2. Outdoor temperature.

Data Collected before interaction with a new fingerprint sensor:

1. Moisture content of skin on finger surface.
2. Oiliness of skin on finger surface.
3. Elasticity of skin on finger surface.
4. Temperature of skin on finger surface.

Data Collected during every interaction with the fingerprint sensors:

1. Peak pressure applied on the fingerprint sensor.
2. Failure to acquire.

Appendix C.

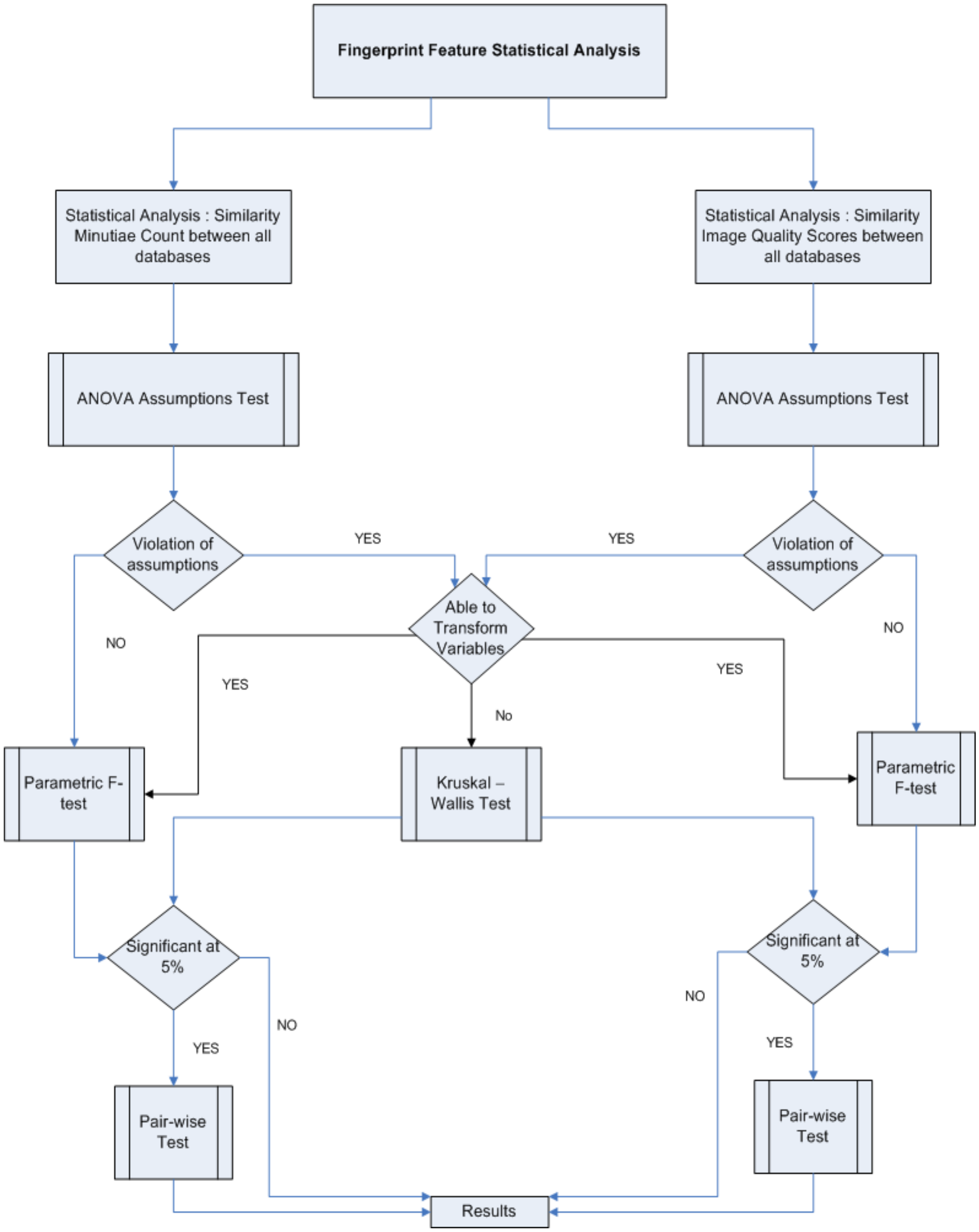

Figure C.1 Flowchart of fingerprint feature statistical analysis

Appendix D.

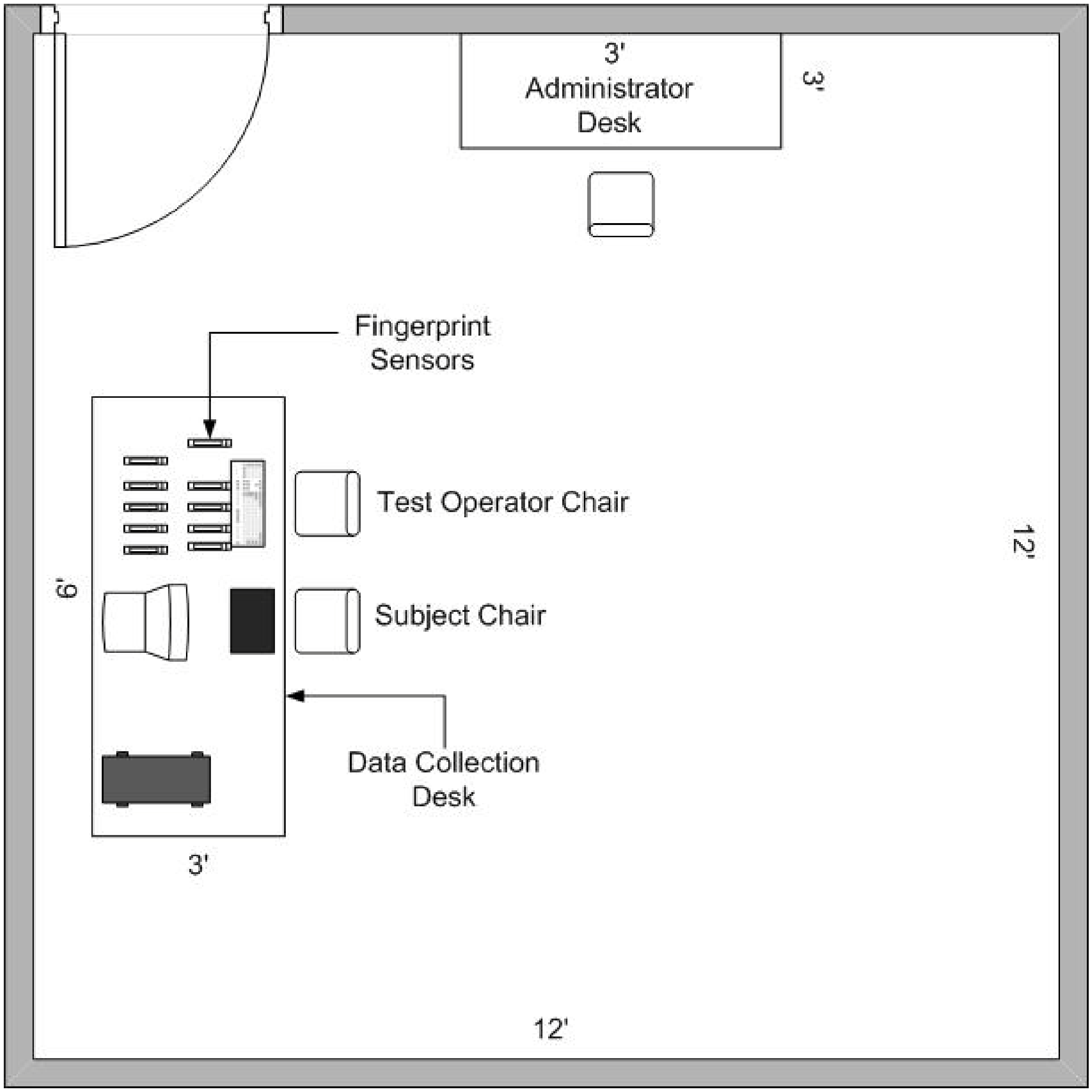

Figure D.1 Physical layout of data collection area

Appendix E.

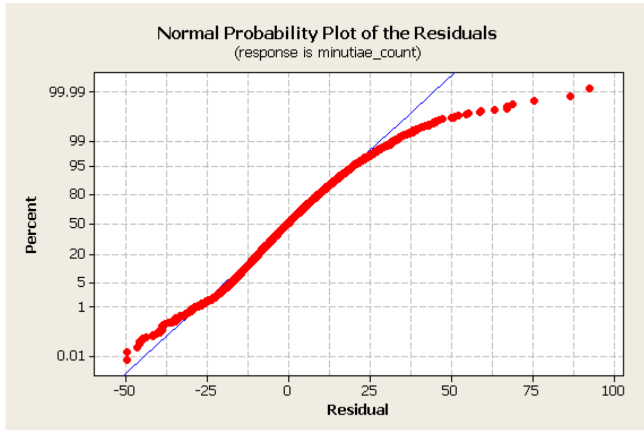

Figure E.1 Normality plot of residuals of Aware Minutiae Count

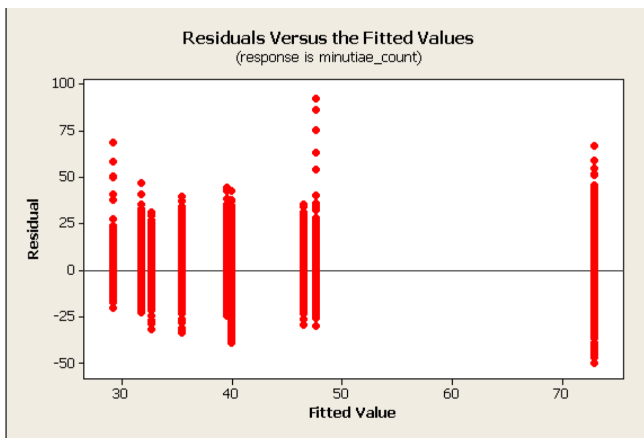

Figure E.2 Residuals vs. Fitted values of Aware Minutiae Count

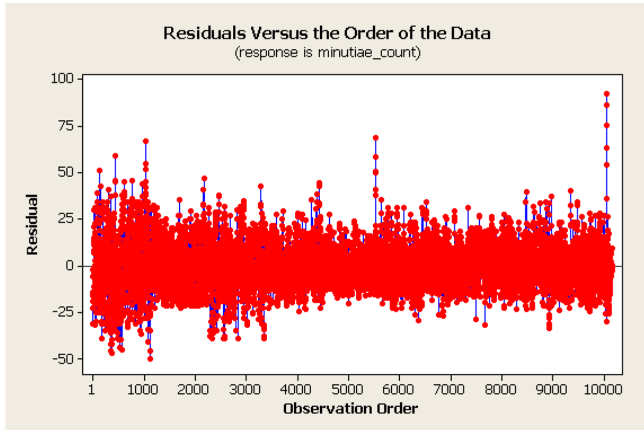

Figure E.3 Time series plot of observations

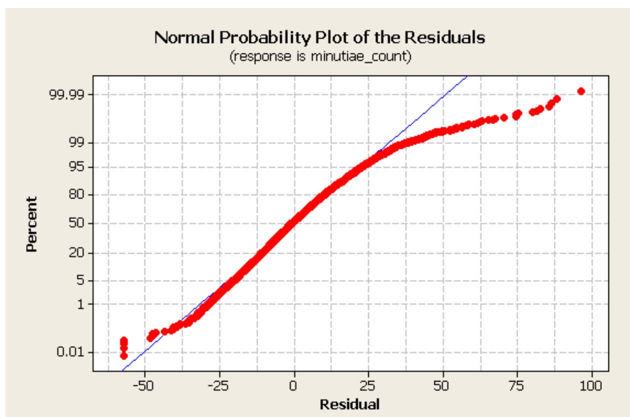

Figure E.4 Normality plot of residuals of NBIS Minutiae Count

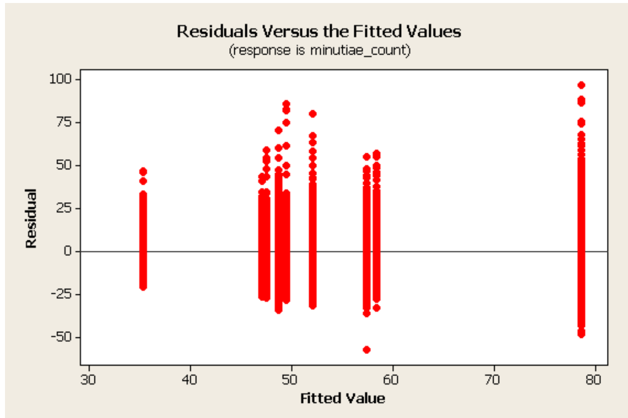

Figure E.5 Residuals vs. Fitted values of NBIS Minutiae Count

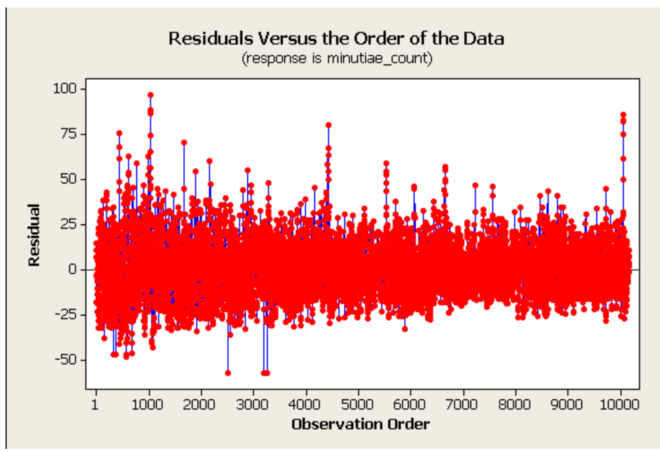

Figure E.6 Time series plot of observations

Appendix F.

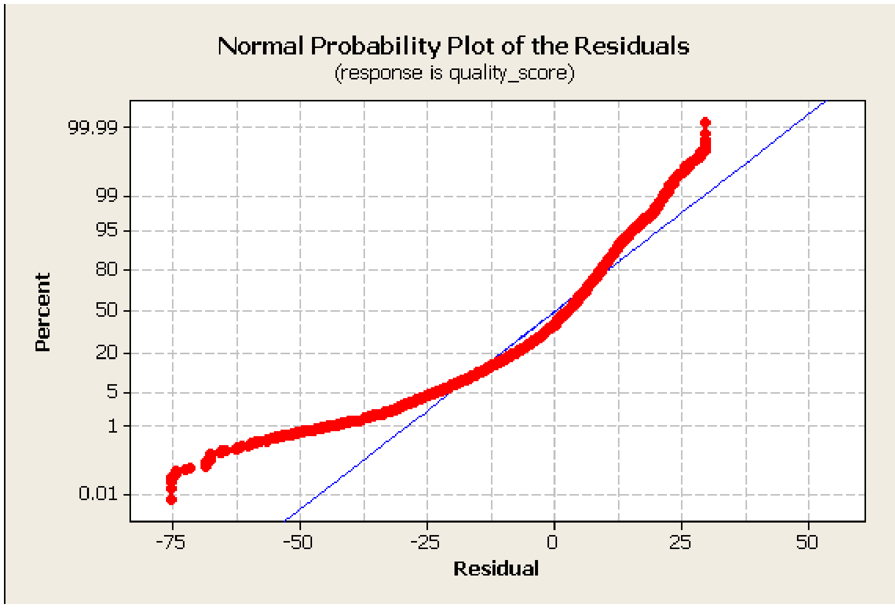

Figure F.1 Normality plot of residuals of Aware Quality Scores

Appendix G.

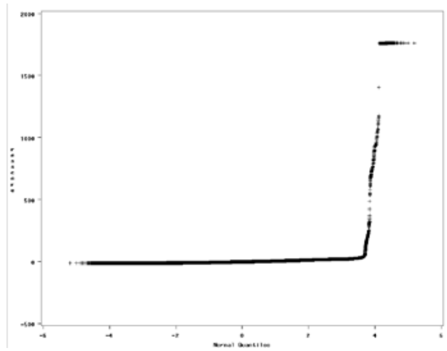

Figure G.1 Normality plot of residuals of match scores from imposter comparisons with S1 dataset

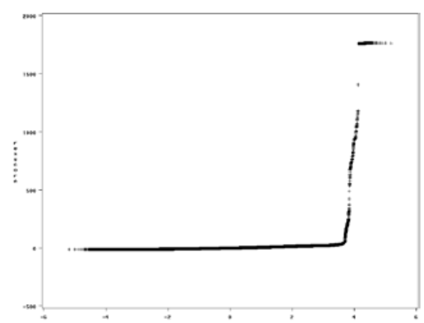

Figure G.2 Normality plot of residuals of match scores from imposter comparisons with S2 dataset

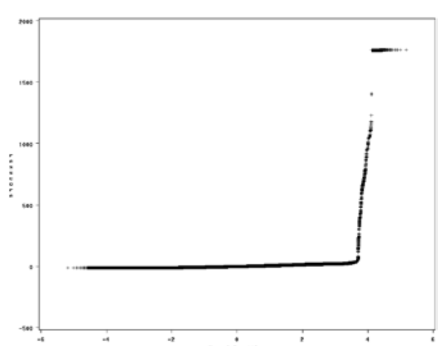

Figure G.3 Normality plot of residuals of match scores from imposter comparisons with S3 dataset

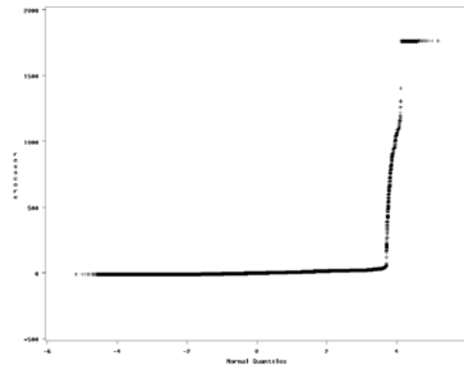

Figure G.4 Normality plot of residuals of match scores from imposter comparisons with S4 dataset

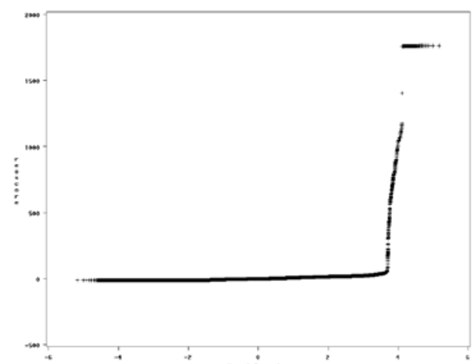

Figure G.5 Normality plot of residuals of match scores from imposter comparisons with S5 dataset

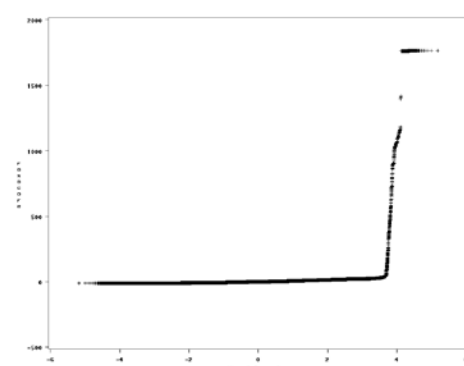

Figure G.6 Normality plot of residuals of match scores from imposter comparisons with S6 dataset

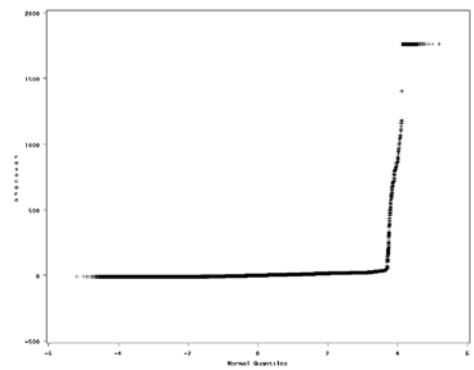

Figure G.7 Normality plot of residuals of match scores from imposter comparisons with S7 dataset

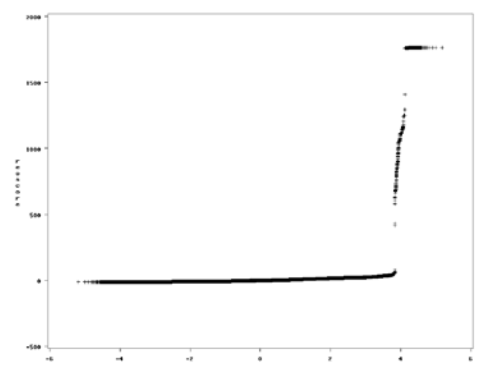

Figure G.8 Normality plot of residuals of match scores from imposter comparisons with S8 dataset

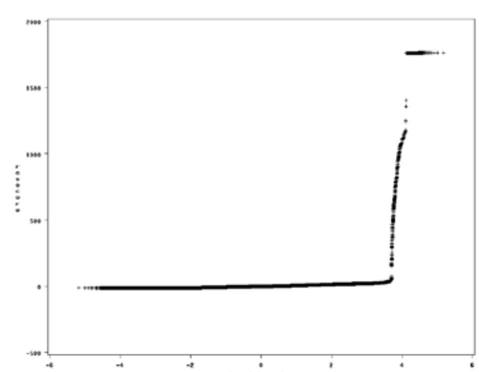

Figure G.9 Normality plot of residuals of match scores from imposter comparisons with S9 dataset

Appendix H.

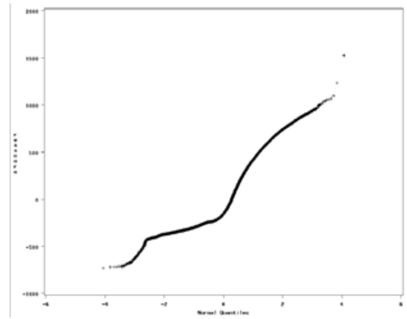

Figure H.1 Normality plot of residuals of match scores from genuine comparisons with S1 dataset

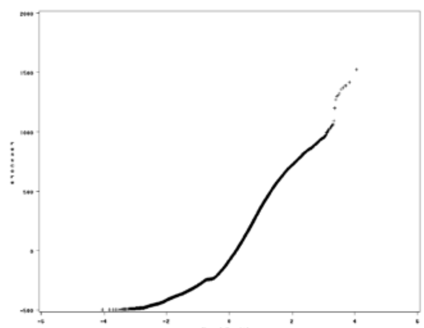

Figure H.2 Normality plot of residuals of match scores from genuine comparisons with S2 dataset

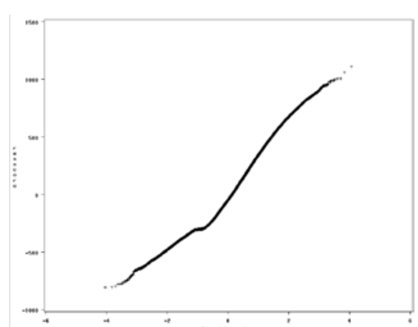

Figure H.3 Normality plot of residuals of match scores from genuine comparisons with S3 dataset

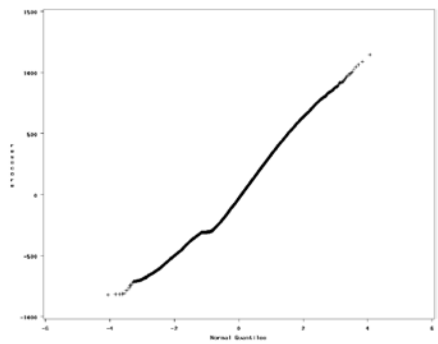

Figure H.4 Normality plot of residuals of match scores from genuine comparisons with S4 dataset

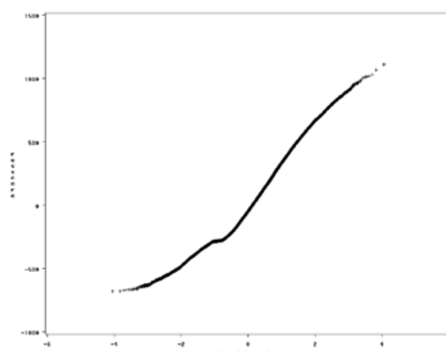

Figure H.5 Normality plot of residuals of match scores from genuine comparisons with S5 dataset

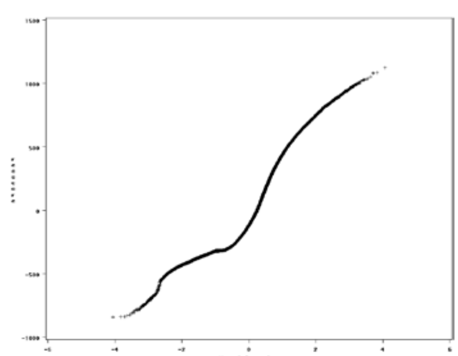

Figure H.6 Normality plot of residuals of match scores from genuine comparisons with S6 dataset

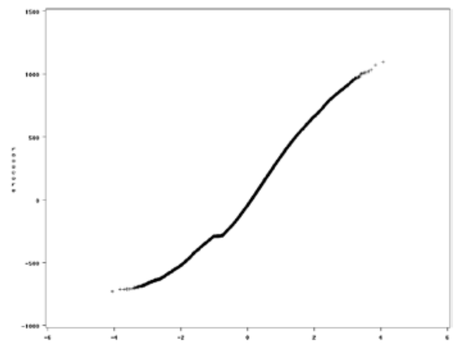

Figure H.7 Normality plot of residuals of match scores from genuine comparisons with S7 dataset

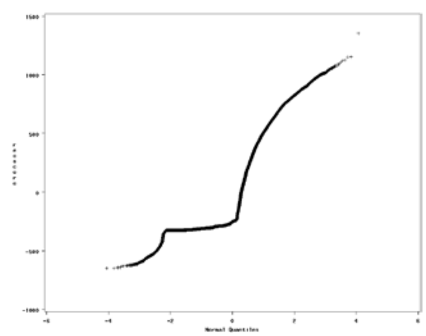

Figure H.8 Normality plot of residuals of match scores from genuine comparisons with S8 dataset

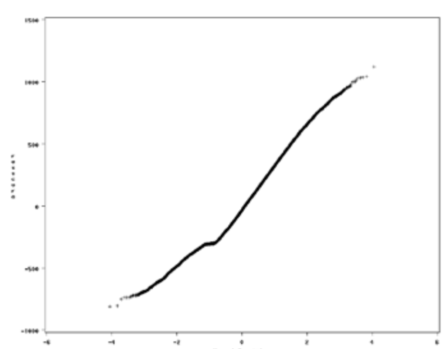

Figure H.9 Normality plot of residuals of match scores from genuine comparisons with S9 dataset

Appendix I.

One hundred and ninety participants completed the survey and the fingerprint collection activity. Fifty nine participants were females and 131 participants were males. Two categories of labor distribution were created based on subjective assessment of the amount of wear and tear an individual puts on their finger skin. The two categories were manual laborer and office worker. Seventeen participants placed themselves in the manual laborer group and 173 participants placed themselves in the office worker group. Three age groups were created to categorize the participants: less than 30 years, between 30 to 50 years, and 50 years and older. One hundred and fifty six participants were less than 30 years old, 23 participants were between 30 and 50 years old, and 11 participants were 50 years and older. One hundred and sixty participants were right handed, 23 participants were left handed, and three participants were ambidextrous. Ambidextrous participants were reassigned according to the dominant hand used for writing. Six categories for ethnicity were used. One hundred and thirty three participants identified themselves as Caucasian, eight as Black, one as American Indian, 34 as Asian, 11 as Hispanic, and three participants as the Other category. Table I.1 provides a summary of the results

Table I.1 Results of Survey

<table>
<tr><td>Total</td><td colspan="6">190</td></tr>
<tr><td rowspan="2">Gender</td><td colspan="3">Male</td><td colspan="3">Female</td></tr>
<tr><td colspan="3">131</td><td colspan="3">59</td></tr>
<tr><td rowspan="2">Occupation</td><td colspan="3">Manual Laborer</td><td colspan="3">Office Worker</td></tr>
<tr><td colspan="3">17</td><td colspan="3">173</td></tr>
<tr><td rowspan="2">Age Groups</td><td colspan="2">< 30 years</td><td colspan="2">30-50 years</td><td colspan="2">> 50 years</td></tr>
<tr><td colspan="2">156</td><td colspan="2">23</td><td colspan="2">11</td></tr>
<tr><td rowspan="2">Handedness</td><td colspan="2">Right</td><td colspan="2">Left</td><td colspan="2">Ambidextrous</td></tr>
<tr><td colspan="2">164</td><td colspan="2">23</td><td colspan="2">3</td></tr>
<tr><td rowspan="2">Ethnicity</td><td>Caucasian</td><td>Black</td><td>Hispanic</td><td>American Indian</td><td>Asian</td><td>Other</td></tr>
<tr><td>133</td><td>8</td><td>11</td><td>1</td><td>34</td><td>3</td></tr>
</table>

Appendix J.

A correlation matrix was generated for examining the relation between quality score of the fingerprint image, temperature, moisture content, oiliness and elasticity of finger skin (Table J.1). The quality score was generated using Aware quality software. The cells with the Pearson correlation coefficient are related to the elements indicated in the corresponding row and column headers.

Table J.1 Correlation matrix of Quality Score, Temperature, Moisture Content, Oiliness, Elasticity

	Quality Score	Temperature	Moisture Content	Oiliness
Temperature	-0.006			
Moisture Content	0.006	0.009		
Oiliness	0.024	0.056	0.10	
Elasticity	-0.20	0.051	0.09	-0.018

A scatter plot of quality score vs. temperature, quality score vs. moisture content, quality score vs. oiliness and quality score vs. elasticity was generated. The data points were grouped according to the type of acquisition technology used to capture the fingerprint images. The scatter plot of quality score vs. moisture did not show a specific relation but the biggest cluster of data points with the higher quality scores was observed between 20 and 50 units on the moisture scale (Figure J.2). The scatter plot of quality score vs. oiliness showed a negative relation between quality score and oiliness units (Figure J.3). The scatter plot of quality score vs. elasticity showed a positive relation between the quality score and elasticity units (Figure J.4). These scatter plots indicated a difference in relation between the finger skin characteristics and the quality scores for images captured from different acquisition technologies. These scatter plots indicate that a framework can be created to improve quality scores based on the finger skin characteristics.

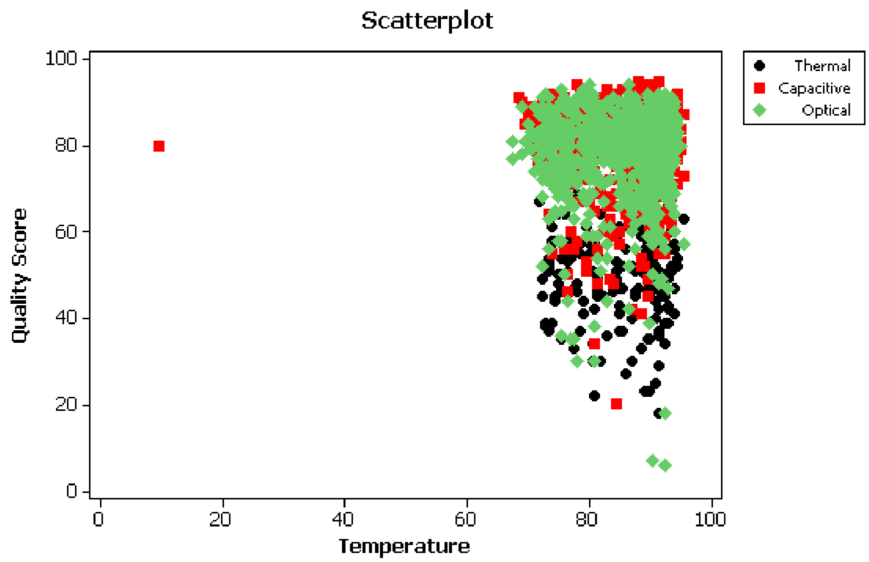

Figure J.1 Scatter plot of Aware Quality Score vs. Temperature

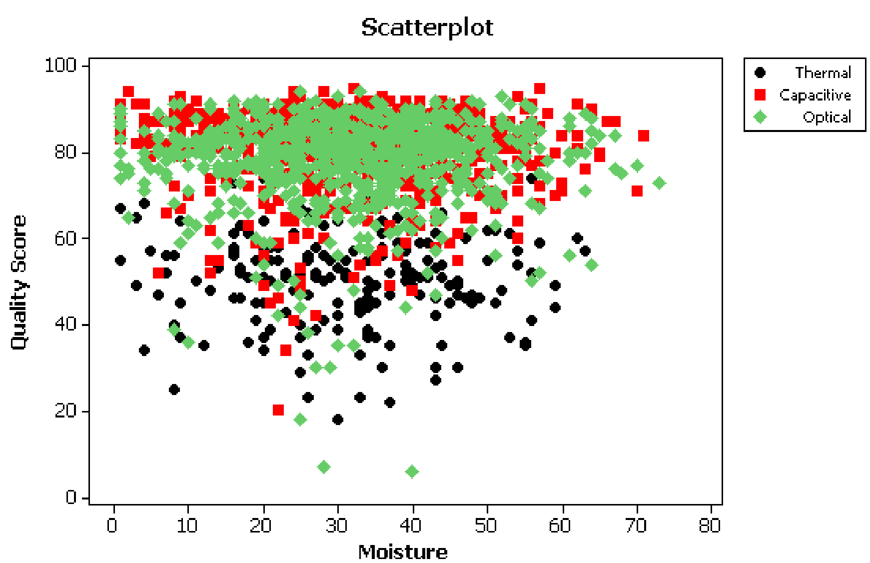

Figure J.2 Scatter plot of Aware Quality Score vs. Moisture

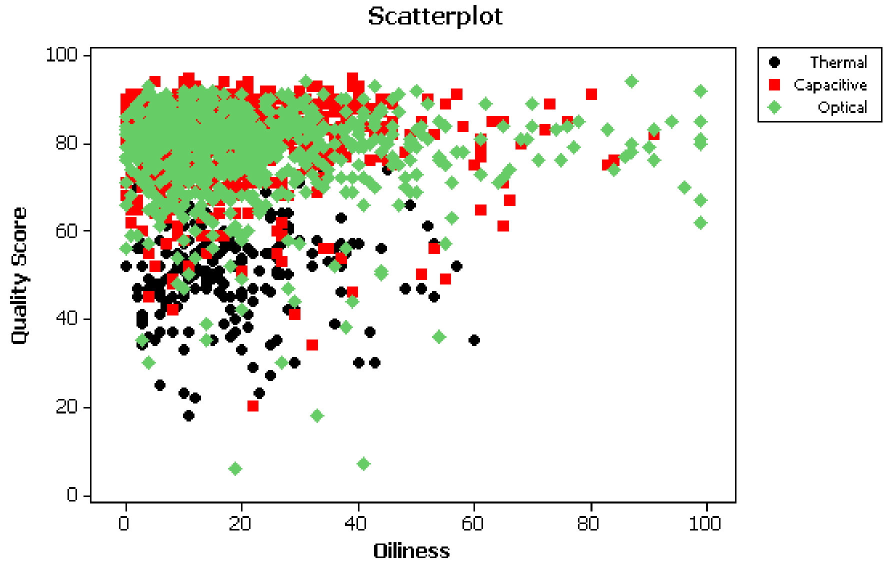

Figure J.3 Scatter plot of Aware Quality Score vs. Oiliness

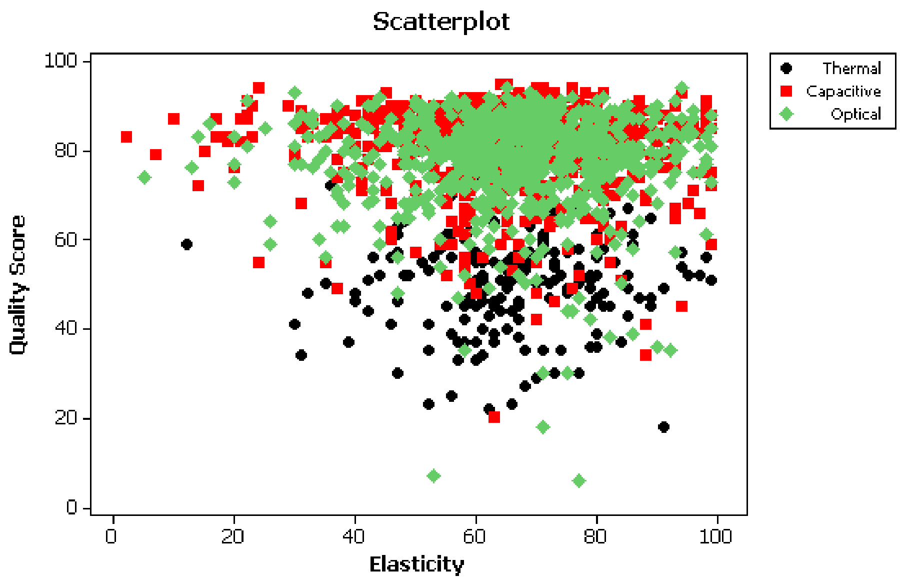

Figure J.4 Scatter plot of Aware Quality Score vs. Elasticity

VITA

SHIMON K. MODI

EDUCATION

Doctorate of Philosophy, Technology, Expected August 2008.
Specialization in Computational Sciences & Engineering
Purdue University.
Dissertation: Fingerprint Sensor Interoperability: Analysis of Fingerprint Sensor Interoperability on System Performance
Advisor: Dr. Stephen J. Elliott

Master of Science, Technology, May 2005.
Specialization in Information Security
Purdue University.
Thesis: Keystroke Dynamics Verification Using Spontaneous Password
Advisor: Dr. Stephen J. Elliott

Bachelor of Science, Computer Science, May 2003.
Dual Minors Management and Economics.
Purdue University

HONORS/AFFILIATIONS

Ross Fellowship for Doctoral Studies 2005-present
Semester Honors Spring ’02, Fall ’02, Spring ’03

STANDARDS INVOLVEMENT

- ISO/IEC JTC1 SC-37 WG3 Data Interchange Formats
- INCITS M1: Biometrics
- Technical Editor for INCITS 1829-D: BioAPI Java Interfaces
- INCITS M1.4 Ad-Hoc Group on Biometrics & E-Authentication

RESEARCH INTERESTS

- Statistical Analysis and Testing of Biometrics Systems
- Biometric sample quality assessment
- Electronic authentication using biometrics, management of federated identity systems, and integration of biometrics and cryptography

INDUSTRY EXPERIENCE

US Biometrics – Biometrics Researcher (May 2007 - August 2007)
Designed and analyzed performance tests for biometric system interoperability. The tests focused on integrating a different accquisition subsytem into the existing security infrastructure. Also responsible for requirements analysis for creating standards compliant physical access and logical access biometrics system. Standards included FIPS-201, BioAPI, and INCITS 378-2004.

RESEARCH EXEPERIENCE

Testing of Fingerprint Recognition Systems
Involved in diverse range of projects aimed at understanding impact of image quality, impact of age, impact of environmental conditions, and impact of finger preference on matching performance.

Biometrics and Cryptography
This is a collaborative research project with Center of Education and Research in Information Assurance and Security (CERIAS) to incorporate biometrics in cryptographic protocols like zero knowledge proofs.

Keystroke Dynamics Verification Using Spontaneous Password
Designed and tested multiple keystroke dynamics verification algorithms for an authentication mechanism that used a spontaneously generated password for verification instead of a pre-known, static password.

Biometrics in Federated Architecture Systems
Joint research with Center of Education and Research in Information Assurance and Security (CERIAS) examining methodologies to incorporate biometrics into federated architecture system, and other identity management system architectures.

Equine Iris Recognition – Iristrac™
This research examined the feasibility of using equine eye features for recognition. Extracting equine eye features, development of pattern matching algorithms, and feasibility testing were the main research goals.

Security Analysis of Different Biometrics Storage/Matching Architectures– NIST
Created a technical report for National Institute of Standards in Technology (NIST) which examined architectures and interoperability of current biometric systems. This research also analyzed the current methodologies for revocation of a compromised biometric identifier.

PUBLICATIONS

Chapter Co-Author:

- Elliott, S. J., Kukula, E., & Modi, S. (2007). Issues Involving the Human Biometric Sensor Interface. In S. Yanushkevich, P. Wang & S. Srihari (Eds.), *Image Pattern Recognition Synthesis and Analysis in Biometrics* (Vol. 67, pp. 400): World Scientific Publishers.

Co-Author for the following locally published books:

- Introduction to Biometric Technology.
- Securing the Manufacturing Environment Using Biometric Technologies.

Conference Proceedings

- Modi, S. K., Elliott, S. J., Kim, H., & Kukula, E. P. (2008). *Statistical Analysis Framework for Biometric System Interoperability Testing.* Paper presented at the ICITA 2008, Cairns, Australia.
- Frick, M. D., Modi, S. K., Elliott, S. J., & Kukula, E. P. (2008). *Impact of Gender on Fingerprint Recognition Systems.* Paper presented at the ICITA 2008, Cairns, Australia.
- Bhargav-Spantzel, A., Modi, S., Bertino, E., Elliott, S. J., Young, M., & Squicciarini, A. C. (2007). Privacy Preserving Multi-Factor Authentication with Biometrics. *Journal of Computer Science*
- Modi, S. K., Elliott, S. J., & Kim, H. (2007). *Performance Analysis for Multi Sensor Fingerprint Recognition System.* Paper presented at the ICISS 2007, New Delhi, India
- Modi, S., Elliott, S. J., Kim, H., & Whetstone, J. (2007). *Impact of Age Groups on Fingerprint Recognition Performance*. Paper presented at the AutoID 2007, IEEE Workshop on Automatic Identification Advanced Technologies, Alghero, Italy.

- Modi, S. K., & Elliott, S. J. (2006). *Impact of Image Quality on Performance: Comparison of Young and Elderly Fingerprints.* Paper presented at the 6th International Conference on Recent Advances in Soft Computing (RASC), Canterbury, UK.
- Modi, S. K., & Elliott, S. J. (2006). *Graduate Course Development focusing on Security Issues for Professionals Working in the Manufacturing Industry.* Paper presented at the World Congress on Computer Science, Engineering and Technology Education, Santos, Brazil.
- Modi, S. K., & Elliott, S. J. (2006). *Keystroke Dynamics Verification Using Spontaneously Generated Password.* Paper presented at the 40th IEEE International Carnahan Conferences Security Technology, Lexington, Kentucky.
- Modi, S. K., & Elliott, S. J. (2005). *Securing the Manufacturing Environment using Biometrics*. Paper presented at the 39th Annual International Carnahan Conference on Security Technology (ICCST), Las Palmas de G. C., Spain