Effects of sensors, age, and gender on fingerprint image quality

By

Rong Yang

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Master of Applied Science

in

Biomedical Engineering

Department of Systems and Computer Engineering
Carleton University
Ottawa, Ontario

Abstract

The performance of the widely used fingerprint recognition system is heavily influenced by fingerprint quality which in turn is impacted by different factors. Many different algorithms were developed to measure fingerprint quality. This thesis analyzes the impact of different sensors, fingerprint quality algorithms and demographic factors on fingerprint image quality. Three different fingerprint quality algorithms are tested: vendor specific ones from each sensor manufacturer, NFIQ1 and NFIQ2. Our results showed that fingerprint quality decreases with age and males have better fingerprint quality than females on most sensors. The multispectral sensor has the best and stable fingerprint image quality. NFIQ2 worked well with all the tested sensors while NFIQ1 produced anomalous results on sensor 2, 3 and 9. Vendor quality scores from some sensors are either constant or not usable. These findings can help in selecting sensors for biometric systems with targeted subjects and in improving fingerprint sensor design.

Acknowledgements

I would like to thank:

- Dr. Andy Adler of Carleton University, my supervisor for his guidance, inspiration, encouragement and assistance over the years.

- John Harvey for his help throughout my research.

- My wife Mei and my two lovely daughters for providing me support through all my endeavours.

Table of Contents

List of Tables

List of Figures

Nomenclature

AFIS	Automated Fingerprint Identification System
ATM	Automated Teller Machine
BIR	Biometric Identification Record
BKA	Federal Criminal Police Office of Germany
BMS	Border Management System
BSI	British Standards Institution
CAGR	Compound Annual Growth Rate
COM	Centre of Mass
COTS	Commercial Off-the-shelf
DHS	Department of Homeland Security
ERC	Error Rejected Curve
FAR	False Accept Rate
FBI	Federal Bureau of Investigation
FDA	Frequency Domain Analysis
FG	Foreground
FMR	False Match Rate
FNMR	False Non-match Rate
FRR	False Reject Rate
FVC	Fingerprint Verification Competition
GS	Gabor Segment
IDIS	Identification Document Issuance System
IQR	Interquartile Range

LCS	Local Clarity Score
MLP	Multi-layer Perception
MMB	Mu Mu Block
MMSB	Mu Mu Sigma Block
MRZ	Machine-readable Zone
MSB	Mu Sigma Block
NDA	Deoxyribonucleic Acid
NFIQ	NIST Fingerprint Image Quality
NIST	National Institute of Standards and Technology
OCL	Orientation Certainty Level
OF	Orientation Flow
PDF	Probability Density Function
PKI	Public Key Infrastructure
PS	Power Spectrum
RC	Ridge Count
ROC	Receiver Operating Characteristic
RTVTR	Ridge Thickness to Valley Thickness Ratio
RVU	Ridge Valley Uniformity
SOM	Self-organizing Map
SSB	Sigma and Sigma Sigma Block
USG	United States Government
WSQ	Wavelet Scalar Quantization

1   Chapter: Introduction

1.1   Motivation

Fingerprints have been widely used in authentication and identification systems. It can be found in border management systems (BMS), identification document issuance systems (IDIS), automated fingerprint identification systems (AFIS) and other private or commercial security systems. These systems have two important components – fingerprint enrolment (capture) and verification or identification.

The quality of the captured fingerprint images heavily influences the performance of verification and identification, in terms of both accuracy and efficiency [1]. To have the quality score output by a biometric quality assessment algorithm reflect its impact on the recognition performance of the biometric system, the quality score should convey a predicted utility of the biometric sample [2]. Better quality in both enrollment and verification images result in better performance in fingerprint identification and verification. Different fingerprint image quality measuring methods have been studied [3] [4] [5] [6] [7]. Each fingerprint sensor vendor usually has their own proprietary algorithm for measuring captured fingerprint image quality. An open source fingerprint quality measurement algorithm NIST Fingerprint Image Quality (NFIQ) version 1 was developed by NIST and made publicly available in 2004 [1]. The development of NFIQ version 2.0 was announced in 2010 and was formally released to public in 2016. NFIQ provides a uniform interpretation on fingerprint image quality and overcome the

interoperability problem of proprietary vendor implementation [8]. It was widely deployed in many large-scale biometric systems [8].

Fingerprint image quality can be affected by multiple factors. It can be from sensor design, age, gender (sex at birth) and habituation (when users interact with the biometric system quite frequently, they tend to get habituated in providing their biometric data [9]), etc. Understanding how those factors impact fingerprint image quality can help to improve sensor design or help us to make better decisions during sensor selection based on the type of target population. How age, gender and habituation correlate with NFIQ 1 quality scores was explored already [10]. No independent study on how these factors impact NFIQ2 scores and how NFIQ1 and NFIQ2 align with each other is available. A NFIQ score is based on the classification (neural network or other machine learning method) result of a set of measurable quality features. For NFIQ2, the quality features include minutiae count, minutiae quality, ridge-valley structure frequency domain analysis quality, local clarity quality, etc. By learning the correlation between different factors (sensor, age and gender) and quality features, we can provide sensor vendors much more details about how to improve their sensor design and help users to select sensors based on the target subjects. NFIQ2 was trained with images from USG operational finger images, images from Federal Criminal Police Office (BKA) in Germany, and public datasets [8] which are mostly from law enforcement. There is a strong chance of bias toward the male population due to the male dominant training data set. We want to explore this issue also.

1.2   Thesis Objectives

The main objective of this work is to learn how different sensors, age groups, genders and ethnic origins relate to fingerprint qualities and quality features, specifically for NFIQ2 quality features and the overall NFIQ2 quality scores, and how NFIQ2 scores correlate to NFIQ1 scores in different situations. We also want to determine how well vendor quality scores (found in enrolment BIR record) align with NFIQ1 and NFIQ2 scores.

1.3   Outline

This thesis is split into six chapters:

Chapter 1: Introduction, briefly describes the motivation and objectives of this thesis and summarizes the contributions.

Chapter 2: Background, provides a summarized background of biometric, fingerprint, fingerprint image quality, NFIQ1, NFIQ2 and researches on the correlation of different factors with fingerprint image quality. This research is based on these background knowledges.

Chapter 3: Data, describes how the test data were collected, what devices were used to collect the data and what the composition of the test subject population is.

Chapter 4: Research Method, describes data analysis tools and method.

Chapter 5: Results and Discussions, presents the analysis results and discussions around the results. It uses visualized data description and plots to present the findings.

Chapter 6: Conclusion, summarizes the main findings and contributions and lists the potential future works.

1.4   Contributions

This research reported in this thesis lead to following contributions:

- Comparative studies of the impact of different sensors, age groups, genders and ethnic origins on fingerprint qualities described in NFIQ1 scores and NFIQ2 scores

- Comparative studies of the impact of different sensors, age groups, genders and ethnic origins on fingerprint quality features described in NFIQ2 quality feature scores

- Explores the relationships of NFIQ1, NFIQ2 and its quality features, and vendor qualities

- Explores the effectiveness of vendor quality scores

- Compares the effectiveness of NFIQ1 and NFIQ2 quality scores.

- Identifies the sensors with the best quality performance.

2   Chapter: Background

2.1   Biometrics

Biometrics is a pattern recognition system to determine or verify identity through

acquiring biometric data information about characteristics of a person, either

physiologically or behaviourally, extracting a set of salient or discriminatory features

from the data collected, and comparing this feature set against the stored templates set in

the database to generate matching scores [11]. Unlike the conventional knowledge-based

(what you remember, ex. Password) and token-based (what you have, ex. An ID card)

methods, biometrics can provide positive personal recognition.

A biometric system can operate in verification mode or identification mode. Identity

verification is a one-to-one comparison process to validate a person's identity by

comparing the captured biometric data with his/her own biometric template(s) stored in

the system database. It is used for preventing multiple people from using the same

identity. The identification process is a one-to-many comparison to identify an

individual's by searching in the database for a match. It is used for preventing a single

person from using multiple identities [12].

Individual characteristics could be fingerprint, face, iris, voice, DNA, gait, scars, marks,

tattoos, etc. Biometrics has been used in forensics, ID issuance and other applications for

a long time to identify a person. With the improvement of technology, more and more

systems include biometric recognition as a core component. For example, in my previous

working experience, we developed border management system (BMS), automated fast pass gate and used automated fingerprint identification system (AFIS). All are good biometric system examples. BMS matches a traveller's face and fingerprints with the photo image and fingerprint images stored in e-Passports. AFIS matches submitted fingerprints with a set of fingerprints (or templates) stored in database. Automated fast pass gate retrieves the traveller's enrolled photo and fingerprint templates based on the traveller data on a facilitated travel card and matches them with live captured face and fingerprint images. An ATM machine can match the fingerprints with the enrolled fingerprint template [13]. Even mobile device uses biometric login – fingerprint match initially and face match with the new mobile devices as we can see from iPhone evolvement.

2.1.1   Enrolment and Recognition

A biometric system usually has two key components: Enrolment and Recognition (verification) (Figure 2.1).
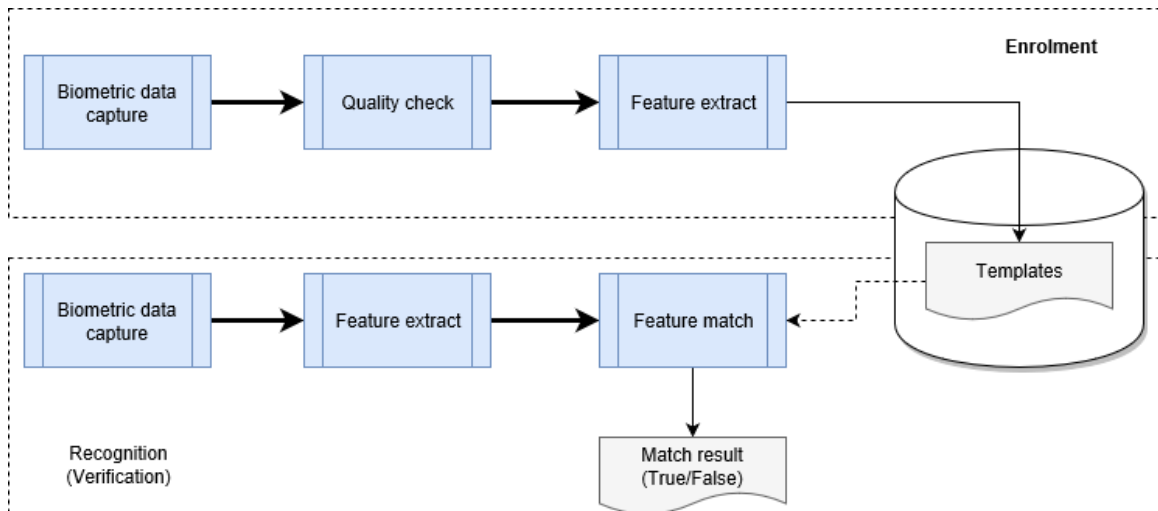
Figure 2.1 Block diagram of the operation of a biometric system, including the enrollment (upper figure) and verification (lower figure)

Enrolment captures the biometric data and generates a feature template. Recognition captures new biometric data, generates a new feature template and then matches the new template with the old one.

2.1.2   False Match Rate and False Non-Match Rate

The accuracy of the biometric system is evaluated by the false non-match rate (FNMR) and false match rate (FMR) which are the probability of type I and II errors of statistics hypothesis testing respectively. In statistical hypothesis testing for biometric system, the null hypothesis is that the input does identify the right person. A type I error is the incorrect rejection of a true null hypothesis (H0), while a type II error is incorrectly retaining a false null hypothesis (

Table **2.1**) **[14]**.

		Null hypothesis (H0) is	
		True	False
Decision About Null Hypothesis (H0)	**Match**	Correct (True Match)	Type II error (False Match)
	Non-Match	Type I error (False Non-Match)	Correct (True Non-Match)

Table 2.1 Table of error types

For a specific biometric system, the match or non-match is determined by matching threshold. FMR and FNMR are related with each other. High FMR means low FNMR. The lower the FMR, the more secure the system but the higher FNMR which means less unauthorized persons get authorized with the cost of more authorized persons getting rejected. The lower the FNMR, the more convenient to process but the higher FMR which means more authorized persons will be accepted with the cost of more unauthorized persons getting accepted also. Every system will have a threshold to balance FMR and FNMR. As Figure 2.2 shows, when threshold moves to higher score, false non-match grows and false match decreases.
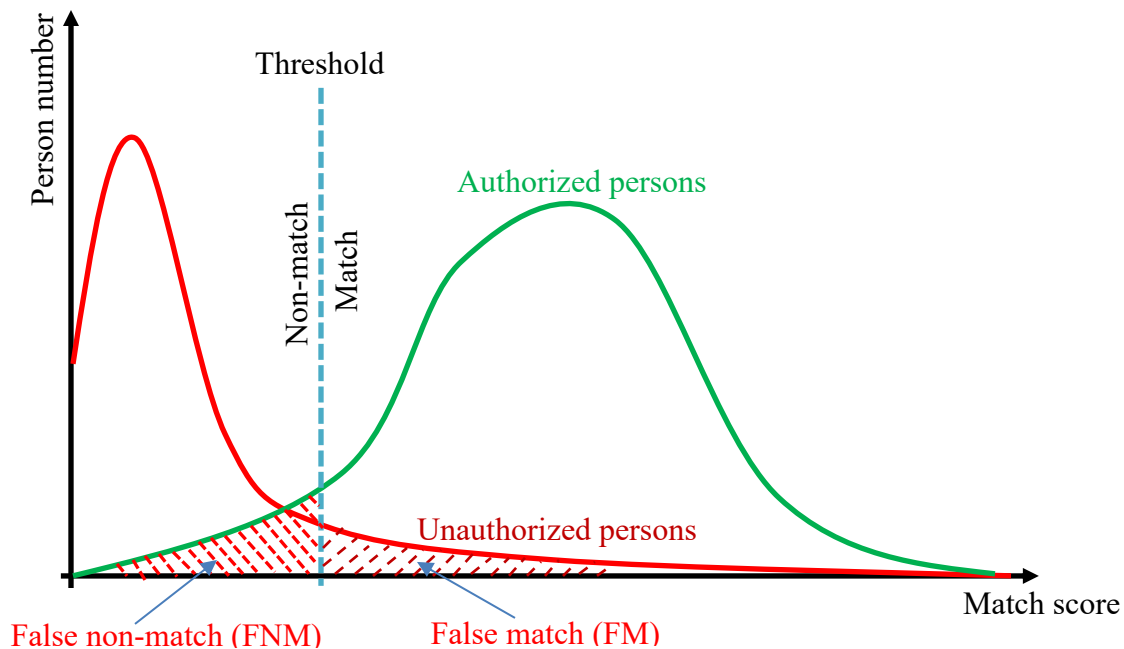
Figure 2.2 Match score distribution of authorized and unauthorized persons

One important component of biometric data quality is utility which reflects the predicted positive (match) or negative (non-match) contribution of an individual sample to the overall performance of a biometric system. Utility-based quality are more predictive of system performance in terms of FMR and FNMR **[8]**. Better quality of biometric data lead to lower FMR and FNMR as Figure 2.3 receiver operating characteristic (ROC) curve shows.
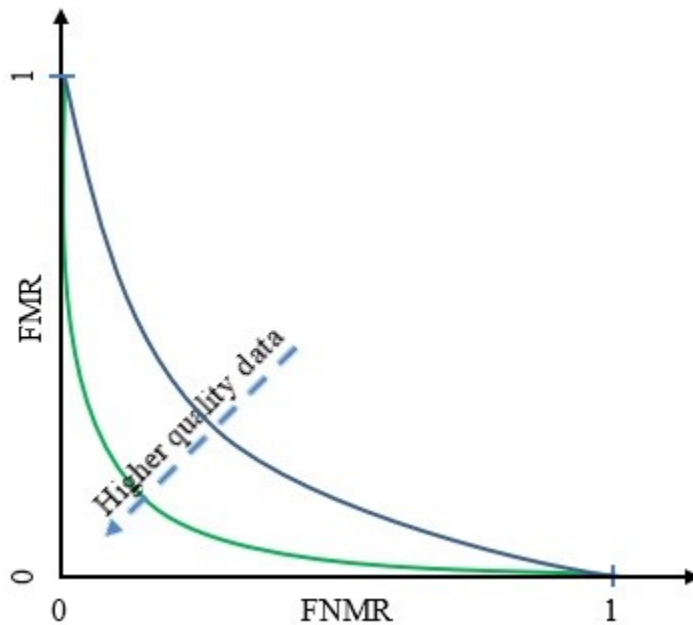
Figure 2.3 ROC curve change when higher quality data is used for matching in a biometric system

2.2 Fingerprint

Fingerprint refers to the friction ridges and valleys on the gripping skin surface of the hands and feet of a person. The ridges have patterns of loops, whorls and arches. The fingerprint is usually defined as an impression left by the friction ridges of a human finger [15]. Human fingerprints are nearly unique, difficult to alter, and durable over the life of an individual, making them suitable as long-term markers of human identity [15]. Fingerprint matching is the most widely used biometrics by law enforcement departments today because of fingerprint's uniqueness and persistence [16]. Edward Henry developed a criminal identification system based on Galton's methods of classifying fingerprints that is still widely used today [17].

10

Live-scan fingerprint recognition systems have largely replaced the traditional ink fingerprint system. Fingerprints can be captured off-line or through live scan devices which capture fingerprint images in electronic format (usually image files). An off-line fingerprint usually is captured by smearing the finger with ink and then press the smeared finger against paper. Then the paper can be scanned to have the fingerprint image enrolled into biometric system. Other methods can be used to capture off-line fingerprint. For example, law enforcement uses forensic technique to get latent fingerprints – impressions left by friction ridge skin on a surface [15]. Most modern biometric systems use live scan devices to capture fingerprints for enrolment and recognition. Even in law enforcement, live scan devices are widely used to capture the fingerprints of criminal subjects.

2.2.1   Identification vs Verification

Fingerprint matching has two operating modes: verification and identification. Verification is a one to one match. The captured fingerprint image matches against specific fingerprint image or template. Verification requires that a target fingerprint image or template be loaded to be matched against. This usually involves claim-based identity (some information used to identify the user) from the user to begin the process. With BMS as an example, the traveller presents e-Passport to start the process. E-Passport readers read the machine-readable zone (MRZ) on the identity page of the passport and then use MRZ data as key and PKI infrastructure to setup a communication session with the chip on the passport to read out face and fingerprint images. The system then requires the traveller to present specific fingers (based on fingerprint position

information read from the chip) to match against the fingerprint image read from the chip to confirm the identity of the passport holder.

Identification is one-to-many matching. The fingerprint features are extracted from the new captured image and then used to match against a set of fingerprint feature templates in a biometric database which are generated from previously captured fingerprints from different subjects. As the templates in biometric database are directly linked to different subjects, when a match is found, the related subject then is identified.

2.2.2 Automated Fingerprint Identification System (AFIS)

An automated fingerprint identification system (AFIS) is a special system specifically used for fingerprint identification. It searches a set of fingerprints against an existing fingerprint database and returns the matching records. AFIS research started at the early 1960s by FBI with the help of National Institute of Standards and Technology (NIST) [18]. Now it's widely used by law enforcement such as RCMP in Canada and other civil applications like border management system (BMS) for criminal identification and civil background clearance. Most law enforcement AFIS systems can have better than 99% accuracy in search a set of known fingerprints [18].

With the increasing threat of terrorism, more and more countries or law enforcement agencies get support from AFIS system. Based on the market forecast from Markets and Markets, the automated fingerprint identification system market size is estimated to reach USD 8.49 Billion by 2020, at an estimated CAGR of 21.0% between 2015 and 2020 [19].

2.3 Fingerprint Quality

As stated in ISO/IEC 29794-1 [20], biometric sample quality is defined having three components:

- Character: inherent features of the source from which the biometric sample is derived. For fingerprint, the ridges and valleys are inherent features.

- Fidelity: the accuracy of the derived features. This reflects the similarity of the biometric sample and the source. For fingerprint, it's the question of whether the fingerprint image looks different from its source finger.

- Utility: the predicted contribution of the sample to the performance (in terms of FMR and FNMR) of a biometric system. Utility-based quality is dependent on both character and fidelity of the sample. It should predict the match score which the biometric system uses to determine match or non-match. Good quality means better match prediction (less FMR and FNMR).

Adler et al [21] define biometric information (BI) as the decrease in uncertainty about the identity of a person due to a set of biometric measurements. Low quality images are less informative.

The score from a good quality measurement algorithm should be accurate in reflecting the utility attribute of the fingerprint quality. Bad quality fingerprint can be rejected at the image capture phase. It means good quality fingerprints get enrolled or used to do fingerprint matching and so allow improved system performance.

The number of fingers used and fingerprint quality captured and present in the database will directly affect the performance of a fingerprint matcher. Poor quality fingerprint images will reduce the system's identification accuracy. High quality of fingerprint images must be ensured in the enrollment process of a biometrics system. Different studies on fingerprint quality assessment have been conducted since 1990s (Section 2.3.1 and 2.3.2). Three types of assessment approaches (block-based or local qualifying approaches, global quality assessment methods, and machine learning-based solutions) were used in the past studies. The first type is to break the fingerprint image into blocks (local features of the image) and obtain the whole image quality via a combination of each block's quality. The second type is to analyze the fingerprint features and qualify fingerprint image globally (global features of the image). To avoid the disadvantages of local (costly) and global level (missing local detail) quality assessment methods, the third type was proposed to consider both local level quality criteria and global quality index either linearly combined or implemented by other solutions.

2.3.1 Fingerprint quality measurement researches

Hong et al [22] proposed a fingerprint enhancement algorithm to quantify the quality of fingerprint using both the local ridge orientation and frequency information. They showed that their algorithm can identify and remove unrecoverable corrupted regions, improve both the goodness index of the extracted minutiae and the accuracy of fingerprint verification through improving the clarity of ridge and furrow structures of fingerprint images. Bolle et al [23] used the ratio of the area of selected contiguous regions of blocks to the total area of the fingerprint image (i.e., the foreground.) to

measure the quality of fingerprint images by dividing the image into directional or non-directional pixels.

Different ways of fingerprint quality measurement have been explored. Ratha and Bolle [24] proposed a wavelet compressed method to estimate fingerprint image quality. Shen et al [3] proposed a Gabor-feature based method to estimate the quality of the fingerprint images by dividing images into blocks. Lim and Yau [4] used orientation certainty to estimate the local and global quality measures of 150 fingerprint images and found that their method is feasible in detecting low quality and invalid fingerprint images. Lee et al [25] used the Fourier spectrum method to measure fingerprint image quality based on global characteristics of the image.

Chen et al [26] developed a global quality index to measure the energy concentration in the frequency domain as a global feature and a local index to measure the spatial coherence in local regions. Luo [27] proposed a fast lifting wavelet algorithm to analyze spatial-frequency of images. An image is divided into blocks and transformed into wavelet domain and sub-band image domain. An artificial neural network is used to optimize weights to obtain quantitative measure of image quality. Lee et al [28] proposed a model-based method to estimate the quality of fingerprint images by comparing the similarity between the 1-dimensional probability density function (PDF) of the sinusoidal wave and the input fingerprint image. The PDF is obtained by projecting the 2-dimensional gradient vectors of the ridges and valleys in the orthogonal direction to the local ridge orientation. Fronthaler et al [29] studied the orientation tensor of

fingerprint images to quantify signal impairments like noise, lack of structure, blur, with the help of symmetry features. They used a multi-algorithm fingerprint recognition system with a trained Bayesian scheme to do quality estimation which can save computation time increase recognition rates.

2.3.2   NIST Fingerprint Image Quality metric (NFIQ)

The NIST Fingerprint Image Quality metric (NFIQ) [30] uses 11 elements (Table 2.2) of the extracted minutiae and a trained neural network to predict the quality (1 for excellent and 5 for poor) of a fingerprint image. Merkle et al. [31] [32] proposed a method to optimize the NFIQ algorithm by re-training the neural network but the improvement is marginal. Olsen et al [33] proposed a fingerprint quality measure based on the Gabor Filter responses by relating the quality of a fingerprint sample to the biometric performance. They suggested this method to be a candidate quality measure for NFIQ 2.0. Ninassi et al [34] proposed a utility-based quality assessment metric combining image quality criterion and pattern-based quality criteria (salient and patch-based features) through a weighted sum optimized by a genetic algorithm using the Fingerprint Verification Competition (FVC) 2002 DB3 fingerprint database. Awasthi [35] used a Logistic regression model to calculate a quality index which takes into consideration of both local and global features of the image and found about two thirds accordance to the NFIQ approach of quality assessment. Yao et al [36] proposed several minutiae-based features to evaluate fingerprint quality based on the triplet representation of minutia point using three FVC databases and quality metric evaluation approach. They used a utility-

based quality metric approach in the feature extraction phase. Their quality value is a linear combination of the features.

Feature	Description
foreground	number of blocks that are quality 1 or better
total #of minutia	number of total minutiae found in the fingerprint
min05	number of minutiae that have quality 0.5 or better
min06	number of minutiae that have quality 0.6 or better
min075	number of minutiae that have quality 0.75 or better
min08	number of minutiae that have quality 0.8 or better
min09	number of minutiae that have quality 0.9 or better
quality zone 1	percentage of the foreground blocks of quality map with quality =1
quality zone 2	percentage of the foreground blocks of quality map with quality =2
quality zone 3	percentage of the foreground blocks of quality map with quality =3
quality zone 4	percentage of the foreground blocks of quality map with quality =4

Table 2.2 NFIQ1 Features (Reproduced from reference [30])

2.4   NFIQ1

The fingerprint verification system will perform well for good quality fingerprint images. An automatic fingerprint recognition system has three subsystems: capturing fingerprint image, extracting fingerprint features and matching the captured image with that in the database. The fingerprint images can be captured by either from the live scan of a person's finger or from an inked impression of a person's finger on paper. Feature extraction is the process of obtaining the features of the captured image. Matching is the process of finding the degree of similarity between the captured image and the gallery sample of the same person in the database. The quality of captured image is essential for the fingerprint matching process.

To measure the fingerprint match performance, the National Institute of Standards and Technology (NIST) developed an artificial neural network based algorithm - Fingerprint Image Quality (NFIQ 1.0) [37] algorithm for the Federal Bureau of Investigation (FBI) and Department of Homeland Security (DHS) in 2004. NFIQ algorithm is a machine learning algorithm which is publicly available and independent to the fingerprint verification software. NFIQ was implemented in the C programming language. It was tested on 20 different live scan and paper fingerprints datasets using 14 fingerprint software development kits supplied by 8 commercial fingerprint vendors. Over 40,000 of fingerprint images from 7 different databases (22 different datasets and each test dataset has 2 fingerprint images of 6000 person) were used to train the neural network and 15 different COTS fingerprint matching algorithms were used to predict the quality of image class from the 11 dimensional features (Table 2.2).

NFIQ uses the ranking of image quality to predict the performance of a fingerprint matching system. The NFIQ algorithm has four parts (Figure 2.4).
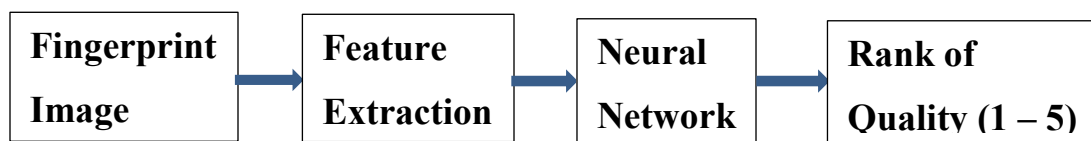

Figure 2.4 Framework of NFIQ1 Algorithm

NFIQ is a fingerprint image quality measuring algorithm and it assumes that higher image quality yields higher genuine (i.e. same person) comparisons match scores. NFIQ's scores are intended to be used to predict the relative performance of a minutiae-based fingerprint matching system. A minutia matching algorithm calculates a similarity score

through comparing local ridge characteristics (minutia) of two fingerprint images (biometric samples). A similarity matrix is produced after comparing all fingerprint images. The normalized match score then measures the quality of fingerprint image. The input of the NFIQ algorithm is a fingerprint image and its output are an integer number as a quality measure for the image. The processing part includes two steps: NFIQ first computes fingerprint image fidelity characteristics and extracts 11-dimensional features such as the numbers of minutiae and image quality index blocks. Then the 11-dimensional fingerprint features are passed to a multi-layer perception (MLP) neural network to classify the image quality into five levels of quality based on various quantiles of the normalized match score distribution. In the end an integer value between 1(highest quality) and 5 (poorest quality) is obtained to show the quality of the image [38]. The input nodes of the neural network are feed-forwardly connected to the hidden nodes, which are feed-forwardly connected to the output node. Figure 2.5 shows the NFIQ2 multi-layer perception (MLP) Neural Network in a simple way. Default hidden layer number 22 is from NFIQ2 source code.
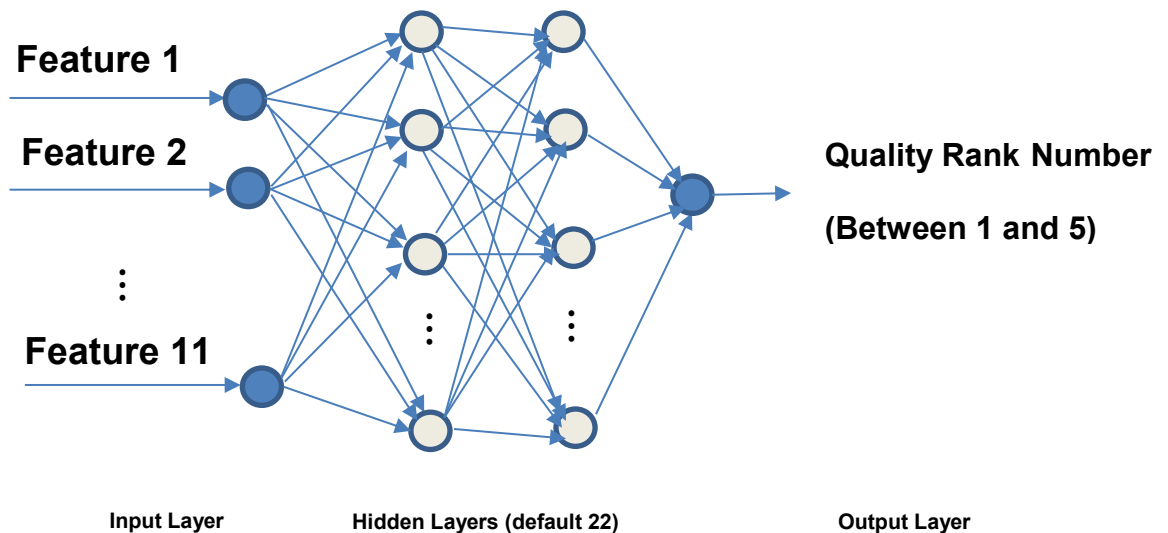

Figure 2.5 Framework of a multi-layer perception (MLP) Neural Network

19

The fingerprint image data and features are split into two groups: training set and test set. The training set is used to train the neural network and the test set is used to check the prediction accuracy and to stop training as soon as the accuracy is reached. A scaled conjugate gradient algorithm [39] combining with limited memory Broyden Fletcher Goldfarb Shanno algorithm [40] is used to optimize the weights. Boltzmann pruning [41] is used for dynamic removal of connections during the training. NFIQ numbers are used as a relative performance prediction of a minutia-based fingerprint matching system. They tell the positive or negative contribution of each sample to the overall performance of the system.

2.5   NFIQ2

NFIQ 2.0 was released in 2016. It is an open source quality assessment algorithm for fingerprint images which is a new and improved version of NFIQ 1.0. The development of NFIQ 2.0 was announced at the NIST March 2010 workshop on "The Future of NFIQ". During the workshop, NIST and BSI requested submission of Software Development Kit (SDK) systems, suggestions to composition and computation of NFIQ 2.0 features and fingerprint images with NFIQ 1.0 anomaly. Then a joint project to develop the new and improved open source NIST 2.0 was set up by National Institute of Standards and Technology (NIST) and Federal Office for Information Security (BSI) and Federal Criminal Police Office (BKA) in Germany, MITRE, Fraunhofer IGD, Hochschule Darmstadt and Secunet as well as research and development entities [8]. In 2011, the fingerprint image training data set and feature set were selected. The quality feature definitions and algorithm processing steps were documented in 2012. The NFIQ

2.0 core technical specification such as framework design and implementation, feature

selection and implementation were completed in 2013. NFIQ II Lite (2014) [42]

introduced a new method of fingerprint quality assessment based on a combination of

(self-organizing map [43]) and supervised (Random Forest [44]) machine learning

techniques. A self-organizing map (SOM) is trained to learn the spatial information

content to cluster blocks of a fingerprint image. This new representation of the finger

image becomes input for the Random Forest trained to predict the relationship between

the SOM output and biometric performance.

NFIQ2 development used a modular approach. Figure 2.6 shows the overall NFIQ2

Framework. It defines four generic interfaces for input/output, quality feature extraction,

utility estimation, and machine learning modules. A module implements one of the

interfaces and plugs into the framework. Input/output modules load fingerprint images or

templates from a file system or database and output the quality and quality feature scores.

Quality feature extraction modules extract feature data from the image. Utility estimation

modules compute the feature utility values, fusion, and histogram bin selection. Machine

learning modules do model training or final NFIQ2 quality score prediction.

NFIQ2 source code was released in two different packages: main NFIQ 2.0 and

operational software package. The main NFIQ 2.0 follows the structure displayed in

Figure 2.6 and all the developing modules. The operational software package has

optimized code which does not follow the structure in Figure 2.6 and can't be used to do

model training. For our study, we only use the operational software to generate NFIQ2 quality scores and quality feature scores.
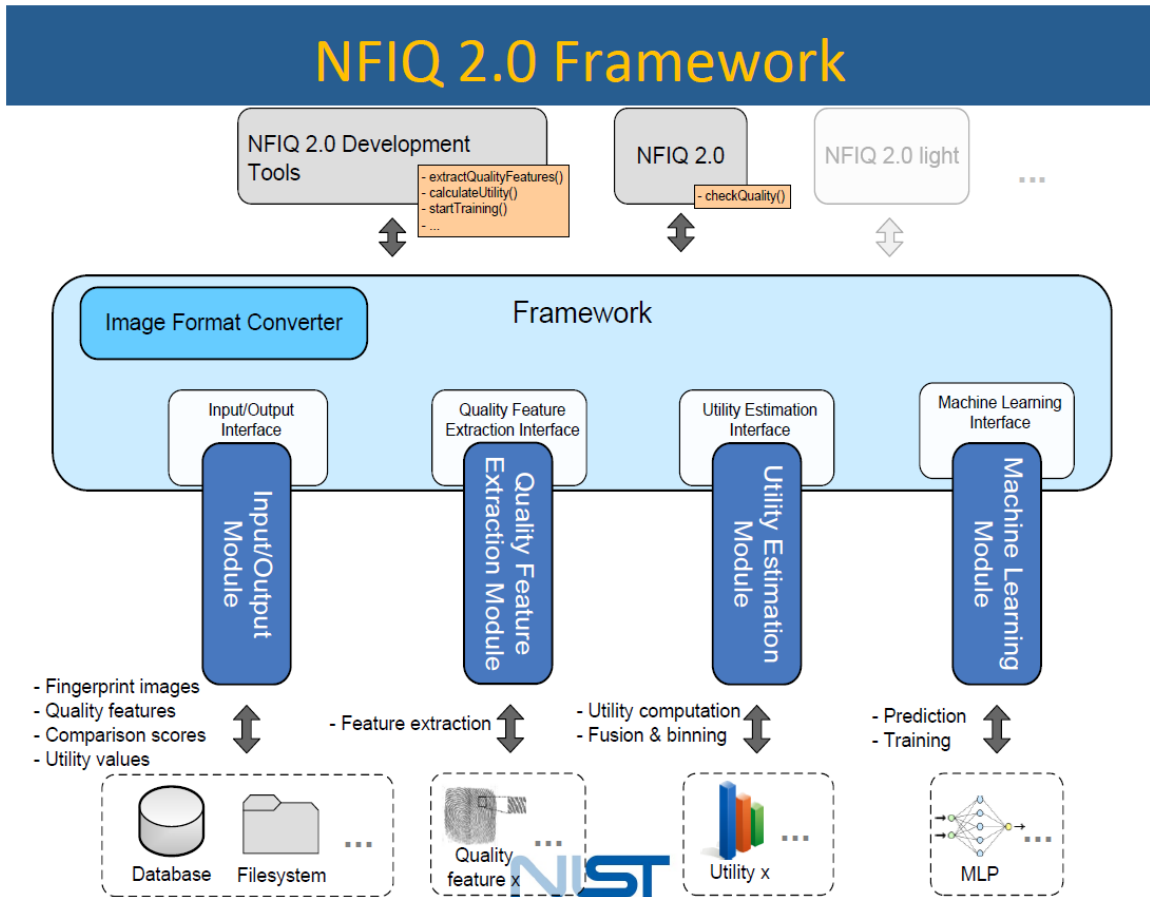
Figure 2.6 NFIQ 2.0 Framework (reproduced from reference [45] presented at NFIQ 2.0 workshop)

Based on [46], compared to NFIQ 1.0, NFIQ 2.0 provides a higher resolution quality score (1 to 100) which aligns with the international biometric sample quality standard ISO/IEC 29794-1:2016, lower computation complexity, as well as support for quality assessment in mobile platform. NFIQ 2.0 quality features are used for revising the international standard of Biometric sample quality – Part 4: Finger image.

NFIQ 2.0 was developed in C/C++ for images captured at 500 dpi and plain impression captured using optical sensors or scanned from inked card. It uses a trained random forest classifier to predict the utility of fingerprint image. Three sources of WSQ compressed images were used in training, validation and evaluation of NFIQ 2.0.

Similar to NFIQ 1.0, NFIQ 2.0 has two major computation steps: feature extraction and training of a machine-learning algorithm. The features were selected by first removing features with low predictive power or above NFIQ 2.0 threshold computation time and then removing redundant features based on Spearman rank correlation. The white pixels on the margins are removed and a local standard deviation algorithm is used to obtain the fingerprint foreground mask in the process of segmenting [8].

NFIQ 2.0 features include following global and local features:

1) NFIQ1 quality features:

   Foreground (FG), Minutiae counts, Quality zone (blockwise) counts

2) FingerJet FX Features:

   Minutiae Count, Fingerprint Quality, Minutiae quality counts, Average minutiae quality

3) Quality features implemented HDA Olsen:

   Frequency Domain Analysis (FDA), Local Clarity Score (LCS), Orientation Certainty Level (OCL), Orientation Flow (OF), (Radial) Power Spectrum (PS), Ridge Valley Uniformity (RVU), Gabor, Gabor Segment (GS), Gabor Shen

(GSh), Mu, Mu Mu Block (MMB), Mu Mu Sigma Block (MMSB), Mu Sigma

Block (MSB), Sigma and Sigma Sigma Block (SSB)

The predictive power of NFIQ 2.0 was evaluated by spearman correlation, receiver

operating characteristic (ROC) curves and error rejected curve (ERC) using 9

impressions (taken with 3 sensors) of 8784 fingers from 1098 individuals from BSI

FingerQS database.

The neural network was trained by 3295 Class 0 and 3334 Class 1 images and validated

by 99797 random images using Random Forest for binary classification (Class 0 - low

utility and Class 1 - high utility). The inputs of the neural network are 69 dimensional

features of images and the output is the score between 1 and 100 of input being Class 1.

2.6　Correlation of Different Factors with Fingerprint Image Quality

There are many factors impacting fingerprint image quality [47], which might be

classified as internal and external factors. The ridge-valley structure and sensing

technologies can be deemed as internal factors. However, the external factors have

broader scopes such as follows:

- environmental conditions - illumination, temperature, or installation height

- demographic impacts - age, gender, or medical/physical impairments

- skin conditions - moisture, oiliness, elasticity, dirtiness, temporary or permanent

  cuts and bruises, and temperature of the skin

- human interaction type with the sensor - swipe or touch

- sensor conditions - dirtiness, noise, and size

- sensor types - optical sensor, capacitive sensor, and thermal sensor, etc.

- design or shape of a sensor

The impact of some of these factors cannot be avoided and some of them vary over time. The sensing technology and interaction type (swipe or touch) are moderately and weakly correlated respectively with image quality scores [48]. There is no significant correlation found between image quality scores and the skin characteristics [48].

One study examined the effects of scanner height on fingerprint capture and found that the fingerprint quality decreased as the height of the work surface decreased [49].

Theofanos et al. [10] confirmed the general consensus that demographic factors do impact fingerprint image quality. People in the age 18-25 age range give consistently good prints in their study, while older individuals have lower print quality. Print quality differed from day-to-day for every individual, but there was no overall trend toward significantly higher or lower quality prints over time. This study also confirmed men give higher quality prints than women. This could due to the fact found in [50] that females have higher Ridge Thickness to Valley Thickness Ratio (RTVTR).

Kang et al. [51] researched 33 habituated cooperative subjects using optical, semiconductor, tactile and thermal sensors throughout a year in uncontrolled environment. It has been observed that the image quality decreases when the temperature goes below zero due to the dryness of the skin [51]. Although all the sensors produce no major image degradation as the temperature changes, they, on the whole, give good

quality images above room temperature [51]. The same goes to air humidity. As far as the pressure of the finger on the sensor is concerned, the image quality is always good with the middle level. For optical sensors, the foreground image gets smaller at low pressure while the fingerprint is smeared at high pressure. Semi-conductor sensors produce good images not only with moderate pressure but also with high pressure. It is very interesting, however, that the tactile sensor gives better images at low pressure than at high pressure [51]. It is also observed that the skin humidity affects the image quality of all the sensors except the thermal sensor which is a sweeping type. Overall, the quality of fingerprint images is more affected by the human factors such as skin humidity and pressure than the environmental factors such as air temperature and air humidity [51].

When a biometric system has a specific target group of subjects, the correlation of demographic factors (age, gender, ethnic origin) with fingerprint quality has a direct effect on the performance of the system. Elliott et al. [52] and Modi et al. [53] compared the fingerprint quality across two populations (18-25 and 62+) with vendor algorithms and NFIQ1 as the quality measuring methods. Modi et al. [54] then extended the research to 4 age groups (18–25, 26–39, 40–62, and 62+). These research results show that fingerprint image qualities are different across age groups, especially 18-25 and 62+ groups. Aging causes collagen loss, skin loose and dry and so less structural firmness, and reduced capability to interact with the sensor due to medical conditions [54]. Figure 2.7 reproduced from [54] shows representative samples of high and low fingerprint quality for each of the four test groups of [54]. Figure 2.8 [54] shows a boxplot of quality scores across the four test groups of [54]. From the boxplot we can see that the

fingerprint quality scores decrease with the increasing age and the difference between age

group 26-39 and 40-62 is not so apparent though.
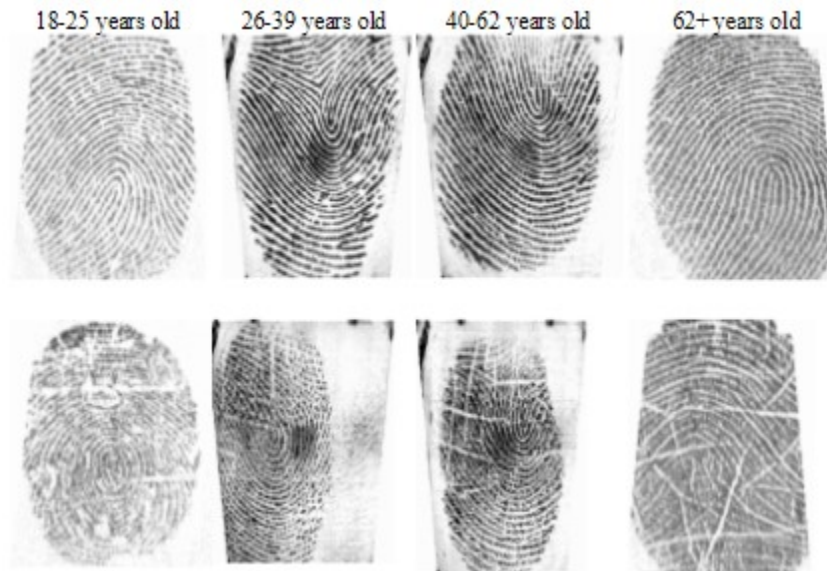
Figure 2.7 Representative high quality (top row) and low quality (bottom row) fingerprints
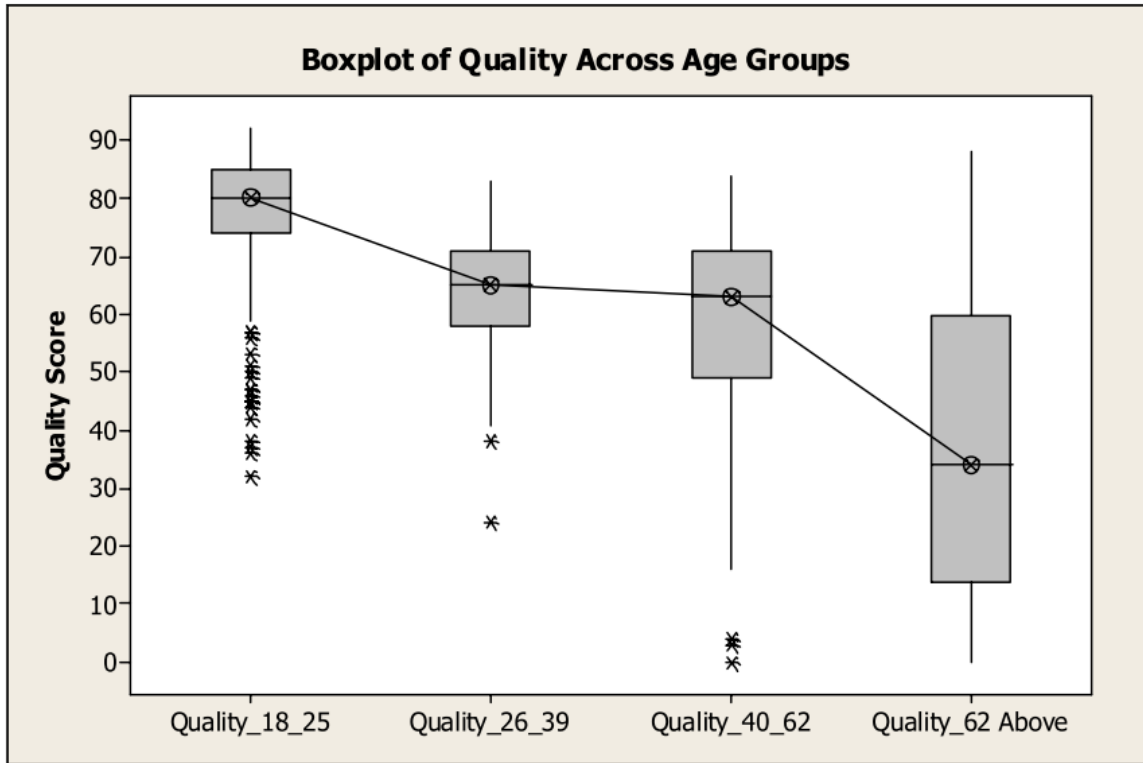
(reproduced from reference [54])

Figure 2.8 Box-plot of quality scores from research (reproduced from reference [54])

Thakkar [55] describes how optical, capacitive based and multispectral imaging sensors work and how the captured fingerprint image quality is affected with different sensors.

NFIQ2 score and its subsidiary quality feature scores have not been widely investigated in researches as fingerprint quality reference. How different factors correlate with NFIQ2 quality score and quality feature scores are unknown yet. How NFIQ2 scores align with NFIQ1 scores and other quality measurement scores is also need to be studied.

2.7 Summary

This chapter presents some basic concepts of biometric, fingerprint and fingerprint quality. Fingerprint quality measuring methods are introduced. Two fingerprint quality measuring algorithms (NFIQ1 and NFIQ2) are briefly described in terms of their evolution and working mechanism. Related researches on the correlation of different factors with fingerprint image quality are also explored.

3   Chapter: Data

This chapter describes the data, data collection methodology and population, on which the analyses of this thesis are performed.

3.1   Data Capture

Fingerprint images were captured in a relatively fixed number of subjects in two visits each year in 2006, 2008, 2012 and 2013 at Bion Biometrics offices on Glencoe Street in Ottawa and Carleton University campus. The two visits in each year are two to three weeks apart and the same capture and verification processes were performed in each visit - enroll each subject on each biometric sensor with two index fingers (finger position 2 and 7) and then verify them with two index fingers against each biometric sensor. For enrolment, 3 fingerprints were captured for each finger per sensor (9 sensors in total). For verification, 6 fingerprints were captured for each index finger per sensor. Ex. for subject x who appears in visit a, there will be 3 left index fingerprint images and 3 right index fingerprint images captured for each sensor during enrolment, 6 left and 6 right index fingerprint images for each sensor during verification. So, a subject will have 18 fingerprint images captured for each sensor so total 162 images on 9 sensors for one visit. For data accuracy, the test subjects would be instructed which sensor to use, which finger to present, and when to put the finger and remove it [5]. If the skin of the test subjects were too dry and the sensor software won't accept the fingerprint image, the subjects were asked to moisten their hands to facilitate the fingerprint capturing. If it's hard to have two index fingers enrolled, other fingers could be asked to be used for enrolment.

For enrolment, a fingerprint BIR (Biometric Information Record) was generated from the fingerprint images of one or two fingers (one image from each finger) for the subject on a specific sensor and so total 9 BIRs for a subject in a visit. BIR is a standard based data structure which holds biometric data in a standard format so to support international interoperability. Our fingerprint BIR is FMR (fingerprint minutiae record) which contains fingerprint template which has finger position, fingerprint image quality score, and fingerprint minutiae data (position, angle) for each finger in the pair. The fingerprint image quality score in the template was calculated by sensor vendor proprietary algorithm based on captured images so to be called vendor quality score in this thesis. Vendor quality scores were extracted out from BIR data in 2018 for this research. All the captured fingerprint images and BIRs were stored in Microsoft SQL database during image capturing for easy query. During verification, there is no BIR generated and so no vendor quality scores can be retrieved from existing verification image data.

NFIQ1 tool then was run against all the fingerprint images in 2013 to calculate NFIQ1 score for each image. These NFIQ1 fingerprint image quality scores were stored in database tables with related fingerprint images.

NFIQ2 tool was run against all the fingerprint images in 2018 to calculate NFIQ2 quality score and NFIQ2 quality feature scores for each image. Those scores were saved in database tables with related fingerprint images.

Figure 3.1 shows the data capture workflow. During a visit, first subject information, visit and sensor information is collected. Then the subject puts fingers on the sensors one by one for enrolment. Enrolment images and templates are generated during enrolment. Following that the subject presents fingers on each sensor for verification match. Verification images are captured and verification scores are recorded. NFIQ1 tool runs against enrolment and verification images to generate NFIQ1 quality scores. NFIQ2 tool then runs against enrolment and verification images to generate NFIQ2 quality scores and quality feature scores. Vendor quality scores are retrieved from enrolment templates. In the data analysis phase, subject, visit and sensor information along with NFIQ1 quality scores, NFIQ2 quality scores, NFIQ2 quality feature scores and vendor quality scores are studied to draw interesting conclusions.
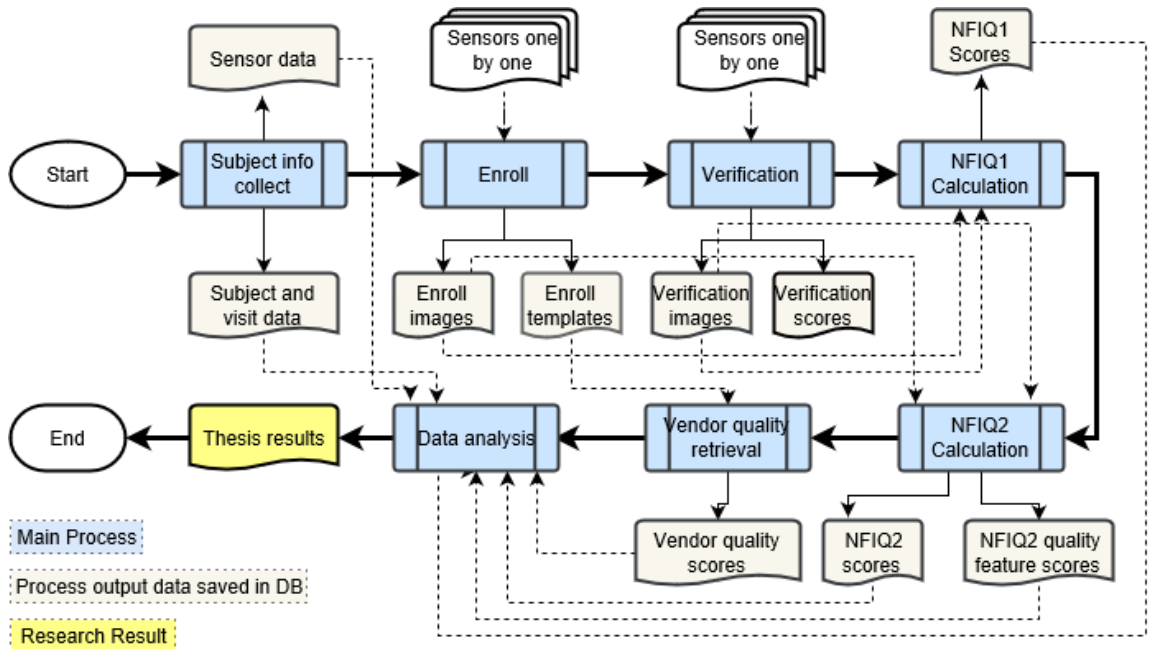

Figure 3.1 Data capture workflow

3.2   Capture Devices

Three of the most commonly used fingerprint sensor types were used in this study. They are optical sensor, capacitive based sensor and multispectral imaging sensor [13]. Optical sensor is the most common and economic one. It uses LED light to illuminate the finger pressed on glass plate and CCD or CMOS based detectors to detect the light and dark areas corresponding to fingerprint ridges and valleys so to generate the fingerprint images. The fingerprint image quality of this type of sensors can be affected by other light sources or surface contamination.

Capacitive based sensors use a set of capacitor plates to sense the capacitance difference of the ridges and valleys of the finger. It solves the other light source and surface contamination issue but can be worn out so image quality can get worse over time.

Multispectral imaging sensor is a kind of optical sensor but it can sense the through the skin surface so to have a second representation of the pattern on the fingerprint surface to compensate the regular surface features. It can get good quality fingerprint images under a variety of conditions [14]

Nine fingerprint sensors were used for capturing the image data for this research. These sensors are from different vendors. Seven of the sensors are optical sensors, one capacitive semi-conductor and one multispectral sensor. The captured fingerprint image sizes are different also. Each sensor came with its own vendor specific software set which are used to do image capturing, quality checking, enrolment (template BIR generating)

and verification (matching with enrolled template). Thus, the generated quality score

calculation, fingerprint minutiae extraction and matching algorithm are proprietary and

might be different from sensor to sensor. The captured image resolution of all the sensors

except Multispectral sensor are the same - 500 dpi (dots per inch). Multispectral sensor

captures images in 500 dpi but resamples the images and so it has more pixels. Each of

the sensor was assigned a sensor id (Table 3.1) instead of using sensor model to represent

the sensor to keep the sensor detail information private. Sensor technology, size of

captured image and a sample image with explicit consent from the subject are also listed

in Table 3.1 Test fingerprint capture sensors.

Sensor Id	Sensor Technology	Size of Captured Image	Sample Image
02	Capacitive Semi-Conductor	256 × 360	

03	Standard Optical	456 × 480	
05	Standard Optical	620 × 620	
06	Standard Optical	512 × 512	
07	Standard Optical	524 × 524	

08	Standard Optical	400 × 512	
09	Multispectral Sensor	1788 × 2764	
10	Standard Optical	512 × 512	
12	Standard Optical	608 × 480	

Table 3.1 Test fingerprint capture sensors

3.3 Test Subjects

Subjects of a wide range in age, gender and ethnic origin were recruited for the research. Subjects were asked to sign the consent form after reading the experimental procedure and project summary prior to start data collection session [5]. Subjects were reimbursed in the form of gift cards for their involvement in data collection at the end of every data collection session. Each subject was assigned a unique identification number for tracing data collection. Subject birth year, gender and ethnic origin were collected and linked with the assigned unique id number at the time of the subject's first visit.

The data were captured through 8 visits in over 4 years (two visits for each year). Each visit was assigned a unique visit id (Table 3.2). Initial design was to keep the same subjects through all the eight visits so other research on fingerprint aging can also be carried out on the same data set. Due to various reasons (moving away, loss of interest in participating, sick, etc.), some subjects dropped off across the visits. For some visits, new subjects could be recruited. Figure 3.2 and Figure 3.3 are the total number of subjects for different enrolment and verification visits.

VisitID	Date
1	Feb 2006
2	Mar 2006
3	Sep 2008
4	Oct 2008

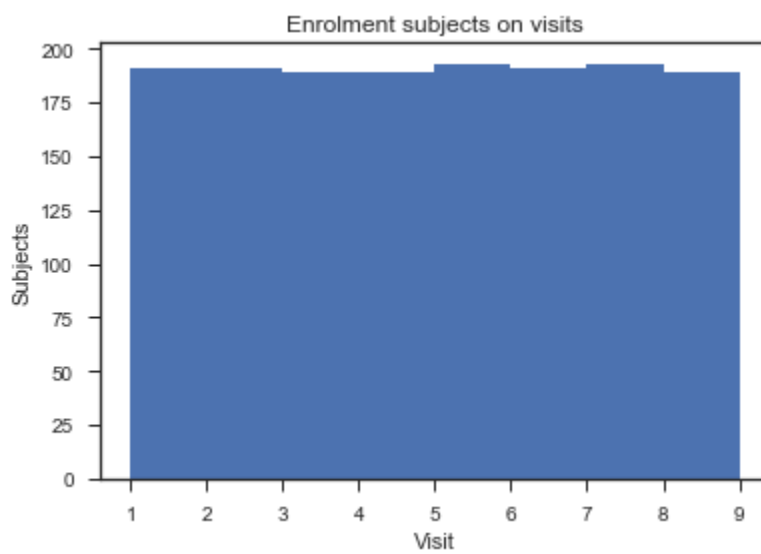|	
5	Feb 2012
6	Mar 2012
7	Mar 2013
8	Apr 2013

Table 3.2 Data capture visits

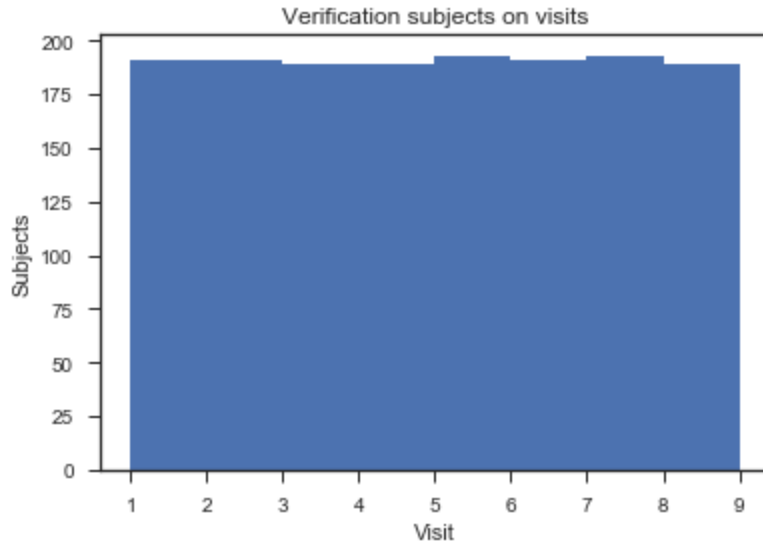**Figure 3.2 Histogram of subjects on enrolment visits**

Figure 3.3 Histogram of subjects on verification visits

There are 191 subjects for visit 1,2 and 6, 189 subjects for visit 3, 4 and 8, 193 subjects for visit 5 and 7.

Here is the subject distribution in each visit for age groups (Figure 3.4). There are more subjects in 45 to 59 years old than other age groups and less in 60 and up years old. We can also see that almost no change for the two visits in the same year though (1-2, 3-4, 5-6, 7-8).
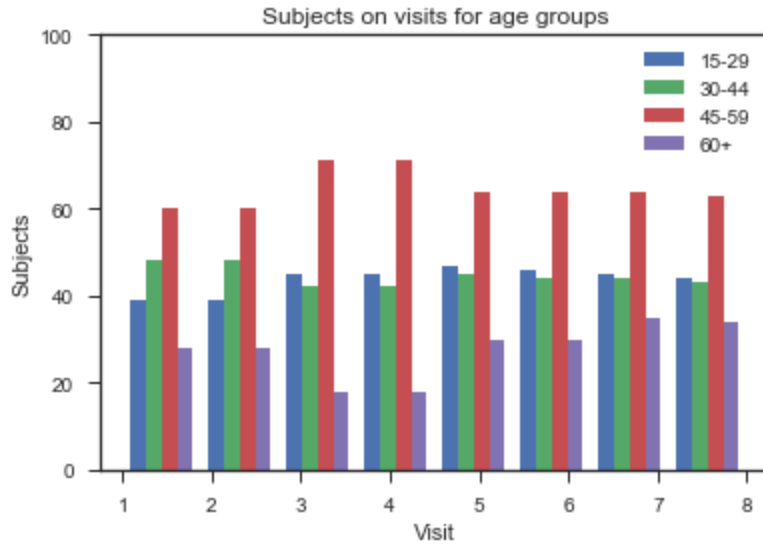
Figure 3.4 Subject distribution in each visit for age groups

Figure 3.5 is the subject distribution in each visit for male and female. The gender ratio in the population is quite balanced across the visits.
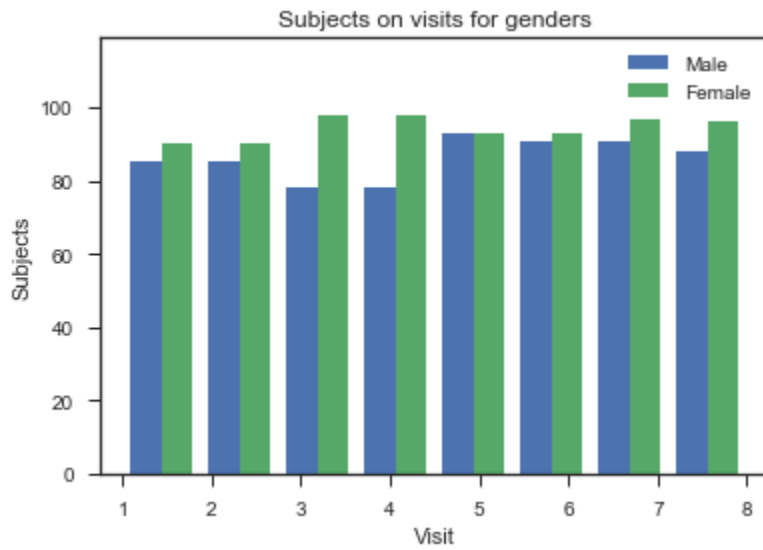

Figure 3.5 Subject distribution in each visit for genders

Figure 3.6 is the subject distribution in each visit for each claimed ethnic origin. It shows that the recruited subjects are mainly from North America, Europe and Asia.
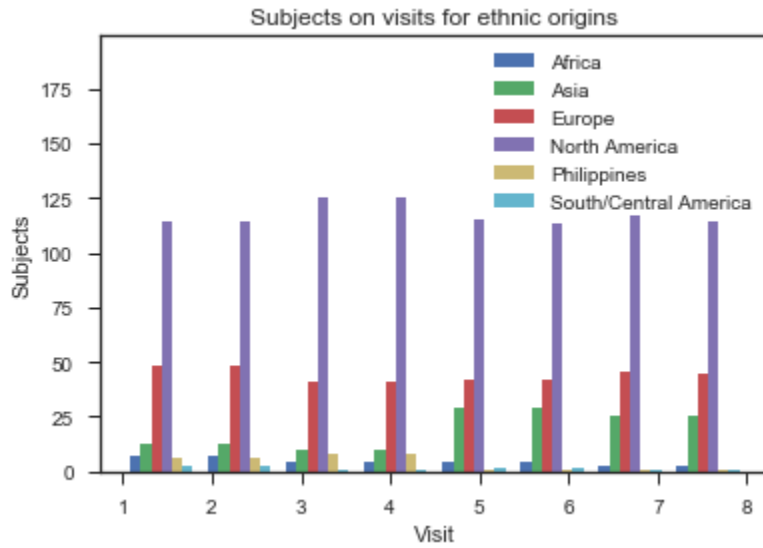

Figure 3.6 Subject distribution in each visit for ethnic origins

4 Chapter: Research Method

This study uses a comparative analysis method to investigate the impact of different sensors, age groups, gender and ethnic origin on fingerprint quality represented by NFIQ2 quality score and fingerprint quality features represented by NFIQ2 quality feature scores. The correlations of NFIQ1 score with NFIQ2 score and NFIQ2 quality features are also studied.

Python was used to write different tool methods to generate different plots to visualize the comparisons. The boxplot and histogram images used in this study are mostly generated by these codes. Several statistics related python libraries are used by those tool methods, such as scipy, matplotlib, sklearn, seaborn, pandas and numpy.

The research data was extracted as csv file from the database based on specific queries so to have all the required data aggregated together. Inside the output data, ages are calculated with birth year of the subject and the time of the visit. Age groups then come from the age and are divided into "15-29", "30-44", "45-59" and "60+" four groups witch 15 years as the range of each group. ProductId is used to represent assigned sensor id. Two csv files were generated, one for enrolment data and one for verification data. These two csv files have records for similar data items (enrolment data have one extra vendor quality score item) such as SubjectId, VisitId, ProductId, FingerPosition, ModifiedOn, GroupRowId, BirthYear, Age, Origin, EthnicOrigin, NFIQ1, NFIQ2, NFIQ2Range, AgeRange, AgeGroup, NGender, Gender and 73 NFIQ2 quality feature scores.

NFIQ2 uses random forest for a binary classification: Class 0 represents images of very low utility and Class 1 represents images of very high utility. The trained random forest outputs class membership along with its probability. NFIQ2 score is the probability that a given image belongs to class 1 multiplied by 100 and rounded to its closest integer [8].

In random forest decision trees, each node is a condition on a single feature which split the dataset into two subsets so that similar classes of data go to the same subset. The measure of how mixed the classes are in the two subsets is called impurity. NFIQ2 uses Gini impurity to measure the split. Gini impurity for a set of items with J classes is calculated with Equation 4.1 where $i \in \{1, 2, ..., J\}$ and p_i is the proportion of items labeled with class I in the set.

$$I_G(p) = 1 - \sum_{i=1}^{J} p_i^2$$

Equation 4.1 Gini impurity

In NFIQ2, the mean of the impurity decreases (mean decrease of Gini impurity) in the tree from each feature was computed and used to rank the feature. A higher mean decrease of Gini impurity indicates that a particular predictor variable plays a greater role in partitioning the data into the defined classes ([8]), which means higher power in predicting NFIQ2 final scores. Table 4.1 is the feature rank list for NFIQ2.

Rank	Feature Name	MeanDecreaseGini
1	Frequency Domain Analysis Standard Deviation	140.76
2	FingerJet FX OSE COM Minutiae Count	92.089

3	FingerJet FX OSE OCL MinutiaeQuality	83.027
4	Ridge Valley Uniformity Mean	69.517
5	Frequency Domain Analysis Mean	62.229
6	FingerJet FX OSE Total Minutiae Count	57.565
7	Ridge Valley Uniformity Standard Deviation	50.946
8	Local Clarity Score Bin 7	50.688
9	Local Clarity Score Bin 8	50.1
10	Frequency Domain Analysis Bin 9	47.844
11	ROI Orientation Map Coherence Sum	38.104
12	Orientation Flow Bin 2	37.172
13	Local Clarity Score Mean	36.483
14	Ridge Valley Uniformity Bin 5	35.617
15	Ridge Valley Uniformity Bin 3	35.139
16	ROI Area Mean	34.932
17	Orientation Flow Bin 1	33.751
18	Orientation Flow Bin 0	33.513
19	MU	32.914
20	MMB	32.625
21	FingerJet FX OSE Mu MinutiaeQuality	32.316
22	Frequency Domain Analysis Bin 8	31.428
23	Frequency Domain Analysis Bin 7	31.236
24	Orientation Flow Mean	31.172
25	Ridge Valley Uniformity Bin 4	30.801
26	Orientation Certainty Level Mean	30.035
27	Orientation Flow Bin 3	29.721
28	Local Clarity Score Standard Deviation	28.777
29	ROI Relative Orientation Map Coherence Sum	28.7
30	Orientation Certainty Level Standard Deviation	28.429
31	Orientation Flow Standard Deviation	27.556
32	Orientation Certainty Level Bin 8	26.425
33	Frequency Domain Analysis Bin 6	25.161
34	Local Clarity Score Bin 6	23.837
35	Orientation Flow Bin 5	23.431
36	Local Clarity Score Bin 9	23.283
37	Orientation Flow Bin 4	22.883
38	Ridge Valley Uniformity Bin 2	22.843
39	Ridge Valley Uniformity Bin 1	22.38

40	Orientation Flow Bin 9	21.413
41	Ridge Valley Uniformity Bin 6	21.306
42	Orientation Flow Bin 7	21.296
43	Orientation Flow Bin 6	20.939
44	Local Clarity Score Bin 1	20.91
45	Orientation Flow Bin 8	20.867
46	Ridge Valley Uniformity Bin 7	20.798
47	Ridge Valley Uniformity Bin 0	20.47
48	Orientation Certainty Level Bin 0	20.234
49	Orientation Certainty Level Bin 7	19.887
50	Orientation Certainty Level Bin 1	19.434
51	Orientation Certainty Level Bin 6	19.188
52	Local Clarity Score Bin 5	19.073
53	Orientation Certainty Level Bin 4	18.865
54	Local Clarity Score Bin 3	18.861
55	Orientation Certainty Level Bin 2	18.663
56	Frequency Domain Analysis Bin 0	18.657
57	Frequency Domain Analysis Bin 5	18.438
58	Orientation Certainty Level Bin 5	18.262
59	Orientation Certainty Level Bin 9	18.164
60	Local Clarity Score Bin 2	18.136
61	Frequency Domain Analysis Bin 3	17.847
62	Orientation Certainty Level Bin 3	17.834
63	Frequency Domain Analysis Bin 1	17.71
64	Local Clarity Score Bin 4	17.598
65	Frequency Domain Analysis Bin 4	17.283
66	Frequency Domain Analysis Bin 2	17.145
67	Ridge Valley Uniformity Bin 8	0
68	Ridge Valley Uniformity Bin 9	0
69	LCS Local Clarity Score Bin 0	0

Table 4.1 NFIQ2 quality feature rank list (reproduced from reference [8])

Most of the analysis are based on the whole data set (verification or enrolment data set) without differentiating the data by visits. When subset data is needed, the data is retrieved

from the whole data by using Python pandas DataFrame query. For NFIQ2 quality

features, this study only takes the first 6 features in the ranking list (Table 4.1) which

represent frequency domain analysis standard deviation of ridges and valleys (Frequency

Domain Analysis Standard Deviation), Centre of Mass (ROI) block minutiae count (FingerJet

FX OSE COM Minutiae Count), Orientation Certainty Level (OCL) of minutiae (FingerJet FX

OSE OCL MinutiaeQuality), ridge valley ratio uniformity (Ridge Valley Uniformity Mean),

frequency domain analysis mean of ridges and valleys (Frequency Domain Analysis Mean),

and total minutiae count (FingerJet FX OSE Total Minutiae Count).

Table 4.2 shows some sample fingerprint images (with explicit consent from the subject)

for NFIQ2 quality features. Two images are selected for each quality feature and the first

one of the two images has better quality score for that feature than the second one. The

sensor and feature score data for each selected image are listed also.

Feature Name	Sensor	Feature Score	Image
Frequency Domain Analysis Standard Deviation	2	0.066854	

	7	0.23615	
FingerJet FX OSE COM Minutiae Count	5	54	
	3	11	
FingerJet FX OSE OCL MinutiaeQuality	9	0.90244	

	8	0	
Ridge Valley Uniformity Mean	7	1.327	
	9	0.95908	
Frequency Domain Analysis Mean	7	0.48934	

	8	0.35943	
FingerJet FX OSE Total Minutiae Count	2	124	
	3	11	

Table 4.2 Sample images for NFIQ2 quality features

Jupyter Notebook is used to run the Python code. Jupyter Notebook is an open-source web application that can run Python code and display the result lively. The test data is loaded through the web interface. Python code can be input and run interactively and different analysis plots can be generated on the fly (Figure 4.1).

Figure 4.1 Jupyter Notebook interface

Figure 4.2 shows the process for generating plots. Run the helper method definition code segment to load helper method definitions. Then call readData method to read previously generated verification and enrolment csv data file. To draw plots for subset of the data, query data subset using pandas query method. Finally call specific helper to draw needed plots from the whole data or data subset. The following helper methods were defined for this study:

- readData: use pandas read_csv method to load csv file to pandas data frame

- drawBoxPlots: use matplotlib pyplot to draw box plots for specified data items in the data. Most of the plots were generated by this method

- drawBoxPlotsOrdered: revised version of drawBoxPlots. It has an extra parameter for specifying the order of the box plot bars. This is used to handle wrong x parameter order situations, ex. Age groups not show in ascending order.

- drawBoxPlotsOrderedLR: revised version of drawBoxPlotsOrdered to put legend at the lower bottom of the plots if the empty space is at the low bottom.

- drawSimpleHistogram: draw simple histogram bins like Figure 3.2 and Figure 3.3.

- drawMultiBinHistogram: draw multiple bins for one unit of x value, ex. Figure 3.4 and Figure 3.5
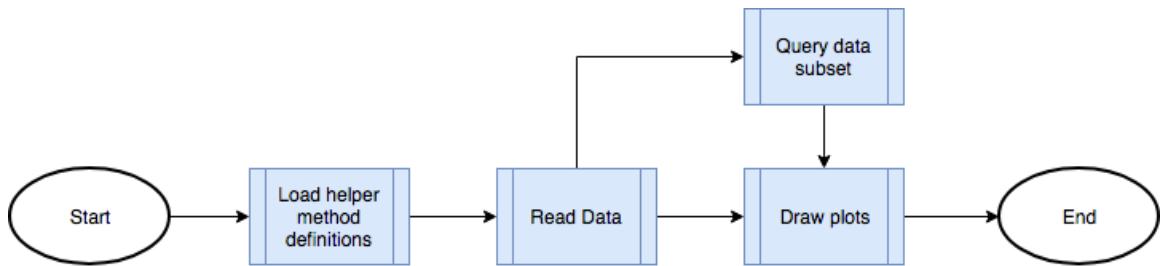

Figure 4.2 Process to draw plots

Plots or images with two or more features that are to be compared are outputted through Jupyter Notebook. By visually looking at the plots or images, we can find various interesting results on whether sensor design affects fingerprint quality and quality features, whether NFIQ2 scores align well with NFIQ1 scores, how different levels of NFIQ1 scores relates to NFIQ2 quality feature scores, how age, gender and ethnic origin impact NFIQ2 quality scores and their quality feature scores.

5 Chapter: Results and Discussions

This chapter presents the analysis of the collected data in terms of sensor, algorithm and demographic factors. This study has some interesting findings on following items:

- NFIQ1/NFIQ2 quality score distributions on different sensors

- NFIQ2 quality scores and NFIQ2 quality feature scores distributions on different NFIQ1 score levels after taking off data for sensor that have abnormal NFIQ1 quality scores

- NFIQ2 quality scores and NFIQ2 quality feature scores distributions on different sensors

- Vendor quality scores vs NFIQ1/NFIQ2 quality scores

- NFIQ2 quality scores and NFIQ2 quality feature scores distributions on different age groups

- NFIQ2 quality scores and NFIQ2 quality feature scores distributions on genders

- NFIQ2 quality scores and NFIQ2 quality feature scores distributions on ethnic origins

5.1 NFIQ1/NFIQ2 quality score distributions on enrolment and verification data

Whether there is a statistically significant difference between NFIQ1/NFIQ2 quality score distribution on enrolment and verification data is investigated. Statistical parameters for enrolment data and verification data are output by Python code through

Jupyter interface (Figure 5.1). A Student's t-test is used for this investigation and 0.01 is selected as the significance level threshold.

The H0 Hypothesis is that there is no statistically significant difference between the NFIQ1/NFIQ2 quality score distribution on enrolment data and that on verification data.

- For NFIQ1 quality scores, we have total 75103 ($N_{e\_nfiq1} = 75103$) values with mean 2.126333 ($\mu_{e\_nfiq1} = 2.126333$) and standard deviation 1.272971 ($\sigma_{e\_nfiq1} = 1.272971$) on enrolment data, and total 153048 ($N_{v\_nfiq1} = 153048$) values with mean 2.045796 ($\mu_{v\_nfiq1} = 2.045796$) and standard deviation 1.248536 ($\sigma_{v\_nfiq1} = 1.248536$) on verification data. t-score on NFIQ1 test is:

$$t_{nfiq1} = \frac{\mu_{e\_nfiq1} - \mu_{v\_nfiq1}}{S_{d\_nfiq1}} = \frac{\mu_{e\_nfiq1} - \mu_{v\_nfiq1}}{\sqrt{\frac{\sigma_{e\_nfiq1}^2}{N_{e\_nfiq1}} + \frac{\sigma_{e\_nfiq1}^2}{N_{e\_nfiq1}}}} = 14.289744$$

The NFIQ1 p-values for one-tailed and two tailed tests are both smaller than the selected significance threshold 0.01. With this result, we reject the null hypothesis and accept the alternative hypothesis: there is statistically significant difference between NFIQ1 quality score distribution on enrolment data and that on verification data. Verification data have better quality in terms of NFIQ1 quality scores (lower mean value $\mu_{v\_nfiq1} < \mu_{e\_nfiq1}$ and standard deviation $\sigma_{v\_nfiq1} < \sigma_{e\_nfiq1}$).

- For NFIQ2 quality scores, we have total 75103 ($N_{e\_nfiq2} = 75103$) values with mean 51.779995 ($\mu_{e\_nfiq2} = 51.779995$) and standard deviation 21.796090 ($\sigma_{e\_nfiq2} = 21.796090$) on enrolment data, and total 153048 ($N_{v\_nfiq2} = 153048$) values with mean 52.661015 ($\mu_{v\_nfiq2} = 52.661015$) and standard deviation 21.080092 ($\sigma_{v\_nfiq2} = 21.080092$) on verification data. t-score on NFIQ2 test is:

$$t_{nfiq2} = \frac{\mu_{e\_nfiq2} - \mu_{v\_nfiq2}}{S_{d\_nfiq2}} = \frac{\mu_{e\_nfiq2} - \mu_{v\_nfiq2}}{\sqrt{\frac{\sigma_{e\_nfiq2}^2}{N_{e\_nfiq2}} + \frac{\sigma_{e\_nfiq2}^2}{N_{e\_nfiq2}}}} = -9.170804$$

The NFIQ2 p-values for one-tailed and two tailed tests are both smaller than the selected significance threshold 0.01. With this result, we reject the null hypothesis and accept the alternative hypothesis: there is statistically significant difference between NFIQ2 quality score distribution on enrolment data and that on verification data. Verification data have better quality in terms of NFIQ2 quality scores (higher mean value as $\mu_{v\_nfiq2} > \mu_{e\_nfiq2}$ and lower standard deviation $\sigma_{v\_nfiq2} < \sigma_{e\_nfiq2}$).

Overall the fingerprint quality on verification data is better than that of enrolment data in terms of NFIQ1 and NFIQ2 quality scores. For enrolment, when the fingerprint images quality passes required threshold, they can be successfully enrolled. For verification, the captured fingerprint images were used to match with enrolled fingerprint template and so

the subject might have tried with more efforts to get better quality of fingerprints. For following analysis, if not specifically mentioned, the verification data is selected.

```
verify_data[['NFIQ1', 'NFIQ2']].describe()
```

	NFIQ1	NFIQ2
count	153048.000000	153048.000000
mean	2.045796	52.661015
std	1.248536	21.080092
min	1.000000	0.000000
25%	1.000000	40.000000
50%	2.000000	56.000000
75%	3.000000	68.000000
max	5.000000	99.000000

```
enrol_data[['NFIQ1', 'NFIQ2']].describe()
```

	NFIQ1	NFIQ2
count	75103.000000	75103.000000
mean	2.126333	51.779995
std	1.272971	21.796090
min	1.000000	0.000000
25%	1.000000	38.000000
50%	2.000000	56.000000
75%	3.000000	68.000000
max	5.000000	98.000000

Figure 5.1 NFIQ1/2 scores in test data

It's quite interesting that the total number of fingerprints on verification data is not the double of the total of enrolment data (Figure 5.2). There is a (153048-75103*2) = 2842 difference. This is due to the rule of 3 enrolment and 6 verification captures weren't

strictly followed. Figure 5.2 shows the captured fingerprints for each subject on verification data and enrolment data. We can see that the double relationship between the fingerprints on verification and enrolment is not strictly kept.
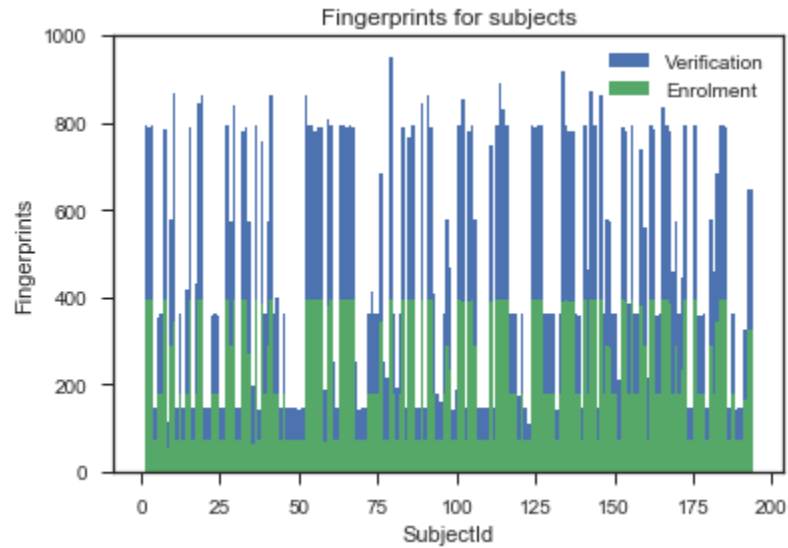

Figure 5.2 Enrolment fingerprint number vs verification fingerprint number

5.2 NFIQ1/NFIQ2 quality score distributions on different sensors

NFIQ1 and NFIQ2 quality score distributions on different sensors have the same behavior for enrolment data and verification data (Figure 5.3, Figure 5.4, Figure 5.5, Figure 5.6). NFIQ1 quality scores on sensor 2 and 3 are similar and quite strange for males with majority of the NFIQ1 quality sores in level 2 except some outliers. NFIQ1 quality scores on sensor 9 for female are also strange in different way with majority of the NFIQ1 quality scores in level 5 with the exception of some outliers.
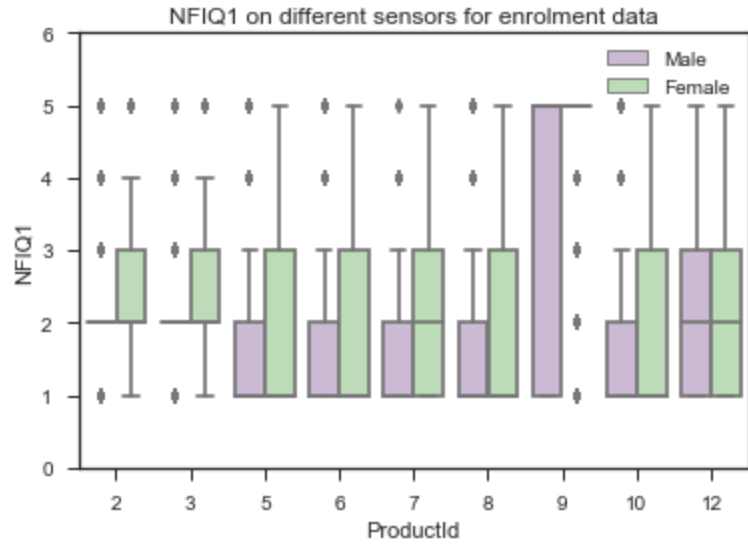
Figure 5.3 NFIQ1 of enrolment data on different sensors
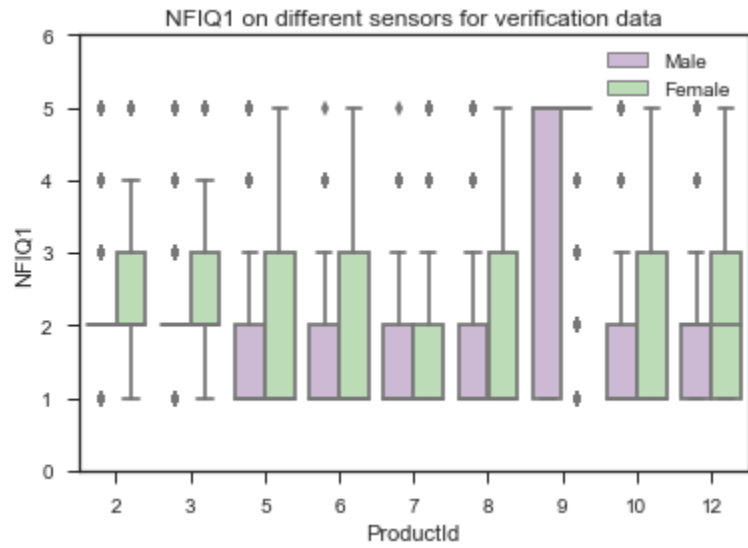
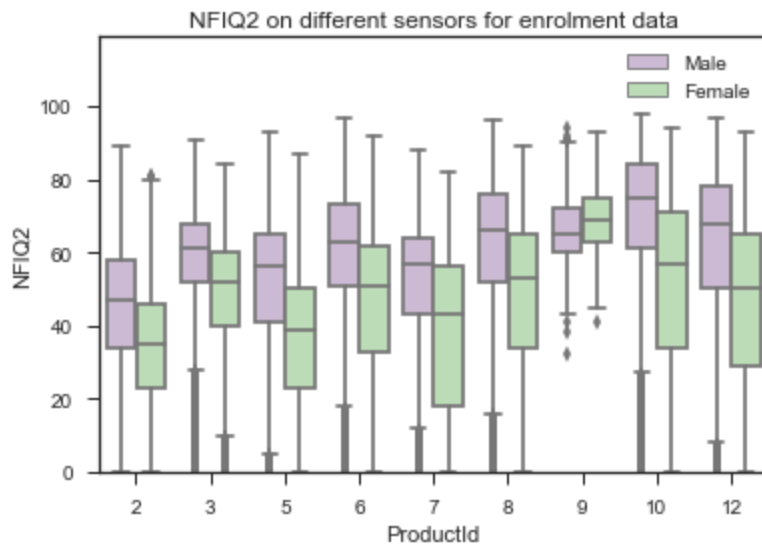**Figure 5.4 NFIQ1 of verification data on different sensors**

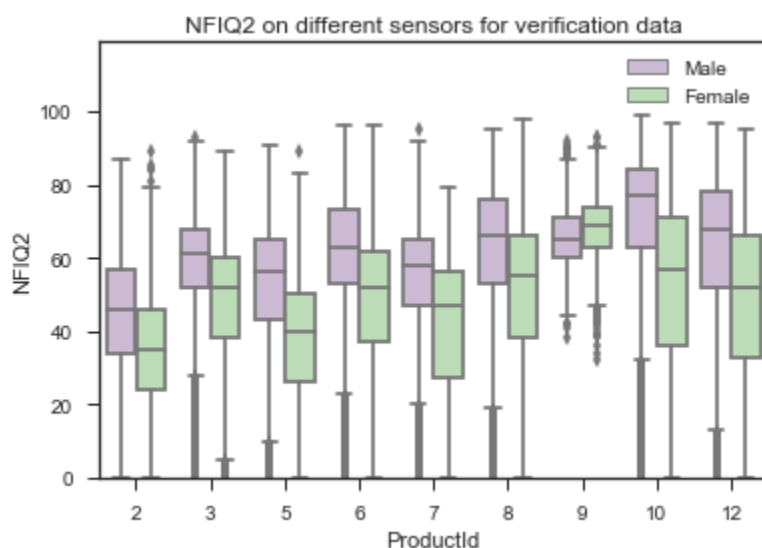**Figure 5.5 NFIQ2 of enrolment data on different sensors**

Figure 5.6 NFIQ2 of verification data on different sensors

For males on sensor 2 and 3, the 25th percentile, 75th percentile and median are all 2 (Table 5.1). Even for female subjects, the NFIQ1 scores are mostly between 2 and 3 with more concentrated on level 2 (Table 5.2). Sensor 2 and 3 are with totally different design - one uses capacitive semi-conductor and the other uses standard optical. Their image

sizes are different also. Currently the real reason why these two devices have most of the

NFIQ1 scores as 2 for male and female subjects is still unknown and needs more

investigation to find out.

Sensor	count	mean	std	min	25%	50%	75%	max
2	8967	2.077	0.799	1	2	2	2	5
3	9000	2.121	0.835	1	2	2	2	5
5	8969	1.487	0.834	1	1	1	2	5
6	8774	1.403	0.731	1	1	1	2	5
7	6496	1.404	0.677	1	1	1	2	5
8	6622	1.578	0.928	1	1	1	2	5
9	6629	3.666	1.790	1	1	5	5	5
10	8926	1.535	0.887	1	1	1	2	5
12	9081	1.878	0.976	1	1	2	2	5

Table 5.1 Male verification NFIQ1 data feature values on sensors

Sensor	count	mean	std	min	25%	50%	75%	max
2	9783	2.459	1.056	1	2	2	3	5
3	9906	2.469	1.064	1	2	2	3	5
5	9767	1.759	1.118	1	1	1	3	5
6	9180	1.676	0.941	1	1	1	3	5
7	6801	1.681	0.884	1	1	1	2	5
8	7238	1.831	1.157	1	1	1	3	5
9	7302	4.163	1.515	1	5	5	5	5
10	9749	1.823	1.130	1	1	1	3	5
12	9858	2.147	1.137	1	1	2	3	5

Table 5.2 Female verification NFIQ1 data feature values on sensors

NFIQ2 quality scores on sensor 2 and 3 are quite reasonable. Sensor 2 has lower NFIQ2

scores and the reason could mainly be from the small image size which caused less detail

in feature patterns.

For sensor 9, most of the NFIQ1 scores are 5 for female subject fingerprints (25th, 75th percentile and median are all 5) (Table 5.2). Many of the fingerprint NFIQ1 scores are 5 also for male subjects (75th percentile and median are 5) (Table 5.1). It means that most of the fingerprints from sensor 9 have very bad quality based on NFIQ1 measurement. This could be due to the sensor technology (Multispectral) or image size (higher dpi). The trained neuro network parameters for NFIQ1 classification algorithm didn't work well with Multispectral sensor images. The big image size or the difference between the fingerprint pattern on the skin and the second represent under the skin surface could also contribute to feature extracting errors with MINDCT algorithm.

Table 5.3 show some sample fingerprint images (with explicit consent from the subject) from sensor 9 (multispectral sensor) with NFIQ1 level 1 (good quality) and level 5 (bad quality). NFIQ1 and NFIQ2 quality scores are listed in the table also. NFIQ2 scores aren't aligned well with NFIQ1 scores for the listed sample images.

Sample Image				
NFIQ1 Score	1	5	1	5
NFIQ2 Score	62	73	82	64

Table 5.3 Sample images from sensor 9 (multispectral sensor)

For sensor 9, NFIQ2 performed quite well. The NFIQ2 quality scores are mostly close to

70 with small interquartile range (IQR) for both male and female subject fingerprints

(Table 5.4, Table 5.5). Its male fingerprint NFIQ2 scores have 65 as median which is less

than those from sensor 10, 12 and 8 but have much less variations, A remarkable

observation with sensor 9 NFIQ2 scores is that female subject fingerprints have better

NFIQ2 scores than those of male subjects while all other sensors behaved the other way.

More than 75 percent of the female NFIQ2 fingerprint scores are above the medians of

the female NFIQ2 scores from other sensors. This could be due to that NFIQ2 training

database has no images from multispectral sensor or image resampling. It could also be

that female tends to have thinner skin surface and so the under-surface pattern dominates

the result image. The fingerprint database used for NFIQ2 model training does not have

fingerprint from Multispectral sensors.

Sensor	count	mean	std	min	25%	50%	75%	max
2	8967	45.256	16.565	0	34	46	57	87
3	9000	57.677	15.678	0	52	61	68	93
5	8969	52.232	17.810	0	43	56	65	91
6	8774	60.991	17.477	0	53	63	73	96
7	6496	53.987	16.666	0	47	58	65	95
8	6622	62.122	18.868	0	53	66	76	95
9	6629	66.176	8.236	38	60	65	71	92
10	8926	69.953	20.652	0	63	77	84	99
12	9081	62.805	20.838	0	52	68	78	97

Table 5.4 Male verification NFIQ2 data feature values on sensors

Sensor	count	mean	std	min	25%	50%	75%	max
2	9783	34.246	17.045	0	24	35	46	89
3	9906	46.772	18.600	0	38	52	60	89
5	9767	36.953	18.145	0	26	40	50	89
6	9180	48.558	19.023	0	37	52	62	96
7	6801	41.270	18.984	0	27	47	56	79
8	7238	50.630	20.613	0	38	55	66	98
9	7302	68.814	8.061	32	63	69	74	93
10	9749	51.933	24.363	0	36	57	71	97
12	9858	48.156	22.217	0	33	52	66	95

Table 5.5 Female verification NFIQ2 data feature values on sensors

Overall, NFIQ2 has less abnormal score distribution than NFIQ1 does on different

fingerprint sensors. NFIQ1 behaved abnormally on sensor 2, 3 and 9. More investigation

is needed in order to find the real reason. NFIQ1 scores were generated years ago. There

might be some code fixes after our NFIQ1 calculation was done. It's worth to regenerate

NFIQ1 scores with most recent NFIQ1 tools. The following sections will remove the data

from sensor 2, 3 and 9 when NFIQ1 related study is performed to avoid bias from

abnormal data. NFIQ2 behaved normally across all the test sensors and so all the data

will be used in all other studies where NFIQ1 is not referred.

5.3   NFIQ2 and NFIQ2 quality feature distributions on NFIQ1 after taking off

sensor 2, 3 and 9

Overall, NFIQ2 quality scores align well with NFIQ1 levels (Figure 5.7). NFIQ1 level 2

for males stay a little higher when mapped to NFIQ2. As both NFIQ1 and NFIQ2

algorithms are from NIST, they share most of the same testing data set. It is reasonable

that NFIQ2 result aligns with NFIQ1 result with normal data.
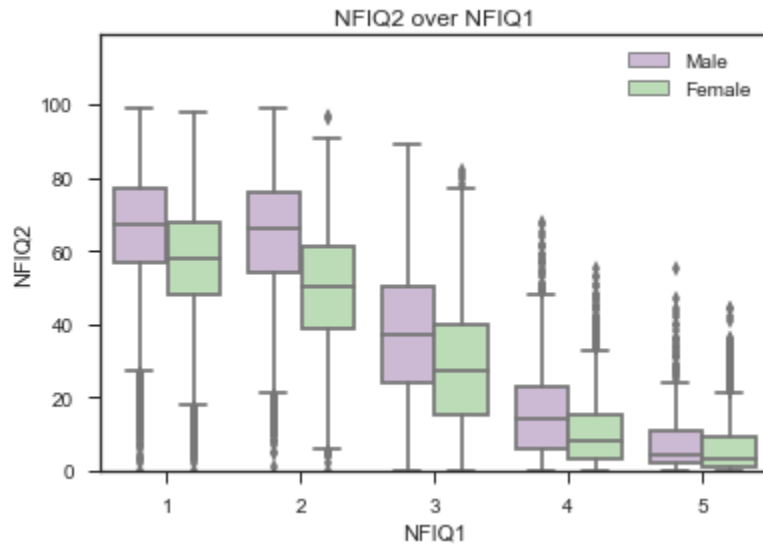
Figure 5.7 NFIQ2 over NFIQ1

NFIQ2 frequency domain analysis (FDA) standard deviation feature scores align well

with NFIQ1 levels also (Figure 5.8). The less the standard deviation score, the better it
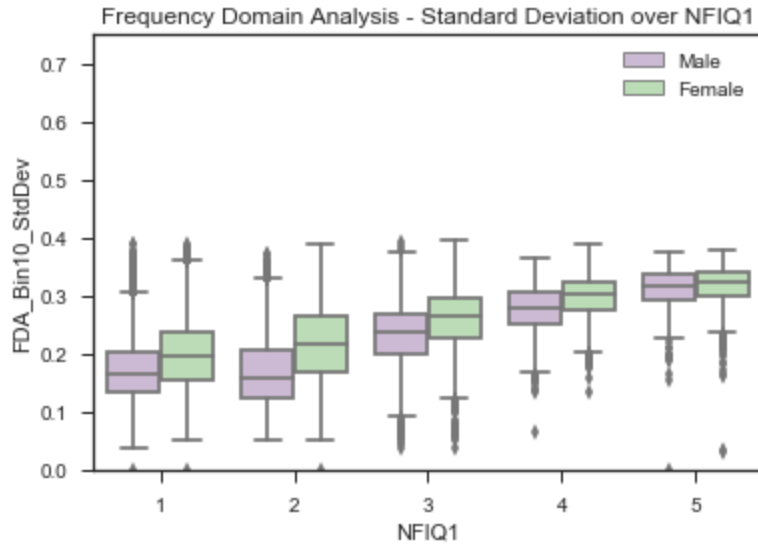
aligns with NFIQ1 score - the lower, the better.

Figure 5.8 FDA standard deviation over NFIQ1

Minutiae count in the Centre of Mass (COM) block aligns with NFIQ1 levels OK (Figure 5.9). Only level 2 behaves a little differently. This contributes to the high overall NFIQ2 quality scores on NFIQ level 1.
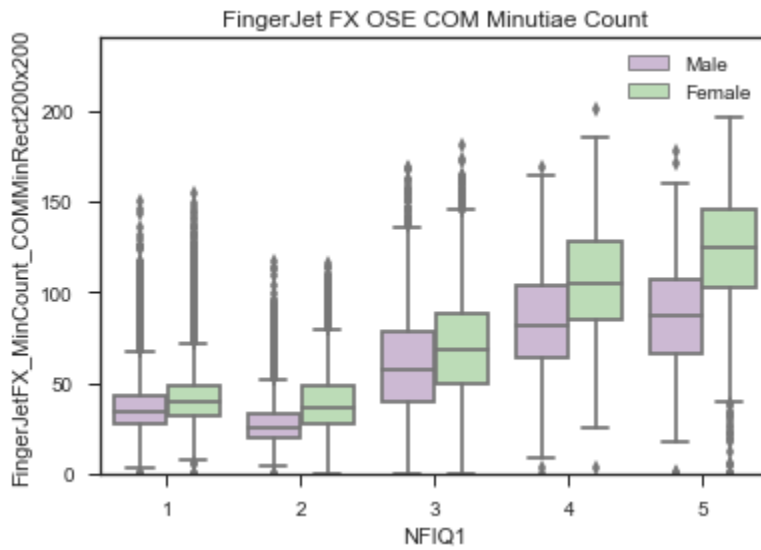
Figure 5.9 Centre of Mass (COM) block minutiae count over NFIQ1

FJFX based minutiae orientation certainty (OCL) quality has apparent changes between

high level (1, 2) and low level (3, 4, 5) NFIQ scores (Figure 5.10). The difference

between NFIQ1 level 1 and level 2 is not much. Overall this feature quality aligns with

NFIQ1 levels.
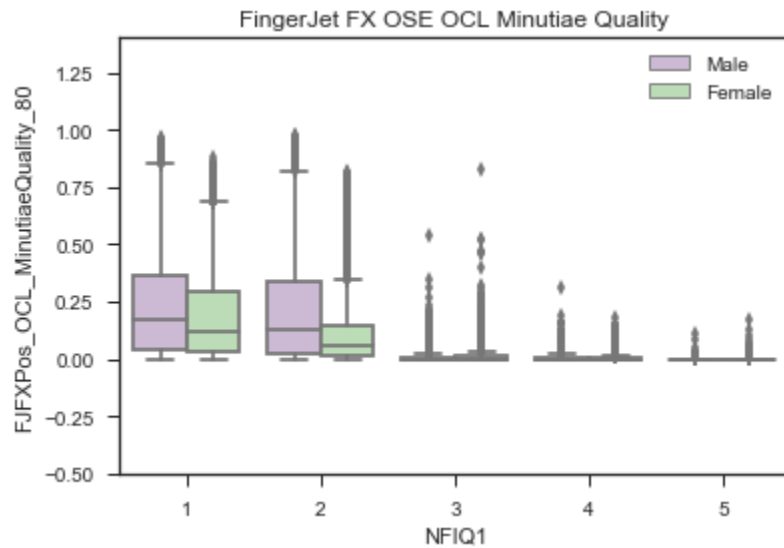
Figure 5.10 OCL minutiae quality over NFIQ1

Ridge valley uniformity aligns well with NFIQ1 levels (Figure 5.11). Again, the

difference between NFIQ1 level 1 and level 2 is not much.
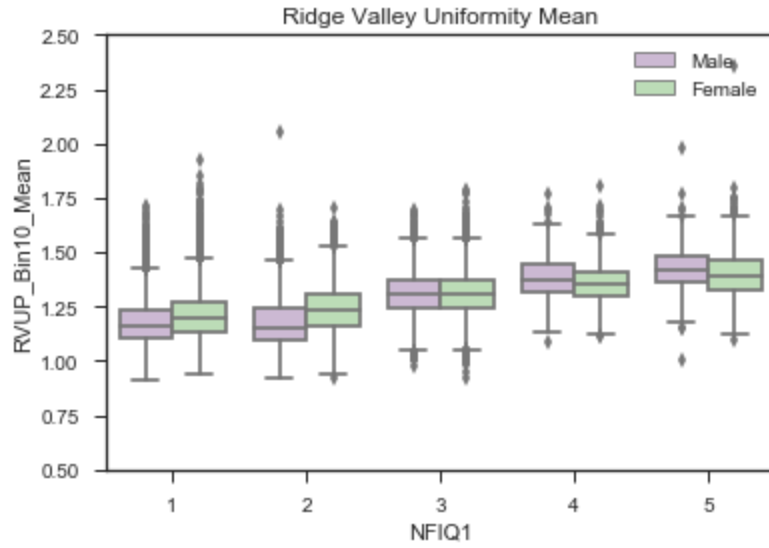
Figure 5.11 Ridge valley uniformity over NFIQ1

Frequency domain analysis (FDA) mean aligns well with NFIQ1 levels (Figure 5.12).

With NFIQ1 quality decrease, FDA mean interquartile range (IQR) grows.
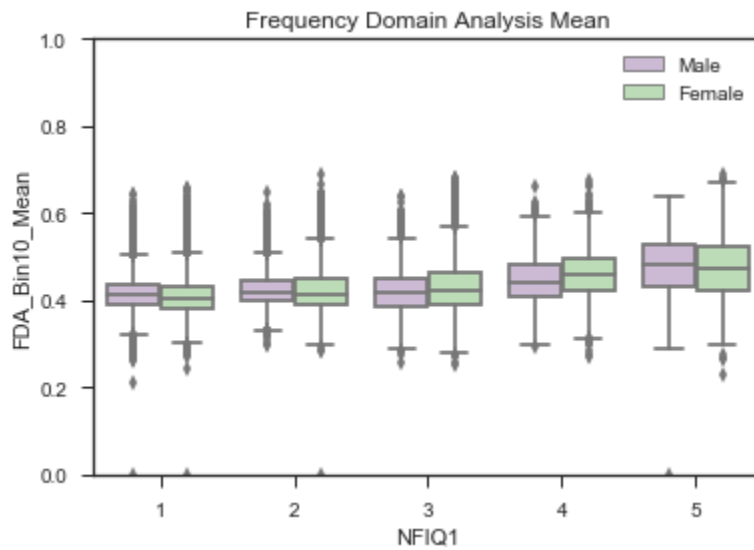
Figure 5.12 Frequency domain analysis mean over NFIQ1

Total minutiae count aligns with NFIQ1 levels. Better NFIQ1 quality scores (lower levels) relate to less minutiae count (Figure 5.13). NFIQ1 level 2 data have less minutiae than NFIQ level 1 data though and this contribute to the high overall NFIQ2 quality scores on NFIQ1 level 2 data.
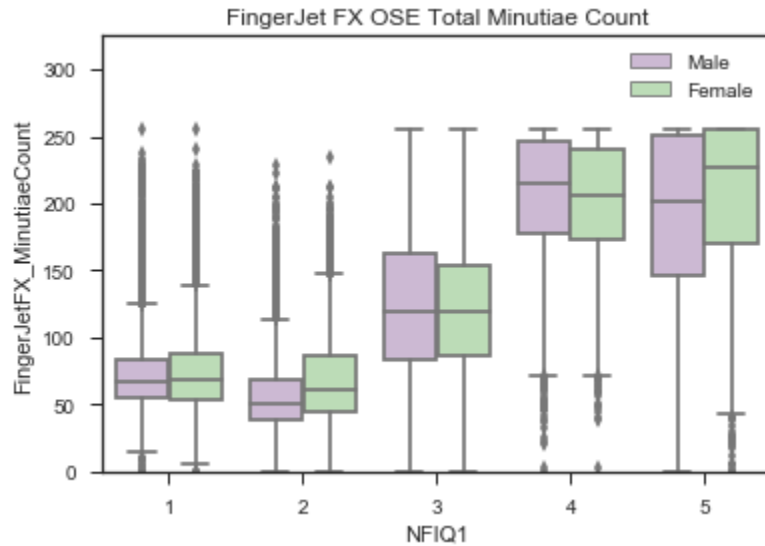
Figure 5.13 Total minutiae count over NFIQ1

Overall, NFIQ2 quality scores and quality feature scores align well with NFIQ1 quality scores. The difference of NFIQ2 quality scores between NFIQ1 level 1 data and level 2 data are not apparent. NFIQ1 classification on level 1 and 2 are not so closely matching NFIQ2 quality scores. The alignment on some quality features are not so apparent. More research on NFIQ1/NFIQ2 algorithms is needed to understand those little alignment issues.

5.4   NFIQ2 quality feature distributions on different sensors

Sensor 9 has best ridge valley FDA standard deviation (Figure 5.14). A possible reason of this behaviour is that sensor 9 returned images interpolated from data captured at

different frequencies so the ridge valley patterns are less variant. Sensor 8, 10 and 12 are

quite good at this feature. Sensor 7 is the worst. More investigation is needed on why it

happened.
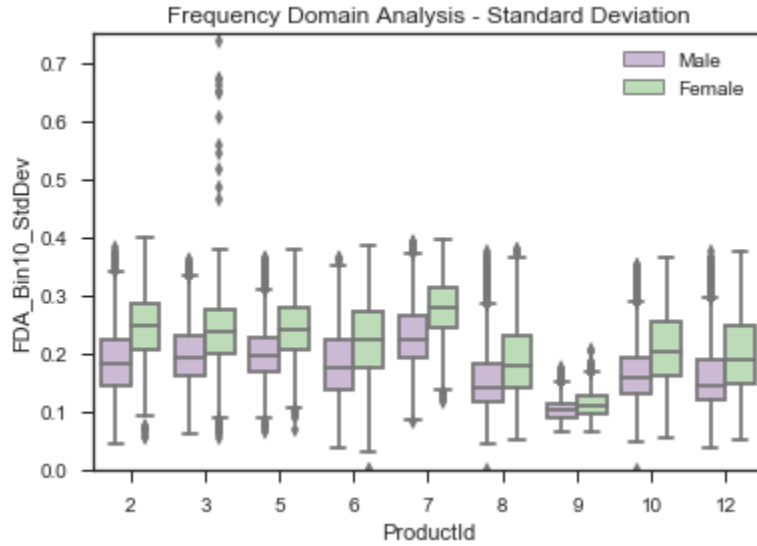
Figure 5.14 FDA standard deviation over sensors

Sensor 9 has least COM minutiae count (Figure 5.15). This might be because its

fingerprint images have less false minutiae due to the enhanced fingerprint image. The

other sensors have similar COM minutiae counts for male subjects and the variations are

not much. Fingerprints for female subjects varies more on this quality feature.
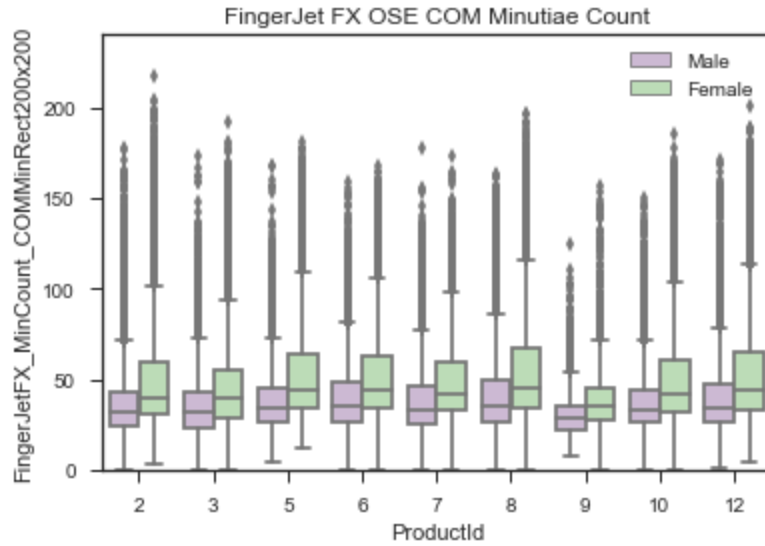
Figure 5.15 COM minutiae count over sensors

Fingerprint data from sensor 9 have much higher minutiae orientation certainty quality than those from other sensors (Figure 5.16). Sensor 2 and 8 behave the worst and more investigation is needed to understand why it happened. Based on the observed behaviour, multispectral sensor is far better than others and capacitive sensor and some optical sensors are not good on this quality feature.
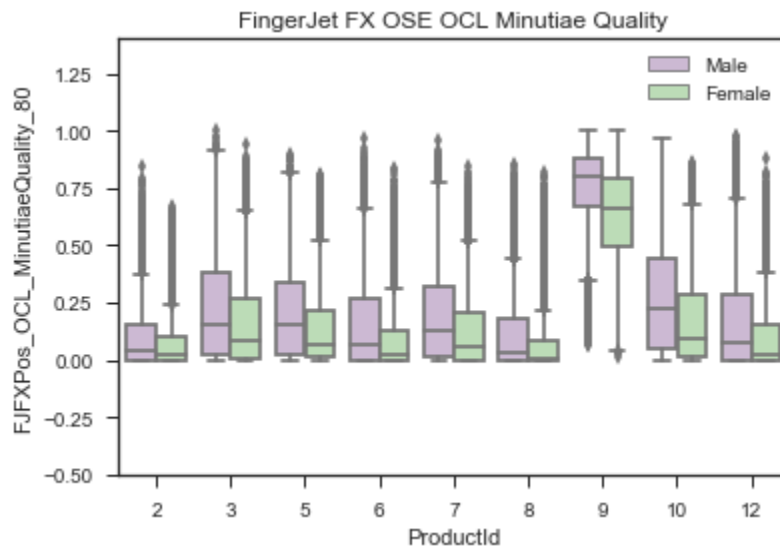
Figure 5.16 OCL minutiae quality over sensors

Sensor 9 outperforms other sensors on the ridge valley uniformity (Figure 5.17). The other sensors behave quite similarly.
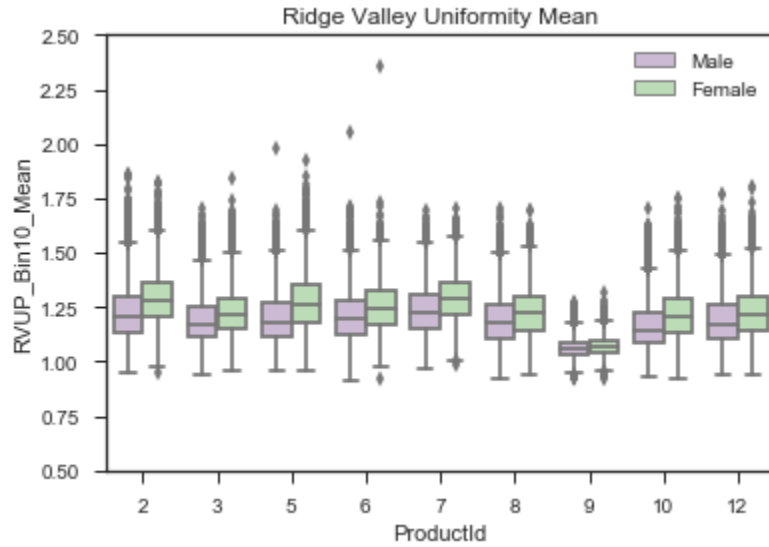
Figure 5.17 Ridge valley uniformity over sensors

Sensor 7 has the highest FDA mean. Sensor 9 has least interquartile range (IQR) for FDA mean (Figure 5.18).
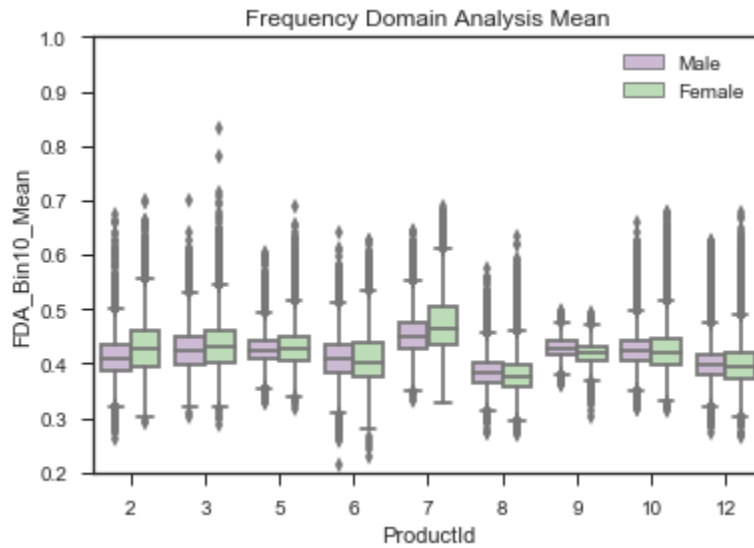
Figure 5.18 Frequency domain analysis mean over sensors

Sensor 9 and 2 have less total minutiae numbers (Figure 5.19). For sensor 9, this might be because that the enhanced fingerprint image led to less false minutiae. For sensor 2, no light source and contamination might lead to less false minutiae.
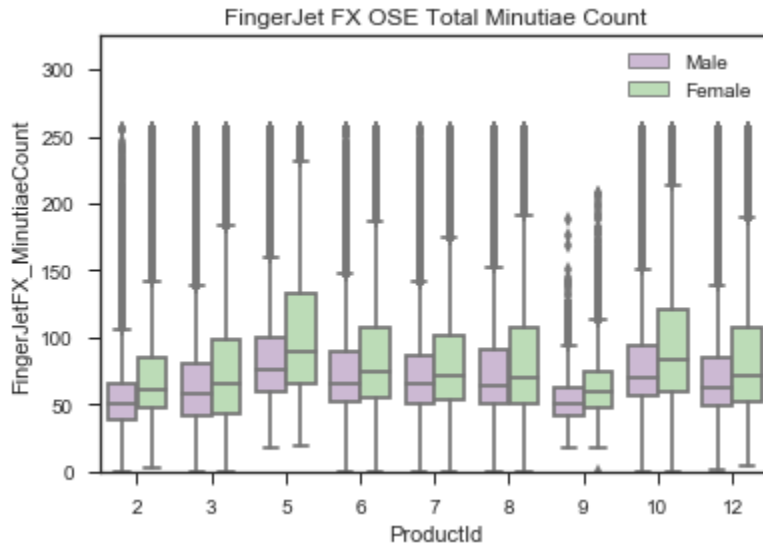
Figure 5.19 Total minutiae count over sensors

5.5 Vendor quality vs NFIQ1/2

Vendor quality scores can only be retrieved from enrolment data, so this test was only on enrolment data that have vendor quality (labeled as EnrolFingerQuality in the test data records).

Sensor 9 uses fixed 40 and sensor 10 uses fixed 100 as the vendor quality and so their vendor qualities are not usable for study (Figure 5.20). Sensor 2 has the vendor quality as 100 with a little strand out and this quality score is not trustable also. Sensor 3 has below 50 vendor qualities but above 50 NFIQ2 quality scores. The maximum possible vendor quality score of sensor 3 is 50.
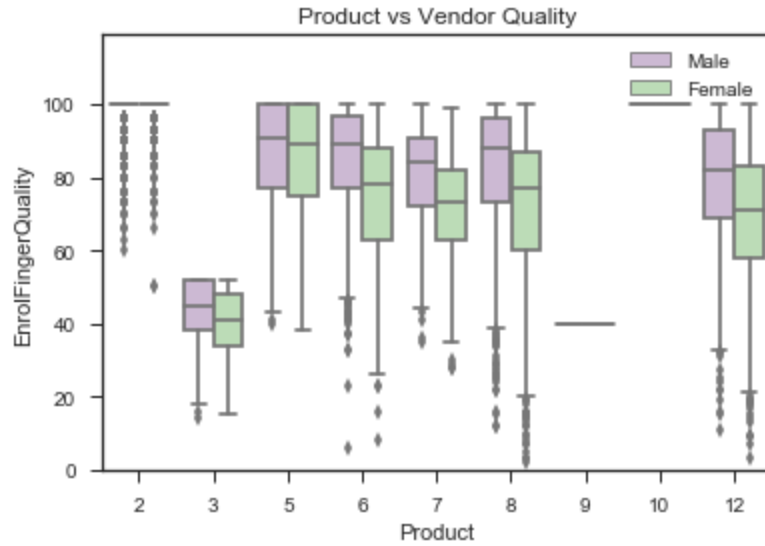
Figure 5.20 Vendor quality over sensors

Overall, vendor quality scores for sensor 3, 6, 7, 8 and 12 align with NFIQ1 quality scores (median value goes down with NFIQ1 score level up) and male vendor quality scores are higher than female vendor quality scores (Figure 5.21, Figure 5.23, Figure 5.24, Figure 5.25, Figure 5.26). For sensor 6, 8 and 12, the vendor quality scores don't change significantly between NFIQ1 level 1 and 2; however, the scores dramatically increase from NFIQ1 level 2 to 3. For sensor 6 and 8, most of the vendor quality scores for male subjects on NFIQ1 level 2 are even higher than the median of male vender scores on NFIQ1 level 1. More investigation is needed for these behaviours. For sensor 5, the vendor quality scores are not well correlated with NFIQ levels (Figure 5.22).
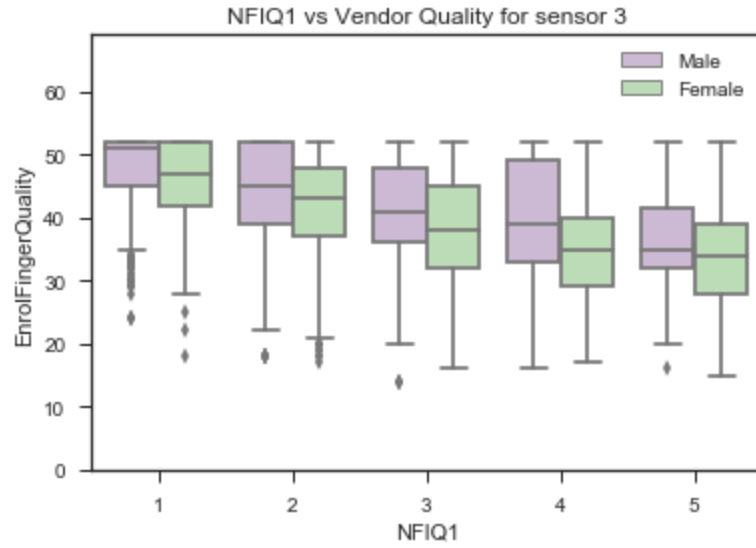
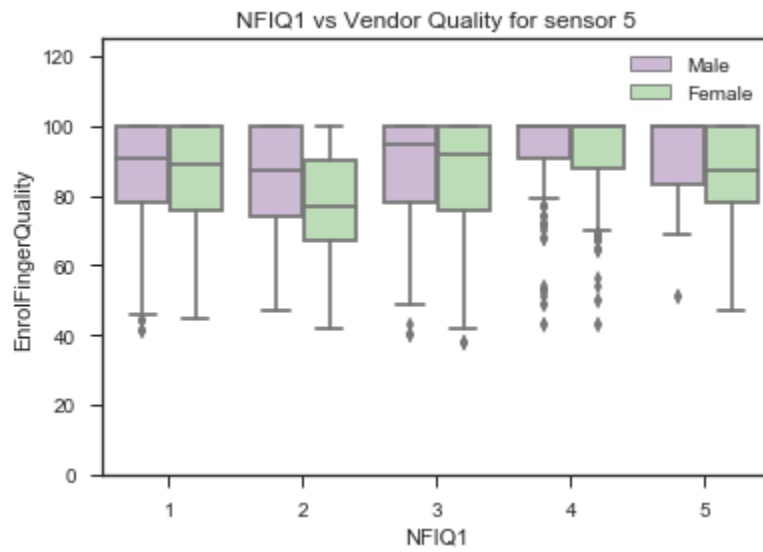**Figure 5.21 Vendor quality over NFIQ1 on sensor 3**
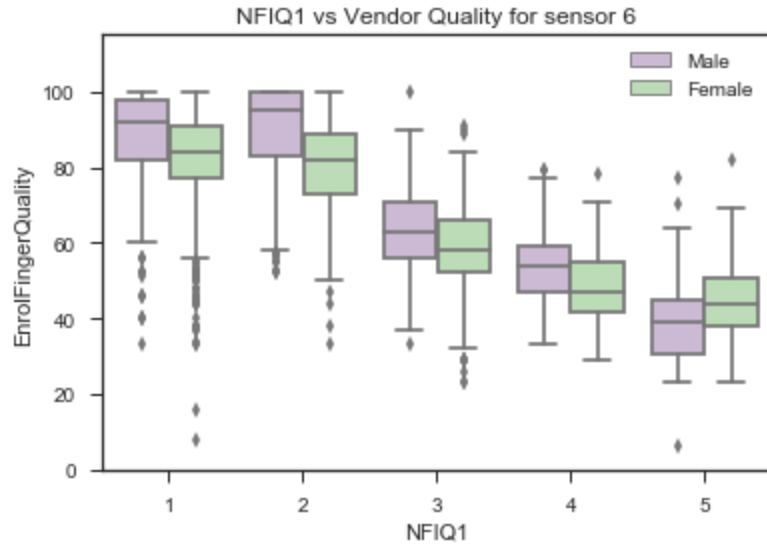

Figure 5.22 Vendor quality over NFIQ1 on sensor 5

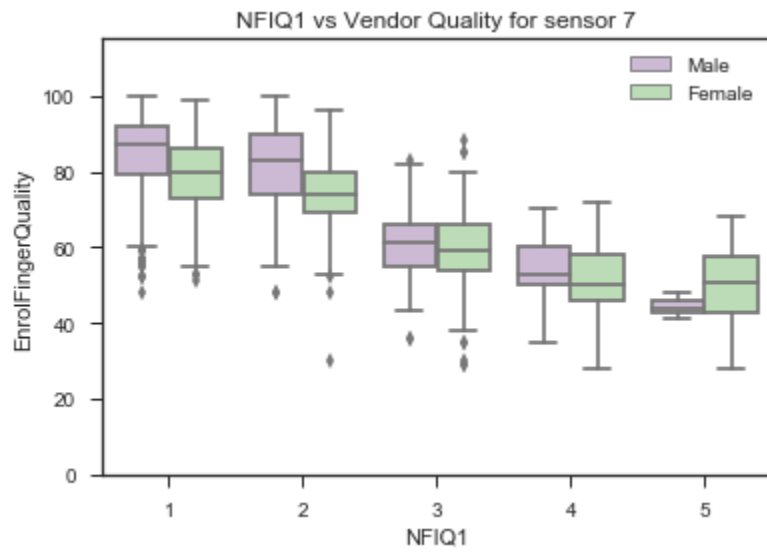**Figure 5.23 Vendor quality over NFIQ1 on sensor 6**

Figure 5.24 Vendor quality over NFIQ1 on sensor 7
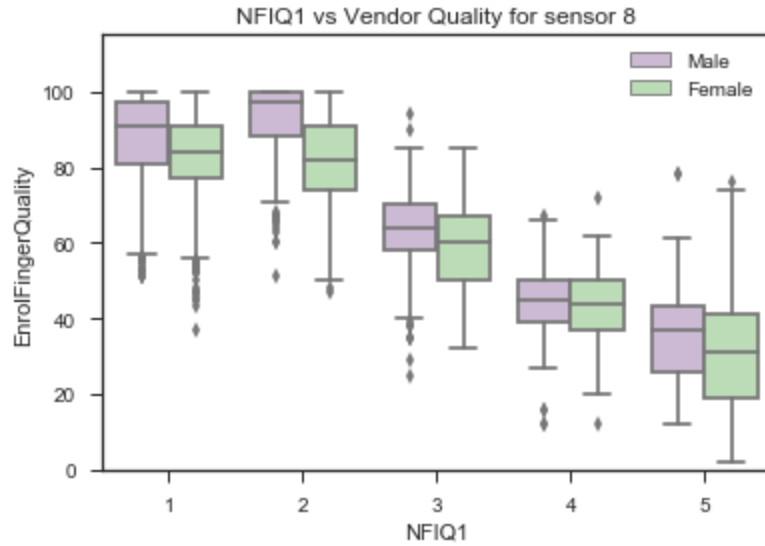
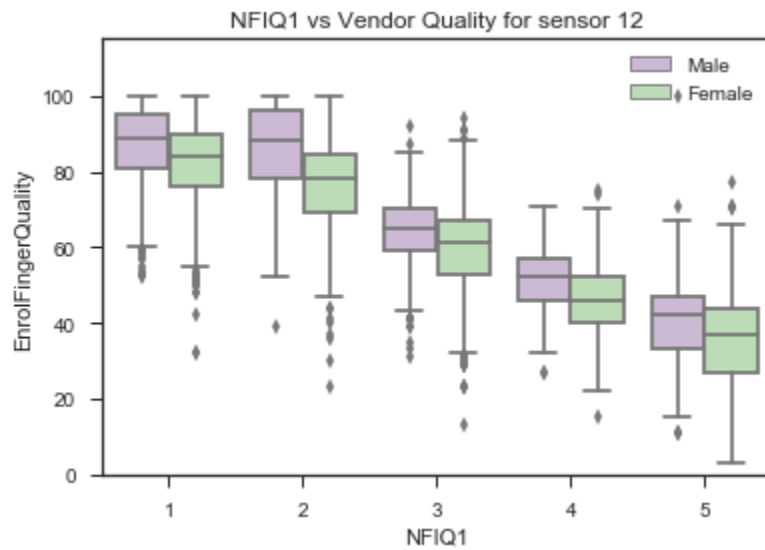**Figure 5.25 Vendor quality over NFIQ1 on sensor 8**

Figure 5.26 Vendor quality over NFIQ1 on sensor 12

Except sensor 5, the other sensors have good alignment with vendor quality scores and NFIQ2 quality scores (Figure 5.27, Figure 5.28, Figure 5.29, Figure 5.30, Figure 5.31, Figure 5.32). Overall the fingerprints of male subjects have better vendor quality scores

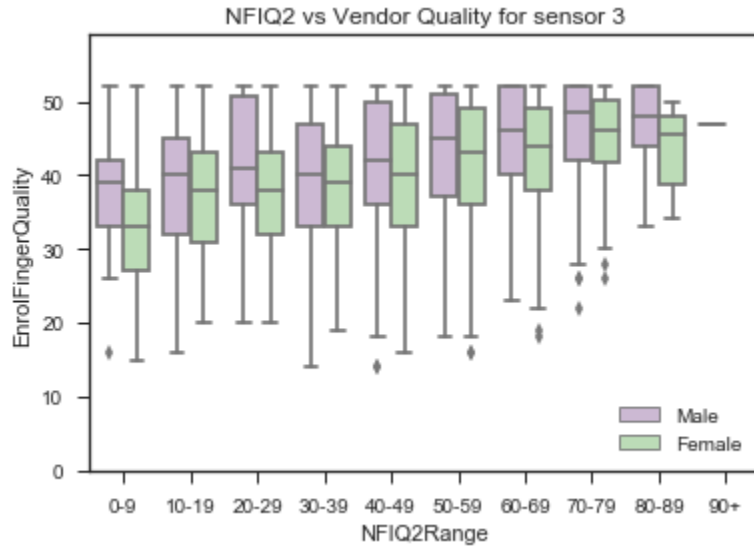across NFIQ2 score ranges. The vender quality scores don't correlate well with NFIQ2 scores.

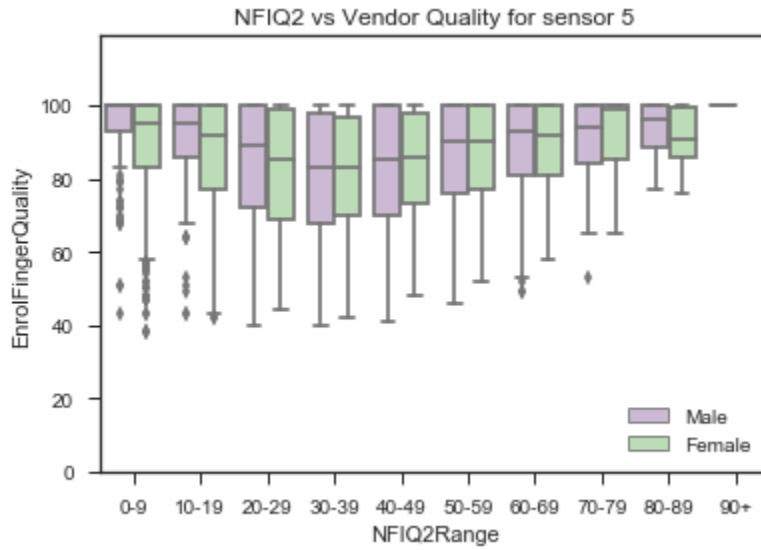**Figure 5.27 Vendor quality over NFIQ2 on sensor 3**

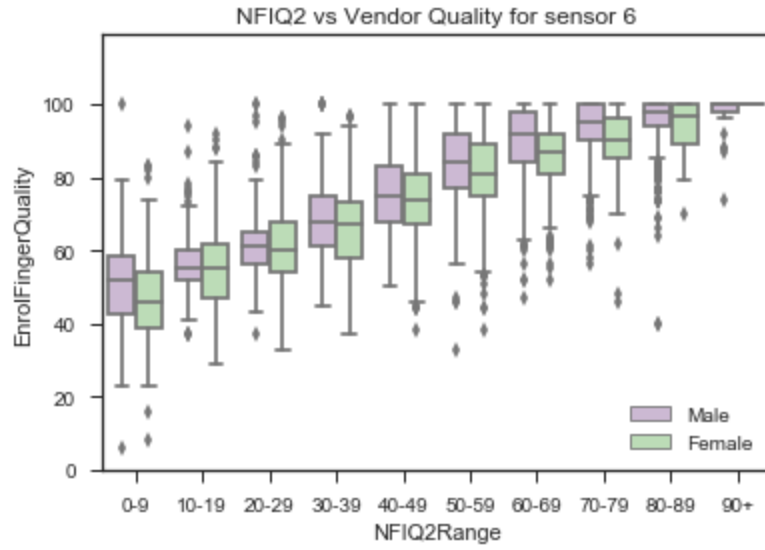**Figure 5.28 Vendor quality over NFIQ2 on sensor 5**

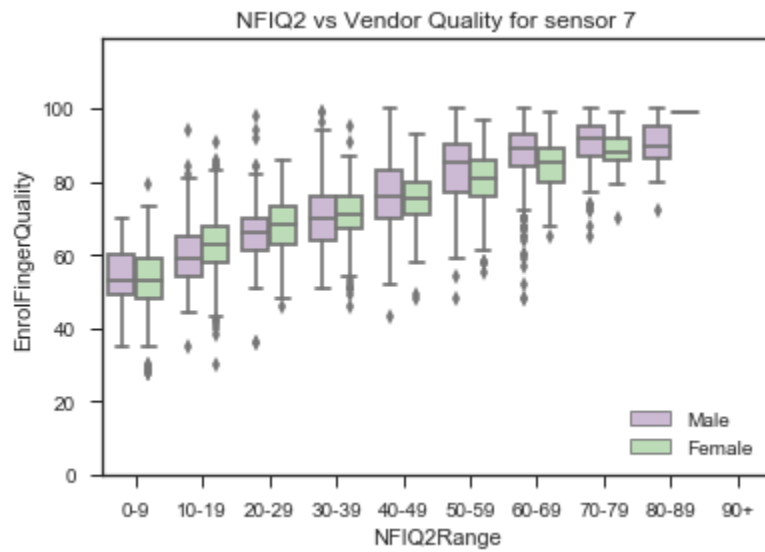**Figure 5.29 Vendor quality over NFIQ2 on sensor 6**

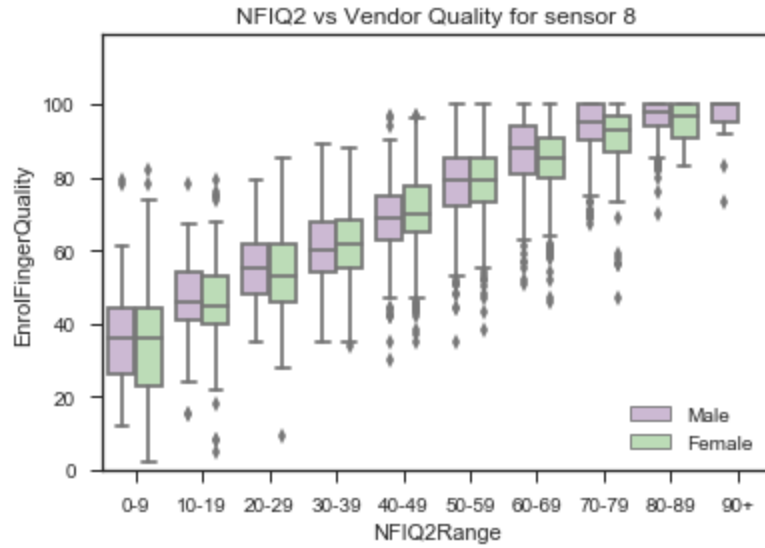**Figure 5.30 Vendor quality over NFIQ2 on sensor 7**

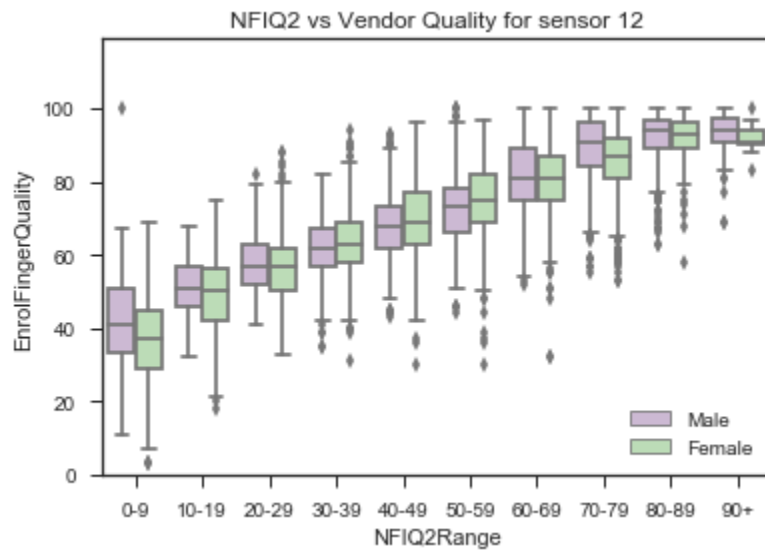**Figure 5.31 Vendor quality over NFIQ2 on sensor 8**

Figure 5.32 Vendor quality over NFIQ2 on sensor 12

With NFIQ1 and NFIQ2 as reference, the vendor quality scores from sensor 5 are also

not reliable. Sensor 3, 6, 7, 8 and 12 have reasonable vendor quality calculation

algorithms.

5.6 NFIQ2 and NFIQ2 quality feature distributions on different age groups

The NFIQ2 to Age Group plot (Figure 5.33) verifies that fingerprint quality decreases with the age growth [54].
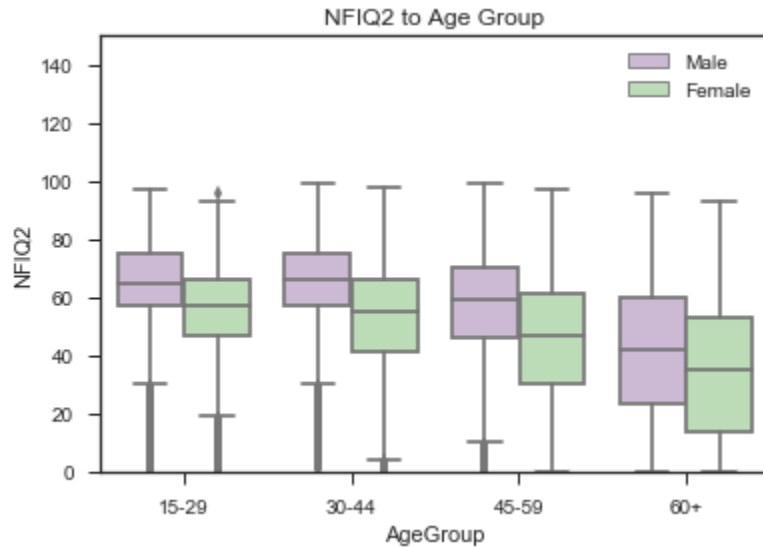
Figure 5.33 NFIQ2 score over age groups

Age effects on NFIQ2 quality features are apparent and quite uniform. NFIQ2 quality features get worse with increasing age (Figure 5.34, Figure 5.35, Figure 5.36, Figure 5.37, Figure 5.38, Figure 5.39).
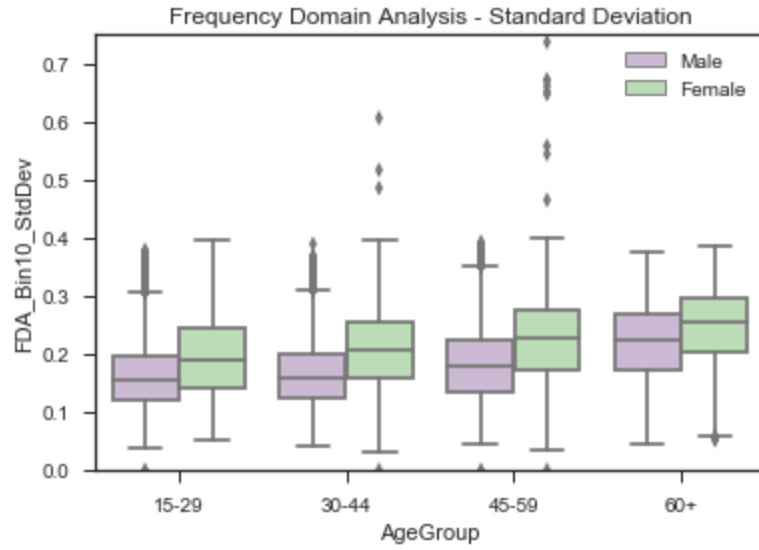
Figure 5.34 FDA standard deviation over age groups
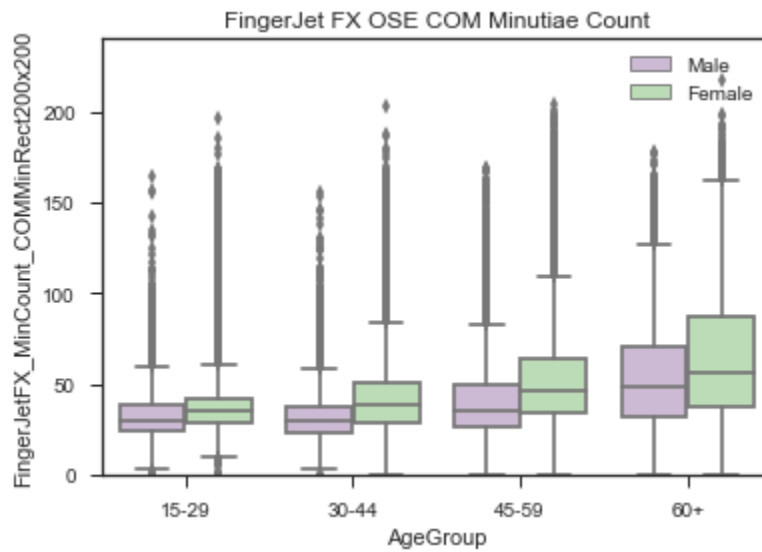
Figure 5.35 COM minutiae count over age groups
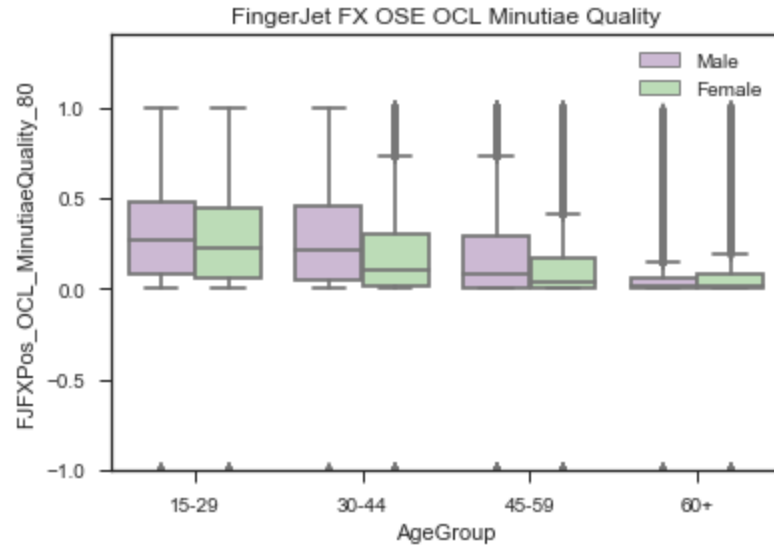
Figure 5.36 OCL minutiae quality over age groups
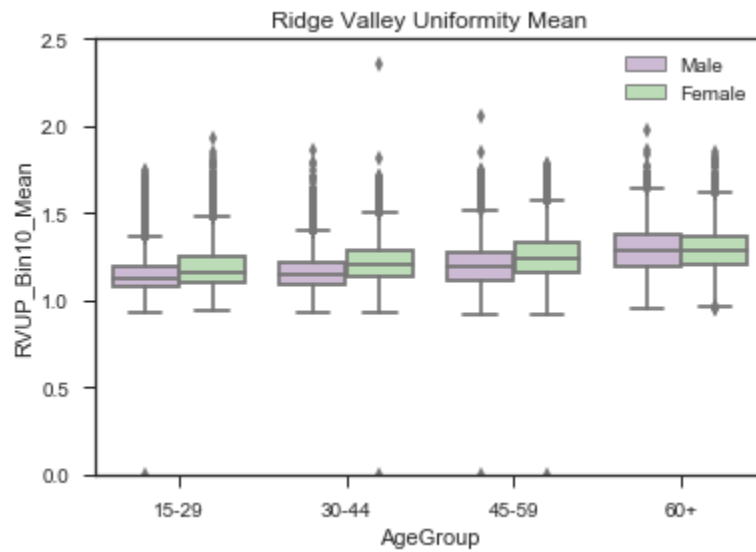
Figure 5.37 Ridge valley uniformity over age groups

Figure 5.38 Frequency domain analysis mean over age groups
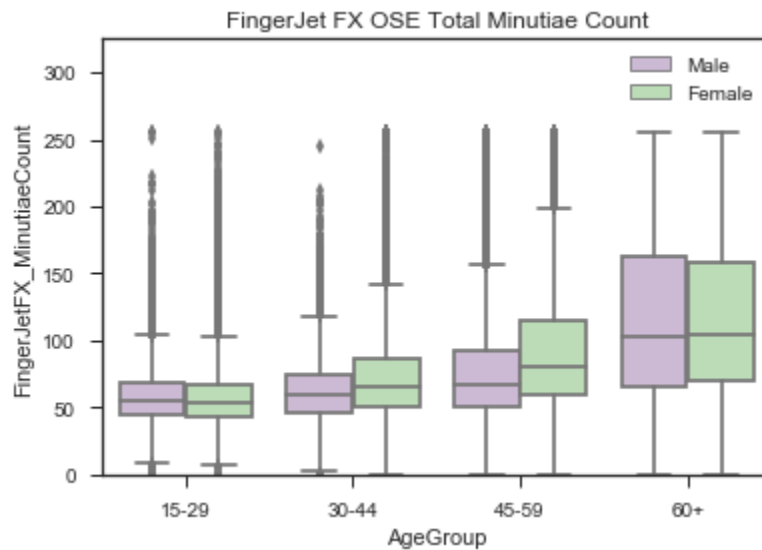
Figure 5.39 Total minutiae count over age groups

Overall, NFIQ2 quality score and NFIQ2 quality feature scores behave consistently without apparent abnormality - fingerprint quality gets worse as the age grows.

5.7 NFIQ2 and NFIQ2 quality feature distributions on genders

It is observed from the above boxplots with comparison of male and female that fingerprint qualities are different for male and female. Overall fingerprint qualities for males are better than those for females.
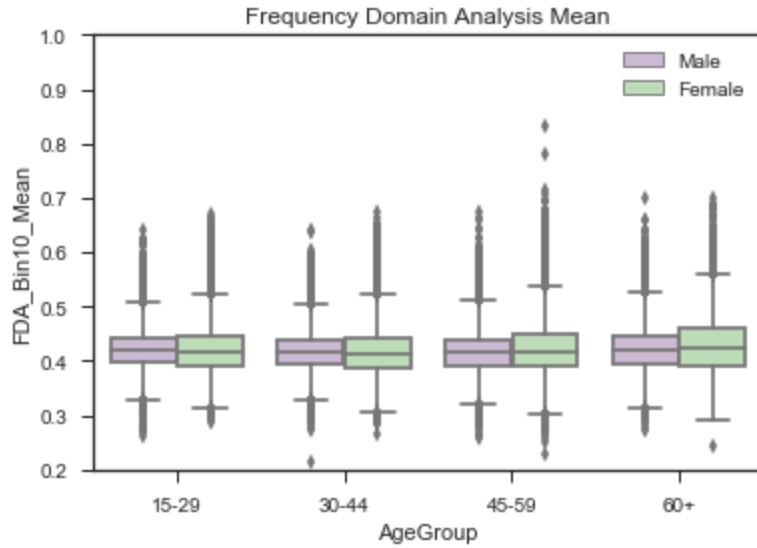
5.8 NFIQ2 and NFIQ2 quality feature distributions on age groups for different sensors

For all the sensors except sensor 9, NFIQ2 quality scores go down with the age growth (Figure 5.40). For sensor 9, the multispectral one, the result is totally reverse - NFIQ2 quality scores go up with the age growth. It could be that the surface skin becomes less opaque and so the under-skin fingerprint patterns becomes more dominant in the synthesized fingerprint image with the age growth. More research on biology is needed to justify this. Again, sensor 9 outperforms other sensors on NFIQ2 quality scores and the age change does not affect the quality scores as much as the other sensors. The overall quality scores vary more on female than male fingerprints by looking at the interquartile range (IQR).
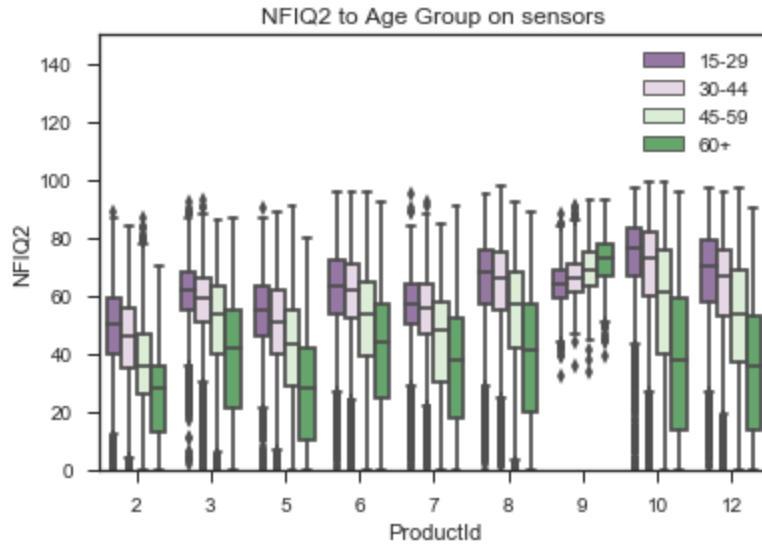
Figure 5.40 NFIQ2 for age groups over sensors

For NFIQ2 quality features, sensor 9 outperforms the other sensors on all the selected quality features except frequency domain analysis mean where sensor 7 has the highest scores, a lot on ridge valley frequency domain analysis standard deviation and minutiae quality (Figure 5.41, Figure 5.42, Figure 5.43, Figure 5.44, Figure 5.45, Figure 5.46). All other sensors have pattern that quality features get worse with the age growth and sensor 9 share the same pattern (not so much apparently) on frequency domain analysis, minutiae quality and ridge valley uniformity but not on Centre of Mass minutiae count and total minutiae count. It seems like that multispectral detect better on minutiae count. There might be more quality features that have different behavior with multispectral sensors.
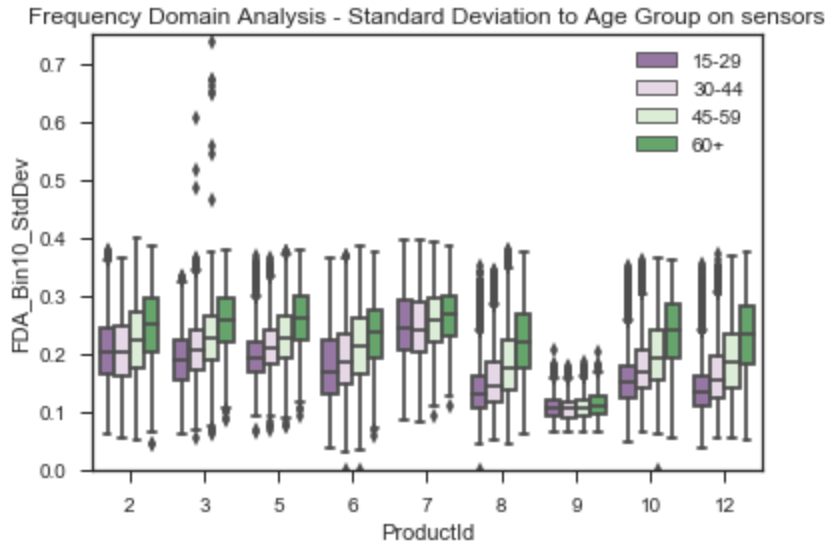
Figure 5.41 FDA standard deviation for age groups over sensors
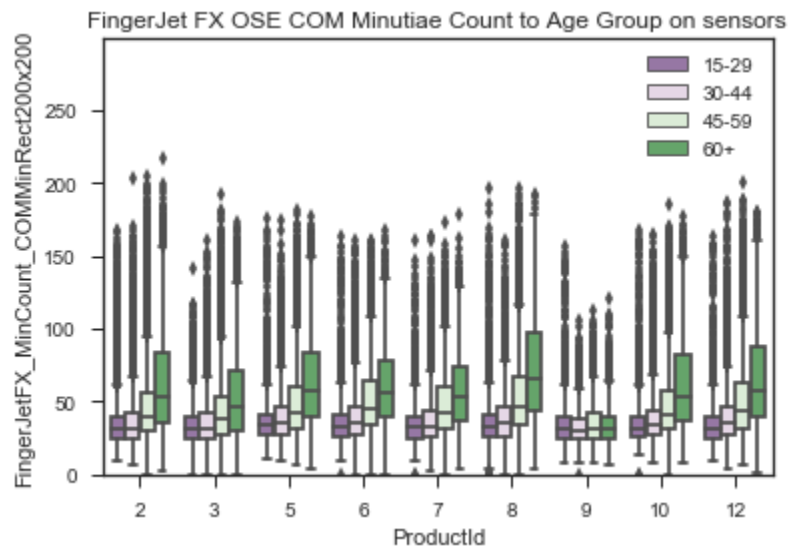
Figure 5.42 COM minutiae count for age groups over sensors
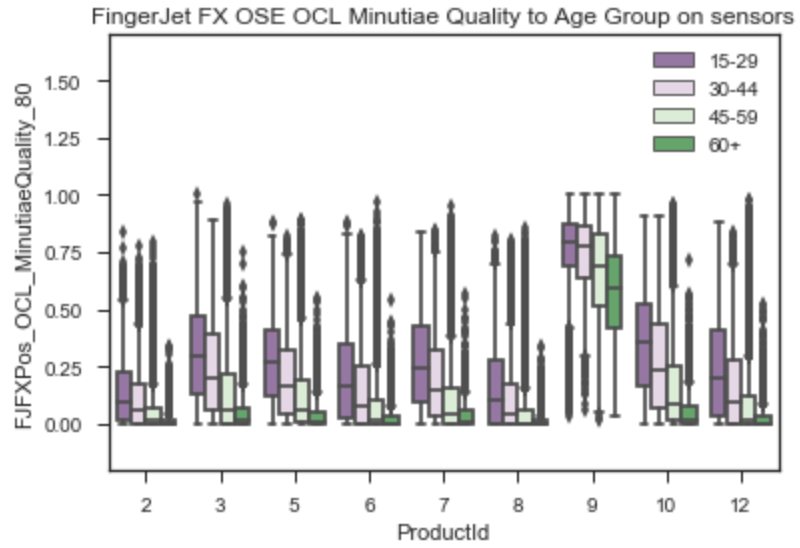
Figure 5.43 OCL minutiae quality for age groups over sensors
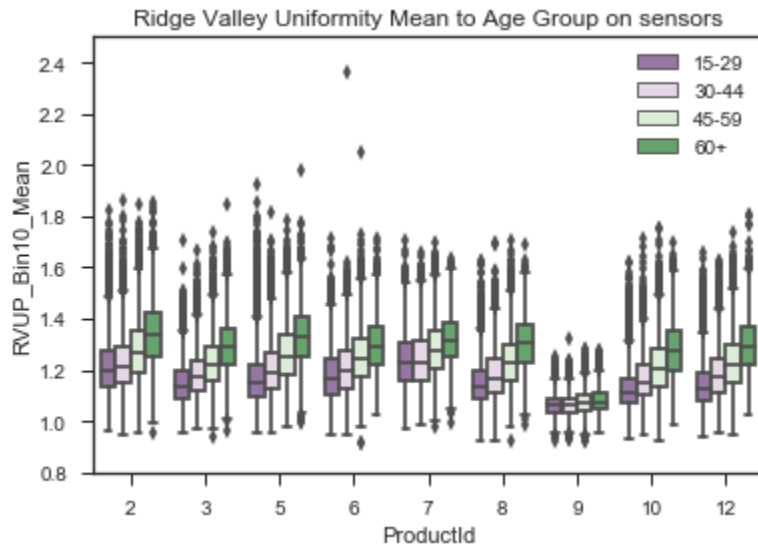
Figure 5.44 Ridge valley uniformity for age groups over sensors
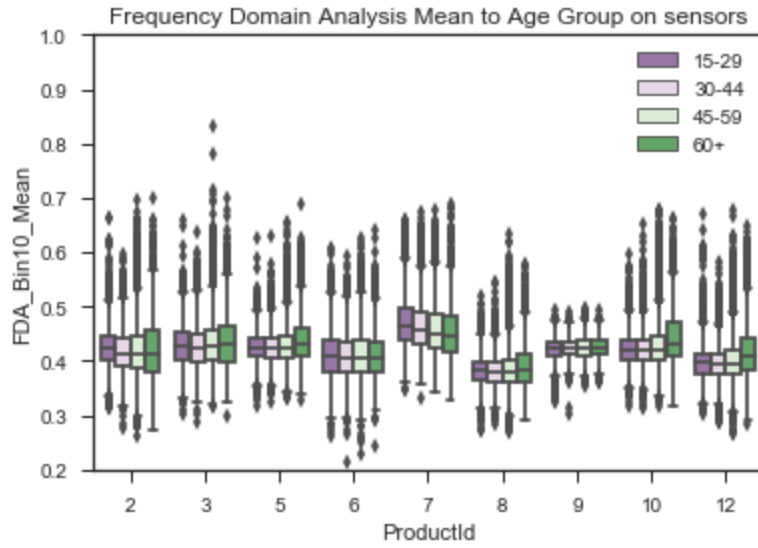
Figure 5.45 Frequency Domain Analysis Mean for Age Group over sensors
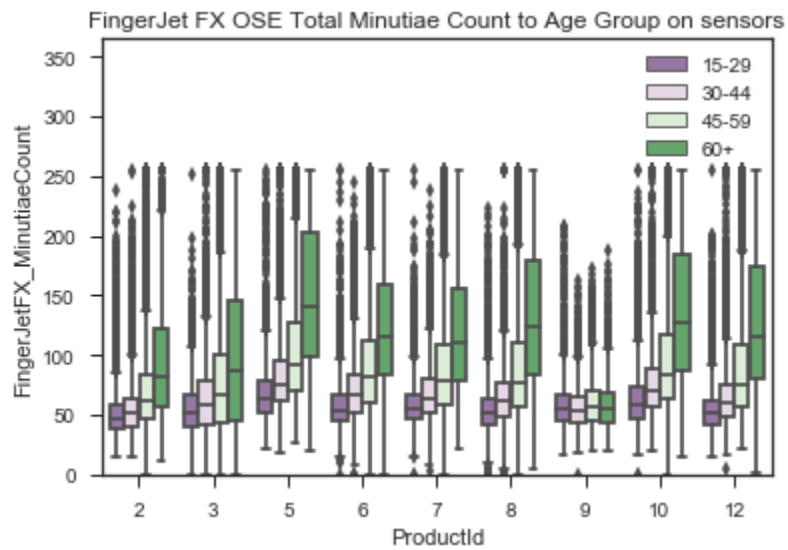
Figure 5.46 Total minutiae for age groups over sensors

5.9 NFIQ2 on different claimed ethnic origins

The subjects' self-identified ethnic origin information was recorded during the data

capturing process. The testing subjects are from six different ethnic origins (the

classification of the ethnic origin might not be optimal). The captured images for Africa, Philippine and Central America are quite few (Figure 5.47) and have no statistical meaning and so were filtered out for this analysis.
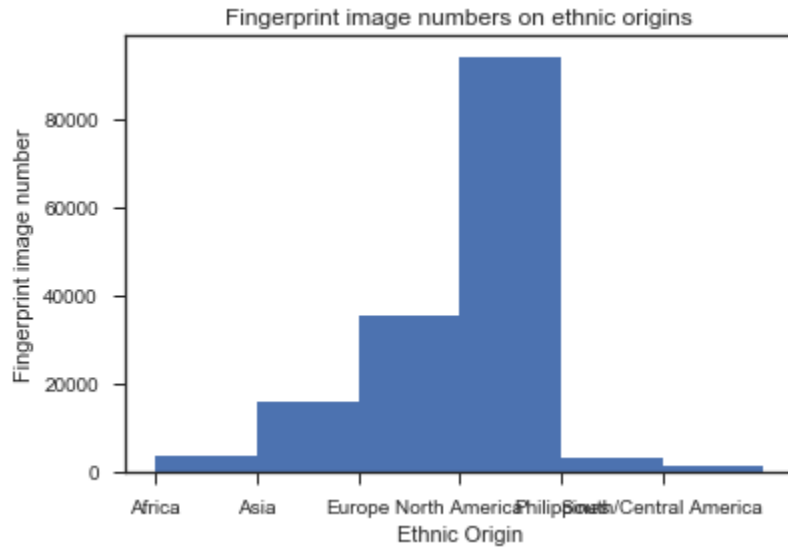
Figure 5.47 Fingerprint image numbers on ethnic origins of verification data

Figure 5.48 shows the subject distributions (same person in different year visits is counted as different subject due to age change for this statistics) in age groups for those from Europe, North America and Asia. The ratio of each age group in different ethnic origins are quite different. The fingerprint quality score distributions on genders of different ethnic origins are not meaningful as the sample population is biased.
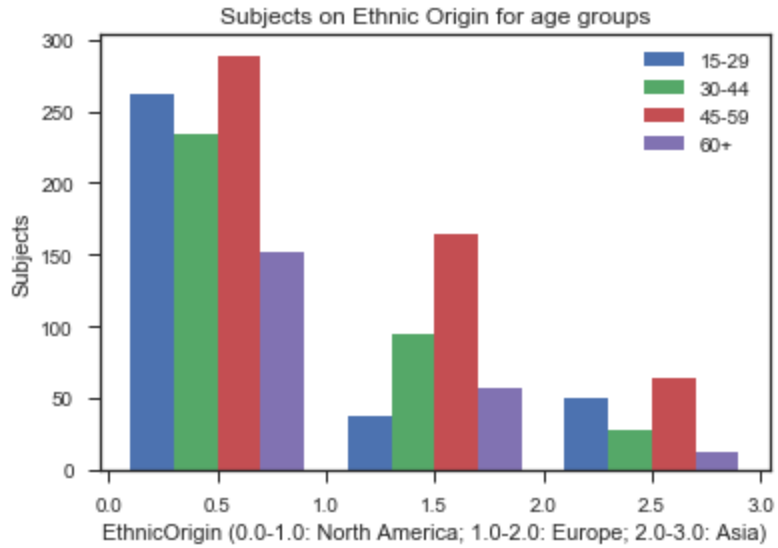
Figure 5.48 Age group distribution for different ethnic origins

Figure 5.49 shows the subject (for all visits) distributions in ethnic origins for those from Europe, North America and Asia. The female/male ratio in different ethnic origins are different but the difference is not too large. The fingerprint quality score distributions on genders of different ethnic origins are more statistically meaningful.
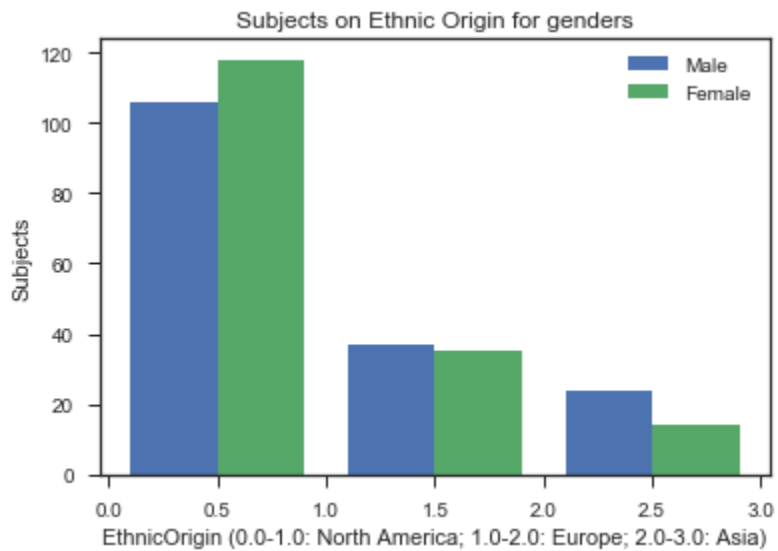
Figure 5.49 Gender distribution for different ethnic origins

Figure 5.50 and Figure 5.51 shows the NFIQ2 quality score distributions on different age groups and different genders for selected ethnic origins. There is no significant difference in quality score distribution on the same gender for different ethnic origins and there are some difference on the same age group though. As the quality score distributions on genders are more statistically meaningful, a conclusion that ethnic origin does not affect fingerprint quality apparently.
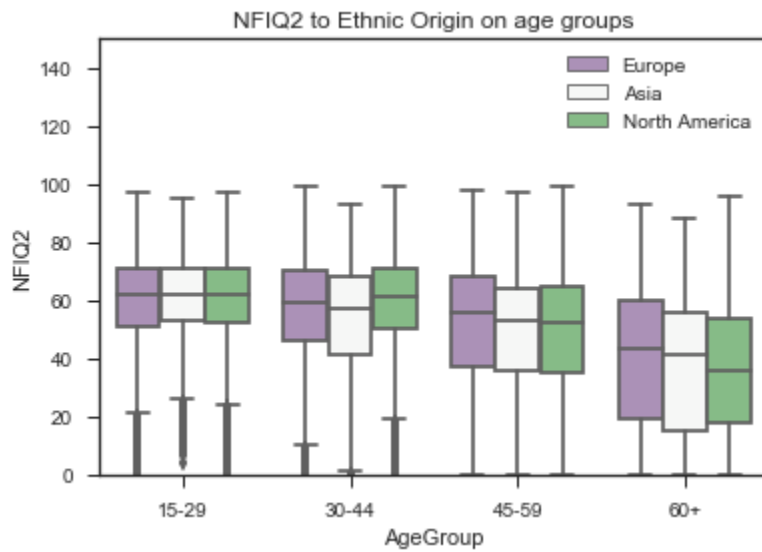
Figure 5.50 NFIQ2 quality score distributions on age groups for selected ethnic origins
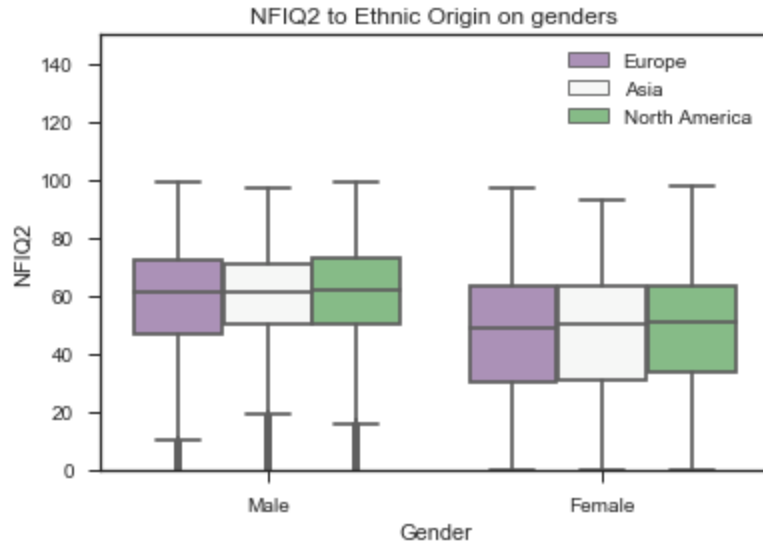
Figure 5.51 NFIQ2 quality score distributions on genders for selected ethnic origins

5.10 Chapter summary

In this chapter, captured data were analyzed on how sensor design, age and genders affect fingerprint quality scores and how the fingerprint quality scores from different measuring algorithms relate to each other. For fingerprint quality scores, verification fingerprint images have better quality than those from enrolment based on t-test result. NFIQ1 scores behave mostly normally on fingerprints from most of the optical sensors but abnormally on those from capacitive based, multispectral and one of the optical test sensors. NFIQ2 scores align well with NFIQ1 scores. When reflecting on NFIQ2, NFIQ1 level 1 and level 2 differ not so much for male subjects in terms of median and IRQ of the boxplots. NFIQ2 scores and NFIQ2 quality feature scores behave more reasonably in most of the sensors. Overall fingerprints have better fingerprint quality on male over female and young over elder but sensor 9 favors more on female fingerprint quality based

on the boxplots. Male fingerprint qualities are more stable than those of females based on IQR of the boxplots. Overall NFIQ2 quality scores are more reliable than NFIQ1 quality scores in terms of abnormality behaviour. The vendor quality scores from some of the sensors are not usable (either fixed or non-reasonable). Different sensors behave differently on NFIQ2 quality features. The multispectral sensor used in this research is better than other used sensors based on the median values and standard deviation, especially for female subjects. Ethnic origin does not affect fingerprint quality much.

6 Chapter: Conclusion

6.1 Thesis conclusion

The main objective of this thesis is to study how sensor design, and age and gender affect fingerprint quality, as measured by the most widely used utility-based fingerprint quality measurement algorithm (NFIQ), with the goal of understanding different sensors' performance and possible impacts from the target population. Different plots (histograms, boxplots), statistical values (mean, standard deviation, etc.) of the data and hypothesis test were used to comparatively study the impact of different factors on fingerprint image quality. Following interesting findings were observed:

➢ Overall

    ✓ Verification images have better quality than enrolment images

    ✓ Sensor 9 generate better quality images for females (vs. males) than the other sensors

    ✓ Males have better quality fingerprint images than females

    ✓ Young subjects have better quality fingerprint images than old subjects

    ✓ NFIQ2 quality scores align well with NFIQ1 quality scores

➢ Exception

    ✓ For sensor 9, fingerprint images for old subjects have better quality than those for young subjects

    ✓ For sensor 9, fingerprint images for female subjects have better quality than those for male subjects

✓ Vendor quality scores of sensor 9, 10, 2 and 5 are not usable

✓ NFIQ1 produces anomalous scores on fingerprint images from sensor 2, 3 and 9

The outcomes of this research provide a lot of information about how different sensors, age, and gender affect fingerprint quality. This study also explores how different fingerprint sensors, subject age and gender on different fingerprint quality features, how different fingerprint quality measurements relate to each other result. The gained knowledge can help sensor selection, sensor design improvement and the improvement of tested quality measurement algorithms.

6.2   Summary of contributions

This research leads to following primary contributions.

- Statistically analyzed the fingerprint quality (in terms of NFIQ1 and NFIQ2 quality scores) distribution differences between enrolment data and verification data. It was concluded that the fingerprint quality on verification data was better than that of enrolment data.

- Comparatively studied fingerprint quality (in terms of NFIQ1 and NFIQ2 quality scores) distributions on different sensors. It was observed that NFIQ2 worked well with the fingerprint images from all test sensors and NFIQ1 ran on all fingerprint images but produced anomalous results on sensor 2, 3 and 9.

- Explored how NFIQ2 quality scores and its quality feature scores aligned with NFIQ1 scores. We saw that NFIQ2 quality scores and NFIQ2 quality feature scores aligned well with NFIQ1 quality scores.

- Comparatively studied NFIQ2 quality feature distributions on different sensors. Sensor 9 outperformed other sensors in most of the selected quality features.

- Checked how vendor quality scores aligned with NFIQ1/NFIQ2 quality scores. It was found that vendor quality scores from some sensors are either constant or not usable. Sensor 9 and 10 uses fixed vendor scores. Sensor 2 has most of the vendor score with value 100. For sensor 5, the vendor quality scores are totally out of sync with NFIQ. Sensor 3, 6, 7, 8 and 12 have reasonable vendor quality calculation algorithms.

- Comparatively studied the impact of ages and genders on fingerprint quality in terms of NFIQ2 quality scores and NFIQ2 quality feature scores. It was learned that fingerprint quality (in terms of NFIQ2 and NFIQ2 quality features) gets worse as the age increases. For gender impact, fingerprint qualities for males are better than those for females.

- Further studied NFIQ2 and NFIQ2 quality feature distributions on age groups for different sensors. A strange behaviour was observed - NFIQ2 quality scores go up with the age growth for sensor 9 while all other sensors go the other way (quality

scores go down with the age growth). For NFIQ2 quality features, sensor 9 outperforms the other sensors on all the selected quality features except frequency domain analysis mean where sensor 7 has the highest scores. Sensor 9 behaves differently on some of NFIQ2 quality features from other sensors.

- Roughly studied the impact of claimed ethnic origins on fingerprint quality (in terms of NFIQ2 quality scores). The population distributions around ethnic origins were explored and only three ethnic origins (North America, Europe and Asia) were selected for the study as the samples for the other ethnic origins are too few. No apparent impact of ethnic origins on fingerprint quality was found based on the limited test data on ethnic origins.

6.3   Future work

There is some additional work that could be done later. Here are the possible studies that can be done later:

- Explore the possibility to design a benchmark platform based on NFIQ2 quality feature scores to evaluate the utility of a fingerprint capturing sensor
- Explore the possibility to use machine learning on NFIQ2 quality feature scores for age, gender prediction
- More investigation on the strange behaviour of NFIQ1 quality scores on sensor 2, 3 and 9. Modifying NFIQ1 code to output more detailed error information might help on the investigation

- Study the impact of manual/chemical on fingerprint qualities and quality features

- Study the impact of age/gender/ethnic origin and sensors on liveness detection

References

[1] O. Barringer and E. Tabassi, "Fingerprint sample quality metric NFIQ 2.0," in *Proceedings of the International Conference of the Special Interest Group on Biometrics*, Darmstadt, Germany, 2011.

[2] NIST, "Development of NFIQ 2.0 – Evaluation of Potential Image Features for Quality Assessment," National Institute of Standard and Technology, Washington DC.

[3] L. Shen, A. Kit and W. Koo, "Quality measure of fingerprint images," 3rd International Conference Audio and Video Based Biometric Person Authentication (AVBPA) 2001, 2001.

[4] E. Lim, X. Jiang and W. Yau, "Fingerprint quality and validity analysis," IEEE International Conference on Image Processing (ICIP '02) 1, 2002.

[5] R. Lolle, "System and methods for determing the quality of fingerprint images". Patent United States patent number US596356, 1999.

[6] T. Chen, X. Jiang and W. Yau, "Fingerprint image quality analysis," in *International Conference on Image Processing (ICIP '04)*, 2004.

[7] S. Xie, S. Yoon, J. Shin and D. Park, "Effective fingerprint quality estimation for diverse capture sensors," *Sensors,* no. ISSN 1424-8220, pp. 7896-7912, 2010.

[8] NIST, "NFIQ 2.0 NIST Fringerprint Image Quality," National Institute of Standard and Technology, Washington DC, 2016.

[9] A. K. Jain, A. A. Ross, K. Nandakumar and b. a, Introduction to Biometrics, Springer Science & Business Media, Nov 17, 2011.

[10] M. Theofanos, R. Micheals, J. Scholtz, E. Morse and P. May, "Does Habituation Affect Fingerprint Quality?," in *The CHI Conference of Human Factors in Computing Systems*, Montréal, 2006.

[11] A. Jain, A. Ross and S. Prabhakar, "An introduction to biometric recognition," *Circuits and Systems for Video Technology,* vol. 14, no. 1, pp. 4-20, 2004.

[12] J. L. Wayman, "Fundamentals of biometric authentication technologies," *International Journal of Image and Graphics,* vol. 1, pp. 93-113, 2001.

[13] N.Geethanjali and K.Thamaraiselvi, "Enhancing the Security of Biometrics in ATM," *International Journal of Scientific & Engineering Research,* vol. 4, no. 4, p. 1192, 2013.

[14] Wikipedia, "Type I and type II errors," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Type_I_and_type_II_errors. [Accessed 5 April 2018].

[15] Wikipedia, "Fingerprint," Wikipedia.org, [Online]. Available: https://en.wikipedia.org/wiki/Fingerprint. [Accessed 27 March 2018].

[16] NFSTC, "A Simplified Guide To Fingerprint Analysis," National Forensic Science Technology Center (NFSTC), Largo, Florida, 2009.

[17] J. V. D. Nardo, "Biometric Technologies: Functionality, Emerging Trends, and Vulnerabilities," *Journal of Applied Security Research,* pp. 194-216, 2009.

[18] K. R. Moses, "Automatic fingerprint identification systems (AFIS)," in *In*

Fingerprint Sourcebook, Washington DC, National Institute of Justice, 2009.

[19] Markets and Markets, "Automated Fingerprint Identification System Market worth 8.49 Billion USD by 2020," Markets and Markets, [Online]. Available: https://www.marketsandmarkets.com/PressReleases/automated-fingerprint-identification-system.asp. [Accessed 6 April 2018].

[20] ISO/IEC 29794-1, "Information technology - biometric sample quality - part 1: Framework," INTERNATIONAL STANDARD , 2009.

[21] A. Adler and R. Youmaran, "Low Quality Images are Less Informative: Measuring Quality via Information Content," 2007.

[22] L. Hong, Y. Wan and A. Jain, "Fingerprint image enhancement: Algorithm and performance evaluation," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 20, pp. 777-789, 1998.

[23] R. M. Bolle, S. U. Pankanti and Y. Yao, "System and method for determining the quality of fingerprint images". US Patent 5963656, 5 October 1999.

[24] N. K. Ratha and R. Bolle, "Fingerprint image quality estimation," IBM TJ Watson Research Center, 1999.

[25] B. Lee, J. Moon and H. Kim, "A novel measure of fingerprint image quality using the Fourier spectrum," in *Proceedings of the SPIE*, 2005.

[26] Y. Chen, S. C. Dass and A. K. Jain, "Fingerprint quality indices for predicting authentication performance," in *5th International Conference Audio- and Video-Based Biometric Person Authentication (AVBPA)*, 2005.

[27] G. Y. Luo, "Objective image quality measurement by local spatial-frequency wavelet analysis," *International Journal of Remote Sensing,* vol. 27, pp. 5003-5025, 2007.

[28] S. Lee, C. Lee and J. Kim, "Model-based quality estimation of fingerprint images," in *IAPR/IEEE International Conference on Biometrics (ICB'06)*, 2006.

[29] H. Fronthaler, K. Kollreider, J. Bigun, J. Fierrez, F. Alonso-Fernandez, J. Ortega-Garcia and J. Gonzalez-Rodriguez, "Fingerprint image quality estimation and its application to multi-algorithm verification," in *IEEE Transactions on Information Forensics and Security*, 2008.

[30] E. Tabassi, C. Wilson and C. Watson, "Nist fingerprint image quality," NIST, 2004.

[31] J. Merkle, M. Schwaiger, O. Bausinger, M. Breitenstein, K. Elwart and M. Nuppeney, "Towards Improving the NIST Fingerprint Image Quality (NFIQ) Algorithm (Short Version)," Proc. BIOSIG 2010, 2010.

[32] J. Merkle, M. Schwaiger, O. Bausinger, M. Breitenstein, K. Elwart and M. Nuppeney, "Towards Improving the NIST Fingerprint Image Quality (NFIQ) Algorithm (Extended Version)," arXiv preprint arXiv:1008.0781, 2010.

[33] M. A. Olsen, H. Xu and C. Busch, "Gabor filters as candidate quality measure for nfiq 2.0," in *NIST International Biometric Performance Testing Conference*, 2012.

[34] M. E. Abed, A. Ninassi, C. Charrier and C. Rosenberger, "Fingerprint quality assessment using a no-reference image quality metric," in *European Signal Processing Conference (EUSIPCO)*, 2013.

[35] A. Awasthi, K. Venkataramani and A. Nandini, "Image quality quantification for fingerprints using quality-impairment assessment," in *IEEE Workshop on Applications of Computer Vision (WACV) 2013*, 2013.

[36] Z. Yao, C. Charrier and C. Rosenberger, "A Preliminary Study of Fingerprint Quality Assessment of Minutiae Template," in *Summer school in Advanced Studies on Biometrics for Secure Authentication*, 2014.

[37] E. Tabassi and C. L. Wilson, "A novel approach to fingerprint image quality," in *International Conference on Image Processing (ICIP)*, 2005.

[38] E. Tabassi, "NIST Fingerprint Image Quality and relation to latent prints," Image Group – NIST, 2006.

[39] J. L. Blue and P. J. Grother, "Training feed forward networks using conjugate gradients," in *Conference on Character Recognition and Digitizer technologies*, San Jose, 1992.

[40] D. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Imathematical programming B,* vol. 45, pp. 503-528, 1989.

[41] O. M. Omidvar and C. L. Wilson, "Information content in neural net optimization," *Journal of connection science,* vol. 6, pp. 91-103, 1993.

[42] E. Tabassi, M. A. Olsen, A. Makarov and C. Busch, "Towards NFIQ II Lite, Self-Organizing Maps for Fingerprint Image Quality Assessment," NIST Interagency Report 7973, 2014.

[43] T. Kohonen, "The self-organizing map," *IEEE J PROC,* vol. 78, no. 9, pp. 1464-

1480, 1990.

[44] L. Breiman, "Random forests," *Mach. Learn,* vol. 45, no. 1, pp. 5-32, 2001.

[45] M. Schwaiger and E. Tabassi, "NFIQ 2.0 Open Source Distribution," NIST, Gaithersburg, 2014.

[46] NIST, "GitHub NFIQ 2.0 project," NIST, [Online]. Available: https://github.com/usnistgov/NFIQ2. [Accessed 20 March 2018].

[47] J. A. Hasselgren, S. J. Elliott and J. Gue, "A Trade-off Between Number of Impressions and Number of Interaction Attempts," in *The 8th International Conference on Information Technology and Applications (ICITA 2013)*, 2013.

[48] C. R. Blomeke, S. K. Modi and S. J. Elliott, "Investigating the Relationship Between Fingerprint Image Quality and Skin Characteristics," in *42nd Annual IEEE International Carnahan Conference on Security Technology*, 2008.

[49] M. Theofanos, S. Orandi, R. Micheals, B. Stanton and N. Zhang, "Effects of Scanner Height on Fingerprint Capture," National Institute of Standards and Technology, Gaithersburg, 2006.

[50] A. S. Falohun, O. D. Fenwa and F. A. Ajala, "A Fingerprint-based Age and Gender Detector System using Fingerprint Pattern Analysis," *International Journal of Computer Applications (0975 – 8887),* vol. 136 – No.4, 2016.

[51] H. Kang, B. Lee, H. Kim, D. Shin and J. Kim, "A Study on Performance Evaluation of Fingerprint Sensors," *Lecture Notes in Computer Science,* vol. 2688, p. 1055, 2003.

[52] S. J. Elliott and N. C. Sickler, "An evaluation of fingeprint image quality across an elderly population vis-a-vis an 18-25-year-old population," in *the International Carnahan Conference on Security Technology*, Las Palmas, Gran Canaria, October 12-14, 2005.

[53] S. K. Modi and S. J. Elliott, "Impact of image, quality on performance: Comparison of young and elderly fingerprints," in *the 6th International Conference on Recent Advances in Soft Computing (RASC)*, Canterbury, UK, July 10-12, 2006.

[54] S. K. Modi, S. J. Elliott, J. Whetsone and H. Kim, "Impact of Age Groups on Fingerprint Recognition Performance," in *IEEE Workshop on Automatic Identification Advanced Technologies*, Alghero, Italy, 2007.

[55] D. Thakkar, "Fingerprint Reader Technology Comparison: Optical Fingerprint Scanner; Capacitive-based Fingerprint Reader and Multispectral Imaging Sensor," Bayometric Technology, [Online]. Available: https://www.bayometric.com/fingerprint-reader-technology-comparison/. [Accessed 3 April 2018].

[56] S. J. Xie, J. C. Yang, D. S. Park, S. Yoon and J. Shin, "Fingerprint Quality Analysis and Estimation for Fingerprint Matching," *State of the Art in Biometrics,* 2011.

[57] E. ONeil, "ATMs Use Biometrics to Combat Fraud," [Online]. Available: https://www.thebalance.com/atms-use-biometrics-to-combat-fraud-315794. [Accessed 23 5 2018].