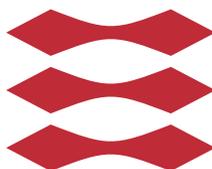


Fingerprint quality assurance using image processing

Marek Dusio

DTU



Kongens Lyngby 2013
IMM-M.Sc.-2013-46

Technical University of Denmark
Informatics and Mathematical Modelling
Building 321, DK-2800 Kongens Lyngby, Denmark
Phone +45 45253351, Fax +45 45882673
reception@imm.dtu.dk
www.imm.dtu.dk IMM-M.Sc.-2013-46

Summary

The subject of this dissertation is measuring fingerprint sample quality using image analysis methods that are fast to compute. Of interest are analysing the impact of fingertip skin moisture on quality and providing feedback if the finger is too dry or too wet. The goal is to propose methods that could be incorporated in NIST Finger Image Quality version 2.0 or in ISO/IEC standards.

A dataset of 6600 samples is collected from 33 subjects using 5 sensors with objective fingertip moisture measurement and varying skin moisture conditions. The impact of skin moisture on fingerprint sample quality is analysed and a Moisture Indication method is proposed and used with thresholds to provide binary indication on skin dryness or wetness.

Three fingerprint Quality Measurement Algorithms are proposed - Ridge Valley Difference, Ridge Line Count and Contrast; their performance is assessed in terms of execution time, output quality correlation with observed utility, and using Error versus Reject Curves. The methods are compared to current state of the art: NIST Finger Image Quality, Orientation Certainty Level, Ridge Valley Uniformity, Local Clarity Score and Gabor Shen.

All proposed methods work and offer acceptable performance - all are fast to compute and provide quality that predicts samples' performance. Some proposed methods are better than state of the art in terms of either execution time or in performance prediction. The Moisture Indication method is successfully used to classify samples as acquired from dry or wet skin with reasonable detection error rates. All proposed methods can possibly be incorporated in the ISO/IEC standards or in NFIQ 2.0.

Preface

This thesis was prepared at the department of Informatics and Mathematical Modelling at the Technical University of Denmark during an exchange visit at the Center for Advanced Security Research Darmstadt in fulfilment of the requirements for acquiring an M.Sc. in Computer Science and Engineering.

This project deals with the quality of fingerprint samples and its relation to fingertip skin moisture during fingerprint sample acquisition.

This document consists of an introduction, description of theory, state of the art methods, proposed methods, experimental setup, results, conclusions and eight appendices.

Lyngby, 15-June-2013

A handwritten signature in black ink, reading "Marek Dusio". The signature is written in a cursive, slightly slanted style.

Marek Dusio

Acknowledgements

The work underlying this thesis was supported by the Center for Advanced Security Research Darmstadt (CASED) and the German Federal Office for Information Security (BSI).

My special thanks to Ralph Leßmann, the Director of Software Solutions in Cross Match Technologies, Inc. for providing me with Cross Match sensors, Patrol ID and the newest Guardian. Thanks to Prof. Dr. Christoph Busch for providing me with L1 DFR-2100 and Cross Match L Scan 100. Thanks to Martin Aastrup Olsen for lending Dermalog ZF-1.

Thanks to Elham Tabassi, working at NIST, for fruitful discussions and providing results of proposed methods assessment on a subset of NIST operational data.

Thanks to the German TeleTrusT Biometrics Working Group and NFIQ development team for valuable comments and interest in the proposed methods.

Thanks to all the 33 participants of fingerprint dataset collection for their time and patience. Special thanks to Prof. Dr. Christoph Busch and Alexander Nouak, head of the Identification and Biometrics competence center at Fraunhofer IGD, for help in finding and motivating these participants.

Finally I would like to thank my supervisors Prof. Dr. Christoph Busch, Prof. Dr. Rasmus Larsen and Martin Aastrup Olsen, M.Sc. for invaluable comments and guidance.

Contents

Summary	i
Preface	iii
Acknowledgements	v
1 Introduction	1
2 Biometrics	5
2.1 Introduction	5
2.2 Biometric modality	6
2.3 Biometric processes	6
2.3.1 Enrolment	8
2.3.2 Recognition	8
2.3.3 Biometric comparison	9
2.3.4 Biometric decision	9
2.4 Biometric system errors	9
2.5 Biometric system performance	10
2.6 Fingerprint recognition systems	11
2.7 Fingerprint sample acquisition	12
2.7.1 Optical sensors	13
2.7.2 Solid-state	13
2.8 Fingerprint comparison	13
2.9 Fingerprint features	14
2.9.1 Local Ridge Orientation	14
3 Biometric sample quality	17
3.1 Introduction	17
3.2 ISO quality definition	18

3.2.1	Quality components	18
3.3	Quality purpose	19
3.3.1	Enrolment	19
3.3.2	Verification and Identification	20
3.3.3	Conditional processing	20
3.3.4	Survey and diagnosis	20
3.4	Observed utility calculation	21
3.5	Quality estimation	22
3.5.1	Quality estimation effectiveness	23
3.6	Fingerprint sample quality	26
4	State of the art in fingerprint quality metrics	29
4.1	Introduction	29
4.2	Orientation Certainty Level	30
4.2.1	OCL calculation procedure	31
4.3	Ridge Valley Uniformity	32
4.3.1	RVU calculation procedure	33
4.4	Local Clarity Score	34
4.4.1	LCS calculation procedure	34
4.5	Gabor Shen	36
4.5.1	Gabor Shen calculation procedure	36
4.6	NIST Fingerprint Image Quality	38
4.6.1	NFIQ calculation procedure	38
4.6.2	NFIQ 2.0 development	39
4.7	Image segmentation	39
4.7.1	Segmentation procedure	40
5	Proposed Fingerprint Analysis Methods	43
5.1	Ridge Line Count (RLC)	44
5.1.1	Block-wise score aggregation	44
5.1.2	Ridge Line Count calculation procedure	45
5.2	Ridge Valley Difference (RVD)	46
5.2.1	Block-wise score aggregation	47
5.2.2	RVD calculation procedure	48
5.3	Contrast (CNT)	49
5.3.1	Block-wise score aggregation	49
5.3.2	CNT calculation procedure	50
5.4	Moisture Indication method	51
5.4.1	MI calculation procedure	52

6	Experimental setup	57
6.1	Project purpose and motivation	58
6.2	Target sensor technology	58
6.3	Procedure of the experiment	59
6.4	Existing fingerprint dataset choice	60
6.5	New fingerprint dataset collection	61
6.5.1	Moisture measurement	62
6.5.2	Fingerprint sensors	64
6.5.3	Skin conditions variation	64
6.5.4	Acquisition procedure	66
6.5.5	Dataset collection summary	68
6.6	Fingerprint sample analysis	68
6.6.1	Comparison score calculation	68
6.6.2	Observed utility extraction	69
6.6.3	Fingerprint quality estimation	69
6.7	Proposed methods performance assessment	71
6.7.1	Average execution time	71
6.7.2	Correlation with utility	71
6.7.3	Error versus Reject Curves	72
6.8	Moisture impact on quality	74
6.8.1	Moisture impact on comparison scores	74
6.8.2	Proposed moisture indication method performance	75
7	Experimental results	77
7.1	Calculated comparison scores	77
7.1.1	Impostor distributions	78
7.1.2	Genuine distributions	78
7.1.3	Summary	78
7.2	Proposed methods evaluation	79
7.2.1	Execution time	79
7.2.2	Error versus Reject Curves	80
7.2.3	Correlation with utility	82
7.2.4	Summary	84
7.3	Moisture impact on quality	85
7.3.1	Measurement statistics	85
7.3.2	Moisture versus comparison scores	86
7.3.3	Summary	88
7.4	Moisture Indication method performance	89
7.4.1	Execution time	89
7.4.2	Detected versus measured moisture	89
7.4.3	Binary classification error	90
7.4.4	Dryness classification	90
7.4.5	Wetness classification	92
7.4.6	Summary	94

8	Conclusions	95
8.1	Future work	96
A	Calculated comparison scores	97
B	Error versus Reject Curves	99
C	Quality and Utility correlation	107
D	Moisture impact on quality	111
E	Moisture impact on quality – sensor differences	119
F	Wetness classification	125
G	Dryness classification	131
	Bibliography	137

CHAPTER 1

Introduction

Personal identity is questioned frequently each day these times. Is a person authorized to access a certain resource? Is this the person he or she claims to be? Or even – who is this person? Throughout the technological development of mankind, several mechanisms were brought to life, be it locks and keys, secret password or PIN-code protection, recently RFID tags and key-cards. Some of these mechanisms are more, some less sophisticated; some offer more security, some are less robust. However, all these methods require special handling – secrets, passwords and PIN codes have to be memorized; keys, key-cards and RFID transceivers need to be carried. All these feature a common characteristic – keys and secrets can be shared and transferred between people, unfortunately not only on purpose.

Natural methods of recognition – recognizing faces and voice, eyes or movement – are subconsciously performed by the human brain. Nowadays, with the development of digital devices, the recognition can be performed automatically, analysing one or more characteristics of the human body, e.g. iris, ear structure, palm vein pattern, fingerprints, etc. Using these for automatically recognizing a person is called biometric recognition or biometrics. Different characteristics give different biometric recognition performance and are differently accepted by people.

Fingerprint recognition is widely accepted as samples are fairly easy to collect.

It is also widely used in the government and industry sectors. USA has introduced the US-VISIT program [usv], India the Unique Identification number [aad] and the United Nations' International Civil Aviation Organization has a raising interest in biometrics [ica].

No biometric system is free of error and it is always of interest to limit the error rates, so that the systems perform better. It is internationally agreed by government, industry and academia that the performance of a biometric system is related the quality of samples it works with [GT07]. Indeed, poor quality samples can generate spurious or missing features which may lead to false rejections and false acceptances of genuine and impostor comparisons respectively.

If possible the sample can be manually assessed to ensure best quality, e.g. during check-in, an airport officer inspects acquired fingerprint samples. Should the quality be insufficient, sample reacquisition may be necessary. Usually the subjects would be instructed how to behave – so that the quality of the acquired samples increases – i.e. put less pressure on the sensor platen, place the finger correctly, dry or moisturise the fingertip skin. Usually these instructions would be accurate if the officer is experienced.

With the development of biometric sensors, unsupervised Automatic Fingerprint Identification Systems (AFIS) were introduced, where an automatic quality measurement is performed. The International Organization of Standardization and the National Institute of Standards and Technology of the U.S. Department of Commerce are working on standardizing the methodologies of automatic quantitative fingerprint sample quality measurements [ISO09] [ISO10] [TW05]. Many fingerprint sample quality measurement algorithms exist, a de facto standard is NFIQ by NIST [TW05].

Measuring the quality allows to decrease error rates of biometric systems by rejecting poor quality samples, yet no feedback is provided to the user on why the quality is poor. If a sample is reacquired with unchanged conditions, the quality may remain the same. An initial study has shown that poor quality samples acquired from dry and wet fingers give a distinct impression when assessed manually. It is interesting to check whether these samples indeed cause performance degradation and if so, if it is possible to automatically distinguish them and provide accurate feedback on finger wetness or dryness to the user.

A new version of the NFIQ standard is in development [oST], where several algorithms are combined together to provide even better quality estimation. The subject of this thesis is to propose Quality Measurement Algorithms that are fast to compute and offer good performance benefits when incorporated in a biometric system. The motivation is to compete with the existing state of the art algorithms and hopefully add to the new version of the standard. Perhaps

it is also possible to relate the quality of acquired samples to the moisture of fingertip skin and if reacquisition is required provide useful feedback to the users so that they can e.g. dry or moisten their hands so that the quality of fingerprint samples acquired afterwards is improved.

Biometrics

This chapter gives an overview of biometric systems. The general concept of biometric recognition is introduced, performance and errors of a biometric system are discussed. Finally, biometric fingerprint recognition systems are described.

The concepts are illustrated accordingly to the ISO/IEC 29794–1 standard [ISO09] developed by International Electro technical Commission of International Organization for Standardization (ISO/IEC). Vocabulary is used accordingly to the ISO/IEC JTC 1/SC 37 Harmonized Biometric Vocabulary defined in SC37 Working Group 1 for the International Standard ISO/IEC 2382–37 [ISO12].

2.1 Introduction

The word *Biometrics* comes from Greek *bios* (life) and *metron* (measurement). Biometric identifiers are anatomical and behavioural characteristics (traits) of a human body.

Biometric recognition (biometrics) is associating these traits with an established identity of a person and using this relation to automatically recognize individuals.

2.2 Biometric modality

Theoretically, any trait or body characteristic (also called modality) could be used in biometrics as long as it is:

- *universal* – every person has it,
- *distinctive* – no two people have it identical,
- *permanent* – unchanged over time,
- *collectible* – it can be measured quantitatively.

In practice, a biometric trait is also evaluated based on the performance it gives in terms of speed and recognition accuracy and ease of circumvention. It is also important to assess how capturing a certain characteristic is accepted by people, e.g. some people may prefer not to have their eyes scanned [MMJP09].

Several characteristics are used in biometrics successfully, some examples of modalities are:

- *Fingerprint* – using ridge line structure of fingertips,
- *Iris* – using iris visual texture,
- *Face* – using facial features size and distance,
- *Vein* – analysing the vein structure of the palm or finger,
- *Voice* – measuring the sound characteristics of speech.

2.3 Biometric processes

Biometric recognition can be applied in several contexts, e.g. granting or denying access, confirming claimed identity or simply for identity recognition. Different applications use slightly different biometric processes, however all systems have several elements in common – all systems capture and store the data, perform signal processing, comparing and make biometric decisions.

Figure 2.1 shows a diagram of a conceptual biometric system structure with the common elements grouped. The process of biometric recognition is also marked

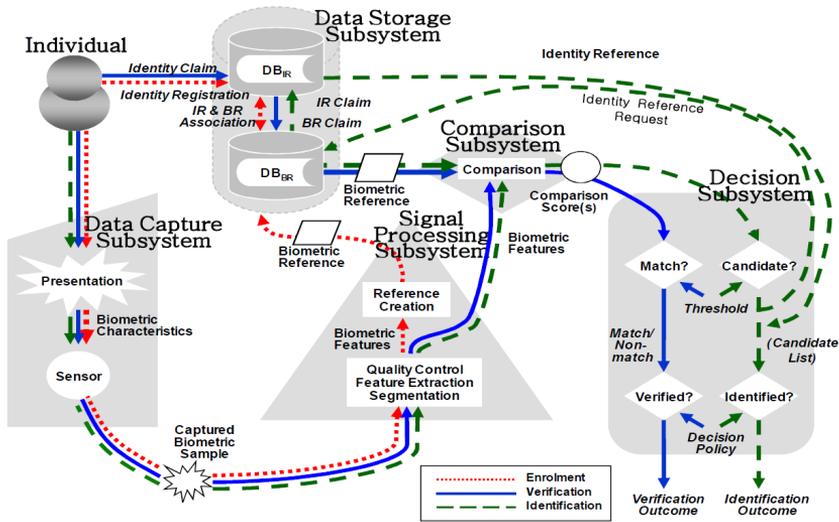


Figure 2.1: Conceptual structure of a biometric system, as defined in ISO/IEC 24745 (source [ISO11]).

with arrows (blue for verification, green for identification, red for enrolment). This process is generally performed as follows:

1. *Data Capture* – biometric sensor acquires a sample from a subject,
2. *Identity Claim* is required in identity verification to get an existing template,
3. *Signal processing* – the sample is analysed and features are extracted,
 - *Enrolment Template* – the resulted features can be saved in the database as a biometric reference,
 - or compared with one or more existing templates during *Verification* or *Identification*,
4. *Comparison Subsystem* determines the comparison score between captured sample and stored reference,
5. *Decision* is made in based on the comparison score.

2.3.1 Enrolment

When a user interacts with a biometric system for the first time, the biometric characteristic is captured to obtain a sample. If the quality of the sample is sufficient, features are extracted to form the feature set (a *biometric reference*) which is registered in the biometric system storage as a biometric enrolment data record. The reference template is associated with the identity of a person, so that it is known whose template it is. This process is called *enrolment*.

2.3.2 Recognition

The second time a person interacts with a biometric system, the biometric characteristic is captured again and similarly as before, sufficient quality of the sample allows feature extraction. This time however, the feature set is a biometric query or a biometric probe. A probe is compared to a reference template to establish a comparison score.

Depending on the application context, a biometric system is either an *identification* or a *verification* system. The term *recognition* is used if the application context is not of interest. The enrolment process can be performed regardless on the application context, but the functions of identification and verification systems are different.

- The *verification* process will authenticate a person and confirm their identity is what they claim to be. This is done by capturing a probe and comparing it to the reference associated with the claimed identity. This is usually a one-to-one comparison.
- The *identification* process is a one-to-many comparison. It recognizes a person by searching an entire database of templates to find a list of possible candidates that match with the captured probe. This can be performed in closed-set, where every person interacting with a system is expected to have been enrolled, and in open-set where it may not be the case.

A closed-set identification can usually give a positive *biometric identification decision* and usually a non-empty set of *biometric candidates*.

Open-set identification may well give an empty set of biometric candidates and a negative identification decision – a person may not be enrolled in the database and no template would be found.

2.3.3 Biometric comparison

The one-to-one comparison of verification and each of the one-to-many comparisons performed during identification are all the process of calculating the similarity or dissimilarity between two biometric samples – the probe and the reference [ISO12]. If a biometric comparison is done between two samples that are taken from the same characteristic of the same person, it is a mated comparison, otherwise the comparison is non-mated. ISO/IEC 2382–37 defines this as follows:

- *mated* – "a comparison of a paired probe and reference from the same characteristic of the same data subject" [ISO12].
- *non-mated* – "a comparison of a probe and a reference from the same characteristic of different data subjects" [ISO12].

2.3.4 Biometric decision

The *comparison score* is used to give the *comparison decision* based on whether the score is above or below a certain *threshold* as follows:

- *match* – score above the threshold gives a positive comparison decision "stating that the biometric probe and the biometric reference are from the same source" [ISO12].
- *non-match* – score below the threshold gives a negative comparison decision "stating that the biometric probe and the biometric reference are not from the same source" [ISO12].

2.4 Biometric system errors

Even though biometric systems use sophisticated hardware and state of the art algorithms, sometimes errors are inevitable. A comparison decision is not always correct – sometimes mated samples are rejected and non-mated accepted, which results in the following errors:

- *False Rejection* – biometric decision of a non-match from a comparison of mated samples. Defined as: "error of rejecting a biometric claim that

should have been accepted in accordance with an authoritative statement on the origin of the biometric probe and the biometric reference" [ISO12].

- *False Acceptance* – biometric decision of a match from a comparison of non-mated samples. Defined as: "error of accepting a biometric claim that should have been rejected in accordance with an authoritative statement on the origin of the biometric probe and the biometric reference" [ISO12].

Errors of biometric systems that were not caused during the comparison procedure are also specified, as defined by ISO/IEC 2382-37 [ISO12]:

- *Fail To Capture (FTC)* – the capturing module may fail to properly capture the sample.
- *Fail To Process (FTP)* – the feature extraction module may fail to process a sample.
- *Fail To Acquire (FTA)* – a combination of first three errors – *FDC*, *FTC*, *FTP* – a general failure in sample acquisition.
- *Fail To Enrol (FTE)* – the template generation module can fail to extract a template resulting in an enrolment failure.

2.5 Biometric system performance

The performance of biometric systems is related to the error rates it gives and it can be assessed in several ways, since there are different errors that can occur. However, most modern systems are sophisticated and do not give a lot of *FTA* or *FTE* errors if the environment conditions are good. Hence, typically the matching error rates are of interest in performance assessment. Figure 2.2 shows the distribution of comparison scores in a theoretical biometric system:

- *genuine distribution* of non-mated samples' comparison scores,
- *impostor distribution* of mated samples' comparison scores.

In a perfect system working with ideal samples, the impostor and genuine comparison score distributions could be separated completely. Practically, both distributions overlap in the middle because some mated samples will give comparison scores relatively lower than some other non-mated samples. The threshold used to determine the decision boundary gives rise to the following error rates:

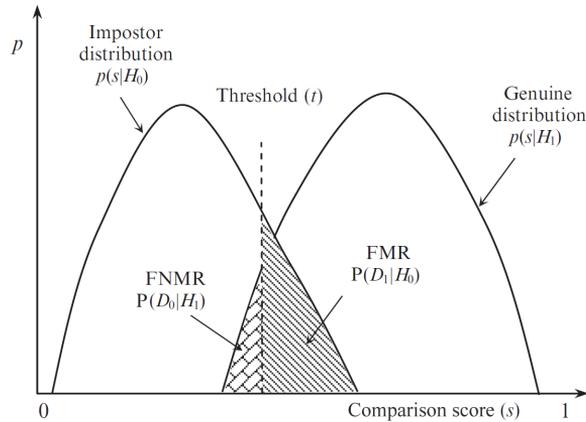


Figure 2.2: Impostor and genuine comparison scores' distributions. FNMR and FMR for a given threshold t . Source [MMJP09] page 17.

- *False Non-Match Rate (FNMR)* – the rate of mated comparisons that get falsely rejected because the comparison score is below the threshold,
- *False Match Rate (FMR)* – the rate of non-mated comparisons that get falsely accepted because the comparison score is above the threshold.

The biometric threshold is chosen accordingly to the required security. The middle point is called *Equal Error Rate (EER)*, where the threshold is chosen such that FNMR is equal to FMR.

If high security is required, e.g. in a bank, the threshold would be chosen high so that False Match Rate is very small and the probability of impostors gaining access is low. On the contrary, low security will require a low threshold so that false rejections will not occur frequently, e.g. for classroom access, where some impostors are allowed.

2.6 Fingerprint recognition systems

Biometric fingerprint recognition systems use the anatomical characteristic of a person's fingertip skin. The surface of the epidermal skin layer that covers fingertips is covered with small ridges and valleys. On a typical fingerprint sample, ridges are represented by dark lines on a white background. Figure 2.3 shows examples of fingerprint samples acquired using different sensors.

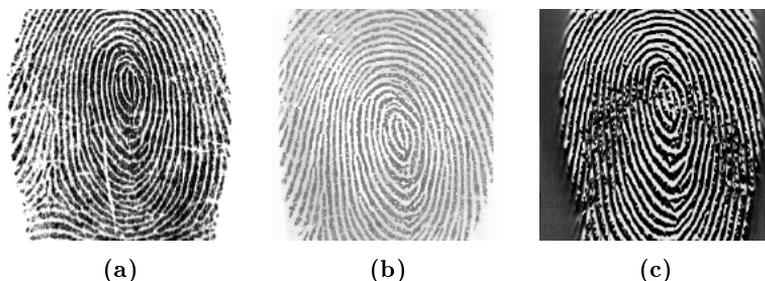


Figure 2.3: Fingerprint sample excerpts acquired from the same finger using different live scanners: **a)** FTIR optical scanner, **b)** solid-state capacitive scanner, **c)** solid-state thermal scanner (source: [MMJP09]).

Identifying criminals by their fingerprints was introduced in 1860s by Sir William James Hershel. The probability of identical fingerprints between two different people was first scientifically analysed by Sir Francis Galton. He also introduced a classification of fingerprints. By 1920s, fingerprint identification was widely used as a form of personal identification by American police and FBI.

2.7 Fingerprint sample acquisition

In order to perform biometric fingerprint recognition, the sample has first to be acquired. Historically before the development of digital systems and Automated Fingerprint Identification Systems (AFIS), this was done off-line by smearing fingertips with black ink and pressing or rolling them against a paper card. Forensic application included collecting latent fingerprints e.g. impressions left on objects at crime scenes.

Modern AFIS utilise scanners (also called sensors) which allow on-line, live sample capturing. This way when the finger is placed on the platen surface, the capture module of an AFIS detects finger presence and immediately captures one or more samples.

Different technologies are used for live-scan fingerprint sensing; most technologically mature and widely used sensors are either optical or solid-state [MMJP09].

2.7.1 Optical sensors

Optical sensors use the *Frustrated Total Internal Reflection* (FTIR) method. The finger platen a side of a glass or plastic prism. Light is projected from one side of the prism to the finger, projecting the image of the fingerprint to the opposing side of the prism which contains a CCD or CMOS camera. Skin ridges touch the surface of the prism but valleys remain in a distance. Light is thus reflected where the valleys are and absorbed (scattered) at ridges. This results in an image of dark ridge lines on a light background.

Optical sensors offer high quality of fingerprint samples and if the size of the prism is large enough, a four-finger slap scan may be possible. A modified version of an optical sensor may use a sheet prism or optical fiber. This way the sensor would be a bit smaller, but the quality of acquired samples may degrade [MMJP09].

2.7.2 Solid-state

Solid-state sensors (also called silicon) are smaller and cheaper to manufacture than optical sensors, but they do not offer the same quality. The sensor platen is an array of several smaller sensors, each producing a pixel of the acquired sample [MMJP09]. Three types of silicon sensors exist: capacitive, thermal and electric field.

Solid state sensors allow to acquire the fingerprint sample by swiping it against the sensor instead of placing the finger on the platen. This was introduced with the thermal type, where the swiping was necessary to keep the temperature of the sensor changing. Swiping has a benefit of self-cleaning the sensor and allows to make the sensor even smaller (a few pixel rows), but requires processing to compose the final sample from the several acquired pixel rows.

2.8 Fingerprint comparison

When a fingerprint probe is acquired, it can be matched with fingerprint templates. Biometric identification and verification are similar in this step as the one-to-many comparisons performed in identification can be viewed as several one-to-one verification comparisons. Comparison of fingerprint template and probe samples can be done as follows:

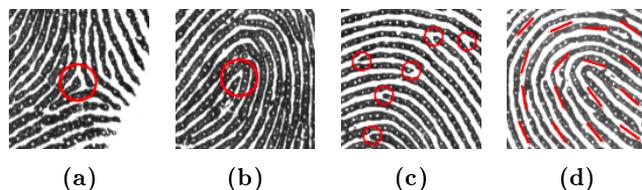


Figure 2.4: Examples of fingerprint features marked in red: **a)** delta, **b)** core, **c)** minutiae (ridge endings), **d)** local ridge orientation.

- *Comparison by correlation* – two images are superimposed and pixel correlation is measured,
- *Minutiae based comparison* – measuring the alignment between probe and reference feature sets,
- *Non-minutiae based comparison* – measuring local ridge orientation or frequency, texture information, etc.

Most widely known and utilised are minutiae comparison methods [MMJP09], but they require feature set extraction performed prior to comparison.

2.9 Fingerprint features

The ridges present on the fingertip skin form into a pattern with characteristic features, that can be categorized into three levels described below. Figure 2.4 shows examples of fingerprint features.

1. *Singular points* – core which is always present and delta which may be present,
2. *Minutiae* – local ridge characteristics, i.e. ridge line endings, bifurcations, lakes, islands, crossovers, etc.
3. *Sweat pores* – or other intra-ridge details – width, shape, curvature.

2.9.1 Local Ridge Orientation

Ridge orientation is a non-minutiae feature that represents the direction of the ridges. It can be measured at each pixel of a sample, or locally for small blocks

of a sample to increase computational efficiency.

Local Ridge Orientation at pixel "represents the angle at which the fingerprint ridges cross with the horizontal axis" [MMJP09].

Biometric sample quality

This chapter discusses the biometric sample quality according to ISO/IEC 29794-1 [ISO09] definitions. First the concept is introduced, secondly the uses of quality measure are discussed with possible applications. Finally, the quality of fingerprint samples is described and related to the earlier introduced concepts.

3.1 Introduction

As described in sections 2.4 and 2.5, the comparison modules in biometric systems are far from perfect and sometimes falsely match non-mated or otherwise falsely reject mated samples. An overlap is seen between the Impostor and Genuine distributions. What is the reason behind it? Some samples give good recognition performance, some lead to false decisions. Depending on several factors, some samples simply have better quality than others.

Indeed, it is now internationally agreed in industry, government and academia that the "quality of a sample should be related to its recognition performance" [GT07]. In other words, a biometric system working with highest quality samples will have better (smaller) error rates than that same system working with poorest quality samples.

3.2 ISO quality definition

One measure of quality could describe some properties of a sample that are useful in one application, but provide no information for a different system. Grother and Tabassi [GT07] give a good example for fingerprints recognition where a sample may have high quality if it has sharp and clearly visible ridges, but perform poor in a minutiae-comparison system if it has few minutiae.

The term "quality" is generally broad and might be used to describe several attributes of a biometric sample. ISO/IEC 29794-1 gives a concrete definition of quality of a biometric sample as the predictor of its biometric performance:

"Biometric sample quality is the degree to which a biometric sample fulfils specified requirements for a targeted application." [ISO09]

3.2.1 Quality components

The performance of a sample in a biometric system depends on the *source of the sample*, and *how well the sample represents its source*.

Low performance might be caused by poor quality of the source, e.g. low quality fingerprint due to skin scars or disease, or by the fact that the sample does not capture the source well, e.g. the sample is blurry. These components of sample quality are defined by ISO/IEC 29794-1 [ISO09] as follows:

1. *Character* of a sample, "attributable to inherent features of the source from which the sample is derived". For example, fingerprints with scars or blisters have poor character.
2. *Fidelity* of a sample to its source, "reflects the degree of similarity between a sample and the source from which it is derived". Several aspects influence fidelity: user behaviour, environment conditions, capturing and feature extraction modules, etc.
3. *Utility* of a sample within a biometric system, the "observed performance of a biometric sample (or a set of samples) in one or more biometric systems". Utility depends on both the character and fidelity of a sample and is intended to predict the *FNMR*, *FMR* and also *FTA* and *FTE* rates. The utility-based quality shows the negative or positive contribution of a sample to the overall system performance.

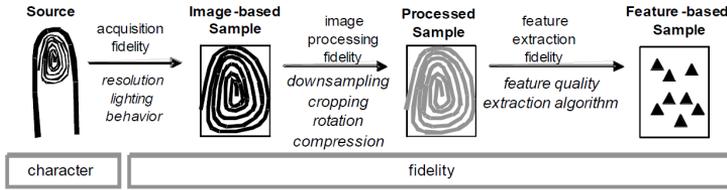


Figure 3.1: Quality reference model as defined in ISO/IEC 29794–1 (source [ISO09]).

Figure 3.1 shows the reference model as defined in ISO/IEC 29794–1 [ISO09]; it shows the relation between quality components character and fidelity, where character depends solely on the source and fidelity is affected by acquisition, image processing and feature extraction.

3.3 Quality purpose

The information about sample quality can be applied in several ways in biometric systems. Quality can be assessed on–line after sample acquisition and utilised immediately, or computed off–line on an existing set of earlier acquired samples.

Output of different quality metrics can be combined together in various manners, since using the same quality metrics in different applications is challenging – as described earlier, different systems may benefit from utilising different quality information [ISO09]. Quality can also be useful in a slightly different way depending on the application context: enrolment, verification, identification, survey, etc. This is described in the following sections.

3.3.1 Enrolment

In the enrolment process, if poor quality is detected the sample can be rejected or recaptured. If this repeats, the *Failure To Acquire* or *Failure To Enrol* errors can be declared. Another possibility is template replacement in case a newly acquired sample has quality higher than that of the current template.

This is the case both in supervised and unsupervised systems. In the first case, the quality information can be helpful for the operator; in the latter it would allow an automatic decision.

3.3.2 Verification and Identification

Generally poor quality samples cause the degradation of biometric system performance. Verification and identification systems benefit from poor quality samples rejection since the *False Non-Match Rate* and *False Match Rate* would be lowered [GT07].

Assuring only well-performing samples are used in identification systems would allow to set the comparison threshold higher. This would reduce the *False Match Rate* [GT07].

In negative identification systems, rejecting poor-performing samples would "prevent attempts of defeating detection", although comparing non-mated samples is expected to give low comparison scores both when the samples' quality is poor and good [GT07].

3.3.3 Conditional processing

Quality could be used to differentiate processing of a sample in the following ways [GT07]. Stronger but slower image enhancement algorithms may be used in case of poor quality in the pre-processing stage. Similarly in the comparison phase, stronger but slower comparison algorithms may be used to correctly match poor quality samples. Decision phase could use a different threshold for poor quality samples.

Additionally, fingerprint sample processing without human supervision (lights-out) in forensics is only justified if sample quality is good.

3.3.4 Survey and diagnosis

Quality information may be extracted from samples which had previously been captured [ISO09]. This is useful for survey statistics where scores from e.g. different operational sites are aggregated and compared or analysed to identify trends or anomalies in performance.

Another use of off-line quality information is correlation between quality and different system metrics. This may help in problems diagnosis or show areas where performance could be improved.

3.4 Observed utility calculation

Quality is a performance predictor and utility-based quality component is more predictive in terms of *FNMR*, *FMR*, *FTA* and *FTE* rates than quality based only on character or fidelity. In fact, utility-based quality depends on character and fidelity as well.

Having a dataset of biometric samples and a set of comparison algorithms, it is possible to extract performance-based quality scores (observed utility) for these samples. This performance-based quality will be directly related to the performance in systems that feature the comparison algorithms used. ISO/IEC 29794-1 [ISO09] defines the procedure of calculating these utility values as follows.

1. Gather the reference dataset and comparators.

A reference biometric dataset containing $N_i \geq 2$ samples, $d_i^{(1)}, d_i^{(2)}, \dots, d_i^{(N)}$, for each of M subjects $i = 1, \dots, M$. Each sample is assumed to contain only one biometric characteristic.

A non-empty set of comparators, where V_k is the k -th comparator from all available $k = 1, \dots, K$ comparators.

2. Generate mated comparison scores for each available comparator V_k and for each sample $d_i^{(u)}$ (u th sample of subject i):

$$\begin{aligned} S_{ii} &= \{s_{i,i}^{u,v} \mid s_{i,i}^{u,v} = V_k(d_i^{(u)}, d_i^{(v)})\} \\ u &= 1, \dots, N_i \text{ and } v = u + 1, \dots, N_i \\ i &= 1, \dots, M \end{aligned}$$

3. For each available comparator V_k , generate all the possible non-mated comparison scores for samples from person i with samples from all $j = 1, \dots, M$ and $j \neq i$ other persons:

$$\begin{aligned} S_{ij} &= \{s_{i,j}^{u,v} \mid s_{i,j}^{u,v} = V_k(d_i^{(u)}, d_j^{(v)})\} \\ u &= 1, \dots, N_i \text{ and } v = 1, \dots, N_j \\ i &= 1, \dots, M \text{ and } j = 1, \dots, M \text{ and } i \neq j \end{aligned}$$

4. Compute the mean $m_{i,u}^{mated}$ of sample $d_i^{(u)}$ mated comparison scores:

$$m_{i,u}^{mated} = \frac{\sum_{\substack{v=1 \\ v \neq u}}^{N_i} s_{i,i}^{u,v}}{N_i - 1}$$

5. Compute the mean $m_{i,u}^{non-mated}$ of sample $d_i^{(u)}$ non-mated comparison scores:

$$m_{i,u}^{non-mated} = \frac{\sum_{j=1, j \neq i}^M \sum_{v=1}^{N_j} s_{i,j}^{u,v}}{\sum_{j=1, j \neq i}^M N_j}$$

6. Compute the standard deviation $\sigma_{i,u}^{mated}$ of sample $d_i^{(u)}$ mated comparison scores:

$$\sigma_{i,u}^{mated} = \sqrt{\frac{\sum_{v=1, v \neq u}^{N_i} (s_{i,i}^{u,v} - m_{i,u}^{mated})^2}{N_i - 1}}$$

7. Compute the standard deviation $\sigma_{i,u}^{non-mated}$ of sample $d_i^{(u)}$ non-mated comparison scores:

$$\sigma_{i,u}^{non-mated} = \sqrt{\frac{\sum_{j=1, j \neq i}^M \sum_{v=1}^{N_j} (s_{i,j}^{u,v} - m_{i,u}^{non-mated})^2}{\sum_{j=1, j \neq i}^M N_j}}$$

8. Finally, for each sample $d_i^{(u)}$, compute the target utility as:

$$utility_i^u = \frac{m_{i,u}^{mated} - m_{i,u}^{non-mated}}{\sigma_{i,u}^{mated} + \sigma_{i,u}^{non-mated}}$$

3.5 Quality estimation

In different biometric systems, quality estimation methods differ depending on the biometric characteristic that is used. However, all systems use the common principle of a *Quality Measurement Algorithm (QMA)* producing a scalar quality score q , given an input sample x :

$$quality = QMA(x)$$

Figure 3.2 shows the relation between quality and system performance. Quality expresses the predicted utility, which is in correlation with the observed utility.

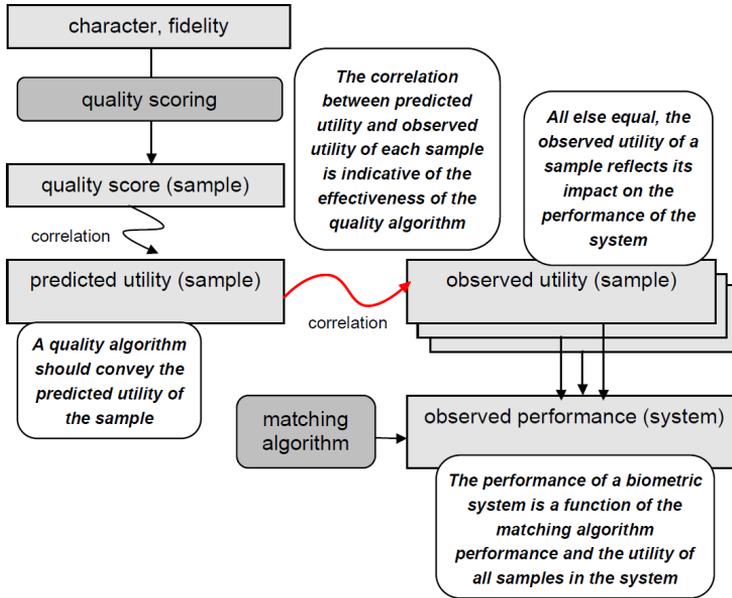


Figure 3.2: Quality, utility and system performance relation [ISO10].

3.5.1 Quality estimation effectiveness

The correlation between quality and observed utility (marked with a red curly arrow in figure 3.2) conveys the performance of the Quality Measurement Algorithm. The better the QMA, the higher the correlation, and hence better prediction of biometric system performance related to the quality of samples used.

3.5.1.1 Spearman correlation

Given a dataset of samples with computed comparison scores and calculated observed utility, one method of measuring the effectiveness of a quality estimation method is to calculate the Spearman's rank correlation coefficient ρ between the quality and the observed utility.

The ρ coefficient will be bigger if the two variables are related monotonically, even if the relation is not linear (which is not a requirement). A positive value indicates that both feature the same monotonicity, whereas negative value indicates opposite monotonicity.

This method is good to indicate whether a certain quality estimation method increases or decreases with utility, but it does not capture their non-monotonic dependency. However, it can also be used to compare two quality methods, i.e. if they show good correlation with utility, but do not correlate between each other, it means they are complementary and could be combined.

3.5.1.2 Error versus Reject Curves

Error vs. Reject Curves (ERC) is a method of QMA evaluation proposed by Grother and Tabassi [GT07] in 2007. As opposed to the correlation coefficient, it can show the non-monotonic dependency between quality and biometric performance. In fact, it shows precisely how rejection of poor quality samples influences the performance of a biometric system, as it plots the False Non-Match Rate against the fraction of samples rejected due to poor quality.

The idea behind *ERC* is to "model the operational case where quality is maintained by reacquisition" [GT07] of poor quality samples. When rejecting and reacquiring such samples until quality is good the performance of a biometric system improves in terms of lowering the False Non-Match Rate. This is because samples of very poor quality give genuine comparison scores in the lowest range – below the biometric threshold – and hence are falsely rejected.

The analysis of *ERC* curves gives a good overview of practical performance benefits. The *ERC* plot can show the behaviour of FNMR depending on the fraction of rejected samples – from zero to one. Rejecting more than one third of the samples (fraction rejected > 0.33) models a situation where on average, more than every third acquisition has to be repeated, probably leading to user frustration. Therefore, the most interesting part of the *ERC* plot is only until fraction rejected is roughly one third, e.g. 35%.

Given a dataset of samples and a quality estimation procedure Q , the procedure of generating an *ERC* curve is as follows ([GT07]):

1. Generate quality scores $q_i = Q(x_i^k)$ for all x_i^k samples where i denotes the subject and k is the acquisition number of the sample.
2. Generate genuine comparison scores $s_{ii}^{(m)}$ for all N pairs of samples (x_i^1, x_i^2) from the same subject
3. For all N pairs of samples (x_i^1, x_i^2) from the same subject take the two quality scores $q_i^{(1)}$ and $q_i^{(2)}$ and to simplify the analysis combine as follows:

$$q_i = H\left(q_i^{(1)}, q_i^{(2)}\right), \text{ e.g. } H = \sqrt{q_1 q_2}.$$

4. Introduce quality threshold u that defines levels of acceptable quality. Define the set of low quality entries as:

$$R(u) = \left\{ j \quad : \quad H(q_i^{(1)}, q_i^{(2)}) < u \right\}$$

5. Using the quantile function (empirical cumulative distribution function) M^{-1} of the genuine comparison scores, determine a biometric threshold t that will result in a reasonable False Non-Match Rate f , e.g. 3% or 10%: $t = M^{-1}(f)$.

6. For each quality threshold qt , determine the set R and calculate:

- set A of samples rejected due to quality below threshold qt ,
- set B of samples left due to quality above threshold qt ,
- *fraction of rejected samples* – the proportion of $|A|$ and the number of all samples,
- set C of samples from B , but that had comparison scores below the biometric threshold t ,
- False Non-Match Rate – the proportion of $|C|$ and $|B|$ – the number of samples with poor comparison scores that should have been rejected due to poor quality:

$$FNMR(t, u) = \frac{|s_{jj} : s_{jj} \leq t, j \notin R(u)|}{|s_{jj} : s_{jj} \leq \infty, j \notin R(u)|}$$

In an ideal case, where the quality is a perfect predictor of performance, picking a biometric threshold to have FNMR of say 10% and rejecting 10% of the lowest quality samples will result in a zero FNMR. This is because all the samples with genuine comparison scores below the threshold would be rejected.

In practice, the quality does not correlate with performance perfectly, but it is a good predictor. Figure 3.3 shows an example ERC plot. The horizontal axis represents the fraction rejected, (from zero to 35%), the vertical axis is the FNMR, starting at zero and reaching 10%.

The ideal case is modelled with a straight green line that goes from FNMR of 10% to zero for fraction rejected of 10%. The remaining two curves, red and blue, represent two QMAs with different performance. The lower curve (red) is lower and closer to the ideal case, the blue curve is higher. Hence, the QMA represented by the red curve performs better.

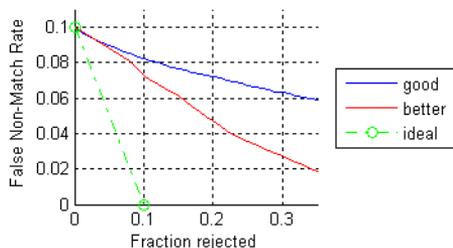


Figure 3.3: Example Error versus Reject curves for two quality measures, blue (good) and red (better). Green line indicates ideal behaviour.

3.6 Fingerprint sample quality

Fingerprint biometric recognition systems use the fingertip skin ridges as the biometric characteristic. Two fingerprint samples are never identical, the captured image depends highly on several factors such as:

- skin defects: blisters, wrinkles, scars, creases,
- condition of the skin: moisture, roughness,
- user behaviour: finger alignment, pressure, elastic deformation,
- environmental conditions: sensor platen cleanness, temperature, lightning.

Some conditions are inherent to the fingerprint source (skin) and cannot be changed, i.e. creases, scars and wrinkles cannot be removed. However, dirtiness, moisture level or user behaviour are not permanent and can be altered to fit the requirements.

Figure 3.4 shows fingerprints acquired from the same finger in varying behaviour (moisture, pressure, cleanliness) conditions and examples of fingerprint samples acquired from fingers with different skin character.

Both behaviour and character can heavily influence the impression. A good character finger with normal moisture level of the skin gives a good sample, as shown in figure 3.4c. Drying the skin and reacquiring results in a very poor impression, shown in figure 3.4b.

If additional moisture or pressure is applied, the samples are also bad, this is shown in figures 3.4e and 3.4a respectively. In fact, it may be difficult to distinguish these cases, as the impressions are very similar.

Dirty skin leads to sample impressions that are very peculiar, such as the one shown in figure 3.4d, where some parts of the foreground ridges are interconnected due to moisture and others are missing due to dirt artefacts present on the skin.

A big difference in fingerprint sample impression is seen due to moisture, pressure and dirtiness variation. However, these factors can be normalized and it is possible to reacquire the fingerprint with an improved quality. Unfortunately, with fingers of bad character, this is not the case and no matter what behaviour, the quality will always be poor if the source is poor.

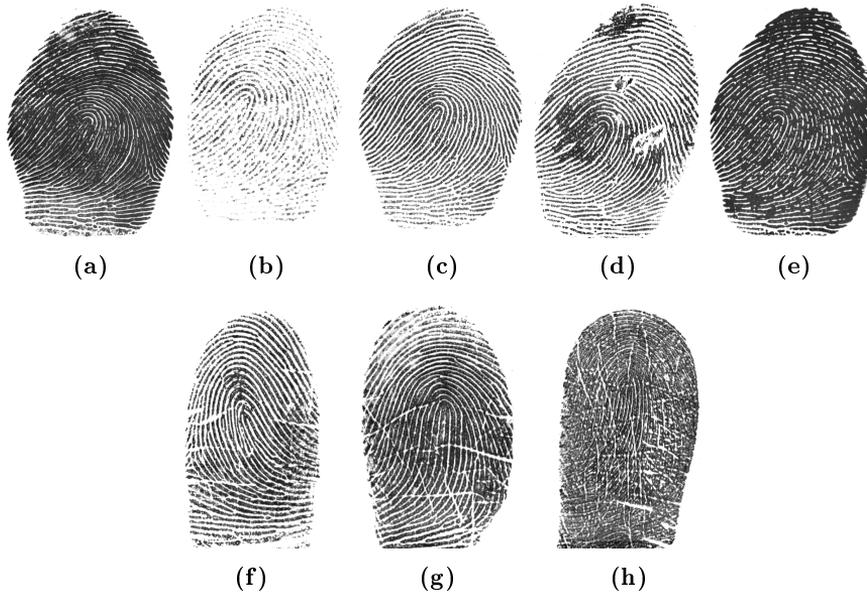


Figure 3.4: Examples of fingerprint samples acquired with an optical scanner from the same finger in different conditions:

- a) too much pressure,
- b) dry skin,
- c) normal conditions,
- d) dirty skin,
- e) wet skin.

Examples of fingerprint samples acquired with an optical scanner from different fingers with varying character:

- f) skin with a few creases,
- g) fingertip with more creases,
- h) very poor fingertip skin

State of the art in fingerprint quality metrics

This chapter describes state of the art algorithms measuring fingerprint sample quality and an algorithm used for fingerprint foreground segmentation. Several methods of fingerprint quality assessment exist [LJY02] [XYYP08] [ISO10] [nfib], the de facto standard is *NIST Finger Image Quality (NFIQ)* [TW05].

4.1 Introduction

As described in section 3.6, fingerprint sample quality varies due to behavioural, environment conditions (moisture, pressure of the finger to sensor platen, etc.) and skin character and conditions. As far as the latter is difficult to improve, the former can be quickly altered so that a sample is recaptured with improved quality to fit the quality requirements.

However, in order to decide whether the quality is poor and recapturing is required, it first has to be measured based on the sample impression. Several methods of measuring fingerprint sample quality exist.

A good overview of existing methods is given in the NFIQ 2.0 Features Defini-

tion document [nfib], with definitions mostly identical to the ISO/IEC 29794–4 [ISO10] specification. A comparative evaluation was performed in NFIQ 2.0 Features Evaluation [nfa] and by Alonso–Fernandez et al. [AFFOG+07], who also introduced a categorization, later followed by ISO/IEC:

- *global* features analysis methods, assessing the sample as a whole and generating the score,
- *local* features analysis methods, analysing smaller portions of the sample individually to generate partial scores and finally aggregating to produce one scalar.
- *classification*, addressing quality assessment as a classification problem (this class is not distinguished by ISO).

The following state of the art fingerprint Quality Measurement Algorithms (QMAs), chosen as best performing accordingly to the NFIQ 2.0 Features Evaluation [nfa] are introduced:

- Orientation Certainty Level, section 4.2,
- Ridge Valley Uniformity, section 4.3,
- Local Clarity Score, section 4.4,
- Gabor Shen, section 4.5,
- NIST Finger Image Quality, section 4.6.

4.2 Orientation Certainty Level

Orientation Certainty Level (OCL) is a method of fingerprint quality estimation first introduced by Lim et al. in 2002 [LJY02], defined in ISO/IEC 29794–4 [ISO10] and described for NFIQ 2.0 Features [nfib].

Fingerprint samples feature dark ridge lines flowing through the image in a certain pattern at varying angles. Local areas of the sample feature a certain orientation of this flow, as shown in figure 4.1.

Good quality samples with well defined ridge flow texture allow to determine this flow with high certainty, whereas poor ridge texture results in uncertain measurement. This certainty is used as a quality indication.

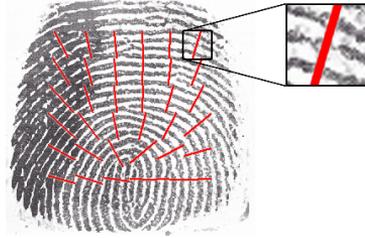


Figure 4.1: Fingerprint sample with ridge orientation marked locally using red lines that are perpendicular to the orientation. Source of the original [ISO10].

4.2.1 OCL calculation procedure

According to the NFIQ 2.0 Features Definition document [nfb], *OCL* is computed as follows. First, the image is divided into blocks of size of 32×32 pixels. For each block, the following steps are performed:

1. Grey level gradient of the image intensity is computed by applying a 3×3 Sobel operator to the sample. The gradient (dx, dy) shows the orientation and strength of the image at the pixel (x, y) .
2. Principal Component Analysis [Pea01] is performed on the gradients of the block to find eigenvalues and eigenvectors. This procedure rotates the data such that "maximum variability is projected onto orthogonal axes". The result is two principal components, first containing largest variance and orthogonal to the ridge orientation, with the corresponding first eigenvalue λ_{max} . The second principal component shows the smallest gradient change and is parallel to ridge orientation, with the second eigenvalue λ_{min} .
3. Covariance matrix is computed as:

$$C = \frac{1}{N} \sum_N \left\{ \begin{bmatrix} dx \\ dy \end{bmatrix} [dx \quad dy] \right\} = \begin{bmatrix} a & c \\ c & b \end{bmatrix}$$

4. From the covariance matrix, the eigenvalues λ are computed as follows:

$$\lambda_{max} = \frac{(a + b) + \sqrt{(a - b)^2 + 4c^2}}{2}$$

$$\lambda_{min} = \frac{(a + b) - \sqrt{(a - b)^2 + 4c^2}}{2}$$

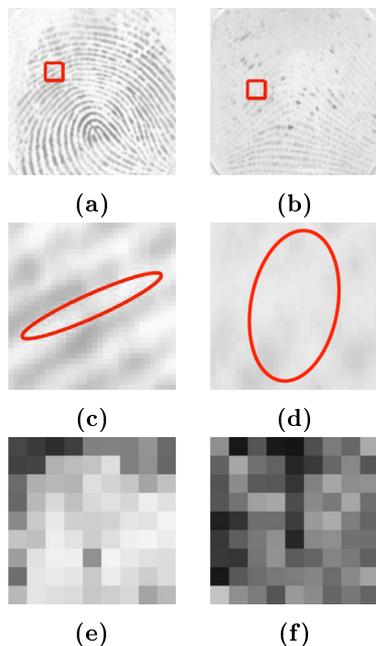


Figure 4.2: *Orientation Certainty level* processing steps for two fingerprint samples of different quality: **a,b**) full sample with marked block, **c,d**) zoomed view of the block with ridge orientation marked as an ellipse with eccentricity corresponding to the *OCL* value of that block. **e,f**) map of *OCL* scores for each block, high intensity corresponds to a high score. Source ISO/IEC 29794-4 [ISO10].

5. The ratio between the eigenvalues is the certainty of orientation:

$$OCL = 1 - \frac{\lambda_{min}}{\lambda_{max}}$$

The final quality of a sample is aggregated by calculating the mean of all blocks' *OCL* values. Figure 4.2 shows an example of the *LCS* quality calculation procedure.

4.3 Ridge Valley Uniformity

Ridge Valley Uniformity (RVU) is defined by ISO/IEC 29794-4 [ISO10] as a quality metric related to the separation of ridges and valleys. The measurement

of quality proposed in the ISO definition is ridge thickness to valley thickness ratio. An assumption is made that a good quality sample should show consistency in this ratio. Therefore small standard deviation of this ratio throughout the image is indicative of good quality.

4.3.1 RVU calculation procedure

The *Ridge Valley Uniformity* QMA is similarly specified in the Quality Feature Definitions for NFIQ 2.0 as the "measure of consistency of the ridge and valley widths" [nffb].

The procedure is defined as follows. First, the image is divided into blocks of size of 32×32 pixels. Then, the following steps are carried out for each block V_0 :

1. Find the orientation angle of ridge lines (as described in section 2.9.1) and create a line perpendicular to this orientation,
2. Rotate the block so that the orientation line is horizontal to obtain V_1 ,
3. From the rotated block V_1 , extract V_2 , a block centred at the orientation line,
4. Compute the average intensity profile $V_3(x) = \frac{\sum_{y=1}^M MV_2(x,y)}{M}$,
5. Apply linear regression to V_3 to determine a threshold DT ,
6. Binarize the image V_3 with the obtained threshold,
7. Find location of background-foreground change, if nothing is found return an empty ratio,
8. Remove incomplete ridges and valleys from the borders of the V_3 block,
9. Return an empty ratio if there were no changes in the last step,
10. Calculate the ratio between ridges and valleys thickness for that block,

The final quality is obtained by calculating standard deviation of all block ratios. Figure 4.3 shows an example of the procedure performed on two samples with different quality.

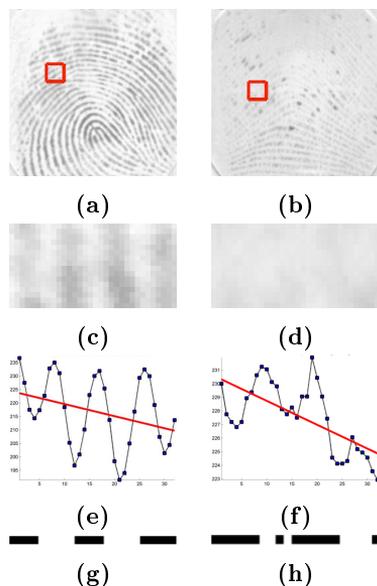


Figure 4.3: *Ridge Valley Uniformity* processing steps for two fingerprint samples of different quality: **a,b)** full sample with marked block, **c,d)** rotated and cropped block, **e,f)** profile of mean intensity with threshold marked in red, **g,h)** cleaned segmentation. Source ISO/IEC 29794-4 [ISO10].

4.4 Local Clarity Score

Local Clarity Score (LCS) operates on the characteristics of ridges and valleys. A normalized ridge and valley width per block is used to assess how many pixels are misclassified as ridge or valley by to the average intensity profile segmentation.

ISO/IEC 29794-4 [ISO10] and the Quality Features Definition of NFIQ 2.0 [nffb] document give a similar definition of *LCS* with a single difference in the final quality score, which is low (ISO) or high (NFIQ 2.0, described below) for good quality samples.

4.4.1 LCS calculation procedure

The image is divided into blocks of size of 32×32 pixels. Then, the following steps are carried out for each block V_0 (first 5 steps are the same as for *RVU*):

1. Find the orientation of ridges (analyse the covariance matrix of the two-dimensional numerical gradient) and create a line perpendicular to this orientation,
2. Rotate the block so that the orientation line is horizontal to obtain V_1 . From the rotated block V_1 , extract V_2 , a block centred at the orientation line,
3. Compute the average intensity profile $V_3(x) = \frac{\sum_{y=1}^M MV_2(x,y)}{M}$ and apply linear regression to determine a threshold DT – see figure 4.4,

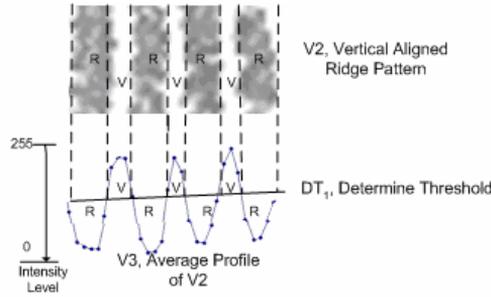


Figure 4.4: Segmentation of V_2 into ridge and valley accordingly to the threshold DT calculated by linear regression of the average intensity profile V_3 .

4. For the block V_2 , count the number of valley pixels ν_T , number of ridge pixels r_T . Count also the number of valley pixels ν_B that are below the threshold DT , and similarly ridge pixels count r_B above that threshold.
5. Determine the proportion of misclassified ridge pixels $\beta = \frac{r_B}{r_T}$ and misclassified valley pixels $\alpha = \frac{\nu_B}{\nu_T}$.
6. Determine the normalized ridge width $\overline{W}_\nu = \frac{W_\nu}{(\frac{S}{125})W^{max}}$ and $\overline{W}_r = \frac{W_r}{(\frac{S}{125})W^{max}}$, where W_r and W_ν are the observed ridge and valley widths; S is the scanner resolution in dpi; and W^{max} is the estimated ridge or valley width a reasonable value of $W^{max} = 5$ for 125 dpi resolution defined by ISO.
7. Determine the minimum and maximum values for the normalized ridge and valley widths as W_r^{nmin} , W_ν^{nmin} , W_r^{nmax} and W_ν^{nmax} and compute the final score:

$$Q_{LCS} = \begin{cases} (1 - (\frac{\alpha + \beta}{2})) * 100, & (W_\nu^{nmin} < \overline{W}_\nu < W_\nu^{nmax}) \\ & \wedge (W_r^{nmin} < \overline{W}_r < W_r^{nmax}) \\ 0, & otherwise \end{cases}$$

The final quality of a sample is aggregated by calculating the mean of all blocks' *LCS* values. Figure 4.5 shows an example of the *LCS* quality calculation procedure.

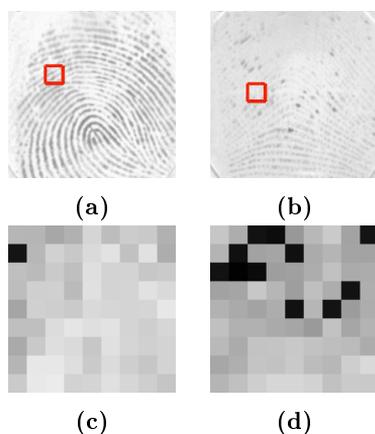


Figure 4.5: Local Clarity Score block wise scores steps for two fingerprint samples of different quality. *a,b*) full fingerprint sample, *c,d*) map of block wise local clarity scores, high intensity is better quality. Source ISO/IEC 29794-4 [ISO10].

4.5 Gabor Shen

First proposed by Shen et al. in 2001 [SKK01], still one of the best performing fingerprint quality estimation methods according to the NFIQ 2.0 Features Evaluation [nfa], where it is named as *Gabor Shen (GSH)*. This method is also defined in the ISO/IEC 29794-4 [ISO10] standard.

Sample quality is estimated by analysing the standard deviation of Gabor filter response at several angles. Image blocks are segmented into foreground and background and separated into two classes – good and bad. Quality is the ratio of poor blocks in the foreground. An example is shown in figure 4.6.

4.5.1 Gabor Shen calculation procedure

The procedure is identical between ISO and NFIQ 2.0 features and performed as follows. The image is divided into $30px \times 30px$ blocks and for each block:

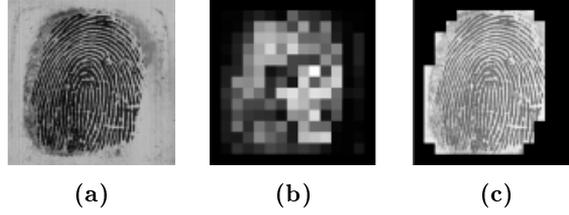


Figure 4.6: *Gabor Shen* quality calculation: **a)** original fingerprint sample, **b)** block wise standard deviation of m Gabor filter responses, **c)** segmentation of image foreground. Source [SKK01].

1. Gabor filters are calculated as:

$$h(x, y, \theta_k, f, \sigma_x, \sigma_y) = \exp\left[-\frac{1}{2}\left(\frac{x_{\theta_k}^2}{\sigma_x^2} + \frac{y_{\theta_k}^2}{\sigma_y^2}\right)\right] \times \exp(i2\pi f x_{\theta_k})$$

- $\theta_k = \pi(k-1)/m$, $k = 1, \dots, m$,
 - $f = 0.12$ is the sinusoidal plane wave frequency,
 - $m = 8$ is the number of orientations,
 - $\sigma_x = \sigma_y = 4$ are the standard deviations of the Gaussian envelope along x and y axes respectively,
 - $x_{\theta_k} = x \cos \theta_k + y \sin \theta_k$ and $y_{\theta_k} = x \sin \theta_k - y \cos \theta_k$.
2. For each computed filter, Gabor filter responses centred at (X, Y) are computed for the block. The standard deviation of the responses is calculated as follows:

$$G = \sqrt{\left(\frac{1}{m-1} \sum_{k=1}^m (g_{\theta_k} - \bar{g}_{\theta})^2\right)}, \quad \bar{g}_{\theta} = \frac{1}{m} \sum_{k=1}^m g_{\theta_k}$$

3. The block is in the set of foreground blocks V_F if the standard deviation is over the segmentation threshold $G > T_b$ where $T_b = 1$,
4. The block is in the set of poor quality foreground blocks V_P if the standard deviation is also below the quality threshold $G > T_b \wedge G < T_q$ where $T_q = 2$,

Finally the quality is the ratio between the size of the V_F and V_P sets:

$$Q_{GABORSHEN} = 1 - \frac{|V_F|}{|V_P|}$$

4.6 NIST Finger Image Quality

Proposed by Tabassi, Wilson and Watson in 2004, *NIST Finger Image Quality (NFIQ)* [TW05] is currently the most popular approach to quality estimation according to Maltoni et al. [MMJP09]. It is also best performing method according to the NFIQ 2.0 Features Evaluation [nfa].

It gives a good performance prediction for diverse comparison algorithms and it performs the quality estimation fast enough to be incorporated in a live fingerprint scanner [TW05] [MMJP09] [nfa].

This method is different from QMAs described earlier as it classifies of a fingerprint sample into five quality levels: excellent, very good, good, fair, poor.

Several features of an image are analysed to produce a feature vector, which is converted to a quality score using an artificial neural network classifier [TW05]. The image features used for NFIQ are extracted using the *NIST Biometric Image Software (NBIS)* [nfi], which includes several general purpose image utilities and a minutiae detector called MINDTCT.

4.6.1 NFIQ calculation procedure

The quality estimation procedure is as follows. First, an image quality map is generated based on inconsistency in the following characteristics: local ridge orientation, local contrast and local curvature.

Then, minutiae are extracted and each one is associated with a quality value accordingly to the measure of mean and standard deviation of pixel intensity in its closest neighbourhood.

Finally, each minutia is given a new quality value calculated as a composition of the old value and a value extracted from the quality map at the minutia location.

The next step is generating the feature vector. This is performed using the minutiae–quality pairs and the image quality map. Table 4.1 describes the 11 features that compose that vector.

The final quality score generation is a classification performed with a neural network. NFIQ was trained on a subset of images extracted from 5 very different

1	number of blocks with quality 1 or more
2	total number of found minutia
3	no. of minutia with quality 0.5 or more
4	no. of minutia with quality 0.6 or more
5	no. of minutia with quality 0.75 or more
6	no. of minutia with quality 0.8 or more
7	no. of minutia with quality 0.9 or more
8	percentage of quality map foreground blocks with quality 1
9	percentage of quality map foreground blocks with quality 2
10	percentage of quality map foreground blocks with quality 3
11	percentage of quality map foreground blocks with quality 4

Table 4.1: NFIQ feature vector components [TW05] [MMJP09]

datasets. This assured different acquisition conditions and a big variation of quality.

4.6.2 NFIQ 2.0 development

A second version of the *NIST Finger Image Quality* algorithm – NFIQ 2.0 [oST] – is currently being developed collaboratively by several institutions in the US and Germany:

- NIST [NIS],
- BSI [BSI],
- BKA [BKA],
- Fraunhofer IGD [Fra],
- Hochschule Darmstadt [hda] – CASED [CASa],
- secunet Security Networks AG [sec].

4.7 Image segmentation

Fingerprint sample images acquired with optical sensors (see section 2.7) contain dark lines (representing fingertip skin ridges) on a bright (usually white) background.

How much of the whole sample is covered with these dark lines depends on finger size, user behavior and sensor platen size. If the finger is small or only a small part of the fingertip touches the sensor, then only part of the sample will contain ridges. Similarly, if a sensor uses a platen that is bigger than the fingertip, the ridges will cover only a portion of the sample.

The background of a fingerprint may not be plainly white but may contain some noise from dirt or latent fingerprint ridges. In order to improve the performance of fingerprint sample comparison, the valid fingerprint region has to be separated from the background.

This way, the comparison procedure will only work on the foreground [MMJP09]. Similarly, for quality value to better predict the comparing performance, it should only assess the foreground.

The process of extracting the foreground region is called segmentation. Several fingerprint segmentation methods were proposed over time [MMJP09], one of them was described in section 4.5 where Gabor filter was used for sample quality estimation – and the response of the filter was also used for segmentation.

4.7.1 Segmentation procedure

Another method, which does not require the computationally expensive Gabor filter response, is based on variance. Fingerprint foreground region shows high variance because ridges are represented by dark lines and valleys by bright lines. Background of a fingerprint sample has lower variance, unless extreme noise is present which is usually not the case.

Therefore it is possible to segment the image based on the local variance. This is performed as follows:

1. divide the sample image into k blocks of size 32×32 pixels,
2. for each block, compute the variance $V(k)$ as:

$$V(k) = \frac{1}{W^2} \sum_{i=0}^{W-1} \sum_{j=0}^{W-1} (I(i, j) - M(k))^2$$

where $I(i, j)$ is the grey level value at pixel (i, j) and $M(k)$ is the mean of grey level of the block k . [Meh93]

3. mark blocks as foreground if their computed variance is over a threshold $V(k) > 0.1$

Figure 4.7 shows example fingerprint samples with segmentation mask superimposed in red. Segmentation is shown for both block sizes 32×32 and 8×8 pixels.

Samples with well defined ridges and a clear white background are easy to segment correctly. Dry impression (figure 4.7d) causes some foreground to be falsely segmented as background.

Segmentation of a sample (figure 4.7c) with a noisy background (older generation optical sensor) results in a poor segmentation where background is marked as foreground.

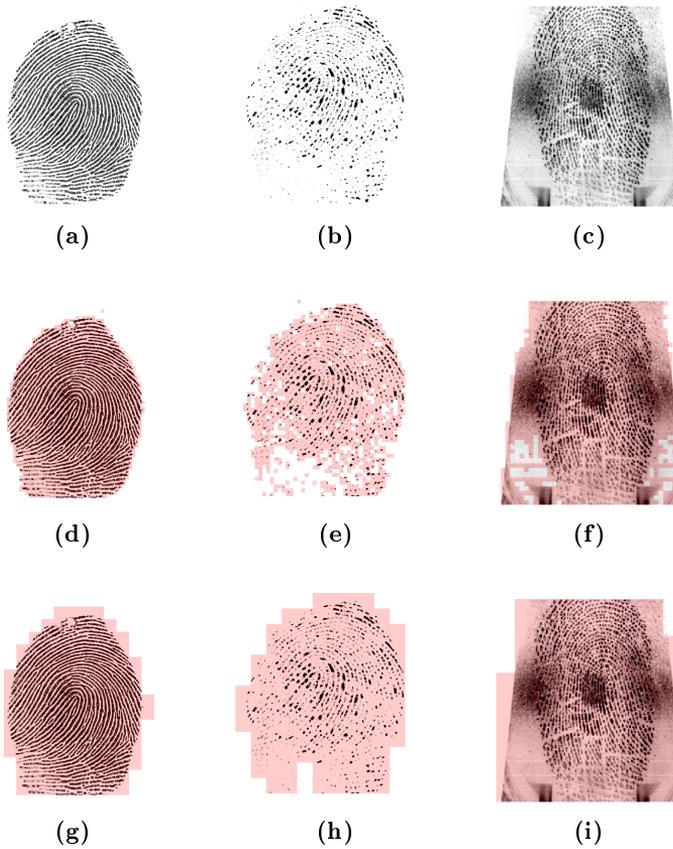


Figure 4.7: Block variance based segmentation procedure:

- a) b) c) Sample before segmentation,
- d) e) f) Segmentation with 8×8 px block (superimposed in red),
- g) h) i) Segmentation with 32×32 px block (red).

Proposed Fingerprint Analysis Methods

This chapter describes the following proposed fingerprint analysis methods:

- Fingerprint sample Quality Measurement Algorithms:
 - Ridge Valley Difference (section 5.2),
 - Ridge Line Count (section 5.1),
 - Contrast (section 5.3).
- Moisture Indication method (section 5.4).

The proposed Quality Measurement Algorithms perform fingerprint sample quality estimation. Given a grey level fingerprint sample image, they produce a quality scalar. The function of these methods is identical as all state of the art QMAs described in chapter 4 – the output quality values are performance predictors of a biometric system working with the analysed fingerprint samples.

The proposed Moisture Indication method performs an analysis of a fingerprint sample and similarly produces a scalar score given a grey level fingerprint image. However, the scalar produced does not convey quality, it expresses skin moisture of the fingertip that touched the sensor platen to acquire the analysed sample.

5.1 Ridge Line Count (RLC)

The first proposed method is based on the count of the dark ridge lines locally passing through each block of a fingerprint sample. The motivation for such measure is that fingerprints with good quality should have a constant number of ridges in the range 3–4, as empirical analysis shows. With the degradation of quality, the local ridges count is expected to fall as the detection will be more difficult on poor quality samples.

The Ridge Line Count Quality Measurement Algorithm analyses the count of ridges for blocks of size 32×32 pixels. The count of ridges is extracted from an edge map of the block. This is performed by counting the edge pixels passing through the middle of a block that is first rotated so that the ridge lines are vertical.

The edges are found using a Laplacian of Gaussian (LoG) convolution filter. This filter applies a Gaussian blur and the Laplacian filter and marks edges where zero crossings occur, i.e. value changing from positive to negative or vice versa.

The LoG edges finding method is best for counting ridge lines in initial tests. The reason for this may be that it has low-pass filtering integrated which reduces noise that otherwise may lead to spurious edges.

5.1.1 Block-wise score aggregation

The Ridge Line Count quality estimation method produces a block-wise score map. To give a final quality score scalar, this map must be aggregated. Three methods are proposed for this aggregation, constituting three QMAs based on Ridge Line Count:

- Ridge Line Count Mean (RLCM),
- Ridge Line Count Number of Good blocks (RLCNG),
- Ridge Line Count Entropy (RLCE),

5.1.2 Ridge Line Count calculation procedure

The procedure of quality estimation using the Ridge Line Count QMA is described below. Figure 5.2 shows example processing on three fingerprint samples with varying quality; the respective quality scores per sample after aggregation with all proposed methods are shown in table 5.1.

1. using blocks of size 32×32 pixels determine the segmentation mask as described in section 4.7.1,
2. convert the greyscale sample image into a binary image, based on a Otsu's threshold [Ots79]¹
3. divide the image into blocks of size 32×32 pixels, for each block marked as foreground:
 - (a) determine the orientation of the block:
 - calculate the grey level gradient of the image,
 - find covariance matrix coefficients of the grey level gradients by performing Principal Component Analysis [Pea01],
 - the covariance matrix is computed as:

$$C = \frac{1}{N} \sum_N \left\{ \begin{bmatrix} dx \\ dy \end{bmatrix} [dx \quad dy] \right\} = \begin{bmatrix} a & c \\ c & b \end{bmatrix}$$

- the orientation angle (in radians) of the block is computed as:

$$angle = \tan^{-1} \left(\frac{c}{\sqrt{c^2 + (a-b)^2}} \frac{a-b}{\sqrt{c^2 + (a-b)^2}} \right)$$

- (b) rotate the block so that the ridges are vertical, crop the block so it does not have invalid regions,
- (c) find edges of the block by finding zero crossings after applying the Laplacian of Gaussian filter shown in figure 5.1 calculated with the following equation:

$$LoG(x, y) = -\frac{1}{\pi\sigma^4} \left[1 - \frac{x^2 + y^2}{2\sigma^2} \right] e^{-\frac{x^2 + y^2}{2\sigma^2}}$$

where $LoG(x, y)$ is the response at pixel (x, y) and $\sigma = 2$.

¹The binarization with Otsu's threshold is chosen to minimize the intra class variance of the black and white pixels, as this has proven to improve the performance of the RLC method in initial tests.

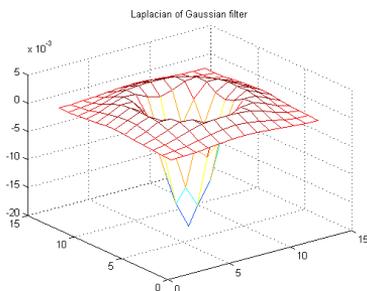


Figure 5.1: Laplacian of Gaussian filter used for edge detection.

- (d) count edge (white) pixels that go through the middle of the rotated image. These are pixels crossing a horizontal line as shown in figure 5.2h,
 - (e) divide the number of edges by 2 to get the number of ridge lines. It is accepted that an image could have half of a ridge, no edges are discarded.
4. From blocks that are marked as foreground with the segmentation mask, aggregate scores via one of the following methods to compute the final quality via one of the following constituted Quality Measurement Algorithms:
 - *RLCM* – mean of all foreground block values
 - *RLCNG* – count of blocks with score above empirically chosen threshold $t = 2.9$
 - *RLCE* – block wise score entropy calculated as:

$$e = abs\left(\sum_{i,j=0}^{i<X,j<Y} S(i,j) * \log_2(S(i,j))\right)$$

where $S(i, j)$ is the block score at block (i, j) , X and Y is the horizontal and vertical sizes of the block wise map respectively.

5.2 Ridge Valley Difference (RVD)

The second proposed method is based on the local differences between the proportion of ridges and valleys of a fingerprint sample. The idea behind this QMA

	RLCM	RLCNG	RLCE
Sample 5.2a	1.9263	0.19658	229.1521
Sample 5.2b	2.8068	0.48485	585.0143
Sample 5.2c	2.0182	0.22667	335.3489

Table 5.1: Fingerprint sample quality scores assigned to three fingerprints from figure 5.2 using Ridge Line Count QMA with different aggregation methods: mean, number of good elements and entropy.

is similar to that of the Ridge Valley Uniformity QMA described in section 4.3 and as follows. Fingerprint samples of best quality feature ridges and valleys of similar thickness. With the degradation of sample quality, the ratio between ridges and valleys is expected to change.

This method analyses the ridge and valley thickness by calculating the difference of black and white pixel proportions in each image block. Since a fingerprint sample is a grey level image, in order to correctly classify the pixels as black or white, a threshold is required. In order to assess different samples with equal conditions, a static binarization threshold $t = 0.5$ is used (for images with luminance in the range 0 to 1).

The difference between white and black pixels ratio may give negative values, hence an absolute value is used to avoid this. The result is subtracted from 1. This way, if the ratio of both white and black pixels is equal, the quality score assigned to the block will be 1 (highest), and if e.g. the block is only black pixels, the quality will be zero (lowest).

5.2.1 Block-wise score aggregation

The Ridge Valley Difference quality estimation method produces a block-wise score map. To give a final quality score scalar, this map must be aggregated. Three methods are proposed for this aggregation, constituting three QMAs based on Ridge Line Count:

- Ridge Valley Difference Mean (RVDM),
- Ridge Valley Difference Number of Good blocks (RVDNG),
- Ridge Valley Difference Entropy (RVDE),

5.2.2 RVD calculation procedure

The procedure of quality estimation using the Ridge Valley Difference QMA is described below. Figure 5.3 shows example processing on three fingerprint samples with varying quality; the respective quality scores per sample after aggregation with all proposed methods are shown in table 5.2.

1. Using blocks of size 32×32 pixels determine the segmentation mask as described in section 4.7.1,
2. convert the greyscale sample image into a binary image based on a threshold of $t = 0.5$,
3. divide the image into blocks of size 32×32 pixels, for each block marked as foreground:

(a) count the proportion of white pixels W in the block as:

$$W = \frac{\sum_{k=0, l=0}^{k<32, l<32} S(k, l)[S(k, l) = 1]}{|S|}$$

where S is the block of size 32×32 pixels, $S(k, l)$ is the value at pixel (k, l) and $|S|$ is the count of all pixels in the block ($32^2 = 1024$),

(b) similarly, count the proportion of black pixels B in the block as:

$$B = \frac{\sum_{k=0, l=0}^{k<32, l<32} S(k, l)[S(k, l) \neq 1]}{|S|}$$

where S is the block of size 32×32 pixels, $S(k, l)$ is the value at pixel (k, l) and $|S|$ is the count of all pixels in the block ($32^2 = 1024$),

(c) calculate the quality score Q of that block as one minus the absolute difference of the white and black pixel proportions:

$$Q = 1 - \text{abs}(B - W)$$

4. From blocks that are marked as foreground with the segmentation mask, aggregate scores via one of the following methods to compute the final quality via one the following constituted Quality Measurement Algorithms:

- *RVDM* – mean of all foreground block values

- *RVDNG* – count of blocks with score above an empirically chosen threshold $t = 0.45$
- *RVDE* – block wise score entropy calculated as:

$$e = abs\left(\sum_{i,j=0}^{i<X,j<Y} S(i,j) * \log_2(S(i,j))\right)$$

where $S(i, j)$ is the block score at block (i, j) , X and Y is the horizontal and vertical sizes of the block wise map respectively.

	RVDM	RVDNG	RVDE
Sample 5.3a	0.11066	0.008547	28.8191
Sample 5.3b	2.8068	0.6303	40.285
Sample 5.3c	0.63961	0.35333	62.6214

Table 5.2: Fingerprint sample quality scores assigned to three fingerprints from figure 5.3 using Ridge Valley Difference QMA with different aggregation methods: mean, number of good elements and entropy.

5.3 Contrast (CNT)

The third proposed Quality Measurement Algorithm is based on the measure of luminance contrast.

In a fingerprint sample ridges form a pattern of dark ridges and bright valleys. The contrast is measured using Michelson’s method, which measures the relation between the spread and the sum of bright and dark luminance values. This should be useful since for good quality samples both bright and dark features cover similar fractions of the sample foreground. With quality degradation the value is expected to fall.

5.3.1 Block-wise score aggregation

The Contrast quality estimation method produces a block-wise score map. To give a final quality score scalar, this map must be aggregated. Three methods are proposed for this aggregation, constituting three QMAs based on Ridge Line Count:

- Contrast Mean (RVDM),
- Contrast Number of Good blocks (RVDNG),
- Contrast Entropy (RVDE),

5.3.2 CNT calculation procedure

The procedure of quality estimation using the Contrast QMA is described below. Figure 5.4 shows example processing on three fingerprint samples with varying moisture level of the fingertip skin during acquisition; quality scores per sample after aggregation with all proposed methods are shown in table 5.4.

1. Using blocks of size 8×8 pixels determine the segmentation mask as described in section 4.7.1,
2. divide the image into blocks of size 8×8 pixels, for each block marked as foreground:
 - (a) calculate the Michelson contrast c of the block:

$$c = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}$$

where I_{max} and I_{min} represent the highest and lowest luminance

- (b) the quality Q indication of the block is equal to the calculated contrast $Q = c$.
3. From blocks that are marked as foreground with the segmentation mask, aggregate scores via one of the following methods to compute the final quality via one the following constituted Quality Measurement Algorithms:
 - *CNTM* – mean of all foreground block values
 - *CNTNG* – count of blocks with score above an empirically chosen threshold $t = 0.6$
 - *CNTE* – block wise score entropy calculated as:

$$e = abs\left(\sum_{i,j=0}^{i<X,j<Y} S(i,j) * \log_2(S(i,j))\right)$$

where $S(i, j)$ is the block score at block (i, j) , X and Y is the horizontal and vertical sizes of the block wise map respectively.

	CNTM	CNTNG	CNTE
Sample 5.4a	0.39257	0.0451	39449.7755
Sample 5.4b	0.60399	0.46221	55778.8541
Sample 5.4c	0.66025	0.67849	49944.2721

Table 5.3: Fingerprint sample quality scores assigned to three fingerprints from figure 5.4 using Contrast QMA with different aggregation methods: mean, number of good elements and entropy.

5.4 Moisture Indication method

The fourth proposed method, Moisture Indication, does not compute the fingerprint sample quality as performance predictor. Instead, it relates the fingerprint sample to the moisture of the fingertip skin presented to the sensor platen during acquisition.

The motivation behind such a method is to provide feedback to users, in case if the fingertip skin moisture is extremely low or extremely high, possibly leading to quality degradation. If the quality of an acquired sample is indeed measured as very poor, reacquisition may be necessary. However, if the cause behind poor quality is extreme skin moisture, without proper feedback the reacquisition will lead to the same poor quality again, because the conditions are unchanged.

Providing proper feedback allows the user to immediately alter the moisture level of the fingertip skin and hence increases the probability of quality improvement upon sample reacquisition.

Empirical analysis of fingerprint samples acquired from fingertip skin with different moisture level shows that dry fingers produce less contact with sensor platen, leading to broken ridges or even missing parts of a fingerprint. Similarly, wet fingers give an opposite effect where the sample looks very dark. An example of this is shown in figures 5.5a and 5.5c showing samples acquired from the same finger when it was dry and wet respectively.

The Moisture Indication method relates this impression to the moisture of the fingertip skin. This is performed similarly as for the Ridge Valley Difference QMA described in section 5.2. The local ratios of white pixels are found for blocks of a fingerprint sample. The indication is the mean of these block-wise values. This way, dry skin gives high indication and wet skin gives low indication. Fingerprints acquired from normal conditions skin are expected to give indications in the middle range.

5.4.1 MI calculation procedure

The procedure of fingerprint skin moisture estimation using the Moisture Indication fingerprint analysis method is described below. Figure 5.5 shows example processing on three fingerprint samples acquired from fingers with different skin moisture.

1. Using blocks of size 32×32 pixels determine the segmentation mask as described in section 4.7.1,
2. convert the greyscale sample image into a binary image based on a threshold chosen to minimize the intra-class variance of the black and white pixels (Otsu's method [Ots79]),
3. divide the image into blocks of size 32×32 pixels, for each block marked as foreground:
 - (a) calculate the moisture indication score MI of that block as the proportion of white pixels in the block:

$$MI = \frac{\sum_{k=0, l=0}^{k<32, l<32} S(k, l)[S(k, l) = 1]}{|S|}$$

where S is the block of size 32×32 pixels, $S(k, l)$ is the value at pixel (k, l) and $|S|$ is the count of all pixels in the block ($32^2 = 1024$).

4. From blocks that are marked as foreground with the segmentation mask, aggregate scores via mean of all block-wise scores to compute the final Moisture Indication scalar score of that sample.

	Moisture Indication	skin moisture impression
Sample 5.4a	0.945	dry
Sample 5.4b	0.668	normal
Sample 5.4c	0.306	wet

Table 5.4: Fingerprint sample Moisture Indication scores of three fingerprints from figure 5.5.

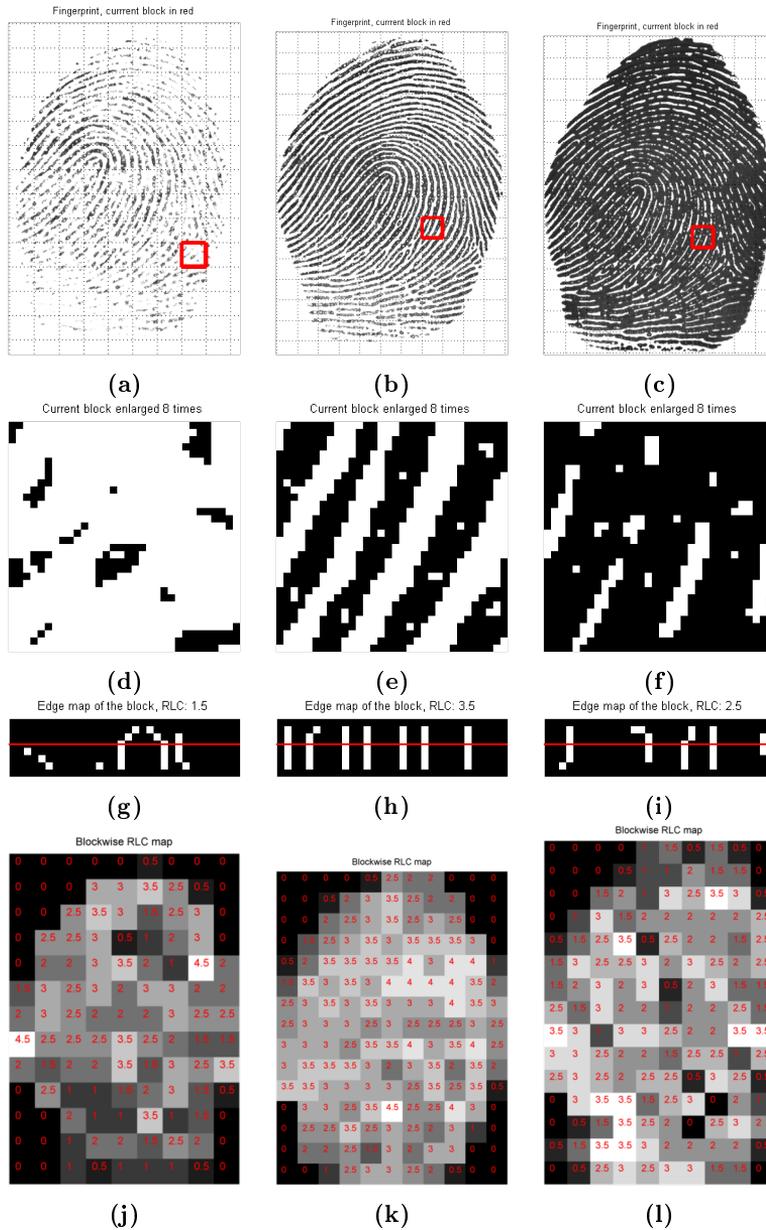


Figure 5.2: Ridge Line Count quality calculation steps for different fingerprint samples:

- a) b) c) Fingerprint sample with analysed block marked in red,
- d) e) f) Analysed block,
- g) h) i) Edges extracted from rotated and cropped block, red line shows which white pixels are counted,
- j) k) l) Block wise RLC map.

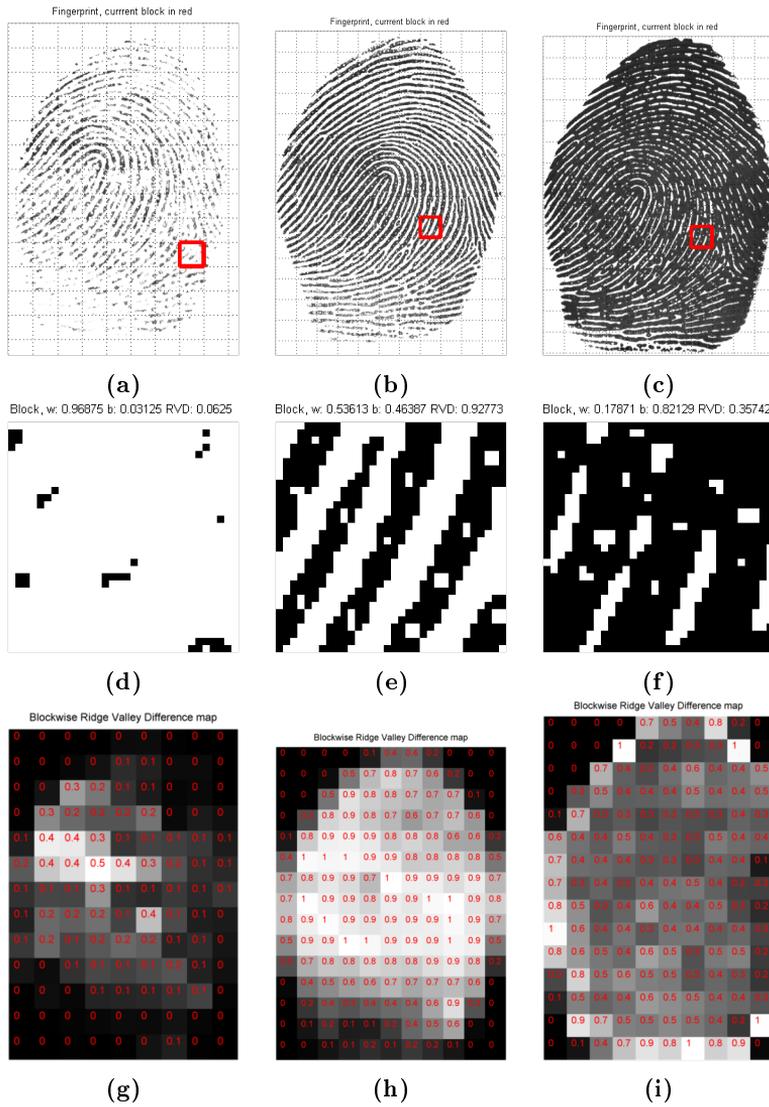


Figure 5.3: Ridge Valley Difference quality calculation steps for different fingerprint samples:

- a) b) c) Fingerprint sample with analysed block marked in red,
- d) e) f) Analysed block,
- g) h) i) Block wise RVD map.

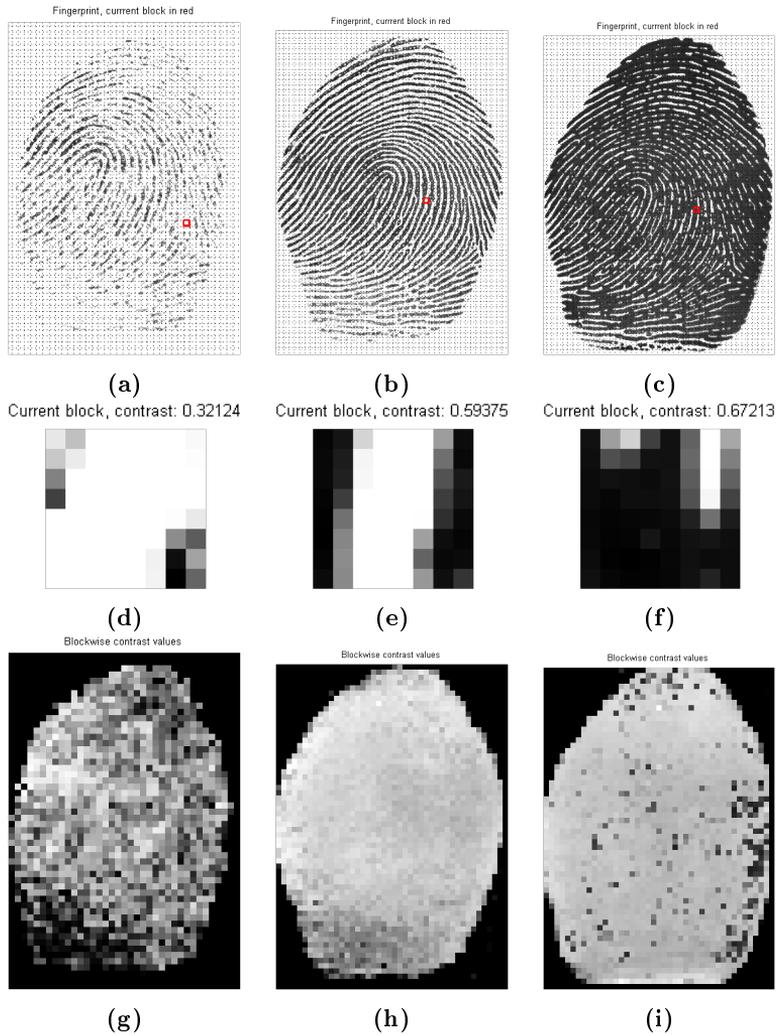


Figure 5.4: Contrast quality calculation steps for different fingerprint samples:
a) b) c) Fingerprint sample with analysed block marked in red,
d) e) f) Analysed block,
g) h) i) Block wise CNT map.

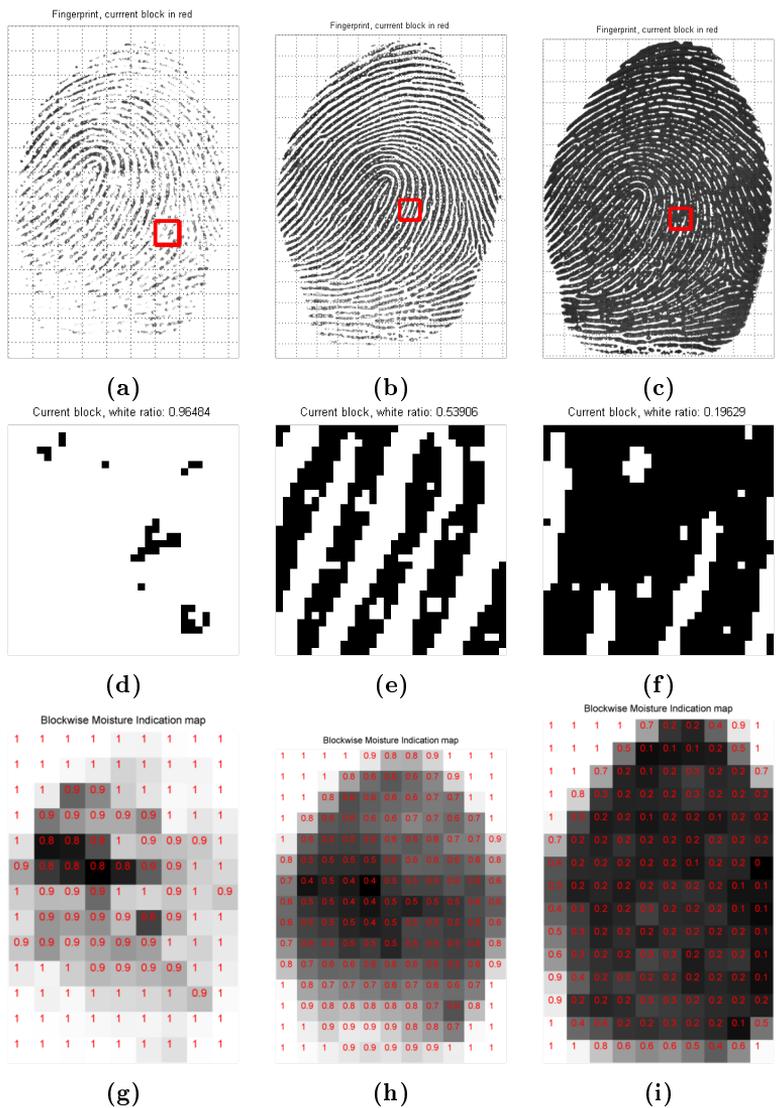


Figure 5.5: Moisture Indication score calculation steps for different fingerprint samples:

- a) b) c) Fingerprint sample with analysed block marked in red,
 - d) e) f) Analysed block,
 - g) h) i) Block wise Moisture–Indication map.
- a) dry finger, b) normal skin moisture, c) wet finger.

CHAPTER 6

Experimental setup

This chapter describes the experiment carried out for this project. Section 6.1 describes the project purpose and motivation behind it. Target biometric fingerprint sensor technology is chosen in section 6.2. The procedure of the experiment is briefly introduced in section 6.3, this discusses preparation, calculation, QMA assessment and moisture impact analysis.

Fingerprint sample dataset choice and collecting new dataset are described in sections 6.4 and 6.5. Calculation of quality values, comparison scores, and utility values is described in section 6.6. Performance assessment of the proposed Quality Measurement Algorithms (see chapter 5) is described in section 6.7.

Finally, section 6.8 introduces the analysis of the skin moisture impact on fingerprint sample quality and section 6.8.2 describes the procedure of performance assessment of the Wetness and Dryness Indication method introduced in chapter 5.

6.1 Project purpose and motivation

The main purpose of this project is to propose new methods of fingerprint quality estimation. The goal is to introduce methods that would be competitive with the current state of the art, so that they could possibly be incorporated in ISO/IEC 29794-4 [ISO10] and NFIQ 2.0 [oST], the second version of NFIQ, which is the de facto standard.

Of additional interest is assessment whether extreme skin moisture indeed causes biometric performance degradation [SKK01] [AFFOG+07] [MMJP09] and finding a method that will allow to indicate such extreme moisture based on the acquired fingerprint sample.

The motivation for the second part is such that in case of poor sample quality due to extreme moisture, reacquisition with unchanged conditions is likely to result in the same impression – over and over again, leading to user frustration. Indicating extreme moisture levels would allow the user to immediately fix the conditions to meet requirements (e.g. wiping the finger with cloth if it is too moist), so that the quality will improve immediately.

6.2 Target sensor technology

As described in section 2.7, several device types can be used in order to capture a biometric fingerprint sample. Examples given in figure 2.3 show that samples differ significantly depending on the technology used for acquisition.

This project focuses on optical sensors, although it is certainly interesting to extend to several sensor technologies. This decision is supported as follows:

- Optical sensors are oldest and most commonly used, offer highest resolution and feature largest platens, allowing to capture four-finger slap impressions as well as single finger samples. [MMJP09] [der] [cro].
- Some optical fingerprint systems are harder to spoof as they feature liveness detection methods [lum07], e.g. Dermalog LF10 [der].
- The Unique Identification Number (UID) project (Aadhaar) [aad] of Unique Identification Authority of India (UIDAI) defines standards for fingerprint sensors [uid] with a recommendation that used biometric devices are certified by Indian Standardization Testing and Quality Certification Direc-

torate (STQC) (most of the certified devices use the optical technology) [stq].

- The US-VISIT program of Department of Homeland Security explicitly defines an optical scanner to be used for the project – a CrossMatch Guardian FW [usv] [cro].

6.3 Procedure of the experiment

The experiment procedure is divided into three parts:

1. *Preparation* of data and algorithms (sections 6.4 and 6.5):
 - fingerprint sample datasets (at least one with moisture information),
 - at least one fingerprint comparator (black box or algorithm),
 - at least one state of the art fingerprint reference QMA.
2. *Calculation* of scores by running algorithms on the data (section 6.6):
 - genuine and impostor comparison scores (via all comparators),
 - observed utility scores of each fingerprint (per comparator),
 - quality indication of each fingerprint sample.
3. *Assessment* of proposed QMAs performance (sections 6.7):

Performance of each proposed Quality Measurement Algorithm is to be analyzed in terms of:

 - improvement in FNMR with poor samples rejection – section 6.7.3,
 - correlation of quality scores with observed utility – section 6.7.2,
 - average execution time – section 6.7.1.
4. *Analysis* of moisture impact on comparison scores and utility (section 6.8).

Assessment of Moisture Indication method performance: (section 6.8.2):

 - correlation of wetness indication with measured moisture,
 - error of threshold based binary labelling of samples as dry or wet,
 - average execution time.

6.4 Existing fingerprint dataset choice

The main part of the project, proposing good Fingerprint Quality Measurement Algorithms, requires a dataset of fingerprint samples.

In order to ensure fair assessment of proposed methods and a good comparison of performance with the reference state of the art, the dataset should feature a significant amount of fingerprints with varying quality of samples, possibly captured from a big number of subjects with varying background (physical or office workers, etc.). This would ensure not only the fidelity component of quality is varied but also the character. As the target technology is optical sensors, the comparison and assessment should be performed on a dataset with samples acquired using such sensors.

Several biometric databases are available in the *Center for Advanced Security Research Darmstadt (CASED)*, including datasets of many biometric traits. Following fingerprint sample datasets are available:

- BIOSECURE multimodal datasets [FaOgTtGr07] of the BioSecure Foundation [bio]. Two multimodal datasets with fingerprint sample subsets (indoor and outdoor) acquired in 2 sessions each, from 667 and 713 subjects respectively. Samples acquired using several sensors, unfortunately mixed such that extraction of images acquired with optical sensors is difficult.
- SD29 dataset of NIST [NIS], containing scans of fingerprints acquired using off-line ink and paper method (does not include on-line acquired images).
- Ministerio de Ciencia y Tecnología (MCYT) [OGFAS+03] database subset with fingerprints, containing samples acquired using capacitive and optical technology from 330 subjects in several sessions resulting in nearly 40 thousand samples. Offers highest sample count but is not publicly available.
- Fingerprint Verification Competition datasets (years 2000, 2002, 2004) – available publicly on-line, offer good quality degradation (also moisture based in 2004). However, small subsets of samples around 880 samples for years 2000 and 2002, 1440 for 2004 would require additional aggregation of results. These datasets include samples from one finger per person. Moisture based quality degradation is not indicated – manual assessment is needed to know which samples are acquired from moist fingers.
- CASIA Fingerprint Image Database Version 5.0 (or CASIA-FingerprintV5) [casb], big database available publicly on-line. Developed in the Institute

of Automation of the Chinese Academy of Sciences, contains samples captured with DigitalPersona URU4000 optical fingerprint sensor (500ppi resolution, size 328×356 pixels) from 500 subjects with different background, i.e. students, workers, waiters, etc. Five images per finger, from 8 fingers (without the small fingers) – 40 samples per subject, 20000 samples in total, acquired with varying rotation and pressure.

Both MCYT and CASIA Fingerprint V5 datasets offer a very big number of samples, significantly higher than other datasets. However, CASIA has images captured from more subjects than MCYT, and is widely available – the choice of a public dataset allows anyone to compare their results easily without registering or paying for access.



Figure 6.1: Example fingerprint samples from the CASIA FP V5 database.

CASIA Fingerprint Version 5.0 offers good character based quality degradation due to a significant variation in subjects background and behaviour (finger rotation and pressure variation). Figure 6.1 shows example fingerprints from this dataset. Additionally, such a big dataset with big variation of quality, acquired from different subjects definitely supports a fair comparison of proposed fingerprint Quality Measurement Algorithms with state of the art methods.

6.5 New fingerprint dataset collection

The second part of this project is assessing whether extreme finger moisture leads to utility degradation of fingerprints acquired from that finger and indicating finger skin moisture based on the impression of the acquired sample.

This requires a fingerprint dataset with ground truth information about the moisture of the fingertip during fingerprint acquisition. No available dataset offers such moisture information. Manual off-line mark-up of samples could be performed on one of the existing datasets, and this is usually the case – a sample is either marked as dry, normal or wet [SKK01] [LJY02] [XYYP08] [AFFOG+07] [MMJP09] by a person.

However, off-line assessment may actually not correlate with the skin moisture of the finger presented during acquisition. In fact, manual mark-up does not indicate whether it was indeed moisture that caused the impression, or for example high force applied to the sensor platen – this is shown in figure 6.2, where a sample acquired from a wet finger (6.2c) looks similar to the sample acquired with pressure (6.2d) to an untrained eye. This project covers moisture variation, though it is certainly interesting to investigate pressure and moisture–pressure variation as well.

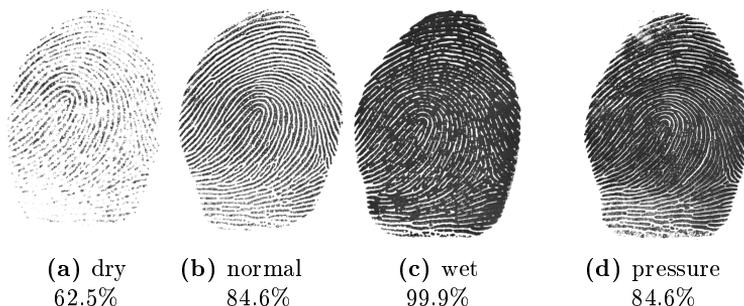


Figure 6.2: Fingerprint samples acquired with an optical scanner from the same finger with different skin moisture and at varying pressure. Giving different impressions: a) dry, b) normal, c) wet, d) pressure.

Therefore, a dataset collection is performed, with ground-truth indication of skin moisture during acquisition. This ground truth can be measured subjectively by a person, but this would be limited to few categories (dry, normal, wet) to avoid errors and thus precision would be low. Thus, the moisture of the fingertip skin is measured objectively using a digital device.

The collected fingerprint sample dataset is referred to as WDSET (abbreviated Wet Dry Dataset).

6.5.1 Moisture measurement

The outermost layer of the human skin is called epidermis. It is composed of several layers with the stratum corneum layer on the surface. Several commercial devices measuring the moisture of stratum corneum skin layer perform the measurement either by analysing the capacitance or the conductance of the skin [CB12] [ANN+].

Clarys et al. [CB12] have analysed both approaches, their analysis can be

summarised as follows:

- measurements done via both types of devices are similarly affected by the dielectric constant of the moisture which covers the skin surface.
- fat (mineral oil) has lower dielectric constant (2) than ethanol (50) or water (80) and thus the measurement is more influenced on the latter,
- capacitance method carries information from deeper parts of the skin than impedance ($45\mu m$ vs. $15\mu m$),
- both measurement approaches show high correlation ($r=0.97$)

Professional skin moisture measurement devices specialized for medical or cosmetic purposes are difficult to purchase and rather expensive. Fortunately, it is possible to obtain cheaper alternatives of Chinese production via most popular auction services. Wide availability and low cost of these would support the ease of reproducibility of the experiment.



Figure 6.3: Face Care Moisture Monitor SK-3 [ske]

For this experiment, a moisture measuring device "Moisture Meter SK-3" produced by Chinese company Face Care in the Shenzhen Kier Electronic Apparatus Factory [ske] is chosen. It is shown in figure 6.3. This device uses the conductance method to give a percentage score; it is cheap to purchase, light (40g), small and portable.

Example fingerprint samples with measurements from this device are shown in figure 6.2. Extreme moisture levels, when the finger skin is covered with water are indicated by 99.9% (over the range).

The Moisture Meter SK-3 takes around 10 seconds per finger to measure skin moisture. It performs calibration before each measurement and initial tests have shown that in cases of extreme moisture where the probe is soaked with water, the device may become inoperable until it fully dries.

6.5.2 Fingerprint sensors

For the data collection the following five optical fingerprint scanners with resolution of 500dpi are used:

- LScan 100 by Cross Match Technologies Inc. [cro],
- DFR-2100 Single Finger Reader by L-1 Identity Solutions Inc. (acquired by Safran S.A.) [mor],
- ZF-1 by Dermalog Identification Systems GmbH [der],
- Patrol ID by Cross Match Technologies Inc. [cro],
- Guardian by Cross Match Technologies Inc. [cro].

Figure 6.4 shows the sensors and technical specifications are given in table 6.1.



Figure 6.4: Fingerprint sensors used for data collection:

- a) Dermalog ZF-1
- b) CrossMatch PatrolID
- c) CrossMatch LScan100
- d) CrossMatch Guardian
- e) Safran (L-1) DFR-2100

6.5.3 Skin conditions variation

For the dataset, of interest are fingerprints acquired from fingers with varying skin moisture under constant force (pressure) applied to the sensor.

Sensor	Platen size	Image size	Operating humidity
L SCAN 100	$31.5 \times 31.5mm$	$620 \times 620px$	10-80% (non-condensing)
DFR 2100	$25 \times 25mm$	$500 \times 500px$	20-80% (non-condensing)
Patrol ID	$81 \times 76mm$	$800 \times 750px$	10-90% (non-condensing)
Guardian	$81 \times 76mm$	$800 \times 750px$	10-90% (non-condensing, splash resistant)
ZF 1	$24 \times 16mm$	$320 \times 480px$	"Excellent quality for dry and wet fingerprints" [der]

Table 6.1: Specifications of fingerprint sensors used for data collection

6.5.3.1 Constant force

In order to fulfil the requirements of invariant pressure during sample acquisition, the dataset collection is performed using a weight of 133 grams placed on a finger prior to acquisition. The subjects are instructed to acquire the sample as follows:

1. put the finger on the sensor platen and do not apply force,
2. put the weight on the finger just above the fingernail,
3. the sample is acquired manually by the software operator.

6.5.3.2 Varying moisture

Usually the hands of a subject do not show extreme moisture levels in office conditions, and therefore additional treatment is required to give sufficiently dry and wet impressions with a limited number of participants and office conditions.

There are four impressions of interest – natural and untreated, artificially dried, artificially moistened, and finally artificially overmoisturized. To obtain these impressions, the following treatment is used:

- For the normal impression, there is no treatment and the fingerprint is captured with subjects natural fingertip moisture.
- For the dry impressions, the treatment is to gently clean the finger with alcohol solvent (68% 760g/l ethanolum, 2% acidum salicylicum, 30% water) and wait a few seconds until the solvent remainings evaporate from the skin surface.

This simulates a practical case when subjects clean their hands with a disinfecting alcohol-based soap.

- To artificially moisturise the finger after drying, water based body lotion (Dove Essential Nourishment Body Lotion) is used – normal lotion would be harder to detect using the moisture measurement device, since water has highest dielectric constant – therefore oil is not used.

This simulates a practical case where subjects would use hand lotion before sample acquisition.

- To ensure over moisturised fingers are also presented to the sensors, the fingers are finally put in water. Prior to this step, remaining excessive lotion is removed with a cloth, but the hands are not thoroughly washed.

This simulates a practical case when subjects present their wet hands when e.g. it was raining or non-dried after washing.

6.5.4 Acquisition procedure

The dataset collection is performed indoors in a calm office to ensure common environmental conditions for all participants. Each subject is instructed to follow a strict procedure, where the skin conditions of fingers are varied as discussed in section 6.5.3, in order.

All of ten fingers are subject to acquisition using all five sensors in a random order. Fingertip skin moisture of each finger is measured prior to acquisition with all five sensors in a row and this value is registered for each fingerprint sample.

The sensor order is random to avoid possible balancing caused by slight moisture variations between sensors – e.g. if acquisition with five sensors causes the moisture to degrade and the order is fixed, then the last sensor in the row would have the driest fingers for the same moisture indication as the first.

To further eliminate the possibility of skin moisture changing between sensors, the delay between acquisitions is minimized and if a pause longer than a few seconds is made between two sensors, an additional measurement is performed before remaining acquisitions and new moisture is registered for these samples.

Before each acquisition, the sensor platen is cleaned to ensure no residual image remains. Similarly, the moisture measurement device probe is cleaned after each moisture measurement.

Initial tests have shown that the impact and moisture change between sensors is very small and randomizing the order is sufficient to avoid balancing error.

The procedure of data collection is as follows:

1. Normal acquisition of each finger (untreated).
 - Perform moisture measurement of fingertip by touching the probe to the finger and pressing the measurement button.
 - Acquisition with all sensors in random order (place finger on platen, place weight on finger, acquire sample and register the measured moisture).
2. Dry acquisition of each finger (dried with alcohol).
 - Clean fingertip with alcohol-soaked cloth.
 - Dry fingertip with dry paper.
 - Wait a few seconds until the alcohol fully evaporates.
 - Perform moisture measurement of fingertip by touching the probe to the finger and pressing the measurement button.
 - Acquisition with all sensors in random order (place finger on platen, place weight on finger, acquire sample and register the measured moisture).
3. Normal acquisition of each finger 2 (moisturized with water-based body lotion).
 - Apply a bit of lotion to fingertip and distribute uniformly so a thin layer is formed.
 - Immediately perform moisture measurement – before the lotion ingredients proportion changes due to evaporation or absorption by skin. This ensures common conditions for all people.
 - Acquisition with all sensors in random order (place finger on platen, place weight on finger, acquire sample and register the measured moisture).
4. Wet acquisition of each finger (treated with water).
 - Apply water to fingertip, shake off excessive droplets of water but ensure the finger is covered with water and not just humid.
 - Acquisition with all sensors (place finger on platen, place weight on finger, acquire sample; moisture level is registered as 99.9%).¹

¹All samples acquired during the last session are registered with moisture level of 99.9% without the actual measurement performed to speed up the procedure.

6.5.5 Dataset collection summary

The fingerprint dataset collection resulted in a total number of 6600 fingerprint samples acquired from 33 subjects – 200 fingerprints per person; 1650 samples for each of the four procedure steps: "normal", "dry", "lotion" and "water".

The fingerprints were acquired from participants of varying age and ethnic origin as shown in table 6.2. The procedure took one hour per person on average.

Prior to acquisition, each of the participants had agreed and signed a form stating that the acquired fingerprints may be used for research purposes.

Age					
max	59	min	23	average	33.5
Ethnic origin					
Caucasian	25	Asian	7	Mid.Eastern	1
Sex					
Female	16	Male	17	total	33

Table 6.2: Fingerprint dataset collection participants data summary.

6.6 Fingerprint sample analysis

This section describes the second part of the experiment procedure – calculation of values – comparison scores, utility and quality.

Three fingerprint comparators are available in the Center for Advanced Security Research Darmstadt (CASED) [CASa], these are extracted from commercially available products, offered for research purposes by their vendors.

The source of these comparators is confidential and thus they are treated as black boxes that produce a comparison score given two fingerprint samples. In CASED these algorithms are known under codenames "28", "63" and "83" and this naming convention is also used in this document.

6.6.1 Comparison score calculation

Comparison scores between all pairs of fingerprints are calculated for each of the two datasets using each of the available comparators such that:

- all genuine comparison scores are calculated for each dataset, with an exception that no sample is compared to itself,
- a limited number of impostor comparison scores is calculated for each dataset.

For CASIA, 100 impostor comparison scores per fingerprint sample are calculated. For the collected dataset, WDSET, this number is 20 per sensor, i.e. for each fingerprint of index finger acquired with sensor 1, it will have 20 impostor scores generated by comparing it with fingerprint samples acquired with sensor 1 from 20 other subjects' index fingers.

6.6.2 Observed utility extraction

For each of the two datasets, all fingerprints of a dataset have their utility related quality value extracted (calculated). This is performed as described in section 3.4, for each of the three fingerprint comparators.

The result is that each fingerprint from both datasets has three utility values related. These values are treated as ground truth observed performance.

6.6.3 Fingerprint quality estimation

The following reference and proposed methods are used to measure fingerprint quality of each fingerprint sample:

- NFIQ 2.0 candidate Quality Measurement Algorithms (see chapter 4):
 - Orientation Certainty Level (OCL),
 - Ridge Valley Uniformity (RVU),
 - Local Clarity Score (LCS),
 - Gabor Shen (GSH),
 - NIST Finger Image Quality (NFIQ).
- Proposed Quality Measurement Algorithms with all aggregation methods (see chapter 5):
 - Ridge Valley Difference Mean (RVDM),
 - Ridge Valley Difference Entropy (RVDE),

- Ridge Valley Difference Number of Good blocks (RVDNG),
- Ridge Line Count Mean (RLCM),
- Ridge Line Count Entropy (RLCE),
- Ridge Line Count Number of Good blocks (RLCNG),
- Contrast Mean (CNTM),
- Contrast Entropy (CNTE),
- Contrast Number of Good blocks (CNTNG).

Running each method per fingerprint sample results in quality indication of that sample. Therefore, each fingerprint is associated with 14 quality values.

Additionally, the proposed wetness detection method is also ran similarly and results in wetness-indication value that is also associated with the samples.

However, the wetness indication method is only run on WDSET, since only this dataset has the objective moisture measurement.

6.6.3.1 Image segmentation

All of the Quality Measurement Algorithms used in this experiment utilize a segmentation algorithm in order to separate the valid fingerprint foreground from the whole sample.

This way, the aggregation of scores assigned to each block of an analysed sample is only performed for blocks that are marked as foreground.

The method described in section 4.7 is used for all the proposed methods and the following reference QMAs: OCL, LCS and RVU. All of the methods use the segmentation procedure with a block size of 32×32 pixels, except for the Contrast QMA, which performs segmentation using blocks of size 8×8 pixels, because it operates on blocks of such size as well.

The Gabor Shen (GSH) QMA uses its own specific segmentation based on the response of the filter (see section 4.5 for the description). The NIST Finger Image Quality (NFIQ) QMA [TW05] calculation procedure does not mention image segmentation.

6.7 Proposed methods performance assessment

The performance of each proposed fingerprint Quality Measurement Algorithm is assessed in terms of:

- average execution time,
- correlation of quality scores with observed utility,
- improvement in False Non-Match Rate with poor quality samples rejection.

6.7.1 Average execution time

The average execution time of each assessed fingerprint Quality Measurement Algorithm is calculated as follows:

1. for each available fingerprint sample dataset randomly choose 400 fingerprint samples,
2. run the method for each of the samples and measure the average execution time,
3. calculate the total average from all available datasets.

6.7.2 Correlation with utility

Similarly as performed for NFIQ 2.0 Features Evaluation [nfiq], each Quality Measurement Algorithm performance is assessed in terms of how the generated quality scores relate to the utility values. Utility score extraction is performed as described in section 3.4.

To assess how well this relationship can be described using a monotonic function, the correlation value is calculated with Spearman's rank correlation coefficient:

$$\rho = \frac{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

Where x_i and y_i are ranks of values X_i and Y_i respectively, taken from the two sets of values X and Y , i.e. the set of quality scores and the set of utility scores.

For ease of readability, each correlation value is multiplied by 100 and rounded, such that all values are in range from -100 to +100.

This is performed per dataset. For each dataset, each QMA quality indication is correlated with all other QMA indications and all three utility values (one utility for each of the three comparators 28, 63 and 83). Additionally, the correlation between utility values is also measured to show their relation.

This allows not only to assess the performance of each method individually, but to relate the methods between each other – high correlation between two QMAs suggests they are very similar metrics, whereas low correlation shows that they are complementary.

6.7.2.1 Collective assessment

In order to finally assess the proposed QMAs using this method, the average of correlation values are calculated per dataset from all three utility values (from comparators 28, 63 and 83). Then the average from both datasets is computed.

Some quality indications may give positive correlation for one provider or dataset and negative for another. Therefore the averages are taken from absolute values of these coefficients.

This way it is possible to assess the proposed methods in terms of correlation with utility across all chosen datasets and available comparators.

6.7.3 Error versus Reject Curves

Finally, the performance of proposed methods is assessed with Error versus Reject Curves, described in section 3.5.1.2. These curves directly show how False Non-Match Rate improves with rejection of poor quality samples.

The ERC plots are generated for all 14 quality metrics for each of the available datasets and each of the three available comparators. The plots are generated for FNMR of 10 per cent, similarly as in [GT07]. The maximum fraction rejected of interest is 35 per cent.

6.7.3.1 Collective assessment

Quantitative analysis of ERC curves is difficult with more than a few lines, i.e. it is possible to say which QMA is better when there are e.g. only three curves to compare. With 14 lines this task is almost impossible, especially if the curves are not straight and cross.

However, it is possible to quantitatively express how low a certain Error versus Reject Curve is, across the whole fraction–rejected range of interest (zero to 35%), by integrating the curve to find the area under it. The smaller the area the lower the curve on average. Even if two curves overlap, comparing their area will show which is better on the full fraction–rejected range of interest.

In order to improve the readability of this area indication and normalize the scores, the area under the ideal case ERC is also calculated and subtracted from all ERC curves' areas. This way the ideal case has an indication of zero and evaluated QMAs are scored in range zero plus.

This is illustrated in figure 6.5, where the lightly shaded triangle represents the area under ideal case and the dark shaded area is the normalized area 'score' of the QMA represented by the blue curve. The score is obtained by rounding the area, which is first multiplied by 100 for readability.

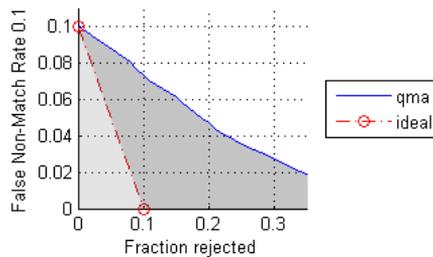


Figure 6.5: Error versus Reject Curve plot with areas used for QMA scoring marked in light grey – ideal case, dark grey – blue curve QMA.

All the generated ERC curves areas are calculated and this information is used to build a score card. Such a score card has values per dataset and per comparators, plus averages per dataset and total average information. This is expected to show which QMA gives best FNMR improvement on average across all datasets and comparators.

6.7.3.2 NIST operational dataset

Additionally to the Error versus Reject Curves generated for the chosen datasets, the proposed and reference quality metrics are executed on a fingerprint data subset owned by the National Institute of Standards and Technology (NIST) [NIS].

This subset is 2000 fingerprint samples extracted from operational data which is confidential – the data is acquired in real scenarios, not laboratory environment.

Fingerprints of high quality variation are featured, chosen specific for Quality Measurement Algorithm evaluation. For this dataset, evaluation is performed using Error versus Reject Curves.

6.8 Moisture impact on quality

The moisture information that is relevant for this project is the relation of the finger skin wetness with the performance of the sample acquired from that finger. This relation is expressed by:

- correlation of measured moisture with observed utility,
- relation between moisture genuine comparison scores².

The analysis is expected to show that extreme moisture levels lead to performance degradation and that samples below or above a certain moisture threshold can be labelled as dry or wet respectively. This may vary between comparators.

6.8.1 Moisture impact on comparison scores

In order to analyse how certain moisture affects the quality of samples, observed performance is related to the measured moisture.

In theory, the observed performance could be extracted utility or comparison scores. However, utility calculation performs high averaging of comparison scores and actually discards the most extreme values. This way, several samples

²Impostor comparison scores are expected to remain low regardless of the quality or moisture and therefore their relation is not analysed

that may indicate correlation would not even be analysed. Initial experiments show that this is the case Hence, utility and moisture correlation is not of interest.

Therefore, the performance of samples is expressed as comparison scores. These values are generated for a pair of samples and thus the results are shown as heat maps. This way X and Y axis show the moisture level and the colour of the map represents the comparison score.

The heat maps are generated per sensor and per comparator, with moisture levels binned into 15 levels (each 4% of moisture indication). This is expected to show differences between sensors and between comparators.

6.8.2 Proposed moisture indication method performance

Performance of the wetness indication method is assessed as:

- average execution time of this method for 400 samples randomly chosen from each dataset and the overall average,
- correlation of the wetness indication score with the registered sample moisture – the higher the correlation, the better the method (this is calculated using Spearman’s rank correlation coefficient),
- error in threshold based binary classification of samples as dry or wet, accordingly to the moisture threshold.

6.8.2.1 Binary decisions

The Moisture Indication method can be used as a binary classifier, when compared to a threshold. This can be performed to detect if the finger is dry, and to detect if the finger is wet – two thresholds are required.

However, before the Moisture Indication thresholds can be chosen, it is crucial to know which moisture levels actually cause performance degradation (if any). The moisture analysis described in section 6.8 is expected to show certain thresholds, i.e. low moisture threshold below which performance is low and similarly high moisture threshold.

When the moisture thresholds are known, e.g. samples with moisture below 69% are considered as dry and above 96% as wet, the thresholds of Moisture Indication for dryness and for wetness can be found.

This is performed by analysing the Detection Error Tradeoff (DET) curves, which show the relation of False Positive (FP) and False Negative (FN) detection rates for all possible thresholds. This way, a threshold of interest is chosen such that both FN and FP remain reasonably low.

Experimental results

This chapter describes the results obtained from executing the experiment procedure described in chapter 6.

Statistics of genuine and impostor comparison scores calculated on the datasets with the available comparators are described in section 7.1. Section 7.2 describes the results of the performance evaluation of the proposed Quality Measurement Algorithms.

The impact of moisture of the skin on fingerprints quality is analysed in section 7.3. Finally, section 7.4 describes the performance measurement results of the Moisture Indication method.

7.1 Calculated comparison scores

The distributions of comparison scores calculated as described section 6.6.1 are shown in figure A.1 in Appendix A. Six histograms showing the number of genuine (blue) and impostor (red) comparison scores are shown, one for each dataset-comparator combination respectively.

Each comparator behaves similarly on both datasets, but the distributions are significantly different between comparators. This is both for genuine and impostor score distributions.

7.1.1 Impostor distributions

Comparators 28 and 63 produce impostor scores in the lowest range, whereas impostor scores from comparator 83 show a nearly normal distribution in the first half of the total range of this comparator.

7.1.2 Genuine distributions

On WDSET, comparator 28 gives a flat distribution of genuine scores with a higher concentration in the highest range. On CASIA, the distribution is similar, but the top range peak is smaller and there is higher concentration in the lowest 20 per cent of the range.

Comparator 63 gives a normal distribution of genuine scores on both datasets in the first three quarters of the range and with a peak in the lowest range. The peak is small on WDSET and reaches half of the total count for CASIA.

Finally, comparator 83 gives genuine scores almost geometrically growing towards the top of the range for both datasets.

7.1.3 Summary

Differences in distributions between comparators show that each performs differently. Differences between datasets for each comparator suggest that the samples in each dataset differ in terms of quality.

The overlap between genuine and impostor comparison scores is high when comparator 63 is used, especially on the CASIA dataset where almost half of the genuine scores are in the same range as impostor scores. A possible explanation is that several fingerprints from the CASIA dataset are acquired with high finger rotation and perhaps comparator 63 does not perform well in such cases.

As for comparators 28 and 83, both give a smaller overlap between genuine and impostor scores than comparator 63. Their performance is similar on WDSET,

but 83 produces slightly less genuine comparison scores in the range of impostor scores with CASIA. Hence, 83 is the best available comparator on the used data.

7.2 Proposed methods evaluation

As described in section 6.7, performance of proposed fingerprint Quality Measurement Algorithms is evaluated in terms of execution time and how well the quality indication of the proposed methods predicts the performance of the analysed fingerprint samples via Error versus Reject Curves, and via correlation with utility values.

7.2.1 Execution time

Average execution time of each proposed and each reference fingerprint Quality Measurement Algorithm method is calculated as described in section 6.7.1. Table 7.1 shows the results.

Full name	code	WDSET	CASIA	Avg.
Nist Finger Image Quality	NFIQ	0.206	0.212	0.209
Gabor Shen	GSH	0.352	0.277	0.315
Ridge Valley Uniformity	RVU	0.229	0.247	0.238
Orientation Certainty Level	OCL	0.144	0.124	0.134
Local Clarity Score	LCS	0.295	0.295	0.295
Ridge Line Count Mean	RLCM	0.605	0.591	0.598
Ridge Line Count Entropy	RLCE	0.607	0.589	0.598
Ridge Line Count No.Good	RLCNG	0.607	0.589	0.598
Ridge Valley Difference Mean	RVDM	0.114	0.093	0.104
Ridge Valley Difference Entropy	RVDE	0.114	0.093	0.104
Ridge Valley Difference No.Good	RVDNG	0.114	0.093	0.104
Contrast Mean	CNTM	0.847	0.662	0.760
Contrast Entropy	CNTE	0.849	0.649	0.749
Contrast No.Good	CNTNG	0.848	0.647	0.748

Table 7.1: Execution time – reference and proposed Quality Measurement Algorithms – per 400 samples from databases: CASIA Fingerprint V5.0, collected dataset – WDSET, average.

All proposed methods and three of the reference methods are generally executing faster on the CASIA database than on WDSET; Ridge Valley Uniformity and

NFIQ are a bit slower. Shorter execution times on the CASIA dataset may be caused by smaller sample image sizes (see sections 6.5.2 and 6.4).

The averaged execution times of reference methods are between 0.134 seconds of the Orientation Certainty Level QMA and 0.315 seconds of Gabor Shen. The NFIQ QMA takes 0.209 seconds on average to analyse a fingerprint sample.

The slowest proposed method is Contrast, as it takes 0.76 seconds with the mean aggregation, 0.749 with entropy and 0.748 using number of blocks with score above threshold. This method is slower than all reference methods probably because of increased number of blocks – this method analyses samples in blocks of size 8×8 pixels, whereas all other methods use blocks of 32×32 pixels.

The fastest proposed method is Ridge Valley Difference – it takes 0.104 seconds to analyse a fingerprint sample using all aggregation methods. This method is also faster than all the reference methods.

The Ridge Line Count QMA takes 0.598 seconds to execute, regardless of the aggregation method used. It is faster than Contrast, yet slower than the reference methods and the Ridge Valley Difference QMA.

7.2.2 Error versus Reject Curves

The proposed Quality Measurement Algorithms are analysed in terms of False Non-Match Rate drop with rejection of poor quality samples, as described in section 3.5.1.2. The results are described in terms of area under the graph, such that the value closest to zero shows best performance of a QMA (see section 6.7.3.1).

Table 7.2 shows the ranking of all reference and proposed methods. The scores shown are averages from all comparators on each database. Additionally, the results of ERC assessment on the operational NIST dataset (see section 6.7.3.2) are shown in the last column of this table.

Full results – per-comparator averages and actual Error versus Reject Curve plots are featured in Appendix B.

Performance of the reference and proposed methods is generally better on WDSET than on CASIA (though there are exceptions, OCL, RVDM and CNTM, which perform better on CASIA). This shows that the datasets differ in terms of quality variation.

QMA name	code	CASIA	WDSET	avg	NIST
Nist Finger Image Quality	NFIQ	2.24	2.03	2.14	N/A
Gabor Shen	GSH	2.07	1.23	1.65	3.36
Ridge Valley Uniformity	RVU	2.19	1.36	1.78	2.96
Orientation Certainty Level	OCL	2.28	2.59	2.44	2.54
Local Clarity Score	LCS	2.43	1.60	2.02	2.36
Ridge Line Count Mean	RLCM	1.97	0.99	1.48	2.65
Ridge Line Count Entropy	RLCE	1.99	0.78	1.39	2.64
Ridge Line Count No.Good	RLCNG	1.96	1.08	1.52	2.75
Ridge Valley Difference M.	RVDM	2.11	2.70	2.41	3.10
Ridge Valley Difference E.	RVDE	2.62	2.03	2.33	3.05
Ridge Valley Difference NG.	RVDNG	2.17	2.64	2.41	3.10
Contrast Mean	CNTM	2.23	2.70	2.47	2.98
Contrast Entropy	CNTE	2.86	2.06	2.46	3.50
Contrast No.Good	CNTNG	2.06	1.67	1.87	2.97

Table 7.2: QMA ranking based on ERC ranking averaged from all three comparators on both datasets: CASIA and WDSET. The fourth column (avg) shows the average from both datasets. Results from the NIST operational dataset (see section 6.7.3.2) are shown in the last, fifth column. Lower values represent better performance. Numbers are calculated as described in section 6.7.3.1 and shown in figure 6.5.

The ERC score averaged from both datasets shows that reference methods score from 1.65 (best, Gabor Shen) to 2.44 (worst, OCL). The proposed methods reach a wider span of scores – from 1.39 (Ridge Line Count Entropy) to 2.47 (Contrast Mean). The worst of proposed methods is only slightly worse than the worst of the reference methods, and the best proposed method outperforms the best reference.

The best proposed method is Ridge Line Count, which with all aggregation methods produces ERC scores lower (better) than the best reference – Gabor Shen. Another very good proposed method is Contrast with number of good blocks aggregation – it outperforms three of the reference methods – NFIQ, RVU and OCL.

Contrast QMA with Mean and Entropy aggregations show worse performance than with Number of Good Elements aggregation with samples from CASIA and WDSET. Ridge Valley Difference QMA is generally the worst of the proposed methods because all aggregation methods give poor results. Although RVD still outperforms the reference Orientation Certainty Level.

7.2.2.1 NIST operational dataset

The ERC results from the NIST dataset show slightly different behaviour of QMAs. On this dataset, the best performing reference method is Local Clarity Score, with a score of 2.36. Next is Orientation Certainty Level with 2.54, which was the worst on the other datasets.

The worst performing (score 3.36) is Gabor Shen, a QMA that performed best on the other datasets. The reason behind this surprising result is unknown. NFIQ was unfortunately not assessed on this dataset.

The best proposed method is Ridge Line Count – with entropy and mean (scores 2.64 and 2.65 respectively) aggregations, it outperforms Gabor Shen and Ridge Valley Uniformity. OCL and LCS are better than this method on the NIST dataset.

The Contrast QMA with number of good blocks aggregation gives a score of 2.97, slightly worse than RVU (2.96). Ridge Valley Difference does not offer a very good performance on the NIST operational dataset.

Reference methods show different behaviour on the NIST dataset than on CASIA and WDSET. Proposed methods behave similarly, Ridge Line Count shows best performance, although CNTNG is not that good on the NIST dataset.

7.2.3 Correlation with utility

As described in section 6.7.2, performance of the Quality Measurement Algorithms is assessed in terms of how well the quality correlates with observed utility. Table 7.3 shows the average correlation results for CASIA, WDSET, and the total average.

The correlation scores on WDSET are better (higher) than on CASIA. This shows that the assessed QMAs perform better on the former dataset.

On average, best performance is achieved by the NFIQ algorithm. This is expected as it is the current de facto standard. The following best performing reference methods are Gabor Shen and Ridge Valley Uniformity with a tied score of 47. LCS and OCL give scores 32 and 30 respectively.

The best score of the proposed methods is achieved by Ridge Line Count with Entropy aggregation (48). With number of good blocks aggregation, RLC has

name	code	CASIA	WDSET	total avg
Nist Finger Image Quality	NFIQ	39	62	51
Gabor Shen	GSH	35	59	47
Ridge Valley Uniformity	RVU	30	64	47
Orientation Certainty Level	OCL	28	31	30
Local Clarity Score	LCS	26	57	42
Ridge Line Count Mean	RLCM	36	41	39
Ridge Line Count Entropy	RLCE	41	55	48
Ridge Line Count No.Good	RLCNG	40	45	43
Ridge Valley Difference Mean	RVDM	33	28	31
Ridge Valley Difference Entropy	RVDE	17	32	25
Ridge Valley Difference No.Good	RVDNG	35	29	32
Contrast Mean	CNTM	25	15	20
Contrast Entropy	CNTE	10	19	15
Contrast No.Good	CNTNG	36	37	37

Table 7.3: Correlation of quality and utility scores. Averages calculated for all three comparators for CASIA and WDSET. Last column shows the total average. Scores are calculated as an absolute value of Spearman’s RHO multiplied by 100 and rounded. Higher values represent better performance

a score of 43. Therefore, the proposed RLCE method is only worse than NFIQ, it outperforms Gabor Shen and Ridge Valley Uniformity.

Proposed Contrast with number of good blocks aggregation scores of 37. Other aggregation methods are worse. Ridge Valley Difference with number of good blocks aggregation it scores 32. Both CNT and RVD with this aggregation method outperform Orientation Certainty Level.

The worst performing proposed methods are Contrast Entropy (15) and Ridge Valley Difference Entropy (25). Considering the scores of all methods assessed as NFIQ 2.0 candidates [nfiq], these are still good results.

7.2.3.1 Inter-method correlation

Tables C.1 and C.2 in Appendix C are constructed as in the NFIQ 2.0 candidate features evaluation document [nfiq]. These tables show the inter-method correlation scores. Analysis of such correlation allows to assess whether two methods give similar indication, or if their scores are complementary.

The highest correlation between a pair of two reference methods (GSH and LCS) is 76 and 82 for CASIA and WDSET respectively. None of the proposed methods gives a higher correlation with the reference methods on WDSET, but on CASIA the Ridge Valley Difference with Mean aggregation has a correlation score of 80 with Gabor Shen.

Generally all proposed methods are considered complementary to the set of reference Quality Measurement Algorithms because their respective inter-method correlation scores are sufficiently low.

7.2.4 Summary

All proposed methods work – performance of the analysed samples is predicted by the quality output when using all of the proposed methods with all aggregation methods tested.

7.2.4.1 Execution time

Average execution times of the proposed methods show that one of the methods is definitely fast to compute, as it outperforms all reference methods. Two other proposed QMAs are twice as slow as the reference methods. However, there are no known boundaries of execution time of a Matlab interpreted-language implementation.

The NFIQ 2.0 project aims at execution time of less than 150ms [Ols13] in order to be incorporated in acquisition loops of live fingerprint sensors. Current de facto standard NFIQ takes 209ms to execute.

With a compiled-language implementation, the execution time of the proposed methods is expected to improve by an order of magnitude, especially for the very simple Contrast QMA. Therefore it is believed that all proposed methods are sufficiently fast to compute to be possible NFIQ 2.0 candidates.

7.2.4.2 Error versus Reject Curves

From the proposed methods, Ridge Line Count with all aggregation methods offers very good performance in comparison to the reference methods on all available data, including results from NIST.

Contrast with number of good blocks aggregation offers good performance, especially on the collected dataset – WDSET.

All Ridge Valley Difference based QMAs and Contrast with Mean or Entropy aggregation offer acceptable performance, but not significantly better than the reference QMAs.

7.2.4.3 Correlation with utility

In this assessment, the Ridge Line Count QMA is better than four of the five reference methods and almost as good as the current de facto standard.

Other proposed Quality Measurement Algorithms offer acceptable performance (non-zero correlation), but in the best case outperform only one of the reference methods.

The proposed methods are complementary to the set of reference methods as all of them have low inter-method correlation scores.

Results of ERC and utility correlation differ since the former analyses the worst-quality 35 per cent of samples, whereas the latter operates on the full range.

7.3 Moisture impact on quality

This section describes the impact of fingerskin tip moisture on the quality of fingerprints acquired, as described in section 6.8.

This analysis is performed on WDSET – the dataset collected accordingly to the procedure described in section 6.5.

7.3.1 Measurement statistics

The distribution of measured moisture values with respect to the procedure steps (untreated, dried with alcohol, moistened with lotion and finally with water) is shown in figure 7.1.

The distribution of measurements from fingers dried with alcohol is in a range lower than those acquired from fingers with body lotion applied. Untreated

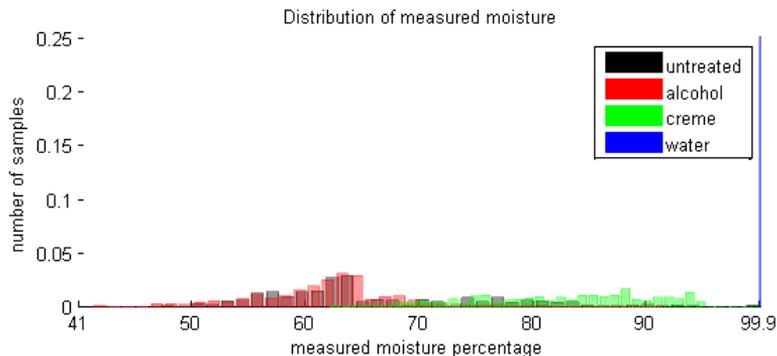


Figure 7.1: Distribution of measured moisture with respect to each step of the acquisition procedure.

fingers have a distribution of measurements similar to that of alcohol and crème combined.

The acquisition procedure worked as expected and WDSSET contains samples acquired from fingers with different skin moisture level. The minimum and maximum measured moisture values are 41 and 99.9 per cent respectively, where the latter denotes an over the range measurement as described in section 6.5.1.

7.3.2 Moisture versus comparison scores

The impact of moisture on calculated genuine comparison scores causes degradation of scores for certain high and low moisture levels. Figure 7.2 shows a heat map that relates comparison scores (block colour) to moisture of compared samples. Moisture is indicated on the x and y axes in the range from lowest measured 41% to highest 99.9%.

This heat map is generated as described in section 6.8.1, for comparator 28 using samples acquired from all sensors. The dark blue areas in the absolute lowest range are caused by the fact that almost no samples were captured with this lowest moisture level below 45% – it does not indicate a drop of comparison scores.

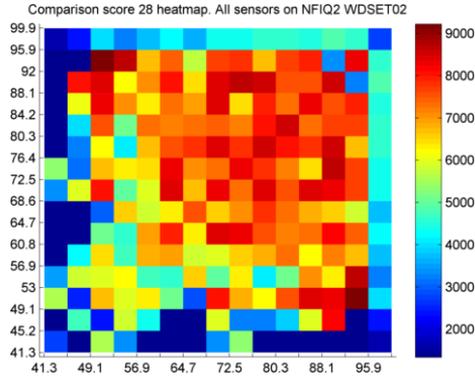


Figure 7.2: Moisture impact on comparison scores heat map. Comparator 28 on all sensors. Colour of the square shows the average comparison score of samples with moisture in the range for which the square is drawn. Horizontal and vertical axes show measured moisture percentage.

7.3.2.1 Fingerprint comparator differences

Heat maps generated for each of the comparators are shown in Appendix D. There are differences between comparator behaviour, according to the collective analysis of heat maps in figures D.1, D.2 and D.3, generated for all samples with comparators 28, 63 and 83 respectively.

Comparators 28 and 63 show a degradation of performance for samples with low moisture indication, whereas comparator 83 does not produce significantly lower scores in this range.

Samples acquired from fingers with moisture level of 99.9% lead to low comparison scores with comparator 63. When comparator 28 is used, the scores are higher – this comparator offers a better performance with wet hands. Such a behaviour is even more visible for comparator 83, which gives even higher scores in the top range. Though it still gives scores lower than non-extreme moisture levels.

7.3.2.2 Sensor differences

Heat maps generated from samples acquired only with a specific sensor for each comparator are shown in Appendix D. There are significant differences in

quality relation to moisture – some sensors do not produce low scores in extreme moisture levels. Samples acquired using different sensors perform differently, which confirms that some sensors are better than others.

Low moisture levels lead to drop to one third of the range in comparison scores (cyan squares) for all used sensors except for L-1 DFR-2100. From the four remaining sensors, samples acquired with CrossMatch Guardian give least low-dry-score (cyan) areas and those from Dermalog ZF-1 give the most.

The reason behind Dermalog ZF-1 performing poorly with samples acquired from very dry fingers may be due to the very small platen, which leads to small foreground portion in case of dry samples – this was observed during dataset acquisition and an example is shown in figure E.1.

Very high moisture levels (over the range, last procedure step where water is applied) cause comparison scores to drop to the bottom of the range for sensors L-1 DFR-2100, CrossMatch LScan100 and CrossMatch PatrolID. Samples acquired with CrossMatch Guardian and Dermalog ZF-1 give good comparison scores in this moisture range, i.e. these sensors perform well with wet fingertip skin.

Appendix E shows concrete examples of fingerprints acquired from one finger with different skin moisture using each of the five sensors. These samples are representative as generally similar impression differences were seen for other subjects during data acquisition.

7.3.3 Summary

Moisture of the fingertip skin does have an impact on the performance of fingerprints acquired.

However, it is possible to build a sensor that works better (produces samples of good performance) with fingers which are wet, e.g. CrossMatch Guardian or Dermalog ZF-1. It is also possible to build a sensor which works better with dry fingers, e.g. L-1 DFR-2100.

There are also differences between comparators. Some comparators perform poorly with wet and dry fingers, but it is possible to build a comparator that will perform very good with dry samples, e.g. comparator 83, regardless of the sensor used.

On the other hand, samples acquired from extremely wet fingers always lead to

performance degradation, although it is possible to build a comparator that will give better scores than other comparators in these cases, e.g. comparator 83.

7.4 Moisture Indication method performance

As described in section 6.8.2, performance of the Moisture Indication method is assessed in terms of execution time, how well its output correlates with measured moisture and the error rates of a binary dry or wet decision.

7.4.1 Execution time

The average execution time of the Moisture Indication method are shown in table 7.4 – it takes 0.101 seconds on average.

The performance is very good, better than all proposed and reference QMAs on average. This is due to the fact that the method is very simple to compute.

	CASIA	WDSET	average
Wetness Detection	0.111	0.09	0.101

Table 7.4: Execution time of wetness detection method per 400 samples from databases: CASIA Fingerprint V5.0, WDSET, average.

7.4.2 Detected versus measured moisture

The correlation between measured moisture values and output generated with the Moisture Indication method is shown in table 7.5. This is performed for each of the sensors and for all of them collectively. Correlation scores are negative since the method produces high scores for low moisture and low scores for high moisture.

Samples acquired with CrossMatch LScan100, L-1 DFR-2100 and CrossMatch Patrol ID give very high correlation between moisture measurement and indication score – -87, -89 and -89 respectively. The Moisture Indication method performs well on these samples.

This is not the case for Dermalog ZF-1 and CrossMatch Guardian, which give scores -50 and -20 respectively. Such low scores are caused by the fact that

	score
Dermalog ZF-1	-50
CrossMatch LScan100	-87
L-1 DFR-2100	-89
CrossMatch PatrolID	-89
CrossMatch Guardian	-20
All sensors	-68

Table 7.5: Correlation score between Moisture Indication and measured moisture, calculated as Spearman’s RHO multiplied by a hundred and rounded.

these sensors are newer and supposed to work well with wet or dry fingers – they behave differently than the other sensors. Examples of fingerprints confirming this observation are shown in Appendix E.

7.4.3 Binary classification error

As described in section 6.8.2.1, the Moisture Indication method can be used as a binary classifier, provided that moisture thresholds are defined for dry and for wet fingers.

7.4.3.1 Measured moisture thresholds

Concrete moisture levels are required to give binary decision on dryness or on wetness using the Moisture Indication method. Heat maps show that comparison scores are lower for moisture levels between the bottom of the measured range (41%) and 57%.

Therefore 57% is chosen as the dryness threshold. As for wetness, only the top of the range in moisture measurement leads to performance degradation. Thus, fingers with moisture measured as 99.9% are considered as wet.

7.4.4 Dryness classification

Detection Error Tradeoff in classifying samples as dry based on the Moisture Indication method output is shown in figure 7.3. Respective False Positive

and False Negative rates versus the Moisture Indication threshold are shown in figure 7.4. Appendix G shows False Positive and False Negative rates per sensor.

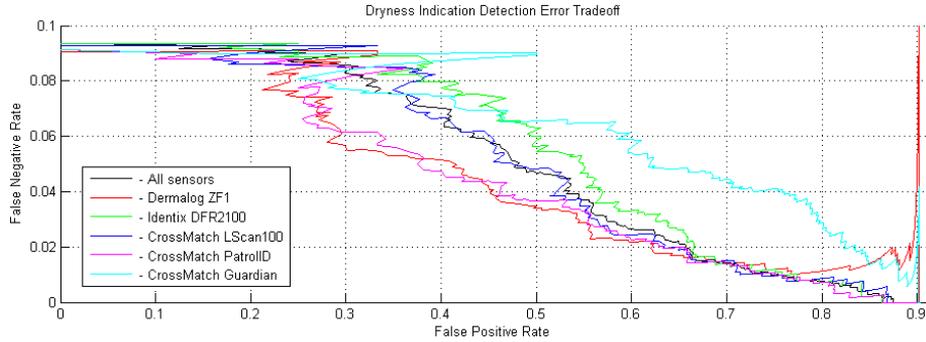


Figure 7.3: Detection Error Tradeoff for dryness classification – classifying samples with moisture below 57%. Note different range of axes.

The Moisture Indication method used as a dryness classification produces a False Positive Error of 30% for a False Negative Rate of 8% on samples acquired using all sensors (black line). In this case the threshold used for Moisture Indication is around 0.81, as read from figure 7.4 (green line).

For this classification it is better to pick a threshold such that False Negative Error is minimal, even if False Positive error is increased until the maximum reasonable boundary. This way samples which are definitely wet will almost always be correctly classified.

On the other hand, the indication of dryness will be shown even if the finger is not very dry (once per three samples on average, with the 30% FP chosen). This is however not a problem, because if a reacquisition is necessary and a subject moistens the finger because dryness is indicated, the quality will not degrade unless the subject puts extreme amount of moistener on the skin, which in operational conditions is not highly probable.

Hence, it is better to indicate dryness more often even if not true, than drop the indication in cases where it should be indicated.

The smallest possible error is produced for sensors Dermalog ZF-1 and Cross-Match PatrolID. CrossMatch LScan 100 gives an error almost identical as the average from all sensors. Using samples acquired with CrossMatch Guardian gives a slightly worse performance, especially for high False Positive Rates. The worst performance of dryness classification is on samples acquired using L-1

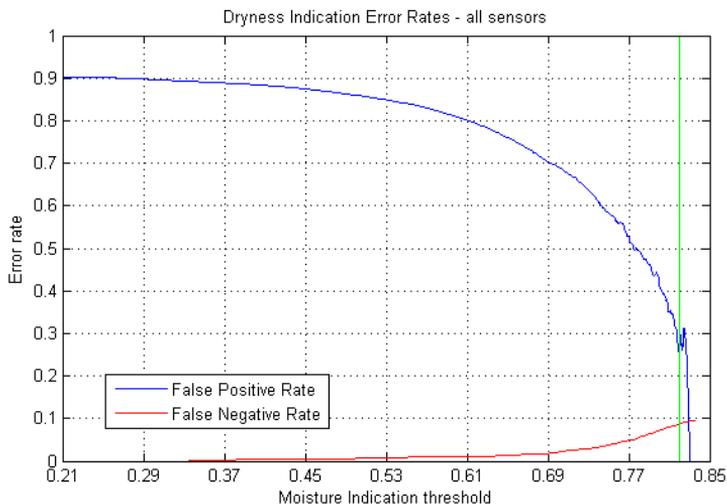


Figure 7.4: False Positive and False Negative error rates in dryness classification versus the threshold used for Moisture Identification. Green line shows the point of 30% FP and 8% FN.

DFR-2100, probably because this sensor features a silicone pad on the platen.

High peak in DET of Dermalog ZF-1 (red curve) is caused by two outlier samples with ground-truth error, as described in Appendix G and shown in figure G.2.

7.4.5 Wetness classification

Classification of acquired samples is performed for sensors which show performance degradation in these cases. These are:

- CrossMatch LScan100,
- L-1 DFR-2100,
- CrossMatch Patrol ID.

The relation between False Positive and False Negative detection rates for wetness classification are shown in figure 7.5. This figure shows the Detection Error Tradeoff (DET) as the for sensors that are known to perform poorly with wet fingers (see section 7.3).

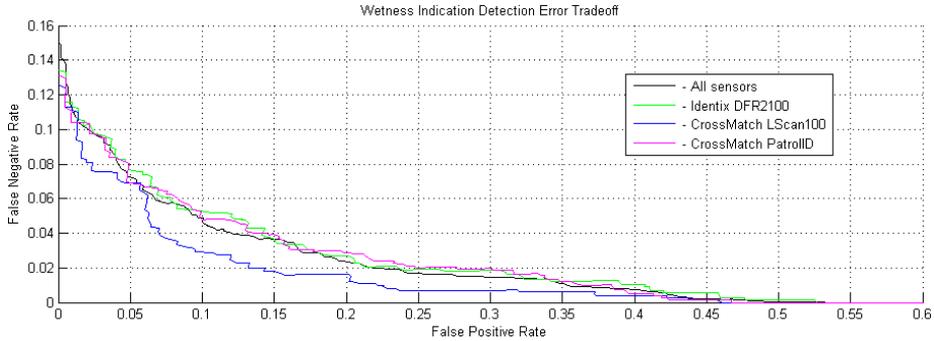


Figure 7.5: Detection Error Tradeoff for wetness classification – classifying samples with moisture above or equal 99%.

Performing the wetness classification for sensors Dermalog ZF–1 and Cross-Match Guardian does not give satisfying results, i.e. the False Positive rate is very high, as shown in figures G.1 and G.3 for Dermalog ZF–1 and CrossMatch Guardian respectively.

The reason behind this is sensor behaviour differences, explained in section 7.4.2 with examples shown in Appendix E. In short, these sensors give samples which look normal even if the fingers are extremely wet.

The False Positive and False Negative rates with respective Moisture Indication threshold used for wetness classification (figure 7.5) are shown in figure 7.6. Appendix F shows False Positive and False Negative rates per each sensor.

The Moisture Indication method used as a wetness classification gives a False Positive Error of 10% for a False Negative Rate of 5% on samples acquired using all sensors (black line). In this case the threshold used for Moisture Indication is around 0.39, as read from figure 7.4 (green line).

Similarly as for dryness indication, it is better to keep the False Negative Rate low even if False Positive Error reaches the reasonable limit. Therefore Equal Error Rate is not used.

This way, samples which are acquired from fingers which are actually wet will be correctly classified in more cases. On the other hand, an indication of wetness in case of reacquisition may lead to changed conditions (e.g. subject would wipe the finger with cloth) which is indeed necessary, since the quality was poor in the first case.

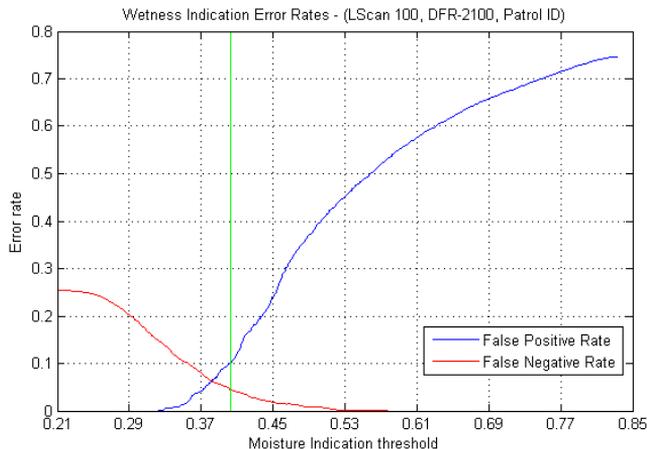


Figure 7.6: False Positive and False Negative error rates in wetness classification versus the threshold used for Moisture Identification. Green line shows the point of 30% FP and 8% FN.

7.4.6 Summary

The introduced Moisture Indication method is very fast to compute, 0.101 seconds on average.

With a threshold it can be used as a binary decision to classify fingerprints as acquired from wet hands, or from dry hands.

In case of wet fingers the error rates are 10% False Positive and 5% False Negative. Dry fingers classification gives a bit higher errors – False Positive of 30% and False Negative 8%.

Conclusions

The main purpose of this project was to propose new fingerprint Quality Measurement Algorithms, possibly to be considered as NFIQ 2.0 candidates [oST] and incorporated in ISO/IEC 29794-4 [ISO10].

Additionally it was relevant to assess whether very dry or very wet skin indeed causes biometric performance degradation and to propose a method which would indicate this via analysis of the acquired fingerprints.

A dataset of fingerprint samples was collected from 33 subjects using 5 sensors with objective fingertip skin moisture measurement in four varying moisture conditions, resulting in 6600 fingerprints with ground-truth information about skin moisture during acquisition.

To analyse fingerprint sample quality, three Quality Measurement Algorithms were proposed – Ridge Valley Difference, Ridge Line Count and a Contrast measurement. To detect extreme finger skin moisture, the impact of skin moisture on fingerprint sample quality was analysed and a Moisture Indication method was proposed.

The proposed quality metrics were assessed using the collected dataset – WD-SET, and a publicly available fingerprint dataset CASIA from the Institute of Automation at the Chinese Academy of Sciences. These methods were com-

pared to reference state of the art methods of fingerprint quality analysis in terms of execution time, quality indication correlation with observed fingerprint utility and via Error versus Reject Curves.

All proposed quality analysis methods were sufficiently fast to compute and offered good performance in terms of performance prediction via quality indication. Moreover, the methods offered quality indication complementary to that of reference methods. Ridge Line Count was the best proposed method, it performed better than several reference state of the art methods in both ERC and correlation analysis. Proposed methods can be considered as NFIQ 2.0 candidates and incorporated in ISO/IEC 29794-4 [ISO10].

The analysis of skin moisture impact on fingerprint sample quality has shown degradation of quality for dry and wet fingers. The proposed Moisture Indication method was fast to compute and successfully used as a binary classifier to detect the extreme moisture cases and indicate if the skin was too dry or too wet. In case of wet fingers the error rates were 10% False Positive and 5% False Negative. Dry fingers classification gave a False Positive error rate of 30% and False Negative error rate of 8%. The Moisture Indication method used as dryness and wetness detection could also be incorporated in ISO/IEC 29794-4 [ISO10] and in NFIQ 2.0.

8.1 Future work

This project analysed the impact of fingertip skin moisture on the acquired fingerprint sample quality. As shown in figure 2.3 and discussed in section 3.6, moisture is not the only factor that causes fingerprint sample degradation.

Following the work underlying this thesis, it would be useful to analyse the impact of other factors, i.e. finger character, finger pressure on the sensor, finger placement and rotation; and most importantly the inter-relation between all these factors combined.

This would allow to propose more methods indicating the causes of sample quality degradation and improve their prevision, leading to better standards in fingerprint quality analysis and to better performance of biometric systems.

APPENDIX A

Calculated comparison scores

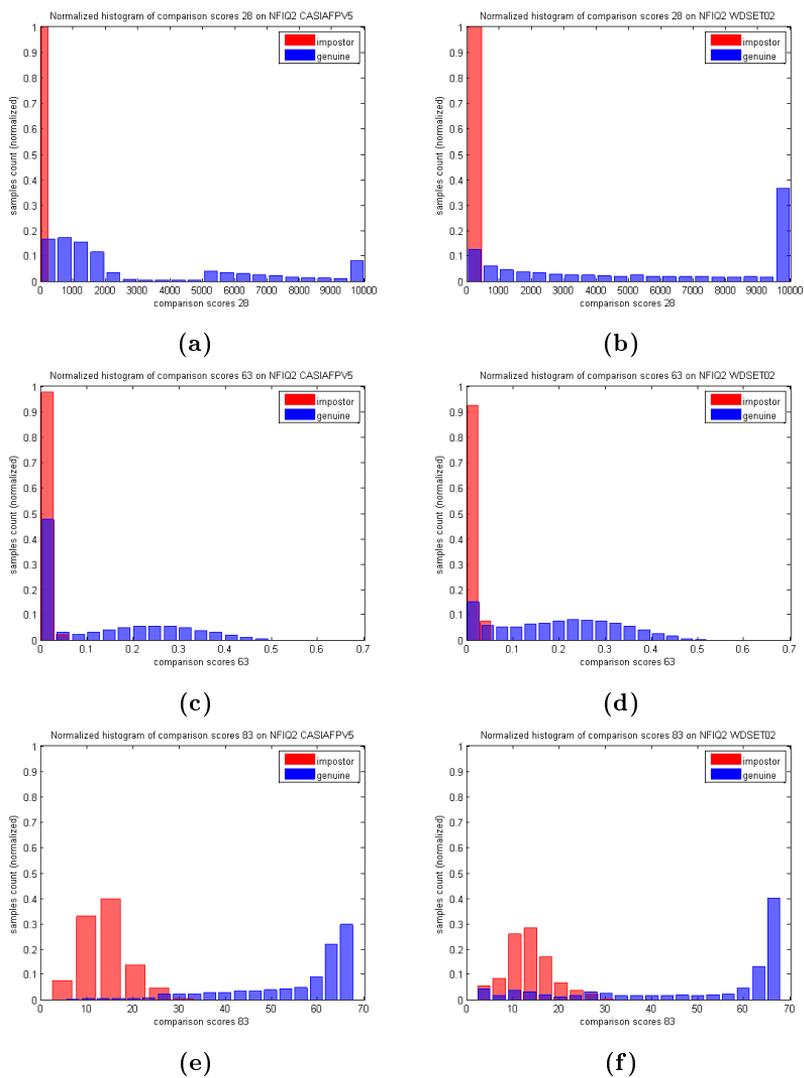


Figure A.1: Genuine and impostor comparison score distribution histograms:

- a) CASIA dataset with comparator 28
- c) CASIA dataset with comparator 63
- e) CASIA dataset with comparator 83
- b) collected dataset with comparator 28
- d) collected dataset with comparator 63
- f) collected dataset with comparator 83.

APPENDIX B

Error versus Reject Curves

	c28	c63	c83	c avg	n28	n63	n83	n avg	nc avg	NIST	all avg
NFIQ	1.98	2.76	1.97	2.24	1.77	2.13	2.2	2.03	2.14	-	2.14
GSH	1.69	2.8	1.72	2.07	1.57	1.18	0.94	1.23	1.65	3.36	2.51
RVU	1.89	2.8	1.89	2.19	1.2	1.39	1.48	1.36	1.78	2.96	2.37
OCL	2.03	2.85	1.95	2.28	2.18	2.76	2.84	2.59	2.44	2.54	2.49
LCS	2.15	2.93	2.22	2.43	1.74	1.67	1.4	1.6	2.02	2.36	2.19
RLCM	1.6	2.65	1.66	1.97	1.49	0.82	0.65	0.99	1.48	2.65	2.07
RLCE	1.72	2.55	1.71	1.99	1.11	0.73	0.51	0.78	1.39	2.64	2.02
RLCNG	1.67	2.57	1.64	1.96	1.38	1.01	0.85	1.08	1.52	2.75	2.14
RVDM	1.75	2.75	1.82	2.11	2.65	2.67	2.77	2.7	2.41	3.1	2.76
RVDE	2.49	2.74	2.63	2.62	2.16	2	1.92	2.03	2.33	3.05	2.69
RVDNG	1.9	2.67	1.93	2.17	2.51	2.69	2.72	2.64	2.41	3.1	2.76
CNTM	1.86	2.87	1.97	2.23	2.95	2.51	2.63	2.7	2.47	2.98	2.73
CNTE	2.9	2.74	2.95	2.86	1.86	2.24	2.07	2.06	2.46	3.5	2.98
CNTNG	1.7	2.7	1.78	2.06	1.86	1.7	1.46	1.67	1.87	2.97	2.42

Table B.1: QMA ranking based on the ERC analysis for all three comparators and both datasets: CASIA and the (new) collected dataset. Additionally, column NIST shows ERC results of QMA execution on the subset of the NIST operational dataset (see section 6.7.3.2). For brevity, columns for CASIA and comparators 28, 63 and 83 are named c28, c63 and c83 respectively. Similarly for the new, collected dataset – n28, n63 and n83. Column c avg is the average on CASIA and n avg on the collected dataset. Average from CASIA and the collected dataset is in column nc avg; the last column is the total average, including the NIST operational dataset.

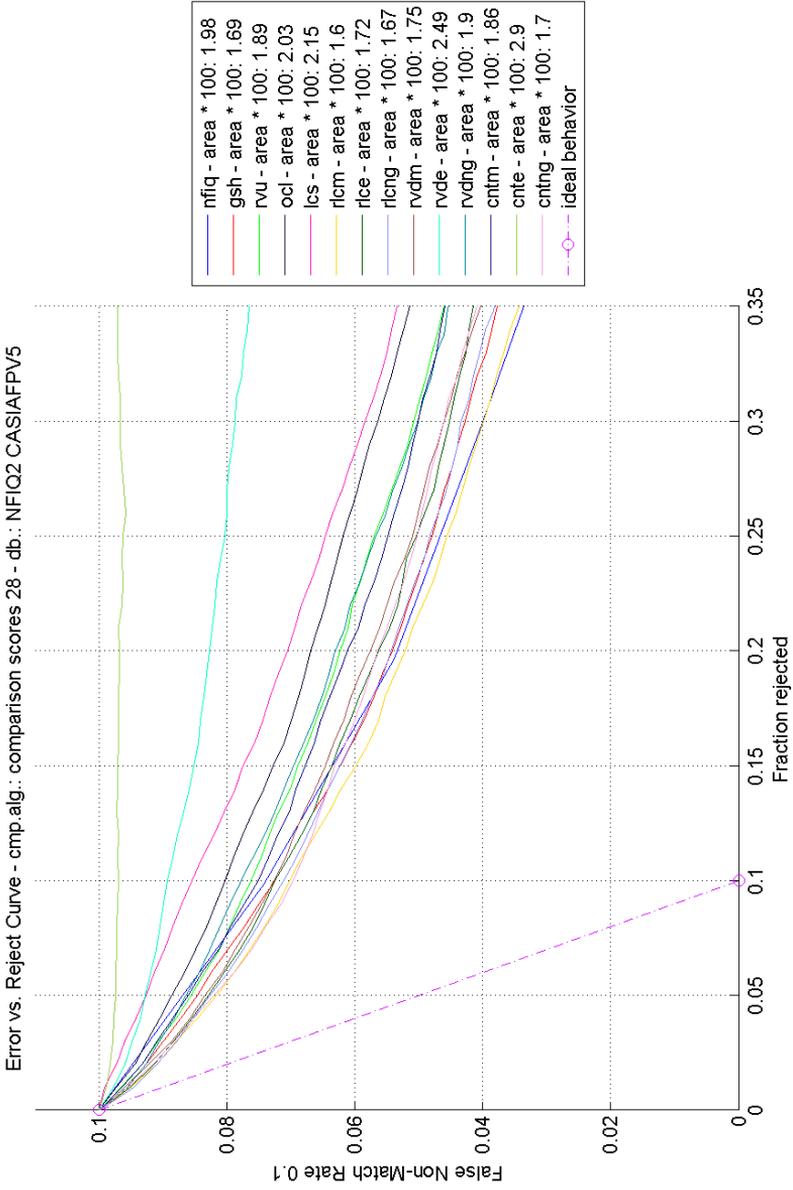


Figure B.1: Error versus Reject Curve plot for comparator 28 on the CASIA dataset.

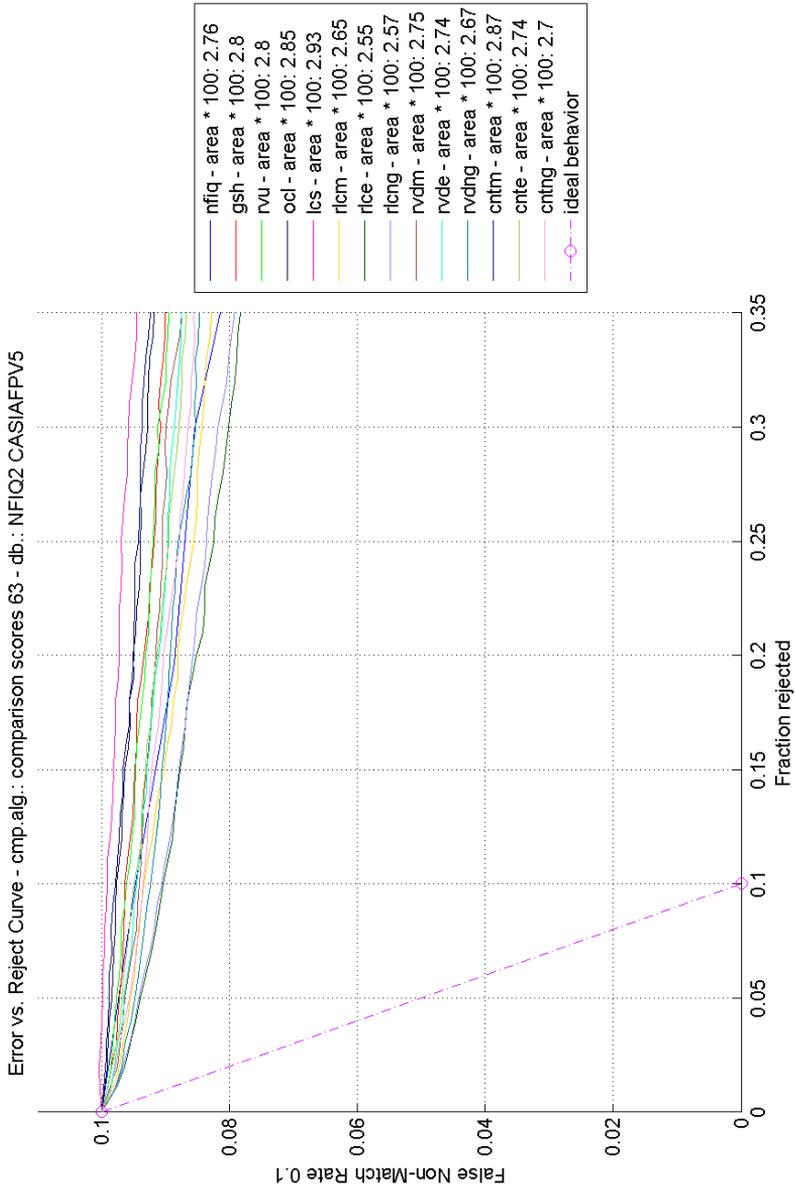


Figure B.2: Error versus Reject Curve plot for comparator 63 on the CASIA dataset.

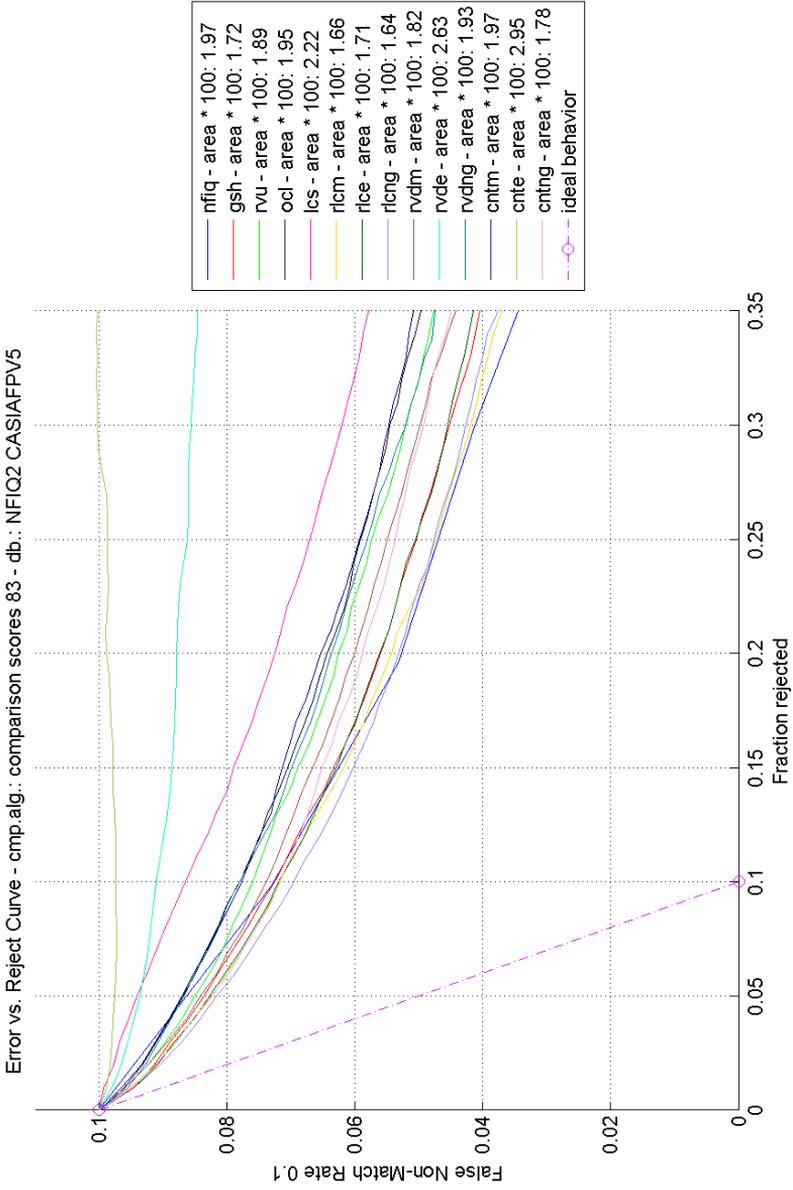


Figure B.3: Error versus Reject Curve plot for comparator 83 on the CASIA dataset.

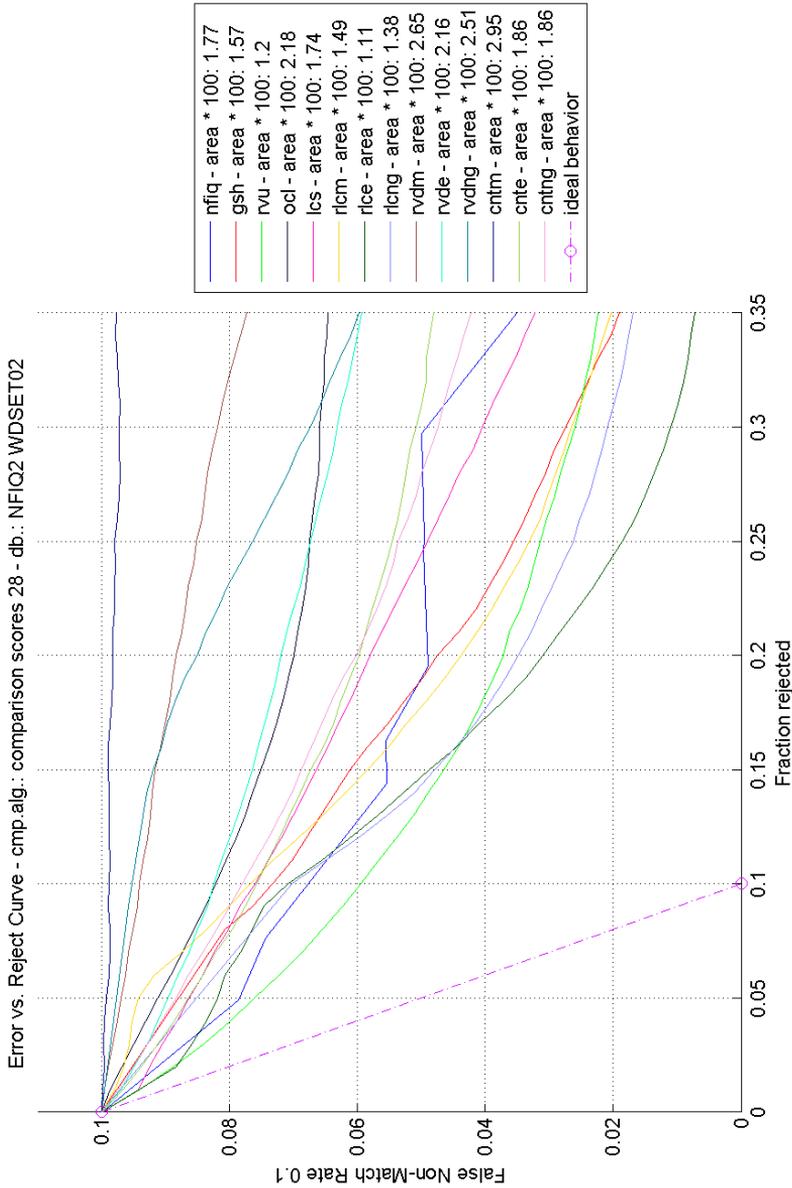


Figure B.4: Error versus Reject Curve plot for comparator 28 on the collected dataset.

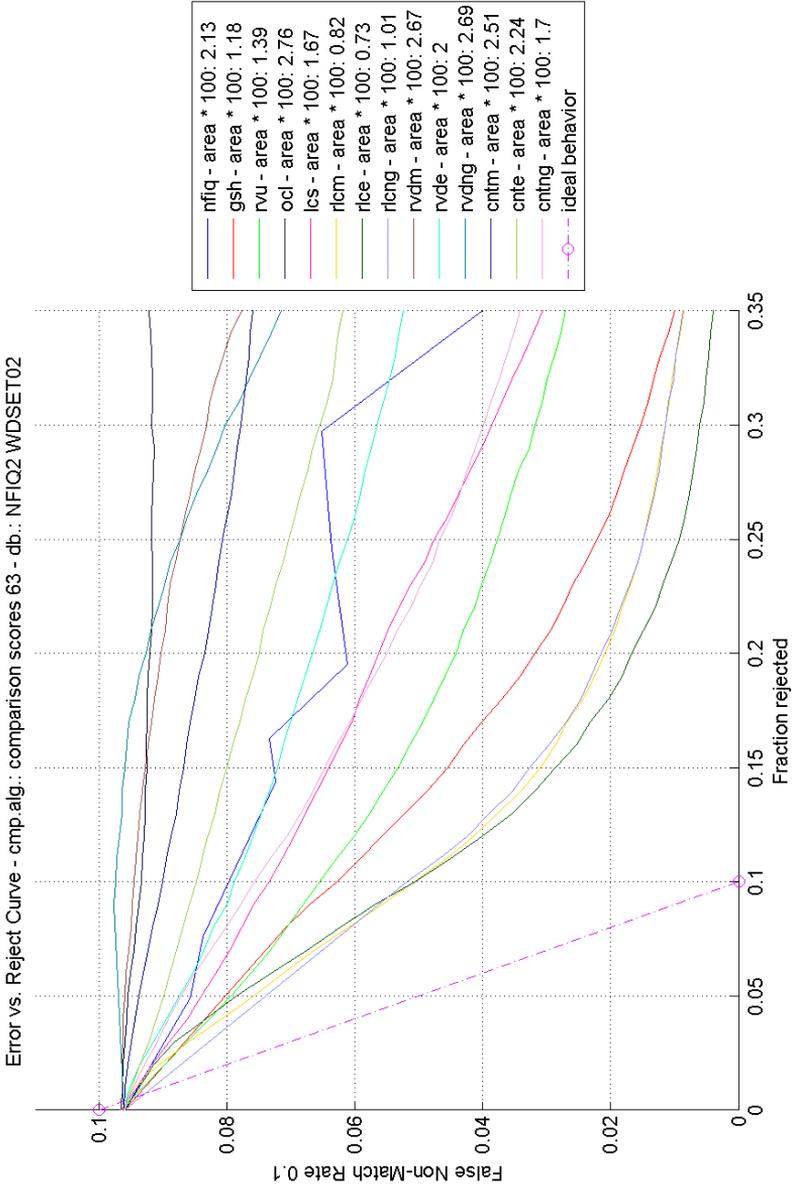


Figure B.5: Error versus Reject Curve plot for comparator 63 on the collected dataset.

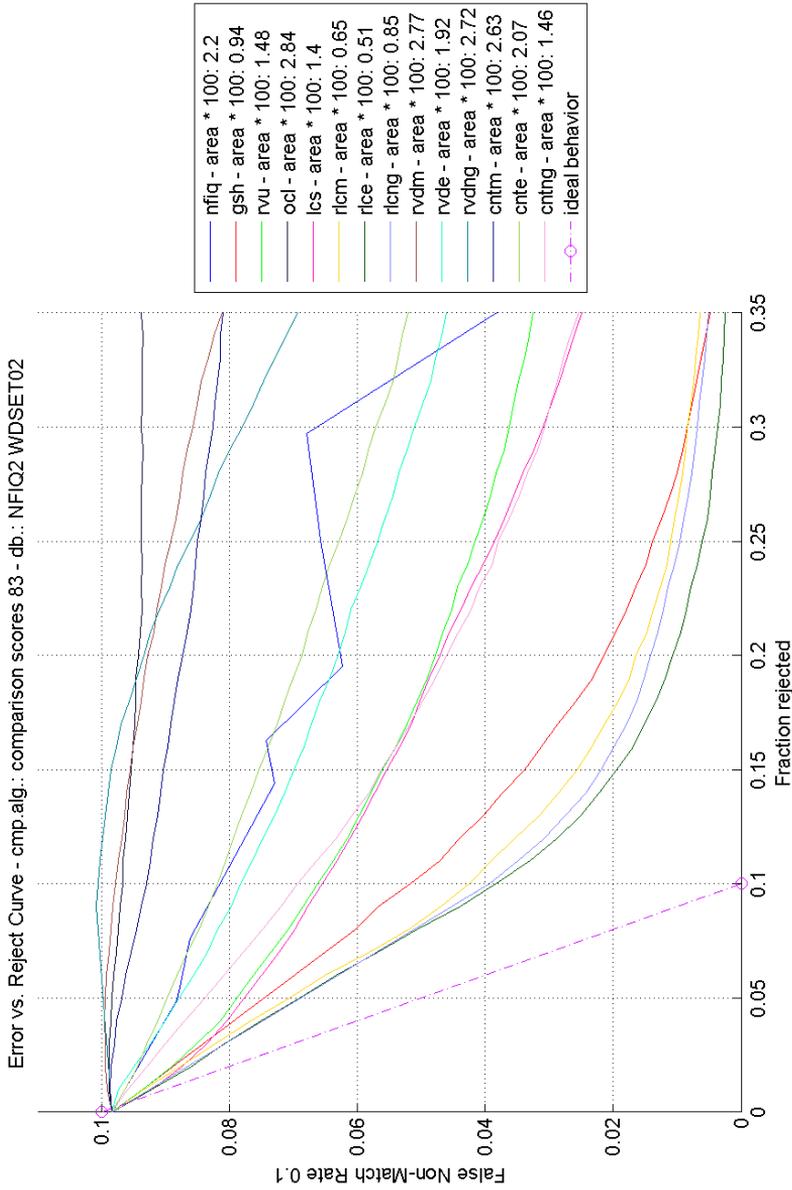


Figure B.6: Error versus Reject Curve plot for comparator 83 on the collected dataset.

APPENDIX C

Quality and Utility correlation

casia	nfq	gsh	rvu	ocl	lcs	rlcm	rlce	rlcng	rvdm	rvde	rvdng	cntm	cnte	cntng	u28	u63	u83
nfq	100	-67	59	-60	-56	-70	-66	-67	-66	-29	-63	-56	-7	-66	-54	-19	-44
gsh	-67	100	-67	67	76	76	62	65	80	23	74	69	0	74	54	12	39
rvu	59	-67	100	-63	-60	-70	-53	-57	-67	-12	-57	-59	7	-57	-45	-10	-35
ocl	-60	67	-63	100	72	69	38	46	63	3	43	69	-33	44	41	9	34
lcs	-56	76	-60	72	100	65	38	41	72	23	59	68	-7	59	43	7	28
rlcm	-70	76	-70	69	65	100	78	80	84	29	71	77	-4	72	51	18	40
rlce	-66	62	-53	38	38	78	100	95	67	41	77	42	39	75	55	26	41
rlcng	-67	65	-57	46	41	80	95	100	67	28	73	44	26	70	55	23	42
rvdm	-66	80	-67	63	72	84	67	67	100	21	89	77	0	81	49	12	38
rvde	-29	23	-12	3	23	29	41	28	21	100	35	29	51	50	25	14	12
rvdng	-63	74	-57	43	59	71	77	73	89	35	100	56	30	87	51	15	38
cntm	-56	69	-59	69	68	77	42	44	77	29	56	100	-36	71	39	6	30
cnte	-7	0	7	-33	-7	-4	39	26	0	51	30	-36	100	21	13	14	2
cntng	-66	74	-57	44	59	72	75	70	81	50	87	71	21	100	53	16	38
util28	-54	54	-45	41	43	51	55	55	49	25	51	39	13	53	100	44	73
util63	-19	12	-10	9	7	18	26	23	12	14	15	6	14	16	44	100	26
util83	-44	39	-35	34	28	40	41	42	38	12	38	30	2	38	73	26	100

Table C.1: Spearman correlation of quality and utility for comparators 28, 63 and 83 on the CASIA v5.0 dataset.

new	nfq	gsh	rvu	ocl	lcs	rlcm	rlce	rlcng	rvdm	rvde	rvdng	cntm	cnte	cntng	u28	u63	u83
nfq	100	-68	55	-45	-61	-58	-65	-51	-42	-29	-39	-30	-8	-41	-64	-65	-57
gsh	-68	100	-52	55	82	63	68	57	76	28	65	57	-11	62	62	60	56
rvu	55	-52	100	-19	-48	-50	-48	-45	-16	-10	-16	-4	-17	-25	-67	-68	-56
ocl	-45	55	-19	100	59	33	27	31	59	9	55	46	-26	44	35	30	27
lcs	-61	82	-48	59	100	52	51	57	72	21	67	48	-8	64	62	56	54
rlcm	-58	63	-50	33	52	100	69	66	56	2	46	35	-3	44	39	49	35
rlce	-65	68	-48	27	51	69	100	57	48	52	42	29	24	43	60	57	49
rlcng	-51	57	-45	31	57	66	57	100	51	14	74	17	21	83	45	48	41
rvdm	-42	76	-16	59	72	56	48	51	100	2	82	66	-29	66	29	28	26
rvde	-29	28	-10	9	21	2	52	14	2	100	13	21	25	27	37	27	32
rvdng	-39	65	-16	55	67	46	42	74	82	13	100	44	-4	89	32	28	28
cntm	-30	57	-4	46	48	35	29	17	66	21	44	100	-73	40	15	14	17
cnte	-8	-11	-17	-26	-8	-3	24	21	-29	25	-4	-73	100	3	20	20	18
cntng	-41	62	-25	44	64	44	43	83	66	27	89	40	3	100	38	35	37
u28	-64	62	-67	35	62	39	60	45	29	37	32	15	20	38	100	81	80
u63	-65	60	-68	30	56	49	57	48	28	27	28	14	20	35	81	100	81
u83	-57	56	-56	27	54	35	49	41	26	32	28	17	18	37	80	81	100

Table C.2: Spearman correlation of quality and utility for comparators 28, 63 and 83 on the collected (new) dataset.

APPENDIX D

Moisture impact on quality

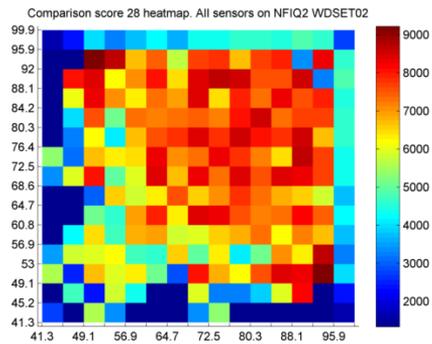


Figure D.1: Moisture impact on comparison scores heat map. Comparator 28 on all sensors.

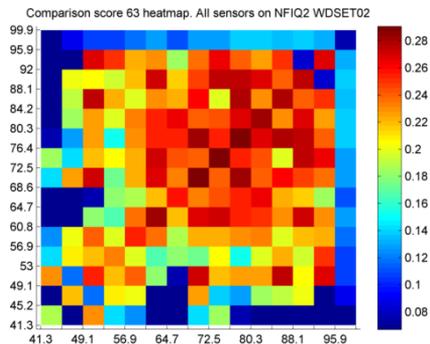


Figure D.2: Moisture impact on comparison scores heat map. Comparator 63 on all sensors.

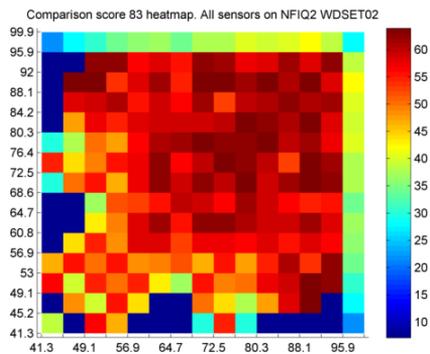


Figure D.3: Moisture impact on comparison scores heat map. Comparator 83 on all sensors.

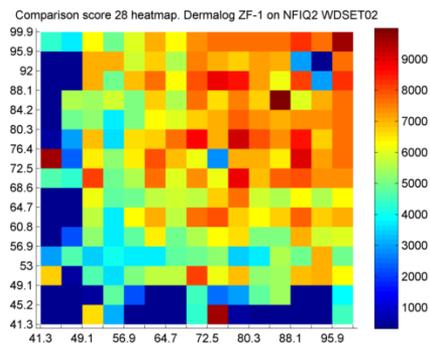


Figure D.4: Moisture impact on comparison scores heat map. Comparator 28 on Dermalog ZF-1.

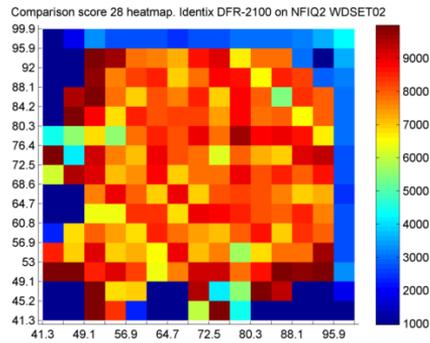


Figure D.5: Moisture impact on comparison scores heat map. Comparator 28 on L-1 DFR-2100.

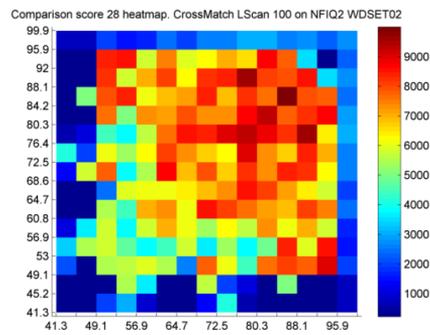


Figure D.6: Moisture impact on comparison scores heat map. Comparator 28 on CrossMatch LScan 100.

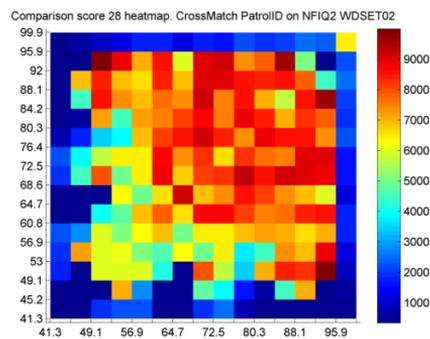


Figure D.7: Moisture impact on comparison scores heat map. Comparator 28 on CrossMatch Patrol ID.

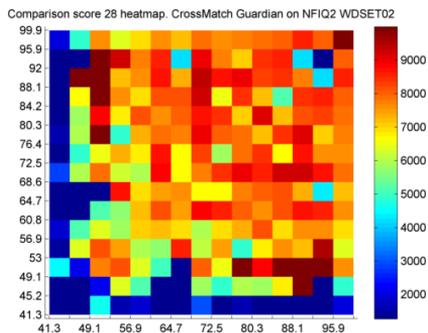


Figure D.8: Moisture impact on comparison scores heat map. Comparator 28 on CrossMatch Guardian.

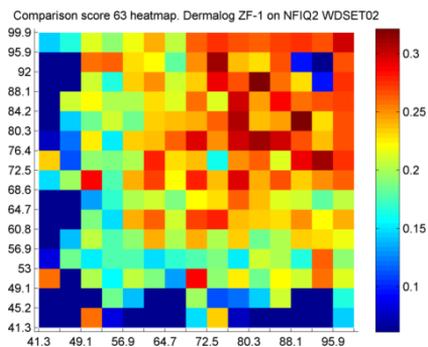


Figure D.9: Moisture impact on comparison scores heat map. Comparator 28 on Dermalog ZF-1.

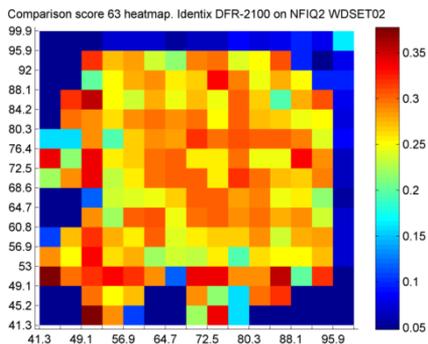


Figure D.10: Moisture impact on comparison scores heat map. Comparator 28 on L-1 DFR-2100.

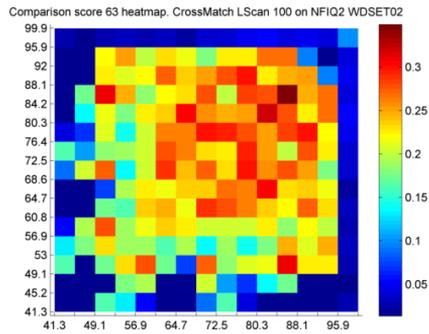


Figure D.11: Moisture impact on comparison scores heat map. Comparator 28 on CrossMatch LScan 100.

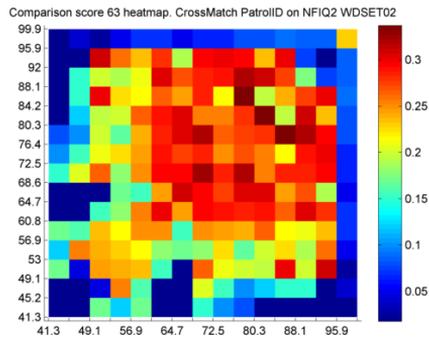


Figure D.12: Moisture impact on comparison scores heat map. Comparator 28 on CrossMatch Patrol ID.

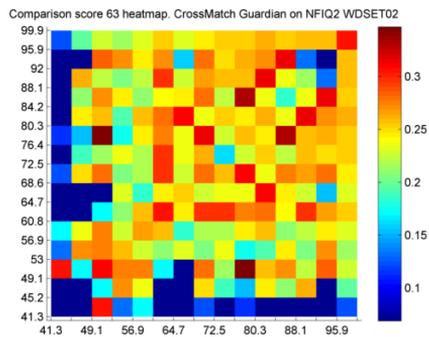


Figure D.13: Moisture impact on comparison scores heat map. Comparator 28 on CrossMatch Guardian.

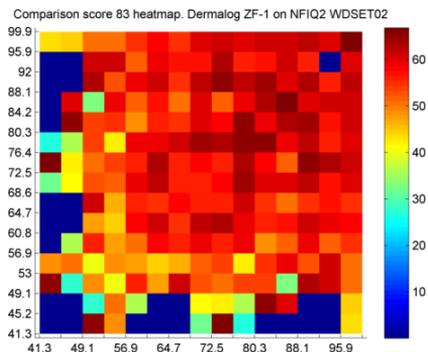


Figure D.14: Moisture impact on comparison scores heat map. Comparator 28 on Dermalog ZF-1.

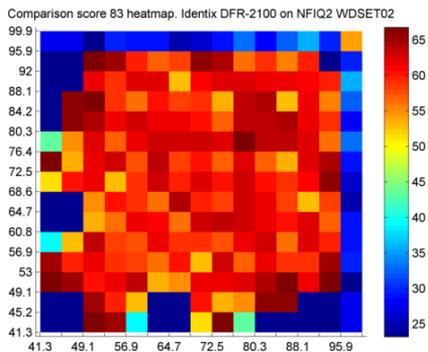


Figure D.15: Moisture impact on comparison scores heat map. Comparator 28 on L-1 DFR-2100.

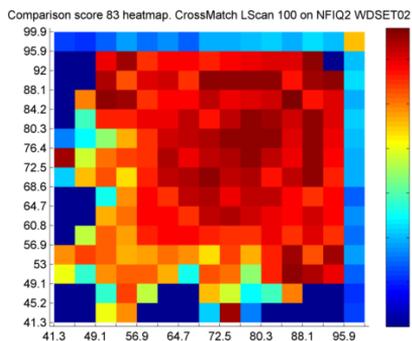


Figure D.16: Moisture impact on comparison scores heat map. Comparator 28 on CrossMatch LScan 100.

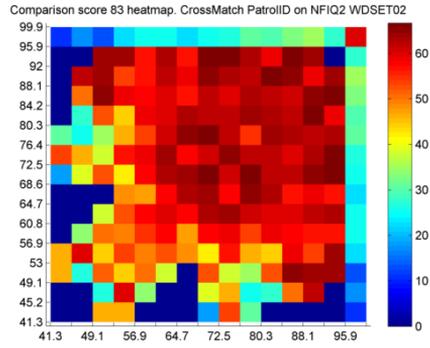


Figure D.17: Moisture impact on comparison scores heat map. Comparator 28 on CrossMatch Patrol ID.

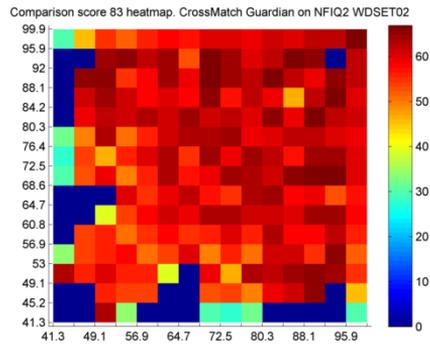


Figure D.18: Moisture impact on comparison scores heat map. Comparator 28 on CrossMatch Guardian.

APPENDIX E

Moisture impact on quality – sensor differences

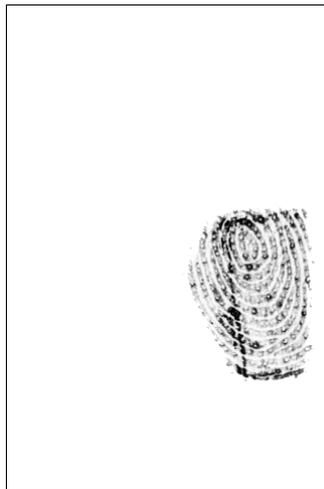


Figure E.1: Sample acquired with Dermalog ZF-1 from a dry finger (measured moisture 57.6%) that features only a very small portion of the ridges. The border of the sample image is painted for better reference.

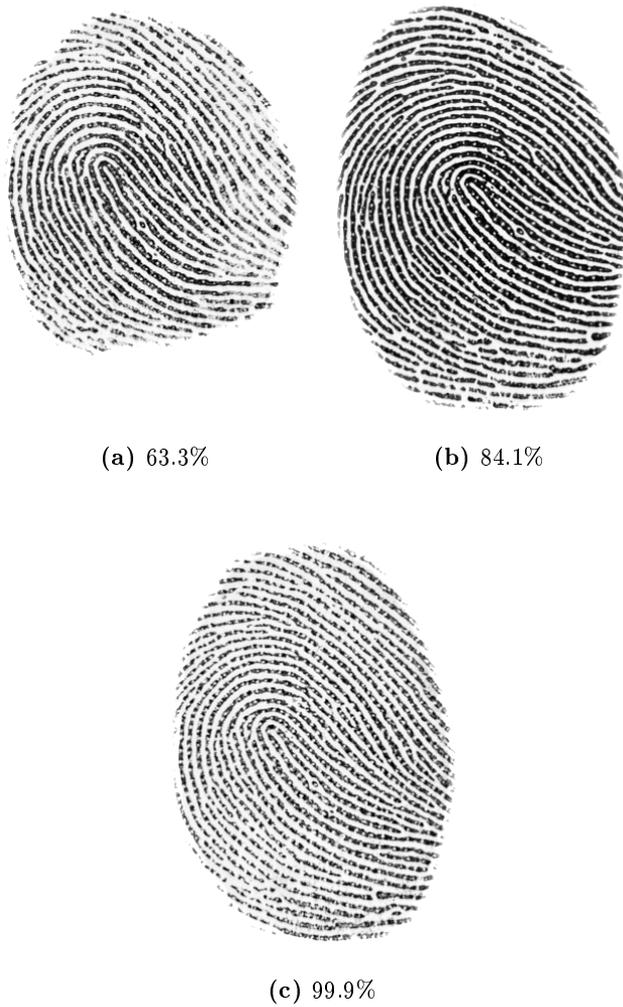


Figure E.2: Fingerprint samples collected from the same finger with different moisture levels using Dermalog ZF-1.

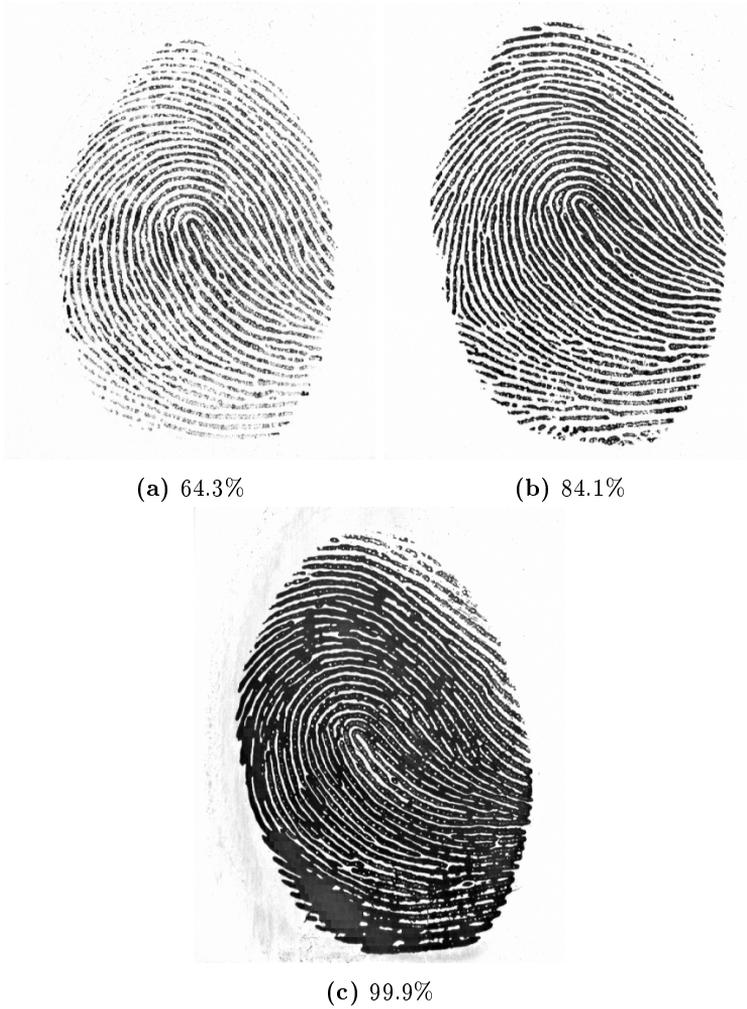


Figure E.3: Fingerprint samples collected from the same finger with different moisture levels using L-1 DFR-2100.



Figure E.4: Fingerprint samples collected from the same finger with different moisture levels using CrossMatch LScan 100.



(a) 64.3%

(b) 84.1%



(c) 99.9%

Figure E.5: Fingerprint samples collected from the same finger with different moisture levels using CrossMatch Patrol ID.

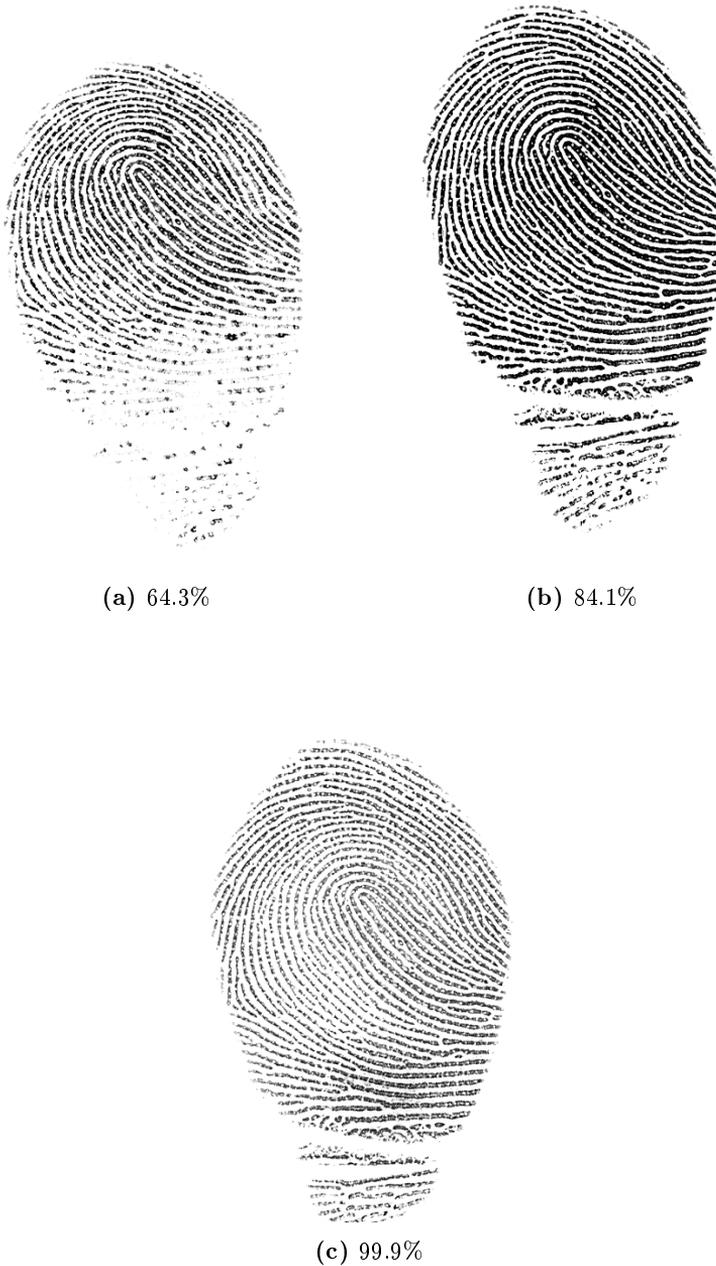


Figure E.6: Fingerprint samples collected from the same finger with different moisture levels using CrossMatch Guardian.

APPENDIX F

Witness classification

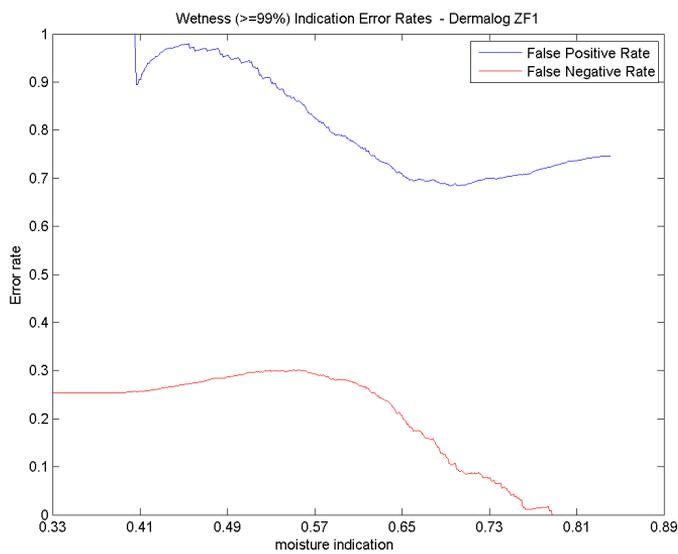


Figure F.1: False Positive and False Negative wetness detection error rates for samples acquired using Dermalog ZF-1. The False Positive Rate is very high since this sensor produces good samples even with wet fingers.

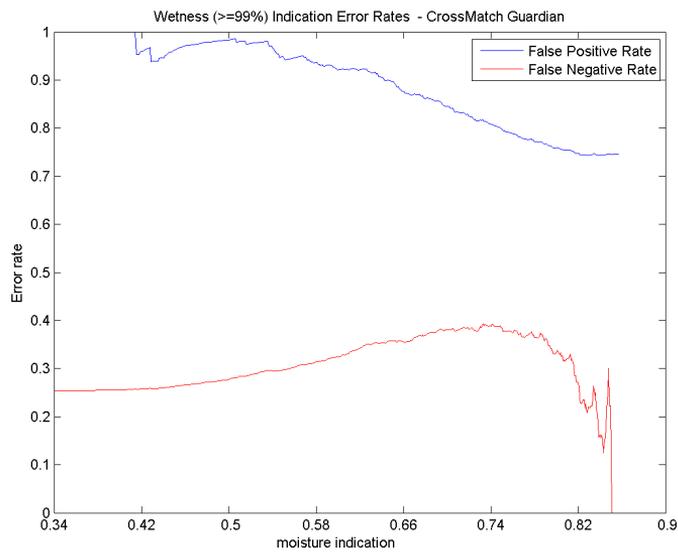


Figure F.2: False Positive and False Negative wetness detection error rates for samples acquired using CrossMatch Guardian. The False Positive Rate is very high since this sensor produces good samples even with wet fingers.

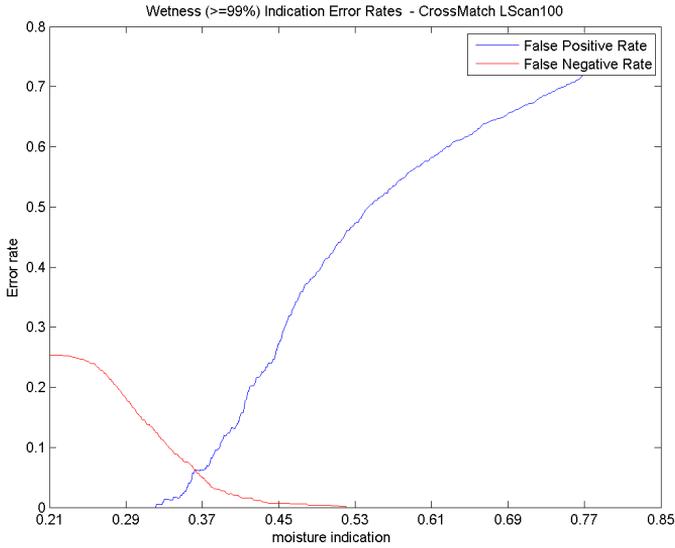


Figure F.3: False Positive and False Negative wetness detection error rates for samples acquired using CrossMatch LScan100.

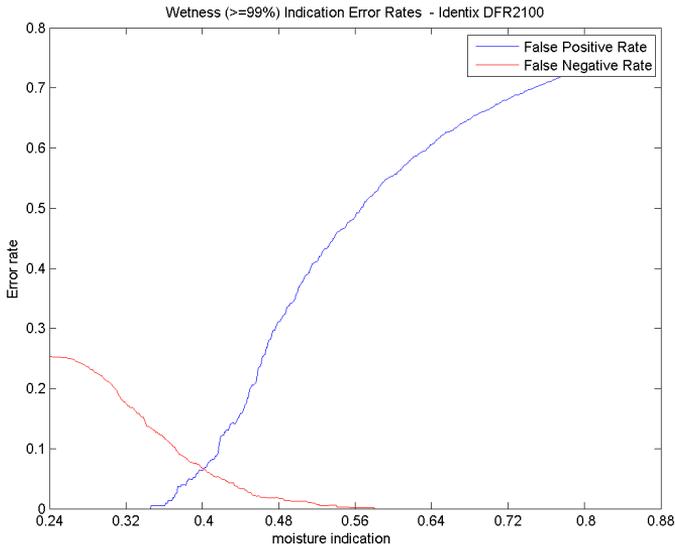


Figure F.4: False Positive and False Negative wetness detection error rates for samples acquired using L-1 DFR-2100.

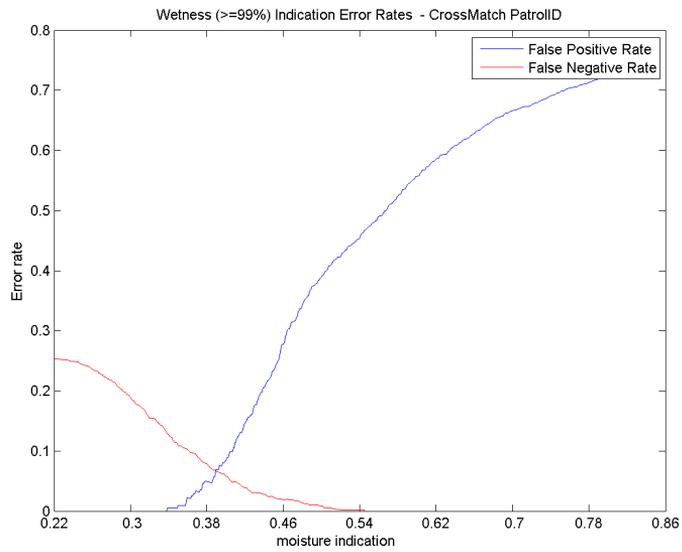


Figure F.5: False Positive and False Negative wetness detection error rates for samples acquired using CrossMatch Patrol ID.

APPENDIX G

Dryness classification

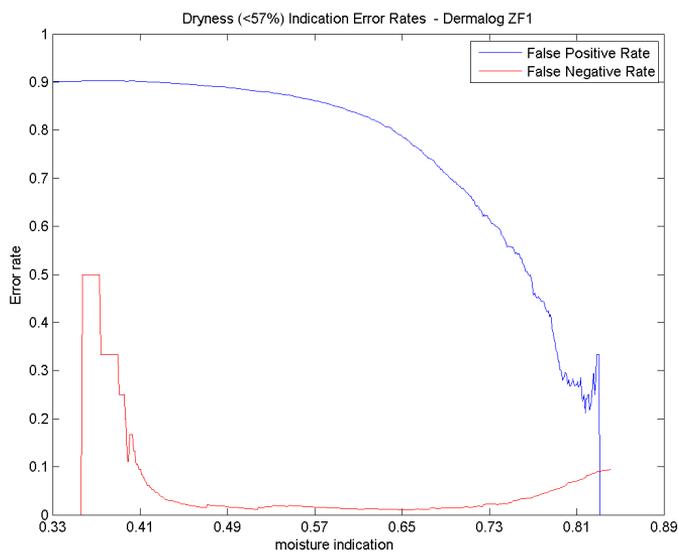
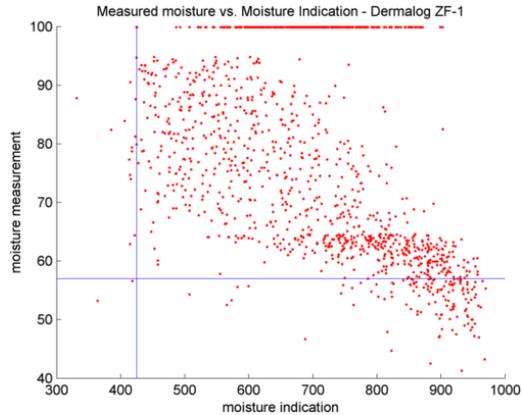


Figure G.1: False Positive and False Negative dryness detection error rates for samples acquired using Dermalog ZF-1. High False Negative indication for low threshold values is caused by two outliers as shown in figure G.2.



(a) Measured Moisture versus Moisture Indication point cloud for samples acquired using Dermalog ZF-1.



Figure G.2: Pointcloud of measured moisture versus Moisture Indication (a) with samples acquired using Dermalog ZF-1. Two outliers below horizontal blue line and to the left of vertical blue line. Two fingerprint samples (b, c) from the point cloud plot having very low moisture measurement (56.6% and 53.2% respectively) but a wet impression – leading to high False Negative rate for low thresholds (see figure G.1). A possible explanation is that the subject used oily lotion before the acquisition, which would not lead to high measurement but cause a wet impression, as described in section 6.5.1.

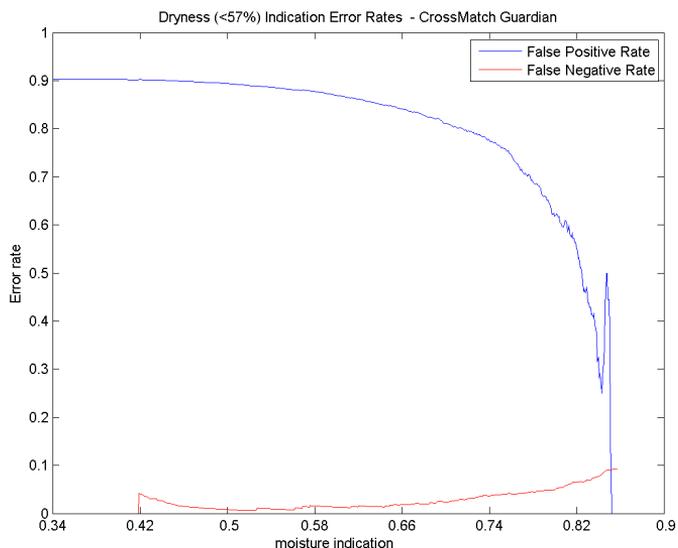


Figure G.3: False Positive and False Negative dryness detection error rates for samples acquired using CrossMatch Guardian.

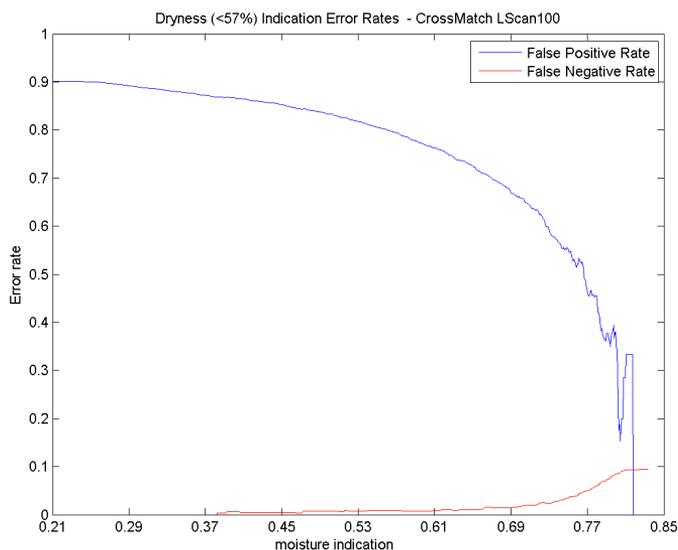


Figure G.4: False Positive and False Negative dryness detection error rates for samples acquired using CrossMatch LScan100.

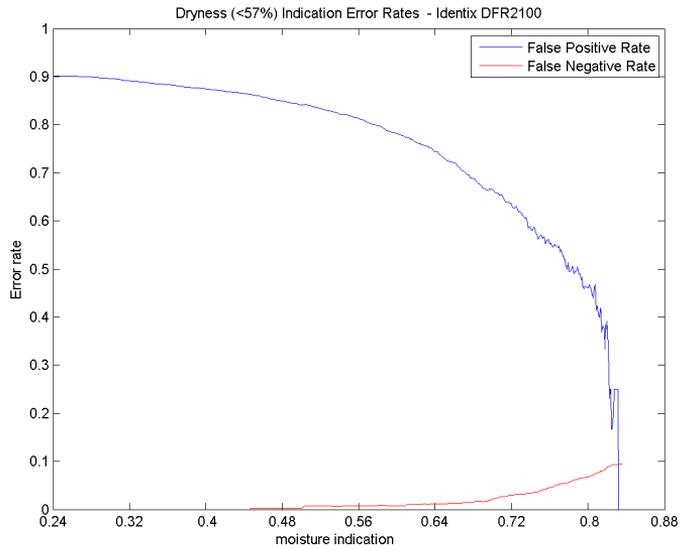


Figure G.5: False Positive and False Negative dryness detection error rates for samples acquired using L-1 DFR-2100.

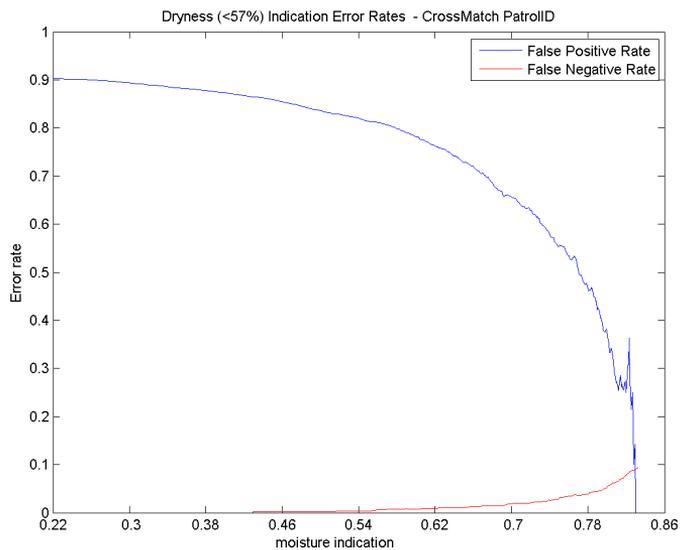


Figure G.6: False Positive and False Negative dryness detection error rates for samples acquired using CrossMatch Patrol ID.

Bibliography

- [aad] Aadhaar, Unique Identification Authority of India. <http://uidai.gov.in/>.
- [AFFOG⁺07] F. Alonso-Fernandez, J. Fierrez, J. Ortega-Garcia, J. Gonzalez-Rodriguez, H. Fronthaler, K. Kollreider, and J. Bigun. A comparative study of fingerprint image-quality estimation methods. *Information Forensics and Security, IEEE Transactions on*, 2(4):734–743, 2007.
- [ANN⁺] Esko Alanen, Jouni Nuutinen, Kirsi Nicklen, Tapani Lahtinen, and Jukka Moenkkonen.
- [bio] Biosecure network of excellence. <http://biosecure.it-sudparis.eu/>.
- [BKA] BKA, Bundeskriminalamt.
- [BSI] BSI, Federal Office for Information Security.
- [CASA] Center for Advanced Security Research Darmstadt.
- [casb] Chinese Academy of Sciences, Institute of Automation (CASIA). <http://biometrics.idealtest.org/>.
- [CB12] Clijsen R. Taeymans J. Clarys, P. and A. O. Barel. Hydration measurements of the stratum corneum: comparison between the capacitance method (digital version of the corneometer cm 825®) and the impedance method (skicon-200ex®). *Skin Research and Technology*, 18:316–323, 2012.

- [cro] Cross match technologies, inc. <http://www.crossmatch.com/>.
- [der] Dermalog identification systems gmbh. <http://www.dermalog.de/en/>.
- [FaOgTtGr07] Julian Fierrez-aguilar, Javier Ortega-garcia, Doroteo Torroledano, and Joaquin Gonzalez-rodriguez. Biosec baseline corpus: A multimodal biometric database. *Pattern Recognition*, pages 1389–1392, 2007.
- [Fra] Fraunhofer IGD.
- [GT07] Patrick Grother and Elham Tabassi. Performance of biometric quality measures. *IEEE Trans. Pattern Anal. Mach. Intell.*, pages 531–543, 2007.
- [hda] Hochschule Darmstadt - University of Applied Sciences.
- [ica] International civil aviation organization.
- [ISO09] Information technology – biometric sample quality – part 1: Framework, 2009.
- [ISO10] Information technology – biometric sample quality – part 4: Finger image data, 2010.
- [ISO11] Information technology – security techniques – biometric information protection, 2011.
- [ISO12] Information technology – Vocabulary – Part 37: Biometrics, 2012.
- [LJY02] E. Lim, Xudong Jiang, and WeiYun Yau. Fingerprint quality and validity analysis. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 1, pages I–469–I–472 vol.1, 2002.
- [lum07] Spoof detection schemes, white paper, 2007.
- [Meh93] B.M. Mehtre. Fingerprint image analysis for automatic identification. *Machine Vision and Applications*, 6(2-3):124–139, 1993.
- [MMJP09] Davide Maltoni, Dario Maio, Anil K. Jain, and Salil Prabhakar. *Handbook of Fingerprint Recognition*. Springer Publishing Company, Incorporated, 2nd edition, 2009.
- [mor] Safran morphotrust. <http://www.morphotrust.com/>.
- [nfia] Nfiq 2.0: Evaluation of potential image features for quality assessment. <http://www.nist.gov/>.

- [nfiq] Nfiq 2.0: Quality feature definitions. <http://www.nist.gov/>.
- [nfic] NIST Biometric Image Software. <http://www.nist.gov/itl/iad/ig/nbis.cfm>.
- [NIS] National Institute of Standards and Technology, U.S. Department of Commerce.
- [OGFAS⁺03] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez-Zanuy, V. Espinosa, A. Satue, I. Hernaez, J.-J. Igarza, C. Vivaracho, D. Escudero, and Q.-I. Moro. Mcty baseline corpus: a bimodal biometric database. *Vision, Image and Signal Processing, IEE Proceedings -*, 150(6):395–401, 2003.
- [Ols13] Martin Olsen. Candidate features for quality assessment, April 2013. Winchester U.K.
- [oST] National Institute of Standards and U.S. Department of Commerce Technology. Development of nfiq 2.0.
- [Ots79] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1):62–66, January 1979.
- [Pea01] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(6):559–572, 1901.
- [sec] secunet Security Networks AG.
- [ske] Face Care, Shenzhen Kier Electronic Apparatus Factory. <http://www.szkie.com/kier/en/about.asp>.
- [SKK01] LinLin Shen, Alex ChiChung Kot, and Wai Mun Koo. Quality measures of fingerprint images. In *Proceedings of the Third International Conference on Audio- and Video-Based Biometric Person Authentication, AVBPA '01*, pages 266–271, London, UK, UK, 2001. Springer-Verlag.
- [stq] Standarization testing and quality certification directorate, department of electronics and information technology, ministry of communications and it, india. <http://www.stqc.gov.in/>.
- [TW05] Elham Tabassi and Charles L. Wilson. A novel approach to fingerprint image quality. In *ICIP (2)*, pages 37–40. IEEE, 2005.
- [uid] Biometrics design standards for uid applications.
- [usv] US-VISIT, U.S. Department of Homeland Security.

- [XYYP08] Shan Juan Xie, Ju Cheng Yang, Sook Yoon, and Dong-Sun Park. An optimal orientation certainty level approach for fingerprint quality estimation. In *Intelligent Information Technology Application, 2008. IITA '08. Second International Symposium on*, volume 3, pages 722–726, 2008.