# VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

## FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
## ÚSTAV INTELIGENTNÍCH SYSTÉMŮ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF INTELLIGENT SYSTEMS

# HODNOCENÍ KVALITY OTISKŮ PRSTŮ
FINGERPRINT IMAGE QUALITY ASSESSMENT

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

AUTOR PRÁCE                                    VLADIMÍR ŠMIDA
AUTHOR

VEDOUCÍ PRÁCE              doc. Ing. MARTIN DRAHANSKÝ, Ph.D.
SUPERVISOR

BRNO 2011

# Abstrakt

Kritickým prvkem biometrického systému pro rozpoznávání otisků prstů je proces snímání. Kvalita snímku totiž ovlivňuje všechny další části systému počínaje zpracováním obrazu, přes extrakci rysů až po samotné rozhodnutí. Přestože bylo navrženo několik metod určování kvality snímku, chybějící formální specifikace kvality otisku nedovoluje ověřit jejich přesnost. Tato diplomová práce se zabývá hodnocením metod určujících kvalitu biometrického signálu otisku prstu. Popisuje jednotlivé faktory ovlivňující kvalitu spolu se současnými přístupy používanými pro její odhad. V práci je rovněž vysvětlena evaluační technika navržená za účelem porovnání schopnosti jednotlivých metod předpovědět výkon biometrického systému. Několik metod pro odhad kvality bylo naimplementováno a ohodnoceno touto technikou.

# Abstract

The crucial factor of fingerprint recognition is image acquisition. Quality of the captured fingerprint image influences all other components of the recognition system, from image processing, through feature extraction to decision making. Therefore, to retain the security of the overall biometric system, it is necessary to determine the quality and validity of the captured fingerprint image. Although several quality estimation techniques have been proposed, a missing general definition of fingerprint quality makes it difficult to verify them. This thesis deals with the problem of evaluation of the fingerprint signal quality assessment methods. Several factors that influence the quality of fingerprints are described. The state-of-the-art approaches for assessing fingerprint quality are reviewed and implemented. An evaluation method based on the biometric performance is suggested in order to determine the capability to predict the performance of the reviewed methods.

# Klíčová slova

biometrie, rozpoznávání otisků prstů, zpracování obrazu, kvalita signálu otisku prstu, výkon biometrického systému

# Keywords

biometrics, fingerprint recognition, image processing, fingerprint image signal quality, biometric performance

# Citace

Vladimír Šmida: Fingerprint Image Quality Assessment, diplomová práce, Brno, FIT VUT v Brně, 2011

# Fingerprint Image Quality Assessment

## Prohlášení

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně pod vedením pana doc. Ing. Dipl.-Ing. Martina Drahanského, Ph.D.
Další informace mi poskytli M.Sc. Martin Aastrup Olsen a Prof. Dr. Christoph Busch.
Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

. . . . . . . . . . . . . . . . . . . . . .
Vladimír Šmida
May 24, 2011

## Poděkování

Foremost I would like to thank to my consultant prof. Dr. Christoph Busch. His lectures I attended during my stay at the Gjøvik University College in Norway raised my interest in biometric research and lead me to this master's thesis. Moreover, I'm grateful to him for giving me the oportunity to join the ongoing research at CASED[1] in Darmstadt, Germany. During this period I benefited from his encouraging philosophy of research work, inspiring discussions and valuable advice.
I also thank to M.Sc. Martin Aastrup Olsen for all the interesting discussions we had, for his recommendations and comments.
I want to thank to my supervisor doc. Ing. Dipl.-Ing. Martin Drahanský, Ph.D. for reviewing this thesis, for his wise comments and suggestions.
A special thank belongs to Barbora Micenková for all the inspiration and encouragement.

*Venované rodine, za jej neutíchajúcu pomoc a podporu pri štúdiu.*

---

[1]Center for Advanced Security Research Darmstadt, http://www.cased.de.

# Contents

# Chapter 1

# Introduction

*"Are you who you claim to be?"* In the effort to fight with identity fraud, organized crime, illegal immigration and other security risks, biometric technology has been increasingly employed. Among others, it has been integrated into identity cards, driving licences and ATM's to prevent them from unauthorized use. From all biometric characteristics, fingerprints are still the most widespread for several reasons. They are unique and permanent, easy to capture and highly reliable. That is why they have been used in criminal prosecution for centuries. Nowadays, their field of application is much wider. Fingerprint recognition has been adopted for physical access control, time and attendance control, e-commerce, transaction authentication and also for border control.

The most evident is the utilization of fingerprints in the US Visit program and, recently, also in the biometric passports used by European Union and all member states of ICAO[1]. As fingerprints are used by wider population, the databases are still growing. For example, the US Visit fingerprint database consists of over 90 million samples by this time [37]. For such large datasets, more reliable recognition algorithms are needed.

Fingerprint recognition systems are able to achieve high accuracy as long as the fingerprint quality is high [42]. Fingerprint images of low quality carry an insufficient amount of information that is essential for the recognition process. Therefore, it is necessary to determine the quality of the captured sample before it is accepted.

For that reason, several methods to estimate the signal quality of fingerprints have been proposed. It has been shown that their incorporation into a biometric system has improved the biometric performance. However, their capability of prediction of the influence of a particular sample on the biometric performance has not been analysed.

This thesis is focused on the evaluation of the quality assessment methods in order to answer the following open question: *Do the methods assessing fingerprint image signal quality predict the performance of a biometric system?*

The structure of the thesis is following. Chapter 2 gives a short introduction into the field of biometrics. Chapter 3 covers a novel ISO/IEC definition of the biometric sample quality which is afterwards interconnected with the fingerprint image quality in Chapter 4. Different aspects that influence the quality of a captured fingerprint image are also outlined in this Chapter. Chapter 5 reviews the state-of-the-art approaches to the estimation of the fingerprint image signal quality while Chapter 6 proposes a quality measure derived accurately from the observed biometric performance. Chapter 7 describes the specific datasets that are later used in the evaluation of quality measures, covered in Chapter 8. Finally, a summary of the work and possible extensions are included in Chapter 9.

---

[1]The International Civil Aviation Organization, http://www.icao.int.

# Chapter 2

# Biometrics Overview

Before immersing into the main topic, proper terminology shall be explained. This chapter is addressed to readers that are not familiar with biometric terminology. Rather than giving an exhaustive explanation, it introduces only the terms that are important to know in order to follow the discussion in the next chapters.

## 2.1   Biometrics

Biometrics is undergoing a process of standardization in the last years. The terminology used in this thesis follows the definitions [17] established by ISO/IEC[1].

The term *biometrics* refers to the automated recognition of individuals based on their behavioural and biological characteristics. *Biometric characteristics* can be perceived in two different manners: *behavioural characteristics* and *biological characteristics*. While the first refers to the behavioural processes created by human body, the second is related to the the physical properties of body parts. For example, signature, voice and gait are behavioural characteristics while fingerprint, face and veins are biological.

For an analogue or digital representation of biometric characteristics, the term *biometric sample* is used. Then, *biometric features* are numbers or labels extracted from the biometric samples that are reproducible for a given biometric characteristic of a person but they differ among the population. *Biometric template* is a set of stored biometric features comparable directly to the probe biometric features.

An implementation of automated recognition based on biometric characteristics is the *biometric system*.

More definitions from the biometric terminology are described in *Harmonized Biometric Vocabulary* [17], available also online [10].

## 2.2   Generic Biometric System

A generic biometric system uses one or more biometric characteristics for recognition. From the operational point of view, a biometric system consists of two main parts: *enrolment* and *recognition*.

---

[1]The ISO stands for the International Standard Organization and the IEC for the International Electrotechnical Commission.

### 2.2.1 Enrolment

Before using the biometric system, each individual has to be enrolled. First, one or more biometric samples of the desired biometric characteristic are captured from the individual. A biometric reference is created that is that biometric samples or biometric templates are attributed to an individual and stored as a biometric reference (in a database or smartcard/token). Biometric reference is later used for recognition. An individual, whose biometric characteristic is used in the biometric system, is defined as a biometric data subject.

### 2.2.2 Recognition

To recognize person's identity, again one or more biometric samples have to be obtained from the data subject. Biometric features are extracted and compared to the biometric reference(s) stored in the database. According to the desired functionality, the recognition phase can be designed in two different manners:

**Verification** – subject claims his identity and a biometric system verifies whether the claim is genuine or imposter by providing *1:1 comparison* with the corresponding reference. Note that the term *genuine* refers to the person who submits a biometric characteristic to a biometric system in an attempt to be recognized as himself. Similarly, the term *imposter* refers to the person who submits a biometric characteristic to a biometric system in an (intentional or unintentional) attempt to be recognized as another person already enrolled in the biometric system.

**Identification** – biometric system compares data subject's input to all the biometric references stored in the database *(1:many comparison)*. The system returns identity of the data subject corresponding to the most similar biometric reference or a decision indicating that the data subject was not enrolled in the biometric system.

## 2.3 Biometric Recognition

A generic biometric system consists of several components in order to work as an identification or verification system. The reference architecture according to ISO/IEC [18] is illustrated in Figure 2.1. Here, we will describe only the main parts of biometric system.

**Data Capture Subsystem** The acquisition device measures the subject's characteristic and creates its digital representation – a biometric sample

**Signal Processing Subsystem** Several processing steps are applied to a biometric sample to derive a concise representation of the biometric sample – biometric features.

**Data Storage Subsystem** During the enrolment, extracted biometric features or a biometric sample attributed to its subject are stored in the database as a reference.

**Comparison Subsystem** Comparison is the process in which the probe biometric sample of one individual is compared against biometric references of one or more individuals. The result of such a comparison is a score that indicates similarity or dissimilarity of two samples.

**Decision Subsystem** After the comparison, the recognition system is capable of deciding according to the comparison score whether a presented sample matches or does not match to a stored reference.
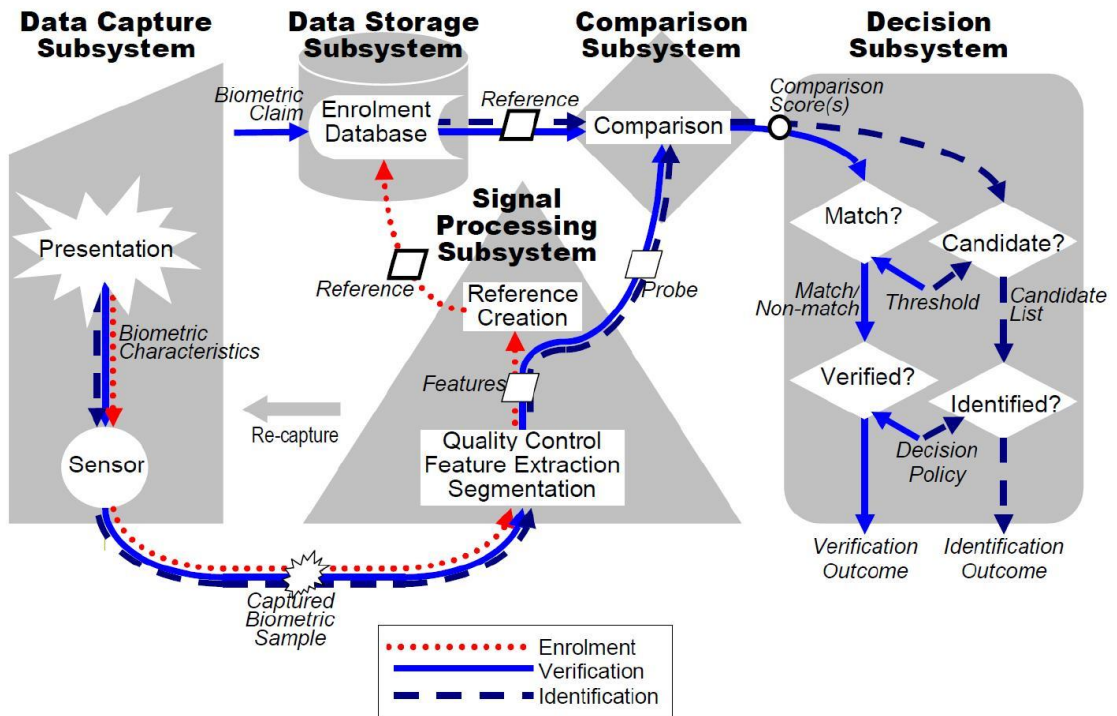
Figure 2.1: ISO reference architecture of the biometric system [18].

## 2.4 Biometric Performance

Biometric performance reflects how precisely the biometric system recognizes individuals. Several metrics that measure the biometric performance have been included into ISO/IEC International Standard (IS) [15]. Only the most important metrics, those that are used later in the thesis, are described here.

### 2.4.1 Failure-to-Capture Rate

The failure-to-capture rate (FTC) is the proportion of failures of the biometric capture process to produce a captured biometric sample that is acceptable for use.

### 2.4.2 Failure-to-Extract Rate

The failure-to-extract rate (FTX) is the proportion of failures of the feature extraction process to generate a template from the captured biometric sample.

### 2.4.3 Failure-to-Acquire Rate

The failure-to-acquire rate (FTA) is the proportion of verification or identification attempts for which the system fails to capture or locate an image or signal of sufficient quality. FTA is caused by either a FTC or an FTX in the in the verification process.

### 2.4.4 False Match Rate

False match rate (FMR) is the proportion of zero-effort imposter attempt samples falsely declared to match the compared non-self reference.

### 2.4.5 False Non-Match Rate

The false non-match rate (FNMR) is the proportion of genuine attempt samples falsely declared not to match the reference of the same characteristic from the same subject.

# Chapter 3

# Biometric Sample Quality

The first component of biometric recognition is the capture subsystem for measuring biometric characteristics. Since all elements of the biometric workflow operate in succession, the initial phase, acquisition of a biometric sample, influences all the following parts of the system. Therefore it is important to place requirements directly on the biometric sample in order to ensure that a sufficient amount of useful information will be obtained. In the biometric terminology, these requirements are conveyed by the expression *biometric sample quality*.

## 3.1 Definition

There is an emerging effort put into defining the biometric sample quality in order to improve the security of biometric systems. ISO/IEC recently established International Standard [19], in which they define different aspects of sample quality that influence the general performance of the biometric system. Three main aspects of quality are described below.

**Character** of a sample is the quality attributable to the inherent features of the source from which the biometric sample is derived. For example scarred finger has a poor character.

**Fidelity** of a sample is the quality that describes the degree of similarity between the biometric sample and its source. The relationship between character and fidelity is illustrated in Figure 3.1.

**Utility** of a sample refers to the predicted impact of an individual sample to the overall performance of the biometric system. It is dependent on the character as well as the fidelity of the sample. More detail description of utility is provided in chapter 6.

In that meaning, quality is considered to be a function on the character and fidelity components. It should convey the predicted utility of the biometric sample.
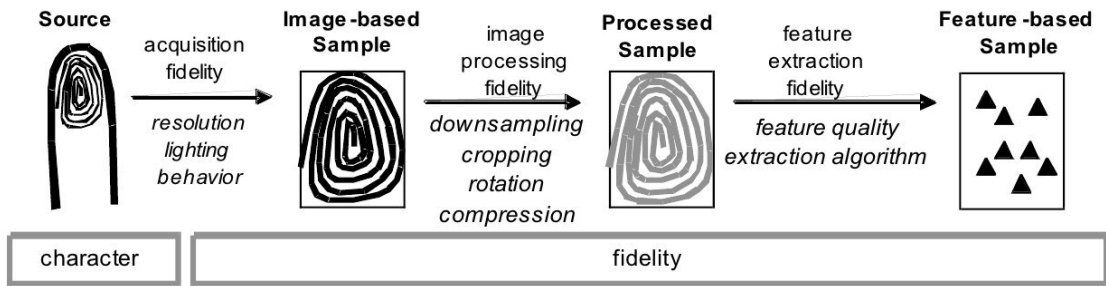
Figure 3.1: Character and fidelity as components of the biometric sample quality. Image adopted from [19].

## 3.2 Relevance of Quality Data

Sample quality measures play different roles in various contexts of biometric operations.

### Sample Submission

Sample submission requires a real-time quality assessment. It helps the operator or the automated system to improve the average quality of biometric samples submitted upon enrolment. Improvement can be done by several decisions based on the quality of the obtained sample:

**Accept** a sample of sufficient quality.

**Reject** a sample because of insufficient quality.

**Reattempt** a capture in order to obtain a sample of better quality.

**Declare a failure to acquire** based on the repetitious rejections.

This feedback is not only related to the operational efficiency. The performance of the overall biometric system could be improved as it depends on the assumption that for all the enrolled subjects suitable samples have been stored from which distinguishing features can be derived.

### Differential Processing

Different enhancement methods can be applied to biometric samples of different qualities in order to improve the accuracy of feature extraction (e.g. applying image restoration to the fingerprint image) or the biometric system can invoke different feature extraction algorithms.

### Conditional Decision

The quality of a sample can also contribute to the decision making. Changing the operational threshold of the recognition system can prevent the imposter attempts at spoofing or defeating detection with discernible samples.

In multimodal biometric systems, the relative qualities of samples of different modalities can contribute to the fusion process as it has been shown by Fierrez-Aguilar et al. [5].

## Reference Replacement

A reference dataset can be improved by tracking the quality of the reference entries. A stored biometric sample or a biometric template can be replaced by newly obtained samples (templates) of better quality. A similar scenario can also negate the effects of aging of the reference.

## Survey Statistics

Aggregated quality data in the operation process can monitor and detect abnormal behaviour, which can for example indicate failures in acquisition device, usability drawbacks etc. Survey statistics can also indicate whether a higher quality sample is likely to be obtained by another capture.

## Benchmarking

Association of the quality data with samples allows us for creating a quality-oriented dataset that can be further used for a specific performance evaluation. Testing biometric systems on samples of specified quality can highlight potential weak areas of the recognition pipeline.

## 3.3 Quality Assessment in the Biometric System

In order to achieve one or more improvements described above it is needed to include the quality apparatus into the biometric system. Figure 3.2 illustrates a general solution. After the biometric sample is acquired, several methods are applied in order to estimate the sample's quality. The output is the quality score which is used by other components of the biometric system. First, the sample has to pass the quality control. Several scenarios of sample submission have been described in section 3.2. However, the quality score can be used even in the other parts of the biometric system – at feature extraction and decision making as described in 3.2.
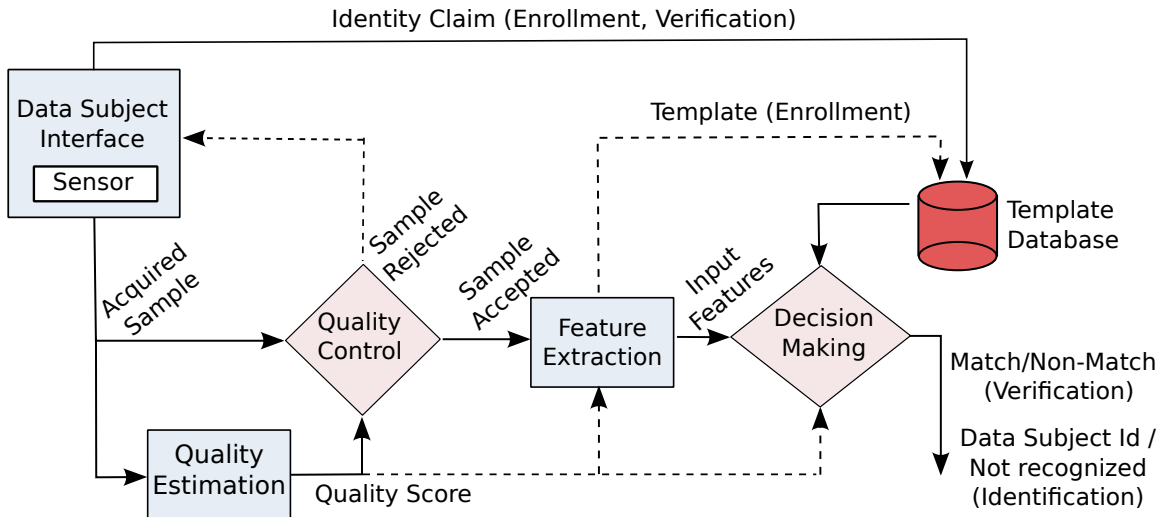


Figure 3.2: A generic biometric system with incorporated sample quality assessment.

# Chapter 4

# Fingerprint Image Quality

In the previous chapter, a general description of the biometric sample quality was given. This chapter is focused on a representation of the biometric sample in the fingerprint recognition system – a fingerprint image. The definitions of sample quality are bound to the fingerprint image in order to describe the relevant factors that influence fingerprint image quality.

## 4.1 Acquisition Fidelity

The growing deployment of fingerprint recognition systems evoked also a technical development in fingerprint acquisition. In the past, digital fingerprint images were acquired by scanning papers with ink finger imprints (Figure 4.1(a)). Nowadays, life-scan acquisition devices are used. They consist of electronic sensors that scan directly the fingertips. These sensors are based on several different technologies.

**Optical Sensors**

The biggest research and development in industry was placed on optical sensors and therefore they are the most used fingerprint sensing devices today. Previously used magnifying glass was replaced by the CMOS[1] optical sensor to capture the image of the fingerprint. The use of angle-wise illumination source causes a total internal reflection at the glass plate. Reflection of light is suppressed where skin contacts the glass. The index of refraction difference is used to obtain an image of the ridge-valley pattern.

An example of a fingerprint image captured by optical sensor is shown in Figure 4.1(b).

**Capacitive Sensors**

Capacitive sensors use dielectric measurements to distinguish between the ridges and valleys of the outer skin. The sensor consists of a CMOS grip chip. Capacity is measured for every cell of the chip via conductance and corresponds to the skin distance. The capacitive cells do not penetrate beneath the skin, they capture only the surface of the finger.

Sample fingerprint image captured by capacitive sensor is shown in Figure 4.1(c).

---

[1]Complementary metal-oxide-semiconductor, a technology for constructing integrated circuits [34].

### Thermal Sensors

Thermal fingerprint sensors measure the heat transferred from the sensor to the fingerprint. The ridges touching the sensor draw heat away from the sensor faster than valleys. Temperature difference between the power drawn by sensor and the finger forms the fingerprint pattern.

An example of a fingerprint image captured by thermal sensor is shown in Figure 4.1(d).

These sensors are mostly implemented as *swipe sensors*. User slides a finger vertically over the sensor surface and repeated temperature measurements are made. Then, a software solution is needed to reproduce the image of the entire fingerprint.

### Ultrasound Sensors

Ultrasound fingerprint sensors were derived from the principles of medical ultrasound. They create visual images of the fingerprints but unlike optical imaging, ultrasonic sensors use very high frequency waves of sound to penetrate the epidermal layer[2]. Ultrasonic beam is scanned across the fingertip and the signal reflected from dermis is measured. The difference between the depth of the valleys and ridges form the fingerprint pattern.

A fingerprint image captured by ultrasound sensor is shown in Figure 4.1(e).

### Electro-optical Sensors

Electro-optical sensors employ polymers that are able to emit light when properly excited with a voltage. The finger acts as a ground – polymer emits light where the ridges touch. CMOS optical sensor is used afterwards to capture the image of the luminous fingerprint pattern.

An example of a fingerprint image captured by electric field sensor is shown in Figure 4.1(f).

### Pressure Sensors

Pressure fingerprint sensors consist of three layers. A tiny silicon layer is placed between two conductive layers and works as a switch. By applying pressure from a finger, two conductive layers are closed. A different pressure generated from the ridges and valleys forms the the fingerprint pattern.

### Electric Field Sensors

Electric field (e-field) sensors use an antenna array that measures the electric field formed between two conductive layers. Field created between the finger and the adjacent semiconductor mimics the shape of the ridges and valleys of the epidermal layer.

### Multispectral Sensors

Multispectral sensors (MSI) capture multiple images of the finger under different illumination conditions (different wavelengths, different illumination orientations, and different polarization conditions). The resulting data contain information about both the dermis

---

[2]The upper or outer layer of two main layers of the cells that make up the skin. The second, inner layer is called *dermis*.

and epidermis features of the skin. Afterwards, data are processed by software to generate a single composite fingerprint image.



(a) Ink         (b) Optical         (c) Capacitive

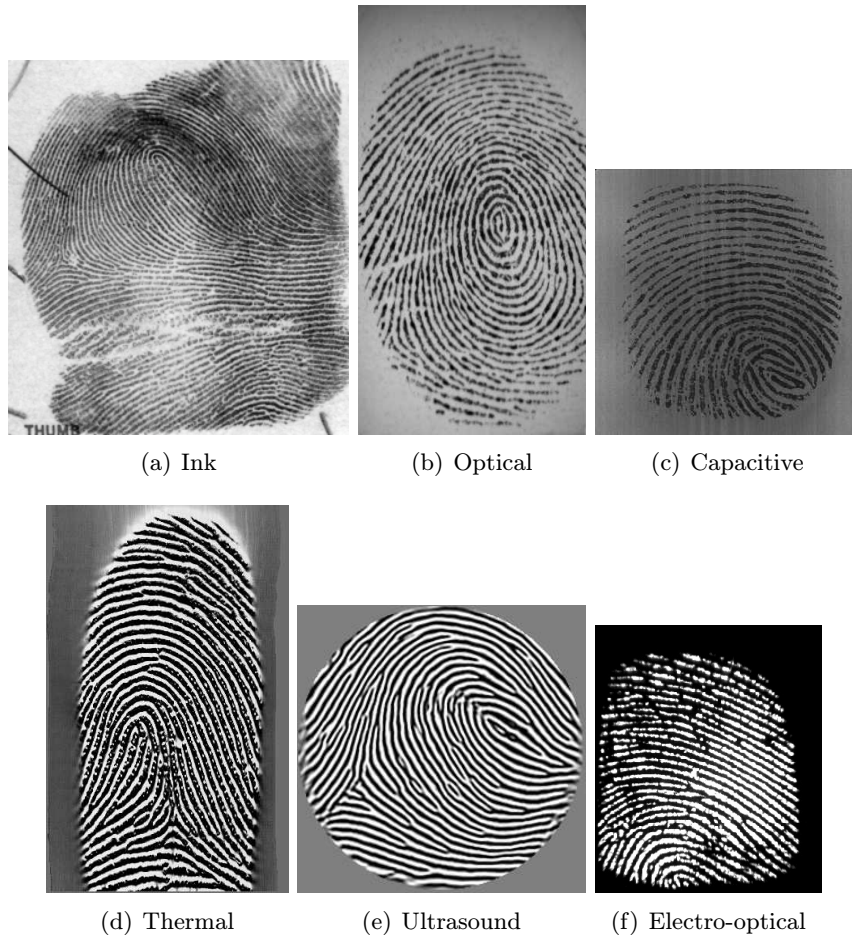(d) Thermal         (e) Ultrasound         (f) Electro-optical

Figure 4.1: An example of fingerprint images acquired by different sensing techniques. Image (a) was obtained from SD27a dataset [30], images (b) and (c) from FVC2002 dataset [26], image (d) from FVC2004 dataset [27] and image (e) from [40]. Image (f) was captured by author.

The appearance of the generated fingerprint image depends on the specific sensor which was used. The deployment of different acquisition technologies results in a great variety of fingerprint images of different parameters. To establish a universal quality standard that will be adaptable for use by all the applications is a challenging task.

The variety of fingerprint images obtained from different acquisition devices can be seen in Figure 4.1. In this thesis, just live-scan fingerprint images will be considered.

## 4.2 Quality Factors

The purpose of explaining all the present acquisition technologies in previous section is to show the important problem of obtaining fingerprint images. Only ultrasonic and multi-spectral sensors are able to penetrate beneath the epidermal layer of the skin. All the other technologies rely purely on surface measurements that can be negatively affected by many

different factors. These factors decrease the fidelity and thus the quality of the captured fingerprint sample. But as shown in Section 3, the quality also depends on the character of the sample.

Different factors that cause defects on fingerprint samples are listed in ISO/IEC Technical Report [19]. Here, a detailed description will be provided.

### 4.2.1 Defect Caused by User Character

*Skin conditions* can degrade the quality of the captured fingerprint image to such an extent that they make it impossible to observe the ridge structure of the fingerprint. Common matters such as dry fingerprints cause that some parts of the ridges do not come in direct contact with the surface of the sensor. They cause that the ridge pattern is intermittent, or the fingerprint area is of low contrast, which leads to feature extraction problems. By contrast, too wet fingertips have the valleys filled with liquid and they produce fingerprint images with blending ridges.

Similar problems can be caused by *character impurities* such as dirt, latent print, etc.

Also other character properties such as *scars, wrinkles, blisters, eczema as well as dermatology factors* negatively influence the fingerprint quality.

An example of a defect on fingerprint image caused by character is shown in Figure 4.3(a).

### 4.2.2 Defects Caused by Imaging

Biometric samples that are represented by digital images are dependent on image sampling and on the quantization process which causes alterations to the original biometric characteristic. These factors are mostly attributed to the properties of the acquisition device. Namely they are:

- sampling error,

- low contrast or signal-to-noise ratio,

- distortion,

- erroneous or streak lines,

- uneven background,

- insufficient dynamic range,

- non-linear or non-uniform grayscale output,

- missing pixels due to hardware failure,

- aliasing problems.

In [36, 35] it has been shown that also the size of the sensing area of a sensor influences the recognition accuracy and therefore, it should also be considered as one of the factors related to imaging that influences the quality. Figure 4.2 shows two fingerprints with different sensing areas.

Figure 4.2: Two fingerprint images acquired by optical scanners with different sensing areas. On the left, a scanner with an insufficiently small sensing area was used. Images obtained from FVC 2000 [28].

### 4.2.3 Defects Caused by User Behaviour

Several scenarios of incorrect user cooperation with the acquisition device can occur. *Improper finger placement* cause that different rotations and positions of a fingerprint pattern occur in the image.

When different parts of a finger are presented on the sensor surface, *insufficient area of the fingerprint pattern* is obtained.

Inadequate pressure of a finger on the sensor surface or pressure in different parts of a finger can cause *elastic deformations*.

An example of fingerprint images affected by user behaviour is shown in Figure 4.3(b).

### 4.2.4 Defects Caused by Environment

Environmental conditions in which the sensor operates can also influence the fingerprint quality. Most known influencing factors are: humidity, light conditions and impurities on the scanner surface.

Sample fingerprint images affected by environmental conditions are shown in Figure 4.3(c).



(a) User character (wet finger)

(b) User behaviour (finger misplacement)

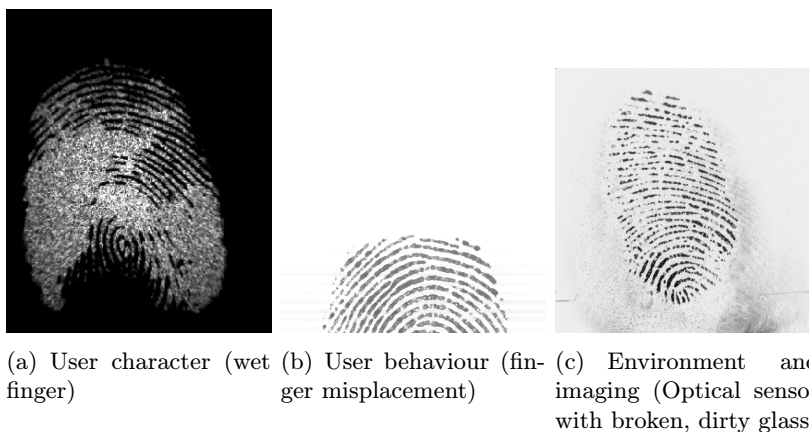(c) Environment and imaging (Optical sensor with broken, dirty glass)

Figure 4.3: An example of fingerprint samples influenced by different factors. Images were captured by author.

## 4.3   Defining Fingerprint Quality

Although several factors that influence fingerprint quality were described by ISO/IEC [19], a general definition of fingerprint quality does not exist. The most used is the description by Chen et al. [4]. They conceive fingerprint quality as "the measure of the clarity of ridges and valleys and the extract-ability of the features used for identification such as minutiae, core and delta points etc".

In this thesis, the perception of fingerprint quality is slightly different from the mentioned one. The author addresses the problem of fingerprint quality not only to the properties of the observed fingerprint pattern but the completeness of the fingerprint is also considered as one of the quality factors. Inspired by Ratha et al. [33], fingerprint quality in this thesis is described from two different points of view: *consistency* and *uniformity*.

### 4.3.1   Fingerprint Consistency

The consistency of the fingerprint image relates to the quality factors influenced by user behaviour. Improper user cooperation with the acquisition device can cause a situation, where the captured fingerprint image carries only partial information about its source.

Inconsistent samples of the same finger have a very small similarity and thus degrade significantly the comparison accuracy. Therefore, the measures of fingerprint consistency should examine the completeness of the fingerprint area. However, an exact specification on the fingerprint completeness does not exist. Several open issues still remain, for example: should there be the delta and core points necessarily present in the fingerprint image?

Sample fingerprints of different consistency are shown in Figure 4.4.



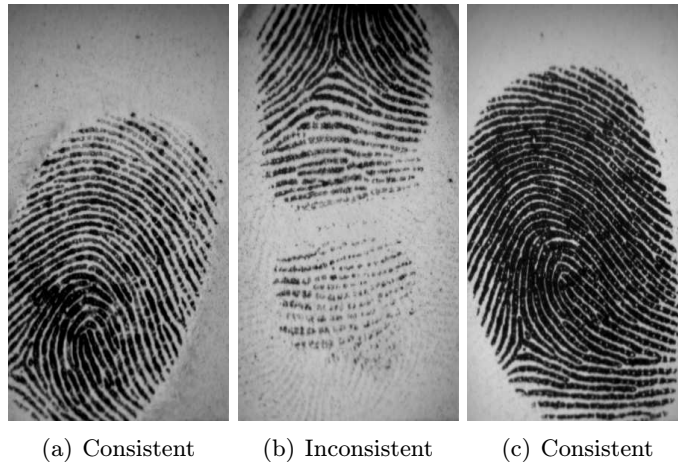(a) Consistent       (b) Inconsistent       (c) Consistent

Figure 4.4: An example of three fingerprint images of the same finger with different consistency. Wrong finger placement on the sensor plate caused the inconsistent fingerprint sample (b). Images were obtained from FVC2002 [26].

Note, that the elastic distortions caused by user behaviour are not related to the fingerprint consistency. Fingerprint consistency, as it is used here, reflects the completeness of the fingerprint image. It does not deal with the quality of the information obtained and thus it should be not attributed to the signal quality of the fingerprint image.

### 4.3.2 Fingerprint Uniformity

Fingerprint uniformity is directly related to the quality of the information already included in the image. Often it is referred to as the "signal quality". It can be understood as the clarity of the ridge-valley pattern and feature extract-ability as described above.

Defects in the fingerprint sample caused by user character, imaging and environment factors but also elastic distortions caused by user behaviour result in a non-uniform fingerprint pattern.

An example of different fingerprint uniformity is shown in Figure 4.5.



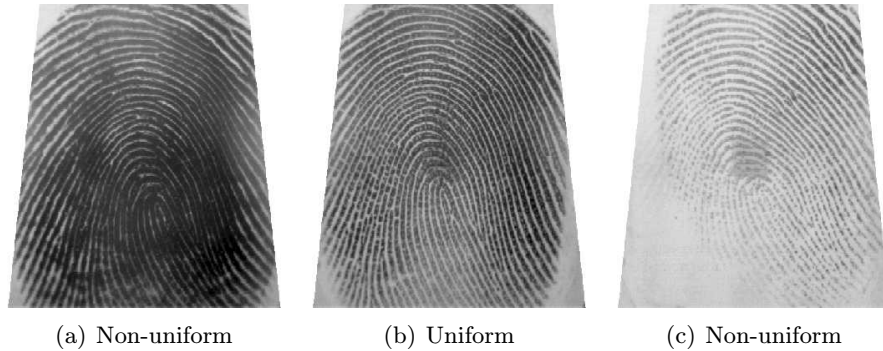(a) Non-uniform      (b) Uniform      (c) Non-uniform

Figure 4.5: An example of three fingerprints of the same finger with different uniformity. Samples were obtained from user character of different skin conditions. Sample (c) is influenced by the moisture present at the finger, while sample (b) wasn't influenced by any of the factors mentioned above. On the contrary too dry finger can also cause a non-uniform captured sample as shown in (a). Images were acquired with an optical scanner and belong to the FVC2004 dataset [27].

In this thesis, signal quality estimation methods will be dealt as the measures of fingerprint uniformity.

# Chapter 5

# Methods for Fingerprint Image Signal Quality Estimation

Several methods for fingerprint image quality estimation have been proposed in the literature. According to ISO/IEC Technical Report [19] they can be divided into *local* and *global* methods.

In local analysis, local features of the fingerprint are used for computation of the quality score [2]:

- local direction of ridges,

- Gabor filters,

- pixel intensity and

- power spectrum.

For global analysis of quality, two approaches are used according to [2]:

- direction field and

- power spectrum.

In this thesis, two methods of local analysis of quality and two methods of global analysis of quality were examined and implemented. Orientation certainty level is based on the local direction of ridges and ridge-valley structure analysis is based on pixel intensity features. From global analysis methods, orientation flow analysis based on the direction field features and radial power spectrum were selected.

## 5.1 Local Analysis

The majority of present methods of the fingerprint signal quality assessment examines local structure of a fingerprint. Local structure is represented by the texture-like pattern of ridges and valleys. An example is shown in Figure 5.1.

Image is partitioned into a grid of non-overlapping square blocks. The size of the block depends on the image resolution. It should be suitably chosen to cover at least 2 ridge lines. Within each block, local features are extracted and used to determine the quality of a particular block. Quality is expressed by the quality score as described above. To obtain
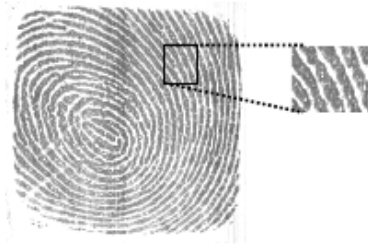
Figure 5.1: Local structure of a fingerprint image. Image taken from [19].

the quality of an entire fingerprint image, quality scores of all blocks are combined into one. Different statistical methods can be used for the score combination, for example majority vote, mean, median etc.

### 5.1.1 Orientation Certainty Level

The Orientation Certainty Level (OCL) measures the energy concentration along the direction of ridges.

In the method presented by Lim et al. [23], first the intensity gradient of each pixel within a block is computed. That is achieved by applying the Sobel operator with two $3 \times 3$ windows [22]. Then, the covariance matrix $C$ of the gradient vector for an $N$-point image block is computed as:

$$C = \frac{1}{N} \sum_N \left\{ \begin{bmatrix} dx \\ dy \end{bmatrix} \begin{bmatrix} dx & dy \end{bmatrix} \right\} = \begin{bmatrix} a & c \\ c & b \end{bmatrix}, \tag{5.1}$$

where $dx$ and $dy$ represent the intensity gradient of each pixel.

From the covariance matrix $C$, two eigenvalues $\lambda$ are derived:

$$\lambda_{min} = \frac{(a+b) - \sqrt{(a-b)^2 + 4c^2}}{2}, \tag{5.2}$$

$$\lambda_{max} = \frac{(a+b) + \sqrt{(a-b)^2 + 4c^2}}{2}. \tag{5.3}$$

For a fingerprint image block, orientation certainty value $(OCL)$ is defined simply as the ratio of $\lambda_{min}$ and $\lambda_{max}$:

$$OCL = \frac{\lambda_{min}}{\lambda_{max}}. \tag{5.4}$$

Since $a, b$ from the Equation 5.1 are greater than 0 and $\lambda_{min}$ is always greater than $\lambda_{max}$, $\lambda_{min}/\lambda_{max}$ is always less than 1.

When two eigenvalues, $\lambda_{min}$ and $\lambda_{max}$, have similar magnitudes, it indicates that there is not a strong energy among the ridge orientation observed and $OCL$ approximates to 1. When the eigenvalues are different, there is a strong energy observed and $OCL$ approximates to 0.

In the concept of fingerprint quality analysis, the energy concentration among the ridge orientation can be used to measure the quality of the fingerprint image block. Higher energy concentration should indicate a higher quality of the fingerprint. In ISO/IEC Technical

Report [19] the $OCL$ value from Equation 5.4 is used to express the quality. Then, the lower the $OCL$ is, the higher fingerprint quality is observed, while the range of quality score remains between 0 and 1. And that is in contradiction with ISO/IEC IS [16] which defines the quality score to be continuously increasing with the higher quality, from 0 up to 100.

On that account, the local quality measure $Q_{OCL}$ proposed here is the inverted and normalized $OCL$ value in order to follow the ISO/IEC requirements [16]:
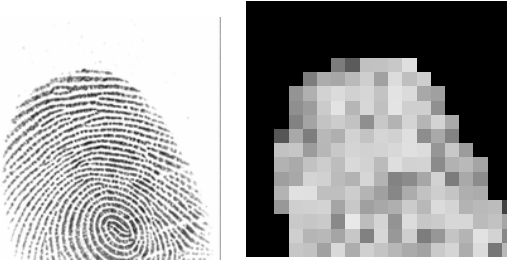
$$Q_{OCL} = \left(1 - \frac{\lambda_{min}}{\lambda_{max}}\right) \times 100, \tag{5.5}$$

where the range of the $Q_{OCL}$ is between 0 and 100.

An example of the OCL computation is shown in Figure 5.2.
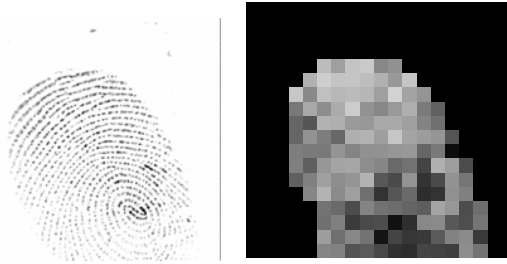


(a) Average $Q_{OCL} = 48$



(b) Average $Q_{OCL} = 72$

Figure 5.2: Computation of the OCL for two fingerprints of the same finger of different quality. Fingerprint images and corresponding block-wise values of $Q_{OCL}$ are shown. Brighter colour indicate higher quality in the region, background area is marked with black colour. The average $Q_{OCL}$ value was computed on the foreground blocks only. Fingerprint images were obtained from FVC2002 dataset [26].

Since the OCL is computed from the grey-level gradient, it can be negatively affected by marks or residuals in the sample with strong orientation strength. Also high curvature areas such as core and delta points have a negative impact on the OCL quality as these regions often do not exhibit a one dominant direction within the block.

### 5.1.2 Ridge-Valley Structure Analysis

The ridge-valley structure analysis indicates the clarity of the local fingerprint structure with the aim to distinguish the ridges and valleys along the ridge direction. Chen et al. [4]

analyse the distribution of the segmented fingerprint structure to describe the clarity of a given fingerprint pattern. Their approach is generally accepted and will be followed in this thesis.

First, inside each image block (denoted as $V_0$) the direction of the ridge flow is measured. An orientation line perpendicular to the ridge direction is computed (Figure 5.3(a)). Then the block $V_0$ is aligned to the horizontal position of the orientation line. The newly created aligned block $V_1$ is shown in Figure 5.3(b). From the centre of the block $V_1$ along the orientation line a 2-D vector $V_2$ is extracted as it is depicted in Figure 5.3(c).

Then, several parameters from the extracted block are computed, such as fingerprint pattern clarity, ridge thickness and valley thickness. For that reason, the vector $V_2$ should be of minimum size of $32 \times 15$ pixels to cover several ridge lines separated by valleys. Therefore, the original block $V_1$ has to be of minimum size of $36 \times 36$ pixels. The whole process is shown in Figure 5.3.



(a) Image block $V_0$ with red orientation line shown.

(b) Block $V_1$, created by aligning the block $V_0$ to horizontal orientation line.
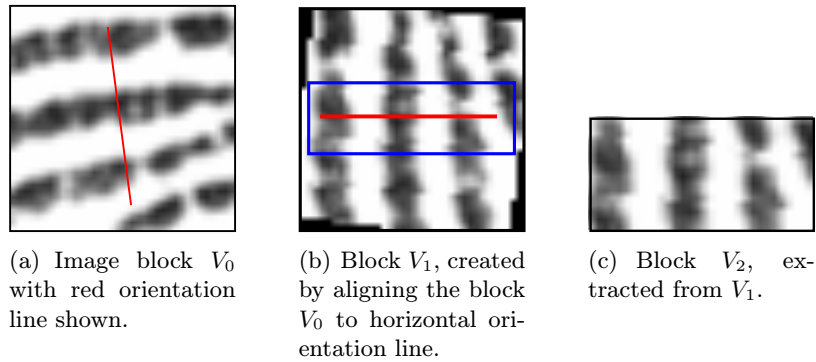
(c) Block $V_2$, extracted from $V_1$.

Figure 5.3: Process of rotation and transformation of image block for ridge-valley analysis.

The average profile is computed from block $V_2$ by equation:

$$V_3(i) = \frac{\sum_{j=1}^{M} V2(i,j)}{M}, \tag{5.6}$$

where $i = 1 \ldots x$ is the horizontal index and $M = y$ is the height of the block $V_3$ of size $x \times y$. It is necessary to determine the threshold $DT$ that will be used to distinguish between ridge and valley regions in $V_3$. For that purpose, linear regression is applied to the average profile. Figure 5.4 shows the process of segmentation of the block using $DT$. Regions with the corresponding values of the profile lower than $DT$ are considered to be ridges, the rest are valleys.

As the ridge and valley have been separated, the clarity test can be performed on each segmented rectangular region of the 2-D vector $V_3$. By the clarity test, the homogeneity of the ridge and valley regions is controlled. Each pixel of the ridge region is checked to have intensity of the ridge pattern – which means that the pixel intensity should be lower than the established threshold $DT$. Similarly, the pixel of the valley region should have intensity higher than the corresponding threshold $DT$.

The total proportion of bad pixels (those with different intensities than expected) can be expressed by:

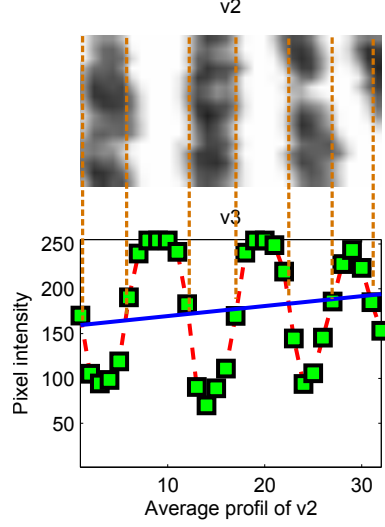$$\alpha = v_B/v_T, \tag{5.7}$$
$$\beta = r_B/r_T, \tag{5.8}$$

Figure 5.4: Segmentation of ridge and valley regions. The upper panel shows the ridge pattern (on the 2D vector $V_2$). The distribution of the ridge pattern is projected as the one dimensional cumulative intensity profile shown in the bottom panel. The $x$ axis is the average profile and the $y$ axis is the intensity level. Green dots represent the elements of the profile, blue line shows the threshold $DT$ used to separate the regions.

where $v_B$ is the number of pixels in the valley region with intensity lower than $DT$ and $v_T$ is the total number of pixels in the valley region. Similarly, $r_B$ is the number of pixels in the ridge region with intensity higher than $DT$ and $r_T$ is the total number of pixels in the ridge region.

Another factor that should be considered by the ridge-valley structure analysis is the ridge thickness and valley thickness. Too thin or too thick ridges (valleys) as well as ridges too close or too far apart indicate the poor fidelity of the captured fingerprint image. In this case, the poor fidelity is probably caused by elastic distortions of the finger (e.g. moistened finger, finger pressed too hard on the sensor, etc.). Thus the nominal value of the ridge and valley thickness can be used to detect the abnormal ridge-valley pattern.

ISO/IEC defines in [19] the normalized valley thickness $W_v^{norm}$ and the normalized ridge thickness $W_r^{norm}$ as follows:

$$W_v^{norm} = \frac{W_v}{(S/125)W^{\max}}, \tag{5.9}$$

$$W_r^{norm} = \frac{W_r}{(S/125)W^{\max}}, \tag{5.10}$$

where $W_v$ and $W_r$ are the observed valley thickness and ridge thickness, respectively. $S$ refers to scanner resolution by which the fingerprint image was obtained and $W^{\max}$ is the estimated ridge (or valley) thickness for image of 125dpi resolution, while according to [19] the recommended $W^{\max} = 5$.

Then the final quality score $Q_{LC}$ based on local clarity is the average value of $\alpha$ and $\beta$ from Equation 5.7, but only for valid ridge and valley regions:

$$Q_{LC} = \begin{cases} (1 - (\alpha + \beta)/2) \times 100 & \text{if } (W_v^{nmin} < W_v^{norm} < W_v^{nmax}) \text{ and} \\ & (W_r^{nmin} < W_r^{norm} < W_r^{nmax}) \\ 0 & \text{otherwise}, \end{cases} \tag{5.11}$$

22

where $W_r^{nmin}$ and $W_v^{nmin}$ are minimum values for the normalized ridge and valley thickness, respectively. Similarly, $W_r^{nmax}$ and $W_v^{nmax}$ are maximum values for the normalized ridge and valley thickness, respectively.

Again, the resulting quality score $Q_{LC}$ proposed here is slightly different from the definition of $LCS$ proposed in ISO/IEC Technical Report [19]. The normalization was performed to achieve that the $Q_{LC}$ will be in the range between 0 and 100. For the ridges with good clarity, $Q_{LC}$ approaches 100, while for unclear ridge-valley structure, $Q_{LC}$ will be close to 0.

An example of the computation of $Q_{LC}$ is shown in Figure 5.5.
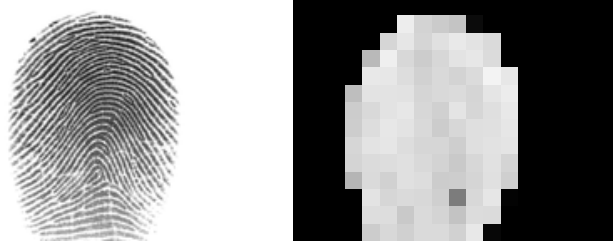


(a) Average $Q_{LC} = 80$
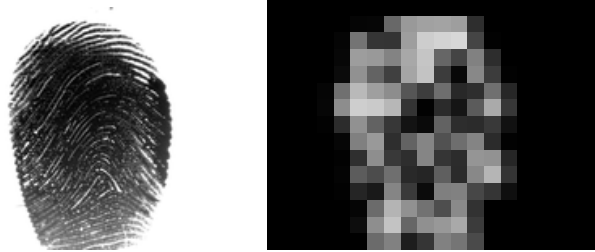


(b) Average $Q_{LC} = 32$

Figure 5.5: Computation of the $Q_{LC}$ for two fingerprints of different quality, obtained from the same finger. Fingerprint images and corresponding block-wise values of $Q_{LC}$ are shown. Brighter colours indicate higher quality in the region, black colours represent the areas with insufficient ridge thickness and background area. The average $Q_{LC}$ value was computed on the foreground blocks only. Fingerprint images were obtained from FVC2004 dataset [27].

There are several factors that can negatively affect the resulting $Q_{LC}$, such as *highly curved ridges, ridge endings, bifurcations, delta and core points*. Since these factors correspond to the character of a sample, they cannot be eliminated. To minimize the probability to encounter these factors in one local fingerprint structure, the size of the block $V_3$ used for analysis needs to be sufficiently small but, on the other hand, it should be possible to observe the ridge-valley structure within one block.

## 5.2 Global Analysis

Fingerprint quality assessment methods that rely on global analysis examine the overall ridge-valley structure of a fingerprint.

### 5.2.1 Orientation Flow Analysis

Orientation flow analysis measures the continuity of the ridge flow in the fingerprint pattern.

The method proposed by Chen et al. [3] uses the local angle information provided by the orientation field. First, the image is partitioned into a grid of non-overlapping blocks. Then, within each block of the grid, the dominant ridge orientation is determined. The ridge direction is given by the principal eigenvector of the covariance matrix computed from the gradient vector as described in Section 5.1.1. A 2D array $V$ is defined to hold all the orientation angles of the ridge directions computed from the fingerprint, as shown in Figure 5.6.

Then, to analyse the ridge flow around the particular block $V$, the absolute difference $D$ of the orientation angle of its surrounding blocks is computed as follows:

$$D(i,j) = \frac{\sum_{m=-1}^{1} \sum_{n=-1}^{1} |V(i,j) - V(i-m, j-n)|}{8}.$$

(5.12)

The sum is divided by 8 because 8-neighbourhood is assumed.

The continuity in the orientation flow is observed where orientation changes gradually between the neighbouring blocks. According to [3], the tolerance of 8°of the angular change is considered. On that account the local orientation quality score $Q_{loc}$ is computed:

$$Q_{loc}(i,j) = \begin{cases} 100 & \text{if } D(i,j) \leq 8° \\ (1 - \frac{D(i,j) - 8°}{90° - 8°}) \times 100 & \text{if } D(i,j) > 8°. \end{cases}$$

(5.13)

Finally, the global quality score $Q_{OF}$ is calculated by averaging the $Q_{loc}$ values:

$$Q_{OF} = \frac{\sum\limits_{i,j} Q_{loc}(i,j)}{N},$$

(5.14)

where $N$ is the number of blocks in $V$. The resulting quality score $Q_{OF}$ proposed here is in the range of $[0, 100]$. Abnormal orientation changes observed in the ridge flow cause that $Q_{OF}$ decreases.

An example of orientation flow computed for two fingerprints of different quality obtained from the same finger is shown in Figure 5.6.

The core and delta points present in the fingerprint cause the significant change of ridge flow and so they negatively influence the global quality score $Q_{OF}$.

### 5.2.2 Radial Power Spectrum

In a grey-scale image, spatial frequency is directly related to the rate of pixel intensity change. For fingerprint image it means that the ridge frequency lies within a certain range.

Chen at al. [4] analyse the robustness of the ridge structure by computing the two-dimensional Discrete Fourier Transform (DFT). They assume that a good quality fingerprint image should yield a strong ring pattern in the Fourier spectrum, indicating the dominant frequency band associated with the frequency of ridges. On the other hand, poor quality
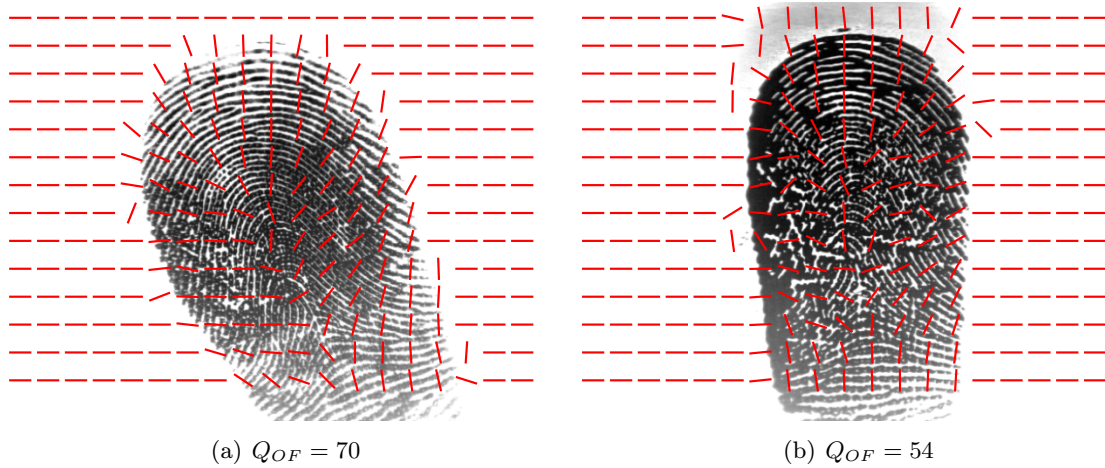
(a) $Q_{OF} = 70$        (b) $Q_{OF} = 54$

Figure 5.6: Computation of the $OF$ for two fingerprints of different quality, obtained from the same finger. On the left image the ridge flow changes in a smooth trend which is reflected by the higher $Q_{OF}$ value. By contrast, in the right fingerprint image, the orientation flow does not change gradually in all the parts, and therefore the $Q_{OF}$ is lower. Fingerprint images were obtained from FVC2004 dataset [27].

image has unclear and non-uniformly spaced ridges which results in a more diffused energy in Fourier spectrum.

The 2-D DFT of a function $p(x,y)$ evaluated at the spatial frequency $(\frac{2\pi k}{M}, \frac{2\pi l}{N})$ is given by the equation [8]:

$$f(k,l) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} p(x,y) e^{-j2\pi(\frac{kx}{M} + \frac{ly}{N})}. \tag{5.15}$$

In the case of a fingerprint sample, $p(x,y)$ refers to the intensity at pixel $(x,y)$ of the grey-scale digital image of size $N \times M$. Components of the Fourier transform in Equation 5.15 are complex quantities. Therefore, to analyse the geometric structure of an image, the magnitude of the Fourier transform should be considered [8]:

$$F(k,l) = |f(k,l)|^2. \tag{5.16}$$

Sample fingerprint images and their corresponding energy concentrations in the Fourier spectrum are shown in Figure 5.8(a) and Figure 5.8(b), respectively.

According to [12], the ridge frequency is generally around 60 cycles per image width or height. Since the quality measure should be invariant to the image size, the range of possible image width (height) is defined to be between 120 and 1000 pixels. Then, the minimum possible ridge frequency is $r_{\min} = 60/1000 = 0.06$ cycle/pixel and the maximum $r_{\max} = 60/120 = 0.5$ cycle/pixel. The term *frequency of interest (FOI)* will be used to express the range of possible ridge frequencies, defined by lower bound $r_{\min}$ and upper bound $r_{\max}$.

Since the DFT is evaluated at the spacial frequency $(\frac{2\pi k}{M}, \frac{2\pi l}{N})$, the ridge frequencies $r_{\min}$ and $r_{\max}$ can be identified in Fourier spectrum on $k$ axis by substitutions: $r_{\min} = \frac{2\pi k}{M}$, $r_{\max} = \frac{2\pi k}{M}$, where $M$ is the image width; and on $l$ axis: $r_{\min} = \frac{2\pi l}{N}$, $r_{\max} = \frac{2\pi l}{N}$, where $N$ is the image height.

To measure the energy concentration, the Fourier spectrum $F(k, l)$ is transformed into the polar coordinate system $(\alpha, r)$:

$$J(r) = \frac{\sum\limits_{\alpha=0}^{\pi} \sum\limits_{r}^{r+\Delta r} F(\alpha, r)}{\sum\limits_{\alpha=0}^{\pi} \sum\limits_{r_{\min}}^{r_{\max}} F(\alpha, r)}. \tag{5.17}$$

The transformation of the Fourier spectrum into the polar coordinate system is visualized in Figure 5.7.

To express the energies in the range of $[0, 1]$, they are normalized by minimum and maximum values occurring within the data.
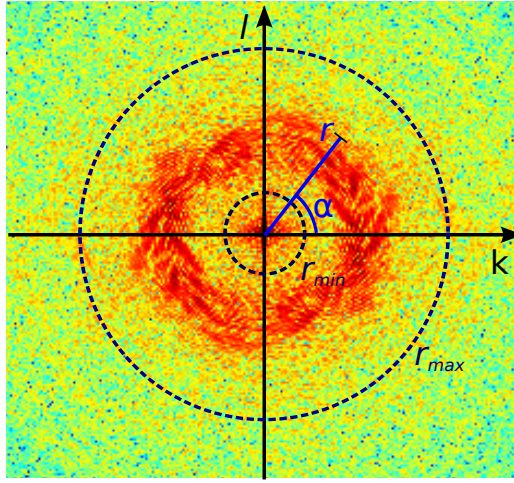


Figure 5.7: Transformation of the Fourier spectrum into the polar coordinate system.

The resulting image quality is defined as the maximum energy distribution in the radial Fourier Spectrum within the FOI:

$$Q_{POW} = \max_{r \in [r_{\min}, r_{\max}]} (J(r)). \tag{5.18}$$

Figure 5.8 shows the computation of $Q_{POW}$ for 3 fingerprint images of different quality. The image quality increases from left to right (Figure 5.8(a)). Corresponding energy concentrations in the Fourier spectrum are shown in Figure 5.8(b). With the increasing fingerprint quality, energy is more concentrated in the ring-shape pattern. Figure 5.8(c) shows the energy distribution within the FOI in the Radial Fourier Spectrum. The energy distribution is more peaked as the image quality improves (from left to right in Figure 5.8(c)).

(a) Fingerprint images of different quality.
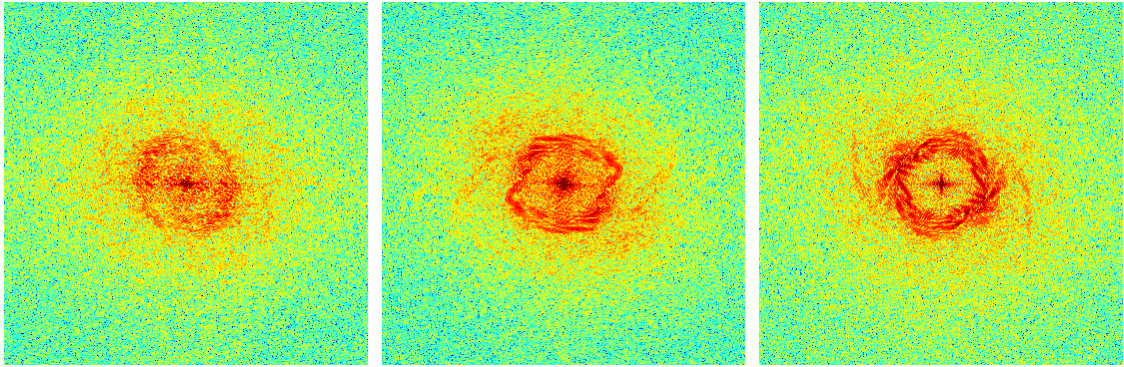


(b) Energy concentration in the Fourier Spectrum.
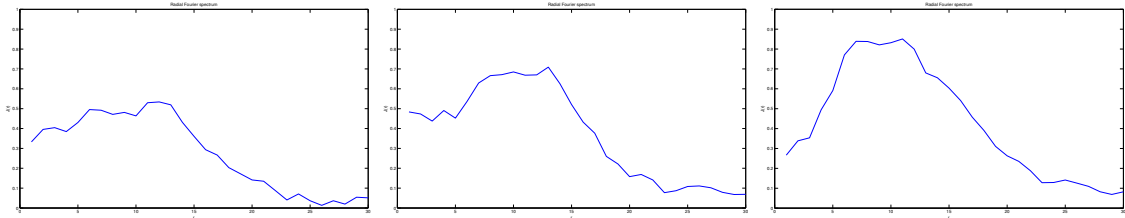


(c) Normalized Radial Fourier Spectrum in the FOI. $x$ axis is the frequency $r$ and $y$ axis the normalized energy value $J(r)$.

Figure 5.8: Computation of the Radial Power Spectrum. In (a), fingerprints subjectively assessed to be of different quality that increases from left to right. Their corresponding energy concentration in the Fourier Spectrum is shown in (b), while (c) displays the energy concentration within the FOI in the Radial Fourier Spectrum. The resulting quality scores are: $Q^1_{POW} = 53$ for the left image, $Q^2_{POW} = 71$ for the middle image and $Q^3_{POW} = 85$ for the right image. Fingerprints were obtained from the dataset FVC2000 [28].

## 5.3   Implementation of the Quality Assessment System

Within this thesis, a system for assessment of the quality of fingerprints has been designed, based on the described methods, and implemented in MATLAB[1]. The workflow diagram of the system is shown in Figure 5.9 and the description follows.
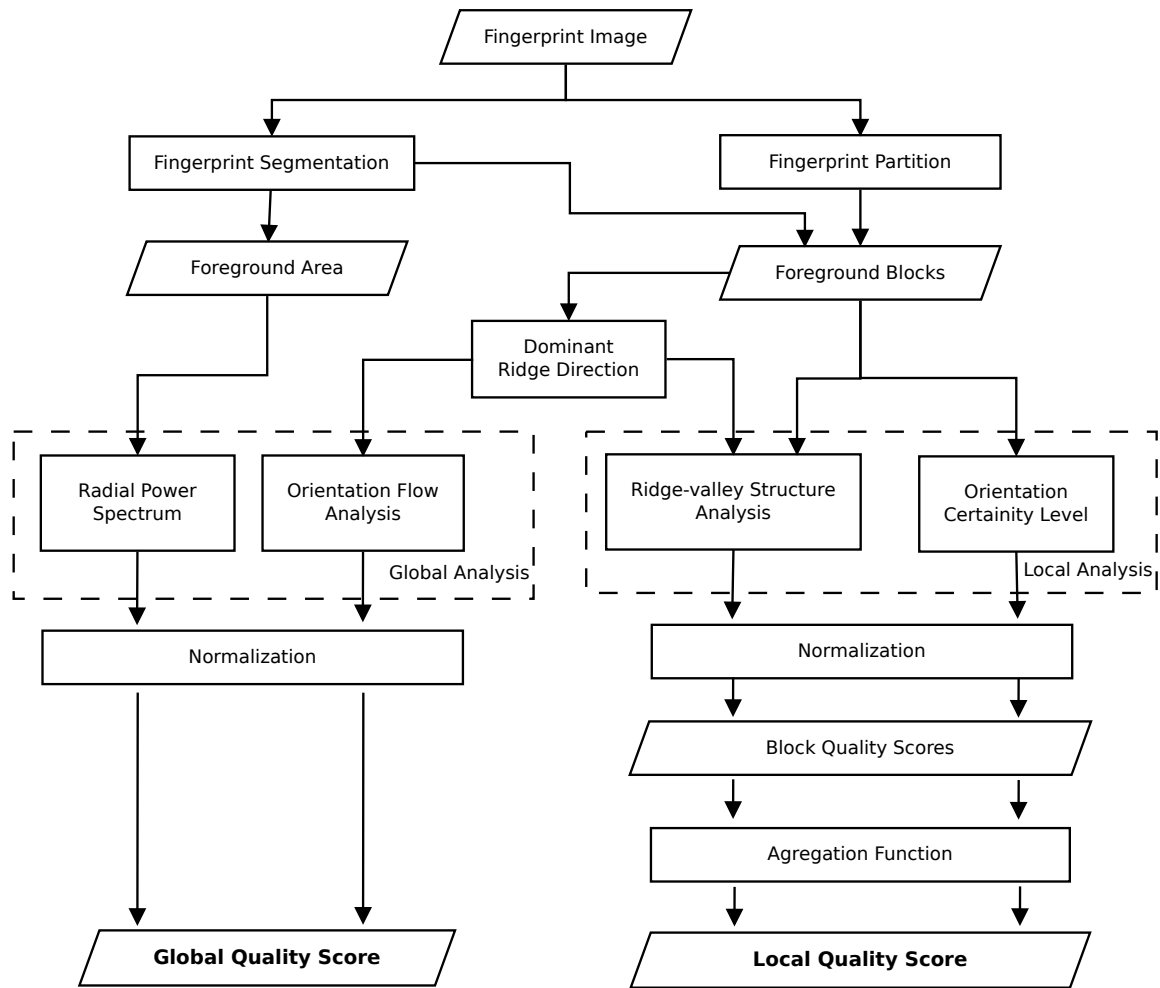
Figure 5.9: The workflow of the proposed quality assessment system.

First, several segmentation algorithms were implemented because the methods for quality assessment need to be applied to the seperated foreground regions. The method based on the standard deviation of pixel intensities within blocks was selected since the segmentation was sufficiently reliable for the most widespread sensoring technology – optical, thermal and capacitive. Further, an appropriate partition of the image was performed and the dominant ridge direction was estimated for each block in the foreground area. Then, the

---

[1]MATLAB is a numerical computing environment and programming language developed by MathWorks, http://www.mathworks.com/products/matlab.

28

local assessment methods (ridge-valley structure analysis and orientation certainty level) determine the quality of each block separately and the final quality score is obtained from their agregation. Global quality scores are output of the radial power spectrum method and orientation flow analysis.

All the assessment methods were implemented according to the description given in the preceding Sections 5.1, 5.2. The figures and results in those sections were generated with help of this system.

# Chapter 6

# Performance-based Quality

An open problem in fingerprint quality assessment is the lack of evaluation techniques that are able to identify the reliability of the given quality estimation method. Most of the current methods were evaluated by their authors on small datasets that are not publicly available and therefore it is not possible to benchmark the algorithms [25, 3, 24].

A possible way out of this problem is to create a dataset which would consist of fingerprint images labelled with quality scores assigned by dactyloscopic experts. Then, the outputs of the quality assessment methods can be easily compared to the assigned labels. But that becomes an impossible task for large-scale datasets one is interested in. Large-scale datasets that consist of fingerprint images obtained from multiple sensors in multiple sessions are the appropriate data for evaluation since they ensure a great diversity of the fingerprint quality.

Another solution can be to determine how the given biometric sample influences the accuracy of a recognition system. The quality assessment methods were designed to refuse samples of poor quality on input. The rationale behind this is that the poor quality samples trigger recognition errors in the biometric system and that results in degradation of the biometric performance. Therefore it is one of the open questions today, how to assign such a quality value to the biometric sample, that it predicts the biometric performance of the given sample in a recognition system. From now on such an evaluation of samples will be denoted as the *performance-based quality*.

The performance-based quality is convenient to use for several reasons. First, such a measure can be automatically computed for a dataset of any size. The results can be directly compared to the quality scores in order to determine to what degree the fingerprint image signal quality estimation methods are capable of predicting the impact of the given sample on the general performance of the biometric system. Further, it allows different quality assessment methods for benchmarking as well as identifying the overall quality of different datasets.

## 6.1 Previous Work

No extensive research has been done to measure the impact of a sample of a given quality on the general biometric performance. Many studies [42, 6] have shown the performance improvement based on rejecting poor quality samples but they did not investigate in quantifying the influence of a particular sample on the recognition performance.

In the study on fingerprint image quality carried out by Alonso-Fernandez et al. [2],

quality assessment methods were evaluated on pairs of samples in comparison. The quality value $Q$ was simply computed for a pair of samples as $Q = \sqrt{Q_e \times Q_t}$, where $Q_e$ and $Q_t$ represent the quality values of the enrolment and test fingerprint image, respectively. Quality value is than compared to the similarity score obtained from comparison of these two samples. But the experiments on the recognition performance [41] have shown that recognition errors were triggered by low quality samples. That indicates that each of the samples of comparison will have a different impact on performance. Therefore, assigning a quality score to the comparison pair is not desirable.

Another approach of fingerprint quality assessment was proposed by Tabassi et al. [38, 9] and recently, it was also included in the in ISO/IEC IS [16]. It measures the fingerprint quality by quantifying the comparator's performance. That makes it the right candidate for the evaluation of quality estimation methods, and therefore this framework will be followed.

## 6.2   Measuring Biometric Performance

A fingerprint image verification algorithm $V$ generates a similarity score $s_{ij}$ of two biometric samples $d_i$ and $d_j$ based on their local ridge characteristics (minutiae):

$$s_{ij} = V(d_i, d_j). \tag{6.1}$$

The similarity score of the samples coming from the same subject and the same instance is denoted as *genuine similarity score* $s_{ii}$. Samples of different instances in comparison result in *imposter similarity score* $s_{ij}$, $i \neq j$.

Consider a biometric dataset containing $T$ fingerprint samples. These samples were obtained from $M$ instances, by capturing $d_i^1, \ldots, d_i^{N_i}$ samples per each instance, $i = 1, \ldots, M$. That is

$$T = \sum_{i=1}^{M} N_i. \tag{6.2}$$

Let's denote $d_i^u$ as the $u$-th sample of instance $i$.

Using a particular recognition system $V_k, k = 1, \ldots, K$ of all $K$ available systems, for each dataset a set $S$ of similarity scores can be computed. Set $S$ will consist of 2 disjoint subsets: a subset of *genuine* similarity scores:

$$S_{ii} = \{s_{i,i}^{u,v} | s_{i,i}^{u,v} = V_k(d_i^u, d_i^v)\}, \tag{6.3}$$
$$\text{for } u = 1, \ldots, N_i \text{ and } v = 1, \ldots, N_i \text{ and } u \neq v$$
$$i = 1, \ldots, M$$

and a subset of *imposter* similarity scores:

$$S_{ij} = \{s_{i,j}^{u,v} | s_{i,j}^{u,v} = V_k(d_i^u, d_j^v)\}, \tag{6.4}$$
$$\text{for } u = 1, \ldots, N_i \text{ and } v = 1, \ldots, N_j$$
$$i = 1, \ldots, M \text{ and } j = 1, \ldots, M \text{ and } i \neq j.$$

The size of the subset of genuine similarity scores is

$$|S_{ii}| = \sum_{i=1}^{M} P(N_i, 2) = \sum_{i=1}^{M} N_i(N_i - 1), \tag{6.5}$$

in words there are $N_i - 1$ genuine scores per each sample of instance $i$.

And the size of the subset of imposter similarity scores is

$$|S_{ij}| = \sum_{i=1, j=1, j \neq i}^{M} N_i N_j, \tag{6.6}$$

which means that each sample $d_i^u$ of the dataset has $\sum_{j=1, j \neq i}^{M} N_j$ imposter scores.

Figure 6.1 shows the histogram of genuine and imposter similarity scores that were computed on the FVC2004 Db3 dataset [27] using the Neurotechnology VeriFinger SDK [29]. There are 8 samples per instance, 110 instances, which is in total 6 160 genuine and 767 360 imposter similarity scores. To visualize them altogether, the frequency of genuine and imposter scores was normalized in order to sum to one in total.
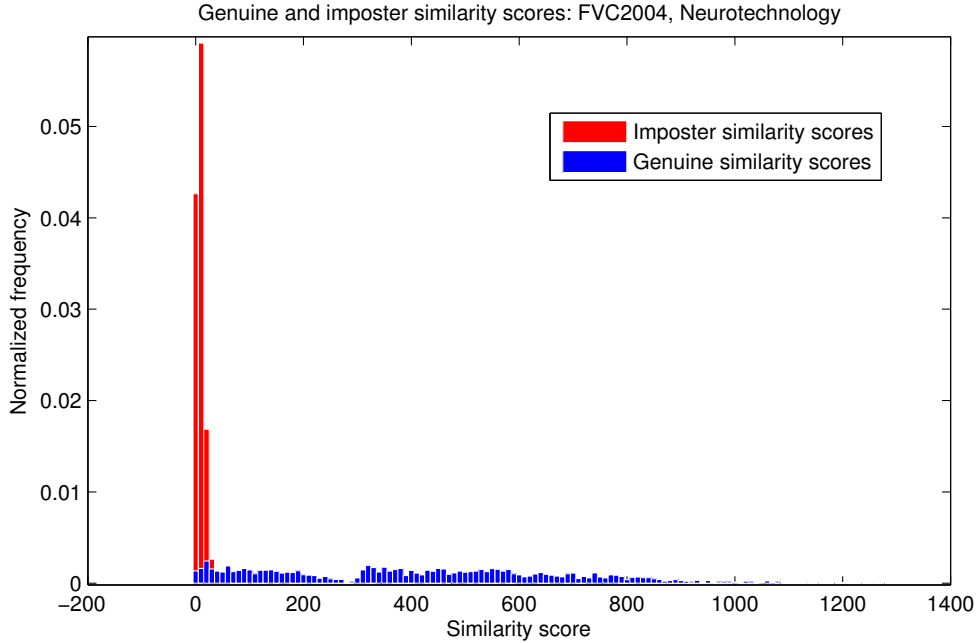


Figure 6.1: Normalized histogram of genuine and imposter similarity scores computed by Neurotechnology VeriFinger SDK [29] on the FVC2004 Db3 dataset [27].

Definition in [10] describes similarity score as a comparison score that increases with similarity. A higher similarity score indicates a higher likelihood that the samples come from the same subject. It is possible to observe this from the higher genuine scores in Figure 6.1. But at the same time you can also see that the genuine score distribution is much wider than the imposter score distribution. And even more, both the distributions are overlapping.

Overlapping distributions mean that a given sample $d_i^u$ results in a false match if there is a smaller genuine similarity score $s_{i,i}^{u,v}$ than some impostor similarity score $s_{i,j}^{u,w}$:

$$s_{i,i}^{u,v} < s_{i,j}^{u,w}, \tag{6.7}$$

$$u \in [1, N_i] \text{ and } v \in [1, N_i] \text{ and } w \in [1, N_j] \text{ and } u \neq v$$

$$i \in [1, M] \text{ and } j \in [1, M] \text{ and } i \neq j.$$

If the cumulative distribution function (CDF) is used to describe the distribution of the imposter similarity scores $P(s_i)$, than it describes the *False match rate (FMR)*:

$$FMR = 1 - P(s_i). \tag{6.8}$$

Similarly, the cumulative distribution function of genuine similarity scores, $P(s_g)$ represents the *False non-match rate (FMR)*:

$$FNMR = P(s_g), \tag{6.9}$$

Both the *FNMR* and *FMR* as described in Section 3 are common indicators of the biometric system performance. It indicates the fact, that the similarity scores can be used to construct a performance-based quality metric based on the genuine and imposter score distributions.

## 6.3  Quantifying Utility

If the performance-based quality should be predictive of the performance of the verification algorithm, *good quality* samples must have high genuine similarity scores and they should be also well separated from the imposter score distribution. On the other hand, *poor quality* samples should have lower genuine similarity scores, some of them being even similar to imposter scores. In that context, the metric describing the performance of the biometric system should measure how much the genuine distribution $P(s_g)$ is separated from the imposter distribution $P(s_i)$.

ISO/IEC IS [16] defines a term *utility* as the predictor of a positive or negative contribution of an individual sample to the overall performance of the biometric system. Moreover, stemming from the formulas described above the *utility* computation is proposed in the informative part of the ISO/IEC IS [16] as follows:

$$utility_i^u = \frac{m_{i,u}^{mated} - m_{i,u}^{non-mated}}{\sigma_{i,u}^{mated} + \sigma_{i,u}^{non-mated}}, \tag{6.10}$$

where $m_{i,u}^{mated}$ is the mean of sample $d_i^u$'s genuine similarity scores. It is a representative of the expected genuine similarity score. Similarly, $m_{i,u}^{non-mated}$ is the mean of sample $d_i^u$'s non-mated similarity scores as expected imposter similarity score. Similarly, $\sigma_{i,u}^{mated}$ is the standard deviation of sample $d_i^u$'s mated similarity score and $\sigma_{i,u}^{non-mated}$ is the standard deviation of sample $d_i^u$'s non-mated similarity score.

The utility value proposed conforms to our requirements on the performance oriented measure of sample quality. In contrast to comparison which involves two samples, utility is defined and measured per sample. Moreover, the quality values are continuous and can be normalized into range of $[0, 100]$ as recommended by the ISO/IEC IS [16]. Afterwards, regression methods can be applied in order to determine a linear or non-linear mapping from the quality score to the utility value.

Note that the utility as described here is not only dependant on the comparator subsystem. Also the process of feature extraction can contribute to the similarity score. The term recognition system was used to cover both.

## 6.4 Quality Levels

Transforming the continuous utilities into several distinct values makes evaluation more robust. It is also pragmatic to do so because of the following facts.

First, the proposed utility computation is dependent on the used recognition algorithm. Therefore different algorithms can yield different comparison scores and thus to result in different utility values.

Figure 6.2 shows the correlation of utilities computed on similarity scores generated by different recognition algorithms. Similarity scores were computed from total 880 samples of FVC2004 Db1 dataset [27]. The similarity scores were generated by two available verification algorithms: Neurotechnology VeriFinger SDK [29] and Innovatrics ANSI/ISO Template generator and comparator [13]. One can observe that there is a significant difference in utility output of two vendors. The linear relationship between the utility values, measured by the Pearson correlation coefficient [32], is 0.6 which gives evidence about the variation of utilities.
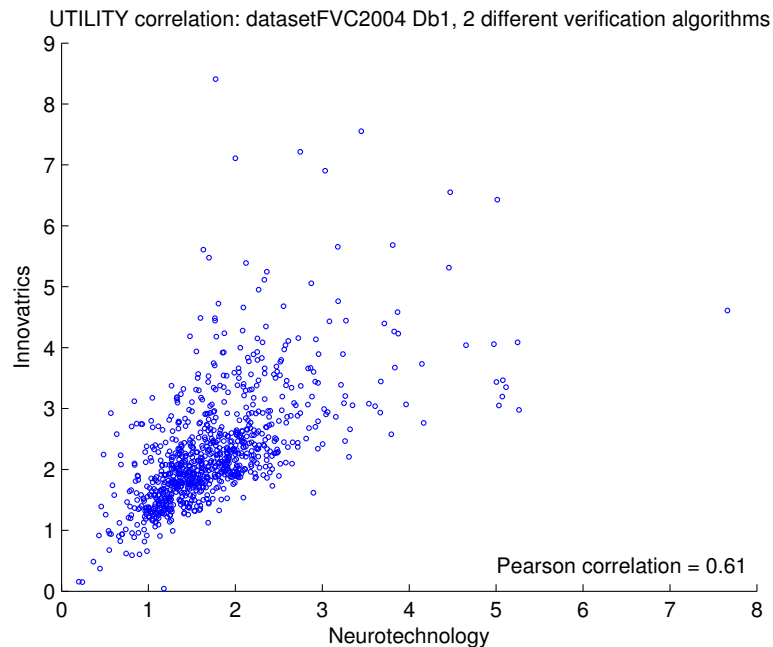


Figure 6.2: Correlation of the utilities computed on the FVC2004 Db1 dataset [27]. There are 880 utilities, one utility per sample, shown. Similarity scores were generated by Neurotechnology VeriFinger SDK [29] and Innovatrics ANSI/ISO Generator & Matcher [13]. To quantify the linear dependence, the Pearson correlation coefficient is shown.

Another fact is that some quality assessment algorithms might only provide a discrete output with few distinct values and such an algorithm would be difficult to compare to a continuous utility.

The target quality resolution must be selected with respect to the dataset size and the sensitivity of the verification algorithm on the input image quality. ISO/IEC IS [16] specifies the minimal resolution of the sample quality that still has the ability to discriminate among distinct levels of performance.

There are four distinct classes:

- "excellent"

- "adequate"

- "marginal"

- "unacceptable"

In this work, 5 classes were considered to provide sufficient quality resolution. They are shown in the Table 6.1.

Mapping utilities into individual levels (=*bins*) can be accomplished according to the ISO/IEC recommendation [16]. It is based on the separation of genuine and imposter similarity scores. But utility value itself does not include the information about which samples resulted in false matches at comparisons. Therefore the utilities will be additionally divided into two sets: *utilities of falsely matched samples* and *utilities of correctly matched samples*. The separation is based on the criteria 6.7.

Binning of the utility values should be designed in such a manner, that sufficient number of samples fall into each bin. That involves binning based on the distribution of population and not the range of utility. Two empirical cumulative distribution functions (CDF) should computed. One for utilities of falsely matched samples $W(.)$ and one for those that were correctly matched in all cases $C(.)$. Then the utilities are binned according to the quantiles of the utility distributions $C(.)$ and $W(.)$. The quantile functions are nothing else than the inverses of their CDF and therefore they are denoted as $C^{-1}(.)$ and $W^{-1}(.)$, where $C^{-1}(0)$ (or $W^{-1}(0)$) and $C^{-1}(1)$ (or $W^{-1}(1)$) denote the empirical minima and maxima, respectively. Possible bin boundaries proposed by ISO/IEC [16] are shown in Table 6.1. According to [9], $x = 0.25$ and $y = 0.75$ were chosen for the utility binning performed in this thesis.

| Quality bin | Label | Range of utilities |
|:---:|:---:|:---:|
| 1 | "unacceptable" | $\{z_i : -\infty < z_i < C^{-1}(0.01)\}$ |
| 2 | "marginal" | $\{z_i : C^{-1}(0.01) \leq z_i < W^{-1}(1)\}$ |
| 3 | "tolerable" | $\{z_i : W^{-1}(1) \leq z_i < C^{-1}(x)\}$ |
| 4 | "adequate" | $\{z_i : C^{-1}(x) \leq z_i < C^{-1}(y)\}$ |
| 5 | "excellent" | $\{z_i : C^{-1}(y) \leq z_i\}$ |

Table 6.1: Binning of the utility values according to the ISO /IEC [16].

Figure 6.3 shows the bin boundaries computed from the empirical CDF of utilities of correctly and falsely matched samples. The CDFs were computed on 678 correctly and 202 falsely matched samples of the FVC2004 Db3 dataset [27].

Finally, the performance-based quality $q_i^u$ is assigned to a sample $d_i^u$ according to the bin to which its $utility_i^u$ falls in.

Note that this quality is bounded to a particular recognition algorithm. To create a system-independent quality measure, the performance-based quality values need to be computed for all of the $K$ available recognition algorithms and afterwards fused into one according to an aggregation function. The aggregation function can be median or arithmetic mean or some other statistical expression of quality values.
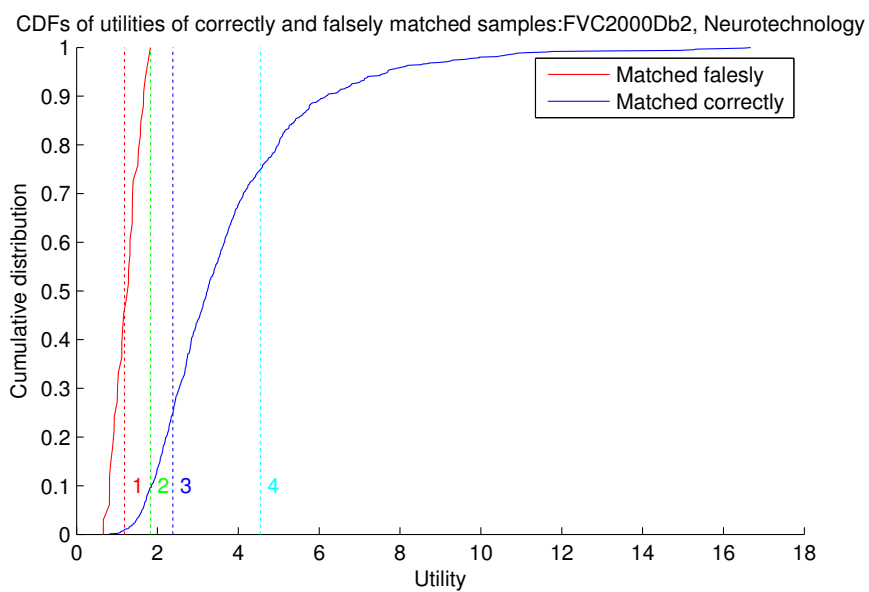
Figure 6.3: Empirical cumulative distribution functions of utilities computed from similarity scores of correctly and falsely matched samples of FVC2000 Db2 dataset [28]. Bin boundaries are shown ($1 = red, 2 = green, 3 = blue, 4 = cyan$). Similarity scores computed by Neurotechnology VeriFinger SDK [29].

# Chapter 7

# Dataset Selection

For a reliable evaluation of quality assessment methods it is important to select appropriate data, from which a general conclusion can be derived. Therefore, several aspects were considered while obtaining datasets.

First, for constructing a robust, performance-based quality measure it is necessary to have fingerprint samples of high inter-subject and intra-subject variability. Translated into English, a large-scale dataset is needed, which consists of many fingers scanned in multiple sessions.

Second, the quality assessment should not be bounded to a particular acquisition device. Datasets should consist of fingerprint images obtained from multiple sensors.

Third, more realistic results are obtained if data is collected in sessions with time intervals in between.

The rest of the chapter describes the datasets that were used for quality assessment as well as for creation of the performance-based quality metric.

The semi-public datasets were provided by the Center for Advanced Security Research Darmstadt, Germany.

## 7.1 FVC Datasets

Four international Fingerprint Verification Competitions (FVC) were organized in 2000, 2002, 2004 and 2006 [28, 26, 27, 7]. For each competition, a fingerprint dataset had been created. In this thesis, the first three datasets are used. The fourth, FVC2006 dataset, is not publicly available and despite of the author's effort towards the BIOSEG[1] group, the dataset was not provided till the submission date of this thesis.

Each dataset consists of four databases (Db1, Db2, Db3, Db4) that were acquired using three different sensors (Db1–3) and the SFinGE synthetic generator (Db4) [20]. This thesis only considers real fingerprints and so from each FVC dataset Db4 was omitted.

Each database has 110 fingers with eight impressions per finger, resulting in 880 impressions.

The databases were obtained with different intentions and so they differ in several aspects. The rest of this section provides a short description of each database with regard to sample quality [1].

---

[1]The Biometric Systems Lab of University of Bologna, http://bias.csr.unibo.it/research/biolab.

### 7.1.1 FVC2000

In FVC2000 [28], the acquisition conditions were different for each database (e.g. interleaving/not interleaving acquisition of different fingers, periodical cleaning/not cleaning of the sensor). For all the databases, no care was taken to assure the minimum quality of the fingerprints. The fingerprints are often rotated, however, it was assured that two impressions from the same finger always have an overlapping area.

Database Db1 was excluded from the quality assessment and evaluation proposed in this thesis as the author suspects it not to pass the requirements specified by ISO/IEC [14], also explained in Section 4. The observed problem is in an insufficiently small sensing area.

Sample images taken from each database are shown in Figure 7.1.



Figure 7.1: Example of a fingerprint image from each database of the FVC2000 dataset [28]. All the images were scaled down by the same factor.

### 7.1.2 FVC2002

In FVC2002 [26], the acquisition conditions were the same for each database: interleaved acquisition of different fingers to maximize differences in finger placement, no care was taken in assuring the minimum quality of the fingerprints and the sensors were not periodically cleaned. Some sessions were focused on an excessive displacement or rotation and some other on dry or moistened fingers.

An example of acquired images from each database are shown in Figure 7.2.
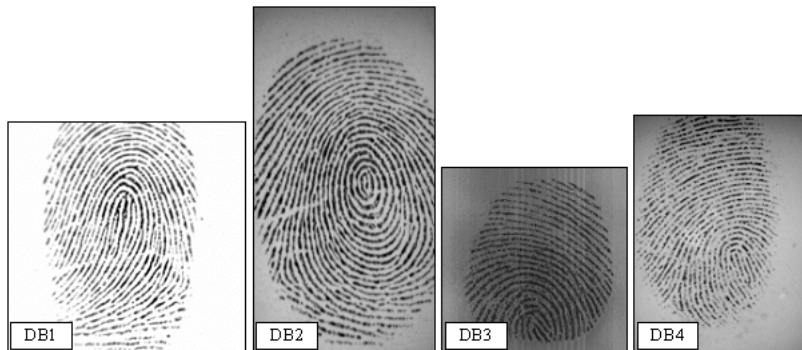


Figure 7.2: Sample images taken from databases Db1, Db2, Db3 and Db4 of FVC2002 dataset [26]. All the images were scaled down by the same factor.

### 7.1.3 FVC2004

The FVC2004 databases [27] were collected with the aim of creating a more difficult benchmark of verification algorithms than in FVC2002. Therefore, more intra-class variation was introduced. Different sessions were focused on

- different vertical positions of finger,

- low or high pressure of finger against the sensor,

- exaggerating skin distortion and rotation,

- dry or moistened fingers.

No care was taken to assure the minimum quality of the fingerprints and sensors were not periodically cleaned. Also, the acquisition of different fingers was interleaved to maximize differences in finger placement.

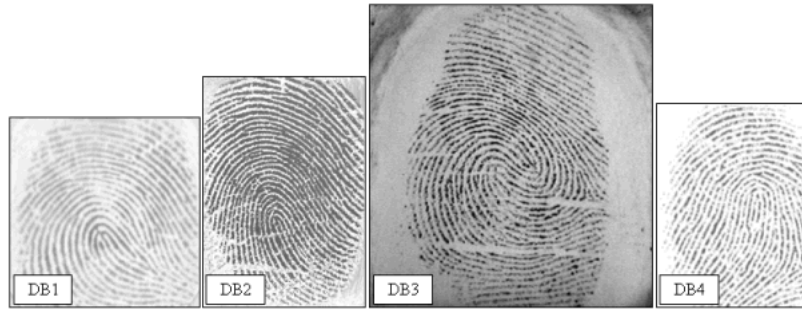Examples of acquired fingerprint images for each database are shown in Figure 7.3.
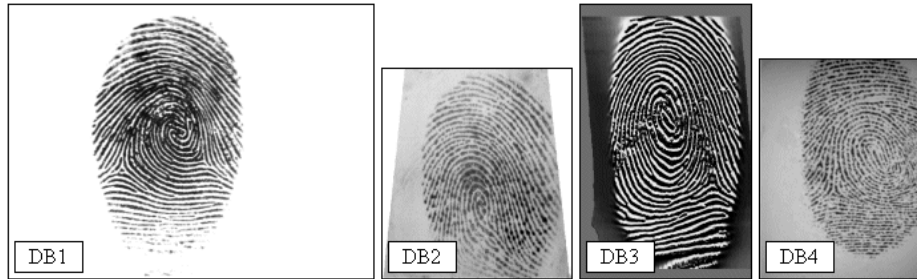


Figure 7.3: Example of fingerprint image from each database of FVC2004 dataset [27]. All the images were scaled down by the same factor.

## 7.2 MCYT Fingerprint Subcorpus

In the MCYT databases [31] the process of the fingerprint capture was accomplished under the supervision of an operator. Two types of acquisition devices were used. With respect to the different sensor properties, the dataset was divided into two sensor-specific databases, one for fingerprints obtained by capacitive sensor (mcyt330pb) and another one obtained by optical acquisition device (mcyt330dp).

In each case, a ten-print acquisition per individual was carried out. For each individual, 12 samples of each finger were acquired under different levels of control (high, medium and low). In total 330 individuals were involved, in all capture sessions each individual provides a total number of 120 fingerprint images to each database, which results in 330 x 10 sessions x 12 samples = 39 600 fingerprint samples for each database.

Example images of MCYT fingerprint subcorpus are shown in figure 7.4.

Figure 7.4: Fingerprint examples from MCYT Fingerprint subcorpus [31]. Fingerprint images are given for the same finger for both capacitive acquisition (left in each pannel) and optical acquisition (right in each pannel), and three different fingerprints in the database, one per column.

## 7.3 CASIA Fingerprint Image Database

The Fingerprint Image Database Version 5.0 [39] was created and provided by the the Chinese Academy of Sciences Institute of Automation (CASIA).

The database contains 20,000 fingerprint images of 500 subjects. The fingerprint images of CASIA-FingerprintV5 were captured in one session using optical fingerprint sensor. From each subject, 40 fingerprint images were acquired from eight fingers (left and right thumb/second/third/fourth finger), i.e. 5 images per finger. The volunteers were asked to rotate their fingers with various levels of pressure to generate a significant intra-class variation.

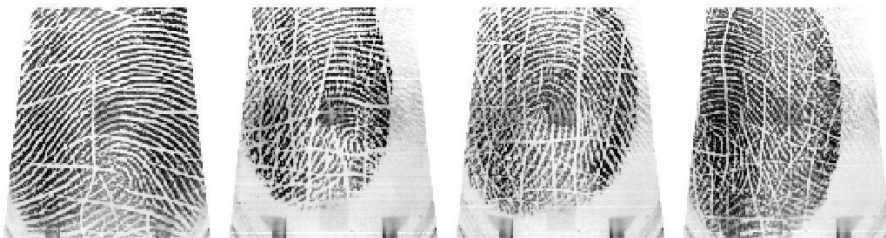Example images from CASIA fingerprint database are shown in Figure 7.5.



Figure 7.5: Example fingerprint images of CASIA-FingerprintV5 database [39].

# Chapter 8

# Evaluation

Previous chapters described the concept of quality estimation measures and performance oriented quality. Now these two concepts will be bound together and evaluated on the databases described in Section 7. The tools of explanatory data analysis will be used to visualize the results.

## 8.1  Quality Estimation

For each sample from all available databases a quality score vector was computed. The vector consists of the output values from global as well as local quality estimation methods described in Section 5.

To validate that the obtained data is correct, a visualization of its probability distribution should be done. E.g. an enormous amount of quality values close to the minimum (0) or maximum (1) value would indicate a possible error in design or implementation of the quality assessment methods or the dataset composition.

Figure 8.1 shows the distribution of quality scores for the largest available database - MCYT330dp. Each of the histograms was computed on 39 600 quality values. From all the histograms a desirable distribution of the data can be observed: the normalized output range (from 0 to 1) is adequately covered by the quality scores and moreover, there is an evident normal distribution of the scores.

A small difference can be seen for the output of the assessment method based on orientation flow where the scores are more concentrated close to the maximum (1). The fact can be assigned to the operational settings of the metric. Decreasing the tolerance of angular change in the ridge flow between the neighbouring blocks from 8°(as explained in the ISO/IEC Technical Report [19]) to a smaller angle would most likely yield lower quality scores. The assumption is based on the definition of the orientation flow metric described in Section 5.

A similar desirable distribution of quality scores has been observed for all available databases.

Plotting the distributions of given quality metrics also allows for a certain comparison of the databases. A comparison between databases MCYT330dp and FVC2002 Db3 is presented in Appendix A. Briefly, higher quality scores were observed in the MCYT330dp database, which testifies to the fact, that the fingerprint samples in MCYT sub-corpus were acquired under the supervision of an operator in the role of quality controller. By contrast, samples from database FVC2002 were acquired without any quality control, with the intent to obtain more challenging fingerprint samples with varying quality.

(a) Orientation Certainty Level $Q_{OCL}$  (b) Ridge-valley Structure Analysis: $Q_{LC}$

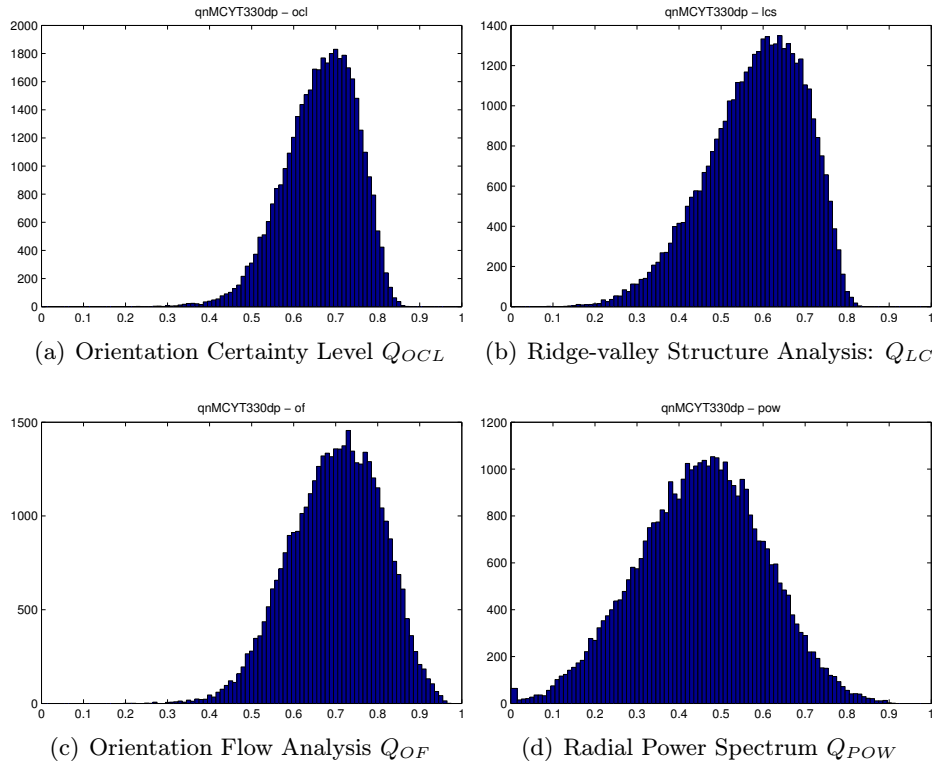(c) Orientation Flow Analysis $Q_{OF}$  (d) Radial Power Spectrum $Q_{POW}$

Figure 8.1: Distribution of the scores of local and global quality analysis methods on the MCYT330dp fingerprint database [31]. There are total 36 900 values in each histogram.

## 8.2 Relationship Among Quality Metrics

To improve the prediction capability of the quality apparatus, several quality metrics can be combined to produce a scalar quality score. High improvement can be achieved only if the metrics describe distinguishing properties of the fingerprint image. Therefore, it is important to analyse the relationship between quality scores of different metrics.

In Figure 8.2, scatter-plots were used to visualize the correlation between all pairs of quality metrics. The quality scores are shown for the FVC2002 Db3 database. Note that it is not necessary to observe a linear correlation between the quality values. The Spearman's rank correlation coefficient was chosen as a non-parametric measure of statistical dependence between two quality values [32].

It can be visually observed from Figure 8.2, that there is quite high correlation between the quality metrics. The strongest correlation is observed between the Ridge-valley structure analysis ($Q_{LC}$) and Orientation Certainty Level ($Q_{OCL}$) (in Figure 8.2(c)). It can be explained by the fact that both $Q_{LC}$ and $Q_{OCL}$ are local quality analysis methods and they use the same eigenvalues computed within blocks as a measure of ridge orientation.

On the contrary, the smallest correlation is observed between the Orientation Flow Analysis ($Q_{OF}$) and Radial Power Spectrum ($Q_{POW}$) (shown in Figure 8.2(f)).

The same degree of correlation between quality metrics was observed in all available databases. The obtained Spearman's correlation values can be a valuable information for the quality metrics fusion.
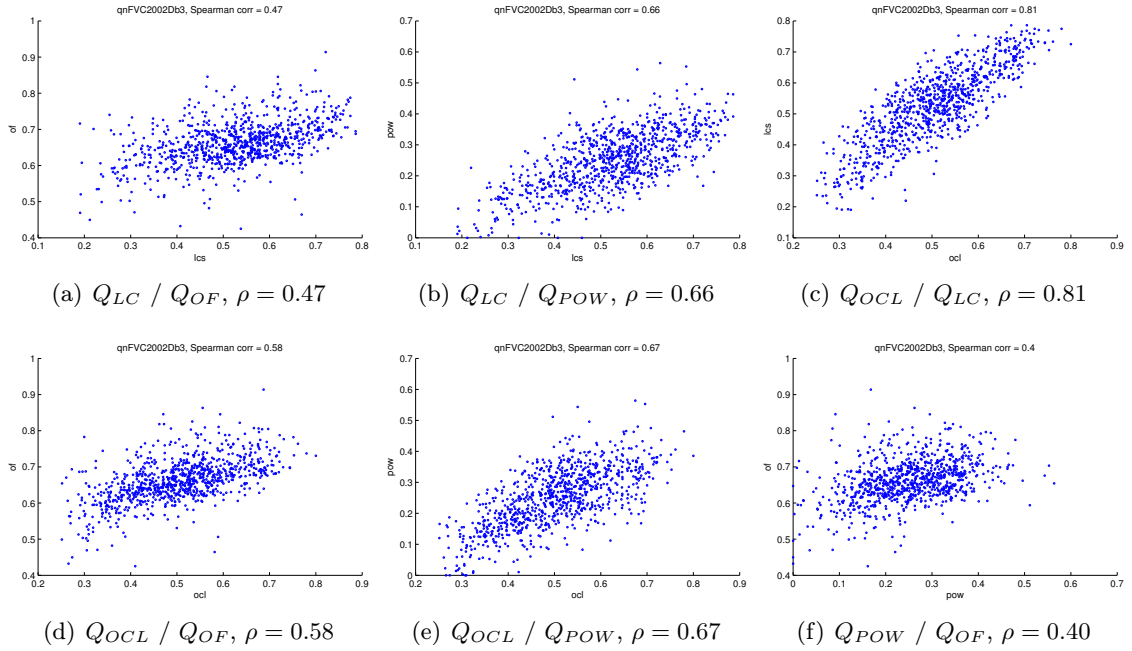
Figure 8.2: Correlation of the quality assessment metrics. $x$ and $y$ are the quality scores of two different methods under comparison. Additionally, the Spearman's correlation coefficient $\rho$ is used to quantify the non-parametric dependency of the quality values. Quality scores were computed on the FVC2002 Db3 [26].

## 8.3 Performance-oriented Evaluation

Important for evaluation of the quality metrics is to observe how the signal quality measures are capable of predicting the performance of a recognition system. In the Section 6, the utility and performance-based quality were defined as a measure of the biometric performance. Now these two measures will be computed for all the available datasets and compared to the output of quality estimation methods.

To build a utility, which will be indicative of general biometric performance, a great inter-class and intra-class variability is needed. The rationale behind is that utility is nothing else than a statistical measure of comparison scores. If this is the case, then more comparison scores mean more accurate results. Table 8.1 shows the amount of comparison scores that were computed and used to build the robust utility measure. In practice, the computation took about 1 month of CPU time, 440 000 of text files with similarity scores were generated and they occupy more than 180 GB of disk space.

Note that for comparison score computation, different instances of the same subject were considered as different subjects (e.g. left index finger and left middle finger of the same person were treated as samples of two different persons).

### 8.3.1 Database Integrity

An interesting artefact can be observed from the cumulative distribution function of utilities computed on MCYT330dp [31] in Figure 8.3. There is an evident change of the behaviour of the CDF for falsely matched samples which indicates an undesirable fact – there are many falsely matched samples with high utility values.

| Dataset | Comparison scores (by one vendor) | |
| --- | --- | --- |
| | Genuine | Imposter |
| FVC 2000 Db1, Db2, Db3 | | |
| FVC 2002 Db1, Db2, Db3 | $6 \times 6.16 \times 10^3$ | $6 \times 767.36 \times 10^3$ |
| FVC 2004 Db1, Db2, Db3 | | |
| CASIA | $80 \times 10^3$ | $79.98 \times 10^6$ |
| MCYT 330dp | $2 \times 435.6 \times 10^3$ | $2 \times 130 \times 10^6$ |
| MCYT 330pb | | |
| **Total**, for 2 vendors | 660 000 000 | |

Table 8.1: Statistics on comparison scores used for utility computation.

Looking back into Section 6, the utility value is computed from sample's genuine and imposter similarity scores. Correctly matched samples have high genuine similarity scores while low imposter similarity scores, which implies high utility. On the contrary, if there is at least one genuine score which is lower than any imposter similarity score, the sample is marked as falsely matched. These rules refer to the top-ranked conditions in ISO/IEC IS 29794-1 [16].
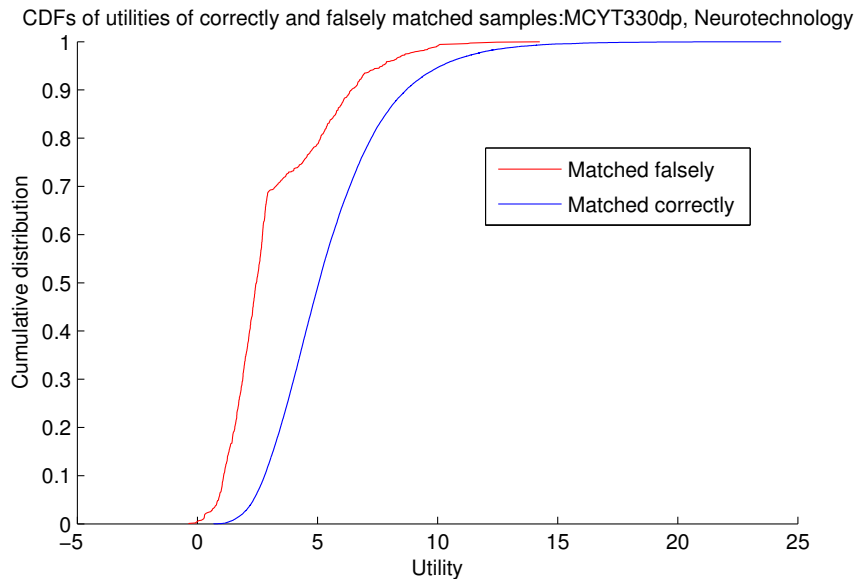


Figure 8.3: Example of the impact of sample labelling errors on the utility computation. Cumulative distribution functions of utilities of correctly and falsely matched samples. Similarity scores were computed by running the Neurotechnology VeriFinger SDK [29] on the MCYT330dp database [31].

If a sample is falsely matched and has a high utility, it indicates that it yields high genuine similarity scores and low imposter scores except for at least one sample. If it is a genuine sample, the similarity score has to be extremely small (smaller than any imposter score). If it is an imposter sample, the similarity score has to be extremely high (higher than any genuine score). And this is a highly unlikely scenario for fingerprints that were

obtained in a controlled environment.

As described in Section 7, fingerprint samples of MCYT sub-corpus were acquired under supervision of an operator who performs the quality check. These facts led the author to the theory that there are labelling errors in the MCYT330dp dataset and these errors can be easily detected by the implemented utility function.

Labelling errors in fingerprint dataset MCYT330dp as well as MCYT330dp were afterwards also confirmed by the University of Twente, Holland[1].

For performance evaluation, it is crucial for each fingerprint image to be correctly assigned to its source. It can be observed from the figure that there are at least 30 % of wrongly *falsely matched* samples. Detection and correction of the mislabelled samples is not in the scope of the thesis and therefore the author decided to exclude the entire MCYT330 fingerprint sub-corpus (dataset MCYT330dp and MCYT330pb) from further evaluation based on biometric performance.



(a) $Q_{LC}$, $\rho = 0.17$

(b) $Q_{OCL}$, $\rho = 0.22$

(c) $Q_{POW}$, $\rho = 0.23$

(d) $Q_{OF}$, $\rho = 0.022$
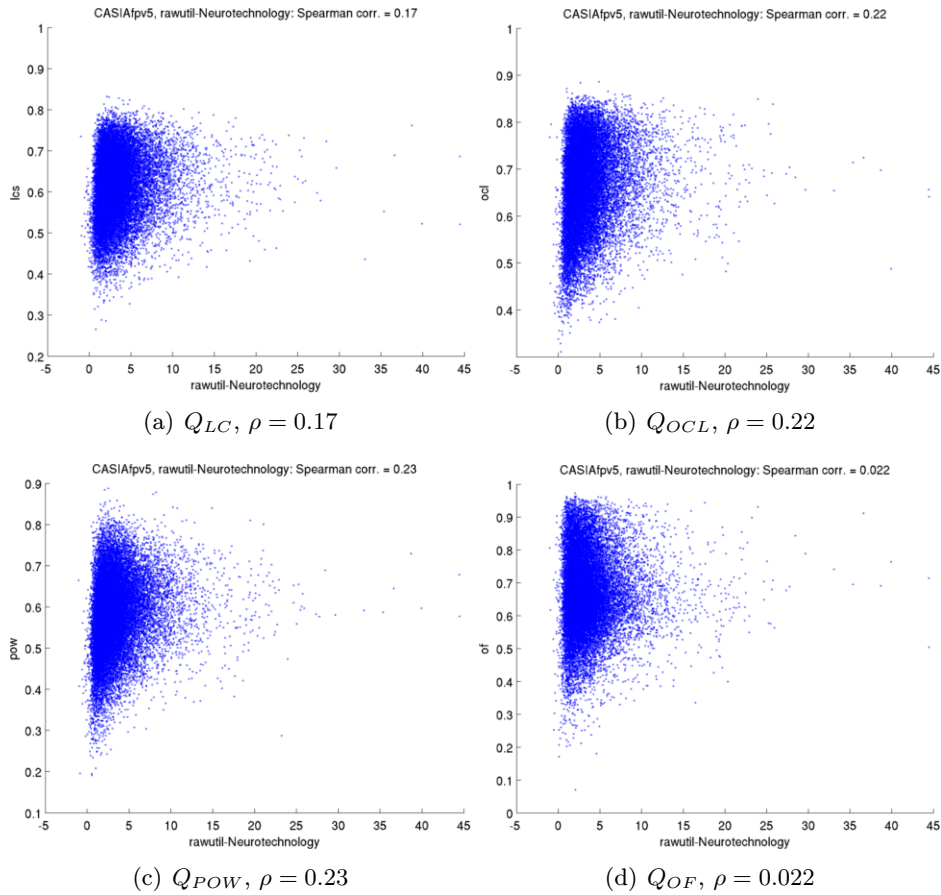
Figure 8.4: Correlation between predicted quality and measured utility. There are utilities on the $x$ axis and quality values of given metric on the $y$. The Spearman's correlation coefficient $\rho$ is shown. For the utility computation, similarity scores obtained from Neurotechnology VeriFinger [29] were used.

---

[1]To the best knowledge of the author, there has not been published any report describing labelling errors in MCYT [31] fingerprint sub-corpus yet.

### 8.3.2 Correlation Between Quality and Utility Scores

Computed utility can be directly plotted against the predicted quality to see how the particular quality measure is able to predict the impact of a sample on the biometric performance.

Figure 8.4 shows the correlation between particular quality metric and utility of the samples from the CASIA dataset [39]. Again, the Spearman correlation coefficient is used to measure the non-linear relationship between the quality and utility.

Very low correlation was observed for all the quality metrics on the CASIA but also the FVC databases. Such a poor dependency makes it impossible to find a regression method which will map quality values to utilities. The same problem was encountered by Tabassi et al. [38]. Thus, with reminder to the Section 6, the utilities will be further divided into fewer classes.

## 8.4 Performance-based Quality

Raw utility scores are not used for the known reasons (see Section 6). Instead, the binned utility was designed and it will be referred to as the *performance-based quality* (PBQ).

However, there is still one more problem with utility. Utility values represent the performance of a particular recognition system. Therefore, to improve generality of the utility measure, another recognition algorithm (Innovatrics ANSI/ISO template generator and comparator [13]) was involved to generate the similarity scores for utility computation. Figure 8.5 shows CDFs of utilities of two different vendors. As one can see, each vendor produces different similarity scores and so also the range of utility values varies for each vendor. To create more robust performance-based quality, utility values from different vendors will be fused.



(a) Neurotechnology  (b) Innovatrics
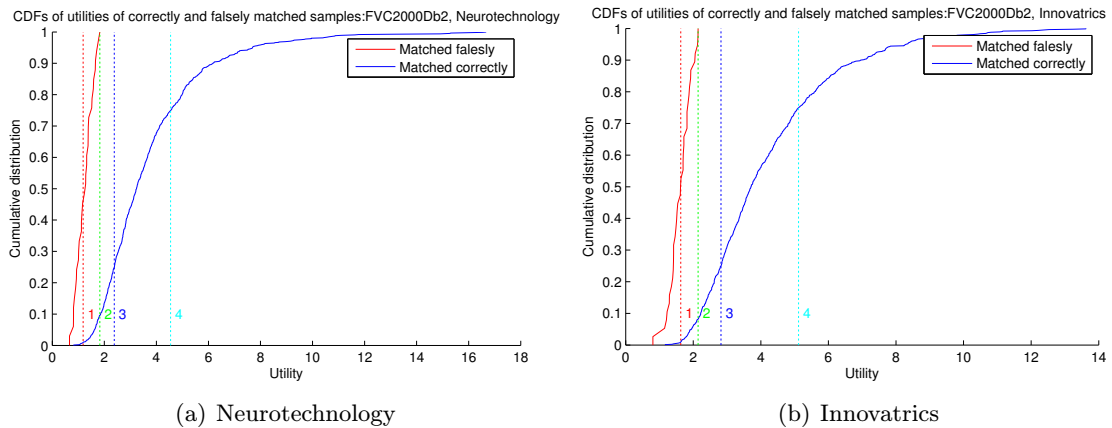
Figure 8.5: CDFs of utilities of correctly and falsely matched samples for two vendors [29, 13]. Boundaries between bins shown ($1 = red, 2 = green, 3 = blue, 4 = cyan$).

For the fusion of utilities, the rule of *unanimity* was chosen. Only the samples with identical utility bins from both recognition algorithms will be considered for further evaluation. The others will be discarded.

Once the PBQ values were computed according to the proposal in Section 6, they can be compared with the output of the signal quality assessment methods. Figure 8.6 shows the

correlation between each quality metric and PBQ on the CASIA dataset. As there are only 5 distinct classes, box-plot figures were chosen for visualization. They show the distribution of each quality measure for each class of PBQ. Only insignificantly higher correlation was achieved compared to to the results in Figure 8.4 although the continuous scale of utilities was quantized into 5 classes and 2 different recognition algorithms were incorporated.
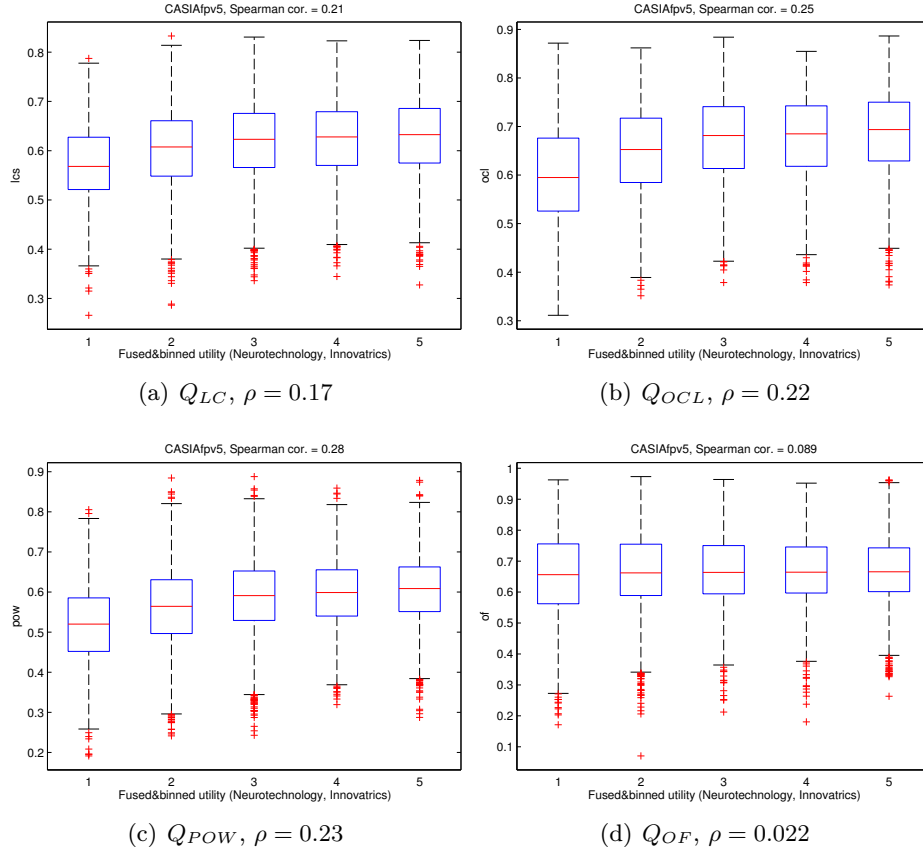


Figure 8.6: Correlation between predicted quality and PBQ. On the $x$ axis there is PBQ, on $y$ axis quality values of given metrics. Box-plots are used to visualize the distribution of quality measure for each class of PBQ. The red line inside the box represents the median, the blue boxes show the inter-quartile range of quality scores, whiskers extend to the most extreme data points not considered to be outliers. Outliers are plotted individually by red points. The Spearman's correlation coefficient $\rho$ is shown. For the PBQ computation, similarity scores obtained from Neurotechnology VeriFinger [29] and Innovatrics SDK [13] were used.

## 8.5  Proposed Utility Modifications

The poor dependency between the signal quality scores and PBQ values indicates a poor capability of the signal quality measures to predict the biometric performance. However, many studies have demonstrated [2, 42, 6] that rejecting samples that were assessed by the signal quality measures as low-quality improves the general performance of the biometric system. This obvious contradiction requires a deeper analysis of the process of utility

computation and also of the impact of one sample to the general biometric performance. The analysis is performed in the next sections.

### 8.5.1 Score Distributions

Looking back into Section 6, concretely at the computation of utility in Equation 6.10, statistical mean is proposed by ISO/IEC [16] as the mathematical expectation of the genuine (imposter) similarity score. And that can be only true if the distribution of genuine (imposter) similarity scores is gaussian. To verify that, a normality test was performed on the similarity scores for each sample. *Jarque-Bera test JB* [21] is defined as:

$$JB = \frac{|S_{ii}|}{6}\left(U^2 + \frac{1}{4}R^2\right),$$

(8.1)

where $|S_{ii}|$ is the number of similarity scores for subject $i$, $U$ is the skewness of sample's similarity score distribution and $R$ is the kurtosis of the sample's score distribution.

| Dataset | Samples with normal distribution | |
|---|---|---|
| | genuine comp. s. | imposter comp. s. |
| FVC 2000 Db1 | 8.07% | 100% |
| FVC 2000 Db2 | 5.68% | 100% |
| FVC 2000 Db3 | 13.41% | 100% |
| FVC 2002 Db1 | 5.88% | 100% |
| FVC 2002 Db2 | 6.59% | 100% |
| FVC 2002 Db3 | 8.64% | 100% |
| FVC 2004 Db1 | 8.18% | 100% |
| FVC 2004 Db2 | 7.84% | 100% |
| FVC 2004 Db3 | 8.23% | 100% |
| CASIA FPv5.0 | 6.19% | 100% |

Table 8.2: Jarque-Bera normality test. Percentage of samples with normal genuine and imposter similarity score distribution. Scores obtained by Neurotechnology VeriFinger [29]. The MCYT databases were omitted as they consist of mislabelled samples that negatively influence the score distribution.

Table 8.2 shows the percentage of samples with normal genuine and imposter distribution. Only around 8% of samples have a normal genuine distribution and therefore the mean value proposed in ISO/IEC is definitely not representing the expected value. The *median* and corresponding *absolute median deviation (MAD)* were preferred by the author as statistically robust measures, being more resilient to outliers than mean and standard deviation [11]. The utility is then computed as

$$utility_i^u = \frac{median_{i,u}^{mated} - median_{i,u}^{non-mated}}{MAD_{i,u}^{mated} + MAD_{i,u}^{non-mated}},$$

(8.2)

with the same notation of sample and subject as in the initial Equation 6.10.

An illustration of the difference between the old and new utility distribution is given in Figure 8.7. It is visible that there are more falsely matched samples with high utility values. The reason of the effect will be described in Section 8.5.2. There are also some

extreme outliers among the falsely matched samples and therefore the corresponding CDF stretches more towards higher utility values. It is out of the interest. Further, it can be noticed that the bins now overlap. It means that for the utility computed from medians, a different binning has to be set which is a reasonable effect.



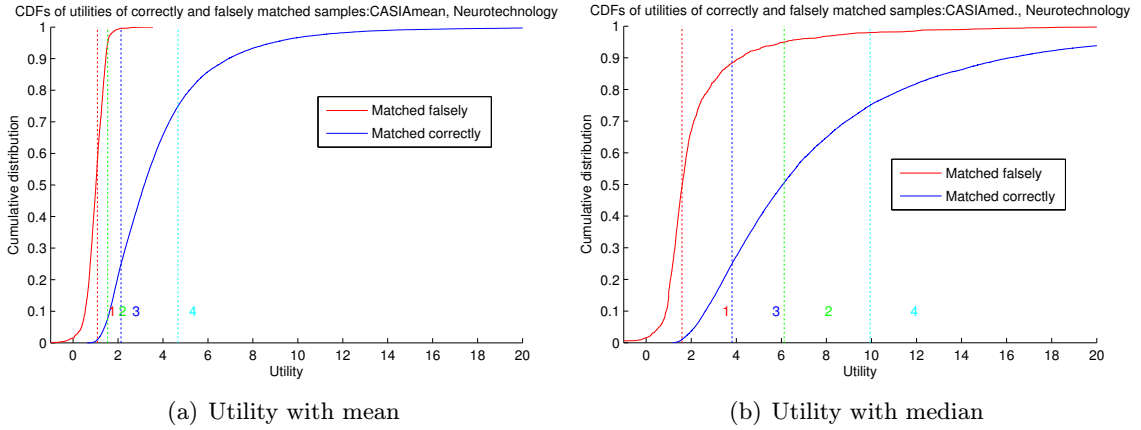(a) Utility with mean    (b) Utility with median

Figure 8.7: CDFs of utilities of correctly and falsely matched samples computed according to the old and newly proposed equation. Boundaries between bins shown ($1 = red, 2 = green, 3 = blue, 4 = cyan$).

## 8.5.2 Falsely and Correctly Matched Samples

Table 8.2 showed that samples genuine similarity scores do not fit the normal distribution. It often means that there was some low genuine similarity score for a subject while the others were high.

As an example of this problem, a similarity score matrix for subject 57 from database FVC2000 Db2 is shown in Table 8.3. In the table, probe 1 yields low similarity scores with most of the other genuine samples. For some of them, it is even lower than some imposter similarity score (red cells in the table). According to the utility division proposed by ISO/ IEC [16], all such samples will be marked as falsely matched (red lines/colloums in table). As a result these samples will end up with quite a low utility and a very low PBQ.

However, it can also be seen from the table that these samples yield high genuine similarity scores with all the other samples which means they provide high recognition performance with all except one sample (in this case it is probe 1). On that account, it is desirable to assign low PBQ only to sample 1 and not to all the falsely matched samples.

In the proposed modified utility computed from median values, all the samples except from #1 would gain a high utility though they would be still marked as falsely matched. If the binning system is adjusted to allow false matched samples with high utility to have high PBQ, the problem is solved.

## 8.6 Results Statement

With respect to the mentioned issues, modifications in the utility computation were made and all utility values were recalculated according to them.

| Probe\Ref. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 | 1629 | 389 | 327 | 143 | 26 | 62 | 17 | 20 |
| 2 | | 2087 | 1103 | 680 | 189 | 179 | 132 | 105 |
| 3 | | | 2126 | 663 | 207 | 200 | 165 | 152 |
| 4 | | | | 2144 | 123 | 104 | 222 | 206 |
| 5 | | | | | 2132 | 348 | 320 | 285 |
| 6 | | | | | | 1925 | 210 | 129 |
| 7 | | | | | | | 2066 | 774 |
| 8 | | | | | | | | 1818 |

Table 8.3: Similarity score matrix for subject 57 from FVC 2000 Db2 [28].

To construct the performance-based quality, a new binning has to be performed. To achieve a bigger quality coarseness, utilities are divided just into two classes (bins) – samples with insufficient quality (labelled as *"poor PBQ"*) and samples with sufficient quality (*"good PBQ"*). The boundary between the classes was set to the utility value reached by 90% of the falsely matched samples. Such a coarse division enables the relationship with signal-quality measures to be analysed in a more robust way.

In Figure 8.8 quality measures were plotted one against another (analysis on the CASIA dataset). Each sample from the dataset was labelled according to its PBQ value (class). An evident correlation between the quality measures can be observed but the samples of different PBQ classes strongly overlap. The conclusion of the study is that the signal quality measures themselves *cannot achieve high accuracy in prediction of the recognition performance.*

Note that this statement is not in contradiction with the studies [2, 6] stating that they have achieved improvement of the biometric performance based on rejecting samples with poor quality. These poor signal-quality samples are in Figure 8.8 scattered in the lower left corner in each plot. For a suitable combination of multiple quality measures, (in our example probably $Q_{OCL}$ with $Q_{POW}$ 8.8(b)), most of the samples assigned as of poor signal-quality will be of poor PBQ, too. Thus sample rejection based on the signal quality measures will certainly achieve some biometric performance improvement.

**Why the signal quality metrics do not predict the recognition performance?**
Further investigations have revealed that the biggest impact on the performance of minutiae-based recognition systems have mutually inconsistent samples of the same subject.

For explanation, subject 57 from FVC 2000 Db2 database was again chosen as an example. All the samples of the subject are shown in Figure 8.9. The genuine similarity scores matrix is provided in Table 8.3. It can be seen immediately from the images that the sample 1 only includes a partial impression of the finger. Looking at the similarity score matrix, it yields low similarity scores with the other genuine samples (red row) and thus it results in low utility[2]. But the low utility is in contradiction with the behaviour of all the signal quality measures since they assigned high quality values for sample 1! A visual control confirms that the output of signal quality assessment methods is correct – the fingerprint image is of a high biometric signal quality. It consists of clear, distinguishable ridges and valleys from which the features can be unambiguously extracted (description of fingerprint quality by Ratha et al. [33]). Therefore, the low utility and the low biometric

---

[2]The new utility was computed.
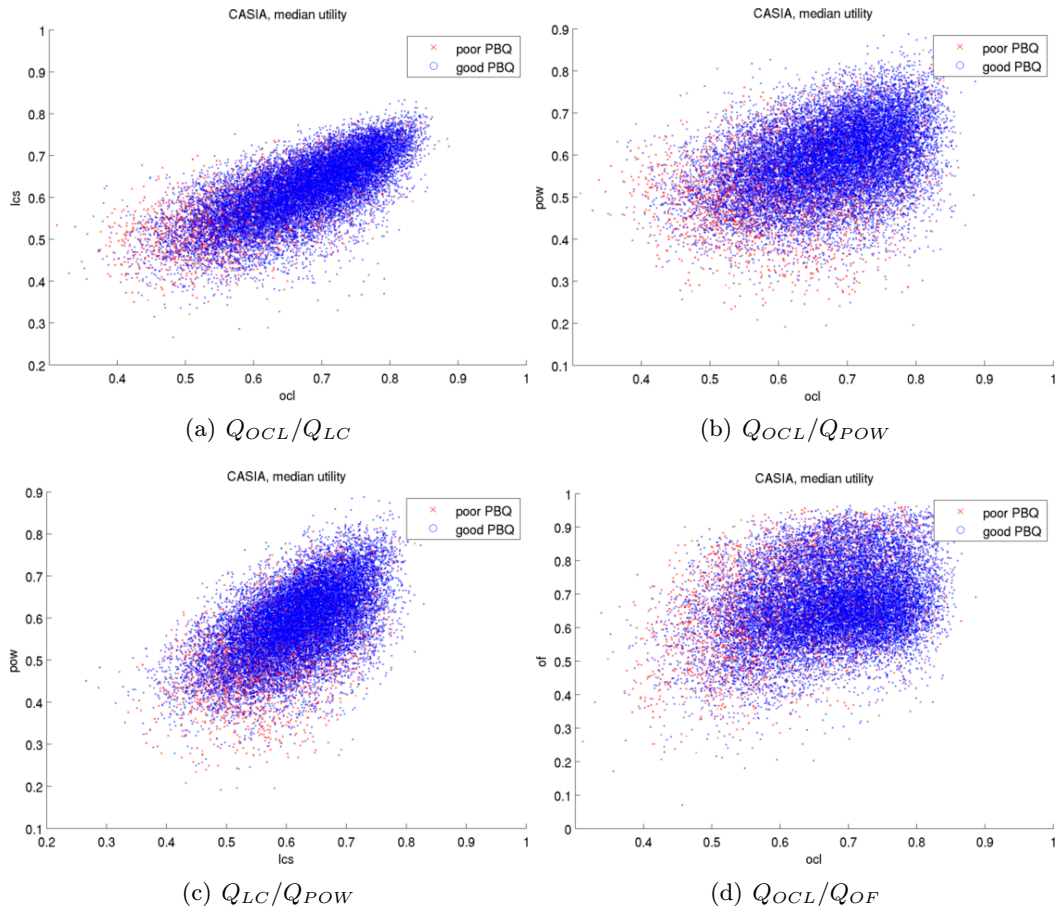
Figure 8.8: Relationship between two given signal quality measures ($x$ and $y$ axes) and performance-based quality, represented by two classes: "poor" and "good". Quality assessment performed on the fingerprints of CASIA dataset [39]. PBQ computed on the similarity scores obtained from Neurotechnology Verifinger [29] and Innovatrics SDK [13].

performance on that sample is caused by the number of minutiae detected and does not refer to the extract-ability of features. This is the reason why signal quality metrics are not in a strong correlation with PBQ.

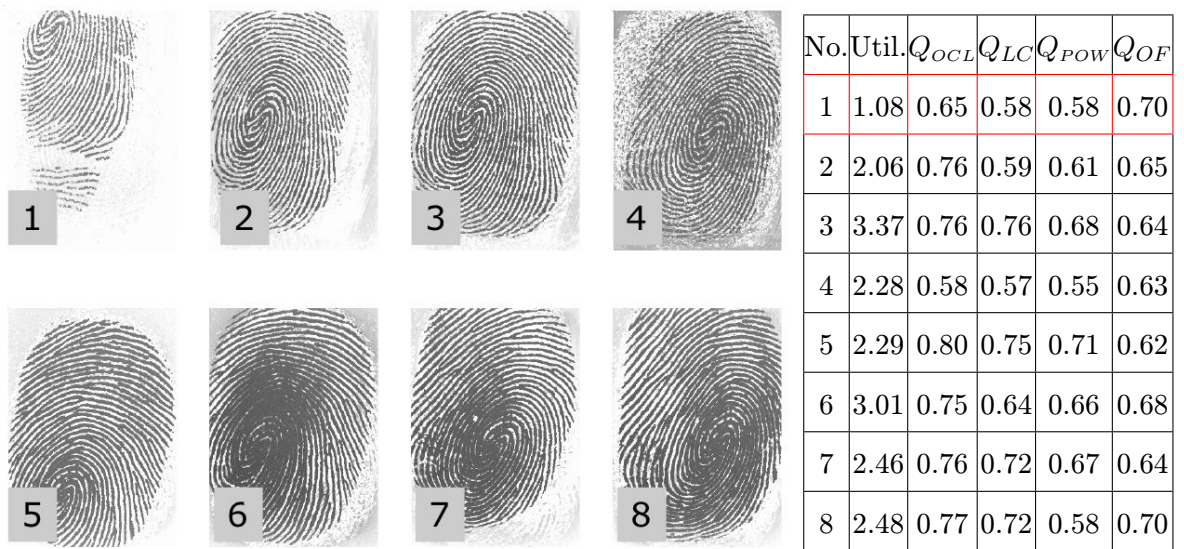| No. | Util. | $Q_{OCL}$ | $Q_{LC}$ | $Q_{POW}$ | $Q_{OF}$ |
|---|---|---|---|---|---|
| 1 | 1.08 | 0.65 | 0.58 | 0.58 | 0.70 |
| 2 | 2.06 | 0.76 | 0.59 | 0.61 | 0.65 |
| 3 | 3.37 | 0.76 | 0.76 | 0.68 | 0.64 |
| 4 | 2.28 | 0.58 | 0.57 | 0.55 | 0.63 |
| 5 | 2.29 | 0.80 | 0.75 | 0.71 | 0.62 |
| 6 | 3.01 | 0.75 | 0.64 | 0.66 | 0.68 |
| 7 | 2.46 | 0.76 | 0.72 | 0.67 | 0.64 |
| 8 | 2.48 | 0.77 | 0.72 | 0.58 | 0.70 |

Figure 8.9: Example of the fingerprint sample-quality assessment. On the left there are fingerprint samples of one subject shown. On the right, sample utility values and quality vectors are computed. The data are from subject 57 from FVC 2000 Db2 [28]. The utilities were computed from similarity scores obtained from Neurotechnology VeriFinger [29].

**What is the solution?**
All the methods that are included in the evaluation only deal with the signal quality of fingerprints and not with the fingerprint area and position. The only metric which is relatively capable to reflect the size of the area of the fingerprint pattern is the method based on the analysis of power spectrum in frequency domain ($Q_{POW}$, described in Section 5). It was designed with respect to the ISO/IEC IS [16] and measures the maximum of the energy concentration in the power spectrum of the frequency band corresponding to a fingerprint pattern. A complete fingerprint pattern gives a stronger energy concentration and thus $Q_{POW}$ achieved a higher correlation with utility and PBQ than any other method.

To sum it up, for systematic evaluation of fingerprint signal quality assessment methods, it is necessary to construct a large-database that consists of images of varying quality but where the completeness of the fingerprint images is guaranteed.

Another possible solution is to combine signal quality metrics with measures dealing with the completeness of the fingerprint pattern. The methods that reflect the presence and position of the core point within the fingerprint image could achieve a higher performance prediction capability, especially for recognition systems based on minutiae detection. A promising technique could be a measure of the distance of the core point from the centre of gravity of the fingerprint pattern.

# Chapter 9

# Conclusion

As long as the fingerprint recognition is used in biometrics, the problem of fingerprint quality remains open. But there is an evidence of change. Standardization process on defining fingerprint quality has begun. In 2010, when this thesis was initiated, ISO/IEC published a Technical Report [19] on fingerprint image quality. This thesis was conducted in consideration of the mentioned report.

The thesis presents a comparative study of fingerprint image signal quality estimation methods. Crucial aspects that influence the fingerprint quality were described and connected to the definition of biometric sample quality. The selected state-of-the-art approaches for fingerprint image signal quality estimation were described and implemented in MATLAB. Further, the problem of evaluation of these methods was explained and a performance-based quality measure was introduced as a possible way out. Finally, the evaluation was performed on large-scale multi-sensor and multi-session databases in order to determine the capability of the fingerprint estimation methods to predict the performance of the recognition algorithm.

While designing and implementing the quality assessment workflow in Section 5, several errors were detected in the published ISO/IEC Technical Report [19]. They refer to the issues of score normalization and aggregation function. The performance-based evaluation in Section 8 disclosed another problem in the publicly accepted utility computation, already included in the informative part of the ISO/IEC International Standard [16]. Improper statistical measures (mean and standard deviation) and an inappropriate design based on the utility binning lead to the inaccuracies in computation of sample's utility, especially for fingerprints with high intra-subject variability. All these comments have been formulated and soon they will be addressed to the ISO JTC1/SC37 technical committee for further investigation.

Also practical applications can benefit from the evaluation proposed in the thesis. Recently, NIST[1] together with BSI[2] initiated a joint project on developing a novel measure of fingerprint quality, more accurate than the current NIST Fingerprint Image Quality (NFIQ) algorithm. This thesis has identified several areas of the potential improvement of the NFIQ and the results proposed here will be considered for the new proposal.

To the best knowledge of the author, no previous studies on the capability of quality assessment methods to predict the biometric performance have been published. It is still a matter of further investigation. To considerably improve the prediction of the biomet-

---

[1]American National Institute of Standards and Technology, USA.
[2]Federal Office for Information Security, Germany.

ric performance, the methods assessing signal quality should be combined together with methods detecting fingerprint consistency and completeness. Besides, experiments have shown that some quality measures are more suitable for a specific acquisition technology. Therefore, future work should also involve the study of differences in behaviour of quality assessment methods.

# Bibliography

[1] ALONSO-FERNANDEZ, F., AND FIÉRREZ, J. Fingerprint databases and evaluation. In *Encyclopedia of Biometrics*. 2009, pp. 452–458.

[2] ALONSO-FERNANDEZ, F., FIÉRREZ-AGUILAR, J., ORTEGA-GARCIA, J., GONZALEZ-RODRIGUEZ, J., FRONTHALER, H., KOLLREIDER, K., AND BIGÜN, J. A comparative study of fingerprint image-quality estimation methods. *IEEE Transactions on Information Forensics and Security 2*, 4 (2007), 734–743.

[3] CHEN, T., JIANG, X., AND YAU, W. Fingerprint image quality analysis. In *Image Processing, 2004. ICIP '04. 2004 International Conference on* (oct. 2004), vol. 2, pp. 1253 – 1256 Vol.2.

[4] CHEN, Y., DASS, S., AND JAIN, A. Fingerprint quality indices for predicting authentication performance. In *In: Proc. AVBPA, Springer LNCS-3546* (2005), pp. 160–170.

[5] FIERREZ-AGUILAR, J., GARCIA A, J. O., AND RODRIGUEZ A, J. G. Discriminative multimodal biometric authentication based on quality measures. *Pattern Recognition 38*, 5 (2005), 777–779.

[6] FRONTHALER, H., KOLLREIDER, K., BIGUN, J., FIERREZ, J., ALONSO-FERNANDEZ, F., ORTEGA-GARCIA, J., AND GONZALEZ-RODRIGUEZ, J. Fingerprint image-quality estimation and its application to multialgorithm verification. *Information Forensics and Security, IEEE Transactions on 3*, 2 (june 2008), 331 –338.

[7] FVC2006. Fingerprint verification competition. http://bias.csr.unibo.it/fvc2006/default.asp, 2006. [online], [accessed 2010-10-15].

[8] GONZALEZ, R. C., AND WOODS, R. E. *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006.

[9] GROTHER, P., AND TABASSI, E. Performance of biometric quality measures. *IEEE Transactions on Pattern Analysis and Machine Intelligence 29* (2007), 531–543.

[10] HARMONIZED BIOMETRIC VOCABULARY. http://www.3dface.org/media/vocabulary.html, [online], [accessed 2010-10-11].

[11] HOAGLIN, D., MOSTELLER, F., AND TUKEY, J. *Understanding Robust and Exploratory Data Analysis*. John Wiley & Sons, New York, 1983.

[12] HONG, L., JAIN, A., PANKANTI, S., AND BOLLE, R. Fingerprint enhancement. 202–207.

[13] INNOVATRICS. ANSI/ISO generator and matcher 1.52. http://www.innovatrics.com/products/fingerprint-identification-sdk/ansiiso-sdk, 2011. [online], [accessed 2011-04-20].

[14] ISO/IEC IS 19794-4. Information technology – Biometric data interchange formats – Part 4: Finger image data, 2005.

[15] ISO/IEC IS 19795-1. Information technology – Biometric performance testing and reporting – Part 1: Principles and framework, 2006.

[16] ISO/IEC IS 29794-1. Information technology – Biometric smaple quality – Part 1: Framework, 2009.

[17] ISO/IEC JTC1 SC37. Biometrics: ISO/IEC JTC1 SC37: Harmonized biometric vocabulary.

[18] ISO/IEC JTC1 SC37. Biometrics: ISO/IEC JTC1 SC37: Text of Standing Document 11 (SD 11), Part 1 Overview Standards Harmonization Document.

[19] ISO/IEC TR 29794-4. Information technology – Biometric smaple quality – Part 4: Finger image data, 2010.

[20] JAIN, A. K., AND MALTONI, D. *Handbook of Fingerprint Recognition*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2003.

[21] JARQUE, C. M., AND BERA, A. K. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters 6*, 3 (1980), 255 – 259.

[22] KANOPOULOS, N., VASANTHAVADA, N., AND BAKER, R. L. Design of an image edge detection filter using the sobel operator. *IEEE Journal of Solid-state Circuits 23* (1988), 358–367.

[23] LIM, E., JIANG, X., AND YAU, W.-Y. Fingerprint quality and validity analysis. In *ICIP (1)* (2002), pp. 469–472.

[24] LIM, E., TOH, K.-A., SUGANTHAN, P., JIANG, X., AND YAU, W.-Y. Fingerprint image quality analysis. In *Image Processing, 2004. ICIP '04. 2004 International Conference on* (oct. 2004), vol. 2, pp. 1241 – 1244 Vol.2.

[25] LIU, L., TAN, T., AND ZHAN, Y. Based on svm automatic measures of fingerprint image quality. In *Proceedings of the 2008 IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application - Volume 01* (Washington, DC, USA, 2008), PACIIA '08, IEEE Computer Society, pp. 575–578.

[26] MAIO, D., MALTONI, D., CAPPELLI, R., WAYMAN, J., AND JAIN, A. K. Fvc2002: Second fingerprint verification competition. In *In Proceedings of 16th International Conference on Pattern Recognition (ICPR2002), Quebec City* (2002), pp. 811–814.

[27] MAIO, D., MALTONI, D., CAPPELLI, R., WAYMAN, J. L., AND JAIN, A. K. Fvc2004: Third fingerprint verification competition. In *in Proceedings of the First International Conference on Biometric Authentication* (2004), pp. 1–7.

[28] MAIO, D., MALTONI, D., WAYMAN, J. L., AND JAIN, A. K. Fvc2000: Fingerprint verification competition. *IEEE Transactions on Pattern Analysis and Machine Intelligence 24* (2002), 402–412.

[29] NEUROTECHNOLOGY. Neurotechnology VeriFinger Software Development Kit. http://www.neurotechnology.com/verifinger.html, 2011. [online], [accessed 2011-04-20].

[30] NIST. The NIST special database 27A. http://www.nist.gov/itl/iad/ig/sd27a.cfm, [online], [accessed 2010-10-11].

[31] ORTEGA-GARCIA, J., FIERREZ-AGUILAR, J., SIMON, D., GONZALEZ, J., FAUNDEZ-ZANUY, M., ESPINOSA, V., SATUE, A., HERNAEZ, I., IGARZA, J. J., VIVARACHO, C., ESCUDERO, D., AND MORO, Q. I. MCYT baseline corpus: A bimodal biometric database. *Vision, Image and Signal Processing, IEE Proceedings-150*, 6 (2003), 395–401.

[32] PAPOULIS, A. *Probability, Random Variables and Stochastic Processes*, 3rd ed. McGraw-Hill Companies, Feb. 1991.

[33] RATHA, N., AND BOLLE, R. *Automatic Fingerprint Recognition Systems*. Springer, 2003.

[34] RAZAVI, B. *Design of Analog CMOS Integrated Circuits*, 1 ed. McGraw-Hill, Inc., New York, NY, USA, 2001.

[35] ROSS, A., AND JAIN, A. Biometric sensor interoperability: A case study in fingerprints. In *in Fingerprints, Appeared in Proc. of International ECCV Workshop on Biometric Authentication* (2004), pp. 134–145.

[36] SCHNEIDER, J. K., RICHARDSON, C. E., KIEFER, F. W., AND GOVINDARAJU, V. On the correlation of image size to system accuracy in automatic fingerprint identification systems. In *Proceedings of the 4th international conference on Audio- and video-based biometric person authentication* (Berlin, Heidelberg, 2003), AVBPA'03, Springer-Verlag, pp. 895–902.

[37] SHAFIR, M. US-VISIT transition to 10-fingerprint collection. http://innovya.com/2009/09/19/us-visit-transition-to-10-fingerprint-collection, 2011. [online], [accessed 2011-02-06].

[38] TABASSI, E., WILSON, C. L., AND WATSON, C. I. Fingerprint image quality: NISTR 7151. Tech. rep., 2004.

[39] TAN, T., AND SUN, Z. CASIA–FingerprintV5. http://biometrics.idealtest.org, [online], [accessed 2011-02-05].

[40] TAYLOR, F., BICZ, W., BANASIAK, D., BRUCIAK, P., GUMULIŃSKI, S., GUMIENNY, Z., KOSZ, D., KRYSIAK, A., KUCZYNSKI, W., PLUTA, M., AND RABIEJ, G. Fingerprint structure imaging based on an ultrasound camera. *Instrumentation Science & Technology* (1999), 295 –303.

[41] WATSON, C., WILSON, A., MARSHALL, K., INDOVINA, M., AND SNELICK, R. Studies of one-to-one fingerprint matching with vendor SDK matchers NISTIR. Tech. rep., NIST, Gaithersburg MD, 2003.

[42] WILSON, C., WATSON, C., AND GARRIS, M. Matching performance for the US-VISIT IDENT system using flat fingerprints. Tech. rep., NIST, 2004.

# Appendix A

# Quality Score Distribution



(a) Orientation Certainity Level $Q_{OCL}$

(b) Ridge-Valley Structure Analysis $Q_{LC}$

(c) Orientation Flow Analysis $Q_{OF}$
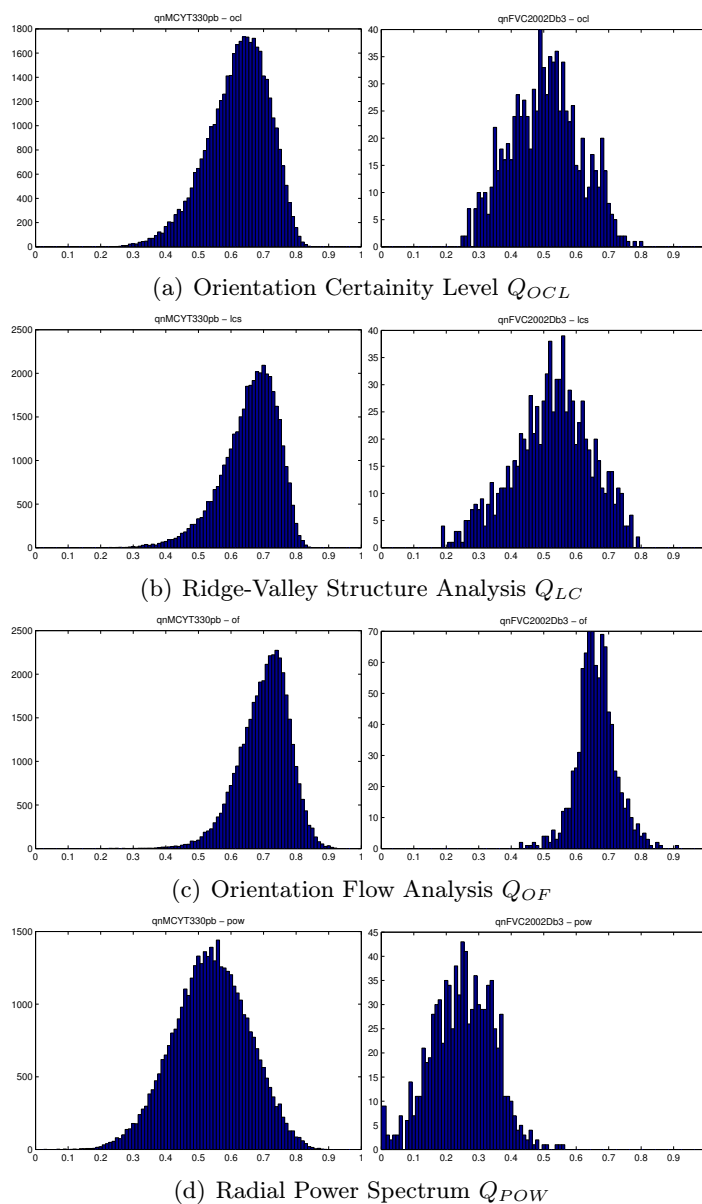
(d) Radial Power Spectrum $Q_{POW}$

Figure A.1: Comparison of quality score distributions for two different fingerprint databases. In both databases, samples were obtained using capacitive sensor. Higher quality scores can be observed in the the MCYT330dp database [31] (left coloumn) than in the FVC2002 Db3 database [26] (right coloumn) because the samples were collected under quality control.

# Appendix B

# CD Content

`/doc`          The electronic version of this report in PDF file.

`/fpquality`   Source files of methods for quality computation (MATLAB code).

`/fputility`   Source files for utility computation (MATLAB code).