Fingerprint Image Quality

Predicting Biometric Performance

Martin Aastrup Olsen

Thesis submitted to Gjøvik University College
for the degree of Doctor of Philosophy in Information Security

2015

Fingerprint Image Quality

Faculty of Computer Science and Media Technology
Gjøvik University College

Fingerprint Image Quality / Martin Astrup Olsen
Doctoral Dissertations at Gjøvik University College 5-2015
ISBN: 978-82-8340-023-6
ISSN: 1893-1227

Declaration of Authorship

I, Martin Aastrup Olsen, hereby declare that this thesis and the work presented in it is entirely my own. Where I have consulted the work of others, this is always clearly stated.

Signed:

(Martin Aastrup Olsen)

Date:

Summary

Biometric systems and fingerprint recognition systems in particular have become very widespread in the recent years, both in mobile devices and through increased usage in border controls and electronic national identification systems. One crucial aspect in any biometric system is that the quality of the data that enters system is of the highest possible quality to facilitate ease of interaction and a high level of biometric performance. In particular, the system error rates are sensitive to the quality of the enrolled sample due to subsequent interactions with the biometric system, results in a comparison being made against the enrolled sample. If the enrolled sample is of poor quality then the comparisons are more likely to result in a false non-match.

A performance evaluation taking the moisture content of skin into account and using modern optical fingerprint sensors revealed that the moisture level in the fingerprint has a bearing on the biometric performance. Biometric performance is sensitive to the moisture level of the skin and it was found that the effect is most pronounced on previous generation of sensors. Newer sensors are constructed in such a way that they are resistant to even high amounts of water on the sensing surface without compromising the fingerprint image quality. The dataset resulting from the investigation is made available.

The influence of the fingerprint position on the sensor platen was investigated and determined to be predictive of the biometric performance. Fingerprints with core points centered receive higher comparison scores on average than fingerprints with non-centered core points in genuine trials .

With respect to the predictive performance of quality algorithms, a performance evaluation of state of the art algorithms showed that the predictive capabilities varies across datasets and comparison subsystem. The implementations of the evaluated state of the art algorithms is made available.

Dactyloscopic examiners assessments of fingerprint quality was investigated and found to have a bearing on genuine comparison scores, and a method towards a system which assists examiners in determining evidential value was proposed.

Concerning quality assessment on devices with limited computational resources, a method based on a clustering using receptive field and random forests was proposed and found to result in predictive performance comparable to state of the art methods.

An algorithm for discriminating between fingerprint image and images of other biometric modalities based on topological invariants was proposed. Experiments on separating fingerprint images and iris and face images show that the classification performance is improved over a state of the art spectral based algorithm.

Acknowledgments

I would like to gratefully thank my advisor, Prof. Dr. Christoph Busch for his time, guidance and advice in bringing this dissertation to fruition and I owe thanks to Prof. Dr. Patrick Bours for being a supportive co-advisor. I would also like to thank the members of the committee, Prof. Dr. Rasmus Larsen, Prof. Dr. Javier Ortega-Garcia, Prof. Dr. Laura Georg and head of the committee Prof. Dr. Stephen Wolthusen for their valuable comments and suggestions.

I am grateful for the opportunity to work with the competent partners in the NFIQ 2.0 project who encouraged me to grow as a researcher and formed an integral part in the definition of my research. These were Christoph Busch, Johannes Merkle, Oliver Bausinger, Michael Schwaiger, Olaf Henniger, Alexander Nouak, Timo Ruhland, and Elham Tabassi whom I owe additional thanks for inviting and accepting me as guest researcher with the NIST ITL/IAD Image Group. With the IAD I am especially thankful for the inspiring discussions with Yooyoung Lee, Mike Garris, Patrick Grother, Jonathon Phillips, Brad Wing and Vladimir Dvornychenko.

I would like to express my appreciation to Prof. Dr. Raymond Veldhuis and his research group with University of Twente for hosting me as a visiting research student.

I owe thanks to the exceptional group of individuals in SC 37 Biometrics with the International Organization for Standardization who have provided ample opportunity to discuss all matters pertaining to biometrics and the intricate process of standardization.

During my time with Hochschule Darmstadt and Gjøvik University College I have had the pleasure to share office space, deadlines, discussions, laughter and more with Daniel, Mark, Claudia, Sebastian, Frank, Anika, Andreas, Christian, Lelai, Xuebing, Jessica, Nicolas, Ctirad, Bian, Soumik, Raghavendra, Guoqiang, Kiran, and Mohammad – thanks for the great times.

I have had the opportunity to supervise and collaborate with Vladimír, Thi Dieu, Andreas, Anton, Marek, Jinghua, Ivan and Martin in their final M.Sc. candidate projects, and from whom I learned a great deal.

I am indebted to my friends and family for their enduring support, understanding and encouragement throughout this undertaking.

Martin Aastrup Olsen

Gjøvik
November 26, 2015

Contents

List of Figures

List of Tables

List of Algorithms

Part I

Fingerprint Image Quality

Chapter 1

Introduction

1.1 Biometric Sample Quality

The first component of a biometric system is the data capture subsystem, which acquires biometric characteristics. All subsystems in the biometric system operate and depend on the successful acquisition of a biometric sample; the acquisition influences all parts of the biometric system and consequently, it is important to specify requirements for the biometric sample to ensure that a sufficient amount of information is available to reach a desired recognition performance. In the biometric terminology, these requirements are conveyed by the concept of biometric sample quality.

Biometric sample quality is a term, which covers any method that is predictive of the recognition performance for a given biometric sample, i.e. a quality assessment is a formulation of the expected recognition performance.

At the highest level, determining sample quality concerns itself with quantifying measurable structures in the object or the digital representation of the object at hand and relating those measurements to observed biometric performance. Biometric sample quality is constrained to only being defined for biometric samples, however, the constraint gives rise to the fundamental question as to how it can be recognized that the presented object is a representation of a biometric characteristic and further, at which point the level of information is sufficient to assert that the object is a biometric sample or a non-biometric object.

Given a biometric sample, either presumed or asserted, it is necessary to determine to which degree the sample is suitable for comparisons with other biometric samples, e.g. in order to avoid storing a sample which has a high likelihood of being rejected when compared with its mate.

A sample, which is suited for biometric comparison, might be used to form a biometric reference attributed to the biometric data subject, however, when a sample is determined to be unsuited for biometric comparison it is crucial to assess which, if any, of the factors contributing to the unsuitability of the sample are temporary or permanent. Temporary factors, such as non-conformant interaction with the biometric data capture subsystem, might be correctable by performing a biometric presentation when conformance has been ensured, whereas permanent factors are those which cannot be changed or corrected through additional subsequent captures, e.g. a severely deteriorated fingerprint. Ideally, when a temporary and correctable factor is deemed to be the cause of the sample being unsuitable, the biometric system should inform the subject about why the sample was unsuitable such that a correction may take place. For permanent factors the possibilities are limited to accepting the unsuited sample as is or to initiate a procedure exterior to the biometric system that manages instances where a sample cannot be acquired. However,

factors which are of temporary or permanent nature might result in a Failure to Enroll (FTE), and thus render the biometric system ineffective for the subject.

In all cases different factors influence biometric performance in different ways, and it is the requirements with respect to the recognition performance that dictate the acceptable level of expected loss in recognition performance and the procedures that need to be in place to handle factors which cannot be corrected within the biometric system.

Readers unfamiliar with biometric sample quality and fingerprints are encouraged to review chapter 2 to develop further context in the reading of the rest of this dissertation.

1.2 Research Objectives

The main objective of this dissertation is to identify factors which are predictive of expected biometric sample recognition performance in biometric mated comparison trials of fingerprint recognition systems. Towards this objective, a number of sub-goals arise and are made concrete in the form of research questions presented in section 1.3. The research is to determine a method which asserts whether an object is resembling a fingerprint; explicit or implicit fingerprint image covariates predictive of biometric performance shall be identified; given a set of image covariates predictive of biometric performance, it shall be ascertained how to combine the covariates to produce a robust classifier for quality assessment; it shall be established to which degree there exists a relationship between subjective and objective quality assessments.

1.3 Research Questions

From the research objectives outlined in section 1.2 a number of research questions forming the main areas of contribution of this dissertation have been formulated as listed below:

$\mathcal{RQ}_1$ **Can it be determined whether a presented object is a fingerprint biometric characteristic?**

$\mathcal{RQ}_2$ **Which measurable fingerprint structures are predictive of biometric performance for fingerprints?**

$\mathcal{RQ}_3$ **Can a computationally efficient quality algorithm be defined?**

$\mathcal{RQ}_4$ **Which environmental factors influence the recognition performance?**

$\mathcal{RQ}_5$ **Are automatic assessments of quality related to human expert assessments of quality?**

1.4 Publications

The following publications are part of this dissertation:

[1] M. Olsen, H. Xu, and C. Busch. "Gabor filters as candidate quality measure for NFIQ 2.0". In: *Biometrics (ICB), 2012 5th IAPR International Conference on*. 2012, pp. 158–163. DOI: 10.1109/ICB.2012.6199802

[2] A. Aabrandt, M. Olsen, and C. Busch. "A topology based approach to categorization of fingerprint images". In: *Biometrics Special Interest Group (BIOSIG), 2012 BIOSIG - Proceedings of the International Conference of the*. 2012, pp. 1–11

[3] M. A. Olsen, E. Tabassi, A. Makarov, and C. Busch. "Self-Organizing Maps for Fingerprint Image Quality Assessment". In: *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*. CVPRW '13. Washington, DC, USA: IEEE Computer Society, 2013, pp. 138–145. ISBN: 978-0-7695-4990-3. DOI: 10.1109/CVPRW.2013.28. URL: http://dx.doi.org/10.1109/CVPRW.2013.28

[4] I. Danov, M. A. Olsen, and C. Busch. *Interpretation of fingerprint image quality features extracted by self-organizing maps*. 2014. DOI: 10.1117/12.2050676. URL: http://dx.doi.org/10.1117/12.2050676

[5] J. Wang, M. A. Olsen, and C. Busch. *Finger image quality based on singular point localization*. 2014. DOI: 10.1117/12.2050145. URL: http://dx.doi.org/10.1117/12.2050145

[6] M. Dusio, M. A. Olsen, and C. Busch. "Fingerprint Sample Quality Assessment via Ridge Line Count Using Laplacian of Gaussian Edge Finding". In: *The 2nd International Workshop on Biometrics and Forensics (IWBF)*. 2014

[7] M. A. Olsen, M. Dusio, and C. Busch. "Fingerprint Skin Moisture Impact on Biometric Performance". In: *International Workshop on Biometrics and Forensics 2015*. 2015

[8] M. A. Olsen, V. Šmida, and C. Busch. "Finger image quality assessment features – definitions and evaluation". English. In: *IET Biometrics* (2015). ISSN: 2047-4938. URL: http://digital-library.theiet.org/content/journals/10.1049/iet-bmt.2014.0055

[9] M. A. Olsen, M. Böckeler, and C. Busch. "Predicting Dactyloscopic Examiner Fingerprint Image Quality Assessments". In: *BIOSIG* (2015)

Each publication and its associated research questions are listed in table 1.1 along with the chapter wherein the publication can be located in this work. A compact description of the main contributions of each publication with respect to the research questions from section 1.3 is located in chapter 3.

Other publications by the author which are not part of this dissertation are the following:

[10] M. A. Olsen, D. Hartung, C. Busch, and R. Larsen. "Contrast Enhancement and Metrics for Biometric Vein Pattern Recognition". In: *Advanced Intelligent Computing Theories and Applications*. Ed. by D.-S. Huang, M. McGinnity, L. Heutte, and X.-P. Zhang. Vol. 93. Communications in Computer and Information Science. Springer Berlin Heidelberg, 2010, pp. 425–434. ISBN: 978-3-642-14831-6. DOI: 10.1007/978-3-642-14831-6_56

[11] M. A. Olsen, D. Hartung, C. Busch, and R. Larsen. "Convolution Approach for Feature Detection in Topological Skeletons Obtained from Vascular Patterns". In: *IEEE Symposium Series in Computational Intelligence 2011 (SSCI 2011)*. Paris, France, Apr. 2011

Citation	Research Questions	Chapter
[1]	$\mathcal{RQ}_2$	10 (p. 133)
[2]	$\mathcal{RQ}_1$	5 (p. 59)
[3]	$\mathcal{RQ}_2$, $\mathcal{RQ}_3$	7 (p. 87)
[4]	$\mathcal{RQ}_2$, $\mathcal{RQ}_3$	8 (p. 103)
[5]	$\mathcal{RQ}_2$	6 (p. 69)
[6]	$\mathcal{RQ}_2$, $\mathcal{RQ}_3$	11 (p. 145)
[7]	$\mathcal{RQ}_4$	9 (p. 123)
[8]	$\mathcal{RQ}_2$	4 (p. 23)
[9]	$\mathcal{RQ}_5$	12 (p. 155)

Table 1.1: Publications and associated chapter and research question.

[12] M. A. Olsen and C. Busch. "Deficiencies in NIST Fingerprint Image Quality Algorithm". In: *12. Deutscher IT-Sicherheitskongress*. Vol. 1. 2011, pp. 251–262

[13] D. Hartung, M. A. Olsen, H. Xu, and C. Busch. "Spectral Minutiae for Vein Pattern Recognition". In: *Proceedings IEEE International Joint Conference on Biometrics (IJCB 2011), October 11-13*. 2011

[14] D. Hartung, M. A. Olsen, H. Xu, H. Thanh Nguyen, and C. Busch. "Comprehensive analysis of spectral minutiae for vein pattern recognition". In: *Biometrics, IET* 1.1 (2012), pp. 25–36. ISSN: 2047-4938. DOI: 10.1049/iet-bmt.2011.0013

[15] G. Li, B. Yang, M. Olsen, and C. Busch. "Quality Assessment for Fingerprints Collected by Smartphone Cameras". In: *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*. 2013, pp. 146–153. DOI: 10.1109/CVPRW.2013.29

[16] C. Gottschlich, A. Mikaelyan, M. Aastrup Olsen, B. Josef, and C. Busch. "Improving Fingerprint Alteration Detection". In: *The 13th IEEE International Symposium on Parallel and Distributed Processing with Applications (IEEE ISPA-15)* (2015)

1.5 Structure of the Dissertation

The dissertation is divided into 2 parts: Overview part I which provides an overview and introduction to the topic, the scope of the dissertation, and summary of main contributions; Publications part II contains a preprint of each refereed publication forming the research aspect of this dissertation.

Chapter 2

State of the Art

2.1 Introduction

This chapter provides an introductory state of the art overview to fingerprints, methods for acquiring and comparing fingerprints, and of aspects of fingerprint image quality. Terms and definitions are as stated in the ISO/IEC International Standard 2382:2012 [17].

2.2 Fingerprint Formation, Persistence and Uniqueness

The formation of fingerprints begins at an early embryonic age where the configuration of epidermal ridges and minutiae positions are determined by week 10 post fertilization and once fully formed by month 7, the ridge configurations do not change further [18]. Several hypotheses on the processes which dictate the formation of the ridge configurations constituent of the fingerprint have been developed; Kücken and Newell formulated a buckling model which takes into account the skin elasticity and rate of growth [19]. Garzón-Alvarado and Ramírez Martinez hypothesize [20] that the formation is based on the reaction-diffusion theory of morphogenesis by Turing [21].

Fingerprint recognition is based on two fundamental premises, namely that the fingerprint patterns are unique across sources, and that the pattern of an individual source is persistent.

The persistence of the fingerprint pattern was observed by Sir William J. Herschel in the late 1800s [22], and investigated in a larger scale study by the German Federal Office for Information Security (BSI) funded BioFinger project [23, 24] and later by Yoon and Jain [25], who found that the genuine comparison scores decrease as the time interval between the two captures increases and that the uncertainty of the temporal stability of recognition accuracy increases when either of the samples are of poor quality.

Uniqueness is a property that was attached to fingerprints for a long time without scientific basis. The use of evidence lacking in scientific basis was disputed in a court of law in the case of Daubert versus Merrell Dow Pharmaceuticals, where the Supreme Court of the United States ruled a set of criteria which must be followed for expert testimony to be considered valid [26]. This ruling was applied in a case from 1999, which challenged the use of forensic evidence based on fingerprints and wherein it was ruled that the fundamental premise asserting the uniqueness of fingerprints had not been objectively tested and that the potential error rates for fingerprint recognition were unknown [27]. This in turn led to an investigation where $50\,000$ fingerprints were compared and a probability of a false match occurring was determined to be in the range of 10^{-27} to 10^{-97}. The investigation has been criticized as not providing a reasonable error estimate when taking into account variable but reasonable size, position and quality [28].

An extensive theoretical quantification of the uniqueness of fingerprint patterns was conducted by Pankanti *et al*. who estimated that when using an Automated Fingerprint Identification System (AFIS) the probability of a false match between two fingerprints, each containing 46 minutiae and obtained from different sources, is on the order of 10^{-77} [29]; reducing the number of minutiae to 12 increases the probability of a false match to an order of 10^{-7}. Empirical observations show that the probabilities obtained by the theoretical model are several orders of magnitude lower than achieved on a collected dataset.

Zhu, Dass *et al*. construct model from a set of mixture models which take into account the distribution and clustering tendencies and dependencies in different image regions of the fingerprint as well as the orientation distributions, to provide a probability of a match between two fingerprints from different sources [30].

Identical twins have the closest genetic relationship and hence any potential genetic and environmental influence on the fingerprint formation is expected to be maximised in fingerprints obtained from identical twins. Comparisons performed between fingerprints obtained from identical twins show that there is a decrease in accuracy (i.e. false matches are more probable), when compared with comparisons between non-twins [31].

2.3 Fingerprint Sensing

Fingerprint sensing broadly covers any device or method which acquires a fingerprint. Sensing techniques can be divided into off-line sensing and live-scan.

Off-line sensing is widely known from law enforcement where fingerprints were acquired by smearing the subject's fingers with ink and rolling the fingers against a paper card which recorded the impression. However, the most used form of off-line sensing are those employed in acquisition of latent fingerprints in forensic investigations, e.g. scenes of crime or disasters. Forensic techniques used in the acquisition process of latent impressions from glass or other surfaces includes powder dusting and iodine fuming, however acquisition from wet surfaces or human skin require other chemicals and procedures to develop the fingerprint [32]. Acquisition of fingerprints post-mortem may be required to determine the identity of the deceased individual, and those cases another set of methods are available to forensic investigators, such as rehydration of the skin to restore the finger texture [33, p.4-10]. In either case the acquired biometric sample is transferred to a digital format such that latent examiners may perform the necessary markup of features and so comparisons using AFIS are possible. While forensic techniques for post-mortem and latent acquisitions are continuously being developed, the practice of smearing ink and recording the impression on papers cards is widely abandoned and replaced by live-scan methods. However, since the method was used for a number of years, the stored cards of acquisitions have in many cases been digitized such that they are available for search using AFIS.

Civil applications using fingerprints for recognition are predominantly, if not entirely, performed using the live-scan sensing method due to user acceptance, convenience, cost and reliability. An acquisition with a live-scan sensor is performed by the user presenting the finger to the device, e.g. by placing the finger on a platen, swiping the finger across the device or by moving the finger through a volume of air where the device performs a touch-less acquisition. ISO/IEC IS 19794-4:2011 [34] specifies identifiers for 20 capture device technologies, with the majority being variations of optical systems, but also capacitive, thermal, ultrasound, electro luminescent devices are represented in the extensive but

not exhaustive list. Apart from the listed, a few others are notable: Optical Coherence Tomography (OCT) is a medical imaging technique which achieves micrometer resolution of the internal structures of the finger tissue and which has successfully been applied to perform fingerprint recognition [35] and presentation attack detection [36, 37].

2.4 Fingerprint Recognition

Biometric recognition involves comparing two biometric samples (a probe and a reference) to obtain a comparison score which is used to decide if the probe and reference are from the same source (a match) or not (a non-match). The comparison is based on representations of the samples in the form of biometric templates. In the case of fingerprints, a template typically contains minutiae position and orientation information, while information about singular point locations, texture information or any other extractable feature may be part of the template.

Either of two fundamental errors may arise when performing a biometric comparison: a false match occurs when a match decision is made on a comparison between two non-mated samples; and conversely, a false rejection occurs when a non-match decision is made on a comparison between two mated samples. Recognition errors can in part be explained by the fact that biometric samples from the same source have inherent variabilities which are not accounted for in feature extraction systems. The variability means that in effect no two biometric samples are identical - variations in sensor noise, minute changes occurring to the tissue over time, and changes in the environment where the acquisition is performed all contribute to this. Recognition is feasible due to the selected features used for recognition possessing an inter-class sample variability much greater than the intra-class variability.

Current implementations of fingerprint recognition systems predominantly rely on minutiae information, e.g. National Institute of Standards and Technology (NIST) provides the NIST Biometric Image Software (NBIS) package which contains a biometric template extractor (MINDTCT) which uses minutiae information and produces templates and a comparison algorithm (BOZORTH3) which compares the produced templates.

Formats for storing fingerprint templates have been specified by International Organisation for Standardization/International Electrotechnical Commission (ISO/IEC) in International Standard 19794:4 [34]. The standard prescribes a standardized format for representing the biometric template and allows for additional proprietary data intended for use by the comparison subsystem. Another standardization effort is through the American National Standards Institute/National Institute of Standards and Technology - Information Technology Laboratory (ANSI/NIST-ITL) standard [38]. The standard provides a record type (Type 9) for storing minutiae information based on the feature set defined in InterNational Committee for Information Technology Standards (INCITS) 378 standard [39], as well as proprietary defined feature sets and an extended feature set.

2.5 Fingerprint Sample Quality

2.5.1 Overview

There is general consensus that sample quality broadly refers to the applicability of a biometric sample in recognition tasks: a sample with high quality is expected to fulfil the requirements of the target application, i.e. it is not likely to increase system error rates. With the development of NIST Fingerprint Image Quality (NFIQ) in 2004 an automated, objective method for establishing the quality of a fingerprint was provided.

A major challenge in achieving high predictive performance of the fingerprint, is the fact that quality is predominantly computed per sample whereas comparisons are always made for pairs of samples. In effect, the conclusion of the expected recognition performance is made based on incomplete information due to the potential for high intra-class variability.

Two high quality samples will give high comparison scores, but when one high quality and one low quality sample is compared it is likely to yield a low comparison score, hence it is hard to assign high quality and low quality appropriately when it is not known a priori which sample the comparison will be made with.

The difficulties related to quality being associated with only a single image are observed in face recognition, where two images in Face Recognition Vendor Test 2006 [40] (FRVT 2006) of the same person receive low similarity score but still receive high similarity scores with other images of the same person from the same location [41].

2.5.2 Distinguishing and Segmenting Fingerprints

Determining the properties which distinguish a fingerprint from another object is a fundamental task associated with general object recognition. This fundamental task is related to segmentation, which associates individual parts of the whole to one or more parts which may be present in the object.

Methods using a global threshold have proved to be ineffective in isolating the fingerprint area from the background mainly due to variations in lighting or noise across the image. Hence most effective methods concern themselves with determining the presence of an oriented striped pattern, i.e. a simplification of the ridge-valley structure which makes up the directional field, to locate the foreground.

Numerous methods are based on this fundamental property of the presence of a directional field as determined by the distinct pattern stemming from the ridge-valley formation [42]. Segmentation methods based on Gabor filters are further a consequence of the inherent directional property in the fingerprint and has been proposed as well [43, 44]. A learning based approach using Support Vector Machine [45] (SVM) was proposed by Bazen and Gerez [46] and extended by Yin *et al*. [47]. Several other learning based methods for segmentation of fingerprints have been proposed since then [48, 49]. A dictionary based method for quality and segmentation of latent fingerprints was proposed by Cao *et al*. [50].

One attempt at determining whether an image contains a fingerprint was proposed by Libert *et al*. in the NIST developed metric Spectral Image Validation and Verification (SIVV) which analyses the 1-dimensional power spectrum of the input image using Discrete Fourier Transform (DFT) [51]. Specifically, the method relies on the assumption that well formed fingerprints exhibit a distinct pair of local minimum and local maximum in

the spectrum which arise from level 1 ridge flow. The method was later modified to provide a measure of image quality for latent fingerprint impressions [52].

Yoon and Jain propose a fingerprintness metric using difference between the observed orientation field of the fingerprint and the orientation field of a fingerprint model [53].

2.5.3 Factors Influencing Fingerprint Sample Quality

Environmental factors play a significant role in the biometric performance for a given biometric system. Studies have shown that fingerprint samples obtained from elderly populations result in inferior quality when compared with a younger 18-25 year old population [54]. Similar observations have been made between healthcare populations and the general population [55]. A number of case studies have been produced from data collected for Unique Identification Authority of India (UIDAI) and the fingerprint image quality of the parts of the enrolled population have been analysed [56, 57].

The impact of fingerprint force during live scan acquisition was investigated by Kukula, Elliott *et al*. [58], and the effect of force during deposition on latent fingerprints was studied by Fieldhouse [59].

With respect to the fingerprint capture device, the fingerprint recognition performance in rugged and cold weather conditions for chip and optical sensors was investigated by Stewart *et al*. who found that biometric performance did not depend on temperature and humidity [60]. Interaction with the sensor device is another crucial aspect of successful acquisition: the influence on biometric performance with varying finger position on the device was studied by Simon-Zorita *et al*. who found that significant performance increases are achievable when the acquisition of reference and probe samples are controlled [61].

The characteristics of the skin, especially the moisture content have been found to have a strong influence on image quality and biometric performance [62]. Drahansky studied fingerprint skin diseases, quality and liveness and the effects on fingerprint recognition in depth in his dissertation [63].

Successful comparison of wet fingerprints has been addressed by Krishnasamy who also made a dataset of wet and wrinkled fingers available [64]. A classifier-based detection of subjective labelled wet, dry and normal blocks within a fingerprint image has been proposed [65].

A correlation between the fingerprint area and the and Equal Error Rate (EER) was established by Schneider *et al*. [66], and Cappelli *et al*. determined in depth the relationships between sensor area, geometric distortion and biometric performance to determine the operational quality of fingerprint capture devices [67].

Intra-class variability causing deterioration in biometric performance can arise from several other factors than the physical sensor. e.g. users unfamiliar with the acquisition process, or a device which does not facilitate easy and correct user interaction, can result in acquisitions which are highly variable due the positioning of the finger on the sensor platen. In some circumstances users might intentionally or unintentionally disrupt the acquisition process such that a low quality sample is acquired. In a border control scenario this disruption could for example be due to tiredness, distractions in the surrounding area, or due to the user being uncooperative and not conforming to the process with the intent to invoke a false non-match.

2.5.4 Explicit Objective Quality Assessment

Explicit quality assessment refer to quality assessment algorithms which quantify specific properties in the fingerprint image, e.g. the ridge sharpness or the number of minutiae. There exists a wide range of quality algorithms which are based on disparate techniques each with the objective to measure one or more properties locally or globally.

Some attempts to distinguish classes of quality algorithms have been made; Fronthaler *et al*. briefly discuss that quality assessment methods may be divided into either of full, reduced, or no-reference approaches, depending on the amount of available prior information [68]. Prior information refers to particular assumptions and criteria that the signal shall conform to, this could for example be that the dominant frequency in a Fourier spectrum of the fingerprint image lies in a specific band which conforms to the empirically observed ridge-valley frequencies in a given population. One may also partition quality algorithms according to the fashion in which they operate: Alonso-Fernandez *et al*. described a possible taxonomy for quality algorithms which divide the algorithms into those which are based on local features, such as local Gabor filter responses or pixel intensities; those that are based on global features, such as the standard deviation of the image gray scale values, or magnitude of a frequency band in the computed Fourier spectrum; and finally those which are based on classifiers, i.e. combine multiple local and global features using for example neural networks to produce a quality value [69]. These categories are useful in the describing the behaviour of a given quality algorithm, however, ultimately the defining criteria for an algorithm being a quality algorithm is that it is predictive of biometric performance [70].

Local quality algorithms are most common in the literature, in part due to the fact that the quality of the fingerprint may vary in different regions and that the majority of current comparison systems following International Organization for Standardization (ISO) standards rely on minutiae information. Many algorithms rely on orientation information, such as the orientation certainty level [71] which quantifies the strength of the local gradient and orientation flow [72] which determines the coherence of dominant gradients in neighbouring blocks. Another approach to determine the local structure is to apply Principal Component Analysis (PCA) to fingerprint blocks [73]. Algorithms can rely on the structure of the ridges, such as the ridge valley uniformity [71], or the local ridge clarity [72], or the magnitude of ridge frequency in a particular band [74].

The Gabor filter has been applied as a contextual filter which enhances particular aspects of ridge frequency and orientation. A Gabor based quality feature which operates in a local region and evaluated on subjective reference data has been proposed [43, 75].

Orientation tensors have been shown to be promising as a quality feature and has similar distribution of quality values as that of human quality assessments [68]. Using Delaunay triangulation of minutiae as a quality through estimating the usable fingerprint area appears as an alternative to estimate the quality when only the minutiae template is available [76].

A local quality algorithm can be made global by use of a summary statistic of the locally computed quality values. For example, an algorithm which determines local filter responses to produce a map of responses can be summarized using the standard deviation of the responses or produce a ratio based on a threshold.

Local quality values can be combined or aggregated by a variety of methods. A classifier based on linear regression using local Gabor filter responses and additional infor-

mation on foreground area and the number of minutiae in the fingerprint was evaluated in terms of reduction in equal error rate and found to be predictive of biometric performance [77]. NFIQ applied a non-linear classifier and utilized a neural network trained on 11 features based on orientation, minutiae to distinguish fingerprints into one of 5 quality categories [78, 79]. NFIQ is widely used in practical applications since 2004 and development of a replacement know as NIST Fingerprint Image Quality 2 (NFIQ 2) is under way [80–82].

Since NFIQ, a number of proposals for classifier based quality assessment algorithms have been developed. Local ridge orientation certainty, uniformity and frequency analysis were used as features in three types of classifier: ïve Bayes, Self-Organizing Map (SOM) and Radial Basis Function Neural Network (RBFNN) to classify subjectively labelled image blocks [83]. The tested classifier achieves good performance on identifying blocks subjectively labelled with a quality value. Another SVM based classifier was trained to output one of three quality levels by using image mean and variance, foreground area, coherence and power spectrum [84]. This system was trained on subjective quality labels. Similarly, a hierarchical K-means classifier using ridge distance, orientation certainty and ridge consistency as features was proposed [85]. A fusion of no reference quality metrics using a genetic algorithm was proposed [86]. Other methods for fusing multiple quality features include using only texture based features [87], or combining no reference features with texture and minutiae-based features [88].

Comparison of quality algorithms when applied to samples captured by different capture technologies and using minutiae-based and texture-based comparison systems [89–91]. Here, the effectiveness of rejecting samples based on quality assessments was found to vary depending on the capture technology used, and quality assessments exhibited a lower discrimination capability for high-quality images over low quality images. Due to the differing behaviour of sensors and comparison systems, an interoperability problem arises where performance is degraded. One proposal to address this problem incorporates sample quality information as a means to perform conditional processing and increase the biometric performance of the overall system [92]. Assessment of quality algorithms and biometric performance by controlled deterioration of fingerprint images underlines the difficulty in a single quality algorithm suited for differing sensor and comparison systems [93]. However, a proposal towards a general biometric sample quality algorithm using SVM with features from Scale-invariant Feature Transform (SIFT) and image quality was tested on face, fingerprints and hand veins and shows that there is a possibility that a general approach to multi modal quality algorithms is feasible [94] A comprehensive review of quality algorithms for fingerprint, iris and face was made by Bharadwaj *et al*. [95, 96] and a comparative assessment study of fingerprint sample quality in minutiae based systems by Alonso-Fernandez *et al*. [97] and by Jin *et al*. [98].

A framework for quality assessment algorithms and specifications of algorithms for several modalities are undergoing a standardisation process primarily through ISO/IEC standards series 29794 on biometric sample quality [70, 99, 100].

2.5.5 Applications of Sample Quality

Quality assessments have a variety of applications, e.g. adaptive image quality based capturing of fingerprints [101], and liveness and presentation attack detection [102, 103]. An extensive review of quality based presentation attack detection for iris, fingerprint and

face modalities is available [104]. The application of biometric sample quality has been studied in context of multi-modal authentication systems based on fingerprints and signatures [105]. Alonso-Fernandez *et al*. found that comparison score normalization based on quality improves EER in the biometric system [106]. With the increased usage of smart phones, one algorithm for assessment for images captured using the phone camera has been proposed [15].

Several attempts on performing comparison score fusion based on quality have been proposed, e.g. an adaptive approach by Fierrez *et al*. [107] and a unified framework for the multi modal case by Poh and Kittler [108]. An adaptive context switching framework using quality assessment to determine the appropriate fusion algorithm and comparison subsystem was proposed by Bharadwaj *et al*. [109].

In the forensic domain it is acknowledged that fingerprint examiners are subjective in their evaluation of the evidential value for a given fingermark. As inter- and intra-examiner evaluations of the fingermark varies, there is a need for objective methods to assist examiners in the evaluation. Pulsifer *et al*. proposed one algorithm to perform an objective quality assessment of fingermarks [110]. Similarly, the Latent Fingerprint Image Quality (LFIQ) algorithm was developed by Yoon *et al*. [111].

An assessment of methods for performance evaluation of quality features in biometric systems was performed by Grother and Tabassi, who proposed Error-Reject Curve (ERC) as a means to evaluate quality algorithm behaviour [112]. The outcome of evaluations based on fixed datasets are likely to have some influence from the dataset itself as differing datasets have different characteristics which can stem from a variety of sources such as sensors used, environmental influences and the population present in the dataset. An estimate of the level of difficulty of a dataset for comparison evaluations can be given by determining the area of overlap and relative distortion between samples [113].

2.6 Context and Application of Fingerprint Sample Quality

The rapid expansion in the use of fingerprint recognition and the increasingly large fingerprint databases created a need for AFIS which serves to aid dactyloscopic experts in finding potential matches in fingerprint databases. Modern fingerprint recognition systems have reached a very high level of biometric performance and have seen widespread use outside of forensic applications in areas as diverse as consumer electronics, time attendance systems, border control and national identification systems.

A central challenge in fingerprint recognition is to maintain a high level of biometric performance in varying environments and for all population groups in biometric systems which cover entire populations. The challenge exists in part due to currently deployed sensing technologies which are primarily focused on capturing the outer fingerprint, and thus susceptible to natural changes of the skin surface due to manual labor surface contaminations. These changes lead to a degradation of biometric performance due to a reduction of the number of features which can be extracted and due to an increased uncertainty of the features that are extracted to form the biometric template. Excessive scarring of the fingerprint results in a number of spurious minutiae along the edge of the scar and poses challenges for ridge flow orientation detection, and extensive manual labor may further result in partial or complete deterioation of the outer ridge structure making conventional feature extraction difficult or infeasible.

One example of a large scale biometric system using fingerprints and which encounters challenging conditions is the Aadhaar project in India. It provides an identification number and enables the identification of any resident in India based on fingerprint, iris and face modalities in addition to demographic information [114]. Residents across the country, from rural regions to populated city centers, are enrolled in Aadhaar as a means to prove their identity when accessing government- and private services which require unique identification of individuals. Enrolment at such a scale requires a distributed approach and the enrolment stations have been outsourced to private companies which enrol local residents. Performing quality assessments of the acquired samples allows for the replacement of a local enrolment supplier if it does not fulfil a specified requirement on the capture quality. Following the enrolment, the resident may determine which of their fingers are best suited for the identification process based in part on measured quality and recognition performance.

European Dactyloscopy (EURODAC) is a European Union-wide electronic system for the identification of asylum seekers and is the first common AFIS within the European Union (EU) [115]. EURODAC was a result of the Dublin Regulation [116] which came into effect in 2003 and covers all EU member states and Iceland, Norway, and Switzerland. A similar large scale system, United States Visitor and Immigrant Status Indicator Technology (US-VISIT), is in place in the United States of America. The Schengen Information System (SIS) is a large database supporting member state police and enabling judicial cooperation and management of e.g. external border control, missing persons or vehicles. The extension of SIS, Second generation Schengen Information System (SIS II) [117] entered operation in 2013 and includes biometric data in the form of fingerprints [118]. In order to ensure a well functioning information system, national police authorities must adhere to a certain quality level as specified by EC Regulation 1987/2006 Article 22(a) when submitting biometric samples [119].

The EURODAC system is in place as a central database for storing fingerprints of visa applicants with the stated purpose of preventing a rejected individual from indefinitely seeking asylum in other countries until accepted. The individual is rejected if the biometric sample provided by the individual matches with a biometric template already registered in the database. Consequently, an individual could supply a low quality sample on purpose in order to minimize their chance of detection and thus apply for a visa multiple times. Through the successful detection of low quality samples, it can be determined whether the supplied sample is of a sufficient standard for recognition purposes and prevent an individual from subverting the system through submission of low quality samples. This scenario is substantiated by the findings of Wilson *et al*. [120] and Tabassi *et al*. [78] who established that there is a strong correlation between fingerprint image quality and biometric performance.

Chapter 3

Contributions and Conclusions

3.1 Contributions

The research questions formulated in section 1.3 form the basis for the research documented in this dissertation. Each publication contained within this dissertation is made with these questions in mind, and each publication contributes towards answering one or more of the questions. The list below contains each of the research questions with a concise description of the contributions made in this dissertation.

$\mathcal{RQ}_1$ **Can it be determined whether a presented object is a fingerprint biometric characteristic?**

- A method for determining whether a presented object is a fingerprint was proposed and the evaluation shows that the method outperforms the NIST provided SIVV algorithm in distinguishing fingerprints from iris images. Existing approaches to determining whether an object is a fingerprint predominantly rely on analysis of the ridge frequency or the orientation properties of the ridge valley pattern, whereas the proposed method is performing the classification based on persistent properties in the form of Betti numbers found in the algebraic topology. The proposed method achieves an equal error rate between 0.2% and 2.7% on mixed iris and fingerprint datasets, where SIVV achieves between 6.2% and 12.9%, and further the performance of the proposed approach appears invariant towards the choice of database used as training set. This contribution has been presented in [2](see chapter 5).

$\mathcal{RQ}_2$ **Which measurable fingerprint structures are predictive of biometric performance for fingerprints?**

- A Gabor filter based method for quality assessment was proposed and objectively evaluated using biometric performance as the evaluation criteria. Existing Gabor filter based approaches have been evaluated against subjective quality assessments, where as the proposed method was evaluated using biometric performance as the criteria. The proposed approach differs from existing Gabor filter based methods by operating on a per-pixel basis. This contribution has been presented in [1](see chapter 10).
- A quality assessment algorithm based on ridge line counts was proposed and evaluated using biometric performance as the criteria using ERC. This contribution has been presented in [6](see chapter 11).
- An extensive evaluation of existing quality features from the literature was performed using ERC as the main evaluation criteria. The evaluated methods have

been made available for the research community as a Matlab based quality computation framework [121]. This contribution has been presented in [8](see chapter 4).

- A quality algorithm based on the location of the singular points in the fingerprint was proposed and evaluated. The location of the singular points are predictive of the biometric performance. This contribution is presented in [5](see chapter 6).
- A quality algorithm based on unsupervised clustering and supervised learning has been proposed. The approach is one step towards automatic learned computationally efficient quality features suitable for embedded devices. This contribution is presented in [3](see chapter 7) and [4](see chapter 8).

$\mathcal{RQ}_3$ **Can a computationally efficient quality algorithm be defined?**

- A novel and computationally efficient method for quality assessment using a Lookup Table (LUT) in the form of a SOM and an ensemble of decision trees via Random Forests was proposed and evaluated. The proposed method sets itself apart from other quality assessment algorithms by learning which features are distinctive in fingerprints and then learning an interpretation of those features with respect to biometric performance. The evaluation shows the predictive performance to be similar to that of orientation certainty level as specified in ISO/IEC Technical Report 29794-4 [99]. This contribution was presented in [3](see chapter 7).
- A continuation of the research on the SOM based approach was conducted and evaluated. The comparative study involved experiments highlighting the influence of the chosen size of the SOM. The study provided results of further experiments on the choice of method for learning the relation between constituent parts of the fingerprint and biometric performance which validate the choice of Random Forest. This contribution was presented in [4](see chapter 8).
- A fast method for quality assessment based on ridge-line count determined by zero-crossings of the Laplacian of Gaussian (LOG) filtered fingerprint image was proposed and evaluated in terms of ERC. This contribution was presented in [6](see chapter 11).

$\mathcal{RQ}_4$ **Which environmental factors influence the recognition performance?**

- An extensive dataset was collected under four controlled conditions with respect to fingerprint skin moisture levels using five different commercially available optical sensors and evaluated with respect to biometric performance for each sensor using three commercially available comparison systems. The dataset is made available to researchers via da/sec website [122].
- The main findings show that the fingerprint moisture level is correlated with biometric performance and that current generation (2011 - 2015) sensors can achieve high biometric performance levels for a wide range of moisture levels including soaking wet. The findings confirm the findings by Blomeke *et al*. [62] and extend them with further analysis on specific sensor characteristics. This contribution was presented in [7](see chapter 9).

$\mathcal{RQ}_5$ **Are automatic assessments of quality related to human expert assessments of quality?**

- The subjective fingerprint quality of a set of fingerprint images was determined by a group of dactyloscopic examiners and found to be indicative of the genuine comparison score.
- Using quality features specified in the literature it was shown that expert assessments are to some extent predictable and thus there is a potential for automated methods to assist human experts in determining which fingerprints are suited for processing in the Analysis, Comparison, Evaluation and Verification (ACE-V) process. These contributions were presented in [9](see chapter 12).

3.2 Conclusions

Automated fingerprint quality assessment continues to be an important part of present and future biometric systems due to its proven predictive capabilities as illustrated herein.

A set of algorithms which measure fingerprint structures inferred from local ridge flows were determined to be predictive of biometric performance. Further, a subset of these were found to be computational efficient. Computationally efficient methods for determining fingerprint quality is particularly useful in capture devices which must decide when to trigger a capture.

Recent developments in hardware construction and fingerprint sensors have reduced the influence of environmental factors such as wet dry fingerprints, however the detection and communication to the user on unintended interactions with the sensor continues to be an open and important problem to be solved. The issue is partly a usability issue where some interaction can be directed through proper ergonomics in the device, however in instances where it is detected that the user interacts in a unintended way, such as not centering the finger on the sensor, it is crucial to provide actionable feedback which can aid the user in interacting with the system in the intended way. This is particularly important as biometric systems in high throughput areas such as in airports are attended only by minimal staff.

Similarly, forensics have seen increased automation with the large presence of AFIS and tools to assist dactyloscopic examiners in identification tasks. The increasing automation is still dependent on sufficient quality levels of the data, and an understanding of the influence that particular image defects may have on the confidence in identification. One further step in this direction is a quality assessment system which assists dactyloscopic examiners in determining evidential value of the respective fingermarks. Understanding of the forensic process and the relation between subjective and objective quality assessments is crucial in producing such automation.

While evaluation of quality features has for the past decade been done with respect to biometric performance, there remains a tendency towards defining quality features in the light of our own subjective understanding and perception of what fingerprint quality means. This is in part attributable to how the comparison and feature extraction subsystem work. There is a largely unexplored area where comparison system sensitivities are discovered by learned approaches such as the SOM and Random Forest based approach proposed here.

Towards new definitions of quality, there is a need for a further understanding and formulation of the characteristics that define an object as being a fingerprint. While one approach has been proposed herein, this fundamental task remains open. Inherent topo-

logical differences were found between fingerprints and non-fingerprint images. These differences contribute to further understanding of the fingerprint as an object and has applications in enabling proper detection of incorrect placement on sensor device, and automatic verification of the presence of fingerprints encapsulated in transferred data formats.

As new sensors enter the market place which enable subsurface and volumetric sensing of the fingerprint, new quality assessment methods must be researched to accommodate the increased dimensionality of the input data as well as to accommodate the new feature extraction and comparison algorithms that naturally follow such development. This poses new problems with respect to computational efficiency which remain of high importance in both civil applications such as integrated devices on mobile phones, and devices at border controls in airports where throughput is of utmost importance.

Given the closed nature of commercial systems the sensitivity towards different image properties change among different vendors, especially when considering the usage of non-standardized biometric templates. These system differences are particularly challenging for the interpretation of the values provided by quality assessment algorithms. A learned approach, which maps a set of distinct quality values associated with a particular fingerprint image with its observed performance across multiple systems can to some extent alleviate the system specific sensitivities. A learned approach further opens up for an adaptive system which takes these sensitivities into account when determining quality values.

Part II

Peer-reviewed Publications

Chapter 4

Finger Image Quality Assessment Features – Definitions and Evaluation

Information about the paper

The predictive capability of quality algorithms vary depending on the dataset and comparison algorithm used for the evaluation. Here we define algorithms as specified in the literature and evaluate them in terms of Error-Reject Curve (ERC) and inter algorithm correlations while providing reference implementations. The paper is a contribution towards research question $\mathcal{RQ}_2$. This original paper was published in: [8] M. A. Olsen, V. Šmida, and C. Busch. "Finger image quality assessment features – definitions and evaluation". English. In: *IET Biometrics* (2015). ISSN: 2047-4938. URL: `http://digital-library.theiet.org/content/journals/10.1049/iet-bmt.2014.0055`.

Abstract

Finger image quality assessment is a crucial part of any system where a high biometric performance and user satisfaction is desired. Several algorithms measuring selected aspects of finger image quality have been proposed in the literature, yet only few of them have found their way into quality assessment algorithms used in practice. We provide comprehensive algorithm descriptions and make available implementations of adaptations of 10 quality assessment algorithms from the literature which operate at the local or global image level. We evaluate the performance on four datasets in terms of the capability in determining samples causing false non-matches and by their Spearman correlation with sample utility. Our evaluation shows that both the capability in rejecting samples causing false non-matches and the correlation between features varies depending on the dataset.

4.1 Introduction

In large scale Automated Fingerprint Identification System (AFIS), there will be a fraction of individuals who are unable to interact with the biometric sensor in the intended way and there will be a certain fraction of individuals who will try to avoid detection. In both cases, quality control in the signal processing subsystem of the biometric system must give an actionable response to the interacting individual or a supervising individual such that errors are addressed immediately.

At the border controls of Japan and the United States of America, an individual may be rejected entry if the biometric probe sample captured from the individual matches with a biometric reference already registered in a database, or a watch-list. Consequently, for negative biometric claims a situation can arise where an individual supplies a low quality probe sample on purpose, with the aim of minimizing the risk of detection. Thus, without a method of determining whether the quality of the captured probe sample is reaching a sufficient level for recognition purposes, an individual can subvert the system. The scenario is substantiated by the findings in [78, 107, 120], where it is established that there is a strong correlation between fingerprint image quality and biometric performance. Determining the quality of a fingerprint finds use in other scenarios such as immigration where an individual can apply for a visa at the embassy or consulate for a given country. In order to verify that the identity of the individual at the border control is indeed the same as the one who received the visa at the country's consulate, a fingerprint capture is performed at the time of the application. Thus the subject is enroled in the biometric system and can be identified at a subsequent border crossing.

In Europe such a system for the exchange of visa data is being implemented for the Schengen area and is known as the Visa Information System (VIS). Its stated purpose is to facilitate the visa application procedure, the fight against fraud and the prevention of visa shopping. Thus the new system shall contribute to the prevention of threats to European countries' internal security. The Council of the European Union (EU) decided to establish the VIS on 8 June 2004 [123] and the European Commission (EC) determined on 30 November 2009 that the VIS should be implemented progressively starting with visa applications first from North Africa, second the Near East and third the Gulf Region [124].

The VIS has a centralized structure with a central information system, the Central Visa Information System (CS-VIS) and the National Interfaces Visa Information System (NI-VIS). Alphanumeric data as well as biometric data, in particular fingerprints and photographs, is stored in the CS-VIS which can be queried through the NI-VIS by authorized staff of the visa authorities in member states. The standard use case for a query is a Schengen border crossing point, where probe fingerprint images of the traveler are compared against the reference fingerprint images of the visa applicant. The biometric data entering the VIS has to pass certain quality requirements on the local and central levels. The EC has provided member states with a dedicated software kit, known as User Software Kit 4 - Quality Check (USK4) [125], which is used to validate the quality of all fingerprints prior to being inserted in the central system. In the period October 2011 until June 2012 results from consular posts show that around 11 percent of all visa applications containing fingerprint image failed to meet the local quality assurance [126].

A different approach to biometric border control exists in the United States and is managed under Department of Homeland Security (DHS). Office of Biometric Identity Management (OBIM) which replaced United States Visitor and Immigrant Status Indicator Technology (US-VISIT) in March 2013 is an immigration and border control system used for registering all non-residents entering or exiting the United States of America. Fingerprints are an integral part of the system supporting the goal of establishing an identity to a subject. Having an identity and subject linked allows for checks against watch lists, verification of existing identity documentation such as passports, and discovery of attempted fraudulent identification. In addition to the fingerprint, a face mug shot is also recorded for manual validation purposes.

The cost of a false reject is high in VIS and US-VISIT for the individual, who is applying for entry or a visa, as he is potentially faced with wasted travel expenses and anguish over being unrightfully rejected, or the border control will have to employ special second line procedures to verify the individuals identity claim through other means if at all possible. On the other side, the authorities may fail to identify an individual who is placed in a watch list or is a known visa shopper with potential negative economic or human consequences.

In India, the Unique Identification Authority of India (UIDAI) is managing the Unique Identification number (UID). 102.5 million UID had been issued by the end of the year 2011 [127] and UIDAI expect 600 million enrolled by 2014 [114]. The goal of the UIDAI is to issue a UID for all Indian residents and thus enable financial inclusion for particularly the rural areas. It is estimated that 40 % of Indian citizens in rural areas do not have a bank account and are thus excluded from large parts of society [128]. The UIDAI has faced the challenge of including individuals who work extensively with manual labor that usually leads to worn down and eventually damaged fingerprint patterns of particularly low quality.

For the UID the cost for the individual is high as the success of a bank transaction or any interaction with government officials is dependent on successfully proving an identity claim. Because of the risks exposed by falsely rejecting or accepting a biometric claim, quality control measures must be implemented in any biometric system.

The rest of the paper is organized as follows: section 4.2 discusses biometric sample quality, section 4.3 contains a detailed overview of finger image quality features with algorithm descriptions, and an evaluation in terms of ERC, correlation and Area Under Curve (AUC) is found in section 4.4. Concluding remarks are found in section 4.5.

4.2 Biometric Sample Quality

4.2.1 Components of Sample Quality

The first component of a biometric system is the capture subsystem, which acquires biometric characteristics. All subsystems operate and depend on the successful acquisition of a biometric sample; hence the acquisition influences all parts of the biometric system. Consequently, it is important to specify requirements for the biometric sample in order to ensure that a sufficient amount of information is available. In the biometric terminology, these requirements are conveyed by the concept of biometric sample quality.

There are different aspects of quality, and one effort to specify those is represented through the International Standard ISO/IEC 29794-1 where three aspects of quality are identified [70]:

- *Character* of a sample is the quality attributable to the inherent features of the source from which the biometric sample is derived. For example a 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. A sensor with low spatial resolution typically results in a low fidelity sample.
- *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 both the character as well as the fidelity of the sample.

It is the utility that is of most interest as it is the objective measure for a biometric sample and directly relates the influence of character and fidelity to expected biometric performance. As an objective measure it has the advantage that the utility derived is fully defined by a set of reproducible algorithms and thus it is possible to trace how a particular sample were assigned a given utility.

For the scope of the definition sample quality assessment features and their evaluation it is essential to correlate the predicted with the observed utility. This relationship is depicted in fig. 4.1.

Current efforts on standardizing fingerprint image quality algorithms are carried out by International Organisation for Standardization/International Electrotechnical Commission (ISO/IEC), subcommittee 37, working group 3. The efforts will result in an update to 29794-4 Biometric sample quality - Part 4: Finger image data

4.2.2 Influence of capture device technology

There exists a diverse set of capture device technologies each with different properties some of which may negatively or positively affect the recognition error. International Standard ISO/IEC 19794-4:2011 [34] lists a wide range of different sensor technologies, and with each technology exists some limitation with regard to sensor surface area and resolution.

The influence of the capture surface area on recognition accuracy has been addressed in the literature where it has been found that the Equal Error Rate (EER) increases as the

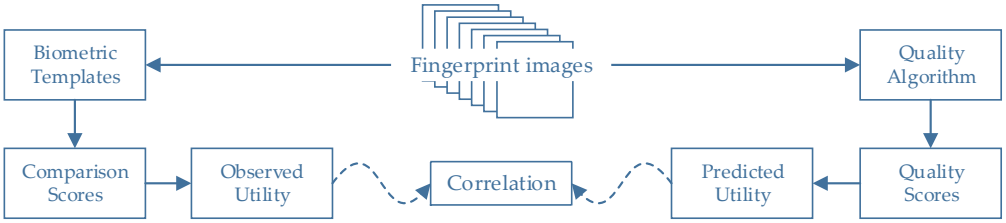


Figure 4.1: Quality scores define predicted utility that should correlate with observed utility derived from comparison scores.

sensor size decreases [23, 66]. The more general problem of recognition errors due to different sensor technology used for probe and reference samples is discussed by Ross and Jain [129] who found that there is a strong need for sensor independent algorithms, in particular with regard to matching algorithms.

4.2.3 Influence of user behaviour and skin condition

User interaction with the acquisition device is another source of recognition error, e.g., placing only the fingertip or side of the finger on the sensor surface is likely to provide insufficient information for the comparator, while pressing the finger too hard on the sensing surface causes elastic deformations, which may exceed compensation capabilities of the feature extraction or matching subsystems. The impact of force on measured image quality and minutia count is addressed by Kukula *et al*. [58].

Degradations in measured quality are expected for certain fingerprint skin conditions [62]. Skin conditions are likely to change due environmental factors such as temperature and moisture levels, as well as based on users profession (e.g. manual labor or office work). Moreover it can be observed that specifically in Western cultures there is an increasing number of individuals that suffer from dermatological diseases, that have a direct temporal or persistent impact on the fingerprint pattern. Examples for such diseases are atopic eczema, hyperceratotic eczema and thromboangitis. The skin condition for persons aged 62 and above is generally less moist than that of persons aged 18 to 25 and the former group generally receives lower quality scores than the latter [54].

4.2.4 Application areas of quality

Quality assessment can provide real-time actionable feedback at the capture level providing useful information that can lead to recommendations for the re-capture process and thus eventually to a higher quality of captured finger images and consequently increased biometric performance over systems not using quality control.

Further quality survey statistics are helpful in system monitoring to assess the performance of individual capture stations or in the evaluation of capture station operators. Distributed enrolment systems that operate e.g. with public-private partnership can be supported by incorporating the quality statistics as one of the criteria, when computing the enrolment fee for the station operator. Measurements over time can also point to systematic or system wide degradation of capture devices as they age and device components such as Light Emitting Diode (LED) deteriorate.

Modern biometric system architectures like the UIDAI operate with multiple biometric feature extractors and comparison subsystems and apply load balancing between these background components. With conditional processing of biometric samples of low quality may be sent to a feature extractor that is particularly robust to low quality but is much slower than the normal feature extractor.

Full-reference and no-reference image quality assessment algorithms have proven useful in the context of fake biometric detection, i.e. the detection of artefacts used in active subversion of a biometric system. An extensive review of the application of this type of image quality measures to perform such detection is addressed in depth by Galbally *et al*. [104].

4.2.5 Measuring biometric performance

4.2.5.1 Overview

Aspects of biometric quality must be expressed in an objective manner to ensure that performance can be measured and compared between different systems.

Tabassi *et al*. [78] proposed with NIST Fingerprint Image Quality (NFIQ) an approach for objective performance assessment based on a measure of the distance between the genuine and imposter comparison score distributions for a given sample; well separated distributions imply that the likelihood of false accept or false reject is low and that it increases with greater overlap between the distributions. This approach is generalized in ISO/IEC 29794-1:2009 [70] which requires that the quality score output by a biometric quality assessment algorithm conveys the predicted utility of the biometric sample. The method suggested for determining the utility of a sample is similar to the one used in NFIQ, i.e. the utility is estimated by the distance between the observed distributions of the sample's comparison scores.

4.2.5.2 Utility

While no specific method is mandated to derive the utility of a sample Annex A of ISO/IEC 29794-1 [99] suggests that the utility value of a sample may be computed as follows.

Given a set of comparison scores computed using a single comparator the utility of a sample j from subject i is

$$utility_i^u = \frac{m_{i,u}^{mated} - m_{i,u}^{non-mated}}{\sigma_{i,u}^{mated} + \sigma_{i,u}^{non-mated}} \tag{4.1}$$

Where $m_{(i,u)}^{mated}$ and $m_{(i,u)}^{non-mated}$ are respectively the mean of the genuine and imposter comparison score distribution of sample j from subject i and $\sigma_{(i,u)}^{mated}$ and $\sigma_{(i,u)}^{non-mated}$ is the standard deviation of respectively the genuine and imposter comparison score distribution. Such a utility value can be computed for each comparator using the comparison score provided by that comparator. Methods for fusing utility values exists and a more detailed description of the entire utility computation process can be found in the annex of ISO/IEC 29794-1:2009 [70].

The Spearman's rank correlation coefficient is used in this work as a non-parametric measure of statistical dependence between the quality score and utility.

4.2.5.3 Error reject curve

Grother and Tabassi [112] introduced the ERC as a method of assessing the predictive performance of a quality algorithm in terms of False Non-Match Rate (FNMR). We adopt the equations for the one-dimensional case, i.e. where a rejection is driven by the minimum quality q_i of the qualities in a pair of samples. In particular, the combination function H is chosen as the min function

$$q_i = H\left(q_i^{(1)}, q_i^{(2)}\right) = min\left(q_i^{(1)}, q_i^{(2)}\right) \tag{4.2}$$

Using the combination function in eq. (4.2) we form the set $R(u)$ containing the pairwise minima less than u

$$R(u) = \left\{j : H\left(q_j^{(1)}, q_j^{(2)}\right) < u\right\} \tag{4.3}$$

We then use $R(u)$ to exclude comparison scores starting with the lowest of the pairwise minimums up to some fixed threshold t which corresponds to a FNMR of interest, f. The threshold is obtained using the empirical cumulative distribution function of the comparison scores.

$$t = M^{-1}(1 - f) \tag{4.4}$$

$$\text{FNMR}(t, u) = \frac{|\{s_{jj} : s_{jj} \leq t, j \notin R(u)\}|}{|\{s_{jj} : s_{jj} \leq \infty\}|} \tag{4.5}$$

The method models the operational case in which samples are rejected due to low quality. The ERC is determined by progressively excluding a fraction of samples and recalculating the FNMR as the proportion of non-excluded scores which are below the threshold. For good algorithms which output quality scores that are monotonically related to the comparison score the desirable result is that the FNMR decreases quickly with the fraction of samples rejected.

To quantify the decrease in FNMR we propose two metrics which consider the AUC of the ERC with respect to the theoretical best case where the decrease in FNMR equals the fraction of samples rejected due to quality. The first metric, η_{auc}^{erc} eq. (4.6) is expressed as the AUC of the ERC subtracted the area under theoretical best.

$$\eta_{auc}^{erc} = \int_0^1 \textit{ERC} - \textit{area under theoretical best} \tag{4.6}$$

The second metric η_{pauc20}^{erc} eq. (4.7) is similar to η_{auc}^{erc} with the modification that only the first 20% the ERC is considered.

$$\eta_{pauc20}^{erc} = \int_0^{0.2} \textit{ERC} - \textit{area under theoretical best} \tag{4.7}$$

4.2.6 Quality in large scale biometric systems

When large biometric systems are analyzed certain properties will become apparent due to the scale of the operative system. Due to wide availability of solutions and systems providing a range of possible performance, usability and conformance constellations German Federal Office for Information Security (BSI) have provided technical guidelines for the usage of biometrics in the public sector [130].

The influence of fingerprint quality on biometric performance has been studied in detail by Wilson *et al*. in their 2004 study on recognition accuracy for US-VISIT [120]. The primary dataset in the study is based on $274\,000$ right index finger pairs sourced from the Mexican visa program. They show that a proprietary quality measure from Cogent Systems (now 3M [131]) with eight quality bins is a good rank statistic with all eight levels ordered such that the error rate of fingerprints with level 1 is always lower than any other quality level, and the error rate of quality level 2 is always lower than the next six quality levels and so on. The same property is demonstrated for the NFIQ, which categorizes finger images into five quality bins. Due to the open source availability of NFIQ this software became widely adopted since its release in 2004. More details on NFIQ will be provided in section section 4.3.2.

On the two comparators tested they show that the performance degrades consistently with lower quality levels for both the NFIQ and Cogent quality algorithms. An interesting aspect shown in the study is that at quality levels 1 to 4 the true accept rate is at least 98.2% at a false accept rate of 1.0% for the comparator used in US-VISIT Automated Biometric Identification System (IDENT). For quality level 5 to 8 the true accept rate further degrades from 95.2% to 53.6%. The implication is that half the quality scale is used for finger images that are highly likely to be correctly matched while much less fidelity is given to the remaining more problematic finger images.

Finally, it is concluded that finger image quality is the most critical single factor impacting the performance of fingerprint recognition systems. Wilson *et al*. recommend that image quality should be the first place to improve if performance of an otherwise reasonably engineered system is poor. One step towards better image quality in biometric systems is to detect when an image is of insufficient quality such that a new capture can be obtained (i.e. actionable quality).

The first study on finger image quality in India was performed by Vatsa *et al*. in a UIDAI case study analyzing fingerprints of Indian population using image quality [56]. Using NFIQ, three datasets are analyzed with the objective of determining quality aspects of fingerprints obtained from urban and rural population. In their analysis it is shown that for the ten-print dataset collected among $20\,000$ individuals in the rural population, 83 % of the $200\,000$ finger images result in a NFIQ score of 1 or 2, indicating good or very good quality. This is unexpected and the authors do show cases where NFIQ has incorrectly assessed the quality of two very poor quality images as being quality score 1 and 2, respectively. This indicates that NFIQ does misclassify finger image quality classes under some circumstances. Using a much smaller dataset of 1620 images in a prepared dataset with finger images from 27 urban individuals and 81 rural individuals an identification accuracy of respectively 90% and 64% was achieved.

It is observed that the finger image quality values reported by NFIQ are highly correlated with comparison scores, which supports the findings by Wilson *et al*. on quality and biometric performance being closely related.

4.3 Review of finger image quality assessment features

4.3.1 Overview

In finger image quality analysis it is important to consider that the finger image can be viewed at a high level (global finger image quality) and a local level (local finger image quality). Analyzing the finger image at the global level can provide a fast assessment but with the tradeoff that local context is neglected. The quality score derived from a global quality feature does reflect the case where e.g. a part of the fingerprint is blurry or missing, but it does not provide information about the spatial location of the defect. This is in contrast to quality features operating on the local level where spatial location is preserved yielding a more nuanced assessment, and potentially providing important information to the minutiae extractor in the form of a quality vector representing a quality map of the finger image. The common approach to compute local level information is to subdivide the image into blocks as illustrated in fig. 4.2. This figure introduces the block and pixel indexing used in the quality feature definitions in sections 4.3.4 and 4.3.5. A taxonomy of finger image quality algorithms has been proposed by Alonso-Fernandez *et al*. [69] where three classes of quality algorithms are described in the context of existing algorithms, namely those which are *a*) based on local features; *b*) based on global features; and *c*) those based on classifiers.

Several fingerprint image quality assessment algorithm have been proposed in the literature, in particular, the classifier based NFIQ algorithm is used extensively in industry and forms the quality control aspect in operational biometric systems. An extensive review of finger image quality algorithms has been performed by Alonso-Fernandez *et al*. [97] and more recently by Bharadwaj *et al*. [95].

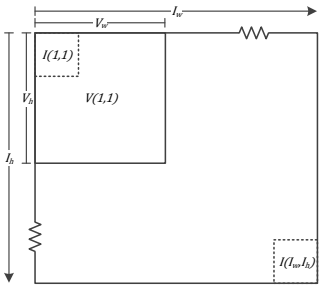


Figure 4.2: Illustration of block and pixel indexing within an image $\mathbf{I}$ with dimensions $\mathbf{I}_w$, $\mathbf{I}_h$. Shown is the pixel $\mathbf{I}(1,1)$, the block $\mathbf{V}$(1,1), with dimensions $\mathbf{V}_w$, $\mathbf{V}_h$

The source code for the quality features specified in sections 4.3.4 and 4.3.5 are made available online [121].

4.3.2 The NFIQ algorithm

The National Institute of Standards and Technology (NIST) developed in 2004 the NIST Fingerprint Image Quality (NFIQ) algorithm and with it a new definition of quality of fingerprint impressions and algorithms for measuring quality was proposed. The work at

the time was the first open algorithm for an assessment of finger image quality, which is a predictor of comparator performance [78].

The NFIQ image features are derived from an interpretation of a fingerprint quality map and minutia counts. Localized quality of the fingerprint is measured by computing four quality maps; a direction map, a low contrast map, a low ridge flow map and a high curvature map. For the local analysis, the image is divided into a grid of blocks, where all pixels within a block are assigned the same value derived from all pixels in the block. The direction map represents the local ridge flow as derived through application of the Discrete Fourier Transform (DFT). The low contrast map separates the background from the fingerprint and detects areas which are blurry. The low flow map is derived from the direction map by assessing if the block has a dominant ridge flow. If no dominant ridge flow exists within a block then minutiae in that block are assumed to be unreliable. The high curvature map is derived from the direction map as the coherence between the directions of one block relative to its eight neighboring blocks giving an indication of the directional stability in a region. The four maps are fused into one single quality map describing the block wise quality of the finger image. The quality blocks and minutiae from the fingerprint are used to compose a quality feature vector which is used as input for a neural network (i.e. a Multi-Layer Perceptron (MLP)) which outputs a corresponding quality number representing the expected utility of the given sample.

4.3.3 Improvements to NFIQ

Since its development, NFIQ has been analyzed and areas of improvement have been identified by Merkle *et al*. [132]. Herein, a two-step process is proposed where the optimization of NFIQ is split into classification and prediction parts which improve NFIQ in the points of fingerprint data basis, similarity score statistic, image feature selection and neural network.

Some of these findings are being applied in the developments of NIST Fingerprint Image Quality 2 (NFIQ 2) [80] which is the evolution of NFIQ initiated by NIST and BSI.

4.3.4 Local finger image quality

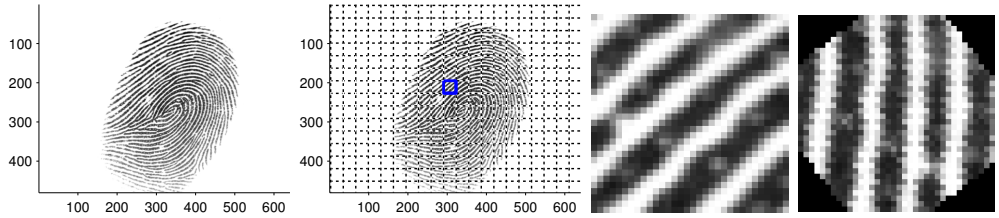
4.3.4.1 Overview

Local finger image quality assessment algorithms operate on local regions of the image where each local region (i.e. an image block) contains at least 2 ridge lines. For images with 500 pixels per inch (ppi) a ridge and valley pair is 8 to 12 pixels wide [49] and consequently, to cover two ridge lines a local region must be at least 24 pixels in either dimension.

Valid for all local finger image quality metrics is that they each assign a quality value to each local region which is aggregated in a specified manner to yield a global quality score for the finger image. It is recommended that the quality algorithm returns a map or histogram of the quality values assigned to each local region. At a later stage, such multiple quality maps can be combined to provide a robust local quality value.

In the remainder of this section state of the art of local finger image quality features are reviewed and detailed algorithm descriptions are provided. Default values provided for the variables are based on an empirical study and shall represent a good balance between computational complexity and accuracy (i.e. utility correlation).

In the following sections, a common input image will be used as example to illustrate the result of specific points in the specified algorithms. Several of the algorithms perform the same initial steps which are shown in fig. 4.3. The input image $\mathbf{I}$ shown in fig. 4.3a is image 8_5.bmp from FVC2004 DB1 [133]. For algorithms operating in a block-wise manner the image is subdivided into blocks according to the overlay grid shown in b. The block $\mathbf{V}(10,7)$ is used as example in local processing and is marked up using a bold line. c shows an enlarged view of $\mathbf{V}(10,7)$ and d shows $\mathbf{V}(10,7)$ rotated according to ridge orientation its dominant ridge orientation as determined using eq. (4.17).



(a) (b) (c) (d)

Figure 4.3: Input image used in illustrative examples of the processing of quality.

Computing the block orientation from gradients From a single block representing a local region of a fingerprint image, the dominant ridgeflow orientation is determined by computing the gradient information and then determine the orientation of the principal variation axis.

The numerical gradient of the block is determined using finite central difference for all interior pixels in x-direction eq. (4.8) and y-direction eq. (4.9).

$$\mathbf{f_x} = \frac{\mathbf{I}(x+1,y) - \mathbf{I}(x-1,y)}{2} \tag{4.8}$$

$$\mathbf{f_y} = \frac{\mathbf{I}(x,y+1) - \mathbf{I}(x,y-1)}{2} \tag{4.9}$$

With $\mathbf{f_x}$ and $\mathbf{f_y}$ the principal axis of variation of $\mathbf{V}$ is determined analytically using the sine and cosine doubled angles determined from the arithmetic means of the image gradient covariances eq. (4.17).

$$a = \overline{\mathbf{f_x}^2} \tag{4.10}$$

$$b = \overline{\mathbf{f_y}^2} \tag{4.11}$$

$$c = \overline{\mathbf{f_x} \cdot \mathbf{f_y}} \tag{4.12}$$

$$\mathbf{C} = \begin{bmatrix} a & c \\ c & b \end{bmatrix} \tag{4.13}$$

$$d = \sqrt{c^2 + (a-b)^2} + \epsilon \tag{4.14}$$

$$sin2 = \frac{c}{d} \tag{4.15}$$

$$cos2 = \frac{a-b}{d} \tag{4.16}$$

$$angle(\mathbf{V}) = \frac{\text{atan2}(sin2, cos2)}{2} \tag{4.17}$$

4.3.4.2 Frequency domain analysis

The Frequency Domain Analysis (FDA) algorithm operates in a block-wise manner. A one-dimensional signature of the ridge-valley structure is extracted and the DFT is computed on the signature to determine the frequency of the sinusoid following the ridge-valley structure [83, 99].

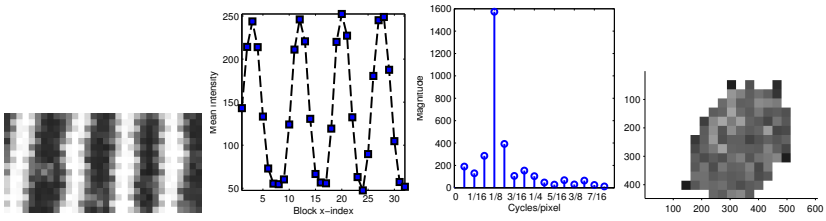
The value of $\mathbf{Q}_{\text{FDA}}^{local}$ is undefined if $F_{max} = 1$ or $F_{max} = A(\text{end})$ as both $A(0)$ and $A(\text{end}+1)$ are not accessing valid indices. Workaround in that case is to set $\mathbf{Q}_{\text{FDA}}^{local} = 1$. Despite International Organization for Standardization (ISO) recommendation [70] of a high quality value indicating a high quality, this is not the case for Frequency Domain Analysis as specified in ISO/IEC TR 29794-4:2010 [99].

Algorithm 4.1: FDA algorithm

Input: Fingerprint image $\mathbf{I}$
Output: FDA quality score Q_{FDA}
for *each block* $\mathbf{V}$ *in* $\mathbf{I}$ **do**
  rotate $\mathbf{V}$ such that dominant ridge flow is perpendicular to x-axis
  crop $\mathbf{V}$ such that no invalid regions are included
  with $\mathbf{V}$ obtain the ridge-valley signature $T(x)$ (eq. (4.18))
  compute the DFT of T to obtain F
  discard the first component of F
  determine F_{max} as the index of the term in F with largest magnitude
  compute $\mathbf{Q}_{\text{FDA}}^{local}$ of $\mathbf{V}$ using F (eq. (4.19))
compute Q_{FDA} as the arithmetic mean of all $\mathbf{Q}_{\text{FDA}}^{local}$ (eq. (4.20))

A visual overview of the algorithm outputs are depicted in fig. 4.4 where a shows $\mathbf{V}$ cropped to contain central area of rotated $\mathbf{V}$; b shows the ridge-valley profile T; c shows the DFT of T after the first component has been removed; d shows $\mathbf{Q}_{\text{FDA}}^{local}$ for each $\mathbf{V}$ in $\mathbf{I}$.

Ridge-valley signature The ridge-valley signature is a projection of the mean values of the local region along the y-axis onto a 1 dimensional vector. This effectively gives an



(a) (b) (c) (d)

Figure 4.4: Processing steps of Frequency Domain Analysis quality algorithm.

approximated representation of the fundamental periodicity within the local region. The signature is computed as:

$$T(x) = \frac{1}{\mathbf{V}_h} \sum_{k=1}^{\mathbf{V}_h} I(x,k) \tag{4.18}$$

Computing the local Frequency Domain Analysis quality score The local FDA quality score, $\mathbf{Q}_{\text{FDA}}^{local}$ is computed as:

$$\mathbf{Q}_{\text{FDA}}^{local} = \frac{A(F_{max}) + C(A(F_{max}-1) + A(F_{max}+1))}{\sum_{F=1}^{N/2} A(F)} \tag{4.19}$$

where $C = 0.3$ according to the definition appearing in ISO/IEC TR 29794-4:2010. The effect of the constant is to retain an attenuated amplitude of the frequency bands immediately surrounding F_{max}

Computing the final Frequency Domain Analysis quality score The final quality score is computed as the mean of local quality scores:

$$Q_{\text{FDA}} = \frac{1}{N * M} \sum_{i=1}^{N} \sum_{j=1}^{M} \mathbf{Q}_{\text{FDA}}^{local} \tag{4.20}$$

4.3.4.3 Gabor quality

The Gabor (GAB) quality feature [1] operates on a per-pixel basis by calculating the standard deviation of the Gabor filter bank responses. The size of the filter bank is used to determine a number of filters oriented evenly across the half circle. The strength of the response at a given location corresponds to the agreement between filter orientation and frequency in the location neighborhood. For areas in the fingerprint image with a regular ridge-valley pattern there will be a high response from one or a few filters likely neighbored orientations. In areas containing background or unclear ridge-valley structure the Gabor response of all orientations will be low and constant.

The Gabor quality feature is resolution dependent.

Algorithm 4.2: GAB algorithm

Input: Fingerprint image $\mathbf{I}$
Output: GAB quality score Q_{GAB}

1 Convolve $\mathbf{I}$ with 2D Gaussian kernel with $\sigma = 1$ and subtract from $\mathbf{I}$ to yield $\hat{\mathbf{I}}$
2 Generate filter bank by computing Gabor filter for each rotation θ (eqs. (4.21) and (4.24)); Compute the Gabor response for each pixel in $\hat{\mathbf{I}}$ for each filter in the filter bank
3 Convolve the magnitude (complex modulus) of each Gabor response with a 2D Gaussian kernel with $\sigma = 4$
4 Compute the standard deviation of the Gabor magnitude response values at each location yielding a map of standard deviations
5 Sum the map of standard deviations and normalize according to number of sample points (typically size of $\mathbf{I}$) to produce the Gabor quality score Q_{GAB}

A visual overview of the algorithm outputs are depicted in fig. 4.5 where a shows the $\hat{\mathbf{I}}$, b to e and f to i depict respectively the real and imaginary part of the Gabor filter at four orientations. The filter response is depicted in j to m and the corresponding Gaussian filtered response is depicted in n to q. Standard deviation of the Gaussian filtered responses are shown in r.

The Gabor filter The general form of the complex 2D Gabor filter h_{cx} in the spatial domain is given by:

$$h_{cx}(x, y; f, \theta, \sigma_x, \sigma_y) = \exp\left(-\frac{1}{2}\left(\frac{x_\theta^2}{\sigma_x^2} + \frac{y_\theta^2}{\sigma_y^2}\right)\exp\left(j2\pi f x_\theta\right)\right) \tag{4.21}$$

where

$$x_\theta = x\sin\theta + y\cos\theta \tag{4.22}$$
$$y_\theta = x\cos\theta - y\sin\theta \tag{4.23}$$

and f is the frequency (cycles/pixel) of the sinusoidal plane wave along the orientation θ. The size of the Gaussian smoothing window is determined by σ_x, σ_y. The filter bank size n is used to compute the differently oriented Gabor filters composing the filter bank. Computing θ given n is done as:

$$\theta = \frac{k-1}{n\pi}, k = 1, .., n \tag{4.24}$$

4.3.4.4 Gabor-Shen quality

Gabor Shen (GSH) [43] is a Gabor based feature separating blocks into two classes: good and bad. The scalar quality is the ratio between number of foreground blocks and number of foreground blocks marked as poor.

The filter response of each Gabor kernel in the filter bank are computed on the pixels in each block and a standard deviation is computed on the responses. Using thresholding

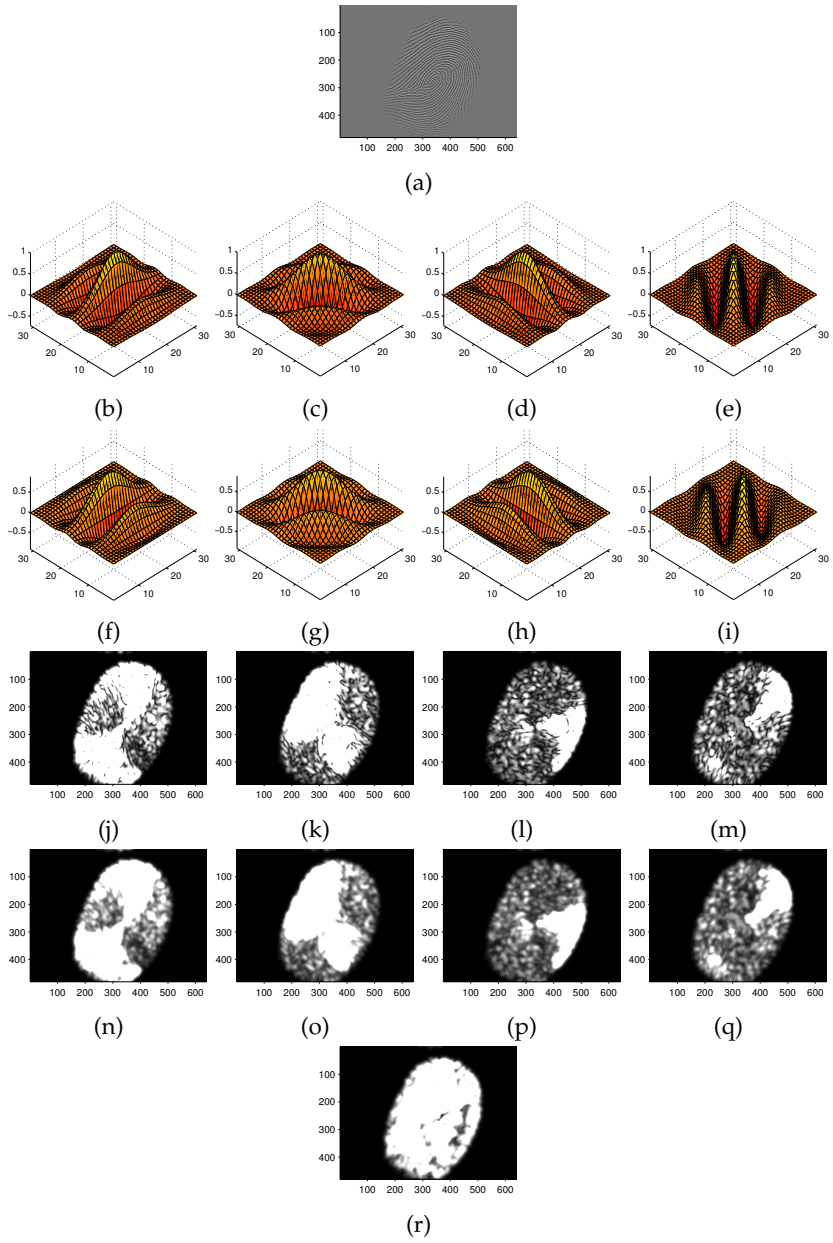


Figure 4.5: Processing steps of Gabor quality algorithm. r indicates in light grayscale values those image regions, where filter responses show large standard deviation and good image quality is assumed.

each block is determined to be foreground, background and poor or good quality. The algorithm is outlined in algorithm 4.3.

The Gabor Shen quality feature is resolution dependent. Shen et al. suggest $\sigma_x = \sigma_y = 4, f = 0.12, n = 8, b = 30$ and that T_b and T_q are empirically determined according to dataset. We have found that setting $T_b = 1$ and $T_q = 2$ to yield good results across several datasets.

Algorithm 4.3: GSH algorithm

Input: Fingerprint image **I**
Output: GSH quality score Q_{GSH}
Generate filter bank of size n by computing Gabor filter for each rotation θ
for *each block* **V** *in* **I** **do**
    Compute the n Gabor responses at each pixel in **V**
    Compute the standard deviation G of Gabor reponses for **V**
With G segment **I** into foreground and background
Determine foreground blocks V_f as those **V** where $\mu_i > T_b$
Determine poor quality blocks V_p as those **V** where $\mu_i > T_b$ and $\mu_i < T_q$
Compute quality score Q_{GSH} as $1 - \frac{V_p}{V_f}$

A visual overview of the algorithm outputs are depicted in fig. 4.6 where a to d and e to h depict respectively the real and imaginary part of the Gabor filter at four orientations. The filter response is depicted in i to l and the standard deviation of the responses in m. The mean of standard deviations in each block is show in n indicating in light grayscale values those image regions, where good quality is assumed. Blocks marked as foreground in o and blocks marked as poor in p.

4.3.4.5 Local clarity score

Local Clarity Score (LCS) [72, 99] computes the block-wise clarity of ridge and valleys by applying linear regression to determine a gray-level threshold, classifying pixels as ridge or valley. A ratio of misclassified pixels is determined by comparing with the normalized ridge and valley width of that block.

Particular regions inherent in a fingerprint will negatively affect Q_{LCS}. For example, ridge endings and bifurcations or areas with high curvature such as those commonly found in the vicinity of core and delta points.

A visual overview of the algorithm outputs are depicted in fig. 4.7 a crop of the current block shown in a with the average profile depicted in b. The grayscale values of the profile and the regression line determined using linear regression is shown in c, with the corresponding binarization into ridge (black) and valley (white) in d. Figures e through g show respectively pixels in the input block determined to be ridge using threshold; gray level threshold for ridges across the block shown; pixels above threshold marked in white. Similarly for valleys in h through j. The local clarity score for each block is visualized in k with white being maximum clarity.

Determining the proportion of misclassified pixels For a block **V** there are v_T pixels in the valley region and v_B pixels in the valley region with intensity lower than a threshold DT. Similarly there are r_T pixels in the ridge region and r_B pixels in the ridge region with intensity lower than a threshold DT. α and β are expressions of these ratios.

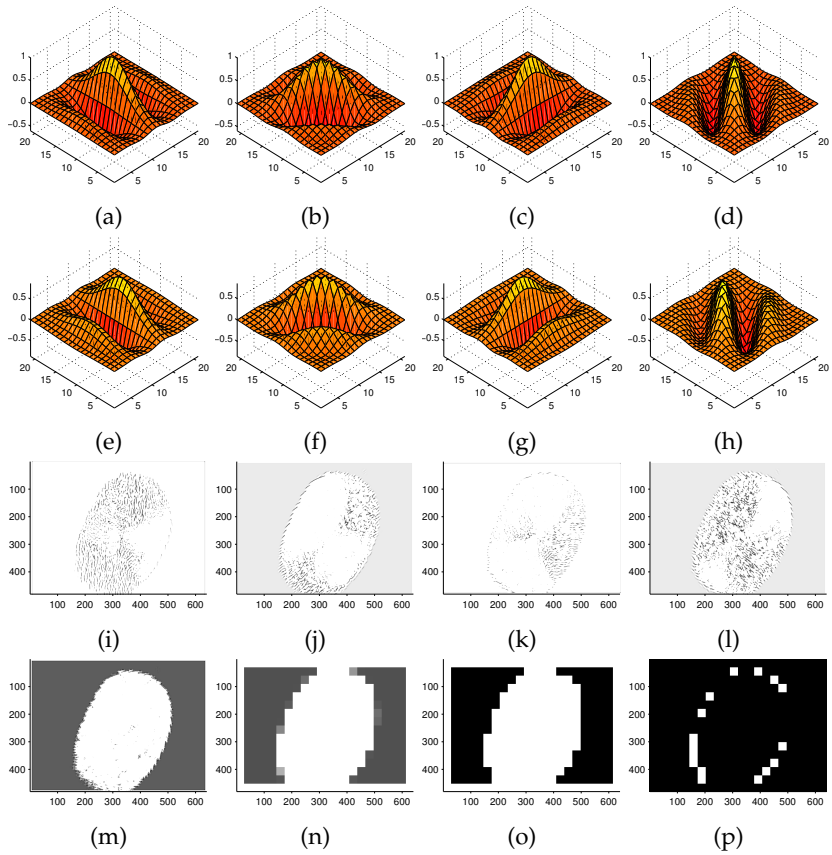


Figure 4.6: Processing steps of Gabor Shen quality algorithm.

$$\alpha = \frac{v_B}{v_T} \tag{4.25}$$

$$\beta = \frac{r_B}{r_T} \tag{4.26}$$

Determining the normalized ridge and valley width The normalized valley width $\bar{W}_v$ and the normalized ridge width $\bar{W}_r$ are determined as

$$\bar{W}_v = \frac{W_v}{\left(\frac{S}{125}\right) W^{max}} \tag{4.27}$$

$$\bar{W}_r = \frac{W_r}{\left(\frac{S}{125}\right) W^{max}} \tag{4.28}$$

where S is the sensor resolution in dpi, W^{max} is the estimated ridge or valley width for an image with 125 dpi resolution, and W_v and W_r are the observed valley and ridge widths. According to [49], $W^{max} = 5$ is reasonable for 125 dpi resolution.

Algorithm 4.4: LCS algorithm

Input: Fingerprint image **I**
Output: LCS quality score Q_{LCS}
for *each block* **V** *in* **I do**
  rotate **V** such that dominant ridge flow is perpendicular to x-axis
  crop **V** such that no invalid regions are included
  with **V** obtain the ridge-valley signature $T(x)$ (eq. (4.18))
  Determine a threshold DT using linear regression on $T(x)$
  Determine proportion of misclassified pixel β and α in the ridge and valley regions (eqs. (4.25) and (4.26))
  Determine normalized ridge width and valley width $\bar{W}_r$ and $\bar{W}_v$ (eqs. (4.27) and (4.28))
  Compute the local quality score $\mathbf{Q}_{\text{LCS}}^{local}$ (eq. (4.31))
Compute quality score Q_{LCS} (eq. (4.32))

For 500 dpi resolution, eqs. (4.27) and (4.28) may be expressed as

$$\hat{W_v} = \frac{W_v}{20} \tag{4.29}$$

$$\hat{W_r} = \frac{W_r}{20} \tag{4.30}$$

Computing the Local Clarity Score The local quality score $\mathbf{Q}_{\text{LCS}}^{local}$ is computed using the average value of α and β in valid ridge and valley regions:

$$\mathbf{Q}_{\text{LCS}}^{local} = \begin{cases} \left(1 - \frac{\alpha+\beta}{2}\right) & , \left(W_v^{nmin} < \bar{W}_v < W_v^{nmax}\right)\left(W_r^{nmin} < \bar{W}_r < W_r^{nmax}\right) \\ 0 & , otherwise \end{cases} \tag{4.31}$$

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

Computing the quality score The quality score Q_{LCS} is computed as the mean of local scores $\mathbf{Q}_{\text{LCS}}^{local}$ using eq. (4.31) where a higher Q_{LCS} score indicates higher quality.

$$Q_{\text{LCS}} = \frac{1}{N * M} \sum_{i=1}^{N} \sum_{j=1}^{M} \mathbf{Q}_{\text{LCS}}^{local}(i,j) \tag{4.32}$$

4.3.4.6 Orientation flow

Orientation Flow (OFL) [72, 99] is a measure of ridge flow continuity which is based on the absolute orientation difference between a block and its neighboring blocks.

In ISO/IEC TR 29794-4:2010 [99] the parameter θ_{min} is a constant such that the angular tolerance is $8\,\text{deg}$ ($\theta_{min} = 8$), however here we set the tolerance to $0\,\text{deg}$. The local orientation quality score in this case is assigned such that $\mathbf{Q}_{\text{OFL}}^{local}(i,j) = 1$ when the local quality

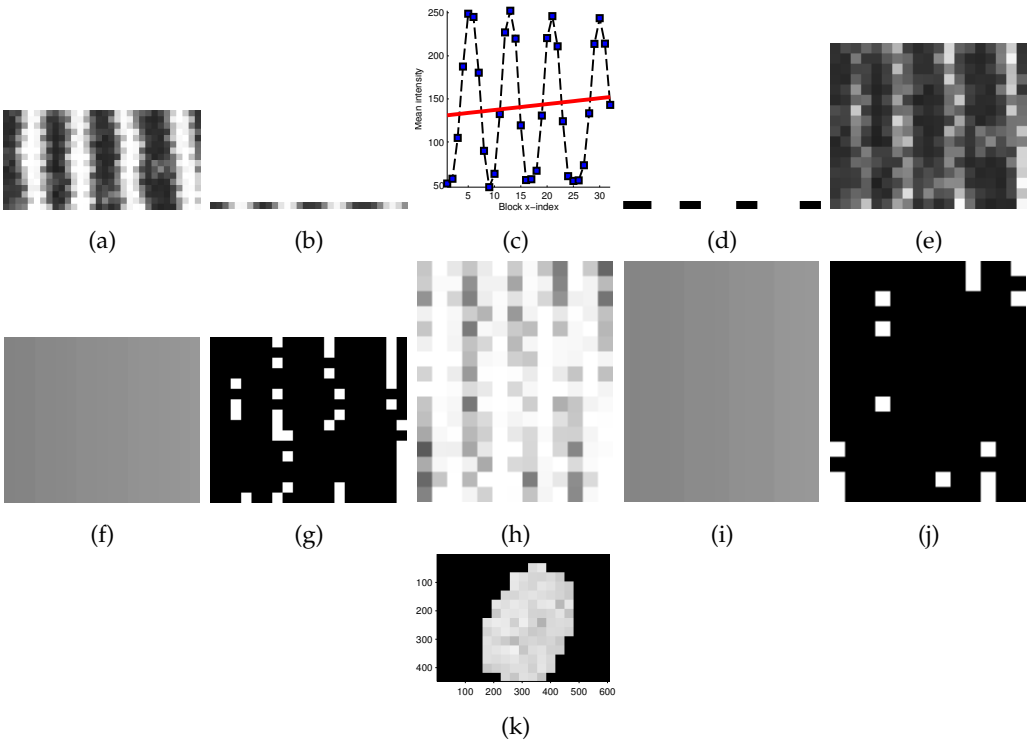


Figure 4.7: Processing steps of Local Clarity Score algorithm. k indicates in light grayscale values those image regions, where good quality is assumed.

is the highest. This is opposite behavior of the ISO/IEC TR 29794-4:2010 definition where $\mathbf{Q}_{\text{OFL}}^{local}(i,j) = 0$ when the local quality is the highest.

Algorithm 4.5: OFL algorithm

Input: Fingerprint image $\mathbf{I}$
Output: OFL quality score Q_{OFL}
for *each block* $\mathbf{V}$ *in* $\mathbf{I}$ **do**
  Determine the dominant ridgeflow orientation (eq. (4.17))
  Compute the absolute orientation difference $D(i,j)$ (eq. (4.33))
  Compute the local orientation quality score $\mathbf{Q}_{\text{OFL}}^{local}(i,j)$ (eq. (4.34))
Compute the quality score Q_{OFL} (eq. (4.35))

A visual overview of the algorithm outputs are depicted in fig. 4.8 where a line marking the normal to the ridge line orientation in a; the blockwise dominant orientations are depicted in b and the orientation differences are shown in c; the local quality scores are shown in d.

Absolute orientation difference The ridge flow is determined as a measure of the absolute difference between a block and its neighboring blocks. The absolute difference for

block $\mathbf{V}(i,j)$ is:

$$D(i,j) = \frac{\sum_{m=-1}^{1}\sum_{n=-1}^{1}|\mathbf{V}(i,j) - \mathbf{V}(i-m,j-n)|}{8} \tag{4.33}$$

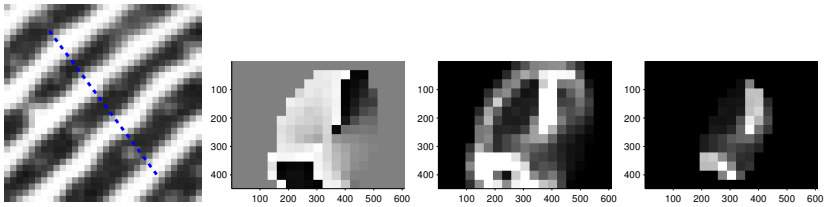
Local orientation quality score The local orientation quality score $\mathbf{Q}_{\text{OFL}}^{local}(i,j)$ for the block orientation difference $D(i,j)$ is determined as:

$$\mathbf{Q}_{\text{OFL}}^{local}(i,j) = \begin{cases} \dfrac{D(i,j) - \theta_{min}}{90\deg - \theta_{min}} & , D(i,j) > \theta_{min} \\ 0 & , otherwise \end{cases} \tag{4.34}$$

where θ_{min} is a threshold for the minimum angle difference to be considered as significant.

Computing the quality score With $N*M$ local orientation quality score blocks the global orientation quality score is computed as

$$Q_{\text{OFL}} = 1 - \frac{1}{N*M}\sum_{i=1}^{N}\sum_{j=1}^{M}\mathbf{Q}_{\text{OFL}}^{local}(i,j) \tag{4.35}$$



(a) (b) (c) (d)

Figure 4.8: Processing steps of Orientation Flow quality algorithm.

4.3.4.7 Orientation certainty level

Orientation Certainty Level (OCL) [71, 99] is a measure of the strength of the energy concentration along the dominant ridge flow orientation. The feature operates in a block-wise manner.

The computation of OCL presented here deviates from ISO/IEC 29794-4:2010 [99] in that we subtract the ratio between the eigen values from 1 such that $Q_{\text{OCL}} = 0$ reflects the lowest quality and $Q_{\text{OCL}} = 1$ the highest quality.

A visual overview of the algorithm outputs are depicted in fig. 4.9 where the ratio of the eigenvalues is shown in a as an ellipse and the local qualities are depicted in b where higher intensity indicates higher $\mathbf{Q}_{\text{OCL}}^{local}$ score.

Algorithm 4.6: OCL algorithm

Input: Fingerprint image $\mathbf{I}$
Output: OCL quality score Q_{OCL}
for *each block* $\mathbf{V}$ *in* $\mathbf{I}$ **do**
  Compute the intensity gradient with centered differences method
  Compute the covariance matrix $\mathbf{C}$ (eq. (4.13))
  Compute the eigenvalues of $\mathbf{C}$ to obtain $\mathbf{Q}_{\text{OCL}}^{local}$ (eqs. (4.36) and (4.38))
Compute the quality score Q_{OCL} as the mean of all $\mathbf{Q}_{\text{OCL}}^{local}$ (eq. (4.39))

Computing the eigenvalues and local orientation certainty From the covariance matrix C the eigenvalues λ_{min} and λ_{max} are computed as

$$\lambda_{min} = \frac{a + b - \sqrt{(a-b)^2 + 4c^2}}{2} \tag{4.36}$$

$$\lambda_{max} = \frac{a + b + \sqrt{(a-b)^2 + 4c^2}}{2} \tag{4.37}$$

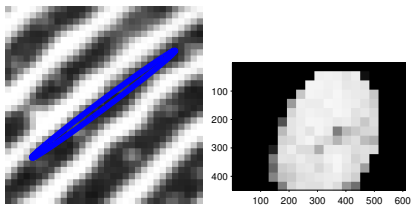
this yields a local orientation certainty level $\mathbf{Q}_{\text{OCL}}^{local}$:

$$\mathbf{Q}_{\text{OCL}}^{local} = 1 - \frac{\lambda_{min}}{\lambda_{max}} \tag{4.38}$$

which is a ratio in the range 0 to 1 respectively the lowest and highest orientation certainty level.

Computing the quality score The quality score Q_{OCL} is computed as the mean of local scores $\mathbf{Q}_{\text{OCL}}^{local}$ using eq. (4.39) where a higher Q_{OCL} score indicates higher quality.

$$Q_{\text{OCL}} = \frac{1}{N * M} \sum_{i=1}^{N} \sum_{j=1}^{M} \mathbf{Q}_{\text{OCL}}^{local}(i,j) \tag{4.39}$$



(a) (b)

Figure 4.9: Processing steps of Orientation Certainty Level quality algorithm. The light grayscale values shown in b indicate regions with good local quality.

4.3.4.8 Ridge valley uniformity

Ridge Valley Uniformity is a measure of the consistency of the ridge and valley widths [71, 99]. The expectation for a finger image with clear ridge and valley separation is that the ratio between ridge and valley widths remains fairly constant and thus the standard deviation of ratios is used as an indication of the sample quality. The Ridge Valley Uniformity quality feature is resolution dependent.

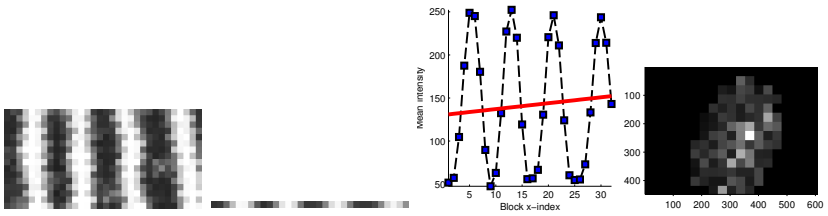
Algorithm 4.7: RVU algorithm

Input: Fingerprint image $\mathbf{I}$
Output: Ridge Valley Uniformity (RVU) quality score Q_{RVU}

for *each block* $\mathbf{V}$ *in* $\mathbf{I}$ **do**
- Determine dominant ridgeflow orientation $\theta(\mathbf{V})$ according to eq. (4.17)
- Rotate $\mathbf{V}$ such that $\theta(\mathbf{V})$ is perpendicular to x-axis
- crop $\mathbf{V}$ such that no invalid regions are included
- with $\mathbf{V}$ obtain the ridge-valley signature $T(x)$ (eq. (4.18))
- Determine a threshold DT using linear regression on $T(x)$
- Binarize $T(x)$ using DT
- Determine transition points in $T(x)$
- Drop first and last transition from $T(x)$ to remove incomplete ridges or valleys
- Determine $\mathbf{Q}_{\text{RVU}}^{local}$ from the ratio between remaining ridge and valley widths

Compute the quality score Q_{RVU} as the standard deviation of all $\mathbf{Q}_{\text{RVU}}^{local}$

A visual overview of the algorithm outputs are depicted in fig. 4.10 where a crop of the valid area of the current block is depicted in a; the ridge-valley signature is shown in b and the result of linear regression on the intensities is depicted in c. The resulting standard deviation of local ratios is shown in d.



(a) (b) (c) (d)

Figure 4.10: Processing steps of Ridge Valley Uniformity quality algorithm. As this algorithm renders local quality not compliant to ISO/IEC 29794-1:2009 the dark grayscale values shown in d indicate regions with good local quality.

4.3.5 Global finger image quality

4.3.5.1 Radial power spectrum

Radial Power Spectrum (RPS) [74, 99] is a measure of maximal signal power in a defined frequency band of the global radial Fourier spectrum. Ridges can be locally approximated by means of a single sine wave, hence high energy concentration in a narrow frequency band corresponds to consistent ridge structures.

Algorithm 4.8: RPS algorithm

Input: Fingerprint image $\mathbf{I}$
Output: RPS quality score Q_{RPS}
1 Filter $\mathbf{I}$ using Blackman window
2 Compute 2D-FFT $\mathcal{F}\{\mathbf{I}(x,y)\}$ of $\mathbf{I}$ (eq. (4.40))
3 Apply Logarithm operator to power spectrum
4 Get range of frequency band of interest
5 Convert to polar coordinates
6 Extract frequency band of interest
7 Compute magnitude of bands (eq. (4.42))
8 Determine Q_{RPS} as the maximum magnitude of any band in frequency range of interest (eq. (4.43))

A visual overview of the algorithm outputs are depicted in fig. 4.11 where the input image convolved with a Blackman filter (algorithm 4.8, line 1) is shown in a. The log transformed spectrum of the DFT is shown in b and the conversion to polar coordinates in c with lines in upper part indicating the frequency band of interest. The extracted band of frequencies is shown in d, and the power is shown in e and the power of binned frequencies is shown in f.

2D Fourier Transform The 2D discrete Fourier transform $F(u,v) = \mathcal{F}\{\mathbf{I}(x,y)\}$ is:

$$F(u,v) = \frac{1}{\mathbf{I}_h \mathbf{I}_w} \sum_{x=0}^{\mathbf{I}_w - 1} \sum_{y=0}^{\mathbf{I}_h - 1} \mathbf{I}(x,y) \exp\left(-j2\pi\left(\frac{ux}{\mathbf{I}_w} + \frac{vy}{\mathbf{I}_h}\right)\right) \tag{4.40}$$

The magnitude of $F(u,v)$ is computed as:

$$|F(u,v)| = \sqrt{\Re(F(u,v))^2 + \Im(F(u,v))^2} \tag{4.41}$$

Magnitude of frequency bands polar coordinates The magnitude of the annular band between r and $r + \delta_r$ in the polar Fourier spectrum $F(\alpha, r)$ is computed as:

$$J(r) = \sum_{\alpha} F(\alpha, r) \tag{4.42}$$

Determine quality score The quality feature Q_{RPS} is found as:

$$Q_{\text{RPS}} = \max |J(r)| \tag{4.43}$$

4.3.5.2 Mu

The MU quality feature is the mean value of the input image.

4.3.5.3 Sigma

The SIG quality feature is the standard deviation of the input image.

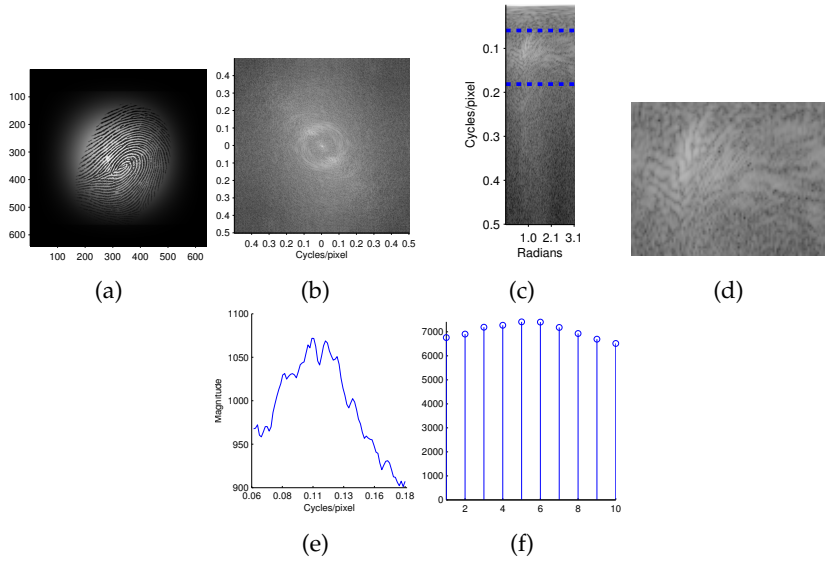


Figure 4.11: Processing steps of Radial Power Spectrum quality algorithm.

Algorithm 4.9: MU algorithm

Input: Fingerprint image $\mathbf{I}$
Output: MU quality score Q_{MU}
1 Compute Q_{MU} as the mean of pixel values in $\mathbf{I}$;

Algorithm 4.10: SIG algorithm

Input: Fingerprint image $\mathbf{I}$
Output: SIG quality score Q_{SIG}
1 Compute Q_{SIG} as the standard deviation of pixel values in $\mathbf{I}$;

4.4 Performance Evaluation

For the performance evaluation we have chosen to assess the predictive performance of NFIQ and the quality features specified in sections 4.3.4 and 4.3.5.

We base our results on finger image data from Fingerprint Verification Competition 2004 (FVC2004) [133], Fingerprint Verification Competition 2006 (FVC2006) [134], MCYT Fingerprint 330 (MCYT-330) [135] and Wet-Dry Dataset 2 (WDSET02) [7] databases which are summarized in section 4.4. For the dataset WDSET02 5 different capture devices were used and the capture protocol specified 4 finger surface conditions: no treatment with capture as the subject arrives; application of an alcohol based solution to dry the skin surface; application of hand-moisturiser; and finally soaking in water. FVC2004 DB1 and FVC2006 DB2A were captured in a supervised manner were subjects were instructed to perform various actions which result in various degradations such as elastic deformation and uneven pressure. Samples in MCYT330 DP were captured using three levels finger placement requirements on the sensor: relaxed but under supervision by operator; placement within

specified rectangle; and high level of control where one core and/or delta must fall within a specified rectangle. The datasets combined provide a total of $48\,760$ fingerprint image samples captured using optical sensors.

For each dataset, comparison scores are computed using 3 commercial state of the art minutiae based comparison systems denoted as A, B, and C as agreements prohibit us from revealing the names. All genuine comparisons were made and imposter scores were determined by randomly selecting a set of probes from other subjects (see section 4.4) and compare with the reference template. In the cases where a provider failed to either compute a template or a comparison score we have set the comparison score to 0.

We evaluate and rank the quality features according to their Error-Reject Curve (ERC) characteristics and Spearman correlation with utility values as computed in eq. (4.1). In the case of ERC, we focus on the decrease in FNMR achieved when rejecting the lowest 20% quality samples as rejecting more than 1 in 5 samples due to quality is unlikely to be acceptable in an operational system (see section 4.2.5.3).

For features Q_{FDA}, Q_{LCS}, Q_{OCL}, Q_{OFL}, Q_{RVU} the quality values were computed on the image foreground as determined by segmentation based on local standard deviation image with a block size of 32 by 32 pixels and a threshold of 0.1 with ridgesegment.m method [136].

Database	FVC2004 DB1 [133]	FVC2006 DB2A [134]	MCYT330 DP [135]	WDSET02 [7]
Sensor	Optical (Single)	Optical (Single)	Optical (Single)	Optical (Multiple)
Subjects	30		330	33
Fingers	4	140	10	10
Impressions	8	12	12	20
Total images	880	1680	39 600	6600
Genuine	6160	18 480	435 600	125 400
Imposter	17 600	840 000	19 800 000	660 000

Table 4.1: Datasets used for performance evaluation of quality algorithms.

4.4.1 Detection error trade-off

In order to assess the characteristics of the datasets and the biometric feature extractors and comparators a benchmark is performed by computing the Detection Error Tradeoff (DET) on each dataset. The DET plots are shown in fig. 4.12. It is observed that at a fixed False Match Rate (FMR) at 10^{-4} the FNMR shifts over several orders of magnitude depending on the dataset used. It is observed that A is performing best across datasets, while C is the worst performing. Further, providers perform the worst on WDSET02 and the best on FVC2006 DB2A.

Providers perform the worst on WDSET02, one reason for this is that the data collection protocol for that dataset included the capture of soaking wet fingerprints which results in very poor quality samples for some capture devices.

4.4.2 Feature correlations

Correlation of features has a significance when assessing which features to include in a composite quality algorithm, i.e. one that uses input from multiple quality algorithms to

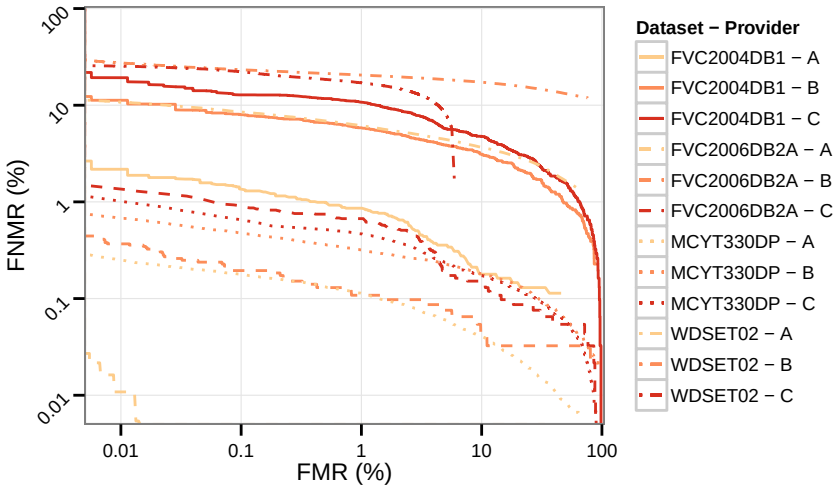


Figure 4.12: Detection error tradeoff curves for datasets FVC2004 DB1, FVC2006 DB2A, MCYT330 DP, and WDSET02 using comparison scores from providers A, B, and C.

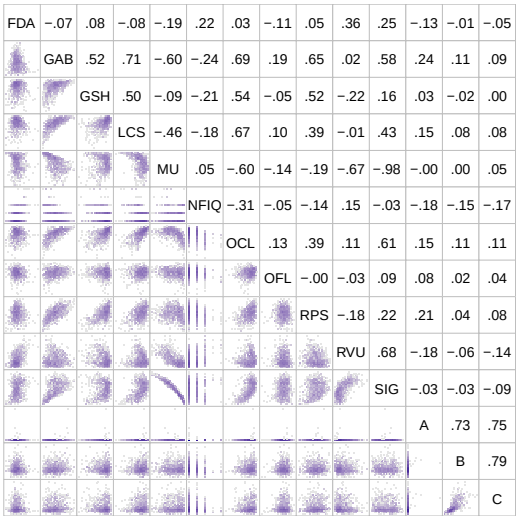
determine a scalar quality value or a vector of multiple selected quality values for a sample. In such case it is advisable to select features which complement each other, i.e. they do not correlate with each other.

Correlation tables for quality values and utility computed on the four datasets are shown in tables 4.2a to 4.2d. The upper half shows the Spearman correlation between two variables, the lower half depicts a scatter plot of the two variables and the diagonal contains the name of the variable in that row and column.

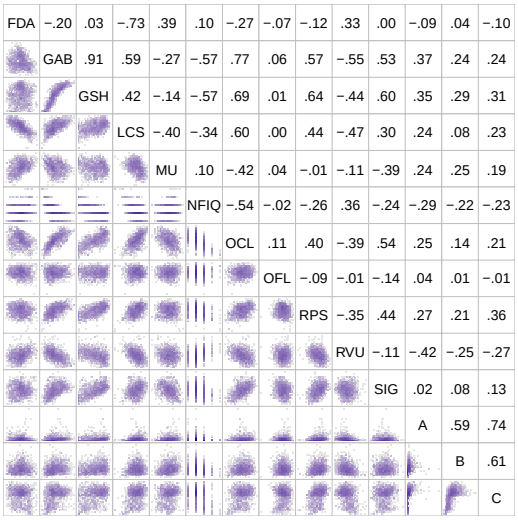
We note that the correlation between features is clearly dependent on the dataset on which they where computed, for example consider the correlation between FDA and MU, which is varying significantly by being positively correlated (0.25) in FVC2004 DB1 and WDSET02 (0.32), not correlated (0.00) in FVC2006 DB2A, and negatively correlated (-0.23) in MCYT330 DP. The correlations between the features and utility computed according to eq. (4.1) for each provider also reveals significant differences across datasets, e.g. the correlation between NFIQ and each of A, B, C is respectively -0.18, -0.15, -0.17 for FVC2004 DB1 while for WDSET02 the correlations with the same providers are -0.65, -0.56 and -0.52. Similarly for GAB we see major differences, and in the case of B on MCYT330 DP there is a correlation of -0.03, while it is 0.49 on WDSET02.

4.4.3 Error reject curves

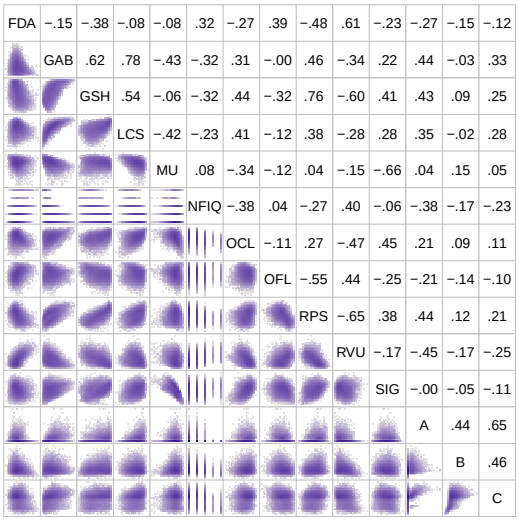
We determine the behaviour of quality features over the range of quality values in relation to the FNMR using ERC (section 4.2.5.3). Ideally, when a sample is removed due to its computed quality value, a corresponding decrease in FNMR occurs, i.e. the idealised quality feature identifies exactly those samples which contribute to FNMR and its behavior in the ERC is that the resulting reduction in FNMR occurs at the same rate as the rejection due to quality. When a sample which did not cause a false non-match is removed due to its quality level, it will result in an increase in the FNMR which occurs because the FNMR is



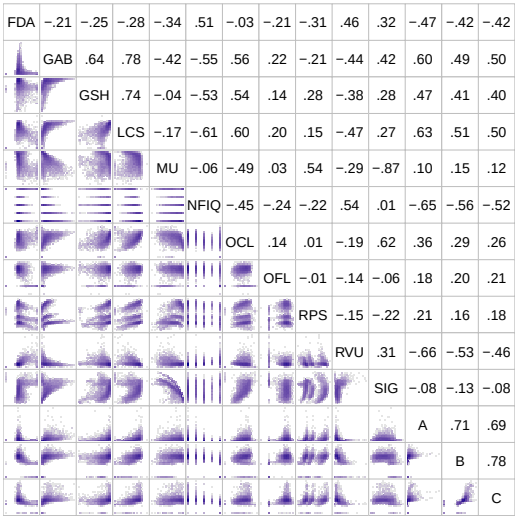
(a) FVC2004 DB1



(b) FVC2006 DB2A



(c) MCYT330 DP



(d) WDSET02

Table 4.2: Spearman correlation table and scatter plots of quality features and utilities for each dataset FVC2004 DB1 a, FVC2006 DB2A b, MCYT330 DP c and WDSET02 d. The upper half shows the Spearman correlation between two variables, the lower half depicts a scatter plot of the two variables and the diagonal contains the name of the variable in that row and column.

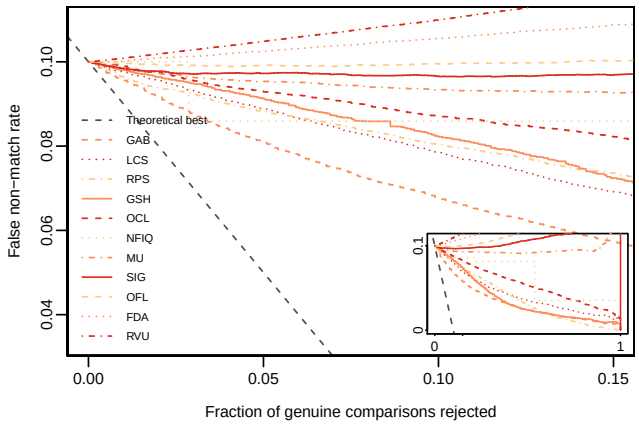
computed using the genuine comparison scores of those samples remaining after samples have been removed due to their inferior quality as determined be the respective algorithm (see eq. (4.5)).

For each feature and combination of dataset and comparison score provider we computed the ERC, noting the η^{erc}_{auc} and η^{erc}_{pauc20} as computed according to eqs. (4.6) and (4.7). The resulting curves for the MCYT330 DP dataset using provider A for comparison scores are displayed in fig. 4.13, where we have used two different values of f, one $f = 0.1$ to simulate an operational case where the FNMR is 10%, and one with $f = 0.01$ where the FNMR is 1% (see eq. (4.4)). We show respectively the first 15% and 1.5% of samples being rejected due to computed quality value. The inlay shows the full ERC from no samples removed to all samples removed. The resulting η^{erc}_{auc} and η^{erc}_{pauc20} computed from ERC for each case of f are summarized in tables 4.3 and 4.4.

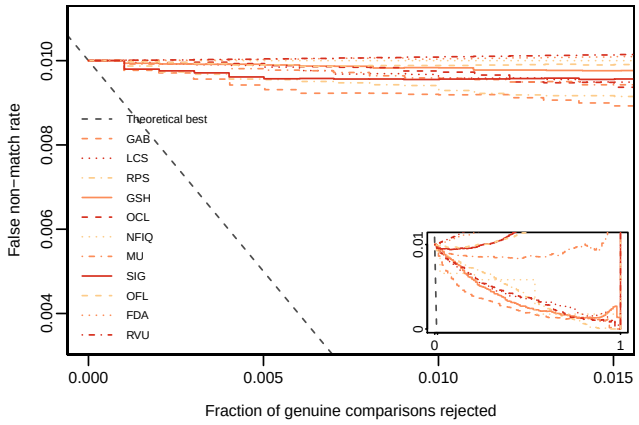
All the presented features provide a wide range of possible quality values, while NFIQ by design provides 5 quality levels. The ERC show that NFIQ is more robust than individual features as the FNMR in the ERC is decreasing as comparisons containing samples associated with the rejected quality level are removed. Only in a single instance, the combination of WDSET02 and provider B, do we note that NFIQ exhibits unexpected behavior where the FNMR is increasing. The cost of the relatively small number of distinct quality levels is that one must always choose to reject an entire level of quality, resulting in a rough step function when then fraction of samples rejected is plotted against the resulting FNMR. This behaviour is visible in fig. 4.13a where the FNMR remains at 0.10 until around 3% of the samples are rejected where after the FNMR drops to slightly more than 8% and stays at that level until around 20% of the samples are rejected and the FNMR is further reduced to around 4%.

Table 4.3: Summary of AUC and Partial Area Under Curve (PAUC) for ERC plots of quality features on FVC2004 DB1, FVC2006 DB2A, MCYT330 DP, WDSET02 computed on ERC using $f = 0.1$.

	FVC2004 DB1		FVC2006 DB2A		MCYT330 DP		WDSET02	
Feature	η^{erc}_{auc}	η^{erc}_{pauc20}	η^{erc}_{auc}	η^{erc}_{pauc20}	η^{erc}_{auc}	η^{erc}_{pauc20}	η^{erc}_{auc}	η^{erc}_{pauc20}
GAB	.068	.014	**.023**	**.010**	**.026**	**.009**	**.017**	**.008**
LCS	.072	.014	.046	**.010**	.036	.011	.023	.010
RPS	.062	.014	.051	.013	.031	.011	.089	.014
GSH	.097	.014	.028	.011	.030	.011	.036	.014
OCL	.081	.015	.039	.011	.049	.013	.063	.012
NFIQ	.068	**.013**	.062	.014	.045	.013	.040	.015
MU	.076	.015	.062	.013	.092	.014	.066	.015
SIG	.115	.016	.077	.014	.105	.014	.120	.014
OFL	**.066**	.013	.083	.012	.140	.015	.048	.011
FDA	.121	.015	.147	.015	.174	.016	.181	.015
RVU	.138	.017	.206	.017	.207	.017	.177	.015



(a)



(b)

Figure 4.13: ERC for features on MCYT330 DP using A as basis for FNMR computation with ERC starting respectively at 10% FNMR ($f = 0.1$) in a and 1% ($f = 0.01$) in b. See tables 4.3 and 4.4 for a summary of each instance of f.

Table 4.4: Summary of AUC and PAUC for ERC plots of quality features on FVC2004 DB1, FVC2006 DB2A, MCYT330 DP, WDSET02 computed on ERC using $f = 0.01$.

Feature	FVC2004 DB1		FVC2006 DB2A		MCYT330 DP		WDSET02	
	η^{erc}_{lauc}	η^{erc}_{pauc20}	η^{erc}_{lauc}	η^{erc}_{pauc20}	η^{erc}_{lauc}	η^{erc}_{pauc20}	η^{erc}_{lauc}	η^{erc}_{pauc20}
GAB	0.008	0.002	**0.002**	**0.001**	**0.003**	**0.001**	**0.001**	**0.001**
LCS	0.010	0.002	0.004	**0.001**	0.004	**0.001**	0.002	**0.001**
RPS	**0.005**	0.002	0.004	0.002	0.004	0.002	0.017	0.003
GSH	0.011	0.002	**0.002**	0.002	0.004	0.002	0.004	0.003
OCL	0.007	0.002	0.004	**0.001**	0.004	0.002	0.013	0.002
NFIQ	0.007	0.002	0.005	0.002	**0.003**	**0.001**	0.004	0.002
MU	0.009	0.002	0.007	0.002	0.011	0.002	0.004	0.002

Table 4.4: *(continued)*

	FVC2004 DB1		FVC2006 DB2A		MCYT330 DP		WDSET02	
Feature	η^{erc}_{lauc}	η^{erc}_{pauc20}	η^{erc}_{lauc}	η^{erc}_{pauc20}	η^{erc}_{lauc}	η^{erc}_{pauc20}	η^{erc}_{lauc}	η^{erc}_{pauc20}
SIG	0.011	0.002	0.006	0.002	0.014	0.002	0.022	0.003
OFL	0.011	0.002	0.008	**0.001**	0.014	0.002	0.002	**0.001**
FDA	0.010	0.002	0.017	0.002	0.021	0.002	0.017	0.002
RVU	0.016	0.002	0.029	0.002	0.026	0.002	0.008	**0.001**

4.5 Conclusion

Fingerprint sample quality continues to play a large role in biometric systems and the presented review and performance evaluation of existing quality metrics has shown that the algorithms have varying performance across dataset and comparison score vendor. This considerably complicates the task of selecting a subset of features which are suitable for a combined or aggregated quality metric. Inter-feature Spearman correlation give indications as to which features are overlapping with respect to the aspect of the fingerprint they measure.

An important point to address in future research is how to best combine quality features in a way which minimizes the dependence on individual comparison score providers while maintaining a sufficient predictive performance with respect to the biometric performance.

The selection of quality features is further complicated by their non-monotonic behavior, i.e. the feature might only be indicative of sample utility in a range of the values it produces, and it is not guaranteed that a high quality value indicates higher biometric performance across all ranges within the possible values. Using η^{erc}_{lauc} and η^{erc}_{pauc20} as indicators for feature provides insights to the behavior of quality features throughout the quality values it provides in the context of FNMR, however it is clear from out results that the ordering of the features using these indicators changes depending on the chosen f parameter in the calculation of ERC. With $f = 0.01$ and using η^{erc}_{lauc} as selection criteria, we conclude that GAB, LCS, GSH provides consistent high performance across the tested datasets, and that RPS, OFL and OCL appears sensitive to the dataset used.

Our results show that correlation with utility as computed using the definitions by ISO/IEC is not indicative of the behavior of the quality feature when evaluated using ERC and our recommendation is to apply feature correlations only in the context of identifying potentially overlapping methods of quality measurements.

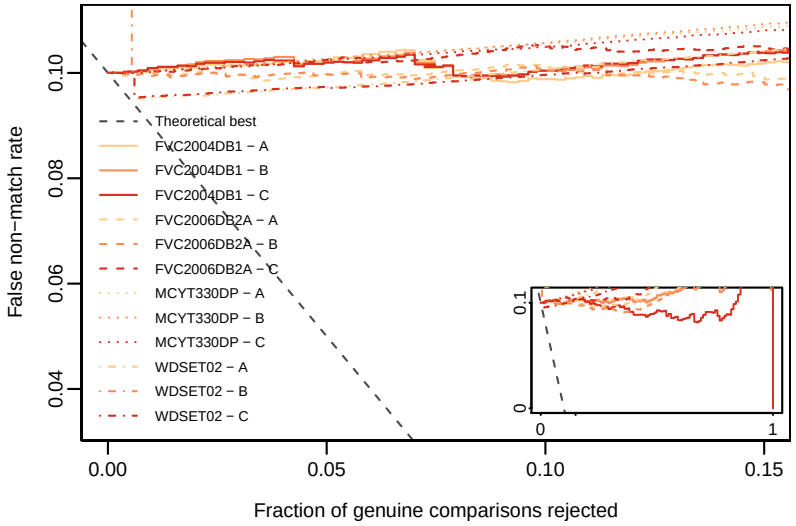
The quality features we have specified and evaluated are made available to the community [121]. We anticipate that parameters in the quality features can be tuned to achieve a higher performance on specific datasets or a better generalisability when faced with new datasets.

Acknowledgements

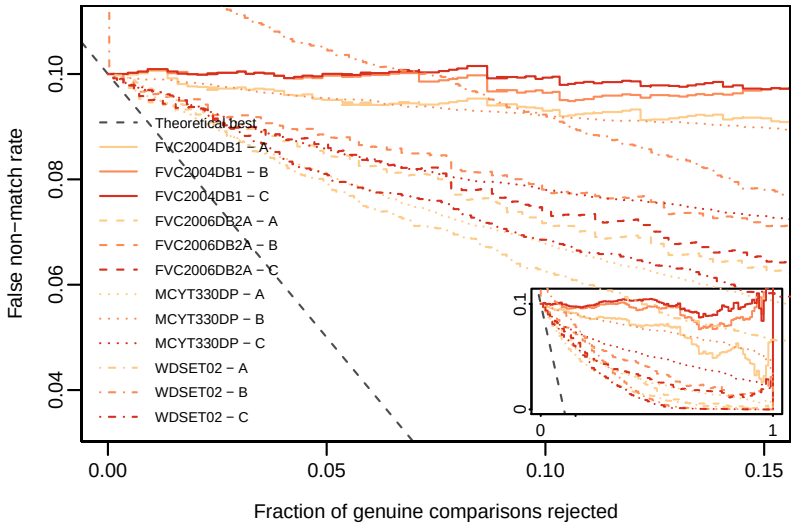
The authors would like to acknowledge Center for Advanced Security Research Darmstadt (CASED), BSI, NIST, secunet Security Networks AG (SECUNET), and Fraunhofer-Institut für Graphische Datenverarbeitung (IGD) for their support in the work presented here. We

thank Universidad Autónoma de Madrid (UAM) for providing the MCYT-330 dataset used in our evaluation of sample quality features.

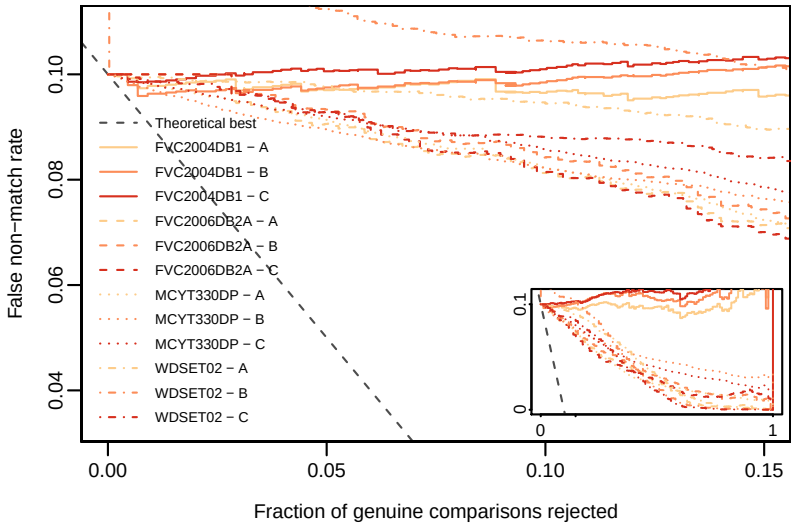
4.6 Error-reject curves for quality features



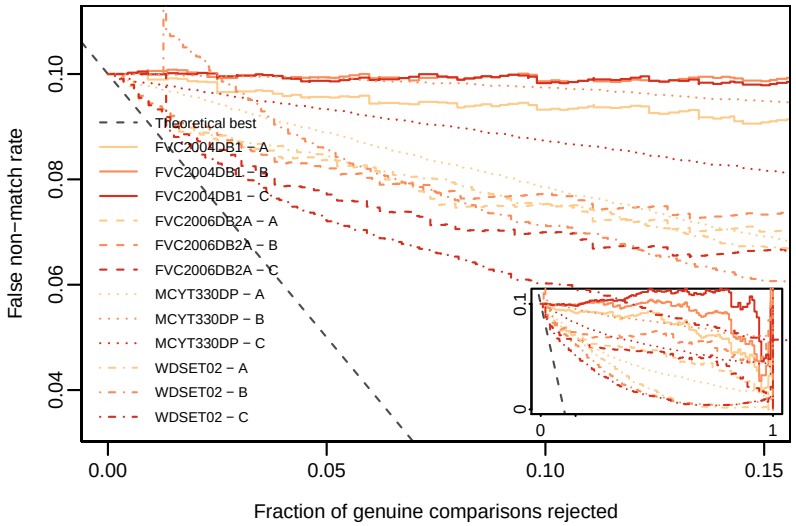
(a) FDA


False non−match rate
0.04
0.06
0.08
0.10
0.00
0.05
0.10
0.15
Fraction of genuine comparisons rejected
Theoretical best
FVC2004DB1 − A
FVC2004DB1 − B
FVC2004DB1 − C
FVC2006DB2A − A
FVC2006DB2A − B
FVC2006DB2A − C
MCYT330DP − A
MCYT330DP − B
MCYT330DP − C
WDSET02 − A
WDSET02 − B
WDSET02 − C
0
0.1
0
1

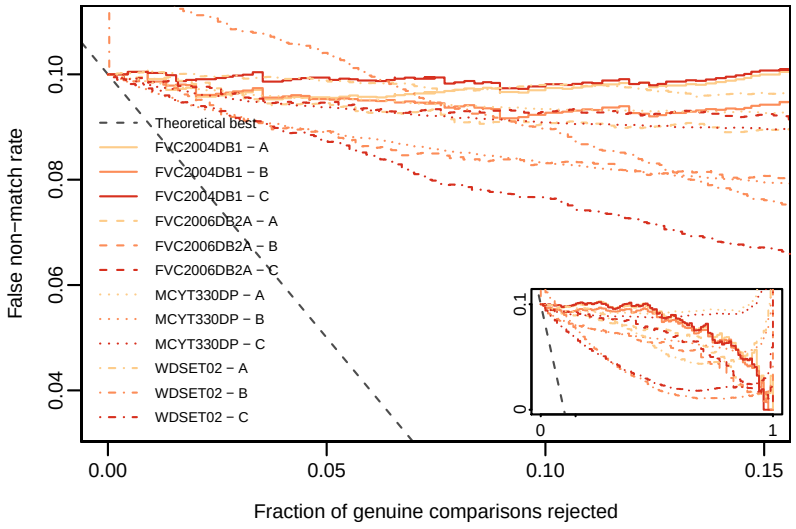
(b) GAB


False non−match rate
0.04
0.06
0.08
0.10
0.00
0.05
0.10
0.15
Fraction of genuine comparisons rejected
Theoretical best
FVC2004DB1 − A
FVC2004DB1 − B
FVC2004DB1 − C
FVC2006DB2A − A
FVC2006DB2A − B
FVC2006DB2A − C
MCYT330DP − A
MCYT330DP − B
MCYT330DP − C
WDSET02 − A
WDSET02 − B
WDSET02 − C
0
0.1
0
1

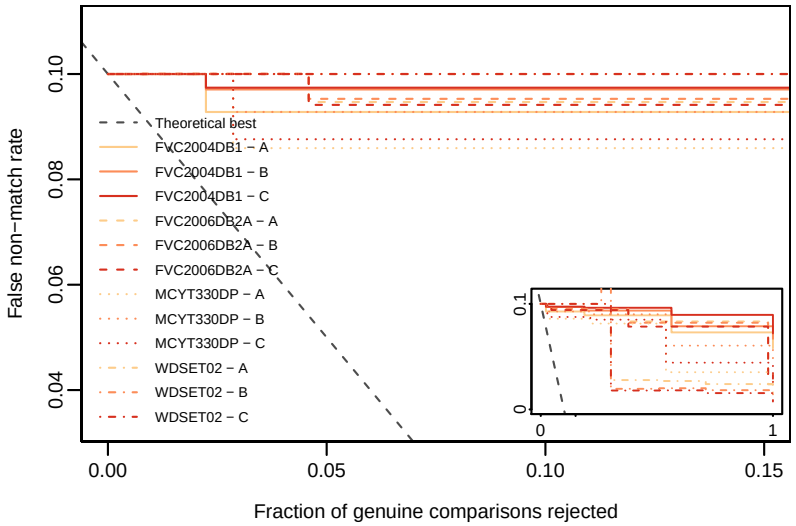
(c) GSH


False non−match rate
0.04
0.06
0.08
0.10
0.00
0.05
0.10
0.15
Fraction of genuine comparisons rejected
Theoretical best
FVC2004DB1 − A
FVC2004DB1 − B
FVC2004DB1 − C
FVC2006DB2A − A
FVC2006DB2A − B
FVC2006DB2A − C
MCYT330DP − A
MCYT330DP − B
MCYT330DP − C
WDSET02 − A
WDSET02 − B
WDSET02 − C
0
0.1
0
1

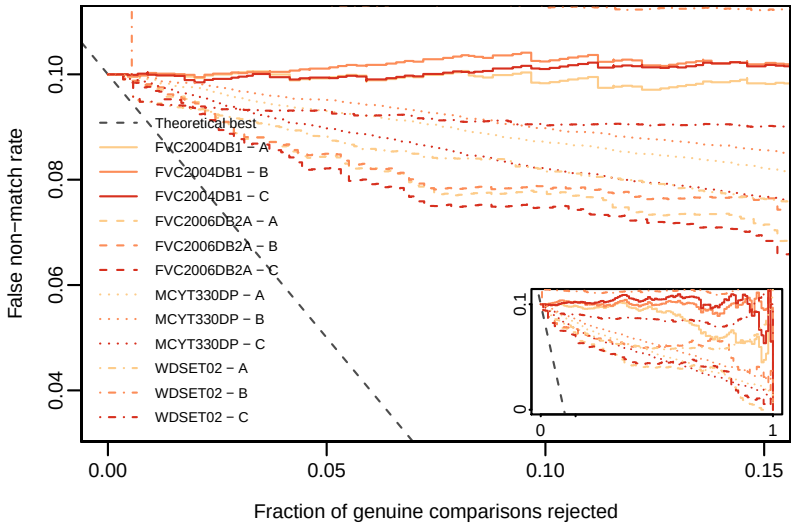
(d) LCS


False non−match rate
0.04
0.06
0.08
0.10
0.00
0.05
0.10
0.15
Fraction of genuine comparisons rejected
Theoretical best
FVC2004DB1 − A
FVC2004DB1 − B
FVC2004DB1 − C
FVC2006DB2A − A
FVC2006DB2A − B
FVC2006DB2A − C
MCYT330DP − A
MCYT330DP − B
MCYT330DP − C
WDSET02 − A
WDSET02 − B
WDSET02 − C
0
0.1
0
1

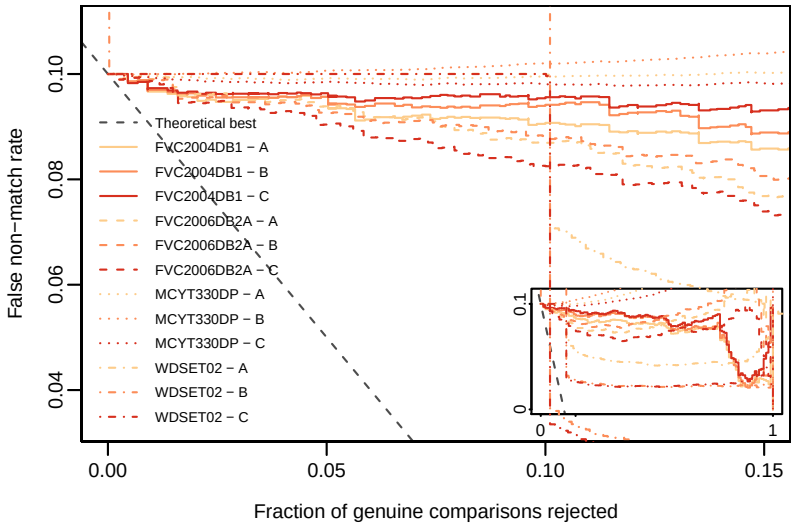
(e) MU


False non−match rate
0.04
0.06
0.08
0.10
Theoretical best
FVC2004DB1 − A
FVC2004DB1 − B
FVC2004DB1 − C
FVC2006DB2A − A
FVC2006DB2A − B
FVC2006DB2A − C
MCYT330DP − A
MCYT330DP − B
MCYT330DP − C
WDSET02 − A
WDSET02 − B
WDSET02 − C
0
0.1
0
1
0.00
0.05
0.10
0.15
Fraction of genuine comparisons rejected

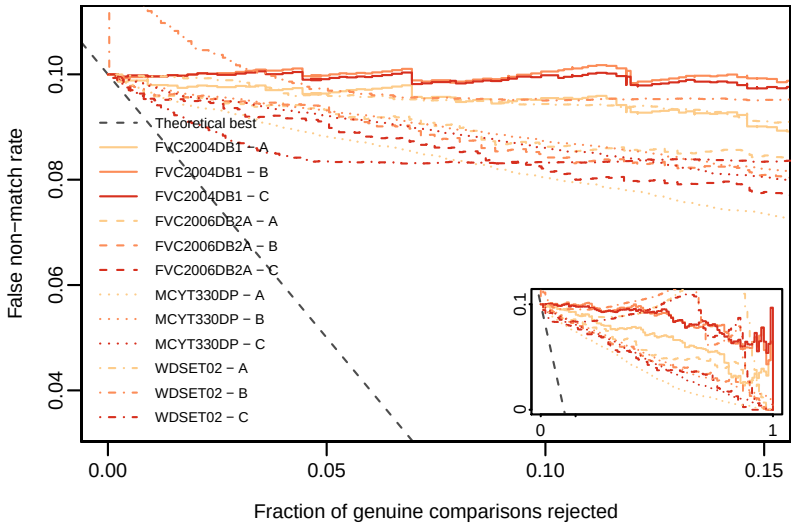
(f) NFIQ


False non−match rate
0.04
0.06
0.08
0.10
Theoretical best
FVC2004DB1 − A
FVC2004DB1 − B
FVC2004DB1 − C
FVC2006DB2A − A
FVC2006DB2A − B
FVC2006DB2A − C
MCYT330DP − A
MCYT330DP − B
MCYT330DP − C
WDSET02 − A
WDSET02 − B
WDSET02 − C
0
0.1
0
1
0.00
0.05
0.10
0.15
Fraction of genuine comparisons rejected

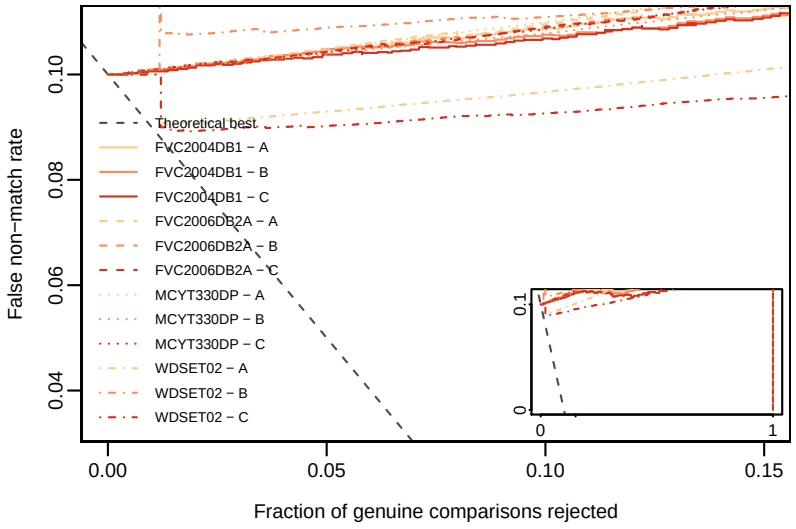
(g) OCL


False non−match rate
0.10
0.08
0.06
0.04
Theoretical best
FVC2004DB1 − A
FVC2004DB1 − B
FVC2004DB1 − C
FVC2006DB2A − A
FVC2006DB2A − B
FVC2006DB2A − C
MCYT330DP − A
MCYT330DP − B
MCYT330DP − C
WDSET02 − A
WDSET02 − B
WDSET02 − C
0.1
0
0
1
0.00
0.05
0.10
0.15
Fraction of genuine comparisons rejected

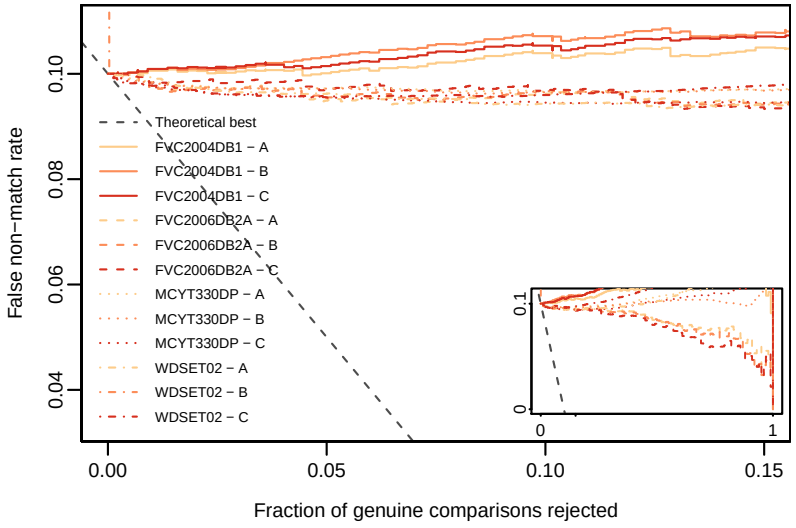
(h) OFL


False non−match rate
0.10
0.08
0.06
0.04
Theoretical best
FVC2004DB1 − A
FVC2004DB1 − B
FVC2004DB1 − C
FVC2006DB2A − A
FVC2006DB2A − B
FVC2006DB2A − C
MCYT330DP − A
MCYT330DP − B
MCYT330DP − C
WDSET02 − A
WDSET02 − B
WDSET02 − C
0.1
0
0
1
0.00
0.05
0.10
0.15
Fraction of genuine comparisons rejected

(i) RPS



(j) RVU



(k) SIG

Figure 4.14: Error-reject curves for each feature on each dataset FVC2004 DB1, FVC2006 DB2A, MCYT330 DP, WDSET02 with each provider A, B, C and $f = 0.1$.

Chapter 5

A Topology Based Approach to Categorization of Fingerprint Images

Information about the paper

Detecting whether a fingerprint is present in an image is of fundamental importance in capture devices and also in maintenance of existing fingerprint based biometric systems. We propose a method to detect whether an object is a fingerprint, and show it applicability in a scenario where iris images and fingerprint images are mixed. The paper is a contribution towards research question $\mathcal{RQ}_1$. This original paper was published in: [2] A. Aabrandt, M. Olsen, and C. Busch. "A topology based approach to categorization of fingerprint images". In: *Biometrics Special Interest Group (BIOSIG), 2012 BIOSIG - Proceedings of the International Conference of the*. 2012, pp. 1–11.

Abstract

This paper discusses the use of betti numbers to characterize fingerprint and iris images. The goal is to automatically separate fingerprint images from non-fingerprint images; where non-fingerprint images of special interest are biometric samples which are not fingerprints. In this regard, an image is viewed as a triangulated point cloud and the topology associated with this construct is summarized using its first betti number - a number that indicates the number of distinct cycles in the triangulation associated to the particular image. This number is then compared against the first betti numbers of "n" prototype images in order to perform classification ("fingerprint" vs "non-fingerprint"). The proposed method is compared against SIVV (a tool provided by NIST). Experimental results on fingerprint and iris databases demonstrate the potential of the scheme.

5.1 Introduction

Recent developments in engineering and other sciences have presented the need for new mathematical tools in order to tackle even the simplest manual tasks like categorization of biometric samples. The human brain can easily distuingish between an iris scan and a fingerprint image; yet doing so for millions of images will undoubtedly take too long to be practically possible. So a need for automatic processing of images by means of computers is present. This area of research requires skills from many different disciplines, including mathematics, computer science and biometrics; where computer science and biometrics are multi-disciplinary sciences.

Classical tools in mathematical analysis, like Fourier analysis, are often the primary tools in applying mathematics to a given problem. Unfortunately requirements of continuity and rigid geometric properties may not always be viable. Mathematicians have recently been more active in refining considerably more advanced tools from mathematics to suit applicational needs. The success of these efforts show promising results in areas like cryptography, statistical mechanics and robotics, to name a few.

The purpose of this article is to present some results indicating the use of modern mathematics to tackle a concrete problem, namely categorization of fingerprint images from a database of mixed biometric samples. Instead of using representations like Fourier transforms, a fingerprint will be represented as a network in 3-dimensional Euclidean space. In topology, a network is called a cell complex which is a generalization of a triangulation. Once a cell complex is constructed, it no longer matters how long the edges are or if the cell complex is deformed into another geometric shape, provided edges cannot be crossed. This means that the analysis will be the same if a given cell complex is equivalent to a triangulation of a cube or if it is equivalent to a triangulation of a sphere. More so the choice of triangulation is not important, e.g. using 18 or 36 triangles to cover a torus would yield the same topology. Topology is sometimes referred to as rubber geometry because it studies properties which persist throughout continuous deformation of objects; such deformations are called homeomorphisms, which are continuous mappings of topological spaces with continuous inverses. Metrics are obviously sensitive to such deformations and therefore a metric is a geometric property. In topology it is not important how far two points are from each other, what matters is how they are connected. In relation to biometric samples,

the main idea is to construct an information network, a cell complex, which accurately captures features unique to the particular type of biometric sample, e.g. iris or fingerprint.

5.1.1 Related work

In 2009 Libert et al. published a detailed validation metric for fingerprint images using Fourier analysis [51]. The peak height of a specially derived power spectrum was found as the most significant classifier based on their analysis on multiple datasets. Moreover applying a windowing function, specifically the blackman window function, the accuracy of the Spectral Image Validation and Verification (SIVV) method was greatly improved, see [51]. The authors illustrated results by thresholding on the peak height. In 2010 an analysis was conducted which is somewhat similar to the usage of the method proposed here, see [137].

The method SIVV, in its current implementation in NIST Biometric Image Software [NIST-NBIS-2012], was used in this work as baseline algorithm to benchmark our own method.

5.2 Fundamentals of topology and homology

One of the most challenging aspects of the method proposed here is that the mathematical machinary for conducting such an analysis requires knowledge of algebraic topology, specifically homology theory. These topics in mathematics are considered modern topics which means they employ advanced algebraic methods developed as late as the second half of the twentieth century. It is therefore recommended that the interested reader seek information from one of the numerous textbooks on the subject. It is recommended that the reader start with the article [138].

The way the method presented here differentiates from the usual methods in applying mathematics is that instead of considering a function which maps some domain to some range, say the discrete Fourier transform; a sequence of functions H_n are to be considered. These functions work on both spaces and maps since they induce a structure on the maps between two spaces. As an example consider the graph in figure 5.1. Denote the triangulation by X.

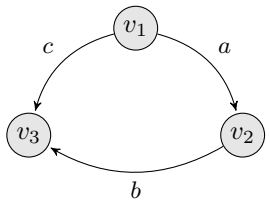


Figure 5.1: Example triangulation with three edges and three vertices.

Let $\mathrm{C}_1(\mathrm{X})$ be a vector space spanned by the edges $\{a, b, c\}$ and $\mathrm{C}_0(\mathrm{X})$ the vector space spanned by the vertices $\{v_1, v_2, v_3\}$. Assume that the coefficients in the two vector spaces are real numbers. Let the map

$$f : \mathrm{C}_1(\mathrm{X}) \to \mathrm{C}_0(\mathrm{X}) \tag{5.1}$$

be one which sends edges of the graph to their vertices. In general homology theories there are analog maps for sending surfaces to edges etc. but in the current context, where we only study graphs, it will suffice to only consider the map f. The homology group $\mathrm{H}_1(\mathrm{X})$ consists of 1-dimensional loops in X. Homology ensures that loops in spaces commute, e.g. it does not matter if one traverses paths abc^{-1} or $b^{-1}a^{-1}c$, i.e.

$$abc^{-1} = b^{-1}a^{-1}c, \tag{5.2}$$

where the inverse notation means traversing a path in the opposite direction of the arrows in the graph. The elements in the kernel $\ker f$ of the map f are called cycles. The idea of homology is to count distinct cycles regardless of starting point or direction taken. In the above example both starting point and direction of the two traversals are different, yet the cycle is the same. For a finite graph like X the first homology group is given by

$$\mathrm{H}_1(\mathrm{X}) = \ker f = \{x \in \mathrm{C}_1(\mathrm{X}) \mid f(x) = 0\}. \tag{5.3}$$

A certain number is associated to each of the groups $\mathrm{H}_n(\mathrm{X})$. It is called the n'th betti number and it is given by the rank of the n'th homology group,

$$\beta_n = \operatorname{rank} \mathrm{H}_n(\mathrm{X}). \tag{5.4}$$

This number is equal to the number of distinct n-dimensional "holes" in the space. For the example above, the betti numbers are $\beta_0 = 1$, $\beta_1 = 1$ and $\beta_n = 0$ for $n \geq 2$.

In relation to biometric samples, the betti number of special interest is β_1. It is important to note that a homology theory ensures that topological distinct spaces produce algebraically distinct homology groups, in particular the set of all betti numbers are not the same for topological distinct spaces.

It should be clear that empirical data is not ideal, i.e. there will be some variation in the observations made. In relation to fingerprints and the methodology proposed here, this means that two biometric samples will most likely produce two distinct topological spaces, even though the samples are acquired from the same source. As will be observed in a later section, the variation in the population will typically be centered around a mean.

Since (connected) graphs are our primary focus of investigation (1-dimensional cell complexes) a much simpler way of calculating homology can be used as opposed to the cases where surfaces and the higher dimensional analogs are taken into account. First of all only one map is of particular interest, namely the one sending edges to its corresponding boundary vertices. This is exactly what we illustrated in the above example, yet the actual map is somewhat unclear at this point. It happens that we can construct this linear map in such a way that the transformation matrix become explicit. One definition of f could be to use the orientation of the edges, e.g. define the boundary of an edge by a linear combination of the vertices of the edge with the sign of the individual terms depending on orientation. Define the starting vertex to have a negative coefficient and the ending vertex to have a positive coefficient. Figure 5.2 illustrates the idea.

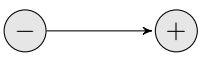


Figure 5.2: Definition of boundary map.

With this definition the edges in X are mapped to their boundary by

$$\begin{aligned} f(a) &= v_2 - v_1, \\ f(b) &= v_3 - v_2, \\ f(c) &= v_3 - v_1. \end{aligned} \tag{5.5}$$

The map f can then be described by a matrix $\mathbf{F}$ given by

$$\mathbf{F} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ -1 & 0 & 1 \end{bmatrix}. \tag{5.6}$$

The kernel $\ker f$ of the map f is the null space of the matrix $\mathbf{F}$. It is a subspace of $\mathrm{C}_1(\mathrm{X})$ and the dimension of this subspace is the betti number β_1 of the graph in figure 5.1, i.e.

$$\beta_1 = \dim \ker f = 1. \tag{5.7}$$

We now state a simple relation without proof, which gives us a very easy way to compute the betti number β_1 of a finite graph, e.g. a 1-dimensional triangulation of a point cloud in 3-dimensional Euclidean space. For a graph it holds that

$$\begin{aligned} \beta_1 &= 1 + \dim \mathrm{C}_1(\mathrm{X}) - \dim \mathrm{C}_0(\mathrm{X}) \\ &= 1 + (\text{number of edges}) - (\text{number of vertices}). \end{aligned} \tag{5.8}$$

Finding the higher betti numbers in higher dimensional cell complexes requires more advanced constructions from homology theory.

5.2.1 Application to biometric samples

Consider each grayscale image as a point cloud in $\mathbb{R}^3$, one point per pixel, i.e. the first two coordinates is the position of the pixel and the third coordinate is the grayscale intensity. A special way of constructing a triangulation of a point cloud called a Witness complex will be used, see [139]. In loose terms, it limits the number of vertices and edges in a triangulation by choosing the vertices according to distances to other vertices in a point cloud. It is a cell complex where a small number of points from the point cloud are choosen as vertices and edges are then constructed iteratively according to a set of criterias as described in [139].

Once the triangulation is constructed, the betti numbers are calculated using homology and the metric will be constructed. For simplicity a weighted Euclidean metric is chosen. Initially only two types of images are used, e.g. fingerprint or non-fingerprint. The non-fingerprint database will be either face samples or iris scan samples.

Assume that a vector is given consisting of n numbers acquired from a sample of fingerprint images, each the 1st betti number of a fingerprint; one for each image in the sample. Denote this vector by v and let the variance of v be denoted by σ^2. Given an image, it is possible to calculate the corresponding 1st betti number β_1. Let x denote the vector where each element is equal to this betti number for a given image.

Then for a fixed training set v, a metric can be constructed by

$$D(\beta_1) = D(\beta_1; x, v) = \frac{1}{\sigma} \left(\sum_{i=1}^{n} (x_i - v_i)^2 \right)^{1/2}. \tag{5.9}$$

For the topological method presented here, the metric in equation 5.9 is used. For every image and corresponding 1st betti number β_1, calculate two numbers: 1) the distance $D_f(\beta_1)$ to a sample set of fingerprints. 2) the distance $D_g(\beta_1)$ to the sample set of non-fingerprints.

Now in order to be able to control the error rates of the overall analysis, a few simple functions are introduced. The two functions are

$$F_1(s, \beta_1) = sD_g(\beta_1) \quad \text{and} \quad F_2(s, \beta_1) = (1 - s)D_f(\beta_1)\ , \quad s \in [0; 1]. \tag{5.10}$$

Given a value of s, let $F_1 \geq F_2$ determine a false match increase by 1. Otherwise a false non match increase of 1. The proportion of the number of false matches (resp. false non matches) with respect to the number of non-fingerprints (resp. fingerprints) is the corresponding error rate. The value of s can be determined experimentally and thus depends on the choice of training sets. The variable s is used to calculate the ROC curve in the section presenting the results.

This method shall be thought of as a way to coarsely characterize a biometric sample or rather a point cloud.

5.3 Results

The results presented here will highlight both strengths and weaknesses of the method derived from topology which will be denoted TOP and the already known method SIVV. Some of the most interesting results are found when considering mixed biometric databases, i.e. databases where both fingerprints, iris and other biometric samples are contained. In order to see how the constructed metrics work on such databases when categorizing joint data sets with various fingerprint databases a DET curve is determined and the equal error rate EER is calculated.

Portions of the research in this article use the database CASIA-FingerprintV5 [140] and CASIA-IrisV4 [141] collected by the Chinese Academy of Sciences' Institute of Automation (CASIA). A fingerprint database acquired by Association BioSecure is used and referred to here simply as Biosecure [142]. Also used is the MCYT database involving fingerprints which was collected in a project conducted in 2003 [135].

All calculations are carried out on a computer running Linux (kernel 3.3.2-1) with a Quad-core (3.10 GHz each) Intel Xeon E3-1225 processor and 4 GB memory. The software used for computation are Javaplex version 4.0 and the SIVVUtility package which is part of NIST Biometric Image Software version 4.0.1. All algorithms were run in parallel on three of the cores due to the large number of images needed to be processed. User interfaces have been rewritten in order to control output formats for both software packages.

5.3.1 General observations

There are a number of general observations to be made. The classifier used in this section is the 1st betti number. Considering the betti number of fingerprint images yield a bell shaped curve when approximating to the histogram in the figure below. A similar yet skewed result is found for non-fingerprints. The databases used to produce figure 5.3 are the CASIA-FingerprintV5 and CASIA-IrisV4 databases.

These bell curve shapes, although one is skewed, may be a product of the method by which the betti numbers are calculated.

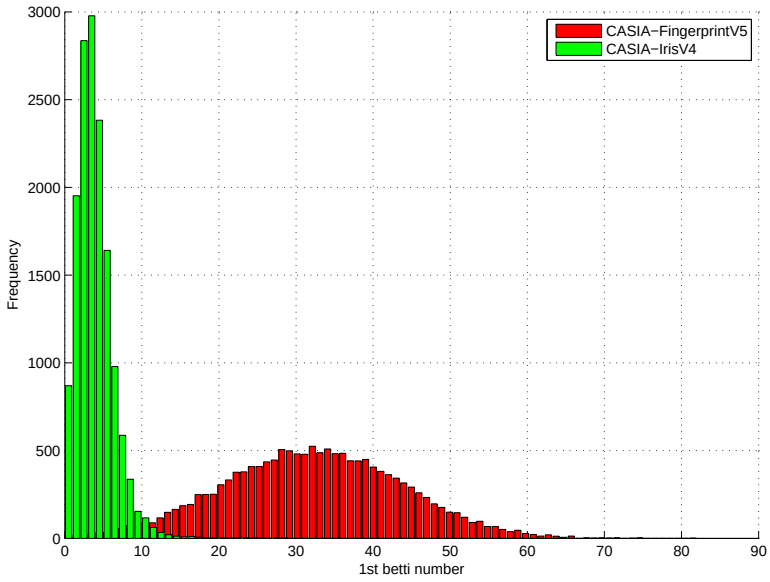


Figure 5.3: Histogram of the 1st betti numbers of CASIA fingerprint and iris images

5.3.2 Mixed biometric (iris vs. fingerprint) database

In this section an analysis of the method is done using the iris database CASIA-IrisV4. The fingerprint databases are analyzed separately and the results are highlighted in a table in the end of this section. To illustrate the difference in the betti numbers a couple of examples are shown. The fingerprint is taken from the public database FVC2000Db2, see [143], and the iris is the one from one of the authors. Note that this is meant merely as an example. The iris images used in the experiment are non-segmented ocular images.

(a) 1st betti number 57

(b) 1st betti number 2

Figure 5.4: Examples of 1st betti numbers of some images via witness complexes.

The examples in figure 5.4 indicate the difference of the betti numbers for the two types of biometric samples.

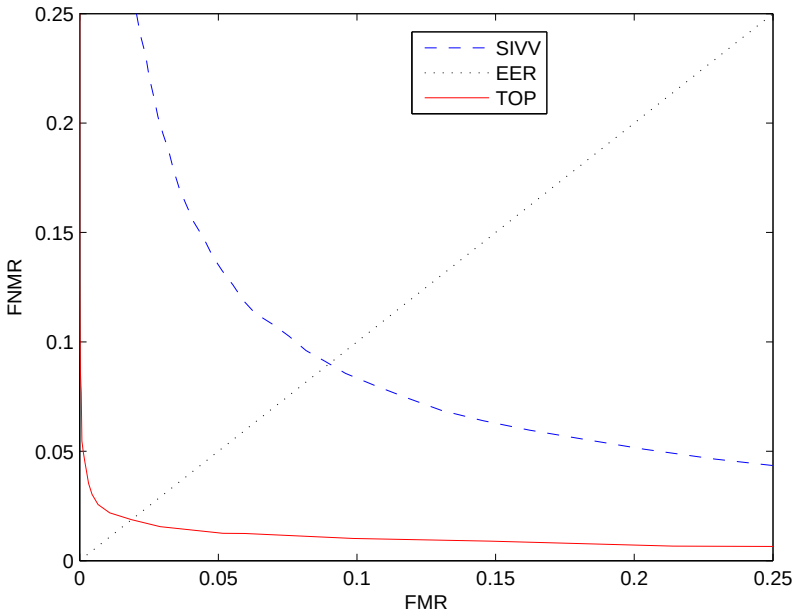


Figure 5.5: fingerprint vs iris database: EER(TOP) = 2.7%, EER(SIVV)=9.0%.

Database	SIVV	TOP	Summary
MCYT dp	12.9%	0.2%	Largest observed difference between methods.
Biosecure	6.2%	1.9%	Smallest absolute difference.
CASIA	9.0%	2.7%	Medium difference between methods.

Table 5.1: EER for SIVV and TOP on varying fingerprint datasets vs. iris images.

Comparing with the SIVV method it becomes clear that the topological approach performs better for fingerprints vs. iris databases (see fig. 5.5). The most interesting aspect of considering mixed biometric databases is that when Fourier analysis seem to produce misleading features; coarse topological features tend to correctly partition the database into biometric types.

Table 5.1 illustrates how the equal error rate varies between fingerprint databases for the two methods. The results clearly suggests that using the TOP-method will produce much better results when categorizing fingerprints and iris scans. The most extreme case is the MCYT dp database. It is the database with best performance using the topological approach and the worst performance when using SIVV. An interesting observation has been made; the choice of fingerprint reference database, also called training set, does not

		Training set		
		MCYT dp	Biosecure	CASIA
Database	MCYT dp	0.2%	0.2%	0.2%
	Biosecure	1.9%	1.9%	1.9%
	CASIA	2.7%	2.7%	2.7%

Table 5.2: EER for SIVV and TOP on varying fingerprint datasets vs. iris images in different training set configurations.

Database	SIVV	TOP	Summary
MCYT dp	5.0%	1.0%	Largest observed difference between methods.
Biosecure	2.5%	3.0%	Medium difference between methods.
CASIA	3.7%	4.1%	Smallest absolute difference.

Table 5.3: EER for SIVV and TOP on varying fingerprint datasets vs. face images.

produce a significant change in the results. So one may choose a subset from either MCYT dp, Biosecure or CASIA-fingerprintV5 as a sample reference to a fingerprint database. Table 5.2 illustrates the observed equal error rates when utilizing various training sets.

At least from a statistical point of view it seems that the TOP-method is invariant of fingerprint reference database chosen. That is the equal error rate remains the same under different training sets. The table only illustrates this observation for the combinations of the three databases, MCYT dp, Biosecure and CASIA. One should be very careful before making any inference based on these preliminary observations. Note that the non-fingerprint image training set is not varied in this experiment.

5.3.3 Mixed biometric (face vs. fingerprint) database

A similar analysis as the one from the previous section will be carried out. The non-fingerprints in this section consists of facial images from the CASIA-FaceV5 [144] database.

In table 5.3 it is observed that SIVV performs slightly better for both CASIA fingerprints and Biosecure databases. A significant difference is observed when comparing with the MCYT database. In this particular case the method TOP performs better than SIVV. The results shown here simply states that the TOP method does not yield better results than SIVV for all biometric samples when categorizing. In complete analogy with the study of iris vs. fingerprint, the equal error rates can be seen to be robust with respect to the choice of fingerprint training set.

For the databases used in this study and referring to tables 5.2 and 5.4, it can be concluded that the variances imposed by changing the training set for the fingerprint database is not significant.

		Training set		
		MCYT dp	Biosecure	CASIA
Database	MCYT dp	1.0%	1.0%	1.0%
	Biosecure	3.1%	3.0%	3.1%
	CASIA	4.2%	4.2%	4.1%

Table 5.4: EER for SIVV and TOP on varying fingerprint datasets vs. face images in different training set configurations.

5.4 Discussion and conclusions

It has been shown that, given an image, homology of a specially constructed triangulation of the image, viewed as a point cloud will enable the separation of fingerprint images from other biometric samples, in particular when categorizing iris scans and fingerprints. For the iris scans the topological approach seems to perform better than the SIVV method. When categorizing fingerprints and face samples the method SIVV seem to perform better, if only slightly, for the databases Biosecure and CASIA-FingerprintV5. For the database MCYT the method TOP seem to perform significantly better. The result indicate that the acquisition of samples in the MCYT database or some form of processing of the images may cause this significant decrease in equal error rate. Moreover it is observed that the opposite holds for the SIVV method, where a significant increase in equal error rate is observed for the MCYT database compared against both iris and face databases. Our method seem to be invariant with respect to the choice of fingerprint training set.

The method of constructing the triangulation from a biometric sample, in this case a fingerprint image, can be improved. The general construction could be to consider particularly important landmarks in the fingerprint as vertices in the triangulation. Edges should then be assigned between two vertices in such a way that certain information about the sample specific to the landmarks yield a denser network, e.g. more information is retained in the triangulation, which in turn should indicate a higher quality. This latter part is an extensive work in progress, requiring many simulations to be run and much statistics to be collected. It is possible that the method TOP can be extended to categorize the types of fingerprints as well, e.g. identify whorl or arch structures.

Another possible extension of this study would be to study the same source under varying conditions, e.g. time, pressure variation and humidity conditions. This would give us an indication of the ability of the method to identify individual samples as being from captured from the same source. The relation to well-known image features and the importance of these are still unclear at this point, this being yet another interesting topic to study next.

Chapter 6

Finger Image Quality Based on Singular Point Localization

Information about the paper

One aspect of fingerprint image quality is the positioning of the finger on the sensor. Here we investigate the influence of the position of the fingerprint core point on the comparison score, and propose a quality algorithm based on this position. The paper is a contribution towards research question $\mathcal{RQ}_2$. This original paper was published in: [5] J. Wang, M. A. Olsen, and C. Busch. *Finger image quality based on singular point localization*. 2014. DOI: 10.1117/12.2050145. URL: http://dx.doi.org/10.1117/12.2050145.

Abstract

Singular points are important global features of fingerprints and singular point localization is a crucial step in biometric recognition. Moreover the presence and position of the core point in a captured fingerprint sample can reflect whether the finger is placed properly on the sensor. Therefore, the displacement given by detected core points is investigated. We propose pattern-based filters to eliminate the false detection given by state of the art approaches. The experimental results show improvement using different databases. Based on the improved singular point localization algorithm, we explore and analyze the importance of singular points on biometric accuracy. The experiment is based on large scale databases and conducted by relating the measured quality of a fingerprint sample, given by the positions of core points, to the biometric performance. The experimental results show the positions of core points do have influence on the comparison algorithms, but are not as relevant as other benchmarked quality metrics.

6.1 Introduction

Due to the excellent distinctiveness, persistence and ease of collection, fingerprints are the most widely observed biometric characteristic for identification and authentication. In practical biometric systems, the performance depends heavily on the quality of the input samples. That is, a poor quality sample might have negative influence on the biometric system. Hence, quality measurement of fingerprint samples is a crucial step to assure the performance by rejecting poor samples specifically in the enrollment procedure.

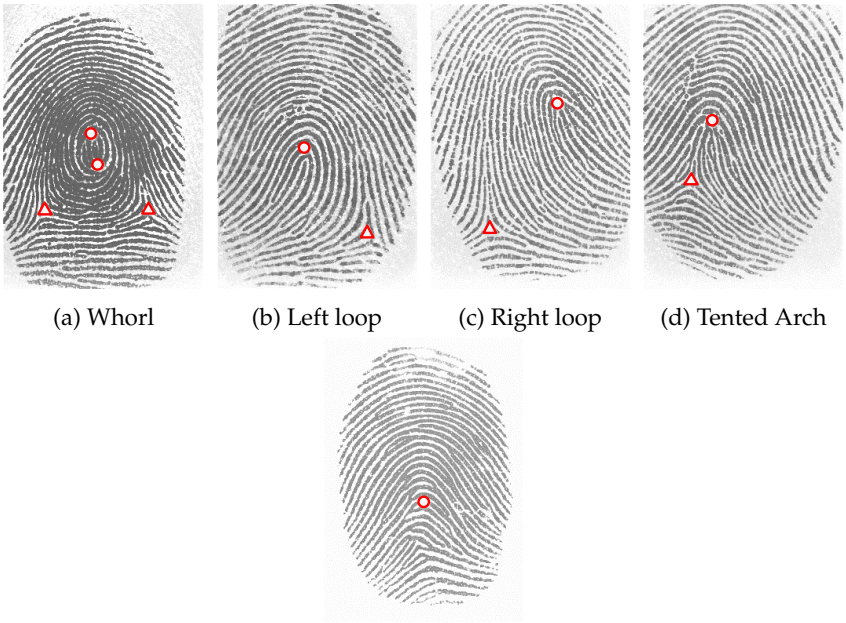
A Singular Point (SP) is the important global feature of a fingerprint pattern. Typically SPs, known as core and delta, are located in the regions which possess the higher ridgeline curvature. The SPs are identified in Fig. 6.1 with regard to the Henry classification system [49]. A core point is defined as the topmost point on the innermost recurving ridgeline of a fingerprint [145] by ISO/IEC[1], thus there is also one core point existed in each arch finger.

In biometric systems, SPs are important in fingerprint classification. Liu et al. [147] proposed an indexing fingerprint based on location, direction, estimation and correlation of SPs, and Wei [148] used the delta direction and the SPs to partition fingerprint classes[149]. Besides, SPs are also significant in fingerprint comparison, which are used as references to line up two fingerprints in one-to-one comparison[150]. Hence recent literature states that SPs plays an important role in fingerprint recognition[151, 152].

The importance of SPs inspires us that a concise description of their properties can be used as metrics to assess fingerprint image quality. Core points are approximately located in the center of a fingerprint pattern so that the positions of cores in a captured fingerprint sample represents the displacement, i.e., whether the finger is placed properly on the sensor. In a mated comparison trial, a low comparison score[2] might be given if one of the fingerprints performs a large displacement. A finger displacement in either the reference or the probe sample or a shift of both in opposite directions will result in a reduced overlapping area to seek minutiae information and thus cause the low comparison score, even though both samples might possess good quality.

[1] International Organization for Standardization/International Electrotechnical Commission

[2] Comparison score is directly proportional to similarity in this paper.



(a) Whorl (b) Left loop (c) Right loop (d) Tented Arch

(e) Arch

Figure 6.1: Identified SPs within Henry classification system, where cores and deltas are marked by circles and triangles respectively. Fingerprints are taken from FVC2000Db2 [146].

However, the mentioned importance of SPs was previously derived by speculatively analyzing the functions in biometric systems without systematic analysis. Furthermore, there is also the question of what is the degree of the importance when benchmarked with other existing quality metrics. The contribution of this paper is the following: develop an algorithm that can increase state of the art SP localization rate analyze the degree of importance for processing SP in biometric systems.

The remainder of this paper is organized as follows: Sec. 6.2 describes the existing techniques of SP localization and quality metrics used in this work; in Sec. 6.3, a few methods are proposed both to localize SPs and explore the importance of the SPs; Sec. 6.4 presents the experimental setup and results; Sec. 6.5 summarizes our contribution in a conclusion.

6.2 State of the Art

6.2.1 Singular Point Localization

In order to localize SPs in a fingerprint, the orientation field (OF) should be estimated as a foundation. Various techniques have been proposed in the literature such as filter bank-based[153] and model-based approaches[154]. However, it is reported that these methods do not provide as much accurate results as the gradient-based methods[155].

Bazen and Gerez proposed a method to estimated the OF on the pixel level[150]. The pixel-wise gradient vectors $[G_x\ G_y]^T$ whose phase angle denoted the direction of the maximum intensity change are given by:

$$\begin{bmatrix} G_x \\ G_y \end{bmatrix} = \nabla I(x,y) \tag{6.1}$$

$$= \begin{bmatrix} \frac{\partial I(x,y)}{\partial x} \\ \frac{\partial I(x,y)}{\partial y} \end{bmatrix} \tag{6.2}$$

where $I(x,y)$ is the intensity at pixel (x,y). The ridge orientation is orthogonal to the gradient phase angle at each pixel, however, the gradient vectors cannot directly be used to compute the OF because opposite gradient vectors will be canceled out with each other although they represent the same orientations. A feasible representation is proposed by doubling the angles and squaring the length of the gradients so that opposite gradient vectors will point in the same direction[156]. While doubling the angle of and squaring the length of gradients we can compute:

$$\begin{bmatrix} G_{s,x} \\ G_{s,y} \end{bmatrix} = \begin{bmatrix} G_\rho^2 cos2G_\psi \\ G_\rho^2 sin2G_\psi \end{bmatrix} \tag{6.3}$$

$$= \begin{bmatrix} G_\rho^2(cos^2 G_\phi - sin^2 G_\psi) \\ G_\rho^2(2sinG_\phi cosG_\psi) \end{bmatrix} \tag{6.4}$$

$$= \begin{bmatrix} G_x^2 - G_y^2 \\ 2G_x G_y \end{bmatrix} \tag{6.5}$$

The estimated orientation at the fine pixel level will be quite noise sensitive, thus the average squared gradient vector $[\overline{G_{s,x}}\ \overline{G_{s,y}}]^T$ of each element should rather be computed according to squared gradient vectors within the surrounding window W:

$$\begin{bmatrix} \overline{G_{s,x}} \\ \overline{G_{s,y}} \end{bmatrix} = \begin{bmatrix} \sum\limits_W G_{s,x} \\ \sum\limits_W G_{s,y} \end{bmatrix} \tag{6.6}$$

$$= \begin{bmatrix} \sum\limits_W G_x^2 - G_y^2 \\ \sum\limits_W 2G_x G_y \end{bmatrix} \tag{6.7}$$

The average gradient direction Φ is given by:

$$\Phi = \begin{cases} \frac{1}{2}tan^{-1}(\overline{G_{s,y}}/\overline{G_{s,x}}), & \overline{G_{s,x}} \geq 0 \\ \frac{1}{2}tan^{-1}(\overline{G_{s,y}}/\overline{G_{s,x}}) + \pi, & \overline{G_{s,x}} < 0 \wedge \overline{G_{s,y}} \geq 0 \\ \frac{1}{2}tan^{-1}(\overline{G_{s,y}}/\overline{G_{s,x}}) - \pi, & \overline{G_{s,x}} < 0 \wedge \overline{G_{s,y}} < 0 \end{cases} \tag{6.8}$$

Then the average ridge-valley direction θ of the fingerprint represented in the OF is given by:

$$\theta = \begin{cases} \Phi + \frac{1}{2}\pi, & \Phi \leq 0 \\ \Phi - \frac{1}{2}\pi, & \Phi > 0 \end{cases} \tag{6.9}$$

Based on the estimated OF, SPs can be localized with many techniques. The Poincare index is an natural and practical method by computing the rotation in each closed curve [157]. However, it is not resistant to noise and the method performs only with poor computational efficiency. Bazen and Gerez [150] proposed the application of the Green's theorem, which improved the Poincare index method by applying a discrete version of Green's theorem and reported better performance regarding the correct detection rate.

The algorithm doubles the OF resulting in an interval $(-\pi, \pi]$, and then computes gradients of the doubled orientation field:

$$\begin{bmatrix} Jx(x,y) \\ Jy(x,y) \end{bmatrix} = \nabla 2\theta(x,y) \tag{6.10}$$

$$= \begin{bmatrix} \frac{\partial 2\theta(x,y)}{\partial x} \\ \frac{\partial 2\theta(x,y)}{\partial y} \end{bmatrix} \tag{6.11}$$

The discrete Green's Theorem can be applied by summing the gradients of the doubled OF $[Jx\ Jy]^T$ over the contour, thus the Poincare indexes change to 2π, -2π, and 0 for, respectively, a core, a delta, and none of these [150]:

$$Index = \sum_{\Delta x, \Delta y\ along\ \partial A} (J_x \cdot \Delta x + J_y \cdot \Delta y) \tag{6.12}$$

$$= \sum_{A} rot[J_x\ Jy]^T \tag{6.13}$$

$$= \sum_{A} (\frac{\partial \omega_y}{\partial x} - \frac{\partial \omega_x}{\partial y}) \tag{6.14}$$

Furthermore, orientations of detected SPs can be estimated by comparing the doubled OF around SPs with precomputed reference models of core and delta respectively:

$$DOF_{core,ref} = \frac{(y,-x)}{\sqrt{x^2+y^2}},\ DOF_{delta,ref} \tag{6.15}$$

$$= \frac{(-y,-x)}{\sqrt{x^2+y^2}} \tag{6.16}$$

The orientation of a core or delta is given by taking the element-by-element product of the estimated doubled OF $DOF_{core,obs}$ and the complex conjugated of the reference model $DOF_{core,ref}$ or $DOF_{delta,ref}$ respectively:

$$\widehat{\alpha}_C = \angle \frac{1}{N} \sum_{x,y} DOF^*_{core,ref} \cdot DOF_{core,obs},\ \widehat{\alpha}_D \tag{6.17}$$

$$= \frac{1}{3} \angle \frac{1}{N} \sum_{x,y} DOF^*_{delta,ref} \cdot DOF_{delta,obs} \tag{6.18}$$

As a result, the orientation of a core is the direction where the core points out to, and for delta it is the direction where one of the three vertexes points to the top of the image.

6.2.2 Quality Measurement Algorithms - State of the Art

The ISO/IEC Technical Report 29794-4 on fingerprint sample quality[99] is the most recent overview of recommended fingerprint quality metrics. Amongst others the following metrics are defined in ISO/IEC 29794-4, which shall serve as baseline in the benchmark with a SP-based sample quality metric in this work:

Orientation Certainty Level (OCL) measures the strength of the energy concentration along the dominant ridge flow orientation within a block by means of computing the block-wise gradient.

Frequency Domain Analysis (FDA) performs a block-wise Discrete Fourier Transform and measures the energy concentration in a frequency band. Dominance in the very low frequencies indicate low quality.

Radial Power Spectrum (PS) is a measure of maximal signal power in a defined frequency band of the global Radial Fourier spectrum. Ridges can be locally approximated by means of a single sine wave, hence high energy concentration in a narrow frequency band corresponds to consistent ridge structures.

Furthermore, NIST[3] Finger Image Quality (NFIQ) as it was released in 2004 is a neural network pattern classifier, which is trained to classify fingerprint quality into 5 values[79] by evaluating the block-based features and the quality of minutiae.

Gabor filter responses (Gabor) measures the responses of a fingerprint by applying Gabor filters with four predefined orientations. For the fingerprint areas with clear ridge pattern, the Gabor responses of one or a few orientations should have larger values than other orientations. Whereas for the background or the poor ridge clarity fingerprint areas, the Gabor responses of four orientations will be low and constant[1].

6.3 Proposed Methods

6.3.1 Semantic Filtering of detected Singular Points

While the method that was reviewed in Section 6.2 can be considered as one of the strongest methods available to date there are known deficiencies[158]. There are numerous spurious SPs being detected in the presented algorithm so that it is an essential post-processing task to eliminate those false detection. We propose a series of sequentially applied semantic filters. The filters are designed under the assumption that for the benign case of a correctly placed finger the core is in the center of the fingerprint area and further that the fingerprint area provides higher quality if it is more approaching to the center. Therefore, if we detect multiple SPs and some subset does not satisfy the semantic conditions outlined in this section, then those SP-candidates, which are near to the boundary of the foreground area, will be removed. The following five semantic conditions are defined:

- **C2D2**: There shall be at most two cores and two deltas.
- **CCDIS**: If there are two cores, then the spatial distance between them should be small.
- **CCORIDIFF**: If there are two cores, then they point out to approximately opposite directions.

[3]National Institute of Standards and Technology

- **DDORIDIFF**: If there are two deltas, then they possess similar orientations.
- **DCDANGLE**: If there are two deltas and at least one core, then the angle $delta1 - core - delta2$ lies in a certain interval. If there are two cores, the midpoint of the two cores is selected as depicted in Fig. 6.2.

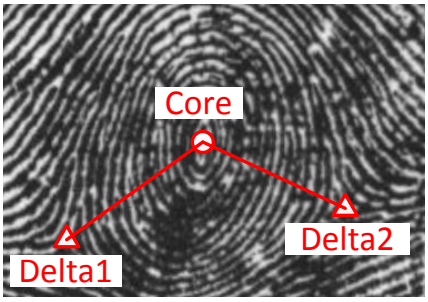


Figure 6.2: Angle $delta1 - core - delta2$ of a image, the midpoint is used as the positions of cores if there are more than one core points.

With the filters as defined we are confident to reduce the error rates in SP detection significantly. Candidates that do not satisfy the filter rules are removed from the candidate list such that the remaining SP can be considered to be reliable. We will show this in Section 6.4.

6.3.2 SP localization-based Quality Measurement Algorithms

The methodology we used to analyze the importance of the SPs is to relate the detected SPs to biometric performance. Precisely the positions of detected SPs and a metric derived from the SP-position with respect to the center of the fingerprint area are proposed as quality metrics.

We propose to use SPs as control points to determine if a finger is placed on the center of the sensor. Intuitively smaller displacement is likely to provide better quality because a non-displaced fingerprint sample is a more complete representation of the fingerprint pattern. Contrarily, if for the same finger instance, the capture subject shifted in two subsequent capture attempts the finger respectively to the left and right, then then two generated captured samples might seem like a pair of non-mated (i.e. imposter) fingerprints.

Core points are located on the center of fingers, whereas delta points does not, so core points are more suitable to indicate wheter a fingerprint has a large displacement. Furthermore in most cases at least one core exists in a fingerprint, thus only information that is derived from the position of the core is proposed as quality metrics. Hence, the estimated quality scores are expected to be decreasing along with an increase in the displacement. If there are two detected core points on a fingerprint, the midpoint of both is used as the core point to measure the displacement.

The segmentation of the fingerprint area in the sample is conducted first, in order to obtain the foreground and then identify the center of the foreground. There is no need to consider the intensity at each pixel because we are not interested in the center of gravity of

the fingerprint. The coordinates of the center (C_x, C_y) of N foreground pixels is computed as:

$$C_x = \frac{1}{N} \sum_{i=0}^{N-1} M_x \tag{6.19}$$

$$C_y = \frac{1}{N} \sum_{i=0}^{N-1} M_y \tag{6.20}$$

where $[M_x, M_y]$ is the index of foreground mask along the horizontal and vertical direction respectively. Then the SP localization-based quality metrics are proposed as:

- **DIS**: the quality score is defined by the distance from the center O of the foreground to the core position C, as depicted in Fig 6.3a;
- **HDIS**: the quality score is defined by the *horizontal* distance from the center O of the foreground to the core position C, as depicted in Fig 6.3b;
- **VDIS**: the quality score is given by the *vertical* distance from the center O of the foreground to the core position C, as depicted in Fig 6.3c.

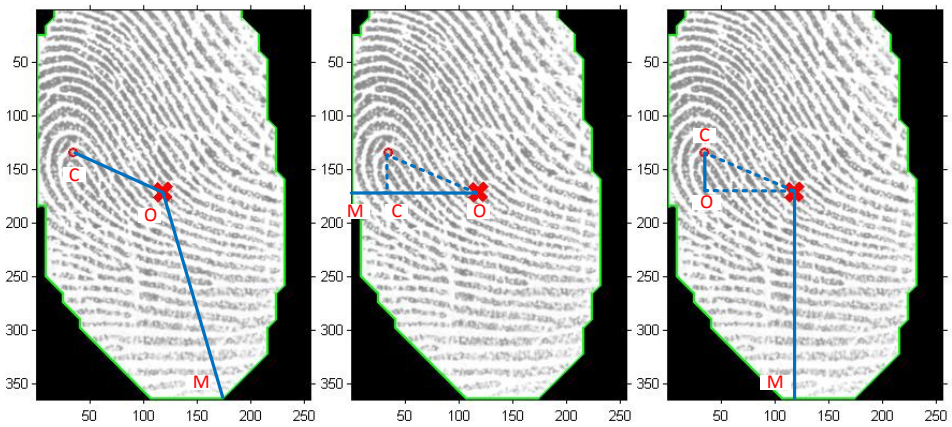
In addition, different fingerprints might likely have different foreground sizes, even though the fingerprint sample is captured from the same fingerprint instance but in different acquisitions. Therefore we consider the size of the foreground as a factor that may have influence and thus additionally consider relative metrics. The respective ratio versions DISR, HDISR, VDISR of the metrics defined above DIS, HDIS, VDIS are computed by dividing the corresponding values by the maximum distance, maximum horizontal distance and maximum vertical distance from center O to the boundary point M of the foreground respectively, as depicted in Fig. 6.3.

6.4 Experiments

6.4.1 Singular Point Localization

For this experiment firstly we use the dataset FVC2002Db2, which contains 880 plain fingerprint samples captured by an optical sensor [143]. The SPs for this dataset were identified manually according to the definition given by ISO/IEC 19794-1:2011[145]. Moreover the ground truth data set SD14-BKA-GTD was used, which is a subset of NIST SD14[159] fingerprint images. The dataset consists of 486 inked rolled and plain fingerprints samples. The foreground, the minutiae , and SPs were marked-up by at least three dactyloscopic experts from the German Federal Criminal Police Office (BKA)[160].

Bazen reported in Ref. [150] that SP localization does not require a high resolution of the fingerprint samples. Thus the samples, which were originally of a spatial resolution of 296×560 pixels (FVC) and 832×768 pixels (SD14), were in our experiment downsized to a quarter size initially. In order to further reduce the interference of noises on the background, segmentation is applied for FVC2002Db2, and ground-truth foregrounds of SK14-BKA-GTD are used. Then a Gaussian filter with $\sigma = 5$ is applied to the foreground.



(a) DIS/DISR (b) HDIS/HDISR (c) VDIS/DVISR

Figure 6.3: Examples of SP localization-based quality metrics on a fingerprint foreground, OC represents the score for DIS, HDIS, VDIS, and OC/OM represents for DISR, HDISR, VDISR.

The experimental results of the conventional Poincare index method, the Bazen method that was based on Green's theorem and the Bazen method with post-processing with the proposed semantic filters are conducted for both datasets FVC2000Db2 and SD14-BKA-GTD and are shown in Tab. 6.1 and Tab. 6.2 respectively.

SP type	Error type	PI	Bazen	Bazen & Filters
Core	False	15.9	6.9	3.3
	Missed	4.8	3.0	3.6
Delta	False	18.5	3.5	3.5
	Missed	2.3	2.3	2.3
SP	False	24.6	8.6	5.9
	Missed	7.1	5.3	5.9

Table 6.1: Results of SP localization for FVC2002Db2 (in %).

SP type	Error type	PI	Bazen	Bazen & Filters
Core	False	58.9	31.3	9.9
	Missed	8.4	9.0	10.5
Delta	False	62.4	23.5	11.3
	Missed	4.7	7.2	9.3
SP	False	66.7	38.5	17.3
	Missed	11.5	14.8	15.0

Table 6.2: Results of SP localization for SD14-BKA-GTD (in %).

Generally the conventional Poincare index method possesses the worst performance,

with relatively low missed rate, but rather high false detection rate. In contrast, the Bazen approach is more robust against noise, resulting in relatively fewer false detections. By filtering out the spurious SPs after applying our proposed semantic filters, the SPs are validated so that the false detection is further decreased. Meanwhile the missed detection rate is slightly higher due to fact that the filter in some cases eliminate a candidate by mistake. As a result, the error rate for SPs decreases, which shows that the proposed semantic filters improves the correct SP localization rate.

6.4.2 Finger Image Quality

Initially the large-scale database CASIAFPV5[140] is used in this experiment, which consists of 20000 fingerprint images of 500 subjects. Each subject provides 8 fingers and each of them is captured for 5 acquisitions by an optical fingerprint sensor with 328×356 pixels. Furthermore, a subset of Ministerio de Ciencia y Tecnología of Science and Technology (MCYT) 330 biomodal Fingerprint subcorpus[135], MCYT330PB is selected. Fingerprints from 330 individuals have been acquired by a capacitive sensor, with a resolution of 500 dpi. Each subject has provided 10 fingers and 12 acquisitions resulting 39600 fingerprints all at 300×300 pixels. In addition, the other subset MCYT330DP is also used, which has the same file structure and is collected by an optical sensor with 296×400 pixels.

Besides, three commercial comparators are used for research purposes which are provided by the industrial vendors with anonymized names A, B, C. For each comparator, comparison scores are computed for each pair of mated fingerprints and thus there are $500 \times 8 \times 5 \times 4 = 80000$ genuine comparisons in CASIAFPV5, and $330 \times 10 \times 12 \times 11 = 435600$ in MCYT330PB and MCYT330DP, respectively.

The experiment consists of two parts: the importance of SPs is explored by analyzing how efficiently SPs can be used to reject finger samples with low comparison scores; what is the influence on utility-based quality scores if the position of core is away from the center of foreground.

6.4.3 Error versus reject curves

The effectiveness of proposed and benchmarked quality metrics are measured by Error versus Reject Curves (ERC)[112], which is a method to evaluate how efficiently poor genuine comparisons will be rejected, which correlate to low comparison scores. Hence, the ideal case is that the rejected comparisons identified by a selected metric are exactly those comparisons with lowest genuine scores. The efficiency of a given metric is evaluated by a curve of False Non Match Rate (FNMR) variation along with the increase of rejections. If the computed quality scores are perfectly correlated with the genuine quality score, then the curve should decrease quickly with the fraction rejected and approach the ideal case.

In this experiment, 10% is set as false non-match rate, so these finger samples are target to be rejected by the quality metrics under investigation. In an operational practice it is not realistic to reject many samples , so at most a fraction of 35% of the samples can be rejected. There might be a variety of quality metrics plotted in the same ERC graph, so that it might get difficult if the ERCs show multiple intersections. In order to improve the readability, the green area under the curve in Fig. 6.4 is used as a single number to indicate the prediction performance. Therefore, a sample quality metric has a better performance if the area is in proximity of 0.

Figure 6.4: Example area of ERC. For the ease of visualization, the area under the curve is enlarged as 100 times. The area in yellow marks the area of the ideal case, and green marks the area under the QMA curve excluding the ideal area.

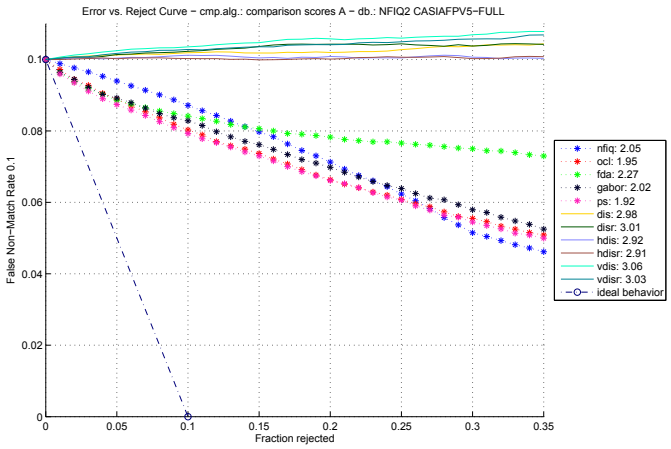
The results given by ERC for CASIAFPV5 are depicted in Fig. 6.5a, Fig. 6.5b, Fig. 6.5c, using comparator A, B, C respectively.

The results given by ERC for MCYT330PB are depicted in Fig. 6.6a, Fig. 6.6b, Fig. 6.6c, using comparator A, B, C respectively, and similarly for MCYT330DP are illustrated in Fig. 6.7a, Fig. 6.7b, Fig. 6.7c

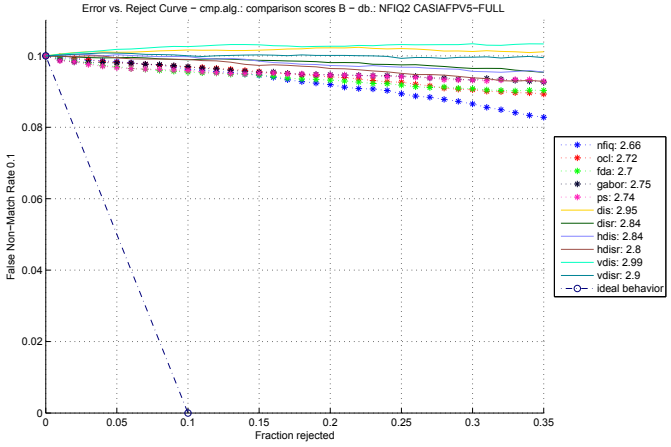
Regarding CASIAFPV5, the ERC is not decreasing for the SP localization-based methods, which seems the proposed approaches are not suitable for CASIAFPV5. As for MCYT330DP and MCYT330PB, however, we can in general observe the ERC for the proposed SP localization-based methods are linearly decreasing. The reason is that the fingerprints in MCYT330PB and MCYT330PB have similar orientations, but there are a large number of systematic rotations occuring in CASIAFPV5, with the result that the distance measurement becomes meaningless. Hence the proposed method is not invariant with respect to the rotation of the fingerprint and should be applied based only on the prerequisite that there is no presence of systematic rotation.

In MCYT330DP and MCYT330PB, although the experimental results vary with dependency on the different databases and comparators, generally the moderate decrease means the low quality samples are rejected as expected. In other words, larger displacement of SP in a fingerprint sample degenerates the biometric performance.

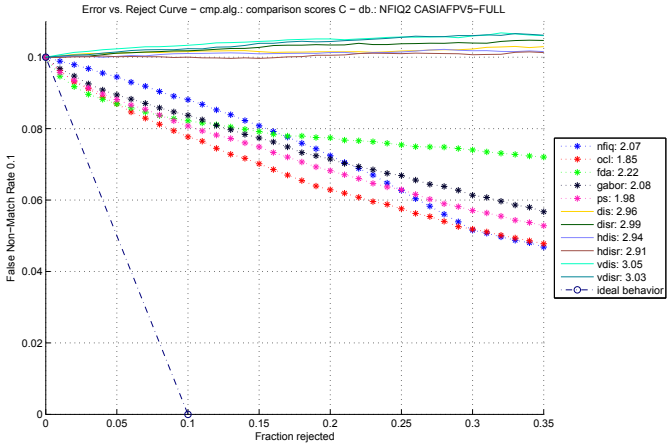
Moreover, we can see that averagely the proposed methods are not so efficient as the benchmarked baseline quality metrics (NFIQ, OCL, FDA, GABOR, PS), which means that the importance of the positions of core points is not as relevant as the measurement of the other features presented in Sec. 6.2.2. For MCYT330DP, the benchmarked quality metrics OCL and GABOR varies seriously among different comparators, but results using SP localization-based methods do not change dramatically, so the quality assessment using the core positions is more reliable.



(a)

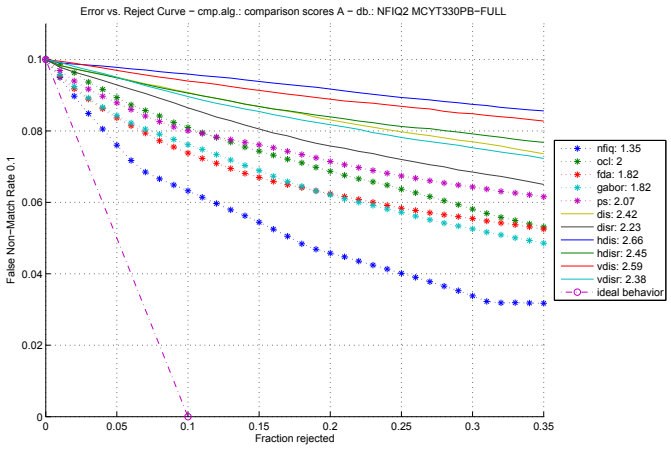


(b)

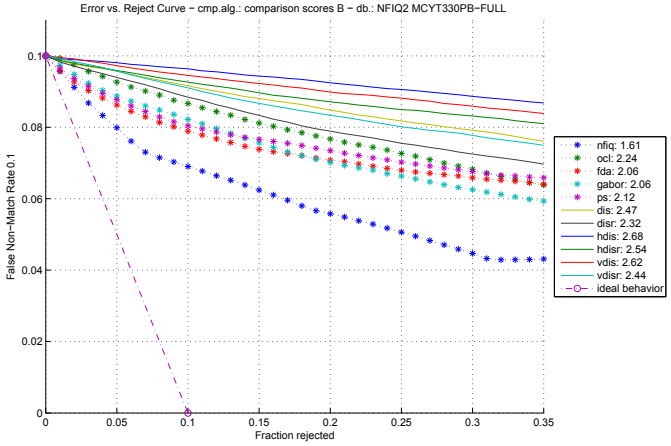


(c)

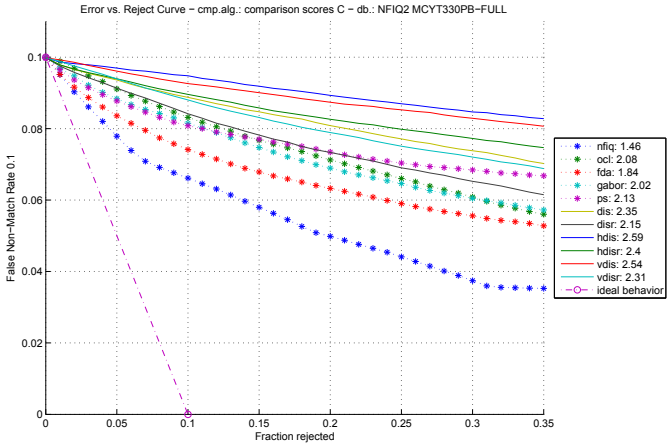
Figure 6.5: ERC for providers A, B, C (left to right) for CASIAFPV5 a, b, c.



(a)

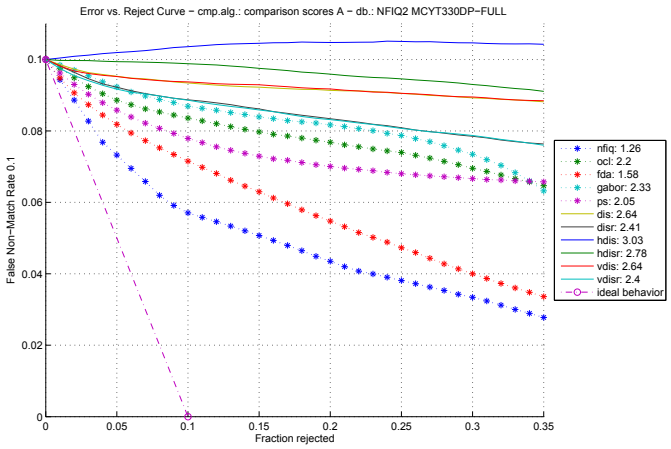


(b)

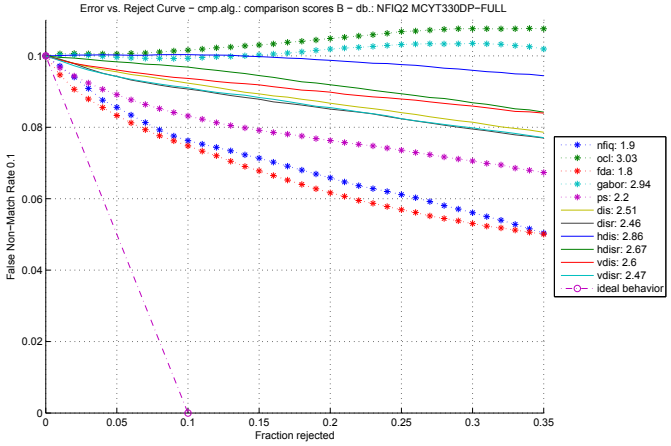


(c)

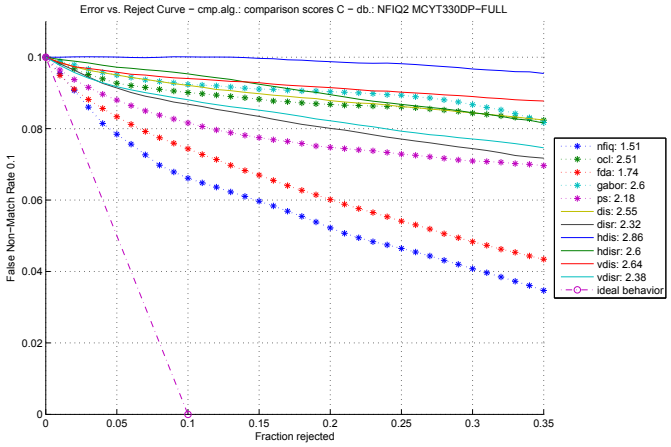
Figure 6.6: ERC for providers A, B, C (left to right) for MCYT330PB a, b, c.



(a)



(b)



(c)

Figure 6.7: ERC for providers A, B, C (left to right) for MCYT330DP a, b, c.

For MCYT330DP, the FNMR is not decreased by HDIS, which indicates that the quality of fingerprint samples in the database is not sensitive to the movement of SP in the horizontal direction. Among the SP localization-based approaches, averagely DIS possesses better performance than HDIS and VDIS. This positive result is depending on the effect of both HDIS and VDIS, which means that the displacements both in horizontal and vertical direction degenerates the biometric performance. Furthermore, the ratio version DISR, HDISR, VDISR has the lower area values than DIS, HDIS, VIDS so that the size of foreground should be considered.

Overall it is obvious that the proposed SP based quality method contains complementary information and fusion with the baseline quality metrics is recommended.

6.4.4 Utility-based quality heatmap

In order to investigate the effect of displacement on the biometric performance, quality heatmaps are generated according to the positions of the core points (the mid-point is adopted if there are two cores) and quality scores.

In this experiment, the utility-based quality score is used to indicate the "good" and "poor" quality of a sample. The utility scores are also genuine because they are computed based on genuine comparison scores using the method defined in ISO/IEC 29794-1 [70]. The utility of the sample $utility_i^u$ (the u^{th} presentation of subject i) is computed by means of the genuine and imposter comparison score distributions:

$$utility_i^u = \frac{m_{i,u}^{genuine} - m_{i,u}^{imposter}}{\sigma_{i,u}^{genuine} + \sigma_{i,u}^{imposter}} \tag{6.21}$$

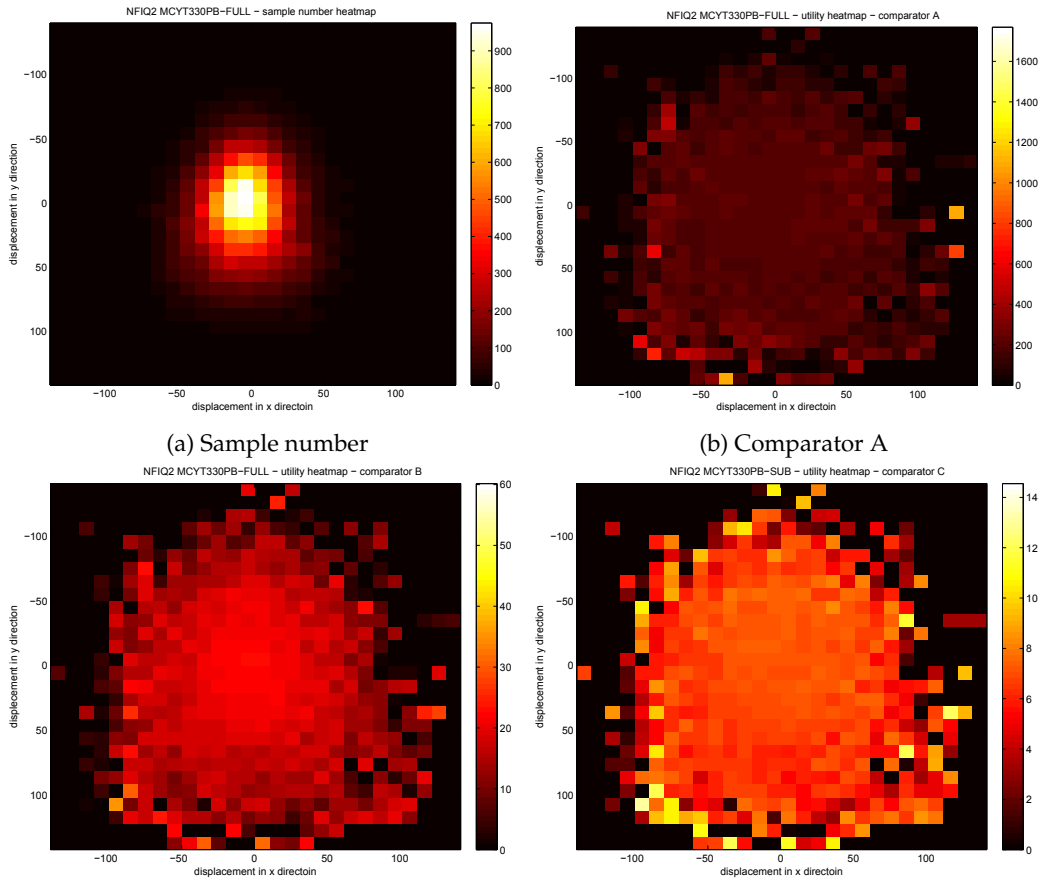
where μ and σ are the mean and standard deviation respectively. Consequently, higher utility score indicates the higher biometric sample quality and vice versa. As mentioned previously we used three top-performing comparators, A, B, C, so there are three utility scores for each finger sample. Each block is 10×10 pixels, and the score in each block is given by the mean of utility scores of samples having an SP in that block. In addition, the number of samples in each block is also plotted in order to find region of interest (ROI), i.e., the utility scores are meaningful only if there are plenty of samples.

Due to the rotations in CASIAFPV5, thus only the other databases are meaningful to be analyzed. The utility-based quality heatmaps are illustrated in Fig. 6.8 and Fig. 6.9 for MCYT330PB and MCYT330DP.

For both databases, generally in ROI the utility-based quality scores are decreased along with the increment of displacement, i.e., the quality score is lower if the core point is further away from the center of foreground. However, the decrease is more apparent with comparator B, which means that this algorithm is more sensitive to the movement of core points. In the other hand, comparator A is more resistant to displacements.

6.5 Conclusion

Our first contribution is a proposed sequential semantic filtering of candidate SPs, which can significantly reduce the presence of falsely detected SPs. Experimental results based on



(a) Sample number

(b) Comparator A

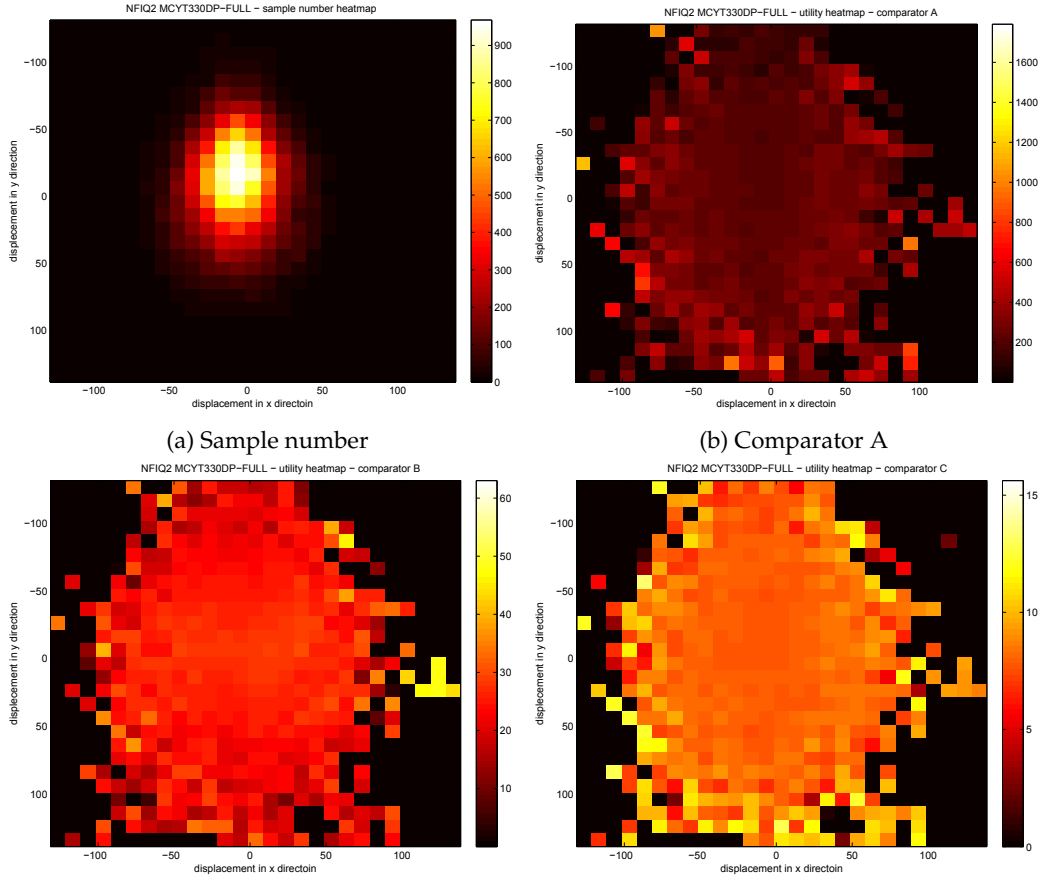
(c) Comparator B

(d) Comparator C

Figure 6.8: Number of samples and utility maps for MCYT330PB. Each block is 10×10 pixels, and ROI is approximately [-60, 60] and [-80, 100] in horizontal and vertical direction respectively.

FVC2002Db2 and SD14-BKA-GTD have shown that the total error rate decreased. Specifically, the SP false detection rate is decreased dramatically with a slight increment of the SP miss rate, which leads to a promising trade-off.

Based on the proposed SP localization approach, we then explored and analyzed in our second contribution, the importance of SP on fingerprint sample quality metrics in terms of the positions of detected core points with respect to the center of the fingerprint area. The experimental results given by ERC show the displacement measurement should be applied without the presence of systematic rotations, and further that the displacement has an influence on the biometric performance, but on average the relevance of the positions of core points is not as significant as the benchmarked baseline quality metrics, i.e, Orientation Certainty Level, Frequency Domain Analysis, Radial Power Spectrum, NIST Finger Image Quality and Gabor filter responses. A better performance is achieved by DISR, HDISR and VDISR so that the size of foreground should be considered.



(a) Sample number

(b) Comparator A

(c) Comparator B

(d) Comparator C

Figure 6.9: Number of samples and utility maps for MCYT330DP. Each block is 10×10 pixels, ROI is approximately [-60, 50] and [-110, 90] in horizontal and vertical direction respectively.

In conclusion, the SP based quality metric DIS is suggested to be used as a NFIQ 2.0 candidate quality metric because it measure the displacement comprehensively and has complementary properties to other NFIQ 2.0 candidate metrics.

Acknowledgments

This research was supported by the Center for Advanced Security Research Darmstadt (CASED).

Chapter 7

Self-Organizing Maps for Fingerprint Image Quality Assessment

Information about the paper

This contribution is represents a contribution to fingerprint quality assessment algorithms which are entirely defined through machine learning as a two tiered unsupervised and supervised approach to result in an algorithm which is computationally efficient. The paper is a contribution towards research question $\mathcal{RQ}_2$ and $\mathcal{RQ}_3$. This original paper was published in: [3] M. A. Olsen, E. Tabassi, A. Makarov, and C. Busch. "Self-Organizing Maps for Fingerprint Image Quality Assessment". In: *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*. CVPRW '13. Washington, DC, USA: IEEE Computer Society, 2013, pp. 138–145. ISBN: 978-0-7695-4990-3. DOI: 10.1109/CVPRW.2013.28. URL: http://dx.doi.org/10.1109/CVPRW.2013.28.

Abstract

Fingerprint quality assessment is a crucial task which needs to be conducted accurately in various phases in the biometric enrolment and recognition processes. Neglecting quality measurement will adversely impact accuracy and efficiency of biometric recognition systems (e.g. verification and identification of individuals). Measuring and reporting quality allows processing enhancements to increase probability of detection and track accuracy while decreasing probability of false alarms.
Aside from predictive capabilities with respect to the recognition performance, another important design criteria for a quality assessment algorithm is to meet the low computational complexity requirement of mobile platforms used in national biometric systems, by military and police forces.
We propose a computationally efficient means of predicting biometric performance based on a combination of unsupervised and supervised machine learning techniques. We train a self-organizing map (SOM) to cluster blocks of fingerprint images based on their spatial information content. The output of the SOM is a high-level representation of the finger image, which forms the input to a Random Forest trained to learn the relationship between the SOM output and biometric performance. The quantitative evaluation performed demonstrates that our proposed quality assessment algorithm is a reasonable predictor of performance. The open source code of our algorithm will be posted at `http://www.nist.gov/itl/iad/ig/development_nfiq_2.cfm`

7.1 Introduction

Recognition performance is increased if fingerprints are of sufficiently good quality and when overall database integrity is improved. Therefore, the ability to automatically examine a biometric image and produce a numerical estimate of the utility of the image to a downstream comparison algorithm (i.e. matcher) is an operationally desirable function. Particularly, if a low quality value is predictive of recognition failure (primarily a false negative, but possibly a false positive too), then a new sample can be collected while the subject is still present. Without a method to determine whether the quality of a captured sample is sufficient for recognition purposes, an individual can be falsely rejected or actively subvert the system. As such, many large scale biometric deployments mandate measuring and reporting of quality scores. Examples are the Unique Identification Authority of India (UIDAI) [127], the European Union Visa Information System (VIS) [123, 124, 161], or the United States Visitor and Immigrant Status Indicator Technology (US-VISIT). While the ability to inspect low utility images is essential for both the server and client side of these applications, the client side demands low computational cost, both in terms of time and processing power. In fact, requirements for next generation quality assessment includes near frame rate throughput [162].

The major contribution of this paper is a novel method for quality assessment which has a very low computation cost; essentially after training the quality assessment is reduce to a look up table. Our proposed method is based on a combination of unsupervised (self-organizing map [163]) and supervised (Random Forest [164]) machine learning algorithm. A self-organizing map (SOM) is utilized as a receptive field to learn the spatial information content of a fingerprint image creating a high level representation in the form of class

labels. Transforming an input from image space to class labels reduces the dimensionality by several orders of magnitude while retaining key characteristics of the finger image. This new representation of the finger image becomes input for the Random Forest which is to give a quality score that is predictive of biometric performance. To the authors' knowledge this is the first application of using a receptive field approach to fingerprint quality assessment.

The paper is organized as follows: section 7.2 provides a quick overview of related work in fingerprint quality assessment. section 7.3 details our proposed method. Discussion of data, experiments and results are given in section 7.4, followed by conclusions and future work in section 7.5.

7.2 Background and Related Work

Several fingerprint quality assessment algorithms have been documented in the literature, a comparative study of them is given in [97]. The majority of them are based on processing an image in either the spatial or frequency domain to compute particular features such as orientation strength. Subsequently a scalar quality value is computed from a feature or a combination of features. The technical report on fingerprint sample quality [99] by the International Organization for Standardization has a recent overview of recommended fingerprint quality features; e.g. Orientation Certainty Level which measures the strength of the energy concentration along the dominant ridge flow orientation within a block by computing the blockwise gradient; Ridge-Valley Structure which computes the blockwise clarity of ridges and valleys by applying linear regression to determine a gray-level threshold and then classify pixels as ridge or valley. A ratio of misclassified pixels is determined by comparing the normalized ridge and valley thickness of that block. Orientation Flow proposed in [72] is a measure of the rate of change in the blockwise ridge flow across the fingerprint. The quality score decreases as the difference between the dominant ridge orientation of the block and its 8 neighboring blocks increases. A block wise Gabor based quality feature was proposed in [43] and a point wise Gabor based quality feature was proposed in [1]. One widely used open source quality assessment algorithm which combines several of the above features is NFIQ [78]. It is based on a neural network trained on an 11-dimensional feature vector derived from finger image characteristics which include minutia count, ridge curvature and local contrast.

Computation of these features happens to be expensive, first a processing unit is required for their computations, and secondly according to our in-house implementations, it takes more than 130 millisecond to compute any of the features. In this paper, we aim to address this shortcoming of the current methods. Once trained, our proposed method is essentially reduced to look-up table operations, which reduces the computation time by about half, and make it suitable for mobile platforms with minimum computational resources.

7.3 Methodology

This section details how we employed the SOM and Random Forest clustering for calculation of fingerprint image quality. fig. 7.1 shows the overview of our proposed method. We give a review of the SOM concept in section 7.3.1, and its application to our proposed

method in section 7.3.2. section 7.3.3 describes how we used Random Forest to classify the output of the SOM into five different quality categories.

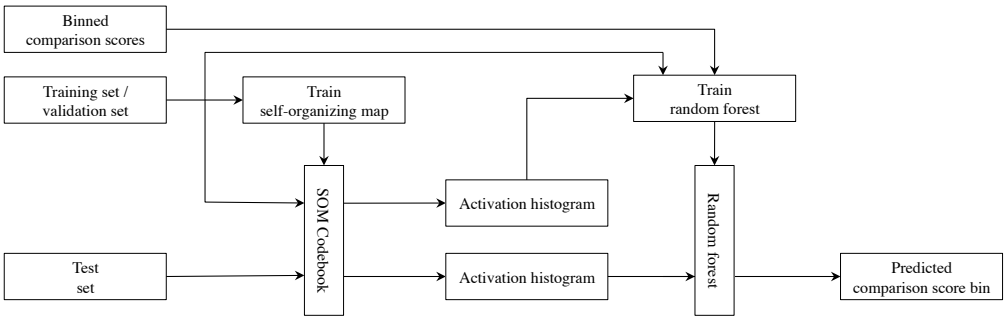


Figure 7.1: Overview of system architecture and components. The training and validation sets and their binned associated comparison scores are utilized to train the self-organizing map and the Random Forest. Training the SOM creates a codebook which is used to create an activation histogram of the training and validation sets. The histogram of activations from the training set is used as the explanatory variables in the Random Forest, while the binned comparison score is the target variable. Once trained, the Random Forest produces a model which is verified using the validation set. The test set, containing the unseen data, is used to determine the accuracy of the pipeline.

7.3.1 Overview of Self-organizing Maps

The Kohonen Map, introduced in [163], is a self-organizing map (SOM) which is trained without supervision. Its objective is to organize the input patterns to a topological structure where the relations between different patterns are preserved. The SOM is a two-layered network consisting of a number of processing units. The first layer of units can be considered as a group of sensors picking up the signal. It is fully connected to a second, two-dimensional layer: the competitive layer. fig. 7.2 shows the topology of the network. The weights associated with the connections are adjusted during training and only one single unit in the competitive layer is active at a time. Due to the training rules explained below, the spatial distance of two units reacting on different input patterns is a measure for the similarity of the two patterns. The training of the network is carried out by presenting data vectors $\mathbf{x}$ to the input layer of the network whose connection weight vectors $\mathbf{m}_i$ of all competitive units i are first initialized by random values. If N is the dimension of the data, we choose N input units of the first layer and define an Euclidean distance d_i between $\mathbf{x}$ and $\mathbf{m}_i$ with

$$d_i = \| \mathbf{x} - \mathbf{m}_i \| = \sqrt{\sum_{j=0}^{N-1} (x_j - m_{ij})^2} \tag{7.1}$$

The input signal activates winner unit c whose weight vector has the minimum distance to the given signal. The winning unit is also known as the best matching unit (BMU).

$$d_c = min(d_i) \tag{7.2}$$

The updating of the weights m_{ij} associated to the units is only performed for units i within a proximity $d(t)$ defining the time-dependant neighborhood $N_c(t)$ of the winning c. This proximity $d(t)$ is reduced as training of the network is progressing with time t (eq. (7.6)). fig. 7.2 illustrates the time-dependent neighborhood. The update of weights follows eq. (7.3), where $a(t)$ represents a time-dependent learning rate.

$$m_{ij}^{(t+1)} = m_{ij}^{(t)} + \Delta m_{ij}^{(t)} \tag{7.3}$$

$$\Delta m_{ij}^{(t)} = \begin{cases} a(t)(x_j - m_{ij}^{(t)}), & \text{if unit } i \in N_c(t) \\ 0, & otherwise \end{cases} \tag{7.4}$$

$$a(t) = a_0(1 - \frac{1}{T}), t \in [0, \ldots, T-1] \tag{7.5}$$

$$d(t) = d_0(1 - \frac{t}{T}), t \in [0, \ldots, T-1] \tag{7.6}$$

The SOM offers two properties during training that are strongly related to our challenge

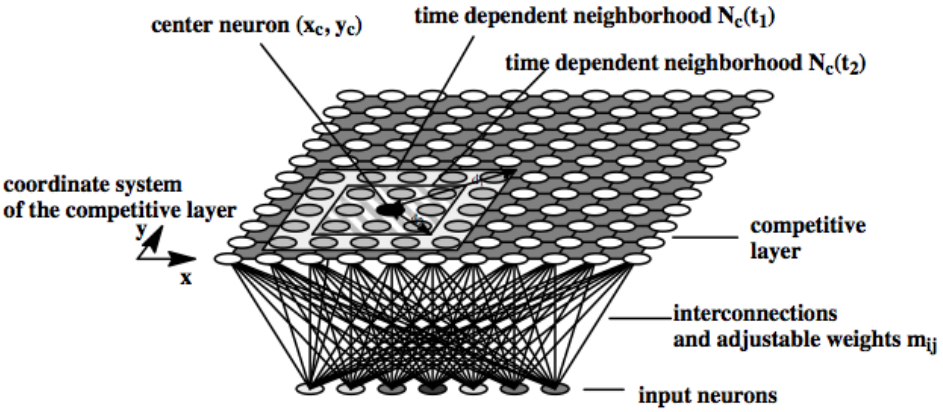


Figure 7.2: Topology of the self-organizing map and the time dependent neighborhood $N_c(t)$ of a unit c.

namely a) an approximation of the presented data by mean vectors $\mathbf{m_i}$ that are associated as weights to the units thus essentially realizing a cluster analysis of the data and b) a topological ordering of the competitive units in such a way that neighboring units in the layer represent similar clusters in multidimensional space and thus a dimensionality reduction. This can be interpreted as a nonlinear, topology preserving, associative mapping process. Generally, the Kohonen Map refers to the well-known K-means clustering [165]. As the units in the competitive layer are ordered topologically, i.e. neighboring units react to data vectors similar in the input space, the mapping can be interpreted as a reduction of any N-dimensional space into two dimensions preserving the topology of the data as much

as possible. The resolution of this discrete 2D-space is given by the number of competitive units, i.e. clusters, which should be configured such that the map serves the requirements with regard to grading (granularity) of the quality score. The limited use case of the SOM for clustering, topological organization and for subspace mapping has been described. Due to the fixed number of units in the SOM and its topology preserving property, some units may have very similar mean vectors $\mathbf{m_i}$. With a trained SOM we then assign a meaning or class to each unit, such that its activation can be related to biometric performance in the form of a comparison score or binned comparison score. Our SOM input is blocks of the fingerprint image, and therefore we do not know the contribution of each individual block towards the resulting comparison score. Furthermore, comparison scores are pairwise while quality scores only relate to a single impression. Thus we cannot know if a low comparison score is due to both impressions being of low quality or only one of them. We uncover the relation between each unit in the SOM and biometric performance through another layer of machine learning. In particular we utilize Random Forests to determine the relation between a certain activation pattern and a given comparison score (see section 7.3.3). During the organization process the goal is to find a limited set of centroids representing the data in a C-means sense. As a result the centroids represent the variety of feature vectors in an optimal way leading to a minimum of the error function, which can be easily calculated by the Euclidean distance between all feature vectors and their nearest codebook vector in the map. This search does not need to cover the entire space of codebook vectors since efficient approximations can exclude the vast majority of codebook vectors from consideration [166]. The unit which is closest to the input vector is known as the best matching unit (BMU).

7.3.2 Application of Self-organizing Maps

First, to provide a fixed length input vector for the SOM the finger images are subdivided without padding or overlap into blocks of size $n \times n$ pixels after intensity normalization. One-dimensional vectors of length n^2 are constructed from the pixel intensities in each block by concatenating the rows. For the training a SOM of size $dim \times dim$ is initialized with randomly generated weight vectors. During the training phase each training input vector (i.e. the pixel intensities of a block) is presented to the network and modification of network weights is performed. After T iterations (assuming T is large enough) the SOM clusters the training input vectors into $dim \times dim$ units representing different spatial characteristics of fingerprints. The trained SOM produces a codebook with $dim \times dim$ entries each of length n^2 where each codebook entry represents a cluster of input training vectors. Due to the update function (eq. (7.3)) the codebook vector (or entry) associated with each unit will stabilize such that nearby units will have similar codebook vectors. Figure 7.3 shows the codebook vectors of a SOM of size 24 by 24 ($dim = 24$) trained using input vectors of length 576 ($n = 24$) for $T = 1 \times 10^6$ iterations. It is clear that the SOM has identified clusters representing various ridge orientations, ridge strengths and background. Variation in image intensity in the codebook is absent or limited due to the image normalization performed as explained in section 7.4.1. The trained SOM assigns an index of the corresponding best-matching unit to each block of an input image. fig. 7.4 shows the unit activation patterns for a finger image. Note that the white background appearing in the input images shown in fig. 7.4a activates the same units in the SOM (shown in fig. 7.4b) indicating that the background has a cluster of its own.

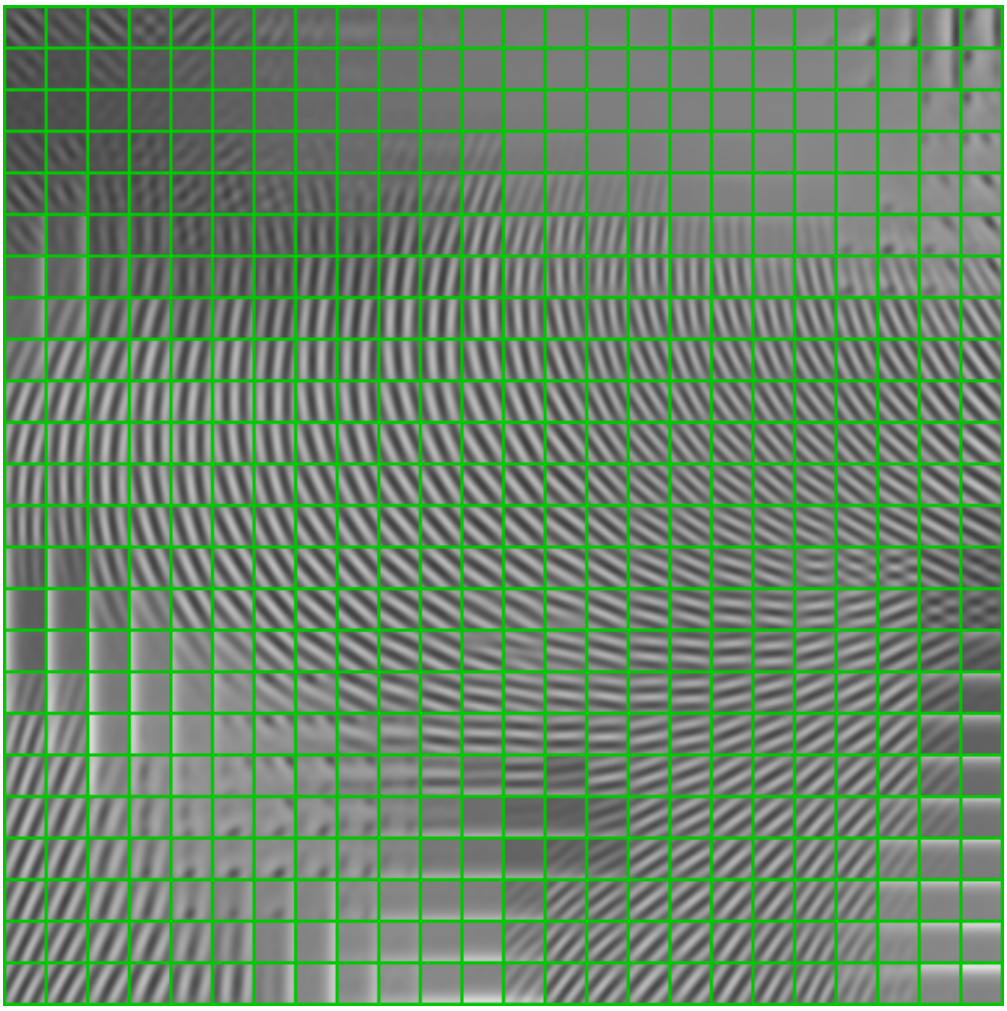
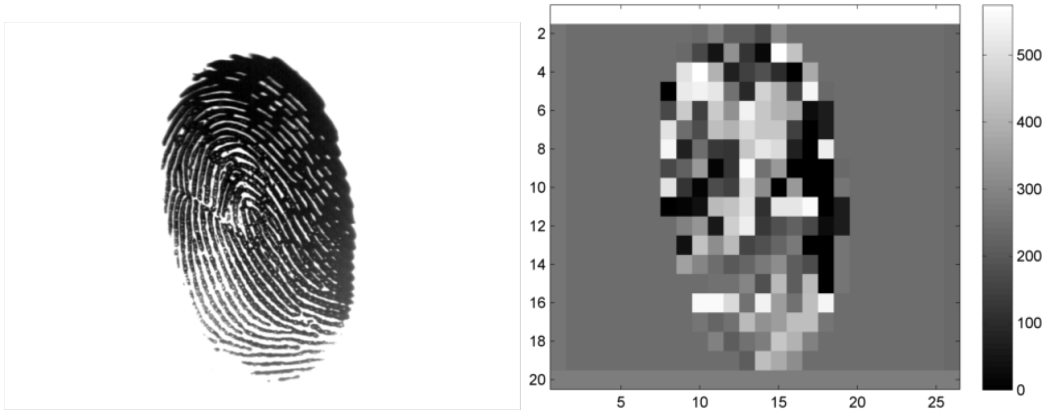


Figure 7.3: Visualization of the codebook for a SOM of size 24 by 24 ($dim = 24$)trained with 1.7×10^7 input vectors of length 576 ($n = 24$) for $T = 1 \times 10^6$ iterations shown in a grid structure. Unit 0 is located in upper left corner and unit 575 is located in the lower right corner.



(a) (b)

Figure 7.4: llustration of relationship between finger image and SOM codebook shown in fig. 7.3. 7.4a is the input images which is divided into blocks of size 24×24. 7.4b shows the best-matching unit for each image block where black is unit 0 and white is unit 575.

To summarize, a trained SOM transforms a fingerprint image from its spatial domain into a vector of the SOM unit activations. The length of this vector varies for different images because the number of blocks depend on the image size. To get a fixed-length representation we compute the histogram of unit activations normalized by the block count per image. Each image block in a finger image has exactly one BMU and thus we can form a histogram of unit activations. This is motivated by the expectation that the activation patterns for high quality and low quality images are different. Therefore, an image of size $U \times V$ is represented by a vector of length $dim \times dim$ with each component being the frequency of occurrence of the codebook entries.

To determine how the SOM behaves when presented with low quality (i.e. difficult to match) and high quality (i.e. easy to match) fingerprint images we select 64 images that result in false non-matches at the threshold corresponding to false match rate of 1×10^{-4} (this set is denoted Low) and 64 images involved in comparisons giving the highest possible comparison score (this set is denoted $High$). Note that the threshold was computed over $120\,000$ impostor comparisons. fig. 7.5 shows 10 randomly chosen examples of activation histograms from each of the Low and $High$ sets. The SOM is of size 32×32 ($dim = 32$) and the network was trained for $T = 1 \times 10^6$ iterations using image blocks of 24×24 pixels ($n = 24$). Visual inspection reveals that the histograms of images in the low and $high$ sets look different. fig. 7.6 shows the sum of SOM unit activations of the 64 images in the $High$ and Low sets respectively. The difference of these two aggregated histogram ($Low - High$) is depicted in fig. 7.7. Bars that extend below zero represent units with stronger activation in the $High$ set than the Low. Similarly bars that extend above zero represents units with stronger activation in the Low set. Since the difference between the two sets is non-zero we conclude that the histograms of the two sets are substantially different. In other words, we have shown that the activation histogram of images involved in a false reject are significantly different from that of those which are highly likely to be correctly recognized. Therefore we used this activation histogram as our feature vector for a supervised

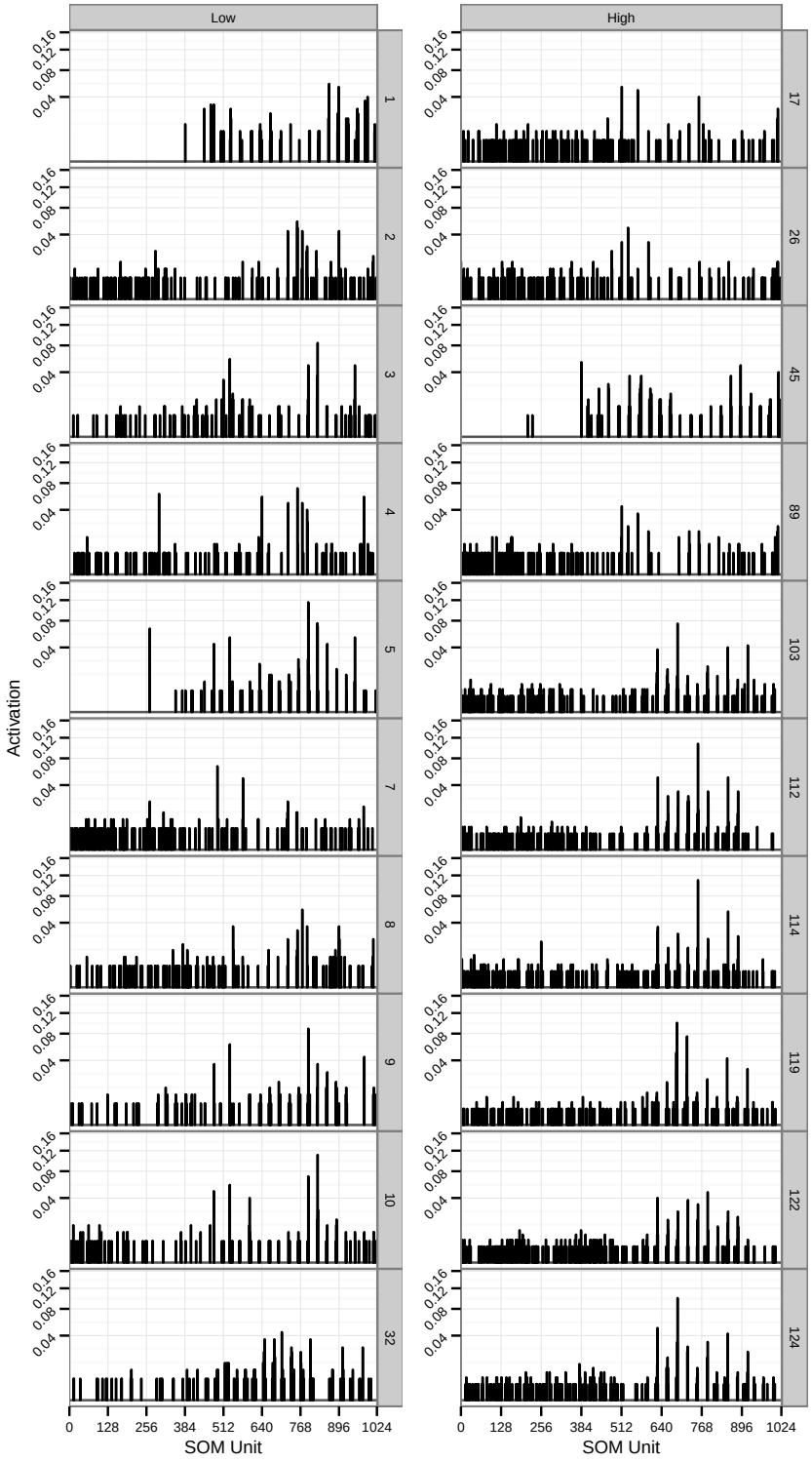


Figure 7.5: On the left, histogram of activations for 10 images from the set of images which occurred in comparisons resulting in false non-matches at FMR= 1×10^{-4} (set $High$). On the right, histogram of activations for 10 images from the set of images which occurred in comparisons with the highest possible comparison score (set Low).

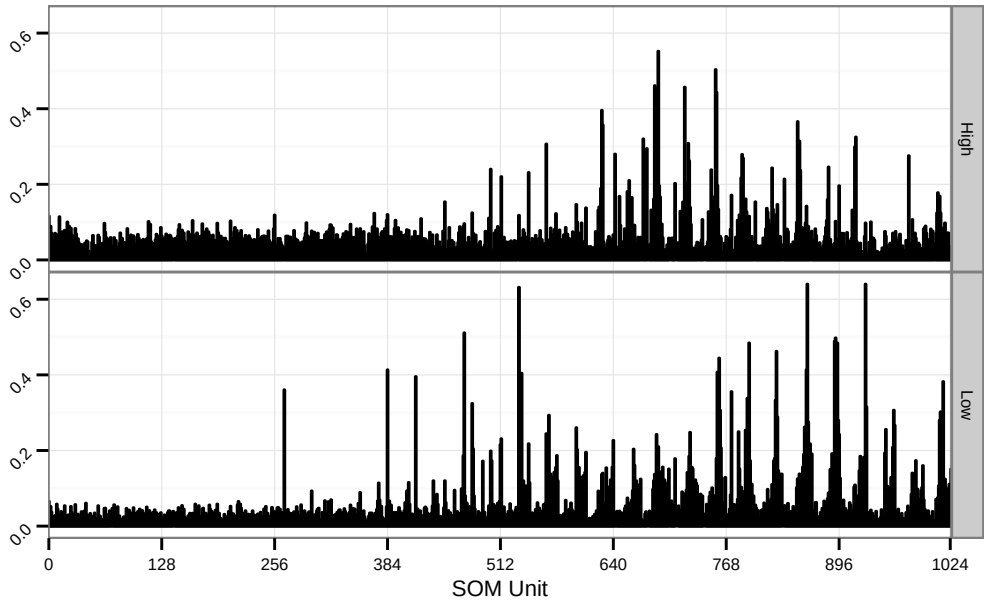


Figure 7.6: Aggregated histograms of the *Low* and *High* sets. The upper figure shows the aggregate histogram of 64 images that gave the highest possible comparison score (easy to match and therefore high quality). The lower figure shows the aggregate histogram of 64 images that produces false non-match at threshold corresponding to false match rate of 1×10^{-4}.

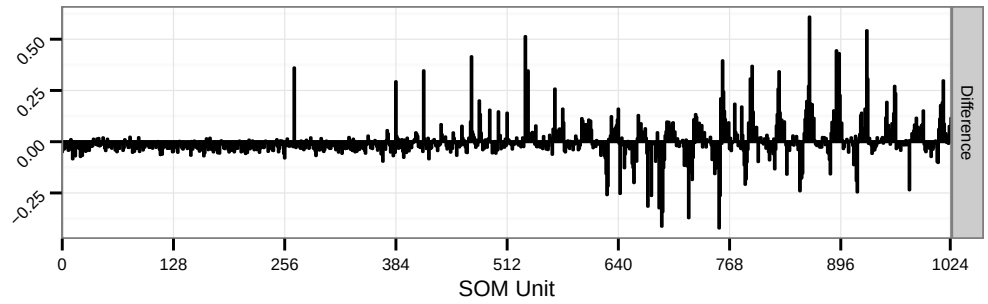


Figure 7.7: Difference between aggregated histograms of the *Low* and *High* sets. Negative values indicate that the unit is more active in images that gave the highest possible comparison score (easy to match and therefore high quality), while positive values indicate that the unit is more active in images that gave a false reject (therefore hard to match or low quality).

learning to predict biometric performance.

7.3.3 Classification using Random Forest

We chose bins of genuine comparison scores as the response variable and tried several classifiers: K-Nearest Neighbors, Support Vector Machine, Recursive Partitioning tree, and Random Forest. Due to the 8-page limit, we only present results of Random Forests which performed better than the others. Random Forests [164] basically grows many classification trees. To classify a new object from an input vector, it puts the input vector down each of the trees in the forest. Each tree gives a classification, in other words the tree votes for that class. The forest chooses the classification having the most votes. Random Forest has the advantages of being fast, capable of handling large datasets, and that it does not over fit.

We investigate network dependence on block size, network size and number of iterations by varying n and dim. table 7.1 summarizes the accuracy of the Random Forest for the different parameters. It shows smaller block size gives better performance. Increasing the size of the network does not seem to have a significant effect, which might be due to lack of sufficient training data. Theoretically, the size of the map is only limited by the memory capacity of the computing unit and availability of sufficient number of training imagery. A larger SOMnetwork is more capable of representing all possible spatial information of its training data, but it requires a larger number of images for training.

7.4 Experiments and Results

7.4.1 Data set

Data we used comprised of 7976 right index finger images of 3988 persons (2 images per person) selected from a large operational repository. We took care to construct a balanced training set (1000 pairs) with a representation of real world difficult to match samples to the extend possible. Comparison scores are generated using one of the leading commercial comparison algorithms we'll reveal the name of the provider in final copy if paper is accepted.We want the training data to be a balanced representation of all possible (or to the extend possible) fingerprint impairments, as well as pristine images. We found that the current available public datasets were not suitable for our work. This is mainly due to small size of these datasets (a few hundred users at the most) and also absence of a reasonably wide range of quality problems in these collections, which is expected and understandable given the careful and at times controlled design of these collections. To ensure inclusion of a wide range of quality in our dataset, we formed a balanced representation from all possible comparison score values generated by the biometric comparator used in our study. Given that different levels of fingerprint quality result in different values of comparison scores, a set of images representing all possible comparison scores, would represent the obtainable diversity in the quality. Specifically, we chose images from a large operational repository using comparison scores of a leading commercial fingerprint comparison provider. First, we quantized the raw comparison scores by dividing their range into 100 bins. Then for each bin of comparison scores, we randomly selected 10 pairs of images with the corresponding quantized comparison score. This gave 2000 fingerprint images (10×100 comparisons, 2 images per comparison) from which we used 1800 for training and 200 for validation. Our

test set consists of images from 2988 comparisons (5976 fingerprint images) selected at random from the initial corpus excluding images occurring in any of the 1000 comparisons selected for the training or validation sets. We further grouped the quantized comparison scores (which are in [1-100]) into 5 levels such that quantized scores in [1, 20] belongs to bin 1, [21, 40] belongs to bin 2 and so on. These 5 levels quantify the biometric performance for each pair of images and are used as the target (response) values for a Random Forest whose explanatory variables are the SOM activation histogram. Images are normalized to give equal intensity according to the pyramid based enhancement method from [167]. An example of the enhancement process is shown in fig. 7.8.

Figure 7.8: Example of normalization process using image from FVC 2004 [168]. (a) is the input image and (b) is the enhanced version.

7.4.2 Performance measurements

This section documents methods and results for the quantitative evaluation of our proposed quality assessment approach. Evaluations are done by quantifying the association between quality scores and the observed matching results. We used three widely accepted and documented methods to assess the goodness of our proposed quality algorithm in predicting performance. We examined the relationship of quality scores with a) genuine comparison scores, b) recognition error rate (FNMR) for each level of quality, and c) improvement of FNMRas poor quality samples are removed.

	$dim = 24$	$dim = 32$	$dim = 40$
$n = 24$	26.38%	26.76%	26.23%
$n = 32$	25.85%	24.32%	26.34%
$n = 64$	23.18%	22.73%	24.18%

Table 7.1: Accuracy of prediction of comparison score bin for various parameter choices for the SOM dimensions (dim) and block sizes (n). The percentage reflects the accuracy of the predicted bin matching the actual bin as determined from the comparsison score.

Per the definition of biometric quality [70], quality values q should be a monotonic increasing function of performance, where high-quality samples give high similarity scores when involved in genuine comparisons. Figure 7.9 shows the relationship of our proposed quality scores with biometric comparison scores, we generated heatmaps (i.e. surface plots) of the three variables q_{verify}, q_{enrol} and $score$ of the images in our test sets. The x-axis represents quality of the enrolment samples. Quality of verification samples are on the y-axis. The color in each cell represents the magnitude of genuine comparison (or similarity) scores. If quality is an indicator of performance, higher similarity scores (shown in yellow) will be achieved for high-quality images of the same finger. The heatmap in the lower right corner shows the magnitude of samples in each of the combinations of $score$ for q_{verify}, q_{enrol}. Of the block sizes investigated, block size 24 gives the best performance; images with higher quality score (4 or 5) result in the highest similarity scores (visualized by yellow color in the upper right corner) and images with low quality scores (1 or 2) result in the lowest similarity score (visualized by red color in the lower left corner. Block size 64 does not perform well.

A quality assessment algorithm is useful if it can at least give an ordered indication of an eventual performance. This is the case for our proposed method as shown in Figure 7.10. Figure 7.10 shows the FNMRfor each $L = 5$ levels of our proposed quality method. The highest FNMRis observed for the lowest quality samples ($q = 1$), and lowest FNMRis observed for the highest quality samples ($q = 5$). Biometric matching involves at least two samples and the challenge is then to relate performance (which involves two samples) to quality values q_{verify} and q_{enrol}. We simplify the analysis by combining the two qualities according to $q_i = M(q_{verify}, q_{enrol})$. It is usually the case that operationally a quality assessment algorithm can be used to assess whether an enrolment sample is of high quality. The enrolment sample will be compared later with a sample that typically is of less controlled quality. We used $M(x, y) = \min(x, y)$ which captures the concept that the worse of two samples drives the similarity score. Figure 7.10 shows that our proposed method provides a rank order indication of FNMR: the highest error rate is observed for the lowest quality sample ($q = 1$) and the lowest error rate for the highest quality ($q = 5$). As such, our proposed method is successful in predicting performance.

Another metric for comparative evaluation of quality assessment algorithms is the error versus reject curves [112]. The goal is to demonstrate how efficiently rejection of low-quality samples results in improved performance. This models the operational case in which quality is maintained by reacquisition after a low-quality sample is detected. Consider that a pair of samples (from the same subject) with qualities q_{verif} and q_{enrol} are compared to produce a genuine score, and this is repeated for N such pairs. If the quality values are perfectly correlated with the genuine comparison scores, setting threshold τ to give an overall FNMRof x and then rejecting x percent with the lowest qualities should result in FNMRof zero after recomputing FNMR. Figure 7.11 shows the error vs. reject curves of our proposed method as applied to different block, namely 24, 32, and 64. We set the value of τ to give a false non-match rate of ten percent. Pairwise quality is computed using the geometric mean of the quality scores of the two images being compared. Similar results were obtained using $\min(x, y)$. Behavior of a perfect quality assessment algorithm is displayed by the gray dotted line where the rejection of the lowest 10% quality would result in an FNMRof zero. The most operationally relevant part of the error vs. reject curves is usually on the left side where a small fraction, x, of low-quality rejections would be tolerable from the perspective of forcing a second enrolment attempt. We set $x = 0.3$,

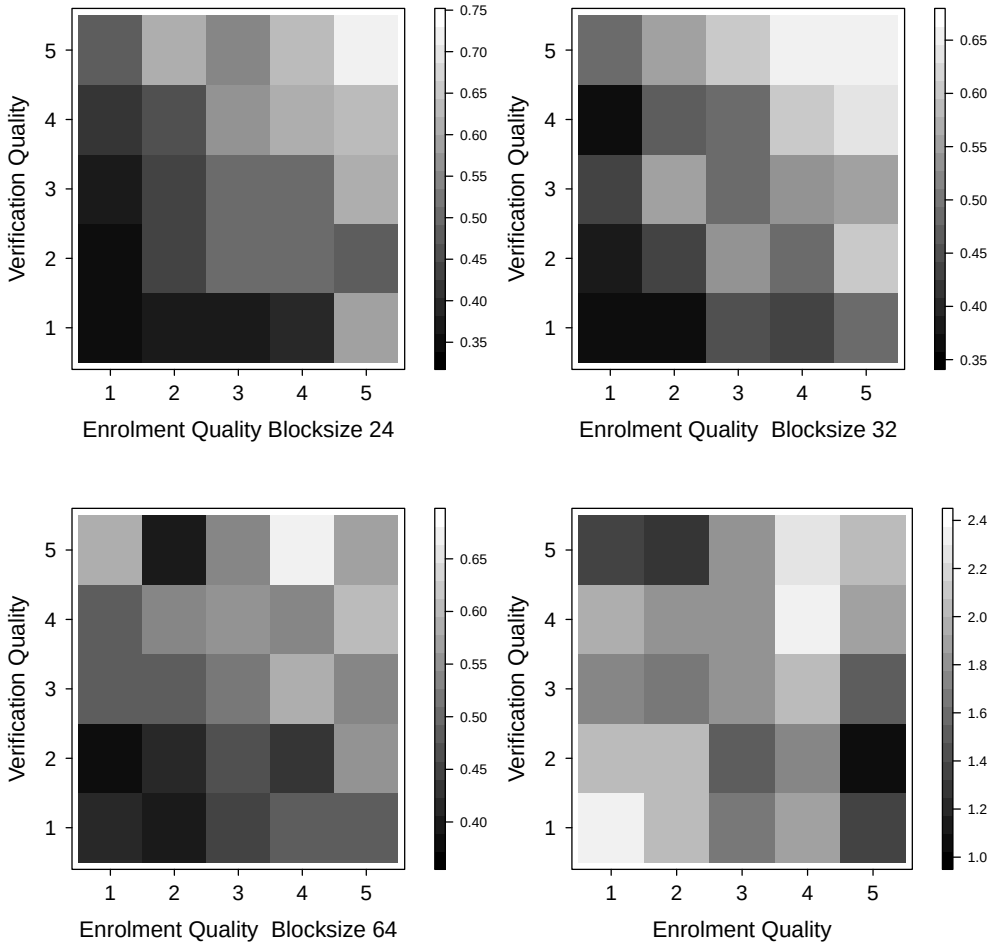


Figure 7.9: Genuine comparison scores vs. (q_{verif}, q_{enrol}) for the 3 different block sizes. The proper behavior is to have low-quality scores for low genuine scores, as is the case for block sizes 24 and 32. The figure in bottom right shows the count of images involved in each cell.

because rejection of more than 30% is not operationally feasible. Block size of 24 performs the best as it is evident by the sharpest decrease in FNMRwhen poor quality samples are rejected. Rejection of 10% lowest quality decreases FNMRto 0.085 from its initial value of 0.1. To give a point of reference for comparison between our method and existing methods, fig. 7.11 includes an existing known quality assessment method - the Orientation Certainty Level [99]. Our proposed method outperforms Orientation Certainty Level for rejection rates higher than 5%.

Computation efficiency: We noted the processing time required to compute quality values on a PC equipped with a 2.3 GHz Intel Core i7 and 16 GB of memory. With a network size of $dim = 24$ the quickest processing time was recorded at a block size of

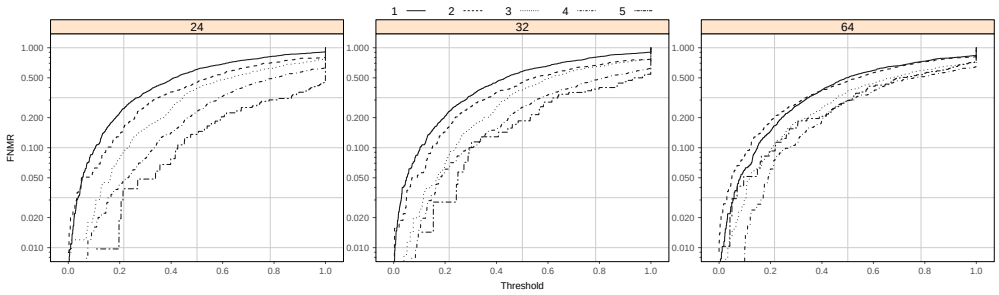


Figure 7.10: FNMRvs score threshold for each level of quality for different block sizes. The SOM-trained on block sizes 24 and 32 performs better than block size 64, as is indicated by a better separation of FNMRfor the five levels of quality.

$n = 24$ with 503 seconds to generate quality values for 7973 images (65 ms/image). With the same network size but block size $n = 64$ we recorded a total processing time of 654 seconds (82 ms/image).

We conclude this section by noting that all three evaluation methods employed here demonstrated that our proposed quality algorithm is capable of predicting performance of a commercial biometric matcher.

7.5 Conclusions and Future Work

We presented a novel approach for fingerprint quality assessment which is fast and of low computational complexity. Our approach is predictive of biometric performance and avoids the explicit quality feature extraction that often has high computational cost. This is achieved by training a self organizing map to represent the variety of spatial information content of fingerprints in our training data. The training of the SOMnetwork may require a significant computational effort, however the training is performed only once, in an offline setting, using suitable computing resources. Once the code book is generated, the operational mode is extremely efficient as the implicit quality assessment is reduced to a loading-task of the fingerprint-image blocks to the input layer of the trained SOM, and a subsequent lookup of the quality level in the associated Random Forest. Thus our suggested approach has the advantage of having a very low computational cost which makes it particularly suitable for applications constrained in processing resources such as low-end mobile fingerprint devices used by the Unique Identification Authority of India (UIDAI). We investigated the effect of finger image block size (input to the SOM), and the size of the SOM network. Our experimental results suggest that smaller image block size (24 pixels by 24 pixels in our case) performed better than larger sizes (32 or 64 pixels). A larger SOMnetwork while more capable of representing all possible spatial information of its training data, requires a larger number of images for training. We tried network sizes of 24 (576 units), 32 (1024 units), and 40 (1600 units). Network size of 32 outperformed the others, noting that our training data might not have been large enough for network size of 40. Quantitative results presented in section 7.4.2 demonstrate the predictive power of our proposed finger quality assessment method.

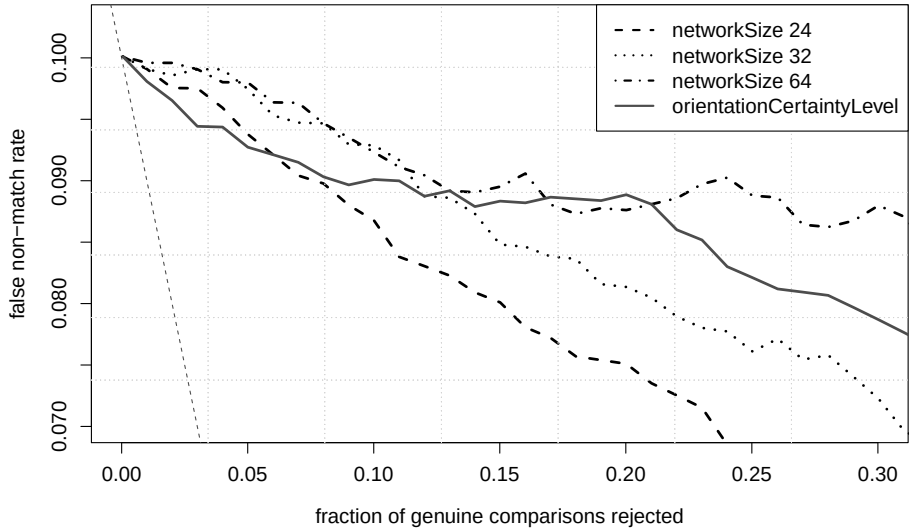


Figure 7.11: FNMRvs reject curves for our proposed method. The threshold is set to give an initial FNMR= 0.1. The gray dotted line shows the ideal case where the rejection of the ten percent lowest quality results in zero FNMR. Blocksize = 24 perform the best as it is evident by the sharpest decrease in FNMRas poor quality samples are rejected. An almost flat curve suggests that the quality algorithm is not effective in predicting the performance. The solid gray line shows the rejection curve for Orientation Certainty Level metric. Our proposed method outperforms Orientation Certainty Level for rejection rates higher than 5%.

Future work will address improvements in the SOM training, presentation of SOMoutput to the learning algorithm (Random Forest), and its fine tuning. A potentially valuable source of information is the distance between an image block and its associated BMU. This information could be used to create a weighted SOM activation histogram. Combining units with low distance between weight vectors may provide more robust classification. Additionally, we plan to repeat the study using comparison scores of one or more comparison subsystem (or algorithms) and investigate its generalizability to comparison algorithms not used at the training time. Finally, we are eager to try this approach for the quality assessment of latent impressions.

Chapter 8

Interpretation of Fingerprint Image Quality Features Extracted by Self-Organizing Maps

Information about the paper

This paper explores potential alternatives to Random Forests in interpreting the feature space produced by the Self-Organizing Map (SOM) approach presented in chapter 7. It is determined that there much of the SOM capacity is spent on encoding non-fingerprint regions and that mean gray level can be used to explain much of the variance in the produced map. The paper is a contribution towards research question $\mathcal{RQ}_2$ and $\mathcal{RQ}_3$. This original paper was published in: [4] I. Danov, M. A. Olsen, and C. Busch. *Interpretation of fingerprint image quality features extracted by self-organizing maps*. 2014. DOI: 10.1117/12.2050676. URL: http://dx.doi.org/10.1117/12.2050676.

Abstract

Accurate prediction of fingerprint quality is of significant importance to any fingerprint-based biometric system. Ensuring high quality samples for both probe and reference can substantially improve the system's performance by lowering false non-matches, thus allowing finer adjustment of the decision threshold of the biometric system. Furthermore, the increasing usage of biometrics in mobile contexts demands development of lightweight methods for operational environment. A novel two-tier computationally efficient approach was recently proposed based on modelling block-wise fingerprint image data using Self-Organizing Map (SOM) to extract specific ridge pattern features, which are then used as an input to a Random Forests (RF)classifier trained to predict the quality score of a propagated sample. This paper conducts an investigative comparative analysis on a publicly available dataset for the improvement of the two-tier approach by proposing additionally three feature interpretation methods, based respectively on SOM, Generative Topographic Mapping and RF. The analysis shows that two of the proposed methods produce promising results on the given dataset.

8.1 Introduction

Fingerprint image quality of the probe and reference samples directly affects the recognition performance of fingerprint-based biometric systems. Detecting low quality samples at acquisition time and reacquiring a new sample with better quality while the subject is still present can prevent the false non-matches. Therefore developing quality measurement algorithms that can predict quality scores is of crucial importance for large scale biometric deployments like the European Union Visa Information System (VIS)[169] or the US VISIT Automated Biometric Identification System (IDENT)[170].

Researchers have contributed to the development of several state of the art algorithms and moves toward standardization have been done[70, 82, 99]. However these algorithms are mostly based on computationally expensive image processing steps, which may make them unsuitable for mobile applications, where processing capabilities are constrained.

Self-Organizing Maps (SOM)[171] have been used for texture analysis[172] over the last couple of decades and that has inspired our research in applying SOM in the field of fingerprint image quality assessment[3]. The achieved operational speed makes the method attractive for applications in mobile environment, however its accuracy, though comparable with the state of the art, still lags behind the best predicting algorithms available.

The contribution of this paper is an investigative comparative analysis for the improvement of the block-wise fingerprint image quality estimation based on SOM. The proposed approach uses two-level modeling for constructing a linearly initialized SOM map to extract block-wise local features. The interpretation of these features to quality scores is done by four different approaches using three different machine learning techniques - SOM, Generative Topographic Mapping (GTM)[173] and Random Forests (RF)[164] with raw features or histogram based feature vectors.

This rest of the paper is organized as follows. Section 8.2 provides information on background and related work, Section 8.3 is an overview of the investigated machine learning techniques. Section 8.4 presents the methodology used for the experiments, whose execu-

tion and results are shown in Section 8.5. Finally, Section 8.6 summarizes conclusions and shares ideas for future work.

8.2 Background and Related Work

8.2.1 Fingerprint sample quality

Fingerprint image samples are compared by comparison algorithms with regard to similarity of their ridge impressions, either by extracting minutiae or some other features. As defined in the international standard ISO/IEC 29794-1 [70], biometric sample quality has two influencing components: character and fidelity; both affecting a third component utility as output. A Quality Measurement Algorithm (QMA) assigns a score to a sample which reflects the positive or negative contribution to the biometrics recognition performance of the biometric system.

Fingerprint sample quality is of high importance to modern fingerprint-based biometric systems, therefore quality estimation and quality measurement have wide range of applications, all aiming to improve recognition accuracy. However a measure of quality is biometric system dependent, as shown by Grother and Tabassi [112], since different comparison subsystems are sensitive to different sample artifacts like blurriness, contrast, translation, rotation, etc. In consequence the endeavour of establishing a universal quality standard defining a singular metric for the utility of the fingerprint sample for all applications remains a challenge[70].

For evaluating the results of the current work, two of the state of the art methods, NIST NFIQ[79] and Orientation Certainty Level (OCL)[71], have been used as baseline methods. NFIQ was proposed by Tabassi[79] and is currently the de facto quality algorithm. Its working principle is based on extracting eleven different features by analyzing the fingerprint image with NIST Fingerprint Image Software (NBIS)[174] to produce a feature vector. The feature vector is given as an input to an artificial neural network to classify the image as one of five quality levels. The neural network is a multi-layered perceptron which has been trained on a subset of images from five datasets with different capture conditions.

OCL is a local feature analysis method for quality estimation introduced by Lim et al. [71]. It is part of both ISO/IEC 29794-4[99] and NFIQ 2.0 Feature Definition Document[82]. This method has grounds in the observation that good quality samples have well defined ridges and valleys. If taken block-wise, ridge and valley lines have the same orientation and determining that orientation for good quality samples could be done with high certainty, unlike for poor quality samples. Therefore the measured orientation certainty is used as quality feature for each block, which are later aggregated into a single quality score.

Definitions of state of the art quality features have been presented in ISO/IEC 29794-4 [99] as well as in NFIQ 2.0 Feature Definitions Document (v0.5)[82]. Their evaluated performance has been shown in NFIQ 2.0 Feature Evaluation Document (v0.5)[82]. However most of the presented methods rely on image analysis techniques for the feature extraction and happen to be computationally expensive. Therefore using machine learning methods to learn the features from a training dataset is a promising novel approach that can speed up the quality assessment.

8.2.2 Holistic approach using Self-Organizing Maps

Self-Organizing Maps[175] is an unsupervised learning method, that has been used for clustering and data visualization. A detailed perspective on the capabilities of this method will be given in the following Section. In the field of image analysis, one of its best use cases is for texture analysis[172]. Therefore using it in the field of fingerprint image sample analysis is a straightforward approach, since fingerprint samples have very specific textures caused by the ridge impressions. Makarov[176] has recently experimented with modeling the whole-image data of a fingerprint sample by applying SOM and then classifying the new samples according to their best-matching unit. Since SOM effectively maps similar samples to the same node, it is expected that similar samples would have similar features, thus similar quality assessment.

The holistic approach is based on the following principle: the pixel values data of the whole-image is taken row-wise to construct a single feature vector. Then a set of samples are used for the training of a randomly initialized hexagonal SOM map with a predefined network size, e.g., 16×16 nodes, 24×24 nodes, etc. The quality scores of the training samples are calculated by binning the observed utility score and after the training has finished for each node a quality class is assigned by the mean quality score of all training samples mapped to that node. A winner takes it all assignment, where the most populous quality class among all training samples mapped to a node is the winner, was also tried in the performed experiments as an alternative quality score node labeling. Also image segmentation for background and foreground separation was applied in some of the experiments and the results were compared with the results obtained from raw image data.

The main goal of the research was to prove that SOM is capable to provide a mapping, such that it would be useful for predicting quality. Producing the results required a validation set to be used and has shown that the samples predicted by the model to be of low quality consisted of more samples with low observed utility; respectively, the samples that were predicted to be of high quality consist of more samples with high observed utility. These results for CASIAFPV5[140] dataset appeared to be quite promising and fueled further research of the usage of SOM for modeling fingerprint image data.

8.2.3 Block-wise approach using Self-Organizing Maps

Even though the holistic approach has shown promising results, the trained models were not able to capture many differences in the images. The global analysis of the images leads to separating them into different groups, but often these groups have large inter-group variations. Therefore, the next advances of the research were done by Olsen et al. [3] and they were based on block-wise fingerprint image analysis and training a randomly initialized SOM model that has to capture the block images data. Before dividing images into blocks, they were normalized using pyramid decomposition[167]. Each of the nodes in the trained SOM is assigned with an index number such that each block of a fingerprint image sample can be assigned with its corresponding best-matching unit's index number.

This is used to construct a vector of indices representing the finger image. The normalized histogram for this vector is computed, i.e., another feature vector with the length of the number of SOM nodes, containing the normalized histogram is constructed. The analysis of the histograms of low and high quality samples has shown that they are different for each of the two groups.

The histogram feature vectors are used for training a supervised Random Forest (RF)[164] model to classify them into five quality classes, determined by the produced comparison scores. Once the two models (the SOM and the RF) are trained, the same feature extraction procedure is done using the SOM model for the validation set and the constructed feature vectors are used to classify the samples into one of the five quality classes by using the RF model. The procedure was applied on a NIST operational dataset and has shown very good results [3], surpassing the performance of the state of the art method Orientation Flow defined in Ref. [82].

8.3 Machine Learning Techniques

8.3.1 Self-Organizing Maps

Self-organizing maps (SOM) have been introduced by Kohonen[171] and are a biologically inspired unsupervised learning neural network model [172]. It is a dimensionality reduction technique, similar in principle to vector quantization, where a number of codebook vectors $\mathbf{m}$ are fitted to represent some larger list of data vectors $\mathbf{x}$. Unlike vector quantization, SOM maintains globally and spatially ordered codebook vectors, usually forming a two-dimensional grid for visualization purposes. The dimensionality of the codebook vectors is the same as the dimensionality of the data and a distance metric is defined, e.g. Euclidean distance:

$$d(\mathbf{x}, \mathbf{m}) = \|\mathbf{x} - \mathbf{m}\| = \sqrt{\sum_{i=1}^{|\mathbf{x}|} (x_i - m_i)^2} \tag{8.1}$$

After the map has reached in the unsupervised learning phase a state of convergence, a non-linear mapping of the high-dimensional data space to the lower-dimensional grid space is achieved. The training data topology is preserved as much as possible by the resulting network (map), where each data sample's best-matching unit (BMU) is determined as follows:

$$c = \underset{i}{\operatorname{argmin}} \{\|\mathbf{x}(t) - \mathbf{m}_i(t)\|\} \tag{8.2}$$

The index t is used to denote the different data samples for $\mathbf{x}(t)$ and the state of the codebook vector $\mathbf{m}_i(t)$ at the moment t of processing data sample $\mathbf{x}(t)$. By considering the coordinates on the map of the sample's BMU, the dimensionality of the data is reduced to the dimensionality of the network lattice. These coordinates could be used as a new set of features themselves.

The network grid arrays could be constructed in a couple of different ways: regular grid array, toroidal or spherical, where the first one is the most commonly used. Depending on the number of node neighbors, a regular grid map could be rectangular (four neighbors) or hexagonal (six neighbors). Hexagonal maps are more accurate and better suited for visualizations, therefore they are recommended [175].

Before the training process a SOM network should be initialized and that could be done either randomly, by randomizing the codebook vectors, or linearly, by spanning the

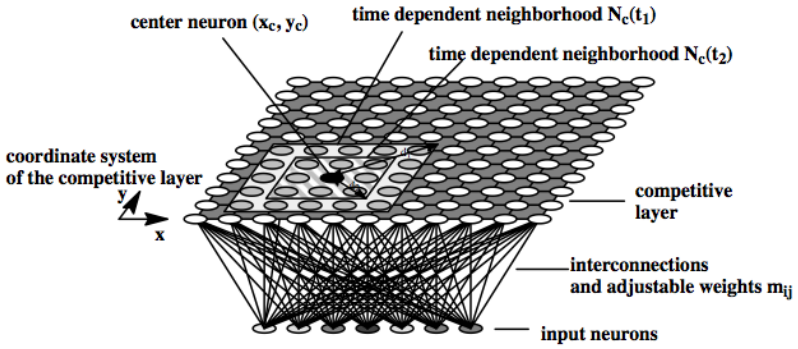


Figure 8.1: Best-matching unit is determined using Eq. 8.2, then the neighborhood is determined by Eq. 8.3 and the winner node and its spatial neighbors get updated.

network grid along the first two principal components of the training data. The training process of a SOM map is controlled by the neighborhood function:

$$h_{ci}(t) = \alpha(t) \exp\left[-\frac{\|\mathbf{m}_c(t) - \mathbf{m}_i(t)\|^2}{2\sigma^2(t)}\right] \tag{8.3}$$

where $\alpha(t)$ is a monotonically decreasing learning rate function affecting the elasticity of the memory and $\sigma(t)$ is a monotonically decreasing radius function, used to determine the neighborhood around the BMU to be updated. This neighborhood is composed of all non-zero values of h_{ci} and when determined by Eq. 8.3 is called Gaussian neighborhood. There exists a bubble neighborhood, for which $h_{ci} = 1$ for a given range of neighboring nodes and $h_{ci} = 0$ otherwise.

The training process proceeds with updating each grid node at a step by using the equation

$$\mathbf{m}_i(t+1) = \mathbf{m}_i(t) + h_{ci}(t)\,[\mathbf{x}(t) - \mathbf{m}_i(t)] \tag{8.4}$$

There are two main SOM algorithms as explained in Ref. [175]. The first one is the stepwise SOM algorithm, where each data sample $\mathbf{x}(t)$ is treated as a separate step in a sequential process. The second algorithm is the batch SOM, which relies on availability of all the training data, therefore no learning rate function $\alpha(t)$ is needed in the neighborhood function $h_{ci}(t)$. A list of best-matched samples $\mathbf{N}_c$ is maintained for each model $\mathbf{m}_c(t)$, where references to all samples, whose BMU is c, are kept. The update of the models is performed in a batch update iteration for all the grid nodes at once, unlike the equential update iterations used in the stepwise SOM.

8.3.2 Generative Topographic Mapping

Generative Topographic Mapping (GTM) was introduced by Svensen[173] as a principled alternative to the SOM. It is a probabilistic counterpart of SOM, that models the distribution $p(\mathbf{t})$ of the D-dimensional data by using a number of L-dimensional latent variables $\mathbf{x}$, where usually $L = 2$ for purposes of visualization. The mapping from the latent space

to the data space is done by using the transformation $\mathbf{y}(\mathbf{x};\mathbf{W})$, where $\mathbf{W}$ is a parameter matrix of some non-linear model, e.g. artificial neural network[177]. Usually $\mathbf{y}(\mathbf{x};\mathbf{W})$ is chosen to be a generalized linear regression model[177] of the form $\mathbf{y}(\mathbf{x};\mathbf{W}) = \mathbf{W}\phi(\mathbf{x})$, where $\phi(\mathbf{x})$ consists of M fixed radial basis functions[177].

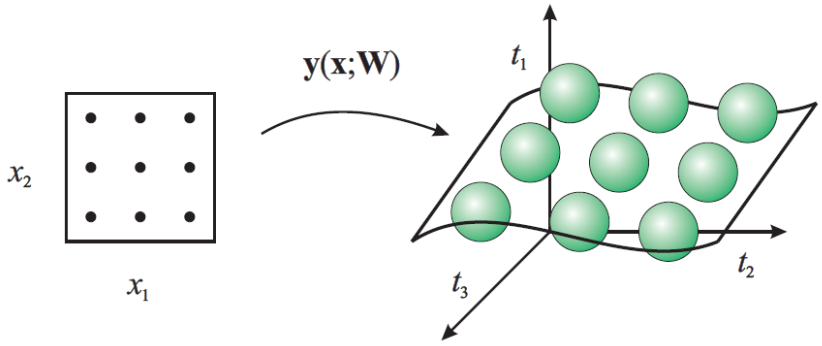


Figure 8.2: The discrete regular grid of nodes in latent space (left), mapped to the data space (right), where they correspond to centers of Gaussians with variance β^{-1}. Taken from Ref. [173].

Defining a probability distribution $p(\mathbf{x})$ in latent space gives a rise to a corresponding distribution $p(\mathbf{t} \mid \mathbf{W})$ in data space, which is confined to the image of the L-dimensional manifold in the data space under the mapping $\mathbf{x} \to \mathbf{y}$ of the non-linear function $\mathbf{y}(\mathbf{x};\mathbf{W})$. Since in reality the data can only approximately lie on a lower-dimensional manifold, a noise model for the $\mathbf{t}$ vector should be included. The distribution for $\mathbf{t}$, given $\mathbf{x}$ and $\mathbf{W}$, becomes radially-symmetric Gaussian centered on $\mathbf{y}(\mathbf{x};\mathbf{W})$ having variance β^{-1} so that

$$p(\mathbf{t} \mid \mathbf{x}, \mathbf{W}, \beta) = \left(\frac{\beta}{2\pi}\right)^{D/2} \exp\left\{-\frac{\beta}{2} \|\mathbf{y}(\mathbf{x};\mathbf{W}) - \mathbf{t}\|^2\right\} \tag{8.5}$$

In order to obtain the distribution in the data space for a given weight matrix $\mathbf{W}$, an integration over the latent distribution should be done

$$p(\mathbf{t} \mid \mathbf{W}, \beta) = \int p(\mathbf{t} \mid \mathbf{x}, \mathbf{W}, \beta) p(\mathbf{x}) \mathrm{d}\mathbf{x} \tag{8.6}$$

However, computing the integration would be infeasible, therefore a discrete form of $p(\mathbf{x})$ is taken, such that

$$p(\mathbf{t} \mid \mathbf{W}, \beta) = \frac{1}{K} \sum_{i=1}^{K} p(\mathbf{t} \mid \mathbf{x}_i, \mathbf{W}, \beta) \tag{8.7}$$

where all $\mathbf{x}_i$ are K nodes, centered on a regular grid in the latent space.

For a given data set the log likelihood is defined as

$$\mathcal{L}(\mathbf{W}, \beta) = \ln \prod_{n=1}^{N} p(\mathbf{t}_n \mid \mathbf{W}, \beta) \tag{8.8}$$

Determining the parameters $\mathbf{W}$ and β is done by maximizing the log likelihood, using an Expectation-Maximization algorithm (EM-algorithm)[178] performed in two steps, called respectively Expectation (E-step) and Maximization (M-step). The algorithm can be explored in details in Ref. Svensen-GTMgenerativetopographic-1998.

8.3.3 Random Forests

Random Forests (RF) was introduced by Breiman[164] as an ensemble machine learning method for supervised classification and regression. A Random Forest in the context of classification can be considered as an ensemble of decision trees[179]. For growing the decision trees, random feature selection is used for the split at any given node by choosing randomly m distinct features out of M features, where $m \ll M$. These m features could be repeatedly chosen among different splits.

The algorithm for constructing a random forest proceeds as follows. Firstly, a number of decision trees D has to be chosen. Then a dataset for each of these trees is bootstrapped by bagging[180]. Each of the trees is fully grown by random feature selection method and not pruned. Classifying an unseen sample involves classifying it using all the trees grown and labeling it with the most populous choice among the classifiers.

8.4 Methodology

8.4.1 Overview

The input of our proposed method is a list of gray-scale fingerprint images and their corresponding averaged genuine comparison scores as a column vector $\mathbf{y}$. For the purpose of modeling and testing our quality algorithm, the data is divided into training and validation sets. The algorithm is divided into two phases, a low-level phase and a high-level phase.

8.4.2 Low-level phase

For the low-level phase, the fingerprint images are divided into blocks and a SOM model is obtained with unsupervised training over the image blocks of the training set of images. Then by using the trained model, feature vectors are extracted for each fingerprint image in the dataset. The algorithm for training the model and extracting the features is as follows:

1. Given the list of training images and the block size b_{size}, all the training images are divided into square blocks of that size. Depending on the image size, different number of blocks are produced along each of the two dimensions $N_b^{(x)}$ and $N_b^{(y)}$.

2. For each block $b_I^{(x,y)}$, its fingerprint image source I and spatial index (x, y) are kept, so later this information is used in feature construction for the high-level modeling.

3. All image blocks are converted into row-vectors $\mathbf{v}_I^{(x,y)}$ of length $b_{size} \times b_{size}$ by taking the pixel values of the blocks row-wise.

4. All N_{train} training vectors are combined in a matrix T. No special ordering of the vectors is required, though random ordering is preferred.

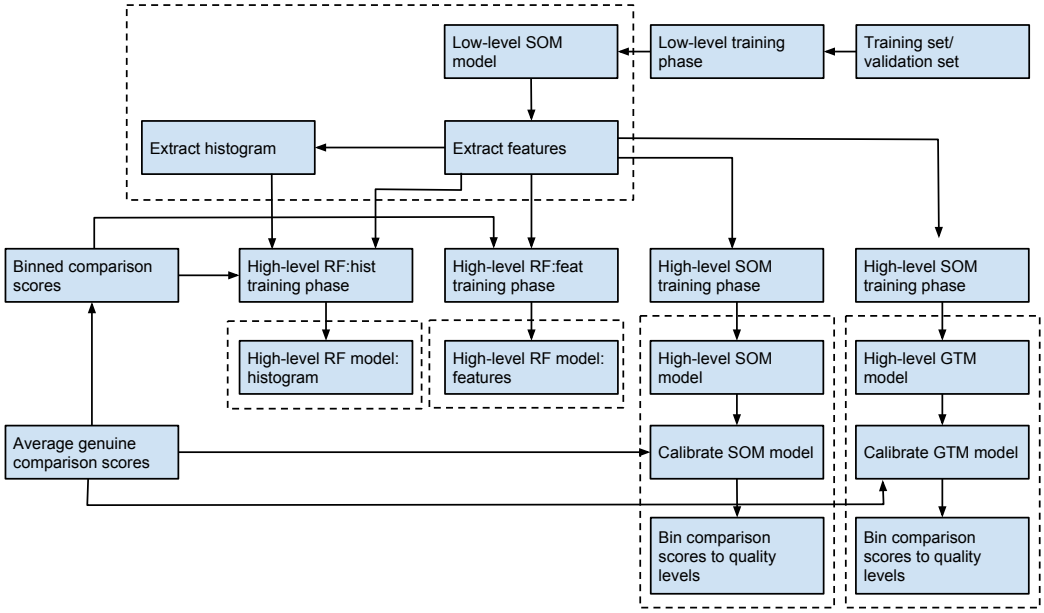


Figure 8.3: Overview of the training process pipeline. Each process or data source is separated in a blue block and the communication between processes is denoted with arrows. Some of the processes are grouped to form a big dashed blocks which are used for unseen data in the prediction mode.

$$T = \begin{vmatrix} \mathbf{v}_I^{(x,y)} \\ \mathbf{v}_I^{(x+1,y)} \\ \cdots \\ \mathbf{v}_{N_{train}}^{(N_b^{(x)},N_b^{(y)})} \end{vmatrix} \tag{8.9}$$

5. Linearly initialize SOM by doing a PCA of the data matrix T. For too big matrices to fit in the main memory, an iterative PCA algorithm [181] should be used. The number of map nodes M_{low} is determined heuristically by $M_{low} = \left\lceil \sqrt{N_{train}^{(blocks)}} \right\rceil$ in dependence of the total number of training image blocks $N_{train}^{(blocks)}$ and is upper bounded by a manually chosen number (e.g. 4000 in the current work), such that a trade-off between computational speed and SOM map smoothness for big datasets is achieved.

6. Stepwise SOM training process is performed with previously determined parameters: initial and final $\alpha(t)$ and $\sigma(t)$, which define the neighborhood function and the training length. These parameters could be either heuristically determined or computed by cross-validation over the training data.

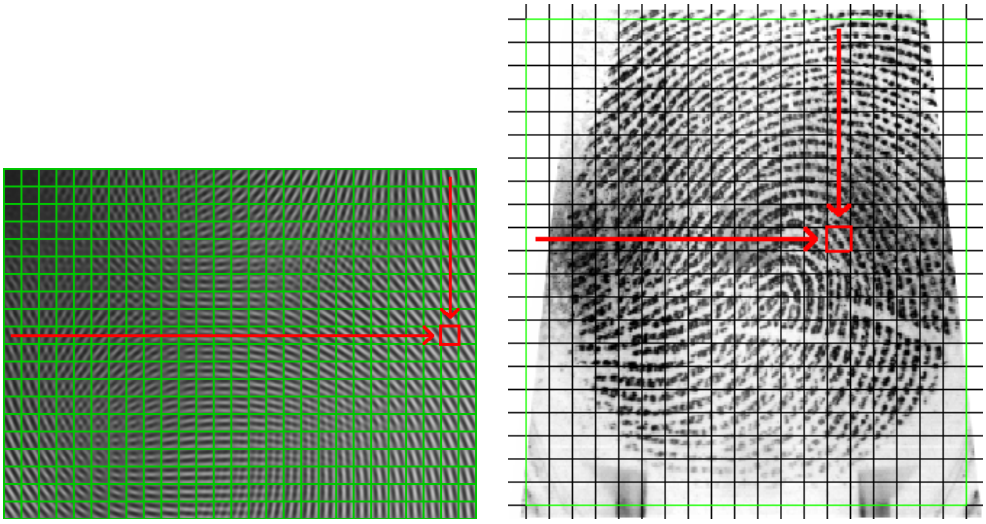


Figure 8.4: The best-matching unit on the upper part of the SOM map (left) and the block of the fingerprint image propagated through the map (right). The coordinates of the best-matching unit are saved in a coordinate row vector $\mathbf{c}_I^{(x,y)}$, indexed by (x, y), which are the coordinates of the block itself on the fingerprint image (right).

7. After the model has been trained, each block vector $\mathbf{v}_I^{(x,y)}$ is propagated through the network, so the best-matching unit is determined and its coordinates on the SOM map produce a two-dimensional coordinate row-vector $\mathbf{c}_I^{(x,y)}$ for each block of the image. See Fig. 8.4.

8. For each image I all blocks' coordinate vectors $\mathbf{c}_I^{(x,y)}$ are concatenated, where block indices in the image are taken row-wise. Thus for each image I a feature vector $\mathbf{f}_I$ of length $2 \times N_b^{(x)} \times N_b^{(y)}$ is constructed.

The output of the low-level training is the low-level SOM model, that is used to extract feature vectors $\mathbf{f}_I$ for all images. These feature vectors are used as input by the high-level phase for predicting quality scores.

8.4.3 High-level phase

Mapping feature vectors to quality scores itself requires the use of a machine learning model. In the current work, several approaches are proposed, discussed and evaluated in a benchmark.

Quantization of the results of the comparison algorithms is done using a binning procedure for converting from genuine comparison scores into quality scores. A cumulative distribution function $C(.)$ of the observed genuine comparison scores of the training set is used[70]. Then a number of levels L is chosen and $L-1$ thresholds in the range $0-1$ are defined. Using the inverse cumulative distribution function $C^{-1}(.)$, these fraction thresholds

are converted into genuine comparison score thresholds. Having these thresholds and a given genuine comparison score, a corresponding quality may be obtained.

8.4.3.1 SOM modelling

One approach for the high-level phase is to train a SOM model for modeling the feature vectors $\mathbf{f}_I$ input data. The relatively small number of components of the feature vectors makes possible the usage of batch SOM algorithm for the training. Thus only the parameters training length and neighborhood range function $\sigma(t)$ are adjusted. The SOM model is linearly initialized and the number of nodes M is heuristically determined by $M = \lceil 5 \times \sqrt{N_{train}} \rceil$.

After the training is done, each node of the constructed SOM map is labeled with the mean average genuine comparison scores of the image samples mapped to it. This labeling process is called SOM map calibration. By propagating a feature vector through the constructed SOM map and assigning an estimated average genuine comparison score, a quality score is determined using the already defined binning procedure.

8.4.3.2 Generative topographic mapping

A different approach is modelling the feature vectors $\mathbf{f}_I$ data by GTM. Parameters like the number of grid nodes K and the number of radial basis functions M are determined heuristically. σ and λ are also determined in this way, while $\mathbf{W}$ and β are estimated by the EM-algorithm.

The training process is the process of estimating $\mathbf{W}$ and β, given all other parameters and the training data feature vectors. The final GTM model is constructed once the EM-algorithm converges and after that a calibration of the GTM map is performed. Since GTM is a mixture of Gaussians, each node $\mathbf{x}_i$ of the GTM map is labeled by computing its expected average genuine comparison score by calculating the Gaussian components' means z_i using

$$z_i = \frac{\sum_n s_n p(\mathbf{x}_i \mid \mathbf{t}_n, \mathbf{w}, \beta)}{\sum_n p(\mathbf{x}_i \mid \mathbf{t}_n, \mathbf{w}, \beta)} \tag{8.10}$$

where s_n is the average genuine comparison score for training sample $\mathbf{t}_n$ and n is iterating over all N_{train} training samples. After having all z_i values, one could predict the expected conditional mean $y(\mathbf{t}_n)$ of an arbitrary sample $\mathbf{t}_n$ by computing the convex combination of the components means z_i in the output space, using its posterior values

$$y(\mathbf{t}_n) = \sum_i^M z_i P(\mathbf{x}_i \mid \mathbf{t}_n, \mathbf{W}, \beta) \tag{8.11}$$

where $P(\mathbf{x}_i \mid \mathbf{t}_n, \mathbf{W}, \beta)$ is the calculated responsibility R_{in}. This way an estimated average genuine comparison score $y(\mathbf{t}_n)$ is calculated and could be converted into a quality score using the defined binning procedure.

8.4.3.3 Random forests classifiers based on features

Another approach for the high-level phase is to use RF to model the data. RF is a supervised learning model and it requires not only the feature vectors, but also the output quality levels of the data prior the training. Since different comparison algorithms result in different average genuine comparison scores for the training samples, for each comparison algorithm a new model should be trained. A classification RF is used, so the binning of the genuine comparison scores is done prior the training as well. Then a RF model is trained, where the parameters m and n are defined by either cross-validation or heuristically.

8.4.3.4 Random forests classifiers based on histogram

The feature vector $\mathbf{f}_I$ is only one way of representing the information retrieved by the low-level phase. Another way is by constructing a histogram of the SOM hits for each image. Each node of the trained low-level SOM map is numbered with an unique index $1 \dots M_{low}$, where numbering the nodes is done column-wise starting from the top-left node. Then each pair of SOM map coordinates in $\mathbf{f}_I$ is translated into index, resulting in a vector of indices ind_I of length $N_b^{(x)} \times N_b^{(y)}$. Then a histogram $\mathbf{h}_I$ of the index numbers is created with bins $1 \dots M_{low}$.

Using the histogram feature vector $\mathbf{h}_I$ for training the RF classification model could be considered as another high-level modelling approach. This approach is the same as the one described in Sec. 8.4.3.3, except that the input data for this one is the histogram vectors $\mathbf{h}_I$, but not the feature vectors $\mathbf{f}_I$. By using the histogram features the information of the spatial coordinates of the blocks is lost and only their best-matching units' indices are considered. However it is not trivial to state whether this information is important or not for constructing better models, therefore the two approaches are examined in the current research.

8.5 Experiments and Results

8.5.1 Dataset

The proposed approach was benchmarked on a publicly available dataset CASIAFPV5, which was separated into training and validation sets.

CASIA Fingerprint Image Database Version 5.0 (CASIAFPV5)[140] is collected by the Chinese Academy of Sciences' Institute of Automation and all fingerprint images are 8-bit gray-scale BMP files with resolution 328×356. The dataset consists of 20000 fingerprint samples, taken from 500 volunteers contributing 40 fingerprint images each. 5 fingerprint samples were captured from all the fingers, except the little finger, on both of their hands using an URU4000 optical fingerprint sensor in a single session. The fingerprint samples are with significant intra-class variations, caused by rotation, translation and high or low pressure. The training set is composed by choosing randomly 12000 fingerprint samples of the dataset and the rest 8000 are used as a validation set.

8.5.2 Low-level model

Training for the low-level is done using the training set of CASIAFPV5. Two models are constructed: one for block size 16×16 pixels and one for block size 24×24 pixels. After the images are cut into blocks, the number of 16×16 training block samples for CASIAFPV5 is 5280000. The number of 24×24 training samples is 2184000.

The training is performed using SOM_PAK[182] with starting parameters $\alpha(0) = 0.05$ and $\sigma(0) = 2 \times unit_{dist}$ for a Gaussian neighborhood function, where $unit_{dist}$ is calculated after the linear initialization of the SOM map by taking the diagonal distance of the map in data space and dividing it by the number of nodes along the longest side of the map.

The total number of nodes M_{low} is determined heuristically by $M_{low} = \left\lceil \sqrt{N_{train}^{(blocks)}} \right\rceil$ in dependence of the number of training image blocks $N_{train}^{(blocks)}$ and is upper bounded by a manually chosen number (4000 in the current work), such that a trade-off between computational speed and SOM map smoothness for big datasets is achieved. The number of nodes along each of the dimensions is calculated respectively to the eigenvalues of the first two principal components. The number of training iterations for block size 16×16 is set to 5000000 and for block size 24×24 it is 2100000.

Analyzing the results from the training is done by observing the constructed SOM maps in Fig. 8.5. Each of the nodes of these maps is a data vector, that could be transformed back into image block. Each of the SOM map nodes has different number of blocks mapped to it and analyzing the number of mapped blocks to each node is done using hitmaps. However, the data from these experiments is not smoothly distributed over the SOM map, therefore a log-hitmap is considered as well.

Observing the results from log-hitmap, we see that much of the data is concentrated in the parts of the SOM map representing image blocks that stem from the background area of the fingerprint sample. This is caused by the fact that there are many more background blocks in the training data than foreground blocks. Moreover, the background blocks differ mostly in mean gray level intensity. The result is that much of the background variance is captured in the training process. The background area is mostly white, whereas the fingerprint area is darker, causing the first principal component to be the brightness of the blocks. Since the SOM maps are linearly initialized, the overrepresentation of background blocks distorts the eigenvalues of the principal components, thus the elongated shape of the SOM is not reflecting the real foreground data distribution.

8.5.3 High-level model

After the low-level model is constructed, both the training and the validation sets are propagated through the trained SOM maps in order to compose the feature vectors that are to be used by the high-level phase of the process. For CASIAFPV5 these feature vectors are of length 880 and 364 for 16×16 and 24×24 respectively.

For obtaining results and converting them to quality scores, an average genuine comparison score is determined for each fingerprint sample by using a single comparison algorithm as a black box for computing all possible genuine comparison scores per image and taking their average value. For this work a selection of three top-performing comparison algorithms has been used, anonymized as {A}, {B} and {C}, so each sample will receive three average genuine comparison scores. Therefore the unsupervised models (SOM and

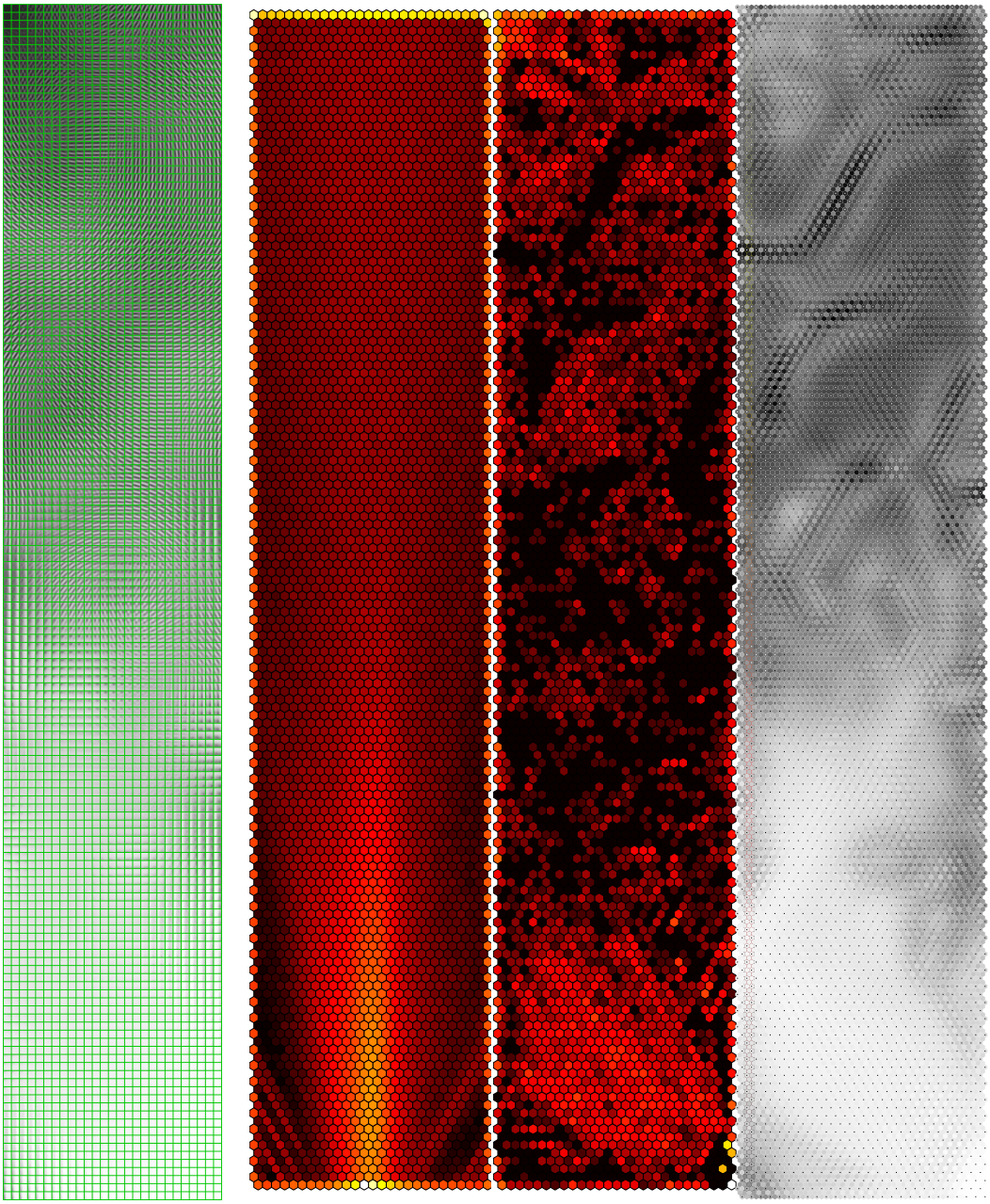


Figure 8.5: From left to right: The trained low-level SOM map, the initial log-hitmap, and the final log-hitmap for the validation set, the distance matrix U-mat. The trained SOM map is hexagonal, since that ensures more accurate visualizations of the hitmaps and the distance matrices. The dataset used for training is CASIAFPV5 and the block size is 16×16.

GTM) have to be calibrated separately for each comparison algorithm and the supervised models (RF-feat and RF-hist) have to be trained for each comparison algorithm separately. The performance evaluation of all high-level approaches is done by using ERC curves and Correlation tables for each comparison algorithm alone.

8.5.3.1 SOM

Since for the high-level SOM training the batch SOM algorithm is used over a linearly initialized map, only the number of batch iterations and the initial $\sigma(0)$ are to be set. The number of batch iterations is set to 30 and $\sigma(0) = 2 \times unit_{dist}$, where $unit_{dist}$ is calculated as noted in Sec. 8.5.2. After the model is trained, three calibrations are done for all three comparison algorithms and evaluated using the validation set of the data.

As it could be seen from the prediction map, the expected average genuine scores are not clustered, but scattered across the map. The same is true for the variance map. That makes the maps harder to analyze and might be caused by poor training process.

The SOM hitmap shows strong bordering effect, meaning that the data is concentrated in the edges of the map. As it seen int the prediction maps as well, there are some nodes which do not have any samples assigned. That might suggest the need for better training parameters or even different map topology, e.g., toroidal SOM, in order to be able to capture the data distribution better.

8.5.3.2 GTM

For constructing the GTM model the EM-algorithm is used to estimate the parameters $\mathbf{W}$ and β. The size of the grid of sample points is heuristically set to 20×20, resulting in $K = 400$ sample points $\mathbf{x}_i$. The form of the RBF grid is set to 9×9 RBFs, so the number of radial basis functions used is $M = 81$. The other two parameters, the common width of the radial basis functions σ and the regularization term λ, are determined heuristically as well and set to $\lambda = 1$ and $\sigma = 2s$, where s is the spacing of the radial basis functions centers. Similarly to the previous Sec. 8.5.3.1, three different calibrations are done for the different comparison algorithms.

The prediction maps here contain more homogeneous regions and the low-valued nodes are grouped together, unlike in the SOM map. That might suggest that GTM is able to capture the data distribution better. Another figure showing how the data is captured by the model is the log-hitmap, where the data distribution could be analyzed by observing the concentration of the data posterior modes.

8.5.3.3 Random Forests: features and histogram

For constructing the RF model, two parameters are to be set. The first one is the number of trees D to be grown, where each tree uses $n = N$ bootstrapped samples. The second parameter is the number of features to be chosen at each split m. For both RF-feat and RF-hist, these parameters are the same and are set to $D = 500$ and $m = 100$. The only difference between RF-feat and RF-hist is that the first one uses the feature vectors extracted from the low-level as input data, where the second one uses a histogram of these features. The performance of the random forests classifiers could be evaluated by observing the confusion matrices for each of the models.

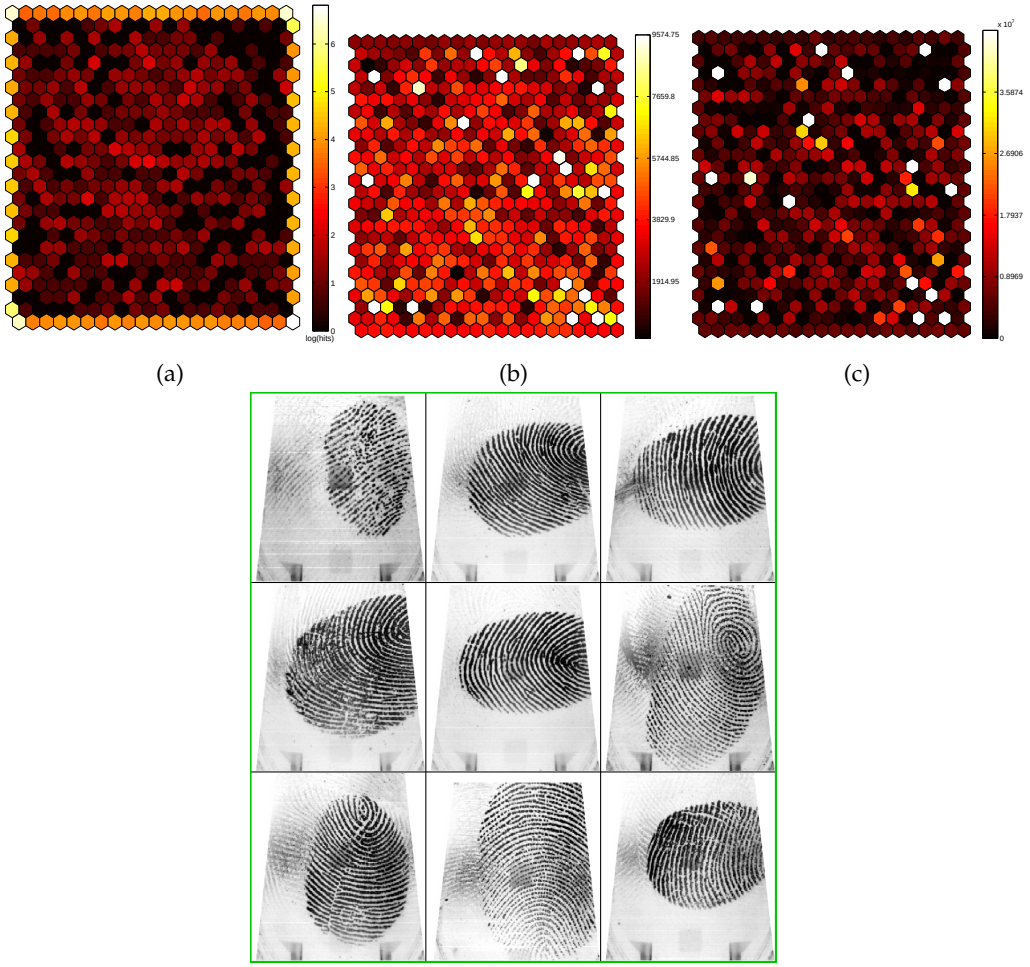


Figure 8.6: The trained high-level SOM log-hitmap (a) for CASIAFPV5 and block size 16×16 with the assigned prediction (b) and its variance (c) of the average genuine score values for each node. (d) contains examples of images associated with the top left node in the SOM. This map is calibrated for provider $\{A\}$.

8.5.4 Final results

In order to evaluate the performance of the proposed approaches, the Error versus Reject Curves (ERC)[112] and the Spearman[183] correlation matrices should be analyzed. ERC curves are an approach to visually evaluate how efficient a given algorithm is for rejecting samples with low genuine comparison scores and Spearman correlation matrices show how similar different algorithms are by measuring their Spearman correlation values. Since the quality levels used by the RF approaches and NFIQ are quantized into five levels, a jittering is added for drawing them. In order to eliminate the fluctuations caused

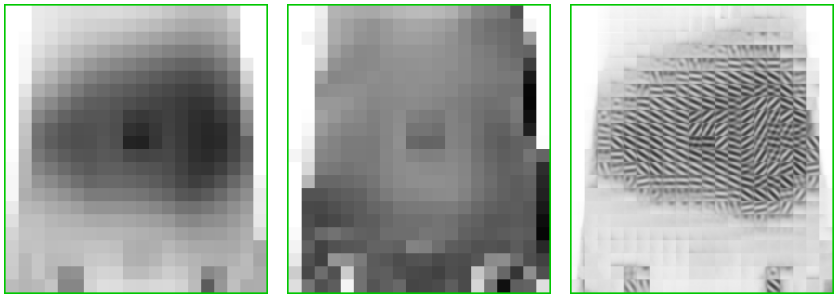


Figure 8.7: Features forming the top leftmost node of the trained high-level SOM map for dataset CASIAFPV5 with block size 16×16.

by the randomness, 20 ERC curves are calculated for each approach and the mean curve is drawn in the figure. Area under the curve is computed and the area of the perfect curve is subtracted, so the approaches could be evaluated quantitatively how close they are to the ideal curve.

The resulting ERC curves for the provider $\{A\}$ is shown in Fig. 8.10. These results show that the worst performing approach is using SOM for the high-level training. Second worst is the high-level approach used in Ref. [3], which is the RF model trained with the histogram as an input, rather than using the extracted features directly. The other two methods are performing quite similar quantitatively, however the GTM approach is better than the RF-feat approach if only the worst 20% of the samples are rejected. The results for providers $\{B\}$ and $\{C\}$ are similar to the ones shown in Fig. 8.10.

Fig. 8.11 shows the Spearman correlation table for the dataset CASIAFPV5 with block size 16×16. From this table it is visible that NFIQ shows highest correlation with the average genuine comparison scores of the samples. OCL and the two RF methods show also quite high correlation with the comparison scores and the NFIQ as well. On the other hand, the GTM approach is very highly correlated with the SOM approach and poorly correlated with the comparison scores. Looking at these table, it is easily visible that the SOM and GTM approaches are somehow related, the same is true for the two RF approaches.

8.6 Conclusions and Future Work

The PCA analysis done for linearly initializing the low-level SOM shows that the first principal component, being the brightness of the block images, explains more than half of the variance in the data. Analyzing the trained SOM shows that at least one third of that variance is contributed by variance in the background blocks. However, the variance information in the background blocks is of limited use as comparison algorithms usually segment the background out prior to feature extraction. Using only foreground information for the low-level training can result in smoother and concise SOM, thus better final results are expected.

The ERC evaluation of different high-level approaches for dataset CASIAFPV5 has shown good results for the methods GTM and RF-feat, which have outperformed the proposed in Olsen et al [3] feature interpretation method RF-hist. Thus the loss of the information about block position in the fingerprint image with constructing histogram features

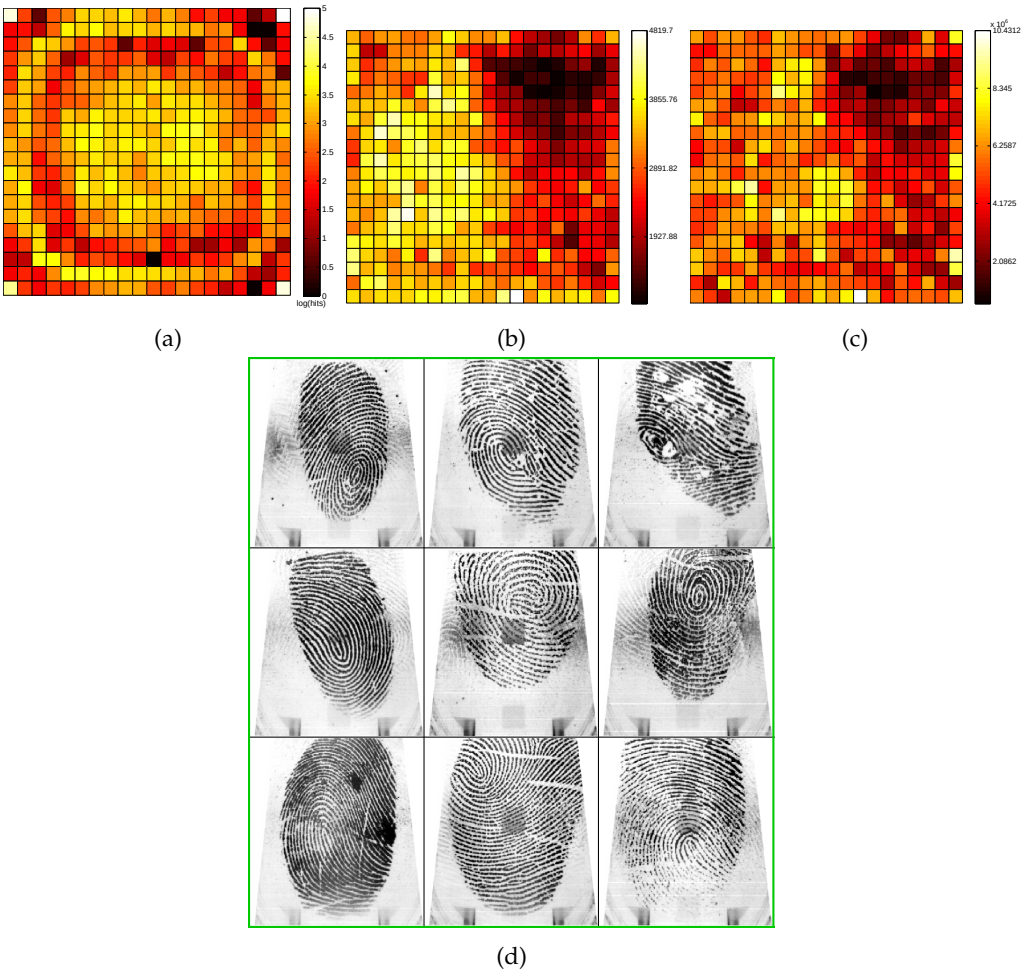
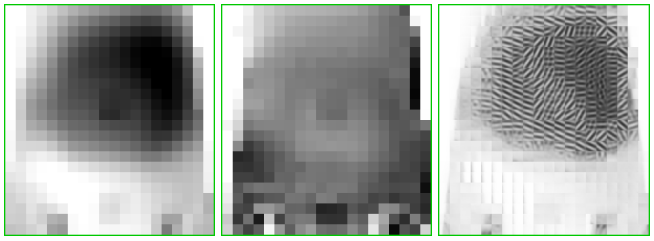


Figure 8.8: The trained high-level GTM log-hitmap (first) for CASIAFPV5 and block size 16×16 with the assigned prediction (second) and its variance (third) of the average genuine score values for each node. This map is calibrated for provider $\{A\}$ and the posteriors' modes are counted as hits.

seems to be crucial for dataset CASIAFPV5, but stating that these observations are valid for all datasets, requires more experiments with better parameter selection for the high-level models.

Operational speed measurements for the high-level phase were not performed, since the propagation time for all the high-level methods is negligible compared to the time consumed by the low-level phase. Thus the gain of optimizing the low-level phase is much bigger and should be first explored. However the RF training requires much longer time, compared to the other two approaches - SOM and GTM. Additionally, for each comparison algorithm a different model should be trained labelled with the binned average genuine comparison scores, unlike the case with the SOM and GTM maps, which only need to be trained once and later calibrated.



(a) vertical (b) horizontal (c) blocks

Figure 8.9: Features forming the top leftmost node of the trained high-level GTM map for dataset CASIAFPV5 with block size 16×16.

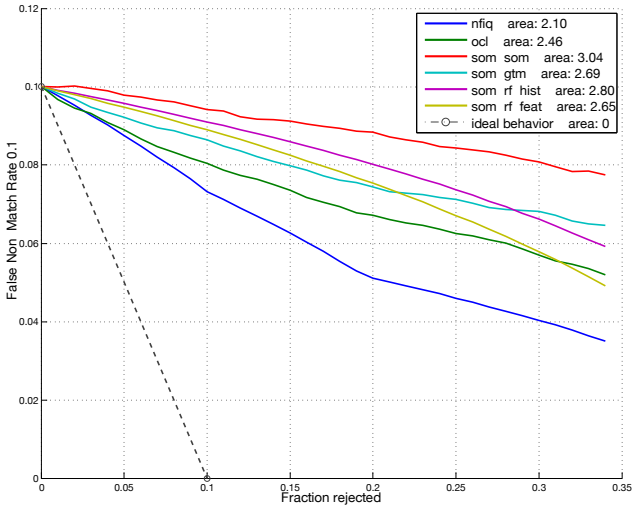


Figure 8.10: ERC curves for CASIAFPV5 with block size 16×16 for provider {A}.

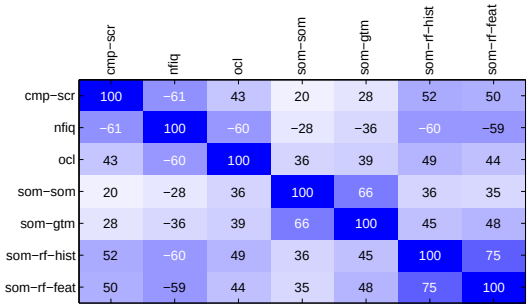


Figure 8.11: Spearman correlations for CASIAFPV5 with block size 16×16 for provider {A}.

According to the ERC plots, GTM and RF-feat have outperformed the previously suggested high-level method RF-hist for CASIAFPV5. The resulting low-level map for block size 24×24 seems to be undertrained, thus more training cycles are needed for this block size. Additionally much better model parameter selection could be performed for both low- and high-level models and this will yield more robust and stable results, that could be used as confirmation to the drawn conclusions. For the low-level phase, training without the background blocks is expected to construct a much finer low-level model, thus improving the results of all high-level methods.

Acknowledgments

This work has been possible thanks to the support of CASED.

Chapter 9

Fingerprint Skin Moisture Impact on Biometric Performance

Information about the paper

One challenge in developing new algorithms for fingerprint recognition systems is the availability test and validation on samples acquired under non-ideal conditions. This paper investigates the relationship between moisture levels and biometric performance and provides a new dataset with ground truth measurements of fingerprint moisture levels prior to acquisition. The paper is a contribution towards research question $\mathcal{RQ}_4$. This original paper was published in: [7] M. A. Olsen, M. Dusio, and C. Busch. "Fingerprint Skin Moisture Impact on Biometric Performance". In: *International Workshop on Biometrics and Forensics 2015*. 2015.

Abstract

Dry or wet fingers have an impact on biometric performance and the measurement of skin characteristics is important for actionable quality. We collected a ground truth fingerprint skin moisture dataset containing 6600 samples from 33 subjects using 5 optical sensors, two claimed to work well in extreme skin moisture conditions. Controlling conditions by applying cream, alcohol and water we performed objective measurements of skin moisture prior to each fingerprint capture. Analyzing the relationship between measured moisture and biometric performance across five sensors and comparison scores from 3 commercial providers confirms claims to resistance to excessive moisture. Results indicate that sensors resistant to excessive moisture perform worse on dry fingerprints than sensors not resistant to excessive moisture. The collected dataset is made available as an in-house dataset.

9.1 Introduction

There is a common agreement in government, industry and academia that the quality of a fingerprint sample has an impact on its recognition performance as Grother and Tabassi discuss [112]. ISO/IEC 29794–1 specifies biometric sample quality as a biometric performance predictor [70]. Grother and Tabassi show [112], the accuracy of a biometric system can be improved by rejecting poor quality samples, resulting in lower False Non-Match Rate (FNMR) and hence an increase in recognition performance as discussed by Wilson *et al*. [120].

Recognition performance may suffer due to poor quality of fingerprint samples caused by one or several factors: *a)* skin defects: blisters, wrinkles or scars; *b)* skin conditions: roughness, dryness or wetness; *c)* user behaviour: bad finger alignment, too little or too much pressure; and *d)* environmental conditions: dirty sensor platen, extreme temperature or lighting. If an acquired sample exhibits poor quality, skin conditions and user behaviour may to some extent be altered if the subject is given sufficient feedback or another finger may be used; then the acquisition process can be repeated.

Several of these defects may be addressed by actionable quality. Providing actionable quality is desirable as information pertaining to how the quality of the fingerprint may be improved in subsequent captures is conveyed. This is applicable in both supervised and unsupervised modes as a means of assistance to the supervisor or simply decreasing the number of recognition or enrolment attempts, leading to improved system performance and better user experience.

The main contribution of this paper is *1)* acquisition of a fingerprint sample dataset with ground truth moisture measurement, available in a semi-public fashion [122]; *2)* analysis of fingerprint skin moisture on recognition performance.

Addressing the lack of ground truth data on fingerprint defects we collect a dataset of fingerprint samples with ground truth measurements of fingerprint skin moisture during acquisition. To our knowledge there exists no dataset offering such ground truth information. Subjective manual mark-up of samples as dry or wet suggested by Shen *et al*. [43], Lim *et al*. [71], Alonso-Fernandez *et al*. [97] and Maltoni *et al*. [49] does not necessarily convey information about skin moisture but categorizes the sample impression, in fact it may not convey moisture at all as shown in Figure 9.1 where the fingerprint impression

of a wet finger and that of a finger forcibly pressed against the platen show very similar characteristics, e.g. overall darker gray levels and spurious ridge line interconnections. Blomeke *et al*. [62] investigated relationships between fingertip skin moisture, oiliness, elasticity, temperature and that of a proprietary quality measure of acquired fingerprints, where it is shown that the characteristics are weakly correlated and proposed that the relationships may be non-linear in nature and warrant further investigation. Sickler *et al*. [54] investigated quality and moisture levels in relation to subject age finding that there is a relation.

We focus on samples acquired using optical fingerprint sensors which is the capture device technology most commonly used and offering highest resolution according to Maltoni *et al*. [49] and used in the Office of Biometric Identity Management (OBIM) program, formerly known as US-VISIT, by Department of Homeland Security [170] and in Aadhaar by Unique Identification Authority of India [184].

The rest of the paper is organized as follows: we provide an overview of fingerprint sample quality in Section 9.1; dataset collection is described in Section 9.2; we evaluate the impact of skin moisture on biometric performance in Section 9.3 before concluding in Section 9.4.

Biometric fingerprint system performance is influenced by the quality of fingerprint samples utilized by the system. Sample quality depends on the scanned finger character and captured sample fidelity; it can be expressed in terms of utility, i.e. the observed performance in one or more systems as defined in ISO/IEC 29794–1 [70]. Generally fingers with clearly defined ridges scanned in suitable conditions provide better quality fingerprint images resulting in lower error rates and better performance. From the factors that influence the quality of a fingerprint sample, skin moisture is a factor that has strong influence (see Figure 9.1), however the moisture can be changed and samples reacquired.

Figure 9.1: Fingerprint samples acquired from same finger with skin subjectively perceived as (left to right): dry (62.5 %), wet (99.9 %), normally moist (84.6 %) and normally moist (84.6 %) with finger strongly pressed on the sensor.

The general term *sample quality* may describe several attributes of a biometric sample yet provide no particular meaning with respect to a concrete biometric system, e.g. a good sample with clearly defined and sharp ridges featuring little or no minutiae will compare

poorly in a minutiae-based comparison, as Grother and Tabassi discuss [112]. Hence the international standard ISO/IEC 29794–1 requires a biometric fingerprint sample quality metric to predict its biometric performance [70].

Based on the requirements in the aforementioned international standard, the state of the art fingerprint sample quality algorithms, e.g. those specified in ISO/IEC 29794–4 [99] and those which are candidates considered for NFIQ 2.0 [8], the successor of NIST Finger Image Quality (NFIQ) [79] quantify sample x quality q via image analysis as $q = \mathrm{QMA}(x)$.

Fingerprint quality is used in Automated Fingerprint Identification Systems (AFIS) to increase system performance by rejecting poor samples. However, they do not explicitly provide information about what actually caused their positive or negative quality assessment, which would be crucial information for improved AFIS user experience and helpful in dataset survey and diagnosis. One reason for this is that ground truth data on particular defects is not available, and as such the effectiveness of methods in detecting these defects cannot be determined reliably.

9.2 Dataset collection

A dataset of 6600 fingerprint samples is collected from 33 subjects using all 10 fingers (see Table 9.1) in four different skin moisture conditions, using 5 biometric fingerprint sensors and registering fingertip skin moisture during acquisition [122]. Prior to sample acquisition, the fingertip is measured using "Moisture Meter SK–3" skin moisture measuring device produced by Shenzhen Kier Electronic Apparatus Factory [185]. It determines a percentage score that conveys objective skin moisture. $99.9\,\%$ is highest possible measurement, and $41.3\,\%$ was lowest encountered in the data collection. In the rest of the paper we will refer to the percentage score as an integer value in the range 0 to 999, e.g., $99.9\,\%$ will be expressed as 999. Skin moisture measurement devices appear to be using either inductance or capacitance methods for the measurement. Alanen *et al*. [186] and Clarys *et al*. [187] show that both methods have very high correlation providing almost identical results and are more sensitive to water than mineral oil, although capacitance method carries information from deeper parts the skin ($45\,\mu\mathrm{m}$ vs. $15\,\mu\mathrm{m}$). The vendor claims to measure skin moisture via the inductance method.

Age		Ethnic origin		Sex	
Max	59	Caucasian	25	Female	16
Min	23	Asian	7	Male	17
Mean	33.5	Mid.Eastern	1		
Median	30				
Total					33

Table 9.1: Self-reported participant information in collected data set.

Capture devices used are all 500 dpi optical fingerprint sensors. We partition the sensors into two groups according to claimed resistance towards sample quality degradation caused by extreme skin moisture levels. Three sensors do not claim resistance to extreme moisture level: Cross Match Technologies LScan 100 ($\mathcal{S}_2$), L–1 Identity Solutions DFR–2100 ($\mathcal{S}_3$), CrossMatch Patrol ID ($\mathcal{S}_4$) and are referred to as $\mathcal{G}_1$; Two sensors do claim resistance

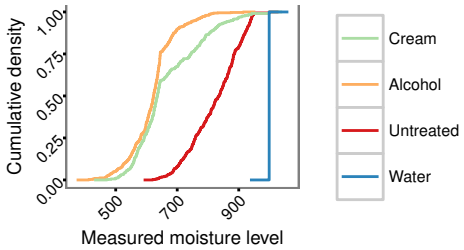


Figure 9.2: Empirical cumulative distribution of measured moisture levels for the 4 treatments fingers were subjected to, respectively cream, alcohol, water and untreated.

to extreme moisture levels: Dermalog ZF–1 ($\mathcal{S}_1$) and Cross Match Technologies Guardian ($\mathcal{S}_5$) and are referred to as $\mathcal{G}_2$ (see Table 9.2).

Sensor	Alias	Group
Dermalog ZF–1	$\mathcal{S}_1$	$\mathcal{G}_2$
Cross Match Technologies LScan 100	$\mathcal{S}_2$	$\mathcal{G}_1$
L–1 DFR–2100	$\mathcal{S}_3$	$\mathcal{G}_1$
Cross Match Technologies PatrolID	$\mathcal{S}_4$	$\mathcal{G}_1$
Cross Match Technologies Guardian	$\mathcal{S}_5$	$\mathcal{G}_2$

Table 9.2: Fingerprint sensors and associated alias and group.

9.2.1 Collection protocol

Fingerprints are collected in office conditions common for all subjects, applying invariant finger–sensor force during acquisition by instructing subjects to gently touch the sensor and placing a weight of $133\,\mathrm{g}$ on top of the fingernail, a weight empirically chosen as producing good sample impressions for non–extreme skin moisture levels. Fingerprints are acquired in four skin conditions: natural moisture, dried, moisturised and soaked in water. The dry conditions are simulated by gently cleaning skin with alcohol solvent ($68\,\%$ ethanol, $2\,\%$ salicylic acid, $30\,\%$ water), after the alcohol has evaporated; moisturising is performed by applying water–based body lotion (since the measurement device is insensitive if oil is applied); over moisturising is performed last by soaking the fingertip in pure water after the skin has been gently wiped with a paper cloth. Empirical testing of the device prior to dataset collection showed that soaking treatment always results in measurement of 999, thus for the fourth condition the measurement is skipped and 999 is registered.

The ECDF of the moisture levels for each of the four treatments is depicted in Figure 9.2. The samples treated with alcohol appear towards the lower end of the moisture range, while those treated with cream appear at a slightly higher lever. Untreated samples, i.e. samples acquired as the subject arrived for enrolment, have distinctly higher measured moisture levels than either of alcohol or cream treatments. Samples treated with water appear only at the highest moisture level.

For each acquisition of fingerprint samples with one finger and using all five sensors the moisture is measured once, prior to applying the finger to the first sensor. In order to reduce sensor bias, we randomize the order of sensors prior to the start of acquisition for each finger. Further we sought to keep the delay between acquisitions to less than four seconds.

Figure 9.3: Collected fingerprint sample excerpts. treatment from left to right: normal, dry, cream and water; top row shows samples acquired using a sensor from $\mathcal{G}_1$, bottom – $\mathcal{G}_2$. Measured moisture from left to right: 630, 610, 778 and 999.

Figure 9.3 shows examples of the influence moisture conditions have on fingerprint sample impression. Sensors in $\mathcal{G}_2$ appear to work well with over moisturised fingers; little difference between normal and dried condition impressions (first and second column) is due to already dry skin in the first measurement.

9.2.2 Partitions

We will assess the impact of moisture levels by determining recognition performance at several levels using intra sensor, intra provider comparison scores. Independent of the 4 treatments, we define 3 moisture groups as follows.

In dataset $\mathcal{D}$ we denote the kth impression belonging to the ith subject captured using sensor s as $d^k_{i_s}$, and the moisture measurement operation as M. The moisture level of $d^k_{i_s}$ is then:

$$m^k_{i_s} = M(d^k_{i_s}) \tag{9.1}$$

In a comparison with two samples there are two associated moisture measurements. We posit that the comparison score is driven by the sample with the highest moisture level. Thus, to perform a ranking we specify that the moisture level in a comparison is deter-

mined by the sample with the highest measured moisture:

$$m_n = \max(m_{i_s}^k, m_{j_t}^l) \tag{9.2}$$

With the moisture level expressed in the range 0 to 999, with 0 being no moisture and 999 being soaking wet, we define three disjunct sets of comparisons according to the ground truth measurements taken during the dataset collection containing comparisons with dry (D), neutral (N), and wet (W) samples:

$$D = \{m_n : m_n < 625\} \tag{9.3a}$$
$$N = \{m_n : 625 \leq m_n < 999\} \tag{9.3b}$$
$$W = \{m_n : 999 \leq m_n\} \tag{9.3c}$$

The range for W is determined as the highest measured value and the limit for N was set by visual inspection under constraint to achieve a set of similar size to W (circa $25\,\%$ of total). Due to the partitioning on moisture level, samples with different treatments might be present in any of the sets.

9.3 Results

For each of the sets D, N, and W we determine the impact of moisture level in context of sensor and provider by computing the DET curve for each provider and sensor combination. The number of acquisitions and comparisons in each set is summarized in Table 9.3. In the cases where a provider failed to produce a biometric template from the captured sample, we assigned that comparison with the lowest comparison score (i.e. 0). Provider A is generally achieving the best performance regardless of sensor and was also the only provider to produce templates for all acquired samples.

		Comparisons		
Set	Acquisitions	Genuine	Impostor	Failed
D	1536	35 724	104 047	692
N	3389	170 526	1 000 024	1362
W	1675	169 950	875 929	35 883

Table 9.3: Properties of ground truth sets for each provider. The Failed column indicates number of comparisons not performed due to Failure-To-Acquire by a provider and which were thus assigned lowest comparison score.

9.3.1 Performance according to measured moisture

The DETs are depicted in Figure 9.4 and show that sensors in $\mathcal{G}_1$ perform worse than those in $\mathcal{G}_2$ (left-most and right-most plots) when dealing with comparisons in W, confirming the statements by those vendors on resistance towards extreme moisture levels. While the overall performance across providers and moisture levels appears to be higher for $\mathcal{S}_5$, we see that for $\mathcal{S}_1$ the highest performance is achieved with comparisons from W using scores from provider A. Turning our attention to the 2nd, 3rd and 4th plots (sensors $\mathcal{S}_2$,

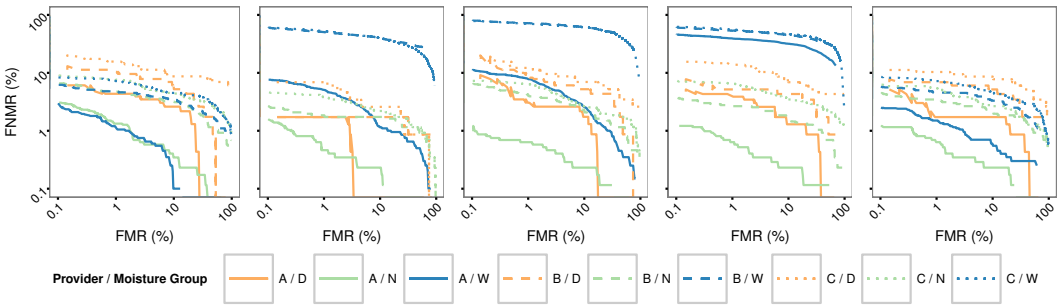


Figure 9.4: Detection Error Tradeoff (DET) plots for each sensor. Left to right: $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$, $\mathcal{S}_4$, $\mathcal{S}_5$. Each curve represents the DET of a provider (A, B, C) grouped by moisture levels D, N, W (see Eq.9.3 for definitions).

Provider	Moisture	$\mathcal{S}_1$	$\mathcal{S}_2$	$\mathcal{S}_3$	$\mathcal{S}_4$	$\mathcal{S}_5$
A	D	4.3	**1.7**	3.0	4.3	**1.7**
	N	1.3	**0.5**	0.6	0.6	0.6
	W	**1.2**	5.1	7.8	39.3	1.5
B	D	10.4	**1.7**	8.3	5.2	6.9
	N	4.7	**1.8**	4.0	2.7	3.5
	W	4.8	50.6	72.0	53.6	**4.6**
C	D	15.7	**6.9**	12.2	13.8	10.3
	N	7.6	**3.8**	6.4	6.5	5.4
	W	**7.2**	52.9	70.3	56.1	**7.2**

Table 9.4: FNMR in % at FMR=1% for each sensor and comparison score provider according to measured moisture group.

$\mathcal{S}_3$, $\mathcal{S}_4$), we see that the comparisons in W are generally associated with poor performance. Provider A (solid lines) extracts enough information from samples captured using $\mathcal{S}_2$ and $\mathcal{S}_3$ to achieve recognition performance similar to that of providers B and C when those consider only comparisons from D or N.

In Table 9.4 we show the FNMR at a fixed False Match Rate FMR $= 1\%$ to highlight the differences across sensors and comparison score providers. Between all sensors, $\mathcal{S}_2$ achieves the lowest FNMR in conditions D and N regardless of comparison score provider used. The lowest FNMR using W are achieved with provider A on samples captured by $\mathcal{S}_1$ closely followed by $\mathcal{S}_5$ with respectively $1.2\,\%$ and $1.5\,\%$. While $\mathcal{S}_1$ achieves lower FNMR in on the combination W and provider A, $\mathcal{S}_5$ has lower FNMR than $\mathcal{S}_1$ in all other combinations of moisture level and provider.

9.3.2 Moisture impact on genuine scores

We quantify the impact on genuine comparison scores by binning the measured moisture levels for probe and reference samples and computing the mean comparison score in each bin. The purpose is to determine if there exist ranges of moisture levels where higher

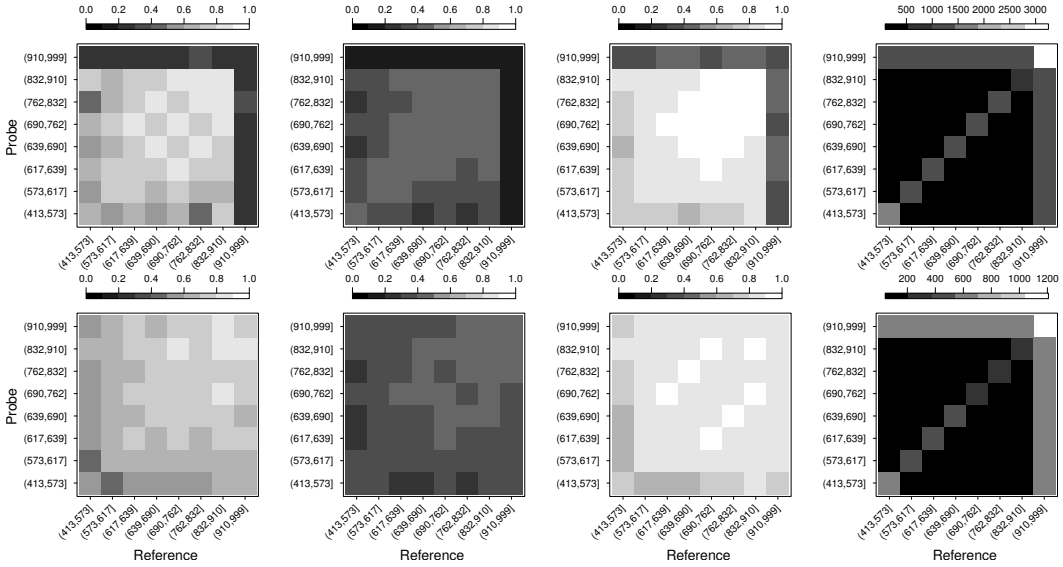


Figure 9.5: Mean genuine comparison score binned by moisture levels for reference and probe samples. Axes show measured moisture and bin gray level represents. Left to right: score from provider, A, B, C, number of samples in each bin. Top row using samples captured using $\mathcal{G}_1$ sensors, bottom row using $\mathcal{G}_2$ sensors. Scores have been scaled to $[0, 1]$ for each provider.

biometric performance can be expected, and to determine how sensor choice influences it.

In Figure 9.5 we divide the moisture range into 8 bins along the measured moisture level of probe and reference samples. For each of the resulting 64 bins we compute the mean comparison score. The initial grouping of sensors has been confirmed by results depicted in Figure 9.4 and we use those same groups to form the sets of samples used in comparisons (only considering scores from comparisons between samples from same sensor). Due to differences in comparison scores for each of the providers, we further group by provider. The right most plot in each row, depicts the number of samples present in a bin. Note that some bins in the lower left corner of the heat maps are depicted as nearly empty, meaning that very few comparison scores are used to form the average score in those bins. There is a large number of comparisons in the upper right corner due to nearly a quarter of the samples being annotated with highest moisture level.

For sensors in $\mathcal{G}_1$ (top row) the genuine comparison score is determined by the sample with the highest moisture level: when the reference sample has measured moisture above 910, the moisture level of the probe sample is inconsequential for the comparison score which stays low. The effect is clearly shown in the previous DET plots. For sensors in $\mathcal{G}_2$ (bottom row) we see that the measured moisture level does not affect the genuine comparison score in the same way as $\mathcal{G}_1$.

Both $\mathcal{G}_1$ and $\mathcal{G}_2$ mean comparison score is always lower when either reference or probe sample has moisture levels below 617 and when the measurements indicate moisture levels above 617 a higher comparison score can be expected. This indicates that our choice of 625 as threshold for set D is fairly close to observations in reduced biometric performance.

9.4 Conclusions

We collected a fingerprint dataset with ground truth skin moisture measurements which is made available to the community for on-site testing at our research group premises. We show that the choice of sensor as well as that of feature extractor and comparator has significant influence on expected performance depending on whether fingertip skin is at wet, dry or normal moisture levels. We demonstrate that sensors in $\mathcal{G}_2$ are resistant to excessive moisture levels in partition W as claimed by vendors. For samples in N and D which are not subjected to excessive moisture we found that one sensor from $\mathcal{G}_1$ performed better regardless of comparison score provider. For sensors in $\mathcal{G}_2$, improvements in biometric performance can be achieved by addressing dry fingerprints. Our ground truth skin moisture measurements indicate that moisture level shall be above 617 for good biometric performance regardless of sensor, and below 910 for $\mathcal{G}_1$ sensors. The collected dataset in this paper may be accessed by following instructions given online via da/sec [122].

Acknowledgements

This work has been partially funded by the Center of Applied Security Research Darmstadt (CASED). The authors would like to express gratitude to the participants of the dataset collection for their time and patience.

Chapter 10

Gabor Filters as Candidate Quality Measure for NFIQ 2.0

Information about the paper

Gabor filters are widely applied in fingerprint systems in recognition, reconstruction and quality areas. Here a variant of existing quality algorithms based on Gabor filters is proposed and evaluated. The proposed method is also evaluated in chapter 4. The paper is a contribution towards research question $\mathcal{RQ}_2$. This original paper was published in: [1] M. Olsen, H. Xu, and C. Busch. "Gabor filters as candidate quality measure for NFIQ 2.0". In: *Biometrics (ICB), 2012 5th IAPR International Conference on*. 2012, pp. 158–163. DOI: 10.1109/ICB.2012.6199802.

Abstract

Quality assessment of biometric fingerprint images is necessary to ensure high biometric performance in biometric recognition systems. We relate the quality of a fingerprint sample to the biometric performance to ensure an objective and performance oriented benchmark. The proposed quality metric is based on Gabor filter responses and is evaluated against eight contemporary quality estimation methods on four datasets using sample utility derived from the separation of genuine and imposter distributions as benchmark. The proposed metric shows performance and consistency approaching that of the composite NFIQ quality assessment algorithm and is thus a candidate for inclusion in a feature vector introducing the NFIQ 2.0 metric.

10.1 Introduction

In large scale Automated Fingerprint Identification Systems (AFIS), it is important to consider that there will be a certain fraction of individuals who will try to avoid detection. At the border controls of Japan and the United States of America, an individual can be rejected entry if the biometric probe sample captured from the individual matches (has a high comparison score) with a biometric reference already registered in a database, or a watch-list. Consequently, for negative biometric claims a situation can arise where an individual will supply a low quality probe sample on purpose, thus minimizing the chance of detection. Without a method of determining whether the quality of the captured probe sample reaching a sufficient level for recognition purposes, an individual can thus subvert the system. The scenario is substantiated by the findings in [78, 107, 120], where it was established that there is a strong correlation between fingerprint image quality and biometric performance.

Determining the quality of a fingerprint also finds use in other scenarios such as in the context of immigration where a subject can apply for a visa at the embassy or consulate for a given country. In order to verify that the identity of the subject at the border control is indeed the same as the one who received the visa at the embassy, a fingerprint capture is performed at the time of the application. Thus the subject is enroled in the biometric system and can be identified at a subsequent border control. In Europe such a system is implemented for all countries of the Shengen area and is known as the Visa Information System (VIS). In a system such as the VIS the cost of a false reject is high: the subject is either faced with a wasted travel expense and anguish over an unrightful rejection, or the border control will have to employ special procedures to verify the subject identity through other means.

This exemplifies that it is desirable to assess the quality of a fingerprint image before any enrolment transaction is initiated, so that for later comparisons, at least the enrolment sample is of suitable quality. High comparator performance is achieved if a fingerprint's quality is sufficiently good and overall database integrity is improved. This requires poor image samples to be rejected before they are enroled into databases. From an operational perspective, such *Failure To Capture* might trigger a re-capture procedure potentially with more support for the capture subject given by the operator personnel.

Here we present an approach based on Gabor filter responses and evaluate the performance relative to the metrics defined in ISO/IEC TR 29794-4 [99] and NFIQ [78]. The

findings presented herein are included in the development considerations of the forthcoming NFIQ 2.0 [188] algorithm, which will be the successor of the widely adopted NFIQ algorithm.

10.2 Background

If fingerprint image quality is to be a predictor of the biometric performance, then the quality metric should reflect the signal quality defining both the local and global characteristics in a fingerprint. Locally, the ridge orientation certainty level [71], local orientation [72], or the blockwise similarity to Gabor filter responses [75] can be applied as indicators of quality. In a similar manner, the discrete Fourier transform can be analyzed [83] to give an impression of the global quality. An extensive comparative study on quality metrics was performed by Alonso-Fernandez et al. [97]. Several of these metrics are included in the ISO/IEC technical report on biometric fingerprint sample quality [99].

Fingerprint image quality as a predictor of biometric performance was first published by National Institute of Standards and Technology (NIST) in the technical report Fingerprint Image Quality [78], which documents the development of the NIST fingerprint image quality (NFIQ) algorithm. Here fingerprints are classified into 5 quality levels, ranging from poor to excellent quality, based on several characteristics: ridge orientation flow, ridge orientation certainty, local ridge curvature, and local contrast. Using an interpretation of these characteristics, NFIQ relies on a neural network to perform the classification. In [132], several deficiencies in NFIQ have been identified, among those the important notion that the the definition of the image features used to predict the NFIQ leaves great optimization potential.

Our proposed method and methodology differs from that in [75] as we do not subdivide the input into blocks and we do not use a subjective measure to compare the performance of the algorithm. Instead we work in a pointwise manner and use an objective performance assessment, thus ensuring reproducability of results.

10.3 Fingerprint Quality Metrics

We propose a fingerprint quality measure based on the Gabor filter responses. In this paper we also evaluate several contemporary quality estimation methods, including *Orientation Certainty Level* (OCL), *Ridge-valley Structure* (LCS), *Ridge-valley Uniformity* (RVU), *Frequency Domain Analysis* (FDA), *Radial Power Spectrum* (POW) and *Orientation Flow* (OF).

10.3.1 Quality Metric Based on Gabor Filter Responses

A Gabor filter is a bandpass filter with local support that enhances some spatial frequencies with a certain orientation and attenuates the others [189]. As fingerprints have parallel ridge patterns with well-defined local frequency orientation, the Gabor filter responses of fingerprints reflects the clarity of the fingerprint ridge pattern along a certain direction. Therefore, we choose the responses of a bank of Gabor filters as one quality measure.

The general form of a complex 2D Gabor filter in the spatial domain is given by [190]:

$$h_{\mathrm{Cx}}(x, y; f, \theta, \sigma_x, \sigma_y) = \exp\left(-\frac{1}{2}(\frac{x_\theta^2}{\sigma_x^2} + \frac{y_\theta^2}{\sigma_y^2})\right) \exp\left(j2\pi f x_\theta\right), \tag{10.1}$$

where

$$x_\theta = x \sin\theta + y \cos\theta, \tag{10.2}$$
$$y_\theta = x \cos\theta - y \sin\theta, \tag{10.3}$$

f is the frequency of the sinusoidal plane wave along the orientation θ, and σ_x, σ_y are the parameters of the Gaussian window.

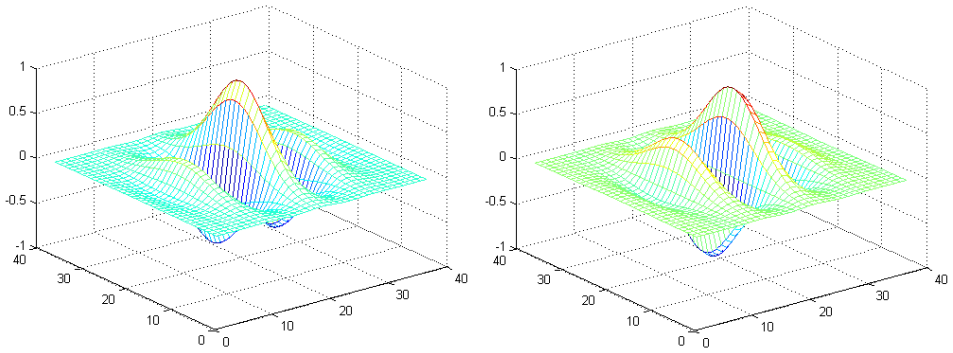


Figure 10.1: One complex Gabor filter ($f = 0.1, \sigma_x = 6.0, \sigma_y = 6.0, \theta = 1/2\pi$). (a) the real part of the complex Gabor filter; (b) the imaginary part of the complex Gabor filter.

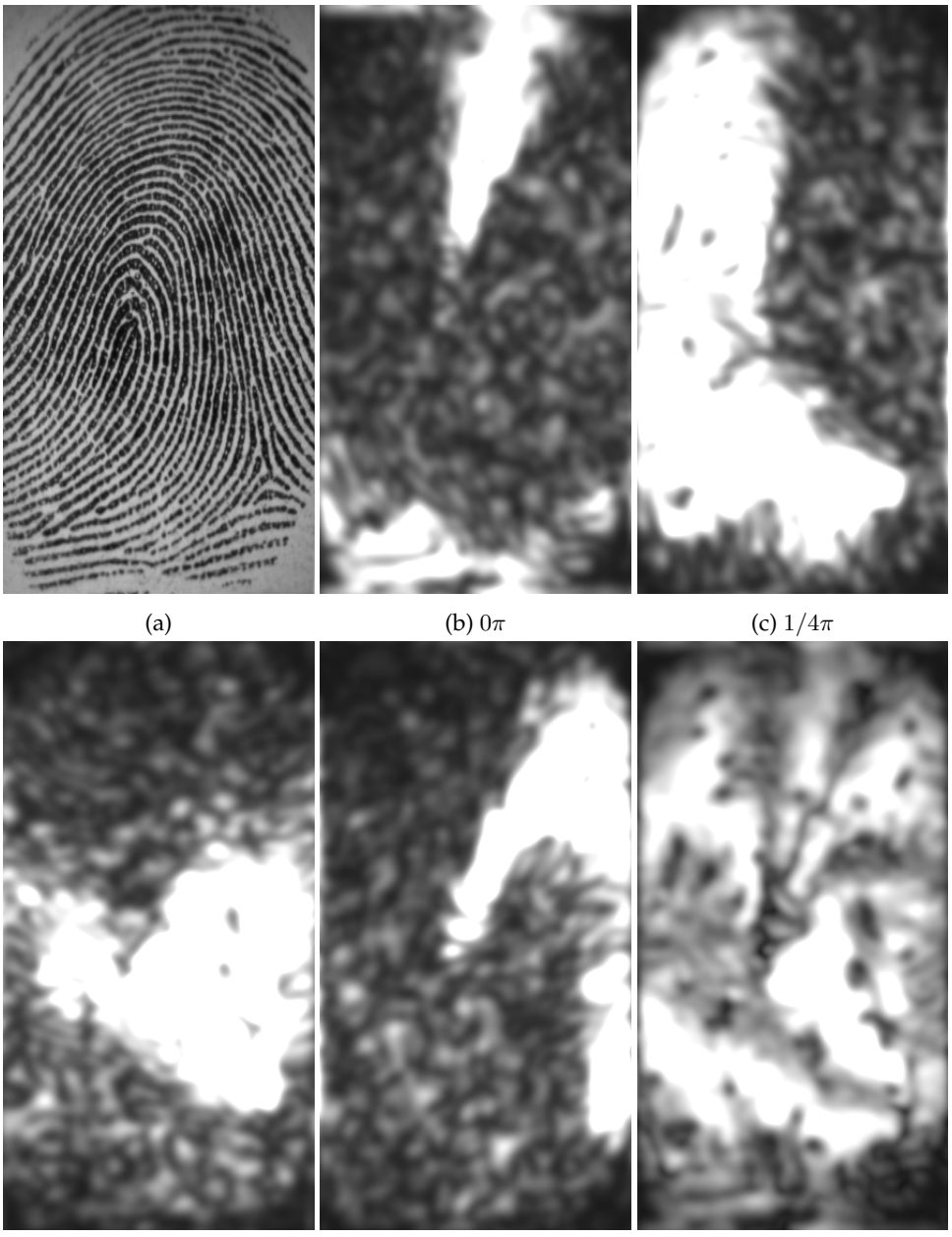
We set the filter frequency f as the reciprocal of the average inter-ridge distance. For fingerprint images captured from an adult population with the 500dpi resolution, the average inter-ridge distance is approximately 10 pixels [191]. Therefore, we set $f = 0.1$ in our experiments.

We use $n = 4$ four different orientations, $\theta = (k-1)/n\pi, k = 1, ..., n$, respectively, for the bank of Gabor filters. Figure 10.1 presents the real and imaginary part of the Gabor filter with orientation $\theta = 1/2\pi$. The Gaussian parameters are set as $\sigma_x = 6.0$ and $\sigma_y = 6.0$.

We apply the above mentioned bank of Gabor filters to fingerprint images. We take the magnitude of the filter responses and apply a small smoothing filter. A fingerprint and its four Gabor filter responses are illustrated in Figures 10.2a-10.2e, respectively.

The Gabor responses reflect the local quality of fingerprints. For the fingerprint areas with clear ridge pattern, the Gabor responses of one or a few orientations should have larger values than other orientations. Whereas for the background or the poor ridge clarity fingerprint areas, the Gabor responses of n orientations will be low and constant. In other words, the variety of the n Gabor responses at a certain area can reflect the clarity of the fingerprint ridge pattern in that area. Therefore, after obtaining the bank of Gabor responses $G_k, k = 1, ..., n$, we take the standard deviation of them, denoted as G_{std} and shown in Figures 10.2f. Finally, the Gabor quality is defined as the mean of G_{std}.

In this paper, we also evaluate another Gabor quality measure proposed by Shen et al. [43] in Section 10.5. The main differences between our method with Shen's are:



(a) (b) 0π (c) $1/4\pi$

(d) $2/4\pi$ (e) $3/4\pi$ (f)

Figure 10.2: Fingerprint from FVC2002Db2 dataset and its Gabor filter responses. (a) fingerprint image; (b-e) Smoothed absolute value of fingerprint image that have been filtered by Gabor filters with different orientations; (f) the standard deviation value of Gabor responses with different orientations.

- In Shen's method, the fingerprint foreground is segmented into blocks, and each block is marked as "good" or "poor" by comparing the Gabor responses with a fixed threshold T_q. However, for fingerprints obtained by different scanners, optimizing such a fixed threshold T_q is difficult and unreliable. During our evaluation, we noticed such unstable behaviour by setting different T_q. In our method, we obtain the Gabor quality measure without thresholds.

- In Shen's method, another fixed threshold T_b is used to segment fingerprint foreground. However, we notice in our experiments that the size of the fingerprint foreground could influence the recognition performance, i.e., a larger fingerprint foreground could lead to a higher performance. Therefore, in our method, we did not include the fingerprint segmentation part in our evaluation, in order to make the Gabor quality measure also reflect the size of the fingerprint foreground.

- In our method, we evaluated the influence of different Gabor filter orientation number n to the quality measure performance. Instead of using $n = 8$ as proposed by Shen, we found out that setting $n = 4$ will be able to obtain equivalent performances. This setting can increase the calculation speed with a factor of two.

10.3.2 Other Quality Metrics

The technical report on fingerprint sample quality [99] by the International Organization for Standardization is the most recent overview of recommended fingerprint quality metrics. We use the metrics defined there as a benchmark with which we compare our approach. We use the following metrics:

Orientation Certainty Level (OCL) measures the strength of the energy concentration along the dominant ridge flow orientation within a block by means of computing the blockwise gradient.

Ridge-valley Structure, also called Local Clarity Score (LCS) computes the blockwise clarity of ridge and valleys by applying linear regression to determine a gray-level threshold, classifying pixels as ridge or valley. A ratio of misclassified pixels is determined by comparing with the normalized ridge and valley thickness of that block.

Ridge-Valley Uniformity (RVU) measures the blockwise consistency of the ratio between ridge and valley. A large deviation from the global mean of ridge-valley ratios indicates low quality.

Frequency Domain Analysis (FDA) performs a blockwise Discrete Fourier Transform and measures the energy concentration in a frequency band. Dominance in the very low frequencies indicate low quality.

Radial Power Spectrum (POW) is a measure of maximal signal power in a defined frequency band of the global Radial Fourier spectrum. Ridges can be locally approximated by means of a single sine wave, hence high energy concentration in a narrow frequency band corresponds to consistent ridge structures.

In addition to the quality metrics included in ISO/IEC 29794-4, we also evaluated the Orientation Flow (OF) proposed in [72]. OF is a measure of the rate of change in the blockwise ridge flow across the fingerprint. The quality score decreases as the difference between the dominant ridge orientation of the block and that of its 8 neighboring blocks increases.

10.4 Quality Assessment

In this paper, we applied the concept of *Biometric Sample Utility* for quality assessment. Biometric Sample Utility is a measure of the biometric performance for a particular sample. The observed utility of the sample d_i^u (the uth presentation of subject i) is computed by means of the genuine and imposter similarity score distributions, e.g. in [70] the utility of a sample is defined as

$$utility_i^u = \frac{\mu_{i,u}^{genuine} - \mu_{i,u}^{imposter}}{\sigma_{i,u}^{genuine} + \sigma_{i,u}^{imposter}}, \tag{10.4}$$

where μ is the mean and σ is standard deviation. $utility_i^u$ is a measure of the distance between the two distributions: a high utility corresponds to the distributions being well separated and therefore that d_i^u has a low likelihood of being falsely rejected and that the likelihood of an imposter being falsely accepted as d_i^u is also low. Therefore, we say that a high utility reflects a high biometric performance. We will use this definition of utility for our performance comparison.

10.5 Experiments

The purpose of our experiments is to quantify the predictive performance of the Gabor based quality metric relative to the metrics from ISO/IEC TR 29794-4 and NFIQ. We do so by investigating the Spearman correlation between sample utility as defined in ISO/IEC standard on biometric fingerprint sample quality [70] and the sample quality value for each quality metric. Further we analyse the inter-metric correlations to uncover possible redundant quality metrics.

10.5.1 Datasets and Protocol

For the experiments we use Db2 (optical) and Db3 (capacitive) of the FVC2002 [143] database and Db1 (optical) and Db2 (optical) of the FVC2004 [133] database. We combine the 8 presentations per subject from the A (100 subjects) and B (10 subjects) sets resulting in 880 samples per dataset.

For each dataset we have computed the utility of each sample according to Eq. 10.4 using comparison scores obtained using the Neurotechnology Verifinger 6.2 SDK. We investigated 36 parameter permutations of the filter bank size, Gaussian and frequency parameters on the four datasets. For the Gabor filter bank size we used $n = \{4, 6, 8, 16\}$, the Gaussian parameters σ_x and σ_y were varied with $\sigma = \{4, 5, 6\}$, while the frequency parameter f was varied with $f = \{0.05, 0.1, 0.15\}$.

We compare our approach with NFIQ, OCL, LCS, RVU, FDA, POW, OF and the Gabor filter approach proposed by Shen et al [43]. Except for NFIQ, we apply 36 different parameter configurations on these metrics. In particular, for the blockbased algorithms (OCL, LCS, RVU, FDA) we use blocksizes of $8, 16, 32, 64$.

10.5.2 Results and Evaluation

The highest Spearman correlation values achieved for each quality metric on each dataset are summarized in table 10.1. The highest correlation with utility was obtained using the Orientation Flow (OF) approach on the FVC2004Db1 dataset. From the table, we observe that all the investigated metrics have an increased correlation with utility when using the FVC2004Db1 dataset compared to the three other datasets. All metrics have low correlation when using the FVC2002Db2 dataset which we believe is due to the FVC2002Db2 dataset containing particularly difficult to match samples (also called *outliers*). The reason is that outliers will make the estimation of the genuine distributions unreliable, thus influence the utility value. We observe that the proposed Gabor filter approach performs consistently and similarly to the NFIQ metric across the datasets.

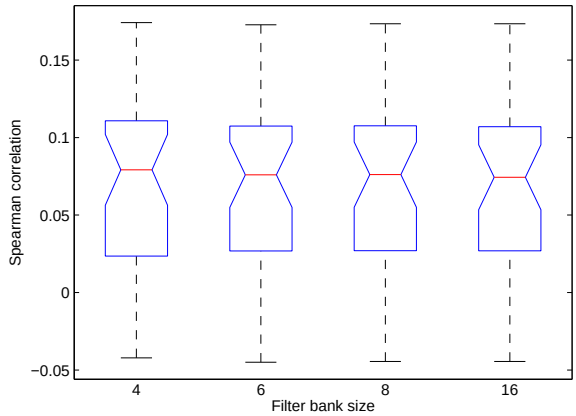
The configuration that resulted in the highest correlations for Gabor was obtained with $n = 4, \sigma_x = \sigma_y = 5, f = 0.1$. In our experiments, we observe that the performance of the quality measure does not increase when using values of n higher than 4. In Figure 10.3 the boxplots (boxes extend to 25th and 75 percentile, with mean marked as a line, and whiskers extending to cover the range excluding outliers) depict the correlation of Gabor quality values across all datasets when fixing respectively the Gabor parameters n 10.3a, f 10.3b, and σ. We observe that when viewed across the four dataset these are indeed the settings that show the highest correlations with utility. The influence of the frequency parameter f is dominating the quality score and thus the whiskers extend across a large range in both the filterbank size n and Gaussian σ box plots.

We also investigate the distribution of Gabor quality scores in relation to utility for each of the datasets. We found that the Gabor quality scores are distinct for each dataset and there are some outliers which show high utility and low quality score and vice versa. This is depicted in Figure 10.4. Neither dataset represents the result of another dataset and hence the consequence is that care must be taken in not extrapolating the results to a general case. The solution is to use datasets containing a larger sample of the population and which contain an even distribution of samples across the entire utility range.

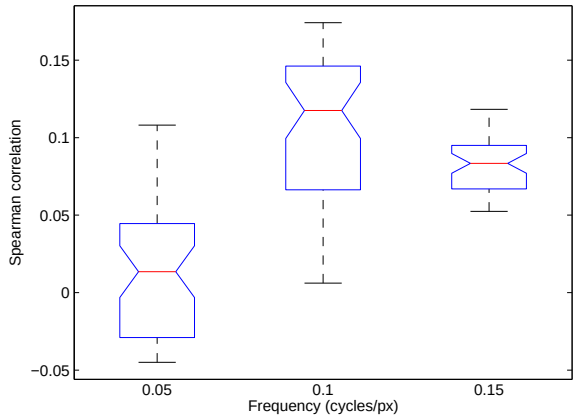
Metric	FDA	Gabor	GaborShen	LCS	NFIQ	OCL	OF	POW	RVU
FVC2002Db2	0.072	0.064	0.075	0.016	-0.039	0.072	0.037	-0.047	0.016
FVC2002Db3	0.025	0.132	0.073	0.056	-0.218	0.055	0.061	0.134	-0.107
FVC2004Db1	0.271	0.174	0.190	0.158	-0.213	0.201	0.211	0.119	-0.143
FVC2004Db2	0.070	0.126	0.113	0.125	-0.208	0.152	0.083	0.085	-0.125

Table 10.1: Spearman correlation between quality metrics and utility for each dataset. The highest correlation found among 36 parameter configurations for each metric is listed. Negative values indicate higher quality score correlate with lower utility (e.g. NFIQ).

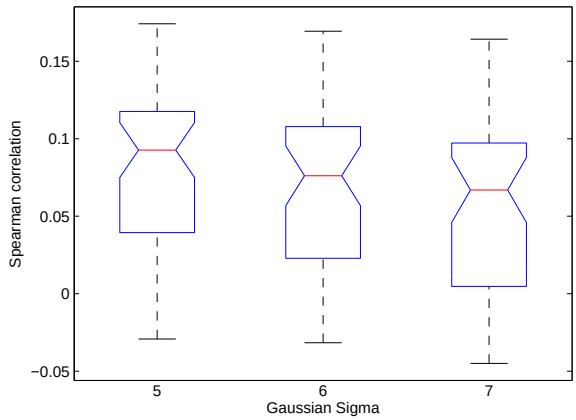
In table 10.2, we summarize the intermetric correlation to give an overview of which metrics may be redundant. We see almost perfect correlation (0.909) between our approach and the approach presented by Shen et al. This is to be expected as both approaches are based on Gabor filter responses, but we can use this information in combination with the utility correlations in Table 10.1 to determine that our approach outperforms the Shen approach when seen across all datasets. We also note that there are relatively high correlations with both the OCL and POW metrics, while the correlation between OCL and POW



(a)



(b)



(c)

Figure 10.3: Spearman correlations between Gabor settings and utility across the parameter space as investigated on the four datasets. (a) for fixed values of n $(4, 6, 8, 16)$; (b) for fixed values of f $(0.05, 0.1, 0.15$; (c) for fixed values of σ $(4, 5, 6)$.

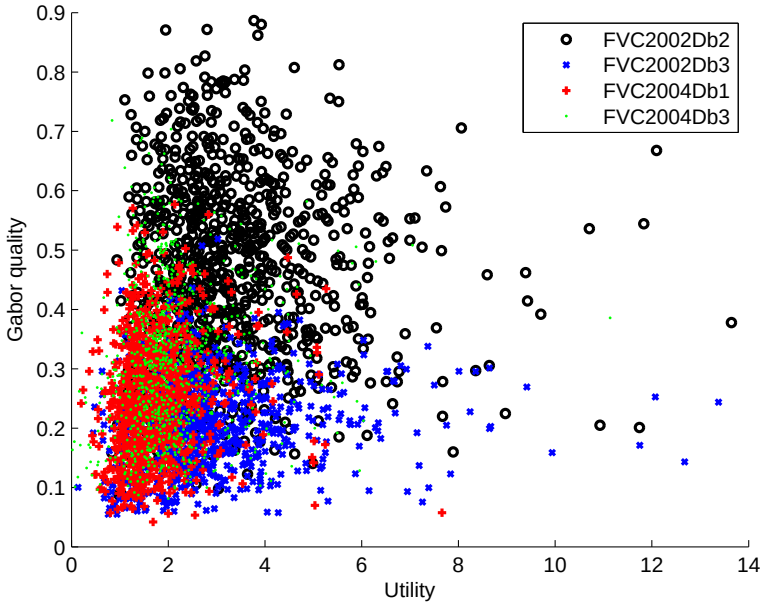


Figure 10.4: Gabor score plotted against utility for each dataset using $n = 4, f = 0.1, \sigma = 5$.

is comparatively lower. This indicates that the Gabor approach may be improved by combining the OCL and POW methods.

Metric	FDA	Gabor	GaborShen	LCS	NFIQ	OCL	OF	POW	RVU
FDA	1.000	0.507	0.464	0.173	-0.068	0.481	0.322	0.364	-0.315
Gabor	0.507	1.000	0.909	0.512	-0.266	0.713	0.449	0.665	0.201
GaborShen	0.464	0.909	1.000	0.522	-0.296	0.709	0.469	0.564	0.188
LCS	0.173	0.512	0.522	1.000	-0.066	0.742	0.441	0.499	0.553
NFIQ	-0.068	-0.266	-0.296	-0.066	1.000	-0.273	-0.319	0.023	0.127
OCL	0.481	0.713	0.709	0.742	-0.273	1.000	0.685	0.482	0.181
OF	0.322	0.449	0.469	0.441	-0.319	0.685	1.000	0.244	-0.006
POW	0.364	0.665	0.564	0.499	0.023	0.482	0.244	1.000	0.284
RVU	-0.315	0.201	0.188	0.553	0.127	0.181	-0.006	0.284	1.000

Table 10.2: Intermetric Spearman correlations on FVC2004Db1.

10.6 Conclusions

We introduced a Gabor filter based approach, which correlates with utility on a level comparable or above that of eight contemporary quality metrics. This shows the viability of using the Gabor filter to predict biometric performance on fingerprint images. While we investigated 36 parameter configurations, there is still room for exploring the parameter space for further optimizations.

As the NFIQ metric is a composite metric containing 11 features based on the fingerprint, we conclude that the Gabor filter approach could be included as part of a quality feature vector in order to be able to outperform and replace the NFIQ metric. In relation to the development of NFIQ 2.0, the successor of NFIQ, we have a positive outlook on the inclusion of the presented approach.

Our future research will be improving the Gabor quality approach, focusing on dealing with different types of fingerprints and/or scanners. Another research topic is how to select and combine quality metrics for generating the NFIQ 2.0 metric. Furthermore, the utility measure as defined in Eq. (10.4) is prone to outliers in the imposter and genuine score distributions, especially the genuine score distributions. This influences the reliability of the utility measure. We can cope with this by using datasets that allow for a larger amount of genuine comparisons, or by using another utility measure, e.g. instead of considering all genuine comparisons, we could consider only the maximum genuine score for a given sample. In our future investigations, we will consider these other options for quality metric performance assessment.

Acknowledgments

This research was supported by the Center for Advanced Security Research Darmstadt (CASED) and the German Federal Office for Information Security (BSI). Further we would like to thank our colleagues from NIST, BSI, secunet, Fraunhofer IGD and BKA for the fruitful discussions which prepared the ground for this paper.

Chapter 11

Fingerprint Sample Quality Assessment via Ridge Line Count Using Laplacian of Gaussian Edge Finding

Information about the paper

This paper is one development towards new quality assessment algorithms which fast to perform on systems with reduced computational resources. The proposed algorithm is based on the assumption that there in any given high quality area within the foreground of the fingerprint will be a certain number of ridge valley transitions. The paper is a contribution towards research questions $\mathcal{RQ}_2$ and $\mathcal{RQ}_3$. This original paper was published in: [6] M. Dusio, M. A. Olsen, and C. Busch. "Fingerprint Sample Quality Assessment via Ridge Line Count Using Laplacian of Gaussian Edge Finding". In: *The 2nd International Workshop on Biometrics and Forensics (IWBF)*. 2014.

Abstract

We propose a new fingerprint sample quality measurement algorithm based on local fingerprint area analysis and ridge line counts. The proposed method determines ridge orientation of local areas with Principal Component Analysis, finds edges using the Laplacian of Gaussian and zero-crossings approach, and counts valley to ridge transitions across the center of the area to assign a local score. The local scores are aggregated to assign a single quality value for the fingerprint. We assess the performance of our method with Error versus Reject Curves obtained from comparison scores from three fingerprint comparators on CASIA Fingerprint V5.0 and FVC 2004 DB 1A datasets. The quantitative evaluation shows that the proposed method achieves predictive performance on level with NFIQ while achieving higher fidelity.

11.1 Introduction

Biometric fingerprint recognition is widely used to check and confirm personal identity by analyzing the pattern of ridges present on a persons fingertip. Biometric systems work by performing a comparison between enrolled reference templates and captured probe samples to generate a comparison score. Based on this score and a biometric threshold, a biometric decision is made - if the score is above the threshold, the decision is a match, otherwise it is a non-match. Mated samples, stemming from the same subject and source, give genuine comparisons; non-mated samples give impostor comparisons. By performing comparisons on all mated pairs of samples and several non-mated pairs in one dataset of samples, two distributions of genuine and impostor comparison scores are obtained. In a perfect world, these distributions could be completely separated by the biometric threshold. In practice however, they do overlap and some genuine comparisons are falsely rejected. The portion of genuine comparison scores that are below the threshold is called the False Non-Match Rate (FNMR) and can be used as a performance metric.

It is currently understood in government, industry and academia that the recognition performance of a sample is, among other factors, influenced by its quality [192]. ISO/IEC 29794–1 [70] gives a concrete definition of the quality of a biometric sample as being the predictor of the biometric performance. Good quality samples give good recognition performance whereas poor quality samples increase the error rates. A biometric system working with highest quality samples will have smaller error rates, i.e., rejecting samples of poor quality will decrease the FNMR [192]. Recognition performance may suffer due to poor quality of fingerprint samples caused by one or several factors: *a)* skin defects such as blisters, wrinkles or scars; *b)* skin conditions such as roughness, dryness or wetness; *c)* user behavior such as bad finger alignment, too little or too much pressure; and *d)* environmental conditions: dirty sensor platen, extreme temperature or lighting. If the acquired sample is of poor quality, *c)* and *d)* may be altered by the user if given sufficient feedback and the acquisition process can be repeated.

The motivation to research fingerprint quality algorithm is driven by a need for higher accuracy in prediction of performance and that ISO/IEC 29794-1 [70] requires 100 quality levels for compliance. The contribution of this paper is a fingerprint sample QMA based on local ridge line counts and which is competitive with state of the art methods and as such is a possible candidate to be incorporated in ISO/IEC 29794–4 [99] and NFIQ 2.0

[80]. Ridge line count is a measure of distance between two points in a fingerprint image. In forensics it has been used in fingerprint classification schemes and the measurement between core and delta has been used in forensic AFIS to improve fingerprint comparison accuracy [49, p.161].

The rest of paper is organized as follows: Section 11.2 describes state of the art quality algorithms. Section 11.3 specifies the proposed method. Experiments and results are discussed in Section 11.4 and 11.5 with conclusions presented in Section 11.6.

11.2 Fingerprint Sample Quality Analysis

State of the art fingerprint quality methods rely on image processing to measure various features of the fingerprint sample to determine if the sample is good or bad quality. Biometric sample quality estimation methods differ depending on the biometric characteristic that is used, but the common principle is to use a *Quality Measurement Algorithm (QMA)* that produces a scalar quality score q given an input sample x: $q = QMA(x)$.

Several state of the art fingerprint QMAs are described in ISO/IEC 29794–4 [99] and the NFIQ 2.0 Features Definition document [82]. From the best performing known methods we have chosen Orientation Certainty Level (OCL) [71] and NIST Finger Image Quality (NFIQ) [79] as baseline algorithms for benchmarking the methods proposed in this work.

OCL is a method of fingerprint quality estimation first introduced by Lim et al. in 2002 [71]. Fingerprint samples feature ridge lines flowing through the image in a certain pattern at varying angles. Local areas of a sample contain a certain orientation of this flow. Good quality samples with well defined ridge texture allow to determine the ridge orientation with high certainty, whereas poor ridge texture results in uncertain measurement. To determine this certainty, OCL applies a Sobel operator to obtain gray level gradients at the local areas and performs Principal Component Analysis to find the covariance matrix and eigenvalues. The ratio between found eigenvalues is the certainty of the orientation flow and is used as the quality indication at the local area. Quality of the whole sample is calculated as the mean of local qualities.

NIST Finger Image Quality, proposed by Tabassi, Wilson and Watson in 2004 [79] is currently the most widespread approach to quality estimation according to Maltoni et al. [49]. It is the best performing method according to the NFIQ 2.0 Features Evaluation [81] and gives a good performance prediction for diverse comparison algorithms. NFIQ uses orientation flow, minutiae count and other image features as a feature vector which is converted into a single quality score using an artificial neural network [79]. NFIQ categorizes fingerprint samples into five quality levels: excellent, very good, good, fair, poor and is the defacto standard for fingerprint quality assessment.

11.3 Ridge Line Count Quality Assessment

The main principle behind the proposed Ridge Line Count (RLC) method is to locally analyze a fingerprint sample and count the number of ridges. We assume that poor quality regions of fingerprint samples do not contain clearly defined ridges and hence does not allow to count them easily. Good quality samples will have more areas where ridges are clearly defined and easily countable. This can be determined by taking into account that for adults ridges/valley cycle takes $500\mu m$ up to $600\mu m$ [49] when the ridges are thick –

with 500dpi sensors corresponds to roughly 3.25 ridge lines over 32px and minimally 2.7 ridge lines over 32px.

In order to calculate the number of ridges passing through a local area extracted from a fingerprint sample, the following approach is proposed. First, the whole sample is binarized using Otsu's threshold in order to minimize intra-class variance of white and black pixels, then the sample is divided into local areas of size 32×32 pixels. The area size is chosen the same as in state of the art methods described in ISO/IEC 29794–4 [99] in order to decrease the computational cost if multiple methods are combined (sample division would not have to be repeated for each method). The upper bound for the local area size is the curvature of skin ridges, the chosen size ensures minimal curvature outside of delta and core fingerprint regions. The lower bound is ridge thickness - with the chosen area size it is ensured that at least two full ridges should be present. Generally larger area size leads to less areas and renders the method faster to compute overall. However, the purpose of this paper is not to present the most optimal approach to quality analysis via ridge line counting, but show that it is indeed possible and offers performance benefits compared to other methods.

The next step is analysis of each area. To count the number of ridges, their orientation is required such that the area may be traversed orthogonal to the ridge line orientation, incrementing a counter whenever a ridge is passed. Area orientation is found by calculating grey level gradient of the area and performing Principal Component Analysis [193] to obtain covariance matrix coefficients $C = \left[\begin{smallmatrix} a & c \\ c & b \end{smallmatrix}\right]$. Area rotation angle is determined from these coefficients as

$$\alpha = tan^{-1}\left(\frac{c}{\sqrt{c^2 + (a-b)^2}} \frac{a-b}{\sqrt{c^2 + (a-b)^2}}\right) \tag{11.1}$$

The area is rotated according to α such that the ridges are vertical. We estimate the ridge count by traversing the area in the horizontal direction and count the number of detected zero crossings after applying a Laplacian of Gaussian filter calculated as

$$LoG(x,y) = -\frac{1}{\pi\sigma^4}\left[1 - \frac{x^2+y^2}{2\sigma^2}\right] e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{11.2}$$

where $LoG(x,y)$ is the response at pixel (x,y). Filter size is chosen as 13×13 and $\sigma = 2$. The LoG approach is chosen as it includes a lowpass filter that removes noise that may lead to spurious edges.

The proposed method produces a map of local area qualities. To produce a final quality scalar q, this map is aggregated as follows – the final quality is the ratio of local areas that feature a ridge count over a threshold to all local areas. The threshold is chosen as 2.7 which as described earlier is the minimum number of ridges expected to pass over 32 pixels.

As a less computationally expensive approach to counting ridge lines we propose an alternative Ridge Line Count simple (RLC simple). The simple alternative approach does not perform Laplacian of Gaussian edges finding, but traverses the rotated area through vertical middle and counts each white to black pixel transition. The simple approach also produces a map of local area qualities that is aggregated as a mean of the local values.

Figure 11.1 shows an example of how a fingerprint sample local area is processed with the proposed method and the corresponding RLC and RLC simple values are shown below

for each processed local area. Five areas extracted from different fingerprint samples are shown, three of them produce a score higher than 2.7 and thus qualify for the threshold-based aggregation method.

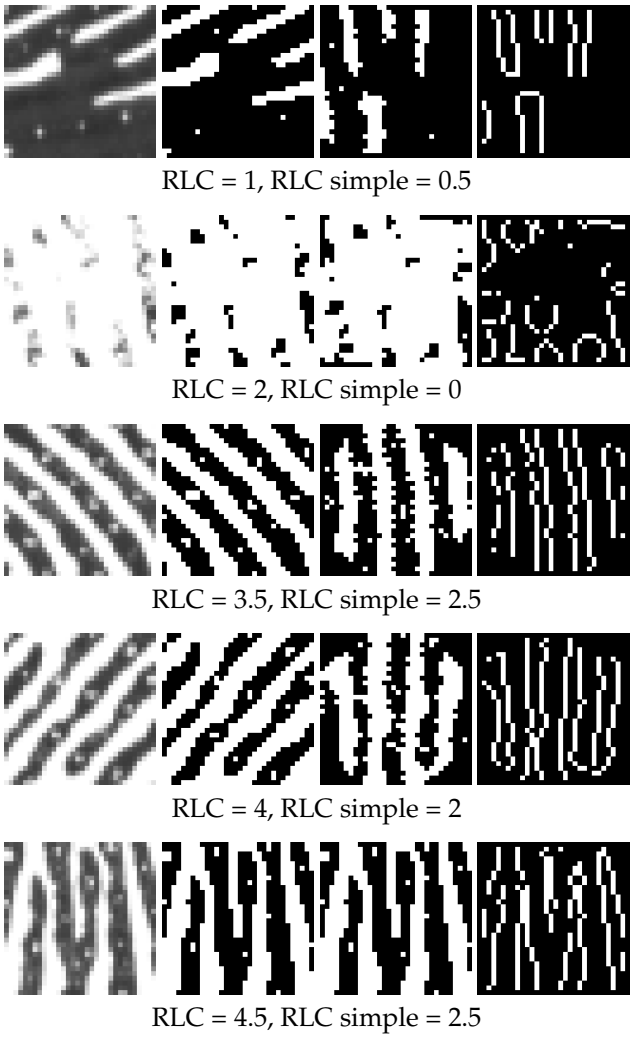


Figure 11.1: Proposed Ridge Line Count method processing performed on five example fingerprint local 32×32 pixels areas. From left: original area, binarized area, rotated area, edges found using LoG. Ridges are counted through horizontal middle of the area. The RLC and RLC simple values for the area are shown below each row.

11.4 Experimental Setup

We focus on optical fingerprint sensors as they are most commonly used, offer highest resolution [49], are used by the Office of Biometric Identity Management (OBIM) program by the Department of Homeland Security [170] and used in the Aadhaar project by Unique Identification Authority of India (UIDAI) [184].

The experiment is divided into three parts: *preparation* of fingerprint dataset, fingerprint comparison algorithms, reference and proposed QMAs; *calculation* of genuine and impostor comparison scores, quality scores of all fingerprints via reference and proposed QMAs; and *assessment* of proposed and reference QMA performance via ERC analysis.

The fingerprint datasets chosen are the publicly available CASIA Fingerprint Image Database Version 5.0 (CASIAFPV5) [140] and DB 1A from the Third International Fingerprint Verification Competition (FVC2004DB1A) [133]. CASIAFPV5 contains 20 000 fingerprint samples acquired with varying rotation and pressure from 500 subjects with different background, i.e. students, workers, waiters, etc. FVC2004DB1 contains 880 fingerprint samples acquired with varying skin moisture conditions, finger pressure and rotation. Both datasets feature samples captured in 500dpi resolution and offer good quality variation to support a fair assessment of the proposed method and comparison of its performance with state of the art.

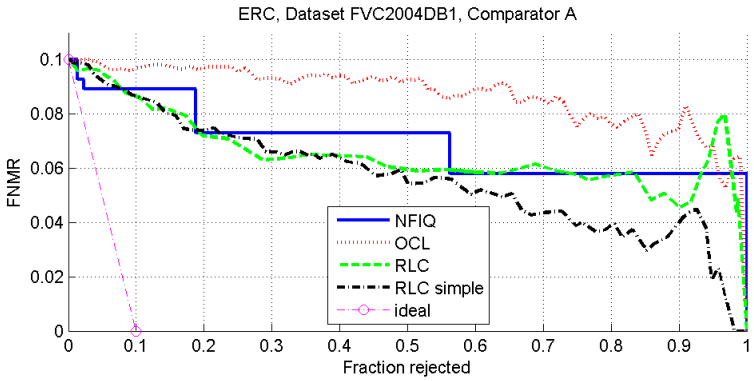
For the experiments we use three fingerprint comparators and treat them as black boxes. They appear in the paper as comparator *A*, *B*, and *C*. The QMAs used to calculate fingerprint quality are the proposed RLC, the variation RLC simple, and state of the art NFIQ and OCL. OCL quality feature extraction is performed accordingly to the procedure described in ISO/IEC 29794–4 [99]. RLC is calculated as described in section 11.3.

Comparison scores between all pairs of fingerprints are calculated using each of the available comparators such that: all genuine comparison scores are calculated with an exception that no sample is compared to itself; a limited number of 100 impostor comparison scores is calculated for each sample. All fingerprints have their quality calculated using the proposed Ridge Line Count QMA, and state of the art Orientation Certainty Level (OCL) and NIST Finger Image Quality (NFIQ).

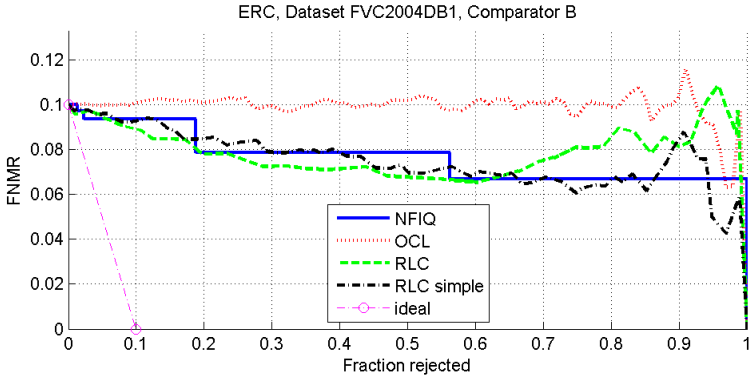
State of the art and proposed method performance is assessed and compared in terms of the improvement in False Non–Match Rate with poor quality samples rejection via Error versus Reject Curves (ERC). ERC was first proposed by Grother and Tabassi [112] in 2007; it shows precisely how removing lowest quality samples (as indicated by a certain QMA) influences system performance viewed as FNMR. Additionally, the quality obtained via the proposed method is compared to OCL and NFIQ in terms of how the results correlate by calculating Spearman's rank correlation coefficient ρ as described in NFIQ 2.0 Features Evaluation [81]. This way it is possible to determine whether two methods are complementary – if the output of a pair of QMAs has low correlation, then these methods are complementary and could be combined, otherwise if the correlation is very high, the methods are likely to measure a similar feature of a sample.

11.5 Results

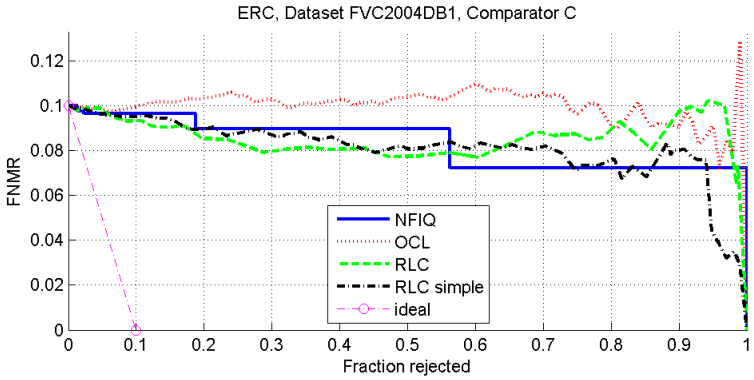
Figures 11.3 and 11.2 show Error versus Reject Curves analysis results calculated on the CASIA Fingerprint V5.0 and FVC2004DB1A datasets using comparators A, B and C using



FVC2004DB1A, Comparator A

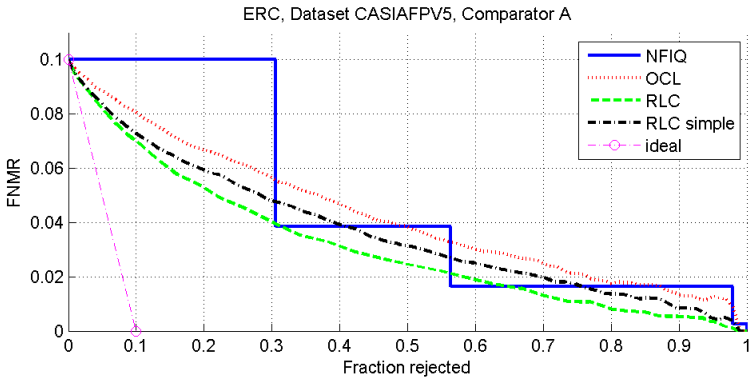


FVC2004DB1A, Comparator B

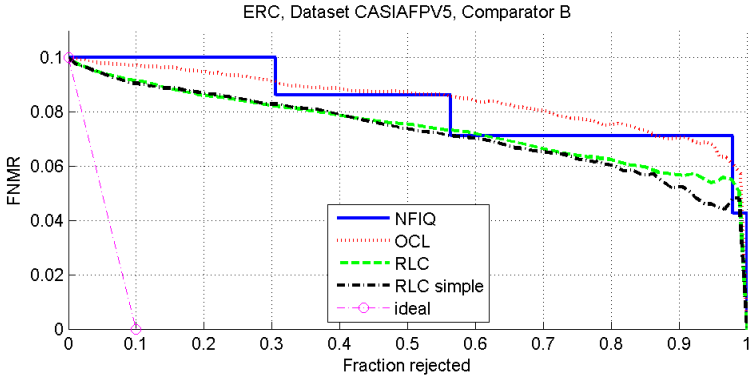


FVC2004DB1A, Comparator C

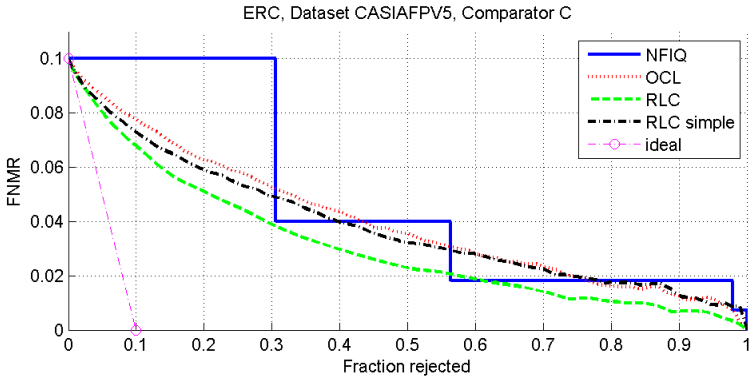
Figure 11.2: Error vs. Reject Curves QMA performance analyzed on FVC2004DB1A using three fingerprint comparators. RLC – proposed method, RLC simple – proposed method simplified version, OCL and NFIQ – state of the art methods.



CASIA Fingerprint V5.0 dataset, Comparator A



CASIA Fingerprint V5.0 dataset, Comparator B



CASIA Fingerprint V5.0 dataset, Comparator C

Figure 11.3: Error vs. Reject Curves QMA performance analyzed on the CASIA Fingerprint V5.0 dataset using three fingerprint comparators. RLC – proposed method, RLC simple – proposed method simplified version, OCL and NFIQ – state of the art methods.

minimum quality score as the criteria for rejecting a comparison. The x-axis shows the fraction of samples rejected according to quality, with lowest quality samples rejected first. The biometric threshold is chosen such that FNMR is ten percent as suggested by method authors [112]. FNMR is shown on y-axis. The ideal case, marked by the straight dashed line, shows the behaviour when up to ten percent of samples are rejected according to quality and when the sole factor in FNMR is quality.

QMA performance as analyzed on the CASIA dataset is as follows: NFIQ allows FNMR drop to 0.04 when 30 per cent of samples are rejected – this is the case for comparators A and C – for comparator B FNMR drops to rougly 0.085. The proposed method offers similar performance with 30 per cent rejected samples for comparators A and B, and slightly outperforms NFIQ when comparator B is used. OCL offers worse performance allowing FNMR drop to 0.055 for comparator A, 0.09 and roughly 0.05 for comparator C. Simplified RLC is worse than the normal version for comparators A and C and performs similarly with comparator B; it is generally better than OCL. The difference in performance between comparators is due to the fact that each comparator produces a different distribution of genuine and impostor comparison scores - on the CASIA dataset comparators A and C behave similarly, whereas when comparator B is used then QMA performance is worse.

Results of ERC analysis of the QMAs on FVC2004DB1A are similar to those of CASIAFPV5. NFIQ and the proposed method give good results – for fraction rejected 30 per cent the FNMR drops to 0.075, 0.08 and 0.09 for comparators A, B and C respectively both for RLC and NFIQ. The simplified RLC is only slightly worse than the normal version until more than 40 per cent of the samples are rejected when comparator A is used and until 65 per cent are rejected for comparators B and C – when these fractions are rejected, the simple method is actually better. It is also worth to note that on this dataset, with comparator C NFIQ is very goot after roguhly 56 per cent of the samples are rejected. With FVC2004DB1A the best QMA performance is observed when comparator A is used, whereas comparators B and C give slightly higher ERC curves denoting worse QMA performance.

	NFIQ	OCL	RLC	RLC simple
NFIQ	100	60	68	62
OCL	60	100	49	32
RLC	68	49	100	82
RLC simple	62	32	82	100

Table 11.1: Inter-method correlation on the quality sets of the CASIA Fingerprint V5.0 dataset. Absolute value of Spearman's ρ calculated using proposed and state of the art QMAs. Value is multiplied by 100 and rounded for reader convenience.

Tables 11.1 and 11.2 show the inter-method correlation calculated as absolute value of Spearman's ρ multiplied by 100 and rounded for reader convenience for datasets CASIA Fingerprint V5.0 and FVC2004DB1A respectively. For CASIAFPV5, state of the art methods OCL and NFIQ show medium correlation of 60, and the proposed method RLC shows medium correlation of 49 with OCL and medium correlation of 68 with NFIQ. The simplified version of RLC shows a correlation of 62 with NFIQ and 32 with OCL, and a correlation of 82 with the normal RLC. For FVC2004DB1A, OCL and NFIQ show a low cor-

	NFIQ	OCL	RLC	RLC simple
NFIQ	100	31	38	32
OCL	31	100	28	38
RLC	38	28	100	78
RLC simple	32	38	78	100

Table 11.2: Inter-method correlation on the quality sets of the FVC2004DB1A. Absolute value of Spearman's ρ calculated using proposed and state of the art QMAs. Value is multiplied by 100 and rounded for reader convenience.

relation of 31, whereas the proposed RLC method shows correlation of 28 with OCL and 38 with NFIQ. The simplified RLC shows a correlation of 32 with NFIQ, 38 with OCL and 78 with the normal RLC. The proposed method and its simplified version are complementary to the analyzed state of the art methods and could be combined with them.

11.6 Conclusions

A fingerprint sample Quality Measurement Algorithm measuring quality based on the local count of ridge lines was proposed, assessed and compared to state of the art methods NFIQ and OCL using Error versus Reject Curve analysis, three fingerprint comparators, a publicly available fingerprint dataset CASIA Fingerprint V5.0 and FVC2004DB1A. The analysis shows that the proposed method offers biometric performance prediction close to that of NFIQ while at the same time increasing the fidelity, i.e., more quality levels are provided so a finer control of the FNMR reduction is achieved. The correlation table shows that the proposed method is complementary to that of orientation certainty level. We show that a variation of the proposed method, which does not use Laplacian of Gaussian for edge detection, can achieve nearly the same predictive performance but at a lower computational cost as the filter convolution is skipped. The proposed method may be further optimized by tuning parameters such as local area size and improving the ridge line counting algorithm in order to minimize the computational cost and increase the performance.

Acknowledgements

This work was partially supported by CASED (Center for Advanced Security Research Darmstadt).

Chapter 12

Predicting Dactyloscopic Examiner Fingerprint Image Quality Assessments

Information about the paper

Dactyloscopic examiners work with fingermarks and one of their tasks is to attribute evidential value to a fingermark. Here we show that quality assessments made by examiners are indicative of comparison scores, and we develop a proof of concept as a step towards an objective quality algorithm which can assist examiners in making decisions on fingermark quality. One major challenge in the development of such an algorithm is in the acquisition of ground truth data which can be applied to develop and verify the proposed algorithm. The paper is a contribution towards research question $\mathcal{RQ}_5$. This original paper was published in: [9] M. A. Olsen, M. Böckeler, and C. Busch. "Predicting Dactyloscopic Examiner Fingerprint Image Quality Assessments". In: *BIOSIG* (2015).

Abstract

We work towards a system which can assist dactyloscopic examiners in assessing the quality and decision value of a fingerprint image and eventually a fingermark. However when quality assessment tasks of datyloscopic examiners are replaced by automatic quality assessment then we need to ensure that the automatic measurement is in agreement with the examiner opinion. Under the assumption of such agreement, we can predict the examiner opinion. We propose a method for determining the examiner agreement on ordinal scales and show that there is a high level of agreement between examiners assessing the ground truth quality of fingerprints. With ground truth quality information on 749 fingerprints and using 10-fold cross validation we construct models using Support Vector Machines and Proportional Odds Logistic Regression which predicts median examiner quality assessments 35% better than when using the prior class distribution.

12.1 Introduction

Fingerprint sample quality in the context of forensic applications where an examiner following the Analysis, Comparison, Evaluation, and Verification (ACE-V) protocol is part of the initial information gathering phase where the examiner studies the impression to quantify the present discriminating information and assess the quality and completeness [194].

The quality assessment will among other factors such as the completeness of the fingerprint have an impact on the decision value of the impression, which can be one of: Value for Individualization (VID), which is used when the quantity and quality of the information present is deemed sufficient to determine if the impression is from the same source as another, yet unseen, impression; Value for Exclusion Only (VEO) is used when sufficient information is present to determine that the impression is not from the same source as another impression; No Value (NV) is used when the impression is deemed unsuitable. The process of assigning VID, VEO, and NV is inherently subjective and requires training and experience to perform accurately and consistently. Ulery *et al*. conducted a study on the accuracy and reliability of 169 forensic examiners who assigned VID, VEO, and NV decisions to 100 latents and found that VID decisions were unanimous in 48% of cases for mated pairs and 33% for non-mated pairs [195]. In a related study on the repeatability and reproducibility of decision by individual examiners, it was found that 93% of VID, 85% NV, and 55% VEO decisions were repeated by individual examiners when presented twice with the same impression after a 7 month interval [196]. These findings, which indicate a high degree of accuracy with respect to VID, and lower accuracy with respect to VEO mirror the findings by Langenburg [197].

It has been demonstrated that when provided with extraneous contextual information, an examiner might change the method of judging and comparing fingerprints [198]. Additionally, the examiners are vulnerable to biasing information such as evidence of confession of a crime, even in cases where the comparison of the impressions is non-difficult [199].

Biometric sample quality has successfully been applied in the context of Automated Fingerprint Recognition Systems (AFIS) to reject samples which are likely to contribute negatively to False Non-Match Rates (FNMR). By rejecting those samples before they are

enrolled, a high level of biometric performance is achievable and with it higher levels of satisfaction by the users interacting with the biometric system [120].

We are motivated by the successful application of automated quality assessment in AFIS and by the findings of Ulery *et al*. and Bradford *et al*., which highlight the subjective nature of the ACE-V protocol, to determine methods which objectively assesses the quality of an impression in the form of a fingerprint or a fingermark. This paper represents one step in this direction and our main objective is to determine methods which predict the quality assessment that a dactyloscopic examiner gives a particular fingerprint. To achieve this, we leverage a dataset which contains ground truth quality labels on inked impressions as assessed by dactyloscopic examiners from the German Federal Criminal Police Office (BKA), and apply quality assessment algorithms identified or developed in the context of NFIQ 2.0 [80].

The rest of the paper is organized as follows: section 12.2 outlines state of the art methods for objective quality assessment. In section 12.3 we propose a method for quantifying examiner agreement, section 12.4 details the ground truth dataset on which we base our experiments that are described in section 12.5. We discuss our results in section 12.6 and conclude in section 12.7.

12.2 Fingerprint quality

There exists a large number of fingerprint image quality assessment algorithms in the literature and several reviews have been made, e.g. in the context of optical and capacitive sensors [90, 91]; relation between quality assessed by human experts and algorithms and comparison scores [89]; and more recently quality assessment using no-reference algorithms have been investigated [86]; a comprehensive review of biometric sample quality is provided by Bharadwaj *et al*. [95] and a quality metric for fingermarks has been proposed [111].

We have selected a subset based on those features specified in the NFIQ 2.0 quality feature definitions document version 0.5 [82] which we summarize here: Frequency Domain Analysis (Q_{FDA}) uses the magnitude of the dominant frequency as determined by discrete Fourier transform fingerprint image as a local quality value. Local Clarity Score (Q_{LCS}) determines the clarity of the fingerprint image by estimating how well each block in the fingerprint image can be segmented into ridge and valley region. Orientation Flow (Q_{OFL}) measures the continuity of ridge flows in the fingerprint image by determining the dominant ridge orientation agreement between one block and its neighbouring blocks. Orientation Certainty Level (Q_{OCL}) is a measure of the strength of the ridge orientation within a image block. A high score indicates that the ridge orientation within the block is well defined. Ridge Valley Uniformity (Q_{RVU}) measures the consistency between ridge and valley widths within each image block. The widths of ridges and valleys are expected to be similar across the entire fingerprint image. Radial Power Spectrum (Q_{RPS}) quantifies the energy concentration within a specified band in the Fourier spectrum. The limits of the band are determined by the expected ridge valley frequency. Image mean (Q_{MU}) is the mean value of pixel intensities across the fingerprint image. Image standard deviation (Q_{SIG}) is the standard deviation of pixel values across the fingerprint image.

The features Q_{FDA}, Q_{LCS}, Q_{OFL}, Q_{OCL}, Q_{RVU} operate on 32 by 32 pixel non-overlapping regions of the image and thus provide a vector of local quality values. Q_{RPS}, Q_{MU}, Q_{SIG}

work on the entirety of the image and provide a scalar value. Based on these two groups we define two sets of features: Set A which contains the mean and standard deviation of each of the local quality vectors of Q_{FDA}, Q_{LCS}, Q_{OFL}, Q_{OCL}, Q_{RVU} giving a total of 10 features; set B contains, in addition to the features in set A, Q_{RPS}, Q_{MU}, Q_{SIG} for a total of 13 features.

12.3 Quantifying examiner agreement on ordinal scales

To quantify examiner agreement, which is an essential task to determine to which degree examiners agree on what quality means and to compare and judge how well automatic quality prediction will perform against its human counterpart, the following requirements are specified for an examiner agreement coefficient:

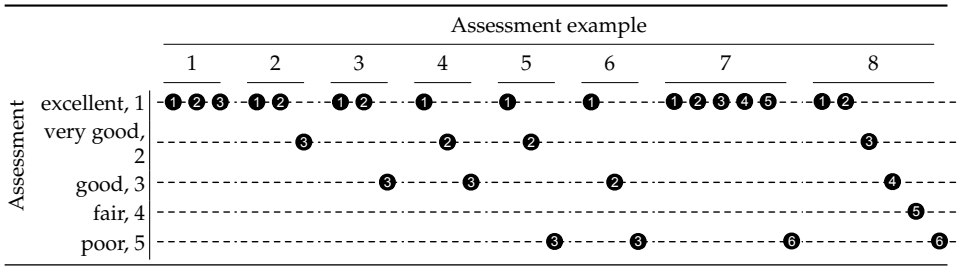
1. an unlimited number of assessments on a single fingerprint shall be considered
2. the agreement of assessments shall be weighted according to their distance
3. assessments that belong to the same decision categories shall be assigned with high weights
4. when the assessment scale varies the measure results shall remain consistent

Ad 1 - the coefficient shall be capable of measuring agreement for fingerprints that were annotated by a minimum of 2 examiners and for fingerprints that are annotated by more than 2 examiners. Ad 2 - assessments that are "closer" to each other shall result in higher agreement. Ad 3 - not only the "distance" of single assessments shall be measured, assessments in equal decision categories shall be weighted higher than assessments that are not in the same decision category. Ad 4 - quality assessments in varying assessment scales shall result in the same agreement if the relative distance of the single ratings are the same. For example, a quality assessment of $x_{11} = 1, x_{12} = 2, x_{13} = 5$ $\{x_{1x} \in \mathbb{N} | 1 \leq x_{1x} \leq 5\}$ and a quality assessment of $x_{21} = 1, x_{22} = 25, x_{23} = 100$ $\{x_{2x} \in \mathbb{N} | 1 \leq x_{2x} \leq 100\}$ shall result in the same agreement coefficient.

Several common statistical measures like the Percentage agreement $\bar{P}$ [196], the Interquartile Range IQR [200, p.27], the Median Absolute Deviation MAD [201, p.220] and the Standard Deviation SD [200, p.26] were investigated to determine if they fulfil the specified requirements. Table 12.1 displays how these measures perform on various assessment examples. It is clear that none of the common measures are able to measure examiner agreement sufficiently as they all violate one of the predefined requirements.

$\bar{P}$ is equal for assessment examples 2 and 3, and examples 4, 5 and 6 thus violating requirement 2. IQR is equal for the assessment examples 3 and 4 violating requirements 2 and 3 as also for assessment example 5 and 6 which violates requirement 2. MAD is equal for the first 3 examples thus violating requirement 2 and 3 and in the 4th and 5th example requirement 2 is violated. SD violates requirement 3 in examples 2, 3, 7 and 8. Requirement 4 is violated by IQR, MAD, SD as the produced values depend on the range of the scale.

None of the measures described above satisfy the task of measuring examiner agreement as specified by our requirements, and hence we propose a new coefficient which does fulfil the specified requirements: Closest-neighbour Median Cluster Agreement ($CMCA$).



Agreement	1	2	3	4	5	6	7	8
$\bar{P}$	1.000	0.333	0.333	0.000	0.000	0.000	0.667	0.067
IQR	0.000	1.000	2.000	2.000	4.000	4.000	0.000	3.000
MAD	0.000	0.000	0.000	1.000	1.000	2.000	0.000	1.500
SD	0.000	0.471	0.943	0.816	1.700	1.633	1.491	1.491
$CMCA$	1.000	0.839	0.689	0.422	0.166	0.125	0.765	0.208

Table 12.1: Examples of agreement values computed using $\bar{P}$, IQR, MAD, SD, $CMCA$ for 8 cases of examiner assessments where the number of experts and their assessments vary. In the top half of the table, each dot represents an examiner assessment ranging from excellent (1) to poor (5). The bottom half of the table shows the agreement value assigned by the 5 metrics for each assessment example.

$CMCA$ consists of multiple parts which are calculated as follows. Let x_j be the j^{th} ascending sorted rating on the fingerprint. The closest neighbour distance $ND_j \in [0, 1]$, one part to fulfil requirement 2, of the j^{th} rating on the fingerprint is:

$$ND_j = \min\left(|x_j - x_{j-1}|, |x_j - x_{j+1}|\right) \tag{12.1}$$

Further, let $\max i$ the maximum possible rating on the rating scale, $\min i$ the minimum possible rating on the scale, $\tilde{x}$ the median of ratings and r the number of ratings on the fingerprint. The distance consensus $D \in [0, 1]$ of the fingerprint made to fulfil requirements 2 and 4 is:

$$D = \frac{\left(\sum_{j=1}^{r}\left(1 - \frac{ND_j}{\max i - \min i}\right) + \left(1 - \frac{|x_j - \tilde{x}|}{\max i - \min i}\right)\right) - 1}{2r - 1} \tag{12.2}$$

Let c_j be the j^{th} cluster (ratings in the same decision category) and let $|c_j|$ be its cardinality. Let nc be the number of clusters on. The average cluster size difference $CSD \in [0, r-2]$, which compares the size of each assessment cluster against each other and calculates the mean of their size differences is:

$$CSD = \frac{\sum_{j=1}^{nc}\sum_{k=j}^{nc}||c_j| - |c_k||}{\frac{(nc-1)\cdot nc}{2}} \tag{12.3}$$

To fulfil requirement 3, the distance consensus D is multiplied by the cluster size power $CSP \in [1, r-1]$, which is calculated as:

$$CSP = nc - \frac{CSD}{r} \tag{12.4}$$

The $CMCA$ on the fingerprint is finally calculated by:

$$CMCA = \begin{cases} 1.0 & , nc = 1 \\ 1.0 - \frac{|x_1 - x_2|}{\max i - \min i} & , r = 2 \\ D^{CSP} & , \text{otherwise} \end{cases} \tag{12.5}$$

The $CMCA \in [0, 1]$ over a set of fingerprints I is calculated as the arithmetic mean over the $CMCA$ agreements of each fingerprint in I.

The measurement results of $CMCA$ applied to the 8 assessment examples are shown in the last row of table 12.1. Assessment examples 5 and 6, show that $CMCA$ fulfils requirement 2 by considering the distance of between single assessments. It also fulfils requirement 3, heavy weighted assessments that belong to the same decision category. The example shows that the $CMCA$ fulfils requirement 1, measuring agreement for any number of assessments on a single fingerprint.

After all, the $CMCA$ was designed to measure examiner agreement per fingerprint, see 12.5. Other coefficients like Cohens Kappa [202] or Fleiss Kappa [203] were designed to measure inter-rater agreement over the whole assessment population which is expressed by the arithmetic mean of the set of $CMCA$. Nonetheless, Cohens Kappa has the disadvantage that it can only measure pairwise rater agreement on nominal data. In addition, Fleiss Kappa is able to measure inter-rater agreement for more than 2 raters, but was also designed for nominal data, so additional information of the natural order of categories on ordinal data like it is present in our case of fingerprint quality, would not be considered properly.

12.4 Ground truth data

The ground truth data of this paper, representing the human examiner quality assessment part, originates form the NIST special database 14 [159], containing 54 000 fingerprints from live-scan and scanned ink impressions and from the NIST special database 29 [204], containing 4 320 fingerprints. In 2009, a team of 9 dactyloscopic examiners from the BKA annotated for the purpose of a conformance testing study [160] several fingerprint characteristics, like minutia points, singular points and the overall fingerprint quality which is investigated in this paper. To establish an objective annotation, the examiner team was equipped with a simple graphical user interface that omitted the support of the automatic or semi-automatic minutia extraction functionality of the AFIS. To further increase the objectivity and anonymity of the process, each examiner was assigned with an ID that was not known to other examiners. The examiners rated the overall fingerprint quality within 5 decision categories, ranging from excellent (1), very good (2), good (3), fair (4) to poor (5). A total of 749 fingerprints were annotated by at least 2 examiners. Figure 12.1 shows the logical partitioning of the annotated samples, where the first level contains the set of 749 fingerprints; the second level shows how many samples were annotated grouped by the number of examiners. The third level shows the number of samples each distinct group of examiners annotated, e.g. there were 713 fingerprints annotated by 3 examiners, which came from two groups of 3 examiners annotating respectively 361 and 352 fingerprints ($\mathbf{S}_2$

and $\mathbf{S}_3$). Table 12.2 shows the number of images that were annotated by each of the 9 examiners. We note that examiners 11 to 16 have each annotated nearly 400 samples, while examiners 17 to 19 have annotated fewer than 20 samples.

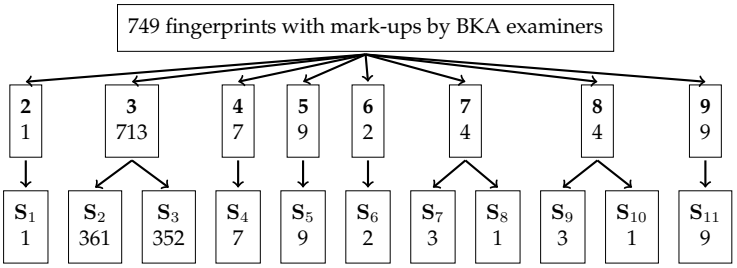


Figure 12.1: Examiner markup tree. The bold digits in the second tree level display how many examiners annotated a specific number of fingerprints. The last tree level shows how many fingerprints where annotated by distinct groups of examiners.

	Examiner								
	11	12	13	14	15	16	17	18	19
Annotations	396	393	397	378	388	371	10	17	17

Table 12.2: The number of images annotated by each of the 9 examiners.

		$CMCA$				
Quality	n	min	max	mean	median	std
1	41	.69	1.00	.82	.84	.08
2	306	.42	1.00	.86	.84	.15
3	305	.42	1.00	.90	.84	.10
4	92	.42	1.00	.88	.84	.11
5	5	.84	.84	.84	.84	.00
All	749	.42	1.00	.88	.84	.13

Table 12.3: Summary statistics of $CMCA$ at each median ground truth quality level and across all samples.

Table 12.3 summarizes the $CMCA$ computed on median ground truth quality levels and for all fingerprint images in the ground truth data set. The mean $CMCA$ is .88 across all fingerprint images, indicating that there is generally consensus between examiners when subjectively assessing the quality level of a fingerprint image. At the individual levels we note some differences in the $CMCA$, in particular that mean $CMCA$ at quality level 1 is .82 while for level 3 it is .90.

Figure 12.2 shows an example fingerprint for each median quality assessment. For median quality levels 1 through 4 the examiners were in agreement with $CMCA = 1.0$ while for the fingerprint illustrating level 5 where $CMCA = 0.838$. We observe that at level 1 (left most figure) the ridge lines are clearly separated around the core, where the

examples for levels 4 and 5 appear without clearly defined ridge lines and with blurry or low contrast regions.

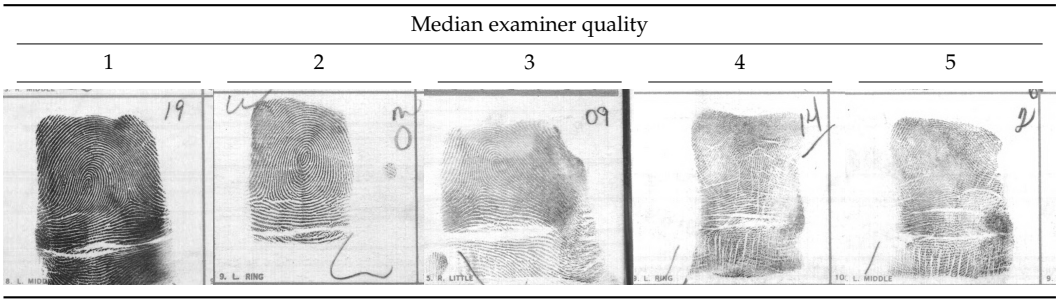


Figure 12.2: Median examiner quality assessment with example fingerprints. All images are from the NIST special database 14 [159] (file names from left to right: f0000118, f0000109, f0000095, f0000969, f0000968).

To assess whether human examiner quality assessments are indicative of the eventual genuine comparison score of the fingerprint sample, we computed the genuine comparison scores for all samples and grouped them according to the examiner who made the assessment and the quality score that was assigned. Box plots showing the relation between assigned quality and genuine comparison score [205] for each of the 6 examiners who annotated the most images (see table 12.2), as well as the median examiner quality are depicted in fig. 12.3. The plots show that generally a higher quality score is associated with a higher comparison score, however, for some cases we note some irregularities. For Examiner 11 (left most plot), we note that no images were assigned quality level 5 and that those samples which received a quality level of 1 were involved in comparisons resulting in scores similar to those samples which were assessed as being quality level 2.

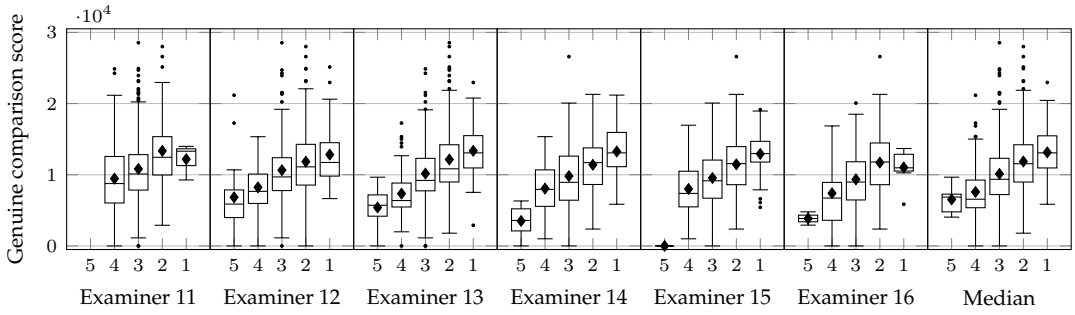


Figure 12.3: Boxplots of genuine comparison score for each of 5 quality levels assigned by 6 examiners and for the median of assigned quality levels.

12.5 Experiments

Our goal is to predict the quality level that an examiner will assign to a given fingerprint. We note from table 12.3 that our proposed $CMCA$ coefficient indicates a high degree of agreement between examiners as to the assigned ground truth quality levels for the data set.

We perform a series of experiments in order to assess to which degree the quality scores assigned by individual examiners or the median quality score is predicted. From table 12.3 we note that the distribution of median quality levels is not uniform with the majority of samples being assigned levels 2 or 3 and with only 5 samples in level 5. Due to the low annotation count by examiners 17, 18, and 19 we do not attempt to predict their assessments (see table 12.2) and instead only assess examiners 11 to 16 individually.

We train our predictive models using Multi-class Support Vector Machine (SVM) [45] and Proportional Odds Logistic Regression (POLR) [206] where the response variable is the assigned quality level (either individual examiner or median of examiners), and the explanatory variables are features in sets A or B (see section 12.2).

The experiments are performed using 10-fold cross validation, i.e., we divide the available data in each experiment randomly into 10 disjunct partitions of equal size. Over the 10 possible permutations we perform training of SVM and POLR on the 9 folds and test the performance on the remaining fold. In the case of SVM we use Radial Basis Function as kernel and perform a grid search for optimal sample influence radius (γ) and cost (C) over $\gamma \in \{0.001, 0.01, ..., 1000\}$ and $C \in \{0.001, 0.01, ..., 1000\}$ given the training folds. In the case of POLR no parameter optimization is performed.

The predictive capability of the constructed models in each experiment setting is determined by calculating the mean and standard deviation of the F_1 score and Cohen's Kappa (κ) [202] over the 10 permutations.

Cohen's Kappa quantifies the class agreement between the model predictions and the ground truth by taking into account the observed probabilities of the classes. When $\kappa = 0$ the agreement is equal to that which can be achieved by random chance based on the priors - when $\kappa = 1$ then the agreement is complete. F_1 is the harmonic mean of precision and recall and the lowest score is achieved when $F_1 = 0$ and highest when $F_1 = 1$.

Experiments were performed using R [207] with SVM from e1071 [208]; POLR from MASS [209]; cross validation and miscellaneous functions from caret [210] and xtable [211] packages.

12.6 Results

Our analysis of the ground truth data set (section 12.4) showed that there is a high level of agreement in assessing quality levels across examiners, and that the higher quality levels are associated with higher genuine comparison scores.

Following the protocol described in section 12.5 we performed a total of 14 experiments and have summarized the mean and standard deviations of F_1 and κ for each of them in table 12.4. Each line in the table alternates between feature set *A* and *B* in *Set* column with the *Target* column indicating what is being predicted where *Median* indicates that it is the median of examiner quality assessments per fingerprint that is predicted, and *Examiner 11* indicates that it is the quality assessments of Examiner 11 which are predicted. The

remaining columns are first grouped by F_1 and κ, next by method *SVM* or *POLR* and finally arithmetic mean (*mean*) and standard deviation (*std*) of testing results across the 10 fold cross validation.

		F_1				κ			
		SVM		POLR		SVM		POLR	
Set	Target	mean	std	mean	std	mean	std	mean	std
A	Median	.58	.05	.59	.06	.31	.07	.33	.07
B	Median	.60	.08	.60	.06	.34	.12	.35	.07
A	Examiner 11	.72	.05	.70	.03	.21	.17	.13	.10
B	Examiner 11	.72	.05	.71	.06	.28	.14	.18	.16
A	Examiner 12	.45	.06	.48	.05	.17	.07	.23	.07
B	Examiner 12	.51	.05	.52	.07	.26	.08	.29	.08
A	Examiner 13	.49	.07	.49	.08	.21	.09	.21	.12
B	Examiner 13	.53	.08	.52	.08	.30	.10	.26	.09
A	Examiner 14	.57	.10	.59	.09	.30	.14	.33	.13
B	Examiner 14	.57	.09	.60	.09	.29	.14	.36	.13
A	Examiner 15	.57	.06	.60	.09	.29	.08	.35	.14
B	Examiner 15	.58	.07	.61	.08	.32	.09	.37	.12
A	Examiner 16	.63	.11	.64	.09	.33	.21	.36	.16
B	Examiner 16	.64	.12	.65	.11	.36	.21	.38	.19

Table 12.4: Results of experiments in predicting median and individual examiner quality assessment using SVM and POLR on feature sets A and B.

The best prediction results when using Median as target is achieved with POLR and feature set B when considering either of F_1 and κ as evaluation criteria. We see that the mean F_1 is .60 for both SVM and POLR, but the standard deviation is smaller in the case of POLR with .06 over .08 achieved with SVM. For κ we note a mean of .35 for POLR and .34 for SVM, again with a smaller standard deviation favouring POLR.

We note that both F_1 and κ spans a wide range when using individual examiner quality assessments as target for the predictions. Examiner 12 appears to be hardest to predict with a mean κ of .17 and .23 and F_1 of .45 and .58 for respectively SVM and POLR when using feature set A. Prediction of the scores assigned by Examiner 16 using feature set B gives the highest mean F_1 of .64 and .65 and κ of .36 and .38, however in both cases the standard deviation over the 10 folds is the highest of the experiments performed.

Generally using the global features present in feature set B lead to marginal increases in the mean F_1 score while κ is increased further for both SVM and POLR.

In addition to SVM and POLR listed in table 12.4 we also used Recursive Partioning, however that algorithm had difficulty working with the relatively small dataset with 10-fold cross validation and was thus not able to complete all intended experiments.

12.7 Conclusion

In this paper we have made steps towards assisting dactyloscopic examiner in assessing the quality of a given fingerprint or fingermark with the aim of determining its decision value in the ACE-V protocol. We address the objective quality of fingerprints and relation to examiner opinion with a continued goal to extend the quality assessment evaluation to fingermarks which pose the greatest challenge in the forensic evaluation.

We proposed the $CMCA$ coefficient as a general method for quantifying examiner agreement on ordinal scales containing any number of categories. Using $CMCA$ we have shown that there is a high level of agreement between examiners as to what constitutes a high quality fingerprint and further that the ground truth assessments made by dactyloscopic examiners are indicative of genuine comparison scores.

On our limited dataset containing 749 finger images and using 13 quality features we have constructed a model which predicts examiner quality assessments around 35% better than random chance given the prior quality class probabilities as measured using Cohen's Kappa.

Our future work includes refinement of the quality feature set to improve the predictive capabilities; evaluation of the importance of the individual features to gain insights as to which image covariates are important to examiners; evaluation of the trained model to determine the degree that the assigned quality levels are indicative of biometric performance; and finally how the trained models can be incorporated in ACE-V protocol to assist the decision making of dactyloscopic examiners when working with fingerprints or fingermarks.

Acknowledgements

We thank the team of dactyloscopic examiners with the German Federal Criminal Police Office for providing ground truth fingerprint information and National Institute of Standards and Technology for providing the selection of images from SD14 and SD29.

Glossary

ACE-V Analysis, Comparison, Evaluation and Verification. 19

AFIS Automated Fingerprint Identification System. 8, 14, 15, 19, 24

ANSI/NIST-ITL American National Standards Institute/National Institute of Standards and Technology - Information Technology Laboratory. 9

AUC Area Under Curve. xi, 25, 29, 50, 51

BSI German Federal Office for Information Security. 7, 30, 32, 52

CASED Center for Advanced Security Research Darmstadt. 52

CS-VIS Central Visa Information System. 25

DET Detection Error Tradeoff. x, 47, 130

DFT Discrete Fourier Transform. 10, 32, 34, 45

DHS Department of Homeland Security. 25

EC European Commission. 24, 25

EER Equal Error Rate. 11, 14, 26

ERC Error-Reject Curve. 14, 17, 18, 23, 25, 29, 47, 48, 50–52

EURODAC European Dactyloscopy. 15

EU European Union. 15, 24

FDA Frequency Domain Analysis. ix, xiii, 34, 35

FMR False Match Rate. 47

FNMR False Non-Match Rate. xi, 29, 47, 48, 50–52

FRVT 2006 Face Recognition Vendor Test 2006 [40]. 10

FTE Failure to Enroll. 4

FVC2004 Fingerprint Verification Competition 2004. 46

FVC2006 Fingerprint Verification Competition 2006. 46

GAB Gabor. ix, xiii, 35–37

GSH Gabor Shen. ix, xiii, 36, 38, 39

IDENT Automated Biometric Identification System. 30

IGD Fraunhofer-Institut für Graphische Datenverarbeitung. 52

INCITS InterNational Committee for Information Technology Standards. 9

ISO/IEC International Organisation for Standardization/International Electrotechnical Commission. 9, 13, 18, 26

ISO International Organization for Standardization. 12, 34

LCS Local Clarity Score. ix, xiii, 38, 40, 41

LED Light Emitting Diode. 27

LFIQ Latent Fingerprint Image Quality. 14

LOG Laplacian of Gaussian. 18

LUT Lookup Table. 18

MCYT-330 MCYT Fingerprint 330. 46, 53

MLP Multi-Layer Perceptron. 32

NBIS NIST Biometric Image Software. 9

NFIQ 2 NIST Fingerprint Image Quality 2. 13, 32

NFIQ NIST Fingerprint Image Quality. 10, 13, 28, 30–32, 46

NI-VIS National Interfaces Visa Information System. 25

NIST National Institute of Standards and Technology. 9, 10, 17, 31, 32, 52

OBIM Office of Biometric Identity Management. 25

OCL Orientation Certainty Level. ix, xiii, 42, 43

OCT Optical Coherence Tomography. 9

OFL Orientation Flow. ix, xiii, 40–42

PAUC Partial Area Under Curve. xi, 50, 51

PCA Principal Component Analysis. 12

RBFNN Radial Basis Function Neural Network. 13

RPS Radial Power Spectrum. ix, xiii, 44–46

RVU Ridge Valley Uniformity. ix, xiii, 44

SECUNET secunet Security Networks AG. 52

SIFT Scale-invariant Feature Transform. 13

SIS II Second generation Schengen Information System. 15

SIS Schengen Information System. 15

SIVV Spectral Image Validation and Verification. 10, 17

SOM Self-Organizing Map. 13, 18, 19, 103

SVM Support Vector Machine [45]. 10, 13

UAM Universidad Autónoma de Madrid. 53

UIDAI Unique Identification Authority of India. 11, 25, 28, 30

UID Unique Identification number. 25

US-VISIT United States Visitor and Immigrant Status Indicator Technology. 15, 25, 30

USK4 User Software Kit 4 - Quality Check. 25

VIS Visa Information System. 24, 25

WDSET02 Wet-Dry Dataset 2. 46

Bibliography

[1] M. Olsen, H. Xu, and C. Busch. "Gabor filters as candidate quality measure for NFIQ 2.0". In: *Biometrics (ICB), 2012 5th IAPR International Conference on*. 2012, pp. 158–163. DOI: 10.1109/ICB.2012.6199802 (cit. on pp. 4, 6, 17, 35, 74, 89, 133).

[2] A. Aabrandt, M. Olsen, and C. Busch. "A topology based approach to categorization of fingerprint images". In: *Biometrics Special Interest Group (BIOSIG), 2012 BIOSIG - Proceedings of the International Conference of the*. 2012, pp. 1–11 (cit. on pp. 5, 6, 17, 59).

[3] M. A. Olsen, E. Tabassi, A. Makarov, and C. Busch. "Self-Organizing Maps for Fingerprint Image Quality Assessment". In: *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*. CVPRW '13. Washington, DC, USA: IEEE Computer Society, 2013, pp. 138–145. ISBN: 978-0-7695-4990-3. DOI: 10.1109/CVPRW.2013.28. URL: http://dx.doi.org/10.1109/CVPRW.2013.28 (cit. on pp. 5, 6, 18, 87, 104, 106, 107, 119).

[4] I. Danov, M. A. Olsen, and C. Busch. *Interpretation of fingerprint image quality features extracted by self-organizing maps*. 2014. DOI: 10.1117/12.2050676. URL: http://dx.doi.org/10.1117/12.2050676 (cit. on pp. 5, 6, 18, 103).

[5] J. Wang, M. A. Olsen, and C. Busch. *Finger image quality based on singular point localization*. 2014. DOI: 10.1117/12.2050145. URL: http://dx.doi.org/10.1117/12.2050145 (cit. on pp. 5, 6, 18, 69).

[6] M. Dusio, M. A. Olsen, and C. Busch. "Fingerprint Sample Quality Assessment via Ridge Line Count Using Laplacian of Gaussian Edge Finding". In: *The 2nd International Workshop on Biometrics and Forensics (IWBF)*. 2014 (cit. on pp. 5, 6, 17, 18, 145).

[7] M. A. Olsen, M. Dusio, and C. Busch. "Fingerprint Skin Moisture Impact on Biometric Performance". In: *International Workshop on Biometrics and Forensics 2015*. 2015 (cit. on pp. 5, 6, 18, 46, 47, 123).

[8] M. A. Olsen, V. Šmida, and C. Busch. "Finger image quality assessment features – definitions and evaluation". English. In: *IET Biometrics* (2015). ISSN: 2047-4938. URL: http://digital-library.theiet.org/content/journals/10.1049/iet-bmt.2014.0055 (cit. on pp. 5, 6, 18, 23, 126).

[9] M. A. Olsen, M. Böckeler, and C. Busch. "Predicting Dactyloscopic Examiner Fingerprint Image Quality Assessments". In: *BIOSIG* (2015) (cit. on pp. 5, 6, 19, 155).

[10] M. A. Olsen, D. Hartung, C. Busch, and R. Larsen. "Contrast Enhancement and Metrics for Biometric Vein Pattern Recognition". In: *Advanced Intelligent Computing Theories and Applications*. Ed. by D.-S. Huang, M. McGinnity, L. Heutte, and X.-P. Zhang. Vol. 93. Communications in Computer and Information Science. Springer Berlin Heidelberg, 2010, pp. 425–434. ISBN: 978-3-642-14831-6. DOI: 10.1007/978-3-642-14831-6_56 (cit. on p. 5).

[11] M. A. Olsen, D. Hartung, C. Busch, and R. Larsen. "Convolution Approach for Feature Detection in Topological Skeletons Obtained from Vascular Patterns". In: *IEEE Symposium Series in Computational Intelligence 2011 (SSCI 2011)*. Paris, France, Apr. 2011 (cit. on p. 5).

[12] M. A. Olsen and C. Busch. "Deficiencies in NIST Fingerprint Image Quality Algorithm". In: *12. Deutscher IT-Sicherheitskongress*. Vol. 1. 2011, pp. 251–262 (cit. on p. 6).

[13] D. Hartung, M. A. Olsen, H. Xu, and C. Busch. "Spectral Minutiae for Vein Pattern Recognition". In: *Proceedings IEEE International Joint Conference on Biometrics (IJCB 2011), October 11-13*. 2011 (cit. on p. 6).

[14] D. Hartung, M. A. Olsen, H. Xu, H. Thanh Nguyen, and C. Busch. "Comprehensive analysis of spectral minutiae for vein pattern recognition". In: *Biometrics, IET* 1.1 (2012), pp. 25–36. ISSN: 2047-4938. DOI: 10.1049/iet-bmt.2011.0013 (cit. on p. 6).

[15] G. Li, B. Yang, M. Olsen, and C. Busch. "Quality Assessment for Fingerprints Collected by Smartphone Cameras". In: *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*. 2013, pp. 146–153. DOI: 10.1109/CVPRW.2013.29 (cit. on pp. 6, 14).

[16] C. Gottschlich, A. Mikaelyan, M. Aastrup Olsen, B. Josef, and C. Busch. "Improving Fingerprint Alteration Detection". In: *The 13th IEEE International Symposium on Parallel and Distributed Processing with Applications (IEEE ISPA-15)* (2015) (cit. on p. 6).

[17] ISO/IEC. *IS 2382-37:2012. Information technology — Vocabulary — Part 37: Biometrics*. Tech. rep. JTC 1/SC 37/WG 1, 2012 (cit. on p. 7).

[18] W. J. Babler. "Embryologic development of epidermal ridges and their configurations". In: *Birth Defects Original Article Series* 27.2 (1991), pp. 95–112 (cit. on p. 7).

[19] M. Kücken and A. C. Newell. "Fingerprint formation". In: *Journal of Theoretical Biology* 235.1 (2005), pp. 71–83. ISSN: 0022-5193. DOI: 10.1016/j.jtbi.2004.12.020. URL: http://www.sciencedirect.com/science/article/pii/S0022519304006198 (cit. on p. 7).

[20] D. A. Garzón-Alvarado and A. M. Ramírez Martinez. "A biochemical hypothesis on the formation of fingerprints using a turing patterns approach". In: *Theoretical Biology & Medical Modelling* 8 (24 2011). DOI: 10.1155/2012/748302 (cit. on p. 7).

[21] A. M. Turing. "The Chemical Basis of Morphogenesis". English. In: *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 237.641 (1952), pp. 37–72. ISSN: 00804622. URL: http://www.jstor.org/stable/92463 (cit. on p. 7).

[22] W. J. Herschel. *The Origin of Finger-Printing*. 1916 (cit. on p. 7).

[23] M. Arnold, H. Daum, and C. Busch. “Comparative Study on Fingerprint Recognition Systems - Project BioFinger.” In: *BIOSIG*. 2003, pp. 33–38 (cit. on pp. 7, 27).

[24] Bundesamt für Sicherheit in der Informationstechnik and Bundeskriminalamt and Fraunhofer Institut für Graphische Datenverarbeitung. “Study: “Evaluation of Fingerprint Recognition Technologies – BioFinger“”. In: (2004) (cit. on p. 7).

[25] S. Yoon and A. K. Jain. “Longitudinal study of fingerprint recognition”. In: *Proceedings of the National Academy of Sciences* (2015). DOI: 10.1073/pnas.1410272112. eprint: http://www.pnas.org/content/early/2015/06/23/1410272112.full.pdf. URL: http://www.pnas.org/content/early/2015/06/23/1410272112.abstract (cit. on p. 7).

[26] Supreme Court of the United States. *William Daubert, et ux., etc., et al., Petitioners v. Merrell Dow Pharmaceuticals, Inc.* 509 U.S. 579, 113 S. Ct. 2786; 125 L. Ed. 2d 469. 1993 (cit. on p. 7).

[27] US District Court for the Eastern District of Pennsylvania. *US v. Byron Mitchell, Criminal Action No. 96-407*. 1999 (cit. on p. 7).

[28] J. L. Wayman. “WHEN BAD SCIENCE LEADS TO GOOD LAW: THE DISTURBING IRONY OF THE DAUBERT HEARING IN THE CASE OF U.S. V. BYRON C. MITCHELL”. In: *Biometrics in Human Services Users' Group Newsletter #17* (2000) (cit. on p. 7).

[29] S. Pankanti, S. Prabhakar, and A. Jain. “On the individuality of fingerprints”. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24.8 (2002), pp. 1010–1025. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2002.1023799 (cit. on p. 8).

[30] Y. Zhu, S. Dass, and A. Jain. “Statistical Models for Assessing the Individuality of Fingerprints”. In: *Information Forensics and Security, IEEE Transactions on* 2.3 (2007), pp. 391–401. ISSN: 1556-6013. DOI: 10.1109/TIFS.2007.903846 (cit. on p. 8).

[31] A. K. Jain, S. Prabhakar, and S. Pankanti. “On the similarity of identical twin fingerprints”. In: *Pattern Recognition* 35.11 (2002), pp. 2653 –2663. ISSN: 0031-3203. DOI: http://dx.doi.org/10.1016/S0031-3203(01)00218-7. URL: http://www.sciencedirect.com/science/article/pii/S0031320301002187 (cit. on p. 8).

[32] H. Lee, R. Ramotowski, and R. Gaensslen. *Advances in Fingerprint Technology, Second Edition*. Forensic and Police Science Series. CRC Press, 2001. ISBN: 9781420041347 (cit. on p. 8).

[33] E. H. Holder, L. O. Robinson, and J. H. Laub. *The fingerprint sourcebook*. US Department. of Justice, Office of Justice Programs, National Institute of Justice, 2011 (cit. on p. 8).

[34] ISO/IEC JTC 1/SC 37. *IS 19794-4:2011. Information technology — Biometric data interchange formats — Part 4: Finger image data*. 2011 (cit. on pp. 8, 9, 26).

[35] A. Bossen, R. Lehmann, and C. Meier. “Internal Fingerprint Identification With Optical Coherence Tomography”. In: *Photonics Technology Letters, IEEE* 22.7 (2010), pp. 507–509. ISSN: 1041-1135. DOI: 10.1109/LPT.2010.2041347 (cit. on p. 9).

[36] Y. Cheng and K. V. Larin. "Artificial fingerprint recognition by using optical coherence tomography with autocorrelation analysis". In: *Appl. Opt.* 45.36 (2006), pp. 9238–9245. DOI: 10.1364/AO.45.009238. URL: http://ao.osa.org/abstract.cfm?URI=ao-45-36-9238 (cit. on p. 9).

[37] C. Sousedik, R. Breithaupt, and C. Busch. "Volumetric fingerprint data analysis using Optical Coherence Tomography". In: *Biometrics Special Interest Group (BIOSIG), 2013 International Conference of the*. 2013, pp. 1–6 (cit. on p. 9).

[38] National Institute of Standards and Technology. "American National Standard for Information Systems: Data format for the interchange of fingerprint, facial other biometric information". In: *ANSI/NIST-ITL 1-2013* (2013) (cit. on p. 9).

[39] American National Standards Instutite. "Finger Minutiae Format for Data Interchange". In: *ANSI INCITS 378-2009,* (2009) (cit. on p. 9).

[40] P. J. Phillips, W. T. Scruggs, A. J. O'Toole, P. J. Flynn, K. W. Bowyer, C. L. Schott, and M. Sharpe. "FRVT 2006 and ICE 2006 large-scale results". In: *National Institute of Standards and Technology, NISTIR* 7408 (2007), p. 1 (cit. on pp. 10, 167).

[41] J. Beveridge, P. Phillips, G. Givens, B. Draper, M. Teli, and D. Bolme. "When high-quality face images match poorly". In: *Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*. 2011, pp. 572–578. DOI: 10.1109/FG.2011.5771460 (cit. on p. 10).

[42] B. Mehtre, N. Murthy, S. Kapoor, and B. Chatterjee. "Segmentation of fingerprint images using the directional image". In: *Pattern Recognition* 20.4 (1987), pp. 429 –435. ISSN: 0031-3203. DOI: http://dx.doi.org/10.1016/0031-3203(87)90069-0 (cit. on p. 10).

[43] L. Shen, A. Kot, and W. Koo. "Quality Measures of Fingerprint Images". In: *IN: PROC. AVBPA, SPRINGER LNCS-2091*. 2001, pp. 266–271 (cit. on pp. 10, 12, 36, 89, 124, 136, 139).

[44] F. Alonso-Fernández, J. Fierrez-Aguilar, and J. Ortega-Garcia. "An enhanced Gabor filter-based segmentation algorithm for fingerprint recognition systems". In: *Proc. 4th Int. Symp. Image and Signal Processing and Analysis ISPA 2005*. 2005, pp. 239–244 (cit. on p. 10).

[45] C. Cortes and V. Vapnik. "Support-Vector Networks". English. In: *Machine Learning* 20.3 (1995), pp. 273–297. ISSN: 0885-6125. DOI: 10.1023/A:1022627411411. URL: http://dx.doi.org/10.1023/A%3A1022627411411 (cit. on pp. 10, 163, 169).

[46] A. M. Bazen and S. H. Gerez. "Segmentation of fingerprint images". In: *ProRISC 2001 Workshop on Circuits, Systems and Signal Processing*. 2001, pp. 276–280 (cit. on p. 10).

[47] Y. Yin, Y. Wang, and X. Yang. "Fingerprint image segmentation based on quadric surface model". In: *Audio-and Video-Based Biometric Person Authentication*. Springer. 2005, pp. 647–655 (cit. on p. 10).

[48] X. Chen, J. Tian, J. Cheng, and X. Yang. "Segmentation of fingerprint images using linear classifier". In: *EURASIP Journal on Applied Signal Processing* 2004 (2004), pp. 480–494 (cit. on p. 10).

[49] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar. *Handbook of Fingerprint Recognition*. 2nd. Springer Publishing Company, Incorporated, 2009. ISBN: 1848822537 (cit. on pp. 10, 32, 39, 70, 124, 125, 147, 150).

[50] K. Cao, E. Liu, and A. Jain. "Segmentation and Enhancement of Latent Fingerprints: A Coarse to Fine RidgeStructure Dictionary". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 36.9 (2014), pp. 1847–1859. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2014.2302450 (cit. on p. 10).

[51] J. M. Libert, J. Grantham, and S. Orandi. *NISTIR 7599 - A 1D Spectral Image Validation/Verification Metric for Fingerprints*. Tech. rep. NIST, Aug. 2009 (cit. on pp. 10, 61).

[52] H. Guan, A. Dienstfrey, and M. Theofanos. "A New Metric for Latent Fingerprint Image Preprocessing". In: *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*. 2013, pp. 84–91. DOI: 10.1109/CVPRW.2013.20 (cit. on p. 11).

[53] S. Yoon and A. Jain. "Is there a fingerprint pattern in the image?" In: *Biometrics (ICB), 2013 International Conference on*. 2013, pp. 1–8. DOI: 10.1109/ICB.2013.6613023 (cit. on p. 11).

[54] N. Sickler and S. Elliott. "An evaluation of fingerprint image quality across an elderly population vis-a-vis an 18-25 year old population". In: *Security Technology, 2005. CCST '05. 39th Annual 2005 International Carnahan Conference on*. 2005, pp. 68–73. DOI: 10.1109/CCST.2005.1594817 (cit. on pp. 11, 27, 125).

[55] C. Blomeke, S. Elliott, B. Senjaya, and G. Hales. "A comparison of fingerprint image quality and matching performance between healthcare and general populations". In: *Biometrics: Theory, Applications, and Systems, 2009. BTAS '09. IEEE 3rd International Conference on*. 2009, pp. 1–4. DOI: 10.1109/BTAS.2009.5339017 (cit. on p. 11).

[56] M. Vatsa, R. Singh, S. Bharadwaj, H. Bhatt, and R. Mashruwala. "Analyzing Fingerprints of Indian Population Using Image Quality: A UIDAI Case Study". In: *Emerging Techniques and Challenges for Hand-Based Biometrics (ETCHB), 2010 International Workshop on*. Aug. 2010, pp. 1–5. DOI: 10.1109/ETCHB.2010.5559279 (cit. on pp. 11, 30).

[57] C. Puri, K. Narang, A. Tiwari, M. Vatsa, and R. Singh. "On Analysis of Rural and Urban Indian Fingerprint Images". English. In: *Ethics and Policy of Biometrics*. Ed. by A. Kumar and D. Zhang. Vol. 6005. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2010, pp. 55–61. ISBN: 978-3-642-12594-2. DOI: 10.1007/978-3-642-12595-9_8. URL: http://dx.doi.org/10.1007/978-3-642-12595-9_8 (cit. on p. 11).

[58] E. Kukula, S. Elliott, H. Kim, and C. San Martin. "The impact of fingerprint force on image quality and the detection of minutiae". In: *Electro/Information Technology, 2007 IEEE International Conference on*. 2007, pp. 432–437. DOI: 10.1109/EIT.2007.4374512 (cit. on pp. 11, 27).

[59] S. J. Fieldhouse. "An Investigation into the Effects of Force Applied During Deposition on Latent Fingermarks and Inked Fingerprints Using a Variable Force Fingerprint Sampler". In: *Journal of Forensic Sciences* 60.2 (2015), pp. 422–427. ISSN: 1556-4029. DOI: 10.1111/1556-4029.12661. URL: http://dx.doi.org/10.1111/1556-4029.12661 (cit. on p. 11).

[60] R. Stewart, M. Estevao, and A Adler. "Fingerprint recognition performance in rugged outdoors and cold weather conditions". In: *Biometrics: Theory, Applications, and Systems, 2009. BTAS '09. IEEE 3rd International Conference on*. 2009, pp. 1–6. DOI: 10.1109/BTAS.2009.5339061 (cit. on p. 11).

[61] D. Simon-Zorita, J. Ortega-Garcia, J. Fierrez-Aguilar, and J. Gonzalez-Rodriguez. "Image quality and position variability assessment in minutiae-based fingerprint verification". In: *Vision, Image and Signal Processing, IEE Proceedings* - 150.6 (2003), pp. 402–408. ISSN: 1350-245X. DOI: 10.1049/ip-vis:20031037 (cit. on p. 11).

[62] C. Blomeke, S. Modi, and S. Elliott. "Investigating the relationship between fingerprint image quality and skin characteristics". In: *Security Technology, 2008. ICCST 2008. 42nd Annual IEEE International Carnahan Conference on*. 2008, pp. 158–161. DOI: 10.1109/CCST.2008.4751295 (cit. on pp. 11, 18, 27, 125).

[63] M. Drahanský. "Fingerprint Recognition Technology: Skin Diseases, Image Quality and Liveness Detection". PhD thesis. Brno University of Technology, 2008 (cit. on p. 11).

[64] P. Krishnasamy, S. Belongie, and D. Kriegman. "Wet Fingerprint Recognition: Challenges and Opportunities". In: *Proceedings of the 2011 International Joint Conference on Biometrics*. IJCB '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 1–7. ISBN: 978-1-4577-1358-3. DOI: 10.1109/IJCB.2011.6117594. URL: http://dx.doi.org/10.1109/IJCB.2011.6117594 (cit. on p. 11).

[65] A Awasthi, K. Venkataramani, and A Nandini. "Image quality quantification for fingerprints using quality-impairment assessment". In: *Applications of Computer Vision (WACV), 2013 IEEE Workshop on*. 2013, pp. 296–302. DOI: 10.1109/WACV.2013.6475032 (cit. on p. 11).

[66] J. Schneider, C. Richardson, F. Kiefer, and V. Govindaraju. "On the Correlation of Image Size to System Accuracy in Automatic Fingerprint Identification Systems". English. In: *Audio- and Video-Based Biometric Person Authentication*. Ed. by J. Kittler and M. Nixon. Vol. 2688. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2003, pp. 895–902. ISBN: 978-3-540-40302-9. DOI: 10.1007/3-540-44887-X_104. URL: http://dx.doi.org/10.1007/3-540-44887-X_104 (cit. on pp. 11, 27).

[67] R. Cappelli, M. Ferrara, and D. Maltoni. "On the Operational Quality of Fingerprint Scanners." In: *Information Forensics and Security, IEEE Transactions on* (2008), pp. 192–202. DOI: 10.1109/TIFS.2008.919336 (cit. on p. 11).

[68] H. Fronthaler, K. Kollreider, and J. Bigun. "Automatic Image Quality Assessment with Application in Biometrics". In: *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW '06. Conference on*. 2006, pp. 30–30. DOI: 10.1109/CVPRW.2006.36 (cit. on p. 12).

[69] F. Alonso-Fernández, J. Fierrez-Aguilar, and J. Ortega-Garcia. "A review of schemes for fingerprint image quality computation". In: *Proc. of the 3rd COST 275 Workshop, COST-275, Hertfordshire, UK*. 2005, pp. 3–6 (cit. on pp. 12, 31).

[70] ISO/IEC. *IS 29794-1:2009. Information technology — Biometric sample quality — Part 1: Framework*. ISO Standard. Jan. 2009 (cit. on pp. 12, 13, 26, 28, 34, 83, 99, 104, 105, 112, 124–126, 139, 146).

[71] E. Lim, X. Jiang, and W. Yau. "Fingerprint quality and validity analysis". In: *Proc. Int. Conf. Image Processing 2002*. Vol. 1. 2002. DOI: 10.1109/ICIP.2002.1038062 (cit. on pp. 12, 42, 44, 105, 124, 135, 147).

[72] T. Chen, X. Jiang, and W. Yau. "Fingerprint image quality analysis". In: *Image Processing, 2004. ICIP '04. 2004 International Conference on*. Vol. 2. 2004, 1253–1256Vol.2. DOI: 10.1109/ICIP.2004.1419725 (cit. on pp. 12, 38, 40, 89, 135, 138).

[73] X. Tao, X. Yang, Y. Zang, X. Jia, and J. Tian. "A novel measure of fingerprint image quality using Principal Component Analysis(PCA)". In: *Biometrics (ICB), 2012 5th IAPR International Conference on*. 2012, pp. 170–175. DOI: 10.1109/ICB.2012.6199804 (cit. on p. 12).

[74] Y. Chen, S. Dass, and A. Jain. "Fingerprint Quality Indices for Predicting Authentication Performance". In: *In: Proc. AVBPA, Springer LNCS-3546*. 2005, pp. 160–170 (cit. on pp. 12, 44).

[75] M. S. Altarawneh, W. L. Woo, and S. S. Dlay. "Objective Fingerprint Image Quality Assessment using Gabor Spectrum Approach". In: *Proc. 15th Int Digital Signal Processing Conf*. 2007, pp. 248–251. DOI: 10.1109/ICDSP.2007.4288565 (cit. on pp. 12, 135).

[76] Z Yao, J. m. Le bars, C Charrier, and C Rosenberger. "Quality Assessment of Fingerprints with Minutiae Delaunay Triangulation". In: *International Conference on Information Systems Security and Privacy*. Angers, France, Feb. 2015. URL: https://hal.archives-ouvertes.fr/hal-01096174 (cit. on p. 12).

[77] J. Qi, D. Abdurrachim, D. Li, and H. Kunieda. "A hybrid method for fingerprint image quality calculation". In: *Automatic Identification Advanced Technologies, 2005. Fourth IEEE Workshop on*. 2005, pp. 124–129. DOI: 10.1109/AUTOID.2005.3 (cit. on p. 13).

[78] E. Tabassi, C. L. Wilson, and C. I. Watson. *NISTIR 7151 - Fingerprint Image Quality*. Tech. rep. NIST, Aug. 2004 (cit. on pp. 13, 15, 24, 28, 32, 89, 134, 135).

[79] E. Tabassi and C. Wilson. "A novel approach to fingerprint image quality". In: *Image Processing, 2005. ICIP 2005. IEEE International Conference on*. Vol. 2. Sept. 2005, pp. II –37–40. DOI: 10.1109/ICIP.2005.1529985 (cit. on pp. 13, 74, 105, 126, 147).

[80] U. D. National Institute of Standards and Technology. *Development of NFIQ 2.0*. Development of NFIQ 2.0. 2014. URL: http://www.nist.gov/itl/iad/ig/development_nfiq_2.cfm (visited on 2014) (cit. on pp. 13, 32, 147, 157).

[81] National Institute of Standards and Technology, U.S. Department of Commerce. *NFIQ 2.0: Evaluation of Potential Image Features for Quality Assessment*. http://biometrics.nist.gov/cs_links/quality/NFIQ_2/IBPC2012/NFIQ2_Feature_Evaluation_v0.5.pdf. 2012 (cit. on pp. 13, 147, 150).

[82] National Institutes of Standards and Technology. "Development of NFIQ 2.0 - Quality Feature Definitions". Version 0.5. In: (2012) (cit. on pp. 13, 104, 105, 107, 147, 157).

[83] E. Lim, K.-A. Toh, P. Suganthan, X. Jiang, and W.-Y. Yau. "Fingerprint image quality analysis". In: *Image Processing, 2004. ICIP '04. 2004 International Conference on*. Vol. 2. 2004, 1241–1244Vol.2. DOI: 10.1109/ICIP.2004.1419530 (cit. on pp. 13, 34, 135).

[84] L. Liu, T. Tan, and Y. Zhan. "Based on SVM Automatic Measures of Fingerprint Image Quality". In: *Proc. Pacific-Asia Workshop Computational Intelligence and Industrial Application PACIIA '08*. Vol. 1. 2008, pp. 575–578. DOI: 10.1109/PACIIA.2008.108 (cit. on p. 13).

[85] M. U. Munir, M. Y. Javed, and S. A. Khan. "A hierarchical k-means clustering based fingerprint quality classification". In: *Neurocomputing* 85.0 (2012), pp. 62–67. ISSN: 0925-2312. DOI: 10.1016/j.neucom.2012.01.002. URL: http://www.sciencedirect.com/science/article/pii/S0925231212000331 (cit. on p. 13).

[86] M. El Abed, A Ninassi, C. Charrier, and C. Rosenberger. "Fingerprint quality assessment using a no-reference image quality metric". In: *Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 21st European*. 2013, pp. 1–5 (cit. on pp. 13, 157).

[87] Z. Yao, C. Charrier, and C. Rosenberger. "Utility Validation of a New Fingerprint Quality Metric". In: *International Biometric Performance Testing Conference (IBPC)*. gattersburg, United States, Apr. 2014, pp. –. URL: https://hal.archives-ouvertes.fr/hal-00988135 (cit. on p. 13).

[88] Z. Yao, J. Le Bars, C. Charrier, and C. Rosenberger. "Fingerprint Quality Assessment Combining Blind Image Quality, Texture and Minutiae Features". In: *ICISSP 2015*. Feb. 2015. URL: https://hal.archives-ouvertes.fr/hal-01097723 (cit. on p. 13).

[89] J. Fierrez-Aguilar, L. M. Muñoz-Serrano, F. Alonso-Fernandez, and J. Ortega-Garcia. "On the effects of image quality degradation on minutiae- and ridge-based automatic fingerprint recognition". In: *Proc. IEEE Intl. Carnahan Conf. on Security Technology, ICCST*. 2005, pp. 79–82 (cit. on pp. 13, 157).

[90] F. Alonso-Fernández, F. Roli, G. L. Marcialis, J. Fierrez, and J. Ortega-Garcia. "Comparison of fingerprint quality measures using an optical and a capacitive sensor". In: *Proc. IEEE Conference on Biometrics: Theory, Applications and Systems, BTAS*. 2007, pp. 1–6 (cit. on pp. 13, 157).

[91] F. Alonso-Fernández, F. Roli, G. Marcialis, J. Fierrez, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. "Performance of fingerprint quality measures depending on sensor technology". In: *SPIE Journal of Electronic Imaging, Special Section on Biometrics: Advances in Security, Usability and Interoperability* 17.1 (2008) (cit. on pp. 13, 157).

[92] F. Alonso-Fernández, J. Fierrez, D. Ramos, and J. Gonzalez-Rodriguez. "Quality-Based Conditional Processing in Multi-Biometrics: application to Sensor Interoperability". In: *IEEE Tansactions on Systems, Man and Cybernetics Part A* 40.6 (2010), pp. 1168–1179 (cit. on p. 13).

[93] J. Hammerle-Uhl, M. Pober, and A. Uhl. "Systematic evaluation methodology for fingerprint-image quality assessment techniques". In: *Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014 37th International Convention on*. 2014, pp. 1315–1319. DOI: 10.1109/MIPRO.2014.6859771 (cit. on p. 13).

[94] M. El-Abed, C. Charrier, and C. Rosenberger. "Quality assessment of image-based biometric information". English. In: *EURASIP Journal on Image and Video Processing* 2015.1, 3 (2015). DOI: 10.1186/s13640-015-0055-8. URL: http://dx.doi.org/10.1186/s13640-015-0055-8 (cit. on p. 13).

[95] S. Bharadwaj, M. Vatsa, and R. Singh. "Biometric quality: a review of fingerprint, iris, and face". English. In: *EURASIP Journal on Image and Video Processing* 2014.1, 34 (2014). DOI: 10.1186/1687-5281-2014-34. URL: http://dx.doi.org/10.1186/1687-5281-2014-34 (cit. on pp. 13, 31, 157).

[96] S. Bharadwaj, M. Vatsa, and R. Singh. *Biometric quality: from assessment to multibiometrics*. Tech. rep. IIITD-TR-2015-003. Department of Computer Science and Engineering, IIIT-Delhi, 2015 (cit. on p. 13).

[97] F. Alonso-Fernández, J. Fierrez, J. Ortega-Garcia, J. Gonzalez-Rodriguez, H. Fronthaler, K. Kollreider, and J. Bigun. "A Comparative Study of Fingerprint Image-Quality Estimation Methods". In: *Information Forensics and Security, IEEE Transactions on* 2.4 (2007), pp. 734–743. ISSN: 1556-6013. DOI: 10.1109/TIFS.2007.908228 (cit. on pp. 13, 31, 89, 124, 135).

[98] C. Jin, H. Kim, X. Cui, E. Park, J. Kim, J. Hwang, and S. Elliott. "Comparative Assessment of Fingerprint Sample Quality Measures Based on Minutiae-Based Matching Performance". In: *Electronic Commerce and Security, 2009. ISECS '09. Second International Symposium on*. Vol. 1. 2009, pp. 309–313. DOI: 10.1109/ISECS.2009.59 (cit. on p. 13).

[99] ISO/IEC. *TR 29794-4:2010. Information technology — Biometric sample quality — Part 4: Finger image data*. Tech. rep. JTC 1/SC 37, 2010 (cit. on pp. 13, 18, 28, 34, 38, 40, 42, 44, 74, 89, 100, 104, 105, 126, 134, 135, 138, 146–148, 150).

[100] ISO/IEC. *29794-6:2010. Information technology — Biometric sample quality — Part 6: Iris image*. Tech. rep. JTC 1/SC 37, 2010 (cit. on p. 13).

[101] S. Shigematsu, H. Morimura, and K. Machida. "A new sensing and digital-conversion scheme with adaptive image-quality adjustment for a fingerprint sensor chip". In: *ASIC/SOC Conference, 2001. Proceedings. 14th Annual IEEE International*. 2001, pp. 396–400. DOI: 10.1109/ASIC.2001.954734 (cit. on p. 13).

[102] J. Galbally, F. Alonso-Fernandez, J. Fierrez, and J. Ortega-Garcia. "A high performance fingerprint liveness detection method based on quality related features". In: *Future Generation Computer Systems* 28.1 (2012), pp. 311–321 (cit. on p. 13).

[103] G. D. C. Cavalcanti, L. Pereira, H. Pinheiro, J. Silva, A. Silva, T. Pina, D. Carvalho, and T. Ren. "A modular architecture based on image quality for fingerprint spoof detection". In: *Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on*. 2012, pp. 258–262. DOI: 10.1109/ICSMC.2012.6377710 (cit. on p. 13).

[104] J. Galbally, S. Marcel, and J. Fierrez. "Image Quality Assessment for Fake Biometric Detection: Application to Iris, Fingerprint, and Face Recognition". In: *Image Processing, IEEE Transactions on* 23.2 (2014), pp. 710–724. ISSN: 1057-7149. DOI: 10.1109/TIP.2013.2292332 (cit. on pp. 14, 28).

[105] F. Alonso-Fernández. "Biometric Sample Quality and its Application to Multimodal Authentication Systems". PhD thesis. Universidad Politécnica de Madrid (UPM), 2008 (cit. on p. 14).

[106] F. Alonso-Fernández, R. Veldhuis, A. Bazen, J Fierrez-Aguilar, and J Ortega-Garcia. "On the relation between biometric quality and user-dependent score distributions in fingerprint verification". In: *Workshop on Multimodal User Authentication, MMUA, Toulouse, France, May 2006*. 2006 (cit. on p. 14).

[107] J. Fierrez-Aguilar, Y. Chen, J. Ortega-garcia, and A. K. Jain. "Incorporating image quality in multi-algorithm fingerprint verification". In: *Proc. IAPR Intl. Conf. on Biometrics, ICB, Springer LNCS-3832*. Springer, 2006, pp. 213–220 (cit. on pp. 14, 24, 134).

[108] N. Poh and J. Kittler. "A Unified Framework for Biometric Expert Fusion Incorporating Quality Measures". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 34.1 (2012), pp. 3–18. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2011.102 (cit. on p. 14).

[109] S. Bharadwaj, H. S. Bhatt, R. Singh, M. Vatsa, and A. Noore. "QFuse: Online learning framework for adaptive biometric system". In: *Pattern Recognition* 0 (2015), pp. –. ISSN: 0031-3203. DOI: http://dx.doi.org/10.1016/j.patcog.2015.05.002. URL: http://www.sciencedirect.com/science/article/pii/S0031320315001673 (cit. on p. 14).

[110] D. P. Pulsifer, S. A. Muhlberger, S. F. Williams, R. C. Shaler, and A. Lakhtakia. "An objective fingerprint quality-grading system". In: *Forensic Science International* 231.1–3 (2013), pp. 204 –207. ISSN: 0379-0738. DOI: http://dx.doi.org/10.1016/j.forsciint.2013.05.003. URL: http://www.sciencedirect.com/science/article/pii/S0379073813002703 (cit. on p. 14).

[111] S. Yoon, K. Cao, E. Liu, and A. Jain. "LFIQ: Latent fingerprint image quality". In: *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*. 2013, pp. 1–8. DOI: 10.1109/BTAS.2013.6712750 (cit. on pp. 14, 157).

[112] P. Grother and E. Tabassi. "Performance of Biometric Quality Measures". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 29.4 (2007), pp. 531–543. DOI: 10.1109/TPAMI.2007.1019 (cit. on pp. 14, 29, 78, 99, 105, 118, 124, 126, 150, 153).

[113] S. Li, H. Kim, C. Jin, S. Elliott, and M. Ma. "Assessing the level of difficulty of fingerprint datasets based on relative quality measures". In: *Information Sciences* 268.0 (2014). New Sensing and Processing Technologies for Hand-based Biometrics Authentication, pp. 122 –132. ISSN: 0020-0255. DOI: http://dx.doi.org/10.1016/j.ins.2013.05.025. URL: http://www.sciencedirect.com/science/article/pii/S0020025513004003 (cit. on p. 14).

[114] Unique Identification Authority of India. *UIDAI Strategy Overview Creating a Unique Identity Number for Every Resident in India*. 2010 (cit. on pp. 15, 25).

[115] European Commission. "Eurodac: a European Union-wide electronic system for the identification of asylum-seekers MEMO/06/334 ". In: (2006) (cit. on p. 15).

[116] European Council. "Council Regulation (EC) No 343/2003 of 18 February 2003 establishing the criteria and mechanisms for determining the Member State responsible for examining an asylum application lodged in one of the Member States by a third-country national". In: (2003) (cit. on p. 15).

[117] European Council. "Council Decision 2007/533/JHA of 12 June 2007 on the establishment, operation and use of the second generation Schengen Information System (SIS II)". In: *Official Journal of the European Union* (2007) (cit. on p. 15).

[118] European Council. "Commission Implementing Decision (EU) 2015/219 of 29 January 2015 replacing the Annex to Implementing Decision 2013/115/EU on the Sirene Manual and other implementing measures for the second generation Schengen Information System (SIS II) (notified under document C(2015) 326)". In: (2015) (cit. on p. 15).

[119] European Commission, European Parliament. "Regulation (EC) No 1987/2006 of the European Parliament and of the Council of 20 December 2006 on the establishment, operation and use of the second generation Schengen Information System (SIS II)". In: (2006) (cit. on p. 15).

[120] C. L. Wilson, M. D. Garris, and C. I Watson. *Matching Performance for the US-VISIT IDENT System Using Flat Fingerprints NISTIR 7110*. Tech. rep. NIST, May 2004 (cit. on pp. 15, 24, 30, 124, 134, 157).

[121] M. A. Olsen, V. Smida, and C. Busch. *Fingerprint Quality Assessment Algorithms*. Online. `http://www.nislab.no`, `https://www.dasec.h-da.de`. 2015 (cit. on pp. 18, 31, 52).

[122] *da/sec WDSET02 multisensor fingerprint database*. Online. `https://www.dasec.h-da.de/research/biometrics/wdset02/` (cit. on pp. 18, 124, 126, 132).

[123] European Council. "Establishing the Visa Information System (VIS) (2004/512/EC)". In: *Official Journal of the European Union* (2004), pp. 5–7 (cit. on pp. 24, 88).

[124] European Commission. "Determining the first regions for the start of operations of the Visa Information System (VIS), C(2009) 8542". In: *Official Journal of the European Union* 23 (2010), pp. 62–64 (cit. on pp. 24, 88).

[125] Morpho. *Secure Biometric Access - Morpho*. 2014. URL: `http://www.morpho.com/identification/secure-biometric-access/` (visited on 2014) (cit. on p. 25).

[126] P. Kneidinger and M. Schwaiger. "Evaluation and monitoring of fingerprint acquisitions for the European Visa Information System experiences from Austria". In: *Biometrics Special Interest Group (BIOSIG), 2012 BIOSIG - Proceedings of the International Conference of the*. Biometrics Special Interest Group (BIOSIG), 2012 BIOSIG - Proceedings of the International Conference of the. 2012, pp. 1–12 (cit. on p. 25).

[127] Unique Identification Authority of India. *Role of Biometric Technology in Aadhaar Enrollment*. 2012 (cit. on pp. 25, 88).

[128] Unique Identification Authority of India. *Exclusion to Inclusion with Micropayments*. http://uidai.gov.in/UID_PDF/Front_Page_Articles/Strategy/Exclusion_to_Inclusion_with_Micropayments.pdf – accessed Februar 2013. 2010 (cit. on p. 25).

[129] A. Ross and A. K. Jain. "Biometric Sensor Interoperability: A Case Study in Fingerprints." In: *ECCV Workshop BioAW*. 2004, pp. 134–145. DOI: 10.1007/978-3-540-25976-3_13 (cit. on p. 27).

[130] Bundesamt für Sicherheit in der Informationstechnik. "Technical Guideline TR-03121-1, Biometrics for Public Sector Applications, Part 1: Framework, Version 2.3". In: 2011 (cit. on p. 30).

[131] 3M Cogent. *3M Cogent*. 2014. URL: Online.\url{http://solutions.3m.com/wps/portal/3M/en_US/Security/Identity_Management/3M_Cogent/} (visited on 2014) (cit. on p. 30).

[132] J. Merkle, M. Schwaiger, and M. Breitenstein. "Towards Improving the NIST Fingerprint Image Quality (NFIQ) Algorithm." In: *BIOSIG*. 2010, pp. 29–44 (cit. on pp. 32, 135).

[133] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain. "FVC2004: Third Fingerprint Verification Competition". In: *ICBA*. 2004, pp. 1–7. DOI: 10.1007/978-3-540-25948-0_1 (cit. on pp. 33, 46, 47, 139, 150).

[134] R. Cappelli, M. Ferrara, A. Franco, and D. Maltoni. "Fingerprint verification competition 2006". In: *Biometric Technology Today* 15.7–8 (2007), pp. 7 –9. ISSN: 0969-4765. DOI: http://dx.doi.org/10.1016/S0969-4765(07)70140-6. URL: http://www.sciencedirect.com/science/article/pii/S0969476507701406 (cit. on pp. 46, 47).

[135] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez-Zanuy, V. Espinosa, A. Satue, I. Hernaez, J.-J. Igarza, C. Vivaracho, D. Escudero, and Q.-I. Moro. "MCYT baseline corpus: a bimodal biometric database". In: *IEEE Proceedings -Vision, Image and Signal Processing* 150.6 (2003), pp. 395–401. DOI: 10.1049/ip-vis:20031078 (cit. on pp. 46, 47, 64, 78).

[136] P. D. Kovesi. *MATLAB and Octave Functions for Computer Vision and Image Processing*. Available from: <http://www.peterkovesi.com/matlabfns/>. 2005 (cit. on p. 47).

[137] B. S. Swann, J. M. Libert, and M. A. Lepley. "Tools for quality control of fingerprint databases". In: *Proc. SPIE* 7667.766709 (2010) (cit. on p. 61).

[138] R. Ghrist. "Three examples of applied and computational homology". In: *Nieuw Archief voor Wiskunde* 5/9(2) (2008) (cit. on p. 61).

[139] V. de Silva and G. Carlsson. "Topological estimation using witness complexes". In: *SPBG04 Symposium on Point-Based Graphics*. 2004, pp. 157–166 (cit. on p. 63).

[140] Chinese Academy of Sciences' Institute of Automation (CASIA). *CASIA-FingerprintV5*. Online. CASIA-FingerprintV5, http://biometrics.idealtest.org. 2010 (cit. on pp. 64, 78, 106, 114, 150).

[141] Chinese Academy of Sciences' Institute of Automation (CASIA). *CASIA-IrisV4*. Online. CASIA-IrisV4, http://biometrics.idealtest.org. 2010 (cit. on p. 64).

[142] BioSecure. *BioSecure*. Online. http://biosecure.it-sudparis.eu/. Association BioSecure, 2006 (cit. on p. 64).

[143] D. Maio, D. Maltoni, R. Cappelli, J. Wayman, A. K. Jain, and A. K. Jain. "FVC2002: Second Fingerprint Verification Competition". In: 2002, pp. 811–814 (cit. on pp. 65, 76, 139).

[144] Chinese Academy of Sciences' Institute of Automation (CASIA). *CASIA-FaceV5*. Online. CASIA-FaceV5, http://biometrics.idealtest.org. 2010 (cit. on p. 67).

[145] ISO/IEC. "ISO/IEC 19794-1:2011 Information technology – Biometric data interchange formats – Part 1: Framework". In: (2011) (cit. on pp. 70, 76).

[146] D. Maio, D. Maltoni, J. L. Wayman, and A. K. Jain. "FVC2000: Fingerprint verification competition". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (2002), pp. 402–412 (cit. on p. 71).

[147] T. Liu, G. Zhu, C. Zhang, and P. Hao. "Fingerprint indexing based on singular point correlation". In: *Image Processing, 2005. ICIP 2005. IEEE International Conference on*. Vol. 3. IEEE. 2005, pp. II–293 (cit. on p. 70).

[148] L. Wei. "Fingerprint Classification using Singularities Detection". In: *international journal of mathematics and computers in simulation* 2.2 (2008), pp. 158–162 (cit. on p. 70).

[149] F. Ahmad and D. Mohamad. "A review on fingerprint classification techniques". In: *Computer Technology and Development, 2009. ICCTD'09. International Conference on*. Vol. 2. IEEE. 2009, pp. 411–415 (cit. on p. 70).

[150] A. Bazen and S. Gerez. "Systematic methods for the computation of the directional fields and singular points of fingerprints". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24.7 (July 2002), pp. 905–919. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2002.1017618 (cit. on pp. 70, 71, 73, 76).

[151] D. Weng, Y Yin, and D. Yang. "Singular points detection based on multi-resolution in fingerprint images". In: *Neurocomputing* 74.17 (2011), pp. 3376 –3388. ISSN: 0925-2312. DOI: http://dx.doi.org/10.1016/j.neucom.2011.05.023 (cit. on p. 70).

[152] C. Jin and H. Kim. "Pixel-level singular point detection from multi-scale Gaussian filtered orientation field". In: *Pattern Recogn.* 43.11 (Nov. 2010), pp. 3879–3890. ISSN: 0031-3203. DOI: 10.1016/j.patcog.2010.05.023. URL: http://dx.doi.org/10.1016/j.patcog.2010.05.023 (cit. on p. 70).

[153] A. K. Jain, S. Prabhakar, and L. Hong. "A multichannel approach to fingerprint classification". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 21.4 (1999), pp. 348–359 (cit. on p. 71).

[154] J. Gu, J. Zhou, and D. Zhang. "A combination model for orientation field of fingerprints". In: *Pattern Recognition* 37.3 (2004), pp. 543 –553. ISSN: 0031-3203. DOI: http://dx.doi.org/10.1016/S0031-3203(03)00178-X. URL: http://www.sciencedirect.com/science/article/pii/S003132030300178X (cit. on p. 71).

[155] Y. Wang, J. Hu, and F. Han. "Enhanced gradient-based algorithm for the estimation of fingerprint orientation fields". In: *Applied Mathematics and Computation* 185.2 (2007), pp. 823–833 (cit. on p. 71).

[156] M. Kass and A. Witkin. "Analyzing oriented patterns". In: *Computer vision, graphics, and image processing* 37.3 (1987), pp. 362–385 (cit. on p. 72).

[157] M. Kawagoe and A. Tojo. "Fingerprint pattern classification". In: *Pattern Recogn.* 17.3 (June 1984), pp. 295–303. ISSN: 0031-3203. DOI: 10.1016/0031-3203(84)90079-7. URL: http://dx.doi.org/10.1016/0031-3203(84)90079-7 (cit. on p. 73).

[158] A. M. Bazen and R. N. J. Veldhuis. "Likelihood-ratio-based biometric verification". In: *Circuits and Systems for Video Technology, IEEE Transactions on* 14.1 (2004), pp. 86–94 (cit. on p. 74).

[159] C. I. Watson. *NIST Special Database 14*. Tech. rep. National Institute of Standards and Technology, 1993 (cit. on pp. 76, 160, 162).

[160] C. Busch, D. Lodrova, E. Tabassi, and W. Krodel. "Semantic conformance testing for finger minutiae data". In: *Security and Communication Networks (IWSCN), 2009 Proceedings of the 1st International Workshop on*. 2009, pp. 1–7 (cit. on pp. 76, 160).

[161] European Commission. "Proposal for a Council Decision establishing the Visa Information System (VIS) COM/2004/0099". In: (2004) (cit. on p. 88).

[162] National Institute of Standards and Technology. *Proceeding of the 2nd International Biometric Performance Conference*. 2012. URL: http://www.nist.gov/itl/iad/ig/ibpc2012_presentations.cfm (cit. on p. 88).

[163] T. Kohonen. "The self-organizing map". In: 78.9 (1990), pp. 1464–1480. DOI: 10.1109/5.58325 (cit. on pp. 88, 90).

[164] L. Breiman. "Random Forests". In: vol. 45. 1. Hingham, MA, USA: Kluwer Academic Publishers, Oct. 2001, pp. 5–32. DOI: 10.1023/A:1010933404324. URL: http://dx.doi.org/10.1023/A:1010933404324 (cit. on pp. 88, 97, 104, 107, 110).

[165] R. Duda and P. Hart. *Pattern Classification and Scene Analysis*. New York, NY, USA: John Wiley and Sons, 1973 (cit. on p. 91).

[166] D. Willett, C. Busch, and F. Seibert. "Fast Image Analysis using Kohonen Maps". In: *Proceedings of the IEEE Workshop NNSP*. 1994, pp. 461–470 (cit. on p. 92).

[167] H. Fronthaler, K. Kollreider, and J. Bigun. "Local Features for Enhancement and Minutiae Extraction in Fingerprints". In: *Image Processing, IEEE Transactions on* 17.3 (2008), pp. 354–363. ISSN: 1057-7149. DOI: 10.1109/TIP.2007.916155 (cit. on pp. 98, 106).

[168] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain. "FVC2004: Third Fingerprint Verification Competition". In: *Biometric Authentication*. Vol. 3072. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2004, pp. 1–7. ISBN: 978-3-540-22146-3. DOI: 10.1007/978-3-540-25948-0_1. URL: http://dx.doi.org/10.1007/978-3-540-25948-0_1 (cit. on p. 98).

[169] European Commission. *Visa Information System*. http://ec.europa.eu/dgs/home-affairs/what-we-do/policies/borders-and-visas/visa-information-system/index_en.htm. [Online; accessed 25-August-2013] (cit. on p. 104).

[170] *OBIM, U.S. Department of Homeland Security*. Online. www.dhs.gov/obim (cit. on pp. 104, 125, 150).

[171] T. Kohonen. "Self-organized formation of topologically correct feature maps". In: *Biological Cybernetics* 43.1 (1982), pp. 59–69. ISSN: 0340-1200. DOI: 10.1007/BF00337288. URL: http://dx.doi.org/10.1007/BF00337288 (cit. on pp. 104, 107).

[172] T. Kohonen, M. R. Schroeder, and T. S. Huang, eds. *Self-Organizing Maps*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2001. ISBN: 3540679219. URL: \url{http://portal.acm.org/citation.cfm?id=558021} (cit. on pp. 104, 106, 107).

[173] J. F. Svensen. "GTM: the generative topographic mapping". PhD thesis. Aston Triangle, Birmingham B4 7ET, United Kingdom: Aston University, 1998 (cit. on pp. 104, 108, 109).

[174] National Institute of Standards and Technology. *NIST Biometric Image Software*. http://www.nist.gov/itl/iad/ig/nbis.cfm. [Online; accessed 15-September-2013] (cit. on p. 105).

[175] T. Kohonen. "Essentials of the self-organizing map". In: *Neural Networks* 37 (2013), pp. 52–65. ISSN: 08936080. DOI: 10.1016/j.neunet.2012.09.018. URL: http://dx.doi.org/10.1016/j.neunet.2012.09.018 (cit. on pp. 106–108).

[176] A. Makarov. "Fingerprint Quality Estimation Using Self-Organizing Maps". MA thesis. Technical University of Denmark, DTU Informatics, 2012. URL: http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=6348 (cit. on p. 106).

[177] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. 1st ed. 2006. Corr. 2nd printing 2011. Springer, Oct. 2007. ISBN: 0387310738. URL: http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20\&path=ASIN/0387310738 (cit. on p. 109).

[178] A. P. Dempster, N. M. Laird, and D. B. Rubin. "Maximum likelihood from incomplete data via the EM algorithm". In: *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B* 39.1 (1977), pp. 1–38 (cit. on p. 110).

[179] J. R. Quinlan. "Induction of decision trees". In: *Machine Learning* 1.1 (Mar. 1986), pp. 81–106. ISSN: 0885-6125. DOI: 10.1007/bf00116251. URL: http://dx.doi.org/10.1007/BF00116251 (cit. on p. 110).

[180] L. Breiman. "Bagging predictors". English. In: *Machine Learning* 24.2 (1996), pp. 123–140. ISSN: 0885-6125. DOI: 10.1007/BF00058655. URL: http://dx.doi.org/10.1007/BF00058655 (cit. on p. 110).

[181] S. Roweis. "EM Algorithms for PCA and SPCA". In: *in Advances in Neural Information Processing Systems*. MIT Press, 1998, pp. 626–632 (cit. on p. 111).

[182] T. Kohonen, J. Hynninen, J. Kangas, and J. Laaksonen. *SOM PAK: The Self-Organizing Map program package*. 1996. URL: http://www.cis.hut.fi/research/som_lvq_pak.shtml (cit. on p. 115).

[183] C. Spearman. "The proof and measurement of association between two things. By C. Spearman, 1904." In: *The American journal of psychology* 100.3-4 (1987), pp. 441–471. ISSN: 0002-9556. URL: http://view.ncbi.nlm.nih.gov/pubmed/3322052 (cit. on p. 118).

[184] *Aadhaar, Unique Identification Authority of India*. Online. uidai.gov.in (cit. on pp. 125, 150).

[185] *Face Care, Shenzhen Kier Electronic Apparatus Factory*. Online. http://www.szkier.com/kier/en/about.asp (cit. on p. 126).

[186] E. Alanen, J. Nuutinen, K. Nicklén, T. Lahtinen, and J. Mönkkönen. "Measurement of hydration in the stratum corneum with the MoistureMeter and comparison with the Corneometer". In: *Skin Research and Technology* 10.1 (2004), pp. 32–37. ISSN: 1600-0846. DOI: 10.1111/j.1600-0846.2004.00050.x. URL: http://dx.doi.org/10.1111/j.1600-0846.2004.00050.x (cit. on p. 126).

[187] P. Clarys, R. Clijsen, J. Taeymans, and A. O. Barel. "Hydration measurements of the stratum corneum: comparison between the capacitance method (digital version of the Corneometer CM 825®) and the impedance method (Skicon-200EX®)". In: *Skin Research and Technology* 18.3 (2012), pp. 316–323. ISSN: 1600-0846. DOI: 10.1111/j.1600-0846.2011.00573.x. URL: http://dx.doi.org/10.1111/j.1600-0846.2011.00573.x (cit. on p. 126).

[188] O. Bausinger and E. Tabassi. "Fingerprint Sample Quality Metric NFIQ 2.0". In: *Proceedings of the International Conference of the Special Interest Group on Biometrics (BIOSIG 2011), September 8-9*. 2011, pp. 167–171 (cit. on p. 135).

[189] J. G. Daugman. "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters". In: *J. Opt. Soc. Am. A* 2.7 (July 1985), pp. 1160–1169. DOI: 10.1364/JOSAA.2.001160. URL: http://josaa.osa.org/abstract.cfm?URI=josaa-2-7-1160 (cit. on p. 135).

[190] Y. Hamamoto, S. Uchimura, M. Watanabe, T. Yasuda, Y. Mitani, and S. Tomita. "A gabor filter-based method for recognizing handwritten numerals". In: *Pattern Recognition* 31 (4 1998), pp. 395–400. DOI: 10.1016/S0031-3203(97)00057-5 (cit. on p. 135).

[191] A. Jain, S. Prabhakar, L. Hong, and S. Pankanti. "Filterbank-Based Fingerprint Matching". In: *IEEE Trans. Image Processing* 9.5 (May 2000), pp. 846–859 (cit. on p. 136).

[192] E. Tabassi and P. Grother. "Fingerprint Image Quality". In: *Encyclopedia of Biometrics*. 2009, pp. 482–490 (cit. on p. 146).

[193] K. Pearson. "On lines and planes of closest fit to systems of points in space". In: *Philosophical Magazine* 2.6 (1901), pp. 559–572 (cit. on p. 148).

[194] Expert Working Group on Human Factors in Latent Print Analysis. *Latent print examination and human factors : improving the practice through a systems approach*. US Department of Commerce, National Institute of Standards and Technology, 2012 (cit. on p. 156).

[195] B. T. Ulery, R. A. Hicklin, J. Buscaglia, and M. A. Roberts. "Accuracy and reliability of forensic latent fingerprint decisions". In: *Proceedings of the National Academy of Sciences* 108.19 (2011), pp. 7733–7738. DOI: 10.1073/pnas.1018707108. eprint: http://www.pnas.org/content/108/19/7733.full.pdf+html. URL: http://www.pnas.org/content/108/19/7733.abstract (cit. on p. 156).

[196] B. T. Ulery, R. A. Hicklin, J. Buscaglia, and M. Antonia Roberts. "Repeatability and Reproducibility of Decisions by Latent Fingerprint Examiners". In: *PLoS ONE* 7.3 (2012). DOI: 10.1371/journal.pone.0032800 (cit. on pp. 156, 158).

[197] G. Langenburg. "A Performance Study of the ACE-V Process: A Pilot Study to Measure the Accuracy, Precision, Reproducibility, Repeatability, and Biasability of Conclusions Resulting from the ACE-V Process". In: *Journal of Forensic Identification* 59.2 (2009), pp. 219–257 (cit. on p. 156).

[198] I. E. Dror, D. Charlton, and A. E. Péron. "Contextual information renders experts vulnerable to making erroneous identifications". In: *Forensic Science International* 156.1 (2005), pp. 74–78. DOI: 10.1016/j.forsciint.2005.10.017 (cit. on p. 156).

[199] I. E. Dror and D. Charlton. "Why experts make errors". In: *Journal of Forensic Identification* 56.4 (2006), p. 600 (cit. on p. 156).

[200] D. Zwillinger and S. Kokoska. *CRC Standard Probability and Statistics Tables and Formulae*. 1st ed. CRC Press, Dec. 1999. ISBN: 9781584880592 (cit. on p. 158).

[201] D. C. Hoaglin, F. Mosteller, and J. W. Tukey, eds. *Understanding Robust and Exploratory Data Analysis (Wiley Series in Probability and Statistics)*. 1st ed. Wiley, Dec. 1982. ISBN: 9780471097778 (cit. on p. 158).

[202] J. Cohen. "A Coefficient of Agreement for Nominal Scales". In: *Educational and Psychological Measurement* 20.1 (1960), pp. 37–46. DOI: 10.1177/001316446002000104. eprint: http://epm.sagepub.com/content/20/1/37.full.pdf+html. URL: http://epm.sagepub.com/content/20/1/37.short (cit. on pp. 160, 163).

[203] J. L. Fleiss. "Measuring nominal scale agreement among many raters." In: *Psychological bulletin* 76.5 (1971), p. 378 (cit. on p. 160).

[204] C. I. Watson. *NIST Special Database 29*. Tech. rep. National Institute of Standards and Technology, 2001 (cit. on p. 160).

[205] id3 Technologies. *id3 Fingerprint SDK v. 1.4.1*. http://www.id3.eu/. 2014 (cit. on p. 162).

[206] R. D. McKelvey and W. Zavoina. "A statistical model for the analysis of ordinal level dependent variables". In: *The Journal of Mathematical Sociology* 4.1 (1975), pp. 103–120. DOI: 10.1080/0022250X.1975.9989847 (cit. on p. 163).

[207] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2014. URL: http://www.R-project.org/ (cit. on p. 163).

[208] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch. *e1071: Misc Functions of the Department of Statistics (e1071), TU Wien*. R package version 1.6-4. 2014. URL: http://CRAN.R-project.org/package=e1071 (cit. on p. 163).

[209] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Fourth. ISBN 0-387-95457-0. New York: Springer, 2002. URL: http://www.stats.ox.ac.uk/pub/MASS4 (cit. on p. 163).

[210] M. Kuhn, J. Wing, S. Weston, A. Williams, C. Keefer, A. Engelhardt, T. Cooper, Z. Mayer, B. Kenkel, the R Core Team, M. Benesty, R. Lescarbeau, A. Ziem, and L. Scrucca. *caret: Classification and Regression Training*. R package version 6.0-41. 2015. URL: http://CRAN.R-project.org/package=caret (cit. on p. 163).

[211] D. B. Dahl. *xtable: Export tables to LaTeX or HTML*. R package version 1.7-4. 2014. URL: http://CRAN.R-project.org/package=xtable (cit. on p. 163).