BioTwist

Overcoming severe distortions in ridge-based biometrics for successful identification

Johannes Kotzerke

Composition of the Graduation Committee:

Chairman and secretary		
Prof.Dr.	P.M.G. Apers	University of Twente, Netherlands
Supervisors		
Prof.Dr.	K.J. Horadam	RMIT University, Australia
Prof.Dr.Ir.	R.N.J. Veldhuis	University of Twente, Netherlands
Co-supervisors		
Dr.Ir.	L.J. Spreeuwers	University of Twente, Netherlands
Members		
Prof.Dr.	D. Meuwly	University of Twente, Netherlands Netherlands Forensic Institute, Netherlands
Prof.Dr.	M. Junger	University of Twente, Netherlands
Dr.	A. Ross	Michigan State University, USA
Prof.Dr.	O.R.P. Bellon	Federal University of Paraná, Brazil
Prof.Dr.	M.J. Sjerps	Universiteit van Amsterdam, Netherlands Netherlands Forensic Institute, Netherlands

The doctoral research of J. Kotzerke was financially supported by ARC Discovery Grant DP120101188 and the Victoria Police.

CTIT Ph.D. Thesis Series No. 16-383
Centre for Telematics and Information Technology
P.O. Box 217, 7500 AE
Enschede, The Netherlands

ISBN 978-90-365-4061-2
ISSN 1381-3617 (CTIT Ph.D. thesis Series No. 16-383)
DOI 10.3990/1.9789036540612
http://dx.doi.org/10.3990/1.9789036540612

Cover: Visualisation of the frequency for significant words to occur in this thesis.

BIOTWIST

OVERCOMING SEVERE DISTORTIONS IN RIDGE-BASED BIOMETRICS FOR SUCCESSFUL IDENTIFICATION

DISSERTATION

to obtain a double-badged degree, namely:
the degree of doctor of philosophy at the University of Twente
and the Royal Melbourne Institute of Technology,
on the authority of the Rector Magnificus of the University of Twente
Prof.Dr. H. Brinksma
and the Council of the Royal Melbourne Institute of Technology
on account of the decision of the graduation committee,
to be publicly defended at the University of Twente
on Thursday, the 19th of May, 2016 at 14:45.

by

Johannes Kotzerke
born on the 17th of January, 1987
in Hanover, Germany

This dissertation has been approved by

Senior Supervisor:	Prof.Dr. K.J. Horadam	School of Mathematical and Geospatial Sciences RMIT University Melbourne, Australia
Joint Senior Supervisor:	Dr. S.A. Davis	
Associate Supervisor:	Dr. A. Arakala	
Senior Supervisor:	Prof.Dr.Ir. R.N.J. Veldhuis	Services, Cybersecurity and Safety University of Twente Enschede, Netherlands
Joint Senior Supervisor:	Dr.Ir. L.J. Spreeuwers	

Declaration

I certify that except where due acknowledgement has been made, the work is that of the author alone; the work has not been submitted previously, in whole or in part, to qualify for any other academic award; the content of the thesis/project is the result of work which has been carried out since the official commencement date of the approved research program; any editorial work, paid or unpaid, carried out by a third party is acknowledged; and, ethics procedures and guidelines have been followed.

Johannes Kotzerke
19th April 2016

Publications

J. Kotzerke, S. Davis, K. Horadam, and J. McVernon, "Newborn and infant footprint crease pattern extraction," in *Image Processing (ICIP), 2013 20th IEEE International Conference on*, Melbourne, Australia, Sept 2013, pp. 4181–4185.

J. Kotzerke, A. Arakala, S. Davis, K. Horadam, and J. McVernon, "Ballprints as an infant biometric: A first approach," in *2014 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BioMS 2014)*, Rome, Italy, Oct 2014, pp. 36–43.

J. Kotzerke, S. A. Davis, R. Hayes, L. J. Spreeuwers, R. N. J. Veldhuis, and K. J. Horadam, "Discriminating fingermarks with evidential value for forensic comparison," in *Biometrics and Forensics (IWBF), 2015 International Workshop on*, Gjøvik, Norway, Mar 2015, pp. 1–6.

J. Kotzerke, S. Davis, R. Hayes, L. Spreeuwers, R. Veldhuis, and K. Horadam, "Identification performance of evidential value estimation for fingermarks," in *Biometrics Special Interest Group (BIOSIG), 2015 International Conference of the*, Darmstadt, Germany, Sept 2015, pp. 1–6.

J. Kotzerke, H. Hao, S. Davis, R. Hayes, L. Spreeuwers, R. Veldhuis, and K. Horadam, "Identification performance of evidential value estimation for ridge-based biometrics," *EURASIP Journal on Information Security (JIS) on "Advances in Biometrics 2015"*, submitted for publication.

J. Kotzerke, S. Davis, R. Hayes, J. McVernon and K. Horadam, "Newborn and infant identification: revisiting footprints," *Forensic Science International*, submitted for publication.

J. Kotzerke, S. Davis, J. McVernon and K. Horadam, "A Solution to the Infant Biometric Problem," *Nature*, submitted for publication.

Acknowledgements

I welcome this opportunity to thank Prof. Kathy Horadam, Dr. Stephen Davis and Dr. Arathi Arakala for their support, guidance and encouragement throughout the entire project. Additionally, I am grateful for Prof. Horadam's continuous support in dealing with administrative issues and numerous coffee breaks with Dr. Davis and Dr. Arakala which often evolved into fruitful discussions.

Similarly, I thank Prof. Raymond Veldhuis, Dr. Luuk Spreeuwers and Prof. Didier Meuwly for the opportunity to spend a research semester at the University of Twente and their support, guidance and invaluable suggestions during this time and afterwards.

I thankfully acknowledge the Victoria Police that funded the research project on fingermarks and the Victoria Police fingerprint examiners who kindly assessed all fingermarks regarding their forensic value. I also thank our Victoria Police contact Robert Hayes for his involvement into the process.

Similarly, I acknowledge Paula Nathan for capturing the Happy Feet database, Nekoosa Coated Products and Belly Art for assistance in supplying the inkless paper system, the Bill & Melinda Gates Foundation that funded the data collection with their Challenges Explorations Grant OPP1046031 and the Australian Research Council that funded the conducted research with the ARC Discovery Grant DP120101188.

Furthermore, I thank Dr. Heinz Hofbauer, Prof. Andreas Uhl and Dr. Peter Wild for supplying the adult ballprint database for reference purposes. I also acknowledge Jacobien Carstens for her help translating the abstract from English into Dutch.

I thank Marc van der Meer for help with accommodation and recreation, and for his warm welcome when I arrived in Melbourne for the first time. I thank my family for their advice, support and encouragement. Finally, I dearly thank Caroline Crites for proof reading this thesis but more importantly for her continuous support. I would not have been able to accomplish this without her love and understanding.

Melbourne, 19th April 2016 J. K.

Contents

List of Figures

List of Tables

List of Abbreviations

ACE-V	Analysis, Comparison, Evaluation, and Verification
AFIS	Automated Fingerprint Identification System
AHE	Adaptive Histogram Equalisation
AUROC	Area Under the ROC curve
BGM	Biometric Graph Matching
ccID	correct and confident identification
CDC	Centers for Disease Control and Prevention
CLAHE	Contrast Limited Adaptive Histogram Equalisation
CRE	Capture Resolution Estimation
DA	Discriminant Analysis
EER	Equal Error Rate
EV	Evidential Value
EVA	Evidential Value Algorithm
FAR	False Accept Rate
FDA	Fisher's Discriminant Analysis
FMR	False Match Rate
FNMR	False Non-match Rate
FNR	False Negative Rate
FPR	False Positive Rate

FRR False Reject Rate

FTA Failure to Acquire

FTE Failure to Enroll

FVC Fingerprint Verification Competition

GAR Genuine Acceptance Rate

HE Histogram Equalisation

IAFIS Integrated AFIS

ICP Iterative Closest Point

IRS inter-ridge spacing

kNN k-nearest neighbours algorithm

LDA Linear Discriminant Analysis

LR Likelihood Ratio

MCS Maximum Common Subgraph

NFIQ NIST Fingerprint Image Quality

NFIQ2 NIST Fingerprint Image Quality 2.0

NIST National Institute of Standards and Technology

NV No Value

OCL Orientation Certainty Level

PDF Probability Density Function

QDA Quadratic Discriminant Analysis

QS Quick Score

RAM Random-Access Memory

RANSAC Random Sample Consensus

RLAPS Radially Limited Averaged Power Spectrum

RLAPS* Binned Radially Limited Averaged Power Spectrum

RLC Ridge Line Count

RLT Ridge Line Tracer

ROC Receiver Operating Characteristic

RPS Radial Power Spectrum

RQ Research Question

sBGM scale-invariant BGM

SD Standard Deviation

SDK Software Development Kit

SE Structure Element

SIFT Scale-Invariant Feature Transform

SVM Support Vector Machine

TAR True Accept Rate

TNR True Negative Rate

TPR True Positive Rate

UNICEF United Nations Children's Emergency Fund

VEO Value for Exclusion Only

VF Neurotechnology Verifinger

VID Value for Individualisation

WHO World Health Organization

Summary

Biometrics are part of everyday life. They are used to solve crimes, to cross borders, to record work attendance and to unlock smartphones. They all rely on a physical trait's permanence and stability over time, as well as its individuality, robustness and ease to be captured. Challenges arise when a biometric's capture suffers from low quality such that accurate extraction of (physical) features is difficult. This issue becomes critical when working with newborns or infants because of the tininess and fragility of an infant's features, their uncooperative nature and their rapid growth. The last of these is particularly relevant when one tries to verify an infant's identity based on captures of a biometric taken at an earlier age. Finding a physical trait that is feasible for infants is often referred to as the "*infant biometric problem*". Similarly, the quality of fingermarks – impressions of ridge skin pattern that are accidentally left at the scene of a crime or accident – is a critical issue for human experts. Low quality has led to several high profile false identifications.

This thesis explores the quality aspect of adult fingermarks and the correlation between image quality and the mark's usefulness for an ongoing forensic investigation, and researches various aspects of the "ballprint" as an infant biometric. The ballprint refers to the friction ridge skin area of the foot pad under the big toe.

It is found that image quality feature sets can be used to infer automatically if a fingermark is of evidential value in an ongoing forensic investigation. If so, it is worthwhile to compare it against a reference database because it is very likely to retrieve its mate correctly and with confidence.

A longitudinal footprint and ballprint database has been collected from 54 infants within 3 days of birth, at two months old, at 6 months and at 2 years. It has been observed that the skin of a newborn's foot dries and cracks so the ridge lines are often not visible to the naked eye and an adult fingerprint scanner cannot capture their features.

Summary

The ballprint exhibits similar properties to fingerprint but the ball possesses larger physical structures and a greater number of features. This increases the likelihood of captures being of reasonable quality and containing enough information for identification or verification when working with newborns and infants.

This thesis presents the physiological discovery that the ballprint grows isotropically during infancy and that it can be well approximated by a linear function of the infant's age. One consequence is that the mature fingerprint technology that has been developed for adult fingerprints can compare ballprints if they are adjusted by a physical feature (the inter-ridge spacing) to be of a similar size to adult fingerprints. The growth in ballprint inter-ridge spacing mirrors infant growth in terms of length/height.

When growth is compensated for by isotropic scaling, impressive verification scores are achieved even for captures taken 22 months apart. The scores improve even further if low-quality prints are rejected. Serendipitously, the infant ballprints could be ranked by quality using the same classification algorithm built for adult fingermarks and when the worst third were removed the Equal Error Rates dropped from $1-2\%$ to 0%.

Additionally, verification is performed on newborn ballprints achieving results comparable to other infant biometrics reported in the literature.

In conclusion, this thesis demonstrates that the ballprint is a feasible solution to the infant biometric problem.

Samenvatting

Biometrische gegevens zijn onderdeel van het dagelijks leven. Ze worden gebruikt om misdaden op te lossen, grenzen over te steken, werktijden te registreren en smartphones te ontgrendelen. De biometrie maakt gebruik van de permanente natuur en stabiliteit van een fysieke eigenschap en de uniciteit, robuustheid en het gemak van opname. Er ontstaat een uitdaging wanneer de opname van een biometrische eigenschap van lage kwaliteit is aangezien dit de extractie van precieze (fysieke) kenmerken compliceert. Dit probleem is kritiek wanneer men met pasgeborenen of zuigelingen werkt, doordat hun fysieke kenmerken klein en kwetsbaar zijn, hun medewerking minimaal is en ze snel groeien. Het laatstgenoemde probleem is in het bijzonder relevant wanneer men de identiteit van een zuigeling wil verifiëren naar aanleiding van een opname van een biometrische eigenschap op jongere leeftijd. De zoektocht naar een fysieke eigenschap die bruikbaar is bij zuigelingen word vaak aangeduid als het "*infant biometric problem*". Lage kwaliteit en bruikbaarheid spelen ook een belangrijke rol bij latente vingerafdrukken – afdrukken van het lijnenpatroon die per ongeluk zijn achtergelaten op een plaats delict – en vormen een belangrijk probleem voor deskundigen.

Dit proefschrift onderzoekt het kwaliteits-aspect van volwassen vingerafdrukken en het verband tussen de kwaliteit van een opname en hoe nuttig deze is in een lopend forensisch onderzoek. Bovendien onderzoekt dit proefschrift verschillende aspecten van de "balafdruk" als een biometrisch gegeven voor zuigelingen. De balafdruk is het lijnenpatroon van de bal van de voet.

We tonen aan dat de beeldkwaliteit van een vingerafdruk kan worden gebruikt om automatisch af te leiden of deze bewijskracht heeft in een lopend forensisch onderzoek. Wanneer dit het geval is, is het de moeite waard om de vingerafdruk te vergelijken met een referentie-database, om correspondenderende vingerafdrukken te vinden.

We hebben een longitudinale voetafdruk en balafdruk database verzameld voor 54

pasgeborenen, met metingen binnen 3 dagen na de geboorte, na twee maanden, na zes maanden en na twee jaar. We hebben waargenomen dat de huid van de voet van een pasgeborene droogt en barst. Hierdoor zijn de lijnenpatronen dikwijls niet zichtbaar met het blote oog en is het onmogelijk om hun kenmerken vast te leggen met een vingerafdrukscanner voor volwassenen.

De balafdruk vertoont vergelijkbare eigenschappen als vingerafdrukken, maar de bal van de voet heeft grotere fysieke structuren en een groter aantal kenmerken. Dit bevordert de kans dat een opname van redelijke kwaliteit is en voldoende informatie bevat voor identificatie en verificatie van pasgeborenen en zuigelingen.

Dit proefschrift beschrijft de fysiologische ontdekking dat de balafdruk van een zuigeling isotroop groeit en dat de groei goed kan worden benaderd door een lineaire functie van de leeftijd van het kind. Dit heeft onder andere als gevolg dat de technologie ontwikkeld voor volwassen vingerafdrukken kan worden toegepast bij het vergelijken van balafdrukken. Om dit mogelijk te maken moeten de balafdrukken worden aangepast zodat ze van vergelijkbare grootte zijn als de vingerafdruk van een volwassene. Dit wordt gedaan door gebruik te maken van de afstand tussen lijnen in de balafdruk. De groei van deze afstand spiegelt de groei in lengte van een zuigeling.

We behalen indrukwekkende verificatie scores wanneer we de groei compenseren via isotrope schaling, zelfs voor balafdrukken die 22 maanden uit elkaar zijn opgenomen. De scores zijn nog beter wanneer afdrukken van lage kwaliteit worden verwijderd. Het bleek mogelijk het classificatie algoritme voor vingerafdrukken van volwassenen te gebruiken om de balafdrukken van zuigelingen te sorteren op kwaliteit. Na verwijdering van een derde van de balafdrukken met de laagste kwaliteit daalde de Equal Error Rate van $1-2\%$ naar 0%.

Bovendien behalen de balafdrukken van pasgeborenen resultaten vergelijkbaar met andere biometrische gegevens die beschreven zijn in de literatuur.

Ter conclusie, dit proefschrift toont aan dat de balafdruk een haalbare oplossing is voor het infant biometric problem.

1 Introduction

This thesis focuses on ridge-based and highly distorted biometrics, the different challenges involved in a verification of identity scenario, and how to overcome them. More specifically, we work on ridge-based biometrics in *two* different contexts: (i) newborn and infant biometrics and (ii) quality estimation for forensic investigations. Finally, we merge both applications and demonstrate that techniques from and optimised for the forensic context can be transferred and applied to newborn and infant biometrics and clearly improve results. The thesis is situated in-between the research areas of engineering, applied mathematics and the forensic and medical sciences.

1.1 Newborn and infant biometrics

One ongoing and pressing global challenge is the formal registration of newborns. The *United Nations Children's Emergency Fund* (UNICEF) estimated, as recently as December 2013, that nearly 230 million children under the age of 5 years "have not had their births registered" and therefore "may be denied health care or education" [149]. UNICEF elaborates "registering children at birth is the first step in securing their recognition before the law, safeguarding their rights, and ensuring that any violation of these rights does not go unnoticed" [149].

The International Labour Organization estimates that 1.2 million children are trafficked for the purpose of slave labour or illegal adoption each year [62] which makes human trafficking the third most lucrative illicit trade after drugs and counterfeiting [150]. Another less global but not less important problem is the accidental swapping of newborns in hospitals [132].

A possible solution might be the registration of a newborn's identity together with

a biometric feature. Biometrics are built upon two main principles. Firstly, certain intrinsic characteristics are assumed to be highly distinct to one individual and secondly, their properties remain constant for the individual's entire life [35]. In 2011, the Bill & Melinda Gates Foundation called for researchers to find a suitable infant biometric as part of its Grand Challenges in Global Health program [14].

There are several established and mature adult biometrics such as fingerprint, handvein and iris that have been developed over the last decades for applications ranging from border security to banking. For infants, however, one is faced with two main problems.

Firstly, the infant's growth imposes change to all of its physical features, particularly during the first years of life. This transformation of the biometric feature is unknown and rapid (cf. Table 5.1). Adult biometrics face this problem to a much smaller degree if at all; for example an individual's fingerprint may be eroded due to hard physical labour. For infants, any physical biometric will change in size; in the first year a newborn will double or triple in size. This presents a tremendous challenge because for example, friction ridge skin biometrics, such as fingerprints, are traditionally matched using a spatial representation (constituted of topological and geometric information) of ridge terminations and bifurcations, also called minutiae (see Figure 2.5). Rapid growth negates this traditional approach and requires a different feature representation because only topological information is maintained. A recently developed technique, Biometric Graph Matching, uses a mathematical graph to compare biometric features. It is based on spatial information (minutiae point pattern) enriched by relational information such as minutia connectivity [56].

Secondly, the capture process and its resulting representation of the biometric feature has to be as accurate as possible while being non-invasive and socially acceptable. Newborns and infants are non-cooperative subjects [155]; they might refuse to open their hands in order to have their fingerprints taken, have their eyes closed preventing any iris or retina captures, or move their legs constantly so that collecting prints of the whole foot or parts of it is difficult. Also, one needs to take the newborn's fragility and the small size of its physical features (cf. Figure 1.1b) or overly protective parents into account. Hence, it is fair to assume that a representation of a biometric feature captured is highly distorted and would be classified as being of low quality if compared to captures of the corresponding adult trait, even though it has been captured under controlled conditions.

In order to maximise the chances to find a robust infant biometric which can be adopted by government and non-government organisations it has to fulfil the following properties: it needs to be robust enough to cope with practical capture challenges,

it needs to be socially acceptable and it needs to be easily captured with (and only with) consent. A common suggestion in the literature is footprints. Whole footprints have been captured at birth for many years in many countries. The ridge spacing in a footprint is larger than for fingerprints and feet are usually protected by socks or shoes (in developed countries), making collection without consent nearly impossible [72]. In this thesis the footprint crease pattern and the friction ridge skin on the ball of the foot (the hallucal area under the big toe, referred to as *ballprint* in accordance with [37, 158]) are studied as possible suitable biometrics.

1.2 Quality estimation for forensic investigations

Newborn and infant biometrics are highly distorted. This applies to fingermarks or other marks at crime scenes as well because they are accidentally left behind and not purposely recorded within a specified environment and under controlled conditions. A mark's quality may range from excellent to poor and determines its further use. Experts evaluate the forensic quality of a mark (or its digital representation which is comprised of the quantity of information available in the mark) and the relevance of the mark at the crime scene. There is no benefit attached to collecting all marks found at a crime scene and submitting them for further analysis regardless of their quality. This only leads to additional workload for forensic specialists who analyse the low quality marks and disregard them afterwards and follow the traditional system that does not use the *Likelihood Ratio* (LR) approach (cf. Section 2.2). Hence, it is desirable to limit the collection of marks to those that are at least of sufficient quality to be of value in an ongoing police investigation.

Therefore, police experts are faced every day with the challenge to determine if a mark is of sufficient quality to be captured and to be of any use in the forensic process. This determination process and its challenges are similar to the quality challenges of a newborn and infant biometric.

This thesis evolves around the challenging problem of infant biometrics. It quantifies the difficulties being faced and points out how to combine mature standard technology with new techniques and novel approaches to overcome these difficulties. It also introduces a novel quality algorithm for adult fingerprints or marks, developed to assist police experts. It demonstrates this algorithm is not limited to adult fingermarks but that it can also be used for ridge-based infant biometrics to improve verification results considerably. This is the case for infant ballprints, despite the fact that the algorithm is trained on and optimised for adult fingermarks.

The thesis answers the following research questions.

(a) Ballprint (b) Fingerprint [155] (c) Footprint (d) Palmprint [28]

Figure 1.1: Captures of ridge-based biometrics; the image sizes of the different modalities are not to scale. However, a size comparison between a newborn's and an adult's fingerprint is shown in (b) and has been published by Weingaertner *et al.* [155].

1.3 Research questions

1. What distortions affect ridge-based biometrics, other than deformed skin introduced by the capture process due to the skin placement on the sensor? What fingerprint quality techniques can be used to identify highly distorted captures? Is it possible to measure a capture's quality globally and locally using the *NIST Fingerprint Image Quality* 2.0 (NFIQ2) feature set?

2. How does infant growth affect ridge-based biometrics such as the ballprint?

   (a) Can the change of the infant biometric as the child ages be modelled or approximated by a similarity transformation?

   (b) How does change of the infant biometric affect verification scenarios (cf. Section 2.1) using different matching techniques?

   (c) Is it possible to overcome the change by applying its (inverse) model?

   (d) If the age is completely unknown, can the age still be estimated in order to compensate for the biometric change which the growth induces?

   (e) Are there growth-invariant features, which allow direct comparison between two biometric specimens captured at different ages (inter-age) without any compensation?

3. How do the different Biometric Graph Matching (BGM) modi compare to non-graph matching approaches in a verification scenario using biometric specimens captured at different ages with respect to different compensation techniques used, in terms of the *Equal Error Rate* (EER)?

1.4 Contents overview

Chapter 2 of this thesis reviews the literature regarding biometrics in general, and ridge-based biometrics in particular, explains their background and highlights the key challenges being faced while developing a biometric system. Furthermore, the main concepts used in the discipline of biometric research and the context of this thesis are explained, and the novel databases used are introduced.

Chapter 3 sets the methods background of this thesis. It presents approaches and techniques of the adjacent research areas, such as computer graphics or machine learning, that are used to process biometric data in form of an image or feature points. Important key concepts include classifiers, point pattern representation and transformation, and image pre- and post-processing. Also, we clarify the software frameworks we applied.

Chapter 4 investigates algorithmic solutions to support police officers in the field when they need to determine the quality of fingermarks. This is to streamline the capture process and reduce the high workload faced by forensic police experts due to the increasing demand for their unique abilities. Therefore, we introduce novel algorithms to reduce the workload by enabling non-specialists to make a basic quality assessment of (adult) fingermarks found at a crime scene or after chemical enhancement. Also we investigate the consequences inferred by the capture device used or by falsely disregarding a certain mark.

Chapter 5 reviews the infant biometric literature for recent or promising approaches. Also, it discusses the necessity for all infant biometrics to be evaluated on long-term longitudinal datasets spanning at least several months (not just a few days) and their availability to the research community. A detailed rationale for the choice of infant footprint crease pattern as a potential biometric is presented, a novel algorithm to extract said pattern is introduced and its results are evaluated.

Chapter 6 focuses on the ballprint as a suitable infant biometric. We follow the methodology employed by Gottschlich *et al.* [47] for fingerprints of juveniles. We use our longitudinal infant ballprint dataset spanning 22 months of age and apply the Eigenvalue difference, Moran's I and Geary's C to distinguish between distortion due to growth and the capture process. Finally, we arrive at a similar conclusion for our dataset: isotropic scaling (based on age) compensates for the growth encountered reasonably well in a real world and practice-driven approach. We also present a linear function to calculate the scaling factor. Additionally, we show that commercial fingerprint algorithms can almost always achieve complete separation on high-quality images in a verification scenario. Low-quality images are disregarded using the quality

estimation techniques from and *trained* for adult fingermarks, their original forensic context. We then show that as a newborn biometric the ballprint has comparable performance with reported results for finger, palm and footprint. Finally we show that the ballprint growth from birth to two years closely follows the external growth in length over this period.

Chapter 7 takes the approach of the previous chapter one step further. It is assumed that when there is no knowledge of the infant's age, it has to be estimated from the images captured in order to be able to use the previous methodology efficiently. Hence, we introduce the novel concept of scale-invariant Biometric Graph Matching which allows us to perform matching and verification across different ages without prior age estimation and image compensation or normalisation. We compare the accuracy of both approaches.

Conclusions and further work are discussed in Chapter 8.

2 Biometrics

This chapter sets the context of the thesis. It explains the history and concept of biometrics and their key characteristics in general (Section 2.1) and friction ridge-based biometrics in particular (Section 2.2). It elaborates on approaches to assess them (Section 2.3), current challenges being faced (Section 2.4) and introduces all biometric databases which are employed in this thesis (Section 2.5).

2.1 Background

Biometrics are the sciences evolving around the measurability of an individual's physical, behavioural or chemical features – not exclusively limited to human beings [66]. The resulting metric of a biometric feature maps the actual feature to a certain representation which allows us to compare features belonging to different individuals. The best-known and most commonly used example is a print of a human fingertip, a so called fingerprint [89, 97]. Other examples include the prints of the foot's sole skin [132], the retina's vascular structure [84, 89] or a zebra's coat markings [83].

The underlying motivation behind the use of biometric features (in short: biometrics, often also referred to as modalities [97], traits [64] or identifiers [66]) is to establish or verify an individual's identity. We distinguish between identification and verification scenarios.

Identification An unknown individual's feature or multiple features are compared against a reference set (one-to-many comparison) to establish one's identity by the inference of common source. This procedure is used e.g. for fingermarks found at crime scenes to find the corresponding individual. The marks are com-

pared against a national fingerprint database (after capturing, development and enhancement, and pre-processing). Furthermore we can differentiate between a closed-set and an open-set scenario [97]. The former assumes that the reference set includes the feature(s) of all potential donors in question. Therefore, there must be a genuine match for the query feature(s). This is not the case for an open-set scenario, there may be no genuine match in the reference set.

Verification The biometric feature of an individual with known or claimed identity is compared against the reference entry stored for that particular identity (one-to-one comparison). The identity is considered to be verified if a certain similarity between the features can be established. Examples are unlocking a phone with one's fingerprint [6] or confirming one's identity at the Australian Taxation Office help line by speaking a certain phrase ("In Australia, my voice identifies me.") [78].

There is confusion around the terms identification and individualisation [8, 103], sometimes they are used synonymously; sometimes individualisation refers to linking a mark to one individual object and identification to a class of objects [24]. The process of inference of common source for a mark and a reference specimen is based on the biometric feature's individuality; it is determined if a certain individual is the source of a mark (linked to a criminal activity) [103]. In our understanding, this process has to be performed by forensic experts and cannot be done by automatic comparison algorithms.

Therefore we use the term *individualisation* for the inference of common source of a mark and a reference feature performed by an expert (with or without the support of automatic tools). The term *identification* is used whenever automatic algorithms are employed to compare a mark against one or multiple reference features and the mark's source is inferred based on those comparisons without any expert involvement. The algorithm establishes that the query image is similar to a given reference class only, the class of biometric features it is unable to distinguish between (from one or multiple sources). We also use the more general term identification if there is no knowledge about expert involvement.

The use of a single feature (*unimodal*) can become challenging due to external influences affecting the capture procedure such as sensor noise. One solution is to use multiple features (*multimodal*) as is the case for the Aadhaar project [148]. The goal is to enroll every Indian citizen with fingerprint, iris and face photo. The robustness, increased accuracy and security comes at a cost of slower response times and more complex capture procedures. Therefore, the use of ancillary information such as hair colour or gender may be a suitable solution for some populations. These ancillary

features are referred to as soft biometrics (in contrast to the *hard* biometrics previously mentioned) because those features alone may not be suitable to individualise an individual. They are less distinctive, and some can easily be forged or manipulated. Jain *et al.* define soft biometric traits as "*characteristics that provide some information about the individual, but lack the distinctiveness and permanence to sufficiently differentiate any two individuals*" [64].

2.1.1 Characteristics

Seven key characteristics have been identified to describe biometric modalities. In most cases, there is an application with specific requirements and the biometric modality needs to be chosen accordingly. These characteristics allow evaluation of its suitability for a certain application and are described by Jain *et al.* [66, p. 15].

Universality means that every individual, in general or only those who access a certain application, should possess the modality.

Distinctiveness refers to the ability to adequately discriminate between individuals of an entire population based on the particular modality.

Permanence means how persistent an individual's biometric modality is over time with respect to the application and the matching algorithm used. If a modality does not possess sufficient permanence and thus changes dramatically over time, it is unsuitable for biometric applications.

Measurability refers to how possible it is to capture the biometric feature using a suitable device without causing harm or undue inconvenience via the capture procedure. The raw data captured must also allow for further processing, such as feature extraction.

Performance describes the recognition accuracy in terms of the resources required and the constraints imposed by the application.

Acceptability refers to the acceptance of the biometric trait by the (target) population and thus their willingness to use the modality.

Circumvention describes how easily an individual's physical or behavioural modality can be imitated by using artefacts or impersonation, respectively.

For example, a low-security application to record work attendance would favour acceptability and measurability over permanence and circumvention. This is the case

because it is more important that the employees are willing to use the system and that no additional burden is imposed on the user than its actual performance. For every application, there are always trade-offs involved as there is no perfect biometric. The biometric used and its characteristics need to be chosen with the application and its requirements in mind. Table 2.1 presents an evaluation of popular biometrics according to their key characteristics; three basic classes (high, medium and low) have been used.

Our assessment of footprints and ballprints are based on the work of Maltoni *et al.* [97, p. 11] for fingerprints. The reasoning behind the differences (highlighted in Table 2.1 by *) is as follows. We consider the biometric footprint to possess only medium permanence because crease pattern studies have shown that major flexion creases are stable but that new ones form and others disappear (within the time span of one to twelve years) [99] (see also Section 5.4 in Chapter 5); this applies especially to infants. This results in decreased performance. Footprints have been taken in hospitals shortly after birth as an identifier for the birth certificate or as a keepsake for well over half a century. Hence taking footprints is widely accepted. The foot's sole is usually protected by socks or shoes making it difficult to obtain a reference print without consent in order to create an imitation. This is not the case for fingerprint. In contrast to the footprint crease pattern, we are going to show the ballprint's permanence for infants in Chapter 6. Hence we evaluate its permanence as *high*. Furthermore, the ballprint is part of the sole (the hallucal area under the big toe) and benefits from high acceptance and footwear protection as footprint does.

In this context, we need to differentiate between uniqueness and distinctiveness and to elaborate on the reasons that the first concept is largely irrelevant but the latter is important in a forensic context. Cole highlights that the philosopher Wittgenstein (as pointed out by Kwan [82] and Meuwly [102]) questions the principle of uniqueness because all objects can be "*the same*" or "*different*" without a specific definitions what these terms mean. It would only depend on the frame of reference. He carries on that "all objects in world are *the same* and all objects in the world are *different*" [24]. In a forensic context this means that all specimens are unique by definition. It is more important to supply tools that allow to assess biometric features on a fine level to ensure that as much (relevant) information is captured as possible. This increases the distinctiveness of the biometric feature representation. Consequently, an informed decision about the feature's source can be derived by comparison to the features of other candidates based on its distinctiveness (see Section 2.2).

Biometric	*Universality*	*Distinctiveness*	*Permanence*	*Measurability*	*Performance*	*Acceptability*	*Circumvention*
Face	+	−	±	+	−	+	+
Hand geometry	±	±	±	+	±	±	±
Hand/finger vein	±	±	±	±	±	±	−
Iris	+	+	+	±	+	−	−
Signature	−	−	−	+	−	+	+
Voice	±	−	−	±	−	+	+
Ballprint	±	+	+	±	+	+*	−*
Fingerprint	±	+	+	±	+	±	±
Footprint	±	+	±*	±	±*	+*	−*

Table 2.1: Subjective evaluation of key characteristics for different biometrics adapted from Maltoni *et al.* [97, p. 11]. We have added ballprint and footprint based on our own experience and their fingerprint assessment. The evaluation categories used are high (+), medium (±) and low (−). The biometrics relevant to this thesis are highlighted in bold. The reasoning why some categories for ballprints and footprints differ from the fingerprint ones (indicated by *) are discussed in Section 2.1.1.

2.1.2 Biometric applications

A biometric system consists of different components or stages which are all affected by distortions in one way or the other; the following four components are well-known [66, pp. 3].

First, a (digital) representation of the physical feature is obtained by the system's sensor module (*capture stage*). This process is strongly affected by the individual's resistance (actively, passively or unknowingly) in terms of their will to cooperate or the feature's source. Criminals might forcefully resist having their fingerprints taken in order to avoid identification or accidentally leaving a mark behind when gripping a window sill as they enter a property. Infants might be scared by the capture device and newborns might be uncomfortable with the sensation of the device touching their skin. They just may be unsettled or playful. Also, individuals might misinterpret directions and e.g. apply too much pressure to the sensor when capturing a fingerprint. All these different causes lead to captures with suboptimal quality. Furthermore, the sensor

(the readout process or its post processing) itself is affected by its age or minor defects (e.g. dead pixels or broken electrical contacts) or by environmental influences such as electro magnetic waves emitted by mobile phones or microwave ovens.

Second, the representation of the biometric feature (the capture) is assessed for its quality, some pre-processing is performed and features are extracted (*feature extraction stage*). This feature representation is also referred to as a *template*. The quality assessment evaluates the capture's suitability for reliable feature extraction. If the quality assessment fails or arrives at the conclusion that the capture does not fulfil the application's quality guidelines, the user might be required to have the biometric re-captured (e.g. for an access system) or it is not processed any further (e.g. for a fingermark taken at a crime scene). The cause, especially for fingermarks, is that low-quality captures or captures of low quality marks lead to unreliable feature extraction, which potentially compromises the underlying main principle, the biometric's distinctiveness. The biometric of two individuals might differ but their low-quality captures might lead to indistinguishable templates. Therefore, it is a common procedure to disregard low quality captures.

Third, the biometric template is compared with either one or multiple references and a decision is made regarding the individual's identity (*matching and decision-making stage*). This stage differs according to the scenario or purpose of its application. Nevertheless, two templates are compared for their similarity and a matching score is returned. In a verification scenario, the query template is compared to the registered template of the nominated identity. The identity is considered confirmed if both templates are very similar and exhibit an absence of significant differences (high matching score). In an identification scenario, the template is compared against all other reference templates and the most similar reference template is considered to be the individual's identity. The decision rules might require, for example, a minimum distance to the second-best match to ensure a decision's confidence [92]. Alternatively, in some cases such as fingerprint comparisons, a certain number n of identities which achieved the greatest similarity to the query template are further processed e.g. in greater detail or manually (n-rank) [63].

Fourth, the database is the last component required for a biometric system (*database component*). The third stage requires the knowledge of previously enrolled or stored identities linked to their biometric template. The enrollment process stores one or multiple templates of the individual in a database and links them to the identity and potentially to some meta information (name, date, or soft biometrics). The database contains the reference templates. Some applications update the templates stored after a successful identification or verification to keep the template up-to-date and

counteract any changes of the physical biometric, or store multiple references. The initial enrollment can be performed either supervised by trustworthy entities to ensure the system's integrity or unsupervised if it is of rather low importance, such as the stored fingerprint to unlock one's smartphone.

In this thesis, we focus on friction ridge-based biometrics such as ballprint, fingerprint and footprint. They all are captured in areas covered by friction ridge skin (finger, foot) and hence their templates are based on friction ridge skin properties.

2.2 Friction ridge biometrics

Friction ridge skin (also sometimes referred to as papillary ridge patterns or dermatoglyphics [26, 103]) is defined as "the skin of the palms of the hands and fingers as well as the soles of the feet and toes" [103]. The skin in these areas forms small ridge-like epidermal extensions forming a three-dimensional structure similar to ridges and valleys. The skin's anatomy reflects its evolutionary purpose. Creases allow the skin to flex, and friction ridges (and the sweat pores within) maximise the friction when gripping or grasping [54, chap. 2]. A print of the friction skin such as a fingerprint refers to the impression left behind by the ridges when they come in contact with a surface such as the sensor. Therefore, the actual structure is not captured but its two-dimensional representation is. Hence the resulting template is subject to the capture procedure and its correct execution; many factors, such as too much or too little pressure, or dry or wet skin, affect the capture quality and therefore the reliability of the resulting template.

The literature hypothesises different theories as to how and when the friction ridge skin develops [26, 54, 69, 89]. Nevertheless all these theories suppose that the ridges form at a very early stage, ranging from a gestational age of 6 – 7 weeks [53] to around 10.5 – 16 weeks [54, chap. 3] and are finalised during the period between the 17th week [69] and the 24th week [9, 54]. These theories agree that the process of creating the distinct flow of ridge lines is "due to developmental noise" [54, chap. 3] and "stresses encountered during growth" [53].

Features

Impressions of the friction ridge skin can be assessed for biometric features on three different levels of coarseness. There are (i) the global level (e.g. ridge flow or crease), (ii) the local level (e.g. minutiae such as ridge endings and bifurcations) and (iii) the detailed level (e.g. pores). Kumbnani lists several studies which investigated

dermatoglyphics such as the sole pattern [81]. Maltoni *et al.* discuss these for fingerprint [97, pp. 39].

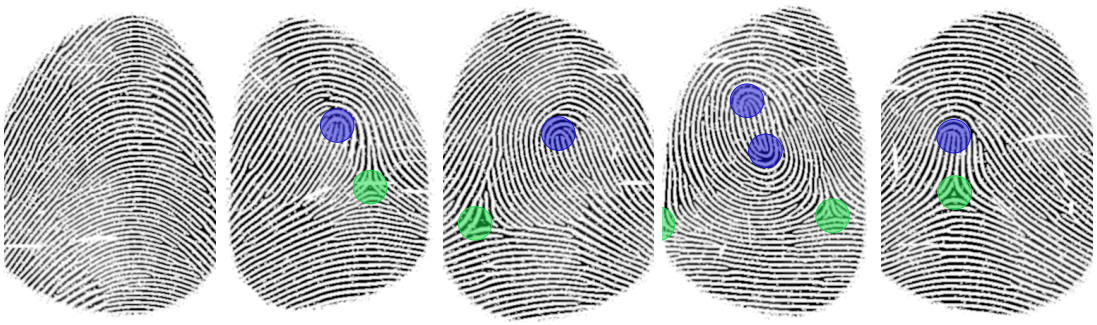
We discuss all three feature levels separately.

Level 1

There are two global features which can be assessed: (i) the overall ridge flow and (ii) gross discontinuities in the ridge pattern. The friction ridges display a regular pattern of approximately parallel lines which in most cases flow around or along certain landmarks, so called singular points. There are two types: cores and deltas (see Figure 2.1) [89].

Henry defined the core as "the north most point of the innermost ridge line" [52]. In practise, the core is often associated with the point of maximum ridge line curvature because not all prints display cores (e.g. arches or unaligned prints). The occurrence of these singular points, and, if present, their spatial position relative to each other, are used to classify fingerprints and ballprints into a relatively small number of types. They can also be used to provide a common frame of reference for registration when attempting to match prints.

Maltoni et al. note that the region around singular points, the singular region, can be classified into three basic typologies: (i) core ($\cap$), (ii) delta (Δ) and (iii) whorl (O) [97, 98]. The whorl may be represented by two cores.



(a) Arch (b) Left-slant loop (c) Right-slant loop (d) Whorl (e) Tented arch

Figure 2.1: Synthetic fingerprints showing the five different Level 1 features; cores and deltas are highlighted in blue and green, respectively. The prints have been generated with the latest version of SFinGe [18].

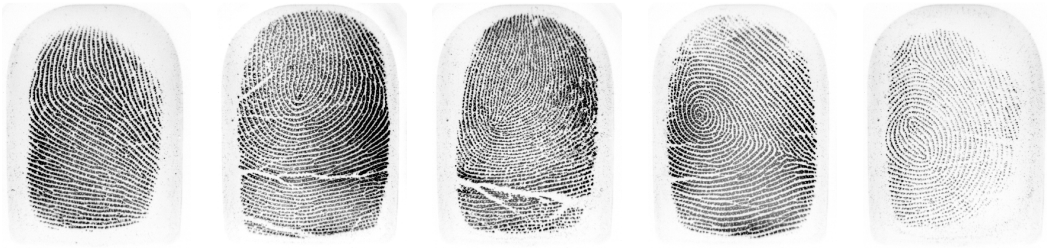
The FBI's *Automated Fingerprint Identification System* (AFIS) distinguishes between the following four patterns for fingerprint: arch (i), left-slant loop (ii), right-slant loop

(iii) and whorl (iv) [54, chap. 5]; sometimes arches are sub-classified into simple or tented [103]. The Figures 2.1 and 2.2 show the different patterns for fingerprint.

(a) Arch (b) Left-slant loop (c) Right-slant loop (d) Whorl

Figure 2.2: Fingerprints exhibiting the four different Level 1 features (a) arch, (b) left-slant loop, (c) right-slant loop and (d) whorl. The prints are taken from the Victoria Police reference database (see Section 2.5.3).

Most of the literature on footprints and ballprints distinguishes between three patterns formed by the ridge flow in the hallucal area. They are (i) arch (also sometimes referred to as open field; compare [37] and [26, 107]), (ii) loop, (iii) whorl; the usage of arch and open field and the degree of sub-classification depends on the study [26, 42, 107, 157, 158]. The FBI's footprint classification scheme is based on the main three types only [37].



(a) Arch (b) Loop (c) Loop (d) Whorl (e) Whorl

Figure 2.3: Infant ballprints exhibiting the three main Level 1 features (a) arch or open field, (b, c) loop and (d, e) whorl. All images have been captured with a fingerprint scanner when the study participants were about 2 years of age. The images are part of the Happy Feet database (cf. 2.5.1): (a) right foot of infant 048, (b) left foot of infant 017, (c) left foot of infant 045, (d) left foot of infant 009 and (e) right foot of infant 036.

The most frequently found feature in a ballprint is the loop [26, 42, 107] but it can be subject to the population investigated [157]. Montgomery observed correlation to some extent between the feature type of a person's right and the left ballprint [107]. Okajima [119], and Wertheim and Maceo [156] offer an explanation for these findings based on friction ridge skin morphogenesis, its development process.

A template's pre-classification into one of these classes can speed-up the computation time in an identification scenario drastically because it eliminates the need for further comparison if the coarse level features of two templates do not match.

The discontinuities of the ridge pattern mentioned above can be caused by either flexion creases (applies less or hardly at all to fingerprint), scars [54, chap. 2] due to injuries, or deliberate manipulation of the biometric [164]. All three contribute towards the distinctiveness of ridge-based biometrics; the major flexion crease pattern in adults is considered to be exhibit a high degree of distinctiveness (cf. Section 2.1.1) in its own right [99].

Here we would like to emphasise the value of the spacing between the ridges, the *inter-ridge spacing* (IRS). It usually varies across the captured area, especially around cores or deltas, and depends greatly on gender, ethnicity or age [1, 49, 116, 136]. Nevertheless, it is a robust measure that allows the comparison of individuals from the same population [110] and to derive an adult's gender [1]. An example is shown in Figure 2.4.

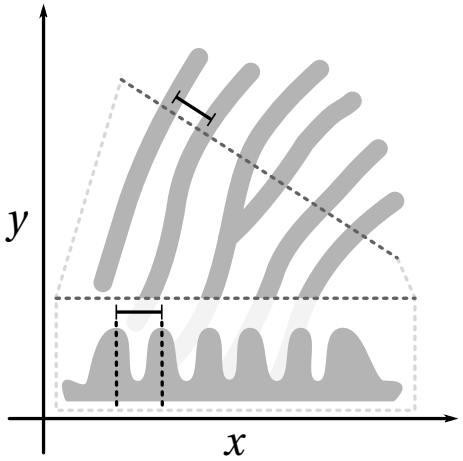


Figure 2.4: Example on how an individual inter-ridge spacing is obtained as top view and cross section. The IRS is the distance between the centre points of two ridge lines. Usually it is measured at multiple points and averaged.

Level 2

Characteristic points of individual ridge lines are called minutiae; sometimes they are also referred to as Galton details for historical reasons. They can be found across the whole fingerprint and ballprint. Their number varies greatly between the fingertip and the ball and even between studies.

Okajima analysed rolled fingerprints of 77 males and 82 females. He counted an upper limit of 112 and 91 minutiae for males and females, respectively. The highest averaged minutia count was found for both sexes and both hands on the thumb and the lowest count in general on the little finger [117]. He only counted minutiae in the centre area of the print. The average count per finger is presented in Table 2.2.

Gender	Side	Finger				
		Thumb	Index finger	Middle finger	Ring finger	Little finger
Male	Right	75.8	56.3	57.5	58.1	48.1
	Left	70.7	53.4	55.3	58.0	46.7
Female	Right	64.1	51.7	52.3	51.4	39.2
	Left	66.9	49.3	51.6	53.7	41.1

Table 2.2: Average minutiae count for 77 Japanese males and 82 Japanese females [117].

Maltoni *et al.* assume the hypothetical number of 36 minutiae on average per fingerprint when estimating its individuality for different models reported throughout the literature [97, p. 351]. Also, Thu *et al.* have shown (for fingerprints only) that minutiae are not uniformly distributed across the fingertip but have the tendency to cluster with respect to their location and orientation [167].

Uhl and Wild found 300 – 400 minutiae in high-quality images of adult ballprints acquired with a flatbed scanner [144]. In contrast, Okajima usually counted 100 – 130 minutiae [118]. He used an ink-based method to obtain the prints from Japanese teenagers and highlighted the hallucal area by projecting a mask on top.

Regardless of the capture techniques used, the adult ballprint competes with or even exceeds the average number of minutiae encountered on a thumb and exceeds that on a (non-rolled) fingerprint.

For matching purposes, the spatial arrangement of the minutiae is most common feature set [97, 101, 153], sometimes enriched by the minutia's type and orientation. The literature lists three basic minutia types: (i) ridge ending, (ii) bifurcation and (iii) dot. Other types are combinations of these basic types such as island (two ridge endings) and enclosure (two bifurcations). Examples of all of these types are shown in Figure 2.5.

Most fingerprint matching algorithms limit the type to ridge ending and bifurcation [115] and ignore dots to increase the likelihood of extracting reliable minutiae. This is because a dot could be caused by suboptimal capture conditions more easily than a ridge ending or a bifurcation. Too little pressure or dry skin can lead to dashed

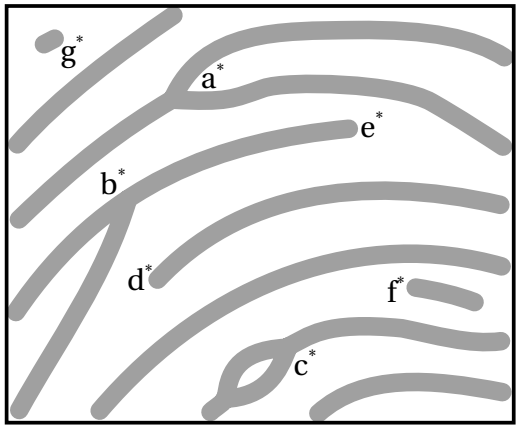
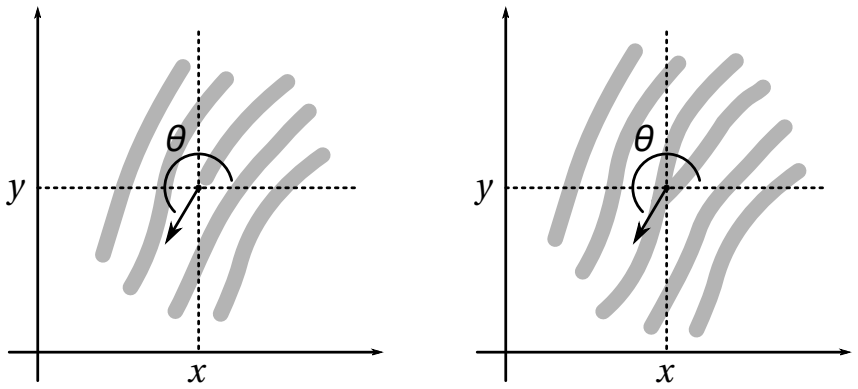


Figure 2.5: Example of different minutia types such as (a*, b*) bifurcation, (c*) enclosure or lake, (d*, e*) ridge endings, (f*) island and (g*) dot.



(a) Ridge ending orientation (b) Bifurcation orientation

Figure 2.6: Determination of the minutia orientation for (a) a ridge ending and (b) a bifurcation. The example is adapted from Maltoni *et al.* [97].

ridges or disconnect one leg of a bifurcation, resulting in additional ridge endings or a mistaken minutia type [13]. The problem of a changed minutia type is referred to as connective ambiguity [97]. A minutia's orientation, θ, depends on its type (shown in Figure 2.6a and 2.6b); it is always calculated as the counter clock-wise angle from the x-axis. The orientation of a ridge ending is the intersecting angle between its tangent and the x-axis. The tangent is the direction into which the terminated ridge line would continue into. Similarly for a bifurcation, it is the direction of the valley ending enclosed by the two ridge lines that become one.

Level 3

The fine ridge line details such as pores, minutia shape, and ridge edge form the third level. Their presence or absence and spatial arrangement can be valuable provided they can be extracted reliably [5], which is rarely the case. However, often they cannot be captured due to the method used (e.g. chemical enhancement of a fingermark found at a crime scene) or insufficient image quality; this issue applies to fingerprint and ballprint similarly and in fact to all of the biometric images analysed in this thesis. We are not aware of any commercial fingerprint matcher that does not disregard third level information automatically as it cannot necessarily be captured reliably.

Forensic applications

The idea of using friction ridge impressions for forensic purposes dates back to the late 19th century when William Herschel and Henry Faulds proposed "the use of fingerprints and fingerprint databases for the identification of serial offenders and the use of fingermarks to establish a link between a scene or an object and an individual" [103]. Again, this application is built upon the principle of distinctiveness; the print belongs to one individual. This is not limited to the impression of the fingertip but can be applied to all other friction ridge impressions as well. Often all three levels of detail are used by experts to arrive at a conclusion but there is some controversy regarding the use of fine detail because chemical enhancement can destroy or modify this information [103].

In this context, we need to understand the difference between a mark and a print and clarify the term *latent* which is often exchanged synonymously with mark. Meuwly defines a *fingermark* as "recovered traces left by unprotected fingers in uncontrolled conditions", whereas he refers to a *fingerprint* as a "standard rolled inked impression captured from the finger papillary ridges" [103]. Latent means non-visible [1] and thus requires development or further processing to make the impression visible. This is only needed in uncontrolled conditions and hence only applicable to marks. In this thesis, we adapt the terms *print* and *mark* for all impressions captured in controlled and uncontrolled conditions, respectively. Hence a fingerprint does not necessarily refer to a rolled ink impression but may also be captured with a scanner or a camera as long as it happens under controlled conditions. This way we also avoid the misleading term "latent fingerprints".

There are different requirements in courts around the world for the process of individualisation (to establish if a certain print or mark belongs to a specific individual).

[1] latent, third-person plural present of the Latin verb *latere* - to be hidden [88]

Some countries require a fixed minimum number of corresponding minutiae (e.g. 12 minutiae in France), some use a more holistic approach and use an implicit threshold (a non-numerical standard, e.g. Australia since 1999) and some rely on a probabilistic approach where the likelihood that a mark and a print are from the same individual is computed based on empirical data (e.g. the Netherlands). The Ne'urim declaration from 1995 states that the requirement of a fixed number of corresponding minutiae lacks scientific foundation. Meuwly lists more than 30 different countries, their official approach and the explicit thresholds used if applicable. He also elaborates on the probabilistic approach, the countries that implemented it and its advantages [103].

The process of making *inferences of common source* [61] (also known as *inference of identity of source* [82]) between a trace and a reference specimen with known origin is often used during police case work [24]. Both approaches that require a fixed minimum number of minutiae or an implicit threshold require the forensic expert to arrive at a decision that she is not supposed to make but leave to a judge or jury. During the comparison process, the expert decides at a certain point that in the absence of difference the two specimen originate from a common source, that there are too many similarities but also differences present to arrive at a conclusion (inconclusive), or that there are too many differences (exclusion). Even before the expert arrives at the official decision, she already has made up her mind based on the rationale that highly distinct patterns cannot be similar if they are not from a common source. Stoney refers to it as a "leap of faith" [135]. Sometimes experts start to rationalise differences away or look for common similarities. This behaviour is encouraged by a threshold requiring a fixed minimum number of minutiae. The *Analysis, Comparison, Evaluation, and Verification* (ACE-V) protocol is supposed to prevent this behaviour.

Meuwly points out that "The challenge for dactyloscopy is about the ability to quantify the information available for the individualization process in a partial distorted fingermark, and not to prove the individuality of the friction ridge skin" [103]. This logically leads to the probabilistic approach because it quantifies the information present in the trace and gives a likelihood ratio stating the probability for the trace and the reference specimen to have a common source. This removes the "leap of faith", adds transparency regarding the certainty or uncertainty present, most importantly it leaves the final decision if there is a common source to the judge or the jury.

Historic overview

Ridge-based analysis has a long (and well documented) history. This especially applies to fingerprints, although their application has expanded over time. In ancient China, fingerprints were used as a trademark symbol to guarantee or highlight that a specific

potter had produced a certain piece of earthware, but the emperor Ts-In-She apparently left the impression of his fingertip on some clay to seal documents and thus prove their authenticity [8, chap. 2]. Furthermore, ancient Chinese books have been found, which strongly suggest that hand impressions have been used (even before Christ) during crime scene investigations [54, chap. 1].

In the western world, in 1788 Mayer was the first to describe the friction ridge pattern as distinct to one individual [8, chap. 2]. Purkinje introduced a basic classification scheme consisting of nine classes in 1823, which is sometimes considered to be the foundation of the famous Henry classification system [54, chap. 1] (Level 1). At the end of the nineteenth century, Herschel and Faulds contributed greatly towards today's knowledge about friction ridge skin, its permanence and distinctiveness [8, chap. 2]. Another important contribution was made in 1892 by Galton [44] who is the first author to describe minutiae and to point out that there are different minutia types (Level 2). Finally, Wilder [158] was the first one in 1918 to mention the pores embedded in the ridge lines and to consider them as both distinct and persistent (Level 3). Recently, Anthonioz and Champod emphasise that pores suffer from a "lack of reproducibility, as well as the effect of distortion" and should only "considered in reference to a minutia" because then "pores have a good discrimination capability" [5].

One of the first murder cases in modern history to be solved by fingermarks took place in 1892 in Buenos Aires, Argentina. Two bloody thumbmarks lead to the confession of a mother to have killed her own two children and the dropping of all false accusations, after matching her fingerprints successfully against the marks found at the crime scene [54, chap. 1].

Since then, forensic expertise has increased significantly and fingermarks and prints are now used commonly in case work to individualise or exclude suspects. In order to compare fingermarks found at a crime scene, reference databases are needed. One of the largest ones is the FBI's *Integrated AFIS* (IAFIS) which contains "more than 70 million subjects in the criminal master file" and "more than 34 million civil prints" [36]. Despite the common practise of using fingerprints in court for over a century (cf. Section 2.2), current practice and knowledge about the biometric's distinctiveness and permanence are being challenged (see Section 2.4).

Footprints and ballprints do not possess an equally well documented history as fingerprints [144]. Footprints have been routinely collected at hospitals for identification purposes or as a memento for parents for decades. Their general ridge line flow has been investigated by Wilders from 1904 onwards [157, 158] and Montgomery studied the Level 1 features extensively during the 1920s [105–108].

Realising the individualisation value of footmarks found at crime scenes, the FBI published an article about the classification of footprints for the first time in 1948 and released an updated version in 1971 [37].

In 1958, a New York police officer published a case study stating that footprints of newborns can be used for individualisation purposes and that abduction cases have been solved due to this biometric [16]. The literature is split regarding footprinting newborns and infants; some see a real added benefit from it, others disregard it as too difficult and too expensive [140]. This topic is discussed in more detail in Chapter 5.

Since the early stages, some researchers have investigated the hallucal area under the big toe, the medial ball and its suitability as a biometric [143, 144] but it remains a topic sparsely studied. Some approaches are presented in Chapter 5 as background information, others are discussed more thoroughly as a main part of the thesis in Chapters 6 and 7.

2.3 Assessing biometrics

Biometrics can be either assessed using the seven criteria defined in Section 2.1.1 or in the context of a certain application or system. Here, we focus on the latter. Regardless of the application, a biometric system will at the very least compare a query and a reference template and express their similarity as a single score. In a real world scenario, every specimen is unique. In the case of ridge-based biometrics, this is due to e.g. the movement to produce the the mark or the print, the force applied and the secretion of the skin. These factors may lead to a different area captured, physical distortions or variation in clarity. Nevertheless, the intra-variability (same subject, a genuine comparison) is expected to be smaller than the inter-variability (different subjects, an imposter comparison) of the templates. Therefore, genuine comparisons result in comparison scores greater than the scores of imposter comparisons. Therefore, it is highly unlikely that two templates of the same individual are identical. In the event that a perfect match is encountered, the template's origin should be questioned and it should be considered that the system is attacked [66, p. 7].

A biometric system's performance can be assessed at any stage (cf. Section 2.1.2). The *Failure to Acquire* (FTA) measures the rate the input device fails to capture the biometric presented to it (capture stage); reasons include but are not limited to the sensor's age, technical malfunctions or very low quality ridge pattern which the sensor does not recognise as a biometric signal. The *Failure to Enroll* (FTE) includes FTA but even after the biometric has been captured successfully, it still might be rejected by the system for quality reasons or the lack of features which could be extracted from it.

This thesis focuses on errors that occur later: at the comparison-and-decision-making stage in a verification scenario. When two templates are compared, they can be from either the same individual or from two different individuals. The first case is a *genuine* comparison, whereas the second one is referred to as an *imposter* comparison. The system's decision can be a matching score ς but is often a binary decision (derived from the matching score and a threshold η). Therefore, there are four different outcomes possible.

True positive The system returns a positive match for a genuine comparison ($\varsigma > \eta$, genuine query template).

True negative An imposter comparison is correctly rejected as not matching ($\varsigma \leq \eta$, imposter query template).

False positive The system falsely mistakes an imposter comparison as being genuine ($\varsigma > \eta$, imposter query template).

False negative A genuine comparison is falsely concluded to be an imposter one ($\varsigma \leq \eta$, genuine query template).

A system's performance is evaluated on a test set consisting of genuine and imposter comparisons. The number of genuine matching scores greater than the threshold η divided by the total number of genuine comparisons is known as *True Positive Rate* (TPR) or *Genuine Acceptance Rate* (GAR); the fraction of genuine scores smaller than or equal to η is often called *False Negative Rate* (FNR) or *False Non-match Rate* (FNMR). Hence, it is possible to derive them from each other as their addition equals 1:

$$\mathrm{TPR} + \mathrm{FNR} = 1. \tag{2.1}$$

Similarly, the number of imposter comparisons exceeding the threshold divided by the total number is referred to as *False Positive Rate* (FPR) or *False Match Rate* (FMR). The fraction of imposter comparisons which are smaller than or equal to η is called the *True Negative Rate* (TNR). An illustration appears in Figure 2.8a.

Both error rates, the FPR and FNR, summarise together a system's performance but are subject to the threshold used. Adjusting η changes the FPR and the FNR at the same time but if one decreases, the other one increases. Therefore, another measure of performance is the minimum value of the sum of the errors.

Sometimes FNMR and *False Reject Rate* (FRR), and FMR and *False Accept Rate* (FAR) are used synonymously. However, there are differences: FRR and FAR take the user's claim of identity into account. This means in an access control scenario (positive claim of identity) that a false match would allow an imposter to gain access (false accept) and that a false non-match would prevent a genuine from gaining access (false reject). In the case of a negative claim of identity (e.g. a black list system preventing certain people from access) a false match would reject a genuine access request and a false non-match would allow access to an imposter [97]. Also FAR and FRR take the entire system into account (including e.g. failure to enroll) in contrast to FMR and FNMR which focus on the error during the decision stage.

In this context, one has to note that the different error types, false reject (false negative) or a false accept (false positive), are not distributed evenly across all users. Some are more prone to a certain error type than others. In 1998, Doddington *et al.* investigated if there are certain user classes in speaker recognition [30]. They encountered four different categories: sheep, goats, lambs, and wolves. Hence, the set of these classes is often referred to as "Doddington's zoo" [122].

Sheep These users possess very distinct biometric features with low intra-class variation. Hence both error rates tend to be low; this is the default case.

Goats These individuals are particularly difficult to recognise. Their features tend to exhibit large intra-class variation, hence these users are likely to encounter false rejects.

Lambs These users or their biometric features can easily be imitated. Therefore, this user group is responsible for disproportionate many false accepts.

Wolves These users imitate (un)knowingly other individuals successfully and tend to be falsely accepted by a biometric system as another user. Their biometric features exhibit great similarity to the ones of enrolled users.

These user classes apply to all types of biometrics and are not limited to the task of speaker recognition [66, p. 8].

Dunstone and Yager extended "Doddington's zoo" by taking genuine and imposter scores into account. They added chameleons, phantoms, doves and worms [34].

Chameleons These users always receive high scores when compared to themselves or to others.

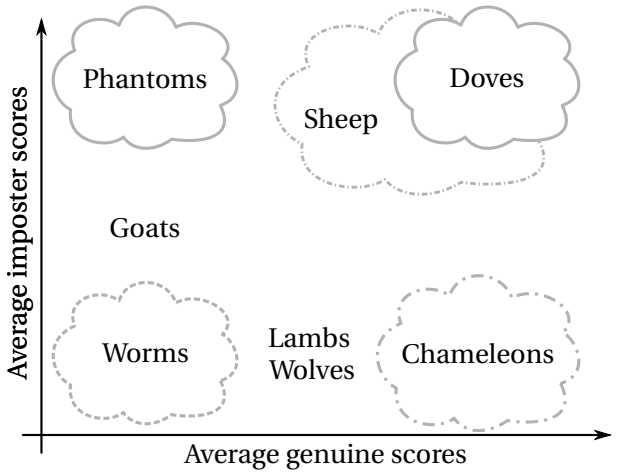


Figure 2.7: Zooplot showing the different animals with their average genuine and imposter scores. The plot is adapted from Alexander *et al.* [2].

Phantoms These individuals always receive low scores and hence don't match against themselves or other users.

Doves These users receive high genuine scores and low imposter scores.

Worms These individuals experience low scores when compared against themselves but high scores when compared against others.

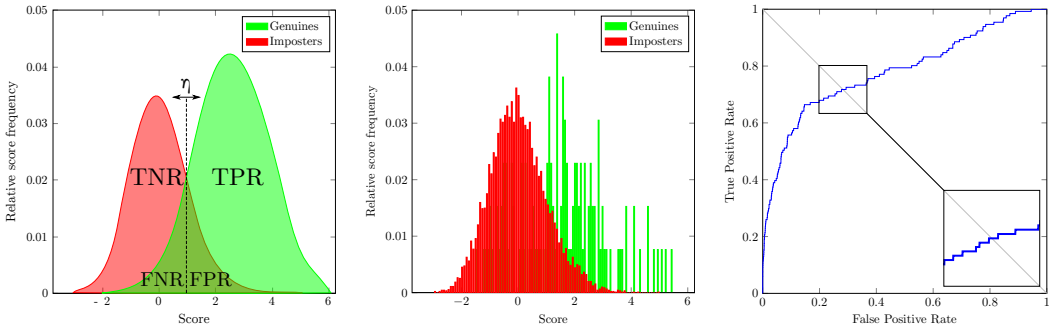
Recently, Alexander *et al.* [2] added the attributes "tall or short" and "fat or thin" to describe the variability of genuine and imposter scores, respectively. A tall thin dove is a speaker with low intra (genuine) but high inter (imposter) variability.

In order to visualise a system's performance, the *Receiver Operating Characteristic* (ROC) curves can be used. The ROC is a plot of the FNR against the FPR for various thresholds. Commonly the TPR (1 – FNR, cf. Equation 2.1) is used instead of the FNR. Hence, the closer the curve runs to the top-left corner of the graph, the smaller the errors and better the performance. An example is given in Figure 2.8c.

ROCs can be difficult to compare and a single performance value might be desired. In this thesis, we use either the Equal Error Rate (EER) or the *Area Under the ROC curve* (AUROC). Both summarise a given ROC with a single value. The EER is calculated at the threshold where FNR and FPR are equal:

$$\text{EER} \coloneqq \text{FNR} = 1 - \text{TPR} = \text{FPR}. \tag{2.2}$$

Visually, it can be obtained by drawing a line from the top-left (TPR = 1) to the bottom-right (FPR = 1) corner. The EER is the FPR value where the line intersects the curve



(a) Theoretical genuine and imposter score distributions

(b) Actual genuine and imposter scores

(c) Receiver Operating Characteristic (ROC)

Figure 2.8: The relationship between the theoretical genuine and imposter score distributions, FPR and FNR and the threshold η are shown in (a). The distribution of the data (b) and the corresponding Receiver Operating Characteristic (c); the Equal Error Rate and Area Under the ROC curve for this example are 0.28 and 0.78, respectively.

(cf. Figure 2.8c). Generally speaking, the lower the EER, the better the system's performance. The AUROC is calculated by computing the area under a ROC curve given; the closer it is to 1, the smaller the errors and hence better the performance. In case of *complete separation* between the genuine and imposter scores, the FNR, FPR and EER are all 0 and the AUROC is 1.

2.4 Recent challenges

Ridge-based biometrics, especially well established traits such as fingerprint, are facing multiple challenges. One of the biggest concerns is the biometric's distinctiveness and how it can be quantified in order to arrive at the best decision possible. It applies particularly to low quality specimens in the context of the constantly increasing world population and the Indian Aadhaar project. The Aadhaar project aims to enroll all Indian citizens. It is the largest biometric system of its kind with potentially *one billion* enrolled users [148]. The main problem is that it is impossible to show that all human individuals possess distinct fingerprints because this would require one to collect all prints and compare them all against each other. Even if it were accomplished for a given moment in time, the population changes and increases constantly; people decease, others are born. Therefore, the base assumption, a biometric's distinctiveness, can only be supported by statistical models estimating the likelihood of randomly matching two fingerprints from two different individuals, or by large sample studies [27]. Maltoni *et al.* [97, pp. 341] and Dass [27] summarise the existing fingerprint literature and confirm the underlying assumption of distinctiveness to some degree.

There are two other challenges. First, it is known that a biometric feature changes over time. This can be due to growth [47], ageing (e.g. reduced elasticity of the skin or wear and tear) [145, 166], injuries or manipulation [164]. Therefore, it is not sufficient to prove only a biometric's distinctiveness but it is also required to consider its (non-)persistence over time. This can be addressed by longitudinal studies. Yoon and Jain [166] and Dass [27] point out that image quality plays an important role. If one compares two specimens of the same individual captured at different ages and one specimen suffers from low image quality it reduces the matching score drastically so that even the "temporal stability of fingerprint recognition accuracy" [166] may be questioned.

Second, the capture quality of a physical biometric feature (image quality for fingerprints) plays a major role. Even if the biometric retains its degree of distinctiveness in the same order of magnitude over time and remains persistent, the capture process might reduce the quantity of information present in a mark. Consequently, the feature extraction produces a template exhibiting a small degree of distinctiveness despite originating from different and highly distinctive traits, due to low quality captures. If the biometric itself is distinct and persistent, its digital template might not be.

In 2004, Mayfield was falsely accused of being involved in the Madrid bombings. His fingerprint had been retrieved (amongst 19 other individuals' prints) from the FBI's fingerprint database as being similar to a low quality fingermark lifted from a bag found at the crime scene [151]. Nevertheless, four experts (three FBI, one court-appointed) testified in court that the mark corresponded indeed to Mayfield's print [103]. Later, an official investigation concluded that this misidentification was due to "human error" and that there are always degrees of similarity when a mark of unknown identity is compared against a large database [151, 152].

Usually, there are guidelines and protocols in place to avoid cases like Mayfield's. In times of high-volume crime and demanding workload for fingerprint experts, due to high demand with plenty of manual work required, sometimes shortcuts are taken and guidelines are unconsciously violated. Nevertheless, it demonstrates clearly the significance of a proper understanding of the process on the inference of identity of source. It also underlines the importance of avoiding to degrade the quantity of information present in a trace compared to its digital representation due to the capture process and image quality.

In summary, the most recent two main challenges in the field of biometrics are understanding a biometric's persistence or how it changes over time (especially during rapid growth as encountered with infants) and developing sensible (automatic) quality algorithms.

2.5 Databases

This thesis uses three databases: a novel longitudinal ballprint and footprint database from birth to 2 years old (Happy Feet, Chapters 5, 6 and 7), an adult ballprint database (Adult ballprint, Chapter 6), and a novel Victoria Police pseudo fingermark database and the corresponding reference prints (Victoria Police, Chapters 4, 6 and 7). Each database and its characteristics is introduced in the following sections. The author assisted and contributed towards the creation of the Happy Feet and Victoria Police databases as part of this thesis.

2.5.1 Happy Feet

The Happy Feet database [71, 72] is the only longitudinal footprint and ballprint database from newborn through infant, to the best of our knowledge, in the world. It contains captures of both feet and balls (right and left) for each of the 54 participants who were enrolled into the study "Happy Feet", which was funded by the Bill and Melinda Gates Foundation to investigate the potential of the footprint and ballprint as infant biometrics.

The data collection time frame spans 2.75 years in total, with data being collected at four ages over the stated period of time: at 2 days, 2 months, 6 months and 2 years. Visit 1 (V1), from 30th May 2012 to 06th December 2012, was to the newborns (ranging from the age of 11 hours to 683 hours with an average age of 67 hours). Visit 2 (V2), from 12th July 2012 to 13th February 2013, was to the young infants (ranging from the age of 31 days to 84 days with an average age of 55 days). Visit 3 (V3), from 07th December 2012 to 22nd May 2013, was to the infants (ranging from the age of 161 days to 211 days with an average age of 183 days). Visit 4 (V4), from 11th August 2014 to 25th February 2015, was to the old infants, (ranging from the age of 702 days to 894 days with an average age of 780 days).

All newborn and infant images were captured in one session per child. The total number of images per visit and capture methodology can be found in Table 2.3.

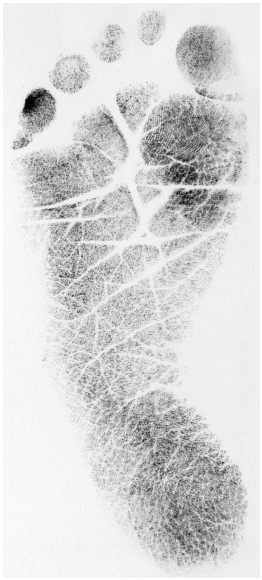
The applied capture methodology slightly varies between the first three visits and the fourth one. For the first three visits, inkless paper (i), a camera (ii) and a fingerprint scanner (iii) have been used, for the fourth one only the fingerprint scanner was used to capture the ballprint. We briefly describe in the following sections the capture protocol used. Example captures can be found in Figure 2.9.

Capture methodology	V1	V2	V3	V4	Total
Inkless paper	427	432	432	–	1,291
Camera	984 = 473 + 511	676 = 331 + 345	494 = 254 + 240	–	2154
Fingerprint scanner	589	625	628	1,147	2,989

Table 2.3: Number of individual images included in the Happy Feet database with respect to the capture methodology used and the infant's age. The total number of camera pictures per visit comprises of the number of photographs taken of the left ball and whole right foot. This is indicated by the addition, the left summand represents the quantity of ball images.

(a) Ball*

(b) Ballprint



(c) Left footprint

(d) Right footprint

(e) Right foot*

Figure 2.9: An infant foot (c, d, e) and its ball, the hallucal area under the big toe (a, b) captured with a camera (a, e), a fingerprint scanner (b) and inkless paper (c, d). The capture process is fully disclosed in Chapters 5 and 6. These particular captures belong to infant 021 and have been taken shortly after birth at the age of 2.37 days. *Please note that photographs (a, e) have been mirrored along the vertical axis to visually match their corresponding prints. This is necessary because prints capture impressions in contrast to photographs.

Inkless paper

Inkless paper prints have been taken during the first three visits only and typically, four prints of each foot were captured. We used Nekoosa Printed Products Identifier™ and captured the prints as follows: the infant's feet were cleaned with baby wipes to remove any vernix caseosa, then wiped with the inkless towelette and pressed onto

the paper in one firm motion, where the paper sheet was clipped to a book to obtain a hard flat surface to work with. Later the prints were scanned at 2400 ppi with a HP Scanjet G4010.

Camera

We used a Nikon D5100 with a Nikkor $f/2.8$ 105 mm-macro lens attached and two SB-R200 flashes mounted to the lens. Following our protocol, the whole right foot had been captured but we focused on the left ball only. The number of pictures taken ranges from none if the infant was very upset or playful up to over 50 if the participant was happily cooperating. It is worth noting that this procedure has only been performed for the first three visits.

Fingerprint scanner

We used a commercial off-the-shelf single fingerprint scanner, the NEC PU900-10, to typically capture 6 (Visits 1, 2, 3) or 12 (Visit 4) ballprint impressions per visit and individual foot. The increased number of captures is to compensate for the infant's growth, to maximise potential overlap of the prints between visits. Examples of ballprint captures are shown in Figure 2.10.

There are three main reasons to choose a commercial off-the-shelf single fingerprint scanner. Firstly, its innocuous and round design with a rather soft capture area minimises the infant's risk of injuries. It also increases both the parents' and their infant's acceptance of the scanner. Secondly, its high native sensor resolution (1000 ppi), even though the output is downsampled in hardware to 500 ppi, maximises the captured details. The print quality ranges from very poor to exceptional and hence the distinctiveness from almost nil to exceptional (cf. Figure 5.1 in Chapter 5) due to, inter alia, uncooperative subjects, the ridges' tininess and dry and cracking skin. Thirdly, the use of commercial "standard" technology is cost-effective and has already proven its suitability for the purpose for which it is marketed. This is not necessarily the case for specialised hardware which has to be developed especially for a study.

In order to capture an infant's ballprint, it has turned out that it is best to sit the infant on one's lap and to press the fingerprint scanner against the ball area. This way, it is possible to control the infant more easily regardless of its mood and encourage it to cooperate.

(a) Left, Visit 1 (b) Left, Visit 2 (c) Left, Visit 3 (d) Left, Visit 4

(e) Right, Visit 1 (f) Right, Visit 2 (g) Right, Visit 3 (h) Right, Visit 4

Figure 2.10: Infant's 021 best prints of its left (a, b, c, d) and right (e, f, g, h) foot at the ages of 2 days (a, e), 2 months (b, f), 6 months (c, g) and 2 years (d, h). These ballprints are of overall high-quality; Infant 021 has the best detail at Visit 1 of all the infants.

Challenges

There are two main challenges, firstly the infant's growth and secondly its will to cooperate. The growth represents a problem even during the capture process when working with a fingerprint scanner. The issue is the small capture area compared to the infant's rapid growth. In order to have a realistic chance to identify the infant or to verify its identity it is necessary to capture the same area of the ball across all visits. This becomes more challenging the more the infant grows because the ridges' size increases and less structure can be captured with one scan. Therefore, it is important to ensure that the same part of the hallucal pad is captured repeatedly, as highlighted in Figure 2.11.

This becomes even more challenging when the infant is unwilling to cooperate because it is scared of the scanning device or just really playful and would rather run or crawl around than sit still.

Additionally, both tiny ridge structures or non-cooperative subjects can lead to low-quality prints which force a heavy burden on the algorithms and procedures (cf.

Section 2.4).

(a) V1: 2 days (b) V2: 2 months (c) V3: 6 months

Figure 2.11: Left footprints of Infant 021 at the age of (a) 2 days, (b) 2 months and (c) 6 months. The area of the ball is circled and the corresponding ballprint can be found in the lower left-hand corner. The footprints are to scale; the change in size due to growth is evident.

2.5.2 Adult ballprint database

In 2007 and 2008, Uhl and Wild published an analysis of different footprint modalities and features regarding their suitability as an distinct biometric in a identification [144] and verification [143] scenario. Their database consists of 160 whole footprint images in total from 27 male and 5 female subjects (5 captures per right foot). They used an HP Scanjet 3500c flatbed scanner to acquire the images at a resolution of 600 ppi; its maximum scanning area was sufficiently large enough for all individuals' feet; some had to be scanned diagonally instead of straight in order to fit within the scanning area. The individual was sitting in front of the scanner while the captures were taken without putting full pressure on the scanner's glass plate.

Figure 2.12: Adult ballprint image after extraction and minutiae detection, showing 352 minutiae. Published by Uhl and Wild [144].

Originally, Uhl and Wild used an automated algorithm to extract the ballprint region. The database, consisting of scans of the entire foot, was made available to us for a "one-time-only" feature extraction under a non-disclosure agreement. In order to not affect verification results, we decided to crop the hallucal area under the big toe manually; no further adjustments such as rotation compensation have been performed. Only the image contrast has been adjusted by the application of the *Contrast Limited Adaptive Histogram Equalisation* (CLAHE) (cf. Section 3.2.2 in Chapter 3). An example of an adult ballprint (and its extracted minutiae) is shown in Figure 2.12.

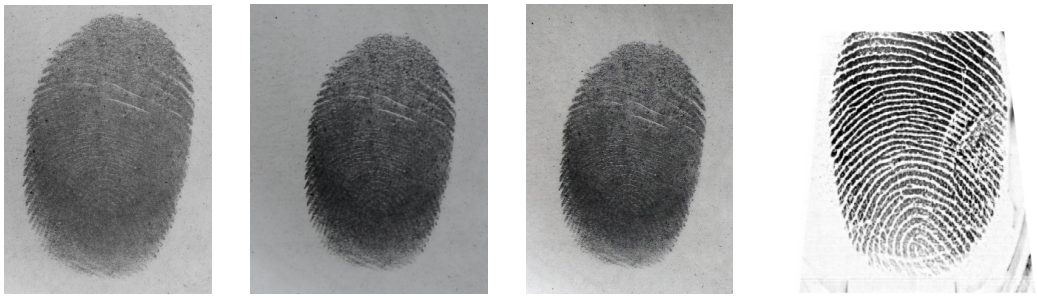
2.5.3 Victoria Police databases

The Victoria Police Forensic Services has supplied a pseudo fingermark database (artificially produced trace marks captured under controlled circumstances). It consists of normal and deliberately distorted marks captured with a variety of different devices as well as a fingerprint reference database to compare the marks against. Example images can be found in Figure 2.13.

Pseudo fingermark database

Two males and two females volunteered to create a large database of pseudo fingermarks, including normal and deliberately distorted marks. There are six categories of deliberate distortion such as finger placed "lightly" on the page or finger twisted

(a) Scanner (b) High-quality camera (c) Mobile phone



(d) Scanner (e) High-quality camera (f) Mobile phone (g) Fingerprint scanner

Figure 2.13: A subject's right middle finger "heavily" placed on the sheet captured by different devices: (a) scanner, (b) high-quality camera, (c) mobile phone. The first three images (a – c) have been cropped more closely before entering them into the database (d – f). The image captured with the fingerprint scanner (g) is used in the reference database and only shown here for reference purposes. It has been captured without any deliberate distortion.

strongly. More details such as the exact class distribution can be found in Table 4.1 in Chapter 4.

After a fingermark (traces of natural oils from the surface of the finger) was left on a paper sheet in predefined marked areas according to the specified distortion, the sheets were brushed using magnetic black powder to make the latent mark visible and conserve it. Finally, each individual sheet was laminated to ensure its permanent integrity. A Victoria Police fingerprint expert supervised the whole process and performed a closer visual inspection of the marks before and after the lamination process to confirm that no major distortions had been added.

Subsequently, all laminated sheets have been individually digitised using 3 different capture devices: (i) a flatbed scanner (HP Scanjet G4010, abbr: Scanner), (ii) a high-

Figure 2.14: The operational Victoria Police Forensic Services camera setup that has been used for capturing pseudo fingermark images.

quality camera (Nikon D3S with a Nikkor $f/2.8$ 60 mm-macro lens attached (cf. Figure 2.14), abbr: DSLR) and (iii) a mobile phone (Apple iPhone 4S, abbr: Phone). The scanner and camera captured the whole sheet at once, whereas the mobile phone photographed each mark individually at an approximate distance of 10 cm. The phone was held mostly parallel to the sheet to avoid light reflections induced by the laminate. A light source was present at all times perpendicular to both the sheet and the phone; and a remote control was used to release the shutter to minimise any movement due to touching the phone during capture. In contrast to Rodriguez *et al.* [125], we did not add any forensic background pattern.

Reference fingerprint database

Additionally, we created a reference database against which to match the pseudo fingermarks. For this purpose, we collected all ten fingerprints of the same four subjects with Digital Persona U.are.U 4000 and Digital Persona U.are.U 4500 fingerprint scanners. We captured four images per print (centre, left and right side, and the top of the finger tip) without any deliberate distortion to imitate a reference scenario. Also, we added imposter images with alike characteristics (no deliberate distortion) which were all captured with optical fingerprint scanners similar to the one we used. Specifically, we used all third prints of FVC2000 DB3 [94] and FVC2004 DB2 [96] and all sixth prints of FVC2002 DB1 [95]. This leads to a reference database consisting of 320 genuine and 330 imposter prints. We verified via the cross verification scores that no duplicate of any imposter is included. One can find a reference image next to its pseudo fingermarks in Figure 2.13.

3 Biometric data processing

This chapter sets the methods background of this thesis. It presents approaches and techniques of the adjacent research areas. This includes different classifiers in the area of machine learning (Section 3.1) or methods of computer graphics (Section 3.2) for data point representation and manipulation (Section 3.2.1) and image processing techniques to enhance biometric images (Section 3.2.2). Last but not least, we clarify the software frameworks we used (Section 3.3).

3.1 Machine learning

Machine learning is a sub-field of computer science, closely related to the fields of pattern recognition and artificial intelligence and often used in the context of computer vision. The basic idea is to show examples to a machine, a so-called training set, let it build a model of this set and apply it to new and unknown data.

More specifically, we need to distinguish between *supervised*, *unsupervised* and *reinforcement* learning. In the first case the training set has been labelled by a higher authority such as a researcher and therefore it forms a ground truth (or desired output) for this particular set of data upon which the computed model can be based. In the second case there is no ground truth available and the algorithm labels the data itself according to some criteria or pattern. This process can be either completely automatic and with no interaction required or semi-automatic for which one needs to guide the algorithm to some degree and select for example the region of interest. In the third and last case the computer faces constantly changing data and has to evaluate itself how close it has come to the desired output or goal to achieve such as performing certain tasks autonomously (controlling an artificial limp, playing a game, etc.).

There is a huge variety of different machine learning techniques. Some are simpler than other and more suited to one or the other learning task. Lately, research focuses on creating learning procedures and algorithms which replicate human learning behaviour and the human brain. One famous state-of-the-art example is "Watson". According to its inventor IBM, the supercomputer "uses the same cognitive framework that humans use to inform their decisions: Observe, Interpret, Evaluate, and Decide" [60]. Watson has been successfully trained to play Jeopardy! (and beat the human competitors) [59] and to create recipes with novel flavour combinations, even resulting in a cook book [58].

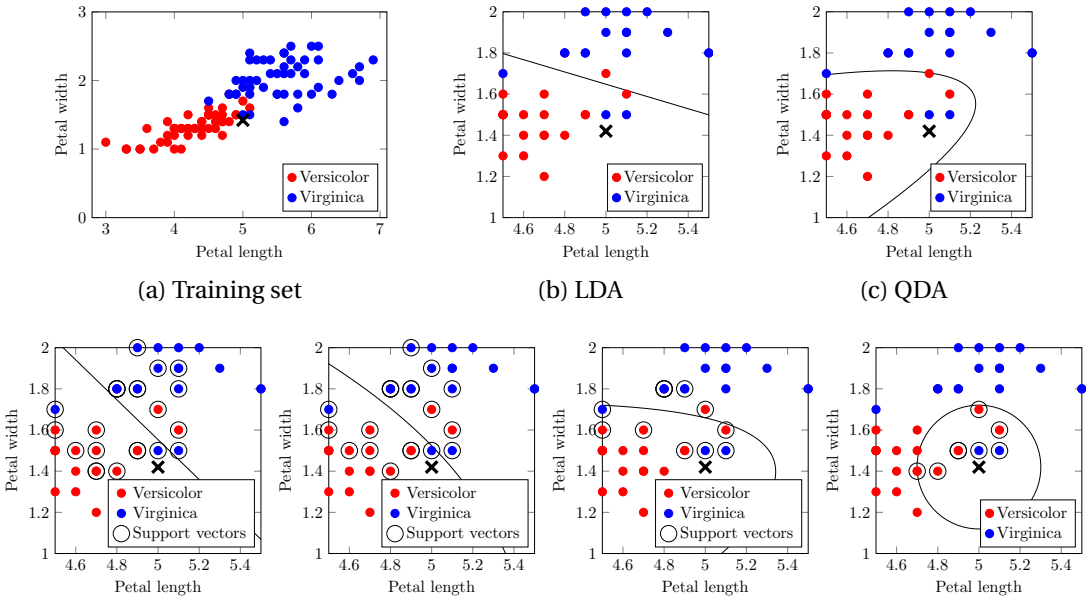
In this thesis, we use basic machine learning techniques to classify data after the classifier has been trained with labelled example data (supervised learning). For this purpose, we consider two datasets: the training set $\boldsymbol{A}^{(k\times n)}$ and test set $\boldsymbol{B}^{(k\times m)}$ in matrix notation. This means that each column represents one individual observation (consisting of k features) and each row a specific feature. Also, there is the ground truth $\boldsymbol{a} \in \{0,1\}^{(1\times n)}$ stating the corresponding class for each observation of the training set and $\boldsymbol{b} \in \{0,1\}^{(1\times m)}$ is the classification result of $\boldsymbol{B}$. The three classifiers (support vector machine, discriminant analysis, k-nearest neighbours) mentioned in the following sections are used within Chapters 4, 6 and 7.

3.1.1 Support vector machine

A *Support Vector Machine* (SVM) is a binary classifier which learns a model from a labelled training set and applies it to new data. It was introduced by Vapnik and Lerner in 1963 [154] and later extended for mapping the data into a higher dimensional feature space by the "kernel-trick" [17], for soft margins to handle mislabelled examples [25], and for the classification of multiple (instead of two) classes [33].

More specifically, the SVM tries to separate given examples linearly by a hyperplane. Now it is possible that the examples cannot be linearly separated in the given space. Hence the kernel-trick maps all training observations into a higher dimensional space which allows for linear separation. The hyperplane is set to maximise the distance (margin) to the class samples which lie the closest, they are known as support vectors. The same transformation or mapping (into the higher dimensional space) is applied to unseen data; those samples are then classified depending on the side of the separating hyperplane they lie on (mathematically speaking: their sign).

In this particular case the ground truth and the classification result are defined slightly differently as $a \in \{-1,1\}^{(1\times n)}$ and $b \in \{-1,1\}^{(1\times m)}$. Hence the classification process



(a) Training set (b) LDA (c) QDA

(d) SVM, linear kernel (e) SVM, Gaussian kernel (f) SVM, polynomial kernel (g) kNN

Figure 3.1: Fisher's well established iris flower dataset [41] is often used to demonstrate a classifier's performance or behaviour. Here it has been limited to Versicolor and Virginica (a). Different classifiers such as (b) LDA, (c) QDA, (d–f) SVM with different kernels and (g) kNN are trained on it and applied to a new data point (marked by X). All classifiers decide that the new point belongs to Versicolor; the decision boundaries are also shown.

can be expressed as follows [137]:

$$b_j = \operatorname{sgn}\left[\sum_{i=1}^{n} \alpha_i a_i \Psi\left(\boldsymbol{B}_{(\cdot,j)}, \boldsymbol{A}_{(\cdot,i)}\right) + c\right], \tag{3.1}$$

where $i \in \{1,\ldots,n\}$, $j \in \{1,\ldots,m\}$, $\alpha_i \in \mathbb{R}^+$, $c \in \mathbb{R}$ and $\Psi(\cdot,\cdot)$ is the kernel used. The sum of the weighted training examples and the mapped new data point states the distance to the hyperplane and therefore the degree of certainty for the classification result, and the sum's sign represents the class assigned. Examples for the SVM using a linear, a Gaussian and a polynomial kernel are shown in Figures 3.1d to 3.1f.

3.1.2 Discriminant analysis

The *Linear Discriminant Analysis* (LDA) is a method to linearly separate two (or more) classes by the linear combination of their features. This mapping aims to discriminate the given training classes as clearly as possible (by maximising the difference of

their means) and can be applied to unseen data which is then implicitly classified depending on the side of the hyperplane onto which it is projected.

Fisher introduced the concept of the LDA as early as 1936 [41]. The terms LDA and *Fisher's Discriminant Analysis* (FDA) are often used interchangeably although the LDA makes a few assumptions, such as normal class distribution or the identical class covariances which are not present in the original paper.

The LDA chooses a statistical approach and assumes that the probability density functions $p(\boldsymbol{A}|a=1)$ and $p(\boldsymbol{A}|a=0)$ of the two classes ($\boldsymbol{A}_{\cdot,a=1}$ and $\boldsymbol{A}_{\cdot,a=0}$) are multivariate normally distributed. They can be described by their mean and covariance matrices being $(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, respectively. The LDA makes now the simplification that $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}$, so only the classes' means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_0$ vary. This leads to the following simplified mapping rule for the data point $\boldsymbol{B}_{(\cdot,j)}$ [98]:

$$\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)\boldsymbol{B}_{(\cdot,j)}{}^T, \tag{3.2}$$

where the result indicates by its sign the side of the hyperplane onto which the point is projected and hence its classification result.

Another variant of a discriminant analysis is closely related to LDA, the *Quadratic Discriminant Analysis* (QDA). The main difference between the two variants is that QDA does not make the assumption of identical class covariances and thereby leads to a quadratically separating hyperplane instead of a linear one. Examples for the LDA and QDA based on Fisher's iris dataset are illustrated in Figures 3.1b and 3.1c, respectively.

3.1.3 *k*-nearest neighbours

The *k-nearest neighbours algorithm* (kNN) is a simple classification approach based on the assumption that similar data points most likely have the same class origin [4,15]. There is no training procedure required.

For a test data point that needs to be classified, its k closest ($k \in \mathbb{N}$) neighbours of training points (according to a given metric) are computed and a majority voting based on their class is performed. Therefore it is reasonable to choose an odd k for a binary classification problem to avoid a draw.

Also, the two classes might be skewed and one might occur much more frequently than the other. One can counteract this effect by normalising the number of votes with the estimated class frequency for example. Another option is to take the distance between

the data point to be classified and the training samples into account and weight their vote by the reciprocal distance. An example without weighting or normalisation can be found in Figure 3.1g.

3.2 Computer graphics

Computer graphics is a sub-field of computer science that evolves around the generation, manipulation and analysis of (digital) visual content such as 2 D imagery. It covers the area of image processing but also mathematical topics such as the mathematical representation, transformation and registration of points and point clouds [46]. In the following sections we introduce these topics and elaborate on them before we expand on basic image processing techniques used.

3.2.1 Point representation and manipulation

Let's assume we have one 2 D point, say $a^{(3\times1)} \in \mathbb{R}^3$ and its spatial position within the Cartesian coordinate system is noted by the vector $\begin{bmatrix} x & y & 1 \end{bmatrix}^T$. The third element or dimension is due to the use of homogeneous coordinates which allow us to apply transformations such as the similarity or affine transformation with a single matrix multiplication [46]. The coordinates x and y are real numbers $x, y \in \mathbb{R}$.

A point cloud $\boldsymbol{A}$ consisting of multiple points, say n, is represented by the column-wise concatenation of the individual point vectors:

$$\boldsymbol{A}^{(3\times n)} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ y_1 & y_2 & \ddots & y_n \\ 1 & 1 & \cdots & 1 \end{bmatrix}. \tag{3.3}$$

Now a transformation can be applied to the whole point cloud by multiplication with the real transformation matrix $\boldsymbol{T}^{(3\times3)}$; the resulting transformed points are stored in matrix $\boldsymbol{B}^{(3\times n)}$ which has the same dimensions as the original matrix $\boldsymbol{A}$.

$$\boldsymbol{B} = \boldsymbol{T}\,\boldsymbol{A} \tag{3.4}$$

The exact definition of $\boldsymbol{T}$ is subject to the transformation used and is explained in the two following sections.

Similarity transformation

The similarity transformation consists only of operations that preserve similarity (in the mathematical sense) between the original and its transformed point cloud or object. Two objects are considered similar if one is the uniformly scaled, translated, rotated and/or reflected version of the other one. This means that the distance of two randomly chosen points expressed as ratio between the original and the transformed version remains constant regardless the chosen similarity transformation:

$$\frac{\boldsymbol{A}_{(\cdot,i)} - \boldsymbol{A}_{(\cdot,j)}}{\boldsymbol{T}\boldsymbol{A}_{(\cdot,i)} - \boldsymbol{T}\boldsymbol{A}_{(\cdot,j)}} = c, \text{ where } i, j \in \{1, \ldots, n\} \text{ and } i \neq j. \tag{3.5}$$

Those requirements can be expressed as follows:

$$\boldsymbol{T} = \begin{bmatrix} k\cos\theta & \sin\theta & t_x \\ -\sin\theta & k\cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{3.6}$$

with θ being the rotation angle (clockwise, in radians), k the scaling factor ($k = 1$ means no scaling) and t_x and t_y being the translation in x and y direction, respectively.

Geometrically speaking, all circles are similar to each other (so are all squares) but not necessarily all ellipses, rectangles or triangles.

Examples of similar triangles (all can be converted into each other by a similarity transformation) are shown in Figure 3.2a.

Affine transformation

The affine transformation allows for all operations used within the similarity transformation plus additional operations such as shearing. Also, all of these operations can be combined in any order. Geometrically speaking, an affine transformation preserves parallel lines but does not preserve angles between lines or distances between points.

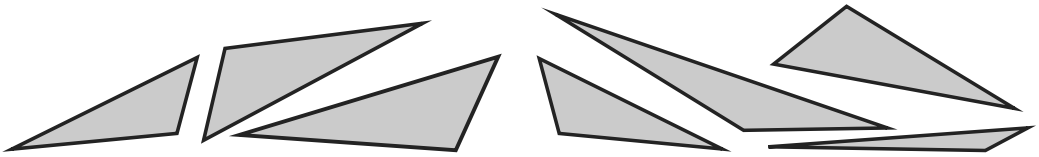
This leads to the following basic transformation matrix:

$$\boldsymbol{T} = \begin{bmatrix} k_x\cos\theta & k_y\sin\theta & t_x \\ -k_x\sin\theta & k_y\cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{3.7}$$

where θ is the rotation angle (clockwise in radians), k_x and k_y are scaling factors and t_x and t_y are the translation in x and y direction, respectively. It is possible to create

more complex transformation matrices by combining (multiplying) several affine transformation matrices.

Some triangles resulting from affine transformations are displayed in Figure 3.2b.



(a) Similarity transformation (b) Affine transformation

Figure 3.2: Triangles resulting from (a) a similarity and (b) an affine transform.

Procrustes analysis

Procrustes analysis is often used for statistical shape analysis. Its name indicates the purpose; it is named after the Greek bandit Procrustes who made his victims fit into his bed by either stretching or cutting off their limbs.

Specifically, the full Procrustes analysis finds the similarity transform that minimises the squared error between two shape configurations [32] whereas the partial Procrustes analysis ignores scaling ($k = 1$). In general, if there are multiple shape configurations of the same shape, it can be used to determine a mean shape and to measure the variability along a certain dimension.

In our case we apply it in Chapter 6 to find the transformation matrix $\boldsymbol{T}$ that projects two point clouds $\boldsymbol{A}$ and $\boldsymbol{B}$ as closely as possible onto each other.

Random Sample Consensus

The *Random Sample Consensus* (RANSAC) was introduced by Fischler and Bolles in 1981 [40]. It is a method or approach to estimate a distribution model for data (say $\boldsymbol{P}$) biased by outliers (cf. Figure 3.3) and can be applied to any distribution model as long as there are enough data points as needed for the estimation of the model's parameters (say n).

RANSAC can be broken down into three basic steps.

1. Choose randomly a subset $\boldsymbol{Q}$ consisting of n points from $\boldsymbol{P}$: $\boldsymbol{Q} \subseteq \boldsymbol{P}, |\boldsymbol{Q}| = n$. Compute the distribution model $\boldsymbol{M}$ based on $\boldsymbol{Q}$. Data points from $\boldsymbol{P}$ which fulfil

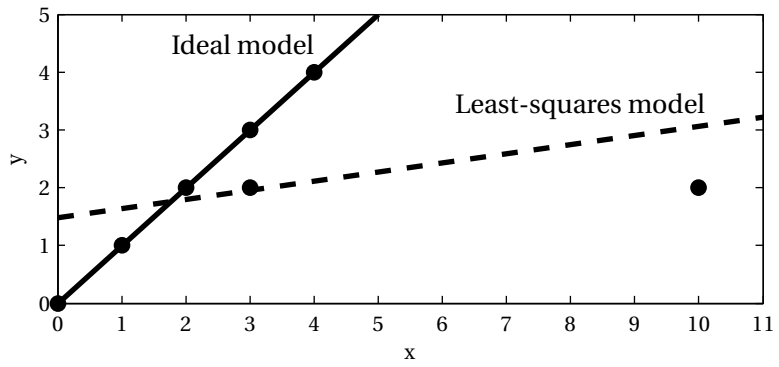


Figure 3.3: Example of different underlying distribution models for a specific dataset, adapted from Fischler and Bolles [40]. One outlier influences a simple model estimation method heavily (method of least-squares, dashed line), whereas RANSAC estimates the model upon which the data is actually based (solid line). One data point is noise afflicted but can still be explained with the ideal model.

the model to a certain degree (e.g. the distance between the estimated point and its original is smaller than some given threshold) form the consensus set $\boldsymbol{Q}'$.

2. If there is great consensus ($|\boldsymbol{Q}'| \geq t$), recompute a refined model based on $\boldsymbol{Q}'$.

3. If there are not enough consenting points ($|\boldsymbol{Q}'| < t$), repeat the first two steps until either a sufficient model has been found or a pre-defined number of repetitions is reached. In the latter case, the model based on the largest consensus set or an error can be returned.

Figure 3.3 highlights the advantages of RANSAC. It is able to estimate the model despite the presence of outliers which lead to a biased estimate when other approaches are used (see Procrustes analysis in Section 3.2.1). In this thesis it is used to reliably estimate a similarity or affine transformation (cf. sections 3.2.1 and 3.2.1) so that one point cloud maps onto another.

Iterative closest point

The *Iterative Closest Point* (ICP) algorithm was introduced in 1992 by two independent research groups [90]: Besl and McKay [11] but also Chen and Medioni [22]. The algorithm is designed to register (roughly aligned) 3 D point sets or clouds onto each other and can handle missing data sufficiently well [126]. Over the last years, multiple varieties and versions have been published. The original version by Chen and Medioni can align point clouds distorted by *one* rigid transformation (rotation, translation and

reflection). Later implementations allow for geometric alignment beyond the rigid transformation, including scaling as well [79].

The algorithm can be broken down into four basic steps. For demonstration purposes, we assume that there are two point clouds $\boldsymbol{A}$ and $\boldsymbol{B}$. $\boldsymbol{A}$ is considered to be the reference and fixed whereas $\boldsymbol{B}$ is the cloud which is aimed to align with the reference cloud.

1. For each individual point in $\boldsymbol{B}$ find the closest (distance-wise) point in $\boldsymbol{A}$ to create corresponding point pairs.
2. Estimate the transformation that allows us to align these point pairs as closely as possible subject to an error function such as the mean squared error. Some ICP versions weight the point pairs or reject specific ones to provide a more reliable transformation estimate.
3. Apply the estimated transformation to $\boldsymbol{B}$.
4. Repeat the previous steps until the error converges or is smaller than a given threshold.

We use the ICP algorithm in Chapter 5.4 to align point patterns of extracted infant footprint crease.

Graph representation and matching

A graph is a mathematical structure comprised of a set of vertices (also referred to as nodes) and edges connecting pair-wise some (not necessarily all) vertices. The edges can be either undirected or directed and may carry additional attributes.

More formally, following the definition by Riesen and Bunke [124], a graph $\boldsymbol{G}$ is defined by $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E}, \mu, \upsilon)$, where $\boldsymbol{V}$ is the set of vertices, $\boldsymbol{E} \subset \boldsymbol{V} \times \boldsymbol{V}$ is the set of edges, $\mu : \boldsymbol{V} \to \boldsymbol{L}_\mu$ is the vertex labelling function which maps each vertex υ onto its label, and $\upsilon : \boldsymbol{V} \to \boldsymbol{L}_\upsilon$ is the edge labelling function that assigns a label to an edge e.

The labels $\boldsymbol{L}_\mu$ and $\boldsymbol{L}_\upsilon$ are not constrained in any way and could be elements of the symbolic alphabet or a vector space $\mathbb{R}^n$ of the arbitrary dimensionality n. A graph becomes unlabelled if all vertices and edges are assigned the same label.

Edges are defined as pairs of vertices $e = (u, \upsilon)$, where u is the source vertex and υ is the target vertex and $u, \upsilon \in \boldsymbol{V}$. This represents the directed case. Riesen and Bunke suggest for the undirected one that the reversed edge has to be included $(u, \upsilon), (\upsilon, u) \in \boldsymbol{E}$ and possess the same labels $\upsilon(u, \upsilon) = \upsilon(\upsilon, u)$ as well [124]. However, we are only using

undirected graphs in this thesis and assume that $(u, v) = (v, u)$ regardless which edge is included in the edge set $\boldsymbol{E}$.

Graph matching refers to the task of comparing two graphs in terms of their structural similarity. There are two basic approaches: the exact matching and the error-tolerant matching [124]. Exact graph matching investigates if the two graphs are identical in structure and labels. This is a non-trivial task because of a graph's unordered nature [120, 124]. In a real world scenario, the supplied labels might differ slightly between the two graphs making exact matching nearly impossible. This is especially the case when they are drawn from actual observations. Hence a less restrictive matching process is needed: the error-tolerant matching which takes variation up to a certain degree into account. Figure 3.4 illustrates some example graphs.

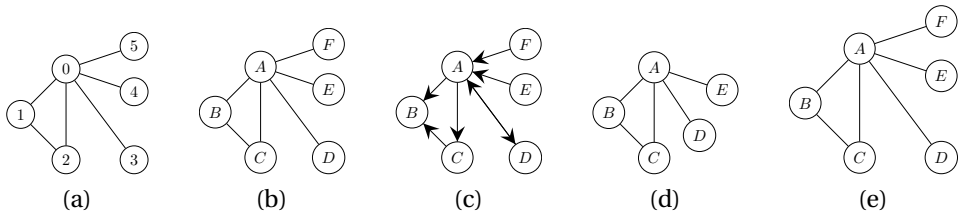


Figure 3.4: Representation of different graphs highlighting the difficulties of exact matching. The graphs (a) and (b) possess only different labels, (b) is the undirected variant of (c), (d) is a distorted version of (b), and (e) is a scaled variant of (b).

In this thesis, we use the recently introduced *Biometric Graph Matching* (BGM) [56]. It builds upon this very basic definition and is suited to the representation and error-tolerant matching of ridge and vessel-based biometrics. It has been applied successfully to fingerprint [56], retina [84], hand vein [85], and palm vein [7].

The exact definition and the graph extraction process for ballprint is stated in Section 7.1.1 in Chapter 7. Additionally, we introduce and motivate *scale-invariant BGM* (sBGM) in Chapter 7.1.2 in Chapter 7 which uses scale-invariant features instead of spatial ones to overcome infant growth implicitly without having to model it explicitly.

3.2.2 Image processing

Image processing is a wide field that is strongly interwoven with machine learning and often considered integral to computer vision. In (digital) image processing an input image is enhanced by applying mathematical functions to it; afterwards either the whole image or some extracted features are returned. Image processing is often used to ensure that input images fulfil certain pre-specified criteria in context of the appli-

cation for which they are going to be used. One example is the contrast enhancement of images, which varies according to the images' purpose. A photographer for example would like to improve the contrast, change the white balance, restore blown-out details or manipulate the colour composition to create a visually more appealing image. In a medical context, images capturing the cardiac-vascular system might instead be optimised for the vessel's visibility without amplifying any background noise, to make it easier for the physician to interpret the images and arrive at a diagnosis.

In this thesis we use some standard image processing techniques which are explained within the different chapters because they are not vital to the understanding of the thesis. Nevertheless, there are two basic reoccurring techniques, the *Contrast Limited Adaptive Histogram Equalisation* (CLAHE) and some basic functions of the sub-field of morphological image processing. CLAHE improves the image contrast in order to make the invisible visible. Morphological image processing is here used to correct an image's interpretation such as the connectivity between recognised objects. A more specific example for the latter is the connectivity between ridge lines in fingerprints, to retain *correct* connections but remove *false* ones.

Contrast limited adaptive histogram equalisation

The *Contrast Limited Adaptive Histogram Equalisation* (CLAHE) is an image processing method to improve the contrast in images while avoiding the amplification of any background noise present at the same time. Pizer *et al.* have introduced it in 1987 [121] as an extended version of the regular *Histogram Equalisation* (HE) to battle some of its shortcomings.

The HE aims to improve the global contrast by spreading the image gray value distribution over the entire range and maximising the distance of different gray values according to their frequency of occurrence.

We assume that there is an image $\boldsymbol{I}^{(n\times m)}$ consisting of n rows and m columns, where every individual pixel, say $I_{(i,j)} \in \{0,\dots,L\}$ with $i \in \{1,\dots,n\}$ and $j \in \{1,\dots,m\}$ can have a (gray) value between 0 (black) and $L \in \mathbb{N}$ (white); usually $L = 255$.

The discrete histogram function $p_{\boldsymbol{I}}(k)$ returns the frequency of a given gray value $k \in \{0,\dots,L\}$ within the image.

$$p_{\boldsymbol{I}}(k) = \sum_{i=1}^{n}\sum_{j=1}^{m} \delta_{(I_{(i,j)},k)}, \text{ with } \delta_{(x,y)} \text{being the Kronecker delta function.} \tag{3.8}$$

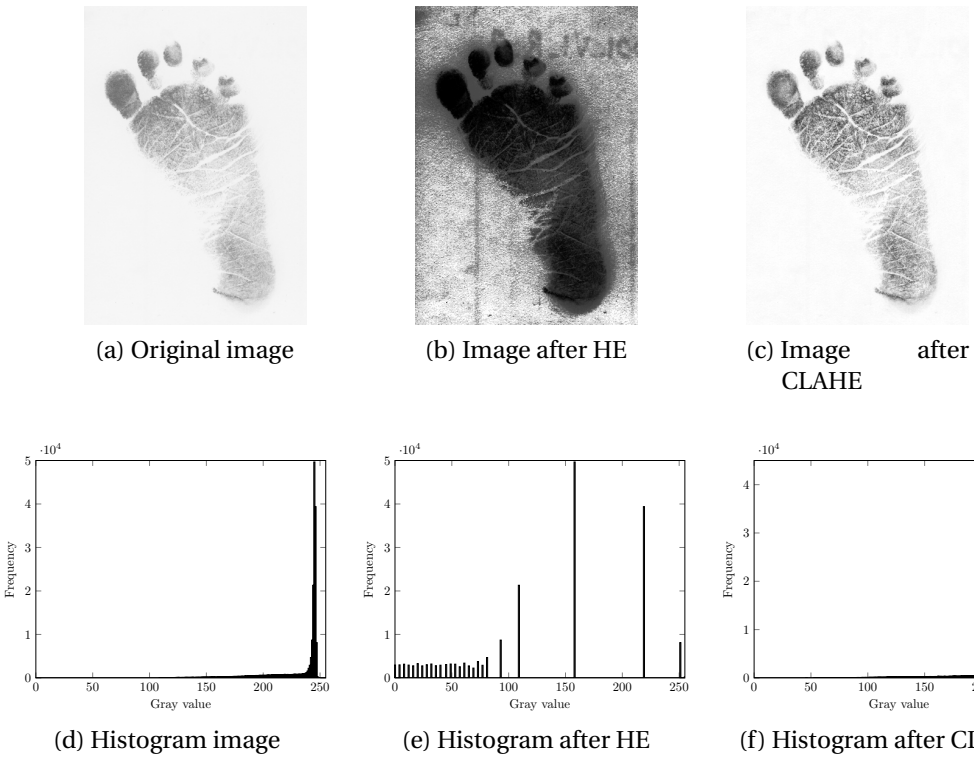
The cumulative distribution function $cdf_{\boldsymbol{I}}$ (here in the context of image $\boldsymbol{I}$) returns the frequency of occurrence for any gray value smaller than or equal to k:

$$\mathrm{cdf}_{\boldsymbol{I}}(k) = \sum_{l=0}^{k} p_{\boldsymbol{I}}(l). \tag{3.9}$$

Now the histogram bins' positions are transformed according to their distribution across the entire image range. All pixels with a certain gray value, say v, are assigned a new gray value v'.

$$v' = \mathrm{he}(v) = \mathrm{round}\left(\frac{\mathrm{cdf}_{\boldsymbol{I}}(v) - cdf_{min}}{mn - cdf_{min}}\right), \tag{3.10}$$

where cdf_{min} refers to the smallest k corresponding to the first non-zero result.



(a) Original image (b) Image after HE (c) Image after CLAHE

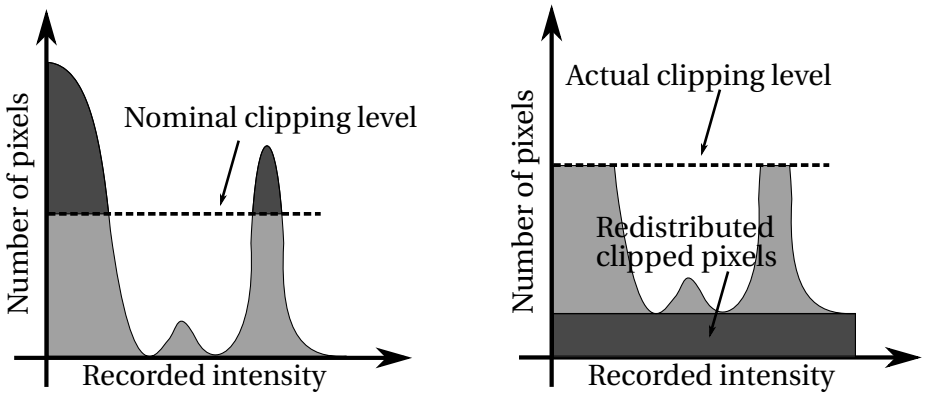
(d) Histogram image (e) Histogram after HE (f) Histogram after CLAHE

Figure 3.5: Right footprint of infant 001 captured shortly after birth shown without further processing (a) and its histogram (d), after histogram equalisation (b) and (e), and after contrast limited adaptive histogram equalisation (c) and (f).

The problem with the regular histogram equalisation is that it optimises the global contrast of the entire image. This can lead to unwanted effects if there are much lighter or darker areas as shown in Figure 3.5. *Adaptive Histogram Equalisation* (AHE)

overcomes this issue because it enhances the contrast locally instead of globally and hence can better adapt to lighter or darker areas [57]. Basically, AHE divides the entire image into smaller areas and performs the regular HE on each of these blocks. Blocks containing homogeneous gray values lead to a histogram with one high peak or a few spikes, all located within the same value range. The histogram equalisation process distributes these spikes across the entire range. This is the reason AHE tends to over-amplify any existing background noise [168].

This is counteracted by CLAHE which limits the maximum contrast. Mathematically, the contrast enhancement is given by the slope of the cumulative distribution function (Equation 3.9). If one now limits the slope so that does not become too steep (and would lead to over-enhancement) it can be interpreted as clipping histogram peaks. In order to keep the number of pixels or area size constant, the clipped counts are uniformly re-distributed across all histogram bins of the neighbourhood (example shown in Figure 3.6) If the re-distribution leads for some bins to a slope greater than the original threshold given, one can repeat the process iteratively until no further clipping occurs.



(a) Contrast limitation (b) Area redistribution

Figure 3.6: CLAHE limits the maximum contrast by (a) clipping high histogram counts and (b) redistributing their counts across all bins. The illustrations are based on [121].

Morphological image processing

Morphological image processing is an area of image processing which is based on the theory of morphology in mathematics [128]. It can be applied to n-dimensional Euclidean spaces [46, chap. 9]. However, this section limits the explanation of the basic operations to binary images because of their relevance to the thesis. More details and complex operations can be found in [31] and [46, chap. 9].

Basic operations Every pixel $I_{(i,j)}$ of a binary image $\boldsymbol{I}^{(n\times m)}$ with n rows and m columns is either 0 or 1.

$$I_{(i,j)} \in \{0,\ 1\},\ i \in \{1,\ldots,n\},\ j \in \{1,\ldots,m\} \tag{3.11}$$

All basic operation operations require a *Structure Element* (SE) which is also a binary image but with smaller dimensions than the image $\boldsymbol{I}$ it is supposed to operate on, $\boldsymbol{S}^{(k\times l)}$, $k < m$ and $l < n$.

The $\circ$ operator represents the Hadamard product, the element-wise multiplication of two matrices. In the case of two binary images, say $\boldsymbol{I}^{(n\times m)}$ and $\boldsymbol{I'}^{(n\times m)}$, it is the pixel-wise logical *and* (&&) and results in a third binary image $\boldsymbol{I''}^{(n\times m)}$ with the same dimensions.

$$\boldsymbol{I''} = \boldsymbol{I} \circ \boldsymbol{I'} \ \rightarrow\ I_{(i,j)} \cdot I'_{(i,j)} = I_{(i,j)}\ \&\&\ I'_{(i,j)} = I''_{(i,j)} \tag{3.12}$$

The ! operator is the pixel-wise binary inversion of an image $\boldsymbol{I}^{(n\times m)}$:

$$\boldsymbol{I'} = !\boldsymbol{I} \ \rightarrow\ I'_{(i,j)} = 1 - I_{(i,j)}. \tag{3.13}$$

Erosion The erosion requires the SE to fit completely into the image structure or in other words: the pixel intersection between the image structure and the SE. Hence this operation trims off the border of objects in the size of the Structure Element or even removes small objects if the SE does not fit. More formally, we assume that there is a copy of the SE $\boldsymbol{S}$ translated by z and centred at pixel $I_{(i,j)}$. The result of the erosion is `true` (equals 1) for that particular pixel if and only if all pixels of the image equal 1 where the pixels of the translated S equal 1.

$$\text{erosion}(\boldsymbol{I},\ \boldsymbol{S}) = \boldsymbol{I} \ominus \boldsymbol{S} = \{\boldsymbol{z} | (\boldsymbol{S})_z \circ !\boldsymbol{I} = \boldsymbol{0}\} \tag{3.14}$$

Figures 3.7d and 3.7h show an example of the application process and its result, respectively.

Dilatation The dilatation basically enlarges an image structure by extending its border by the size of the SE. This resembles the union of the image structure and the Structure Element. $\hat{\boldsymbol{S}}$ is a mirrored version of the SE $\boldsymbol{S}$. Now we assume that $\hat{\boldsymbol{S}}$ is translated by z and centred at pixel $I_{(i,j)}$. The dilatation for this particular pixel is `true` if theSE and the image $\boldsymbol{I}$ have at least one pixel with the value 1 in

common.

$$\text{dilation}(\boldsymbol{I},\ \boldsymbol{S}) = \boldsymbol{I} \oplus \boldsymbol{S} = \left\{ \boldsymbol{z} | (\hat{\boldsymbol{S}})_z \circ \boldsymbol{I} \neq \boldsymbol{0} \right\} \tag{3.15}$$

An example can be found in Figures 3.7e (application) and 3.7i (outcome).

Opening The following equation defines the opening operation:

$$\text{opening}(\boldsymbol{I},\ \boldsymbol{S}) = \boldsymbol{I} \otimes \boldsymbol{S} = (\boldsymbol{I} \ominus \boldsymbol{S}) \oplus \boldsymbol{S}. \tag{3.16}$$

Figure 3.7f demonstrates the operator's application, Figure 3.7j the result.

Closing This equation declares a closing:

$$\text{closing}(\boldsymbol{I},\ \boldsymbol{S}) = \boldsymbol{I} \odot \boldsymbol{S} = (\boldsymbol{I} \oplus \boldsymbol{S}) \ominus \boldsymbol{S}. \tag{3.17}$$

An example for the operator's application and the resulting outcome can be found in Figures 3.7g and 3.7k, respectively.

Combined operations As well as the basic operators, there are more complex operations which are combine basic ones.

White tophat The White tophat is often referred to as *tophat* and is given by the following equation:

$$\text{tophat}(\boldsymbol{I},\ \boldsymbol{S}) = \boldsymbol{I} \circ !(\boldsymbol{I} \otimes \boldsymbol{S}). \tag{3.18}$$

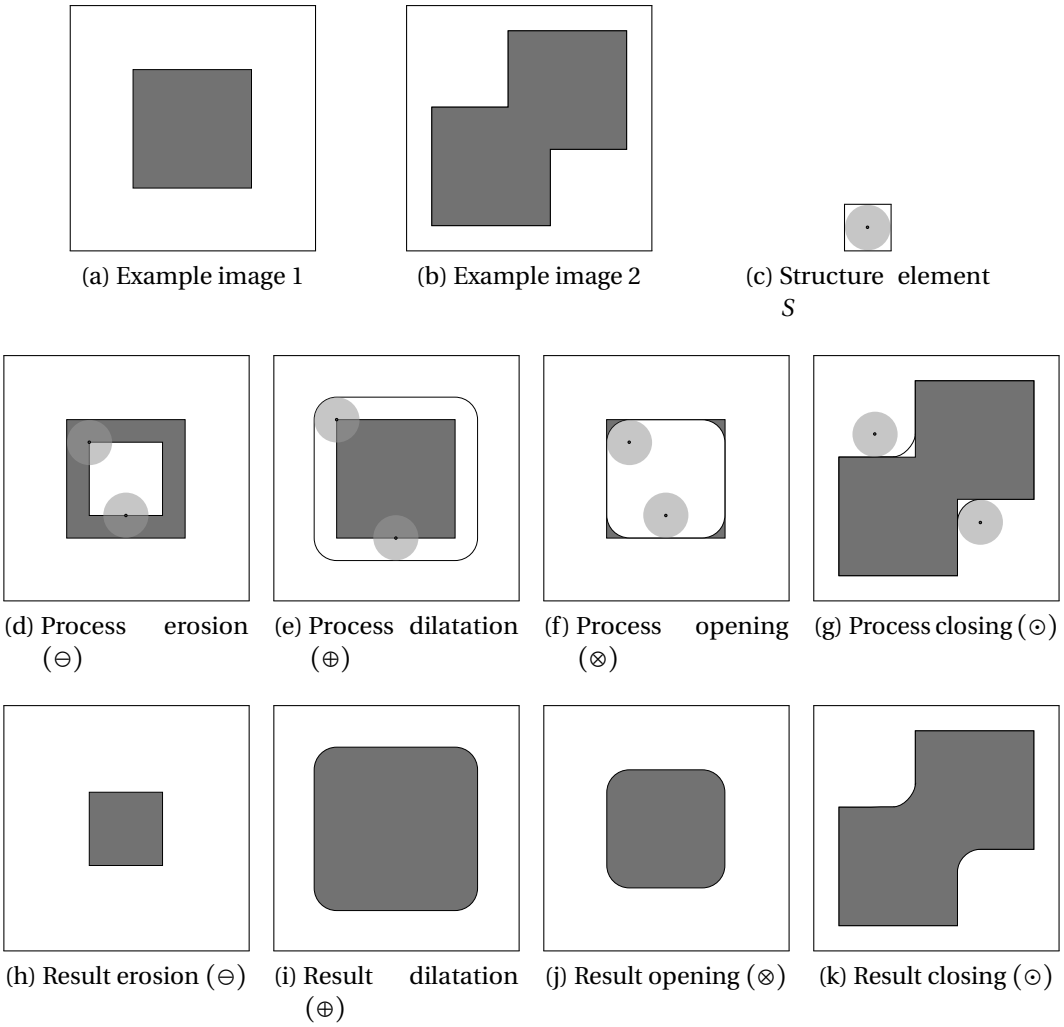
If the tophat is applied to Figure 3.7a it would result in the dark corners shown in Figure 3.7f.

Black tophat The black tophat is also known as *bothat* and defined as

$$\text{bothat}(\boldsymbol{I},\ \boldsymbol{S}) = (\boldsymbol{I} \odot \boldsymbol{S}) \circ !\boldsymbol{I}. \tag{3.19}$$

The white corner areas surrounded by a black border in Figure 3.7g are the result if the bothat is applied Figure 3.7b.

Hit-or-Miss transformation This operation requires the SE $\boldsymbol{S}$ to be slightly redefined. $\boldsymbol{S}$ contains now *two* binary images $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$: $\boldsymbol{S} = \{\boldsymbol{S}_1,\ \boldsymbol{S}_2\}$. Additionally, $\boldsymbol{S}_1 \circ \boldsymbol{S}_2$ is required to result in a matrix (or image) with all zeros. Typically, one SE is the



(a) Example image 1 (b) Example image 2 (c) Structure element S

(d) Process erosion ($\ominus$) (e) Process dilatation ($\oplus$) (f) Process opening ($\otimes$) (g) Process closing ($\odot$)

(h) Result erosion ($\ominus$) (i) Result dilatation ($\oplus$) (j) Result opening ($\otimes$) (k) Result closing ($\odot$)

Figure 3.7: Examples of the basic morphological operations erosion (d, h), dilatation (e, i), opening (f, j) and closing (g, k); split into the process of their application (d, e, f, g) and their respective results (h, i, j, k).
Consider the binary image $\boldsymbol{I}$ as defined in this section. It spans the whole area surrounded by the outer border. The dark part of the image (the centred rectangle (a) or geometric figure (b)) represents all pixels $I_{(i,j)}$ with the value 1. The circle is the binary SE $\boldsymbol{S}$ and here $\boldsymbol{S} = \hat{\boldsymbol{S}}$ because of its symmetry.
The small white areas enclosed by the internal border represent the result after applying the respective operator to the image (d, f) or to the union of the dark and white areas (e, g). The first three operators (erosion, dilatation, opening) are applied to example image 1 (a), whereas closing is applied to example image 2 (b).

inversion of the other ($\boldsymbol{S}_2 = !\,\boldsymbol{S}_1$). The function $homt$ is defined as follows [31]:

$$\mathrm{homt}(\boldsymbol{I},\, \boldsymbol{S}) = \boldsymbol{I} \circledast \boldsymbol{S} = (\boldsymbol{I} \ominus \boldsymbol{S}_1) \circ (!\boldsymbol{I} \ominus \boldsymbol{S}_2) \tag{3.20}$$

The result of this morphological transformation means that the SE $\boldsymbol{S}_1$, with its centre located at $I_{(i,j)}$ is completely contained in $\boldsymbol{I}$ and that at the same time $\boldsymbol{S}_2$ at $I_{(i,j)}$ is completely contained in the background of $\boldsymbol{I}$ ($!\boldsymbol{I}$).

Thinning The SE $\boldsymbol{S}$ contains a sequence of SEs, an original and its rotated copies: $\boldsymbol{S} = \{\boldsymbol{S}_1, \boldsymbol{S}_2, \ldots, \boldsymbol{S}_n\}$. More specifically, every SE is the previous element rotated by $(360/n)°$. Typical choices are $n = 4$ or $n = 8$ depending on what neighbourhood connectivity is supposed to be eliminated [46].

$$\boldsymbol{I} \otimes \{\boldsymbol{S}\} = ((\ldots((\boldsymbol{I} \otimes \boldsymbol{S}_1) \otimes \boldsymbol{S}_2)\ldots) \otimes \boldsymbol{S}_n) \tag{3.21}$$

This operation is repeated until there are no further changes. Furthermore, thinning is a special case of the Hit-or-Miss transformation [46]:

$$\text{thin}(\boldsymbol{I},\ \boldsymbol{S}) = \boldsymbol{I} \otimes \boldsymbol{S} = \boldsymbol{S} \circ !(\boldsymbol{I} \circledast \boldsymbol{S}) \tag{3.22}$$

In contrast to the hit-or-miss transformation, this operation does not perform the background but only the foreground operation.

3.3 Software and feature sets

We use two main *Software Development Kit* (SDK) or frameworks in this thesis. One is the *NIST Fingerprint Image Quality* 2.0 (NFIQ2) (Section 3.3.1), which is an open-source toolbox for estimating and evaluating the quality of a given fingerprint or fingermark. The other one is the commercial fingerprint matcher and feature extractor *Neurotechnology Verifinger* (VF) (Section 3.3.2).

3.3.1 NFIQ

The National Institute of Standards and Technology develops standards, evaluates technologies and shares the results publicly. Its assignment is to "promote U.S. innovation and industrial competitiveness by advancing measurement science, standards, and technology in ways that enhance economic security and improve our quality of life" [113]. The Image Group, a subgroup of NIST's Information Technology Laboratory, responsible for all tasks involving biometric quality, sees the importance of quality estimation for biometric specimens [112]: "Performance of biometric systems is dependent on the quality of the acquired input samples. If quality can be improved, either by sensor design, by user interface design, or by standards compliance, better performance can be realized. For those aspects of quality that cannot be designed-in,

an ability to analyze the quality of a live sample is needed." The difficulties involved in the capture process and the need for robust and efficient quality estimation are also discussed in Section 2.4 in Chapter 2.

In 2004, NIST introduced the first version of *NIST Fingerprint Image Quality* (NFIQ), an open-source framework to predict the performance of minutiae matchers. They extract features much more reliably from high-quality images and thereby can determine much more effectively and accurately (higher score if it is a genuine comparison) if two fingerprints belong to the same individual or not [138]. Hence, if one predicts a matcher's performance, one indirectly estimates the quality of fingerprint images as well.

The NIST Fingerprint Image Quality (NFIQ) framework is comprised of a feature extraction unit and a pre-trained classifier to analyse the image feature set extracted and classify it into one of five categories ranging from 5 to 1, namely poor, fair, good, very good and excellent. The final result is based (amongst other features) on the image percentage that consists of unstable areas (low-contrast, low ridge-flow, high curvature) or the number of high-quality (or low-quality) minutiae in certain regions. Tabassi *et al.* point out that the algorithms have been tuned towards a capture resolution of 500 ppi [138].

So, fingermarks require robust methods to estimate their quality , the amount of (relevant) information present because all factors mentioned above will vary and influence the quality and its estimate. Yoon and Jain demonstrated in [165] that the current NIST quality estimator reference implementation NFIQ is outdated because the FBI's fingerprint database system IAFIS (cf. Chapter 2.1) was able to return a print's mate although it has been classified to have the lowest possible quality (*poor*).

Currently, NFIQ2 is under development and closing this gap; it is scheduled to be released soon. A feature definition guide has been published [139]. It lists features to measure certain fingerprint properties (assuming that there is a fingerprint in the image) and to permit derivation of a quality estimate. Features include for example the "*Radial Power Spectrum* (RPS)", "Mu" or the "*Orientation Certainty Level* (OCL)".

The *Radial Power Spectrum (RPS)* measures the energy in a certain frequency band that usually corresponds with the inter-ridge spacing and therefore carries more energy compared to the neighbouring bands. If this peak does not lie within a certain frequency band or does not exist at all, the print is assumed to be of low quality.

Mu is simply the average grayscale value across the whole image and values too high or too low indicate that the capture is either too dark or too bright as it can happen

when a person's hands are too moist or too dry.

The *OCL* estimates block-wise the ridge orientation by its intensity gradient and derives a certainty level by the ratio of the two gradient eigenvalues. It is assumed that the orientation can be clearly estimated if one eigenvalue is much larger than the other. The final quality score is the average of all OCL values.

All NFIQ2 features can be used with a mask excluding non-fingerprint regions. See the NFIQ2 specification document [139] for a more detailed discussion, examples or an algorithmic description.

However, NFIQ2 is still primarily developed for fingerprints captured at a known resolution (500 ppi) with a contact fingerprint scanner as the parameters within the specification suggest.

The NFIQ2 feature set is used to estimate the quality of fingermarks in Chapter 4 and to decide based on those features if a given mark may be of value in a police forensic investigation. Subsequently these quality features are applied and adjusted to infant ballprints in Chapters 6 and 7 in order to avoid comparing low quality ones (with hardly any information present and low distinctiveness).

3.3.2 Neurotechnology Verifinger

Neurotechnology Verifinger (VF) is a commercial *Software Development Kit* (SDK). The first version named Verifinger was introduced back in 1998 but the company released other fingerprint systems as early as 1991, after it had been founded in 1990 under the name of Neurotechnologija in Vilnius, Lithuania.

According to Neurotechnology [114], their software is used in many end-user products, has won multiple awards in the *Fingerprint Verification Competition* (FVC) [94–96] and is fast but reasonably priced.

This commercial Software Development Kit (SDK) and its matching and image processing algorithms are not disclosed; it is a black box system. Nevertheless, the latest Fingerprint Verification Competition (FVC), FVC-onGoing, shows clearly that fingerprint algorithms developed in academia do not reach the performance or accuracy of commercial products [93]. Furthermore, we are aware of the trade-off involving the use of VF. On the one hand, it might limit some results' reproducibility and make others difficult to explain. On the other hand it enables us to test state-of-the-art fingerprint algorithms and take full advantage of the latest developments without the need to reinvent the wheel (or redevelop and reimplement the algorithms). Also,

because it is an SDK, it can be suited to our existing frameworks and needs. In our opinion, the advantages outweigh the disadvantages.

The VF is used in Chapter 4 to derive quality estimates for fingermarks (again: to decide if those marks may be of use in a police investigation or not) and to compare those marks against reference fingerprints. In Chapters 6 and 7 it is applied (after multiple adjustments) to compare infant ballprints.

4 Adult fingermarks

This chapter studies image quality metrics for fingermarks and their implications in the context of a forensic crime scene investigation by police experts. For this purpose, we limit the application to adult fingermarks and disregard the additional challenges faced by infant biometrics such as the physical tininess of the biometric (cf. Section 2.4). More specifically, we investigate if it is possible to determine (automatically) if a fingermark (more precisely: its digital capture) is of evidential value (i.e. may become useful in an ongoing investigation for the identification or exclusion of a specific individual) or if its quality (quantity of relevant information present) is too low to be of use in an investigation conducted by police officers and forensic experts. It is a simplified scenario because we don't consider a mark's relevance.

The high-level goal is to determine a mark's quality automatically to reduce workload and streamline the capture process by entrusting non-experts with the mark selection and collection.

In this thesis, we do not focus on the final application but the underlying quality feature sets upon which it might be built. Nevertheless with the high-level application in mind, the non-experts would rely entirely on technology due to their lack of training and experience. This puts the algorithm employed in charge of the decision if a given mark is of value in a further investigation and is to be collected or not. Therefore, we investigate the worst case scenario. Is it possible that a mark, which has been flagged by the algorithm to be of no evidential value for the further investigation, still can be used to identify its donor in an automatic identification setup?

The results have been published in [74, 77]. An extended version of this chapter has been submitted to a journal [75].

4.1 Background

Fingermarks are an invaluable form of evidence (or intelligence) that can exclude or identify suspects. Modern law enforcement agencies rely heavily on using fingermarks and automatic systems such as IAFIS (cf. Section 2.2) to assist forensic human experts [97]. Human examiners are expected to follow the *Analysis, Comparison, Evaluation, and Verification* (ACE-V) protocol [8]. During the analysis phase, they decide if the mark at hand is of *Value for Individualisation* (VID), *Value for Exclusion Only* (VEO) or of *No Value* (NV). For example, a mark may be VID if sufficient Level 2 details are captured and the basic main ridge flow (Level 1) can be clearly identified. Without Level 2 features a fingermark might be only suitable for VEO because Level 1 is definitely not distinct and hence does not allow for an individualisation. The mark may be of NV if even the major ridge flow pattern cannot be clearly identified. This is only a basic example relating to the standard details mentioned in Chapter 2.2. Certainly, a forensic fingerprint expert might decide differently according to whatever details are visible based on experience, training and potential court eligibility.

Fingermarks often suffer from low quality (small amount of relevant information present, low degree of distinctiveness) due to being smudged or partial, or if they overlap with other marks [38]. Often too they are distorted by the surface pattern of the object upon which they are found [130]. Their forensic value is difficult to judge for non-experts.

Ulery *et al.* show that accuracy and repeatability varies even for forensic experts and mostly depends on the print quality [146, 147], especially for borderline decisions. Consequently, Kellman *et al.* use image features to predict "expert performance and subjective assessment of difficulty in fingerprint comparisons" [70].

Most quality measures are used to prevent low-quality images from being automatically matched because they tend to produce false minutiae and hence false matches [3, 112]. Here we use the term low-quality images but we actually refer to a two-level problem: the amount of information present in the mark and the actual image quality which can reduce the amount of information even further.

The majority of quality measures are suited and optimised for their employment within the operational setups of modern law enforcement agencies. This means that they are often designed particularly for contact fingerprint scanners [23, 43, 87, 139]. Most scanners still possess a fixed capture resolution of 500 ppi. Therefore, algorithms computing quality measures or features are tuned towards this particular resolution [139]. Fingermarks require a different capture process because they are not taken from a subject. They are (distorted) traces left behind accidentally by a subject on a

physical object.

In conclusion, fingermarks require robust methods to estimate their quality because all factors mentioned above will vary and influence the quality and its estimate.

4.1.1 Potential future application and context

The Victoria Police approached us with the request to research if the automatic decision, if a fingermark is of evidential value based on its photograph taken, could be performed with a smartphone. Victoria Police field officers are issued "intelligent" mobile phones which would make these phones a cost-efficient and sufficient solution. The intended future purpose is to streamline the capture process by entrusting non-experts with the mark selection and collection at high volume, non major-crime scenes. These non-experts would replace highly trained specialists such as fingerprint examiners and make them available for other tasks that require their knowledge and expertise.

Here, we limit ourselves to fingermarks captured with a mobile phone (and the comparison to established capture methods such as a flatbed scanner) and to potential (technical) restrictions faced rather than building and testing a smartphone application in the field.

We focus on the feature sets and the general possibility to derive evidential value from a mark's image in the context of a potential future application, and the consequences or dangers of entrusting an automatic algorithm with the decision.

In the proposed scenario, when fingermarks are captured with a mobile phone, the capture resolution is not fixed or known. Firstly, the phone is hand-held by the operative, most likely at varying distances for each capture. Secondly, the phone model and make can differ and so can the camera module and its resolution. In order to ensure consistent results and, more importantly, to use already existing quality features and algorithms, the approach has to be capture resolution independent.

One way to achieve this independence is to infer the capture resolution (this process is later referred to as *Capture Resolution Estimation* (CRE)) and to scale the image based on this estimate to a standard resolution in order to use existing quality features and algorithms. Other solutions might capture a reference object with known size next to the fingermark or to use the camera's autofocus information such as the distance of the point it is focused on. We decided to base the CRE on image features to preserve the algorithm's generality, transferability and device independence. Also, to the best of our knowledge, there is no smartphone that actually implements a function that

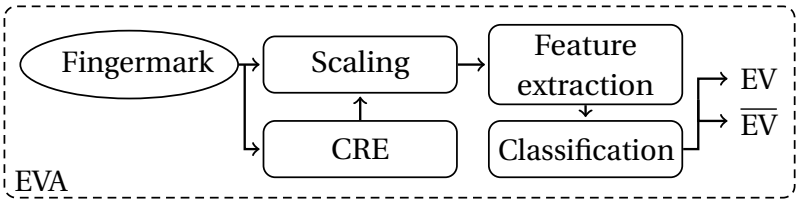


Figure 4.1: Diagram of the *Evidential Value Algorithm* (EVA). A fingermark is captured with an imaging device and scaled based on the *Capture Resolution Estimation* (CRE) to 500 ppi, image features are extracted and classified; resulting in the binary decision if this mark is of sufficient *Evidential Value* (EV) or not ($\overline{EV}$).

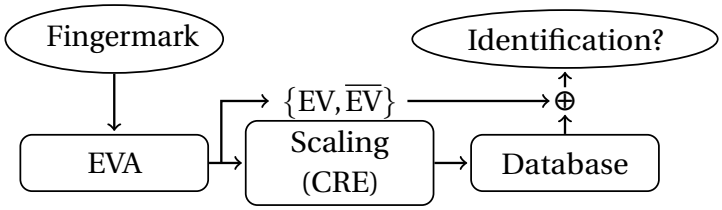


Figure 4.2: Diagram of the experiment performed. A fingermark is captured, its evidential value $\{\text{EV}, \overline{\text{EV}}\}$ is estimated by EVA and scaled in the same way as in the algorithm. The number of *correct and confident identification* (ccID) of the mark compared against a reference database is measured with respect to the image capture device and evidential value estimation method.

returns the autofocus' distance. The proposed framework is shown in Figure 4.1.

There is the need to reject low-quality fingermarks to avoid false identifications. Usually this happens during the comparison stage. The proposed (future) scenario however, takes place much earlier in the process: during the gathering of evidence. It includes the danger of missing potentially valuable marks that could solve a criminal case. Naturally there is a trade-off between the likelihood of missing important marks and capturing fewer marks to reduce workload.

Therefore the following question arises. Are there any marks that can be automatically and with confidence (enforcing a minimum margin to the second best match) identified against a reference database but are not of EV according to either the algorithm or the expert assessment? The proposed experiment is shown in Figure 4.2.

This two-step experiment aims to answer two questions. Firstly, is it possible to determine if a given fingermark is potentially of evidential value based on its image quality features? And secondly, may marks of no evidential value be suitable for comparison against a reference database anyway?

4.1.2 Chapter outline

First, we elaborate on the fingermark database's structure and its ground truth (Section 4.2). Then we introduce our capture resolution estimation algorithm RLAPS (Section 4.3) and the concept of correct and confident identification (Section 4.4). Afterwards we set up the experiments' methodology (Section 4.5), present the results obtained (Section 4.6) and discuss our findings (Section 4.7).

4.2 Database ground truth

We use a fingermark and a reference database. Both have already been introduced in Section 2.5.3. Now we elaborate more on the different deliberate distortions and on the assessment performed by three police fingerprint experts on the fingermark database.

Two males (subjects 2 and 4) and two females (subjects 1 and 3) volunteered to create a large database of pseudo fingermarks, including normal and deliberately distorted marks (1,428 in total). Six categories of deliberate distortion, each containing 168 marks, were defined: (i) finger placed "lightly" on the page, (ii) mark smeared, (iii) finger twisted lightly, (iv) finger twisted strongly, (v) finger placed "heavily" on the page, and (vi) part of the finger placed "heavily" on the page (cf. Figure 2.13 in Chapter 2, Figure 4.3 and Table 4.1).

Type of distortion	Number of marks taken	Prints of sufficient evidential value				
		Assessor 1	Assessor 2	Assessor 3	Ground truth	EVA
(i) light placement	168	48.2%	48.2%	48.2%	48.2%	54.2%
(ii) smeared	168	3.6%	4.2%	3.6%	3.6%	14.9%
(iii) finger twisted lightly	168	4.2%	4.8%	4.8%	4.8%	11.3%
(iv) strong twist	168	0.0%	0.0%	0.0%	0.0%	6.0%
(v) heavy placement	168	69.6%	65.5%	65.5%	65.5%	64.9%
(vi) partial, heavy placement	168	45.8%	48.2%	48.2%	48.2%	50.6%
(vii) normal	420	47.4%	49.0%	50.0%	49.0%	50.7%
Total	1,428	34.1%	34.5%	34.7%	34.5%	38.66%

Table 4.1: A breakdown of the 1,428 marks into the categories of distortion (including no deliberate distortion), and the final status of the assessment of the 3 experts in terms of the proportion of marks found to be of EV. Assessor 1's opinion regarding the marks of categories (iii) and (vii) are respectively 9 and 21 decisions short of the total number. However, the other assessors agree on those marks and therefore a clear decision on ground truth can be made via majority vote. The EV distribution for EVA has been calculated for the mobile phone images which have been scaled using CRE Global and the fused quality feature set at the decision threshold corresponding to the EER because of its performance (cf. Figure 4.5a to 4.5c).

Three Victoria Police experts individually assessed the laminated marks and decided

for each if it is of sufficient evidential value (to be more specific: VID at the analysis stage of ACE-V). The overall agreement between the assessors is high; they reached the same conclusion in 97.8% (assessors 1&2), 95.7% (assessors 1&3) and 99.4% (assessors 2&3) of the marks.

The overall ground truth is obtained via a majority vote; a certain mark is considered to be of sufficient evidential value if and only if at least two assessors agreed that it is indeed of evidential value. According to this methodology, 34.5% of the 1,428 marks are of sufficient evidential value. The exact EV distribution is shown exemplarily in Figure 4.5 and summarily in Table 4.1; all diagrams can be found in the Appendix pp. 176.

4.3 Capture resolution estimation

As already discussed in Section 4.1, the reason to estimate the capture resolution is that most algorithms are tuned towards the specific resolution of 500 ppi and can only handle slight variations sufficiently. Hence, it becomes necessary to scale the image acquired in a unconstrained setup so that it satisfies the resolution requirements.

The main idea behind our CRE algorithm is to measure the inter-ridge spacing (IRS, introduced in Section 2.2) and to scale the image based on the ratio of the measured value to a known reference. To do this, we make two major assumptions regarding the captured image, its quality and fingermarks: first, we presume that an adult fingerprint has been captured and infer an average IRS of $irs' = 9\,\text{px}$ at $c' = 500\,\text{ppi}$; commercial products make similar assumptions [63]. Second, this implies also that there must be visible ridge pattern present.

Our algorithm is very similar to other approaches using the radially averaged power spectrum to either estimate a fingerprint's quality [139] or its inter-ridge spacing [97]. The difference is that it uses only a part of the spectrum, the *Radially Limited Averaged Power Spectrum* (RLAPS). Furthermore, it takes into account that the IRS varies depending on where and how it is determined. The IRS decreases around singular points [97] so it is often measured in areas where none are present [1]. Hence, we choose the spectral peak corresponding to the highest frequency but with an amplitude close to the maximum.

Fingerprint distortions lead to an uneven and non-circular energy distribution of certain fingerprint characteristics at singular points. Because of its radial limitation around the strongest frequency peak, RLAPS features increase robustness against this. It results in accurate estimates and little variation (see Section 4.5.1).

The algorithm consists of three main stages: (i) the computation of the power spectrum and its filtered version, (ii) the computation of the radially averaged power spectrum for a limited area according to the highest energy distribution and (iii) finding the "last large peak" and converting the inter-ridge spacing to a CRE.

We assume that a gray-value image $\boldsymbol{I}$ has the horizontal and vertical resolution of M px and N px, respectively. The pixel at the position (x, y) is denoted by $I_{(x,y)}$ and accesses its assigned (gray) value where $I_{(x,y)} \in \{1,\ldots,255\}$, $x \in \{1,\ldots,M\}$ and $y \in \{1,\ldots,N\}$.

The image is 2 D-DF transformed into the frequency domain, maintaining the same size, and its quadrants are re-arranged so that the zero-frequency component is located at the centre, resulting in $\boldsymbol{F}$. The power spectrum $\boldsymbol{A}$ is computed from the complex modulus of the re-arranged spectrum; $\boldsymbol{A} = |\boldsymbol{F}|^2$. There might be straight lines contained within the spectrum due to spectral distortions. The Hough transform is employed to find and mask those; additionally $\boldsymbol{A}$ is smoothed by convolving it with a square matrix ($w \times w$) consisting only of ones, resulting in $\boldsymbol{A}'$.

The values of $\boldsymbol{A}$ and $\boldsymbol{A}'$ are represented by their polar coordinates ρ and θ rather than their Cartesian coordinates x and y. Finally, the maximum energy peak in $\boldsymbol{A}'$ is located with respect to the constraints that $\theta \geq 0$ (disregarding the third and fourth quadrants due to duplicity) and $r_{min} \leq \rho \leq r_{max}$ (limiting the search range and ignoring the energy peak located at the centre). The corresponding angle is stored in α.

$$\alpha = \operatorname{argmax} \boldsymbol{A}'_{(\theta,\rho)},\ \theta > 0,\ r_{max} \geq \rho \geq r_{min},\ r_{max} = \lfloor 1/2 \min(N, M) \rfloor,\ r_{min} = \lfloor \gamma\, r_{max} \rfloor \tag{4.1}$$

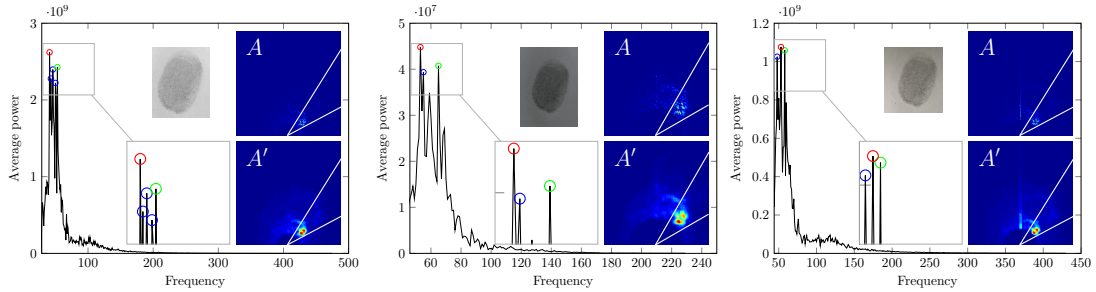
Next, the RLAPS is obtained from $\boldsymbol{A}$ by averaging with respect to the energy values over the angle $\theta = [\alpha - \beta, \alpha + \beta]$ for $\rho = \{r_{min}, \ldots, r_{max}\}$ and results in the vector $\bar{\boldsymbol{J}}$ containing l elements where $l = r_{max} - r_{min} + 1$.

The factor γ and the angle β need to be chosen roughly depending on the expected frequency range and distortion strength, respectively (cf. Section 4.5.1).

We define that all high peaks have at least a certain fraction of the global maximum:

$$\lambda \operatorname{argmax} \bar{\boldsymbol{J}}_{(i)},\ i = \{1,\ldots,l\}. \tag{4.2}$$

The candidate with the highest frequency and hence the greatest index (say r') is chosen and adjusted to $r = r_{min} - 1 + r'$. Therefore the dominant IRS contained in the image is $irs = \frac{2 r_{max}}{r}$. Examples are illustrated in Figure 4.3.



(a) Scanner, $irs = 17.92\,(19.1)$ px (b) DSLR, $irs = 7.05\,(7.3)$ px (c) Phone, $irs = 14.83\,(14.5)$ px

Figure 4.3: The radially limited averaged power spectrum (RLAPS, black curve) for the same mark captured by (a) a scanner, (b) a high-quality camera and (c) a mobile phone as well as the corresponding algorithm (and manual) inter-ridge spacing estimates. The ridge pattern introduces a local increase in power, the last high peak (green circle) indicates the ridge frequency; other individual peaks are highlighted by blue circles and the maximum by a red one. RLAPS is calculated from the spectral part enclosed by the white borders. Lines contained within the spectrum do not have been masked from the images for demonstration purposes but are masked during the calculation.

Lastly,
values for adult IRS irs' at a given resolution c' as previously discussed. This leads to the equation

$$c = \frac{c'}{irs'} irs = \frac{500}{9} \frac{2 r_{max}}{r}. \tag{4.3}$$

There are two kind of uncertainties involved. Firstly, the assumed average IRS of 9 px does not account for individual differences due to e.g. gender, ethnicity or age (cf. Figure 4.4) and secondly, measurement inaccuracies.

4.4 Correct and confident identification

Finally, we would like to introduce the concept of *correct and confident identification* (ccID). It requires a minimum margin (in terms of similarity to the query mark) between the best and the second best match. This concept is also known as distance ratio. It is used for example in the *Scale-Invariant Feature Transform* (SIFT) algorithm [91]. In this particular application of object recognition, the approach "eliminates 90% of the false matches while discarding less than 5% of the correct matches" [92]. This makes it suitable for decisions where a certain amount of confidence is required, such as when individualising the individual belonging to a fingermark, so as to avoid another Mayfield case (cf. Section 2.2 in Chapter 2).

Assuming that a fingermark is compared against a reference database containing L individual fingerprints, a verification score S_i ($i \in \{1,\ldots,L\}$) is returned for every comparison. We define a decision as *correct and confident* (ccID) if and only if the mark and the print with the highest score S_a are from the same subject *and* if the largest score is larger by factor $d > 1$ than any other score:

$$\nexists S_b : S_a \leq d S_b, \;\; b, a \in \{1,\ldots,L\}, \;\; b \neq a. \tag{4.4}$$

This would lead to a correct and confident identification of fingermark a.

The choice of parameter d is crucial and involves a trade-off. If it is chosen too small, one accepts potentially questionable matches as being genuine; if d is chosen too large, one might reject a correct matches which would be valuable in the ongoing investigation.

4.5 Experiments

This section lays out the methodologies used for three experiments. The first experiment demonstrates for the RLAPS algorithm its handling of various distortions of fingermarks (Section 4.5.1). The second investigates EVA's fingermark value prediction performance for various parameters. They include NFIQ2 image features, the features of a commercial fingerprint SDK (Neurotechnology Verifinger) and their fusion with respect to the use of different CREs and image capture devices (Section 4.5.2). The third experiment is set up for the analysis of EVA's influence in an identification scenario enforcing a minimum margin to the second best match. It asks: What consequences do arise if, for marks without evidential value, no comparison is carried out and the marks are disregarded (Section 4.5.3)? Are any confident genuine matches missed?

4.5.1 Capture resolution estimation

In order to demonstrate RLAPS behaviour for fingermarks which vary in pressure and distortion, we apply it to the pseudo fingermark database.

RLAPS has been applied to all images using the following parameters experimentally obtained: $\beta = 15°$, $w = 7$, $\gamma = 0.06$ (scanner), $\gamma = 0.17$ (DSLR), $\gamma = 0.8$ (phone) and $\lambda = 0.8$.

We have manually approximated a *Global* capture resolution of 1200 ppi (scanner), 460 ppi (high-quality camera) and 890 ppi (mobile phone) in order to obtain reference values the estimates can be compared against. Those estimates are based on the

average of 20 randomly chosen images (only for high-quality camera and mobile phone because the scanner's resolution is known a priori) for which we compared the pixel size of the mark to the one of the image scanned as it was captured at the known capture resolution of 1200 ppi.

4.5.2 Fingermark evidential value prediction

The evidential value prediction is subject to two major variables: (i) the image quality feature set and subsequently the CRE method used as it influences the feature extraction and (ii) the image capture device.

We have implemented the NFIQ2 features according to their preliminary specification [139] (NFIQ2) and employ additionally Neurotechnology Verifinger 6.7 [114] and its quality value as well as the number of minutiae found (Verifinger) and all features together (Fusion). In the last case, all features are simply concatenated to form one vector. For more details regarding the different feature sets, refer to Section 3.3 in Chapter 3.

The methodology is set up to evaluate how these image quality feature sets perform for estimating if a certain fingermark is of sufficient evidential value. Different capture devices and their interplay with our CRE algorithm RLAPS (cf. Figure 4.1) are taken into account.

Database partition

The database creation and the ground truth acquisition have been discussed earlier as well as in Section 4.2. In order to run unbiased experiments, the database D consisting of 1,428 marks needs to be partitioned into a training set T_q, a validation set V_q and a test set S_q. We use a q-fold cross-validation (also known as k-fold cross-validation) with $Q = 5$, $q \in \{1,\ldots,Q\}$, following McLachlan *et al.* [100, p. 214]. Hence we perform the partition $Q = 5$ times. We choose all sets randomly but within two main constraints: firstly, the union of all test sets must be the database and the pairwise intersection of the test sets must be the empty set.

$$D \supset S_q,\ \forall q,q'\ S_q \cap S_{q'} = \varnothing,\ D = S_1 \cup S_2 \ldots \cup S_Q \text{ with } q,q' \in \{1,\ldots,Q\}, q \neq q'. \tag{4.5}$$

Secondly, the number of images per set is roughly 60% (training), 20% (validation) and 20% (test). All images are used and there is no overlap between the sets during

each fold.

$$D \supset T_q,\ D \supset V_q,\ D = T_q \cup V_q \cup S_q,\ T_q \cap S_q = \varnothing,\ V_q \cap S_q = \varnothing,\ T_q \cap V_q = \varnothing. \tag{4.6}$$

Experiment

This experiment is performed for three image quality feature sets (NFIQ2, Verifinger, Fusion) subject to different CREs (None, Global, RLAPS) and three capture devices (Scanner, DSLR, Phone). All images are scaled by the factor $500/c$ (see Section 4.1 for the underlying reasoning and cf. Equation 4.3 for the derivation) using the nearest-neighbour interpolation, if applicable.

All feature sets were computed for all images of all capture devices with three different CREs performed: none at all (None) and either global (Global) or individual (RLAPS) estimates. The features obtained are classified by a Support Vector Machine (SVM), the Discriminant Analysis (DA) and the k-nearest neighbours algorithm (kNN) trained according to the database partition and different parameter sets (cf. Figure 4.1).

For each q (Q-fold iteration) the classifier of choice $C \in \{\text{SVM}, \text{kNN}, \text{DA}\}$ is trained on T_q with different sets of parameters $P_i \in \boldsymbol{P}$ and then applied to the validation set V_q. Then the results are scored using the true positive rate for a fixed False Match Rate of 0.01 (TPR@FMR100) and averaged over q and the capture method (scanner, high-quality camera, mobile phone) in order to find the parameter set P_i that leads on average to the best performance.

Now each classifier, trained on T_q and using the parameter set P_i, is applied to each test set S_q individually, resulting in S'_q. The final score is calculated on the concatenation of all test set results $\{S'_1, \ldots, S'_Q\}$. The constraints ensure that there is only one result for every mark captured.

There are two major remarks regarding the experimental procedure to keep in mind. Firstly, we are basically evaluating Q different classifiers instead of one because the classifier is trained differently during each fold due to the different training sets T_q, despite using the same set of parameters. Secondly, this methodology uses a relatively small training set and large test set when compared to other approaches (e.g. leave-one-out cross-validation). Therefore we are establishing a lower performance boundary.

4.5.3 Fingermark verification and identification

This experiment aims to investigate the relationship between a fingermark, which can be automatically matched correctly with a high confidence, and the evidential value assigned to it by experts or EVA (cf. Figure 4.2).

The EVA is predominantly influenced by (i) the image properties such as capture device used and CRE and thereby (ii) the quality features extracted as briefly discussed in Section 4.3. In this context, we obtain the estimated evidential values for three different CREs (None, Global, RLAPS) and three capture devices (Scanner, DSLR, Phone) for all 1,428 fingermarks.

Afterwards, a verification score for every fingermark compared against the reference database is computed. We limited the genuine reference prints to the ones of the central finger tip area captured by the Digital Persona U.are.U 4000 (40 genuine prints in total). This verification process is performed by Neurotechnology Verifinger 7.0. We consider it to be correct and confident identification (ccID) if and only if the highest score returned for one particular fingermark is from the comparison to the same subject's finger and the second highest score multiplied by $d = 1.5$ is still smaller than the highest one (see Equation 4.4). We chose d experimentally, so that it delivers a compromise between rejecting questionable matches while retaining clear ones.

Additionally, we check if the ccID fingermarks are considered to be of EV by either the experts or EVA. In case of the algorithm, the decision threshold for EV corresponding to the EER has been chosen.

The step to determine if a fingermark is of EV is performed first and subject to capture device and CRE. Therefore, the verification scores used for identification are calculated afterwards on the already scaled image (see Figure 4.2). Verifinger failed to compare 20 query fingermarks to the database because of their image resolution; this was only the case for unscaled scanner images. This can be classified as Failure to Enroll. Nevertheless the prints were not identified correctly and with confidence.

Finally, we applied different decision thresholds (instead of just the one corresponding to the EER) to the evidential value raw scores to observe if favouring a certain error type (moving the operating point along the ROC) influences EVA's decision.

4.6 Results

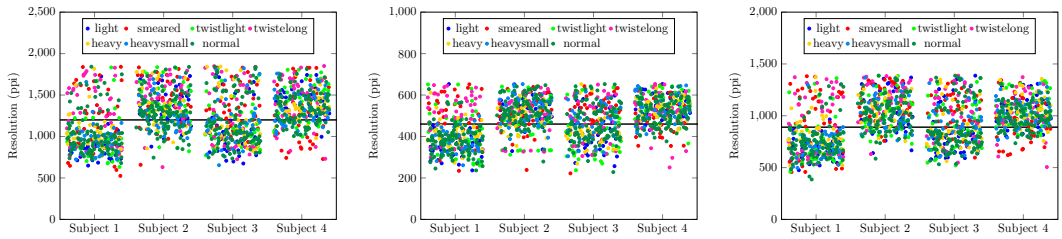
The results follow the same order in which the experiments have been described. First, we present and discuss the results of the capture resolution estimation on our

fingermark database (Section 4.6.1), then the fingermark evidential value prediction (Section 4.6.2) and finally the fingermark verification and identification (Section 4.6.3).

4.6.1 Capture resolution estimation

It is well known in the literature, that the inter-ridge spacing (IRS) for adults differs between the sexes significantly [1, 49, 116, 136]. Depending on the population, men tend to have greater IRS compared to females and it is possible to estimate a subject's gender effectively just from the IRS [110] (see also Section 2.2 in Chapter 2).

However, the ratios calculated for studies found in the literature are 1.1957 (Caucasian [1]), 1.1569 (African American [1]), 1.1406 and 1.1252 (Indian [136] and [116], respectively), and 1.1035 (Spanish Caucasian [49]). Again, all these ratios show some variation depending on the sample population and methodology used. Hence, when we measure the ratio for images obtained with different devices but for the same fingermarks, it should be consistent because the methodology and population do not change.



(a) Scanner (b) DSLR (c) Phone

Figure 4.4: The capture resolution estimate is plotted for each individual subject and capture device to highlight the intra-class variation; the different distortion types are colour encoded and have been spread horizontally for visualisation purposes. The manually estimated capture resolutions for each device are indicated by the horizontal black line; (a) scanner: 1200 ppi, (b) high-quality camera: 460 ppi, (c) mobile phone: 890 ppi. As the algorithm uses a generalised scaling factor, the individual subject resolutions differ.

Its results are visualised in Figure 4.4 with respect to the capture device, individual subject and type of distortion.

The CRE distribution (see Figure 4.4) shows a significant difference related to the known different IRS between the sexes [1].

Therefore, the consistency of the CRE or IRS ratio (cf. Equation 4.3) between males

Measure	Capture device		
	Scanner	DSLR	Phone
Assessed	1200.0 ppi	460.0 ppi	890.0 ppi
Mean	1240.0 ppi	477.9 ppi	946.9 ppi
SD	297.9 ppi	96.9 ppi	219.5 ppi
♂/♀	1.2115	1.1983	1.2213

Table 4.2: The average CRE, its standard deviation and the ratio of the average male and female CRE for different capture devices as measured by RLAPS. Furthermore we include our manual assessment of the capture resolution for reference purposes. See Figure 4.4 for the estimate distribution.

and females across different devices (cf. Table 4.2) indicates the algorithm's stability on a wide range of different images.

4.6.2 Fingermark evidential value prediction

The EV experiments indicate (cf. Table 4.3 and Figure 4.5a to 4.5c) that it is possible to determine if a query mark is of sufficient evidential value based on its image quality features as long as the capture resolution is normalised to 500 ppi. In the case that this is unknown, it can be estimated by the CRE algorithm introduced in Section 4.3 without any significant performance setback. Also, the algorithm's accuracy is not overly important but its consistency is. NFIQ2 performs surprisingly well for unadjusted images; however this is only true for DSLR captures when using Verifinger because their capture resolution is close to 500 ppi.

The NFIQ2 feature set performs better than the Verifinger features but their Fusion shows the most promising and consistent results.

The verification experiment shows a strong correlation between the automatically estimated evidential value and if a confident identification is possible to be performed for a particular fingermark. This is partially due to the setup used because both the score computation and EV estimation are based on image features.

Further limitations of the matching system used became evident. Verifinger is very resolution dependent and requires marks or prints to be in a very narrow capture resolution window (around 500 ppi) with as little variation as possible to perform properly. This is the reason that a global scaling factor and the high-quality camera images without any scaling (their resolution is approximately 460 ppi) work well.

Feature set	CRE	Capture device					
		Scanner		DSLR		Phone	
NFIQ2	None	15.65%	(kNN)	16.24%	(kNN)	16.56%	(SVM)
	Global	11.75%	(kNN)	**15.17%**	(DA)	**14.23%**	(SVM)
	RLAPS	**11.18%**	(DA)	16.13%	(DA)	15.60%	(SVM)
Verifinger	None	32.10%	(kNN)	17.07%	(SVM)	30.75%	(kNN)
	Global	**13.25%**	(SVM)	**15.85%**	(SVM)	**14.43%**	(SVM)
	RLAPS	13.82%	(SVM)	16.56%	(kNN)	15.87%	(kNN)
Fusion	None	13.35%	(SVM)	13.01%	(kNN)	16.46%	(SVM)
	Global	**8.54%**	(DA)	**12.20%**	(SVM)	**13.62%**	(DA)
	RLAPS	9.29%	(DA)	13.03%	(SVM)	14.21%	(SVM)

Table 4.3: EER with respect to capture device (Scanner, DSLR, Phone), quality feature set (NFIQ2, Verifinger, their fusion) and CRE algorithm (None, Global, RLAPS) for the classifier achieving the best performance.

Nevertheless, the image quality due to the use of different capture devices is not a major drawback. The mobile phone performs more strongly than the DSLR but falls shy of the scanner, under the condition that the capture resolution is adjusted properly. The difference between the quality feature sets with scaling is rather small but should be considered in a real world framework.

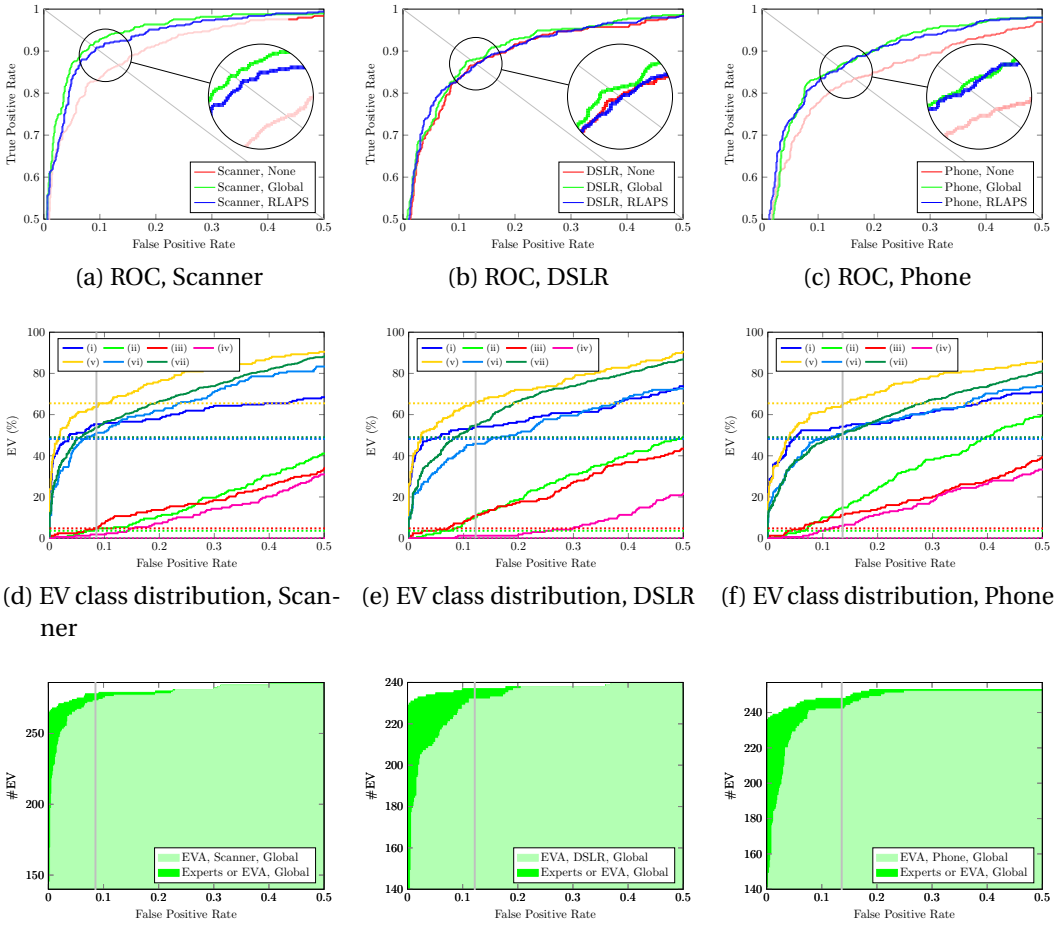
One has to keep in mind that firstly, the fingermarks vary in quality without any strong background patterns or surface distortions, and secondly, that the image quality feature algorithms used have all been developed for images acquired by contact scanners. Furthermore, we are aware of the database's limited size and that inter-age and inter-racial cross validation are needed to draw any further conclusions.

4.6.3 Fingermark verification and identification

The results are shown in Figure 4.5 and Table 4.4; the complete set of figures can be found in the Appendix A.4. These are corrected from [77] following recalculation after a coding bug was fixed.

However, there are marks that though they can be automatically identified with confidence, they do not have sufficient evidential value according to the experts' assessment. Again, this might be due to experimental setup that heavily favours image processing algorithms or to the limited size of the test population and database. Additionally, it is worth noting that some of the identified fingermarks are only considered to be of evidential value by one of the experts or the algorithm.

Encouragingly, only *six* of the 257 marks being identified with confidence and having



(a) ROC, Scanner (b) ROC, DSLR (c) ROC, Phone

(d) EV class distribution, Scanner (e) EV class distribution, DSLR (f) EV class distribution, Phone

(g) ccID and EV, Scanner (h) ccID and EV, DSLR (i) ccID and EV, Phone

Figure 4.5: The first row ((a) to (c)) shows the top left corner of the Receiver Operating Characteristics (ROCs) for all capture devices calculated on the Fusion image quality feature set with global scaling. The second row ((d) to (f)) shows the EV distribution according to the mark's distortion class which has been computed by EVA (solid) and the experts (dashed). The third row ((g) to (i)) shows the number of ccIDs classified as EV by EVA (light colour), the additional ones by the experts (dark colour) and the ones not classified as EV by either (white). The gray line is the threshold corresponding to the EER when the operating point moves along the ROC.

Evidential Value according to the experts were missed by our algorithm in a mobile phone scenario with Global scaling and the Fusion feature set.

Table 4.1 and Figure 4.5b also indicate that EVA works rather conservatively and tends to flag a fingermark as being of sufficient evidential value slightly more often than an expert who applies other considerations (such as court eligibility) than just image quality. Nevertheless, an expert's accuracy and repeatability can vary mostly due to

	CRE	Capture device		
		Scanner	DSLR	Phone
ccID	None	1	244	10
	Global	286	240	257
	RLAPS	115	122	119
Experts	Global	263	228	236
EVA_{NFIQ2}	Global	268	228	246
$EVA_{Verifinger}$		269	224	242
EVA_{Fusion}		273	232	242

Table 4.4: Number of fingermarks which have been correctly and with confidence identified (ccID). Numbers of these classified by experts or EVA to be of sufficient evidential value (EV) with respect to capture device (Scanner, DSLR, Phone), CRE Global and quality feature set (NFIQ2, Verifinger, Fusion) if applicable. EVA uses the threshold corresponding to the EER; Figure 4.5 visualises the effect of using other thresholds. The EV results for the CREs None and RLAPS are not reported separately due to their much smaller numbers compared to Global (cf. columns reporting ccID). The Figures A.4, A.5 and A.6 in Appendix A illustrate this in more detail.

the print quality [146, 147] regarding borderline decisions.

Our results underscore the importance of capturing all fingermarks of sufficient evidential value in the field. They might not be suitable for individualisation but still have value for exclusion.

4.7 Conclusion

The experiments performed demonstrate that it is possible to use existing or proposed fingerprint quality features to determine automatically if a fingermark is of sufficient evidential value. This applies regardless of the image capture device; even a mobile phone can be used. Already today, the mobile phone's image quality is not a limiting constraint. As technology develops, mobile phone cameras will improve further and should be able to deliver consistently high image quality even under more difficult circumstances than we have evaluated in this case study.

The results are encouraging, particularly as they establish a lower performance boundary. One major drawback is that the image capture resolution needs to be estimated in order to be scaled to 500 ppi for optimal performance as most fingerprint image quality features are optimised for this particular resolution. However, the experiment emphasises that a rough approximation is reasonable if no setup with a known resolution and fixed distance is available.

Also we found a strong correlation between the fact that a fingermark can be automatically matched correctly with confidence and its inferred evidential value. Therefore, it is sensible to run fingermarks with sufficient evidential value against a reference database to potentially obtain a correct match.

In this chapter, we have introduced two algorithms, one to estimate the capture resolution of a friction ridge skin image (RLAPS) and one to distinguish for said image if it is of evidential value (EVA) and hence of any use for an ongoing investigation. Both algorithms RLAPS and EVA have been developed for the use of adult fingermarks but are based on image features and therefore age independent. In Chapters 6 and 7, we will use RLAPS to estimate the inter-ridge spacing of infant ballprints and EVA to derive a quality estimate for these ballprint images, via their inferred evidential value based on adult fingermarks.

5 Infant biometrics

This chapter is about infant biometrics. It motivates the necessity of these, reviews the literature and investigates the newborn and infant footprint for this purpose.

The urgent need for a reliable infant biometric and its possible applications have been discussed in Section 1.1 in detail. Most cases are centred on the identification or individualisation of an infant. Reasons may include (but are not limited to) the fight against human child trafficking, the prevention of newborn mix-ups in hospitals, proof of citizenship (and the rights associated with it) and access to medical records.

Firstly, all of these applications require the biometric to be persistent over time (permanence) in order to be able to perform an identification by comparison to a biometric reference template collected at an earlier age. The trait's permanence is challenged by the infant's rapid growth and fast physical development.

Secondly, an infant biometric requires acceptability from a parent's point of view. To create a template that can be used as a reference for an infant's identification it is necessary to be able to capture the biometric feature satisfactorily. There are multiple aspects involved. The biometric should be easily captured by untrained personnel without having to apply any force. This is an issue with using palmprint or fingerprint for newborn registration. Usually newborns have their hands curled up to a fist and are unwilling to open them. Thus one is required to open them (gently) to be able to capture the biometric. This leads to problems with acceptability. A parent could refuse to have the newborn's fingerprints taken if it requires any force and potentially

The introduction and Section 5.1 have been published in [71]; all other sections except the automatic experiment have been published in [72]. The entire chapter has been submitted to a journal [76].

could harm the newborn or even just make it feel uncomfortable. This changes as the newborn grows and becomes an infant. Most parents adjust to their new role and are more relaxed about someone handling their child. Therefore, the problem changes slightly: now the infant's will to cooperate or its playfulness becomes the main challenge which can make it equally difficult to capture the biometric successfully.

Here it becomes necessary to distinguish clearly between an infant biometric and an infant biometric which also fulfils all requirements for *newborns*; or more specifically, infants under 28 days old (according to the *World Health Organization* (WHO) [161]). Work with newborns is especially challenging. First, a newborn's body is very fragile and tiny and so are all physical features. It requires on the one hand a very gentle and careful capture procedure; on the other hand, the biometric sensor equipment used has to be able to pick up the biometric signal, despite its tininess, consistently and accurately. Second, the protectiveness of a new parent must be taken into account. They will prevent any procedure, such as opening the newborn's fists in order to take a fingerprint, that might make the newborn feel uncomfortable, even if it is not the case. Third, the newborn has spent the last months in the mother's womb surrounded by amniotic fluid and is now exposed to air for the first time. The skin is adjusting to the new environment it is facing, often resulting in the newborn's skin drying and peeling. The skin adjustment makes it especially challenging to capture any physical skin features such as the structure of the friction ridge skin.

The most important application of a newborn biometric is to verify one's citizenship later in life [132] but it also can be used immediately to prevent mix-ups in hospitals. However, an infant biometric aims for identification (or individualisation) rather than verification. One example is the digital vaccination pass [14] in order to deliver the correct vaccine to the correct child, which helps avoid wasteful area-wide vaccination programs, and also minimises the child death-rate due to vaccine-preventable diseases. Another application is the identification of kidnapped children. National IDs can easily be falsified, stolen or lost, whereas a biometric that is electronically stored and globally accessible by police forces permits worldwide infant identification, even across borders.

One possible scenario is to use two different biometric traits; one as a newborn and the other one as an infant biometric if there is none that is suitable for both applications. This is a special case of a multimodal biometric (cf. Chapter 2.1). The Indian national ID program Aadhaar uses fingerprint and iris (also a pin code, a picture and a signature). Nevertheless, considering that there is not one biometric which fulfils the requirements of a newborn and an infant biometric, the switch from one to another represents a huge hurdle and would only work if the recording process allows

for a well documented hand-over or switch between the two traits. New problems and challenges may arise if the hand-over document is lost or vanishes as the connection between the two templates cannot be established any more. Therefore it is a major motivation to find a biometric modality that is applicable to both newborns and infants to avoid these challenges.

This chapter is structured as follows. In Section 5.1, we review the literature for potential newborn and infant biometrics and discuss results published in the literature with respect to the biometric modalities foot, palm, ear, face and finger and highlight the challenges being faced when trying to establish a biometric suitable for newborns. In Section 5.2 we point out the reasoning behind the choice of the footprint crease pattern and introduce a novel algorithm for crease pattern extraction in Section 5.3. In Section 5.4 we perform and analyse a manual and an automatic verification experiment in footprint crease pattern. Finally, we draw a conclusion and propose the direction for further research in Section 5.5.

5.1 Background

We reviewed the literature for different biometric modalities which may suit infant and especially newborn identification; those traits are ear, face, finger, foot and palm. The results achieved by their authors are summarised in Table 5.1. However, it is worth mentioning that only footprints and fingerprints have been explored using longitudinal datasets spanning time periods of several weeks. This is important because growth and its challenging effects can only be encountered when longitudinal data is collected. Other challenges arise due to image quality because of the structure's tininess or the subject's will to cooperate. Nevertheless, the use of a longitudinal dataset is most relevant to the biometric's planned application unless the use is only to prevent mix-ups in hospitals.

5.1.1 Footprint

Footprints of newborns are recorded by hospitals all around the world to help identify infants in the rare case of a mix-up. More recently, the lack of a reliable biometric for infants has been identified as a roadblock for improving vaccination programs in developing countries since it is often the case that the vaccination history of a child is unknown and paper-based systems typically fail [14]. There is considerable controversy over the usefulness of collecting footprints, especially for identification purposes. The literature regarding the value of infant footprinting is split into two opposing positions:

Biometric	Author	Dataset	Reported Score
Footprint	Blake [16], 1959	1,388 **newborn** prints collected directly after birth.	"79% of the original footprints could be identified [individualised] by flexure crease alone."
	Shepard *et al.* [129], 1966	51 **newborns** printed directly after birth and 5 – 6 weeks later.	10 out of 51 were manually matched by the California State Department of Justice.
	Thompson *et al.* [140], 1981	100 full-term **newborn** footprints and 20 footprints from premature infants, a second set at discharge 4 – 8 weeks later.	11% of the full-term prints were "technically acceptable" and *one case* court acceptable. No premature prints were matched correctly.
	Jia *et al.* [67], 2010	101 **newborn** feet captured within 2 days after birth.	3.82% EER (intra-visit).
	Balameenakshi and Sumanthi [10], 2013	240 images from 40 **newborns**, collected within 2 days after birth.	130 of 200 images are matched correctly.
	Lemes *et al.* [29], 2014	519 footprints from 173 **newborns** collected within 2 days after birth. The experiment is limited to high-quality prints ("good, normal and dry"), resulting in 231 footprints belonging to 77 newborns.	93.72% TAR with 0% FAR.
Palmprint	Weingaertner *et al.* [155], 2008	First and second set of prints from 106 **newborns** collected within 24 h and 48 h after birth, respectively.	On average 65% correct manual pair matching.
	Lemes *et al.* [28], 2011	1,221 palmprints from 250 **newborns** collected within 2 days after birth. Only prints from 20 newborns were usable; used were the 3 best ones, 60 prints in total.	78% correct recognition rate.
	Lemes *et al.* [29], 2014	519 palmprints from 173 **newborns** collected within 2 days after birth. The experiment is limited to high-quality prints ("good, normal and dry"), resulting in 297 palmprints belonging to 99 newborns.	91.53% True Accept Rate (TAR) with 0% FAR.
Ear	Tiwari *et al.* [141], 2011	210 **newborns**, 5 pictures per ear.	83.67% identification accuracy.
	Kumar *et al.* [80], 2015	750 images of children aged 0 – 6 years, up to 5 pictures per ear, no further details given.	87.54% AUROC; we measured an EER of approximately 20%.
Face	Bharadwaj *et al.* [12], 2010	In total 374 images from 34 **newborns** were captured in two sessions: within 2 h after birth and at discharge after 8 – 15 h.	86.9% identification accuracy.
	Tiwari and Singh [142], 2012	2,800 images in total from 280 **newborns** at the ages of 4 h and 20 h to 70 h, depending on type of birth.	87.04% identification accuracy for newborns with neutral face expression.
Fingerprint	**Gottschlich *et al.*** [47], 2011	48 criminal records of juveniles initially aged 6 – 15 and a second set of prints collected between one and many years later at the ages 17 – 34.	1.8% EER.
	Jain *et al.* [63], 2014	East Lansing data: 1,600 images from 20 infants aged 0 – 4 years collected over 5 sessions about one week apart.	83.8% (98.97%) rank-1 identification accuracy with a commercial fingerprint SDK (state-of-the-art latent fingerprint SDK).
		Benin data: 420 images from 70 infants, one session, no age group specified.	40.0% (67.14%) rank-1 identification accuracy with a commercial fingerprint SDK (state-of-the-art latent fingerprint SDK).

Table 5.1: Infant biometric results reported throughout the literature. Authors using longitudinal datasets spanning several weeks, months or even years are highlighted as growth effects become more important. All studies are working with newborns (highlighted in column three), except for the two investigations reviewing the fingerprint.

Shepard *et al.* [129] printed 51 newborns directly after birth and collected a second print after 5 – 6 weeks. The California State Department of Justice was assigned the task to match the two sets; only 10 of 51 were matched correctly. Thompson *et al.* [140] collected 100 full-term infant footprints and 20 footprints from premature infants (newborns weighing less than 1,500 gm [140]). A second set of prints was collected at discharge 4 – 8 weeks later. They claim that only 11% of the full-term prints were "technically acceptable" and only in *one case* was it conceded that identification was possible and that the footprints might be used as evidence in court. Furthermore, none of the first prints from premature infants were matched correctly with those taken at discharge.

Montgomery [105] captured footprints from 191 newborns within 1 – 7 days after birth. He was able to obtain visible ridge structure using an undisclosed technique. Based on this data, he proposed a classification system. Blake [16] collected 1,388 footprints and concluded that "79% of the original footprints could be identified [individualised] by flexure creases alone and that ridge structure was present in 85% of the impressions". Furthermore, Blake states that cases of infant abduction had been solved using footprints and the prints formed part of the evidence used to prosecute the offenders. Stapleton [134] stresses the importance of taking footprint captures at birth, because they are a fast, reliable and cost-efficient tool that allows for infant identification. He sees insufficiently trained personal as the main problem for the failure to use footprints for identification purposes. Sinclair and Fox [132] report their success for one infant-to-adult footprint individualisation intended to establish U.S. citizenship. They used the infant's footprints from her hospital birth certificate that were "completely unsuitable for comparison purposes" [132, p. 487] and managed to obtain a positive identification.

More recently, Jia *et al.* [67] collected on average 19 or 20 footprint images of 101 newborns; two people were needed, one to take a picture with a digital camera and one to hold and calm down the newborn. The authors would have liked to collect a second set of images of the same subjects but they state that due to the importance of a newborn in China, most parents declined their request. Furthermore, they emphasise the importance of choosing a convenient time to capture the images. This is when the infant sleeps or is very calm and is not crying because of hunger or moving its hands and feet around playfully. They were able to achieve an EER of 3.82% on their dataset.

Balameenakshi and Sumathi [10] collected 240 images from 40 newborns within 2 days after birth. They followed the two-person approach chosen by Jia *et al.* and confirmed the importance of the newborn's mood. The capture process becomes much easier and more reliable when the newborn sleeps or is calm and does not cry

or move its extremities. The authors state that most images were taken from calm subjects and they were able to match 130 out of 200 images correctly.

Last and most recently, Lemes *et al.* [29] collected 519 footprints of 173 newborns within the first 48 hours after birth. They used a high resolution sensor that captures the images with 2400 ppi. Their authentication experiment is limited to high-quality captures classified as good, normal or dry following Wu *et al.* [162]. This subset consists of 231 footprints of 77 individuals. The authors achieved 93.72% TAR at 0% FAR.

It is safe to say that matching infant footprints is challenging due to poorly captured prints, the ridges' tininess, and capture in the wrong period after birth. Blake [16] notes the importance of the latter because a few days after birth the skin starts to dry and crack, obscuring the ridge pattern. The flexure creases however are not affected by this effect and are claimed to "remain persistent over time" [99] based on the author's study observations.

5.1.2 Palmprint

In 2008, Weingaertner *et al.* [155] studied the possibility of using palmprint for newborn identification thoroughly. They found that commercially available off-the-shelf fingerprint or palmprint scanners were unable to pick up any biometric signal from newborns due to the ridge lines' tininess and the fragility of the skin. A flatbed scanner would capture the ridge lines but they found three disadvantages of this method: firstly, the whole process is too slow and requires the newborn to hold still during the entire capture time of two minutes; secondly, the results tend to lack contrast making the separation between ridges and valleys difficult, and thirdly, it is required to apply some pressure to the newborn's hand in order to keep it from moving, which drains the blood from the hand and decreases the contrast even further. Hence they developed a scanning device based on a high-quality camera and a glass plate angled at 45° leading to a resolution of approximately 1400 ppi. This device enabled the researchers to capture the palm at an acceptable quality. They collected palmprints from 106 newborns within the first 24 hours and again before the newborns were 48 hours old; both times two captures were taken.

Experts classified 83% of the captures as good or excellent and therefore suitable for manual individualisation, despite the fact that newborns are unwilling to open their hands. However, fingerprint experts were given the prints (one each) from both sessions of 30 randomly chosen newborns and had to find the corresponding matches. This experiment was carried out twice. The experts achieved on average 65% correct

individualisation.

In 2011, Lemes *et al.* [28] employed a commercial palmprint scanner with a resolution of 1000 ppi. They found that the scanner chosen is not suitable for newborns for two reasons: firstly, the low resolution (they estimated that 1500 ppi would be sufficient), and secondly, the fact that its ergonomics are not suitable for newborns. Nevertheless, the researchers collected 1,221 palmprints of 250 newborns within the first 48 hours after birth. The initial quality estimation brought to light that only 5% of the prints acquired were of sufficient quality to perform automatic matching. The authors found four reasons: (i) the newborns moved their hands during the capture process leading to smudged images, (ii) the pressure applied could cause parallel ridge lines to appear as one (too much pressure) or prevent them from being picked up at all (too little pressure), (iii) alcohol cleans the skin efficiently but also causes it to be dry, and (iv) water and soap do not clean the palm sufficiently enough and the skin remains too wet if not dried properly.

Due to the image quality issues, the authors were only able to compare 60 images from 20 newborns automatically with a 78% correct recognition rate.

In 2014, Lemes *et al.* [29] followed up on their previous work [28] and collected additional 519 palmprints of 173 newborns within the first 48 hours after birth. They used a high resolution sensor that captures the images with 2400 ppi (patent pending). The palmprint authentication experiment applies the same quality constraints as their footprint work. Hence the experiment uses a subset consisting of 297 palmprints of 99 individuals resulting in 91.53% TAR at 0% FAR. Also the authors extended their analysis to the dataset used in their previous work and have been able to improve those results to 98% TAR while maintaining 0% FAR. Finally, the authors stress the importance of high resolution sensors to capture the biometric features of newborns accurately.

5.1.3 Ear

The ear has been suggested as an infant biometric. It has been shown that the ear shape is distinctive and can identify newborns, identical twins or triplets, and adults [39].

Tiwari *et al.* [141] acquired 5 images per ear from 210 newborns with slight variations such as changes in illumination or partial occlusion. They emphasise that the capture process can be carried out from a distance as no contact sensor is required and a picture taken with a standard camera is sufficient due to the size of the ear. This implies increased hygiene which is an attractive reason for using the ear as a biometric

for newborns. The authors achieve 83.67% identification accuracy.

In 2015, Kumar *et al.* [80] used an undisclosed or unspecified database consisting of 750 photographs of children's ears. The children were aged between 0 and 6 years and the database contains up to five mobile phone images of the same individual. They report an AUROC of 87.54%. For a more direct comparison, we measured an EER of approximately 20% from their published results by computing the intersection between their curve and the diagonal line reaching from the top-left to the bottom-right corner of their figure of the ROC curve.

Nevertheless, there is no longitudinal investigation of ear shape and whether it changes over time as the infant grows, and if so, how.

5.1.4 Face

Face recognition is becoming more popular, even for infants and newborns. In 2010, Bharadwaj *et al.* [12] investigated the identification performance for the biometric trait face within an unconstrained setup. This allows other than neutral facial expressions, change in pose, illumination or environment. They collected a database of 374 images of 34 individuals within two sessions (10–14 images per subject). The first one was held within 2 hours after birth and the second one at discharge (at the age of 8 – 15 hours); one pair of twins is included in the dataset.

The authors note the uncooperative nature of the subjects and point out that they are rather difficult to deal with because of constant movement (motion blur) and their "refusal" to make a neutral facial expression. Nevertheless, 86.9% identification accuracy were achieved.

In 2012, Tiwari and Singh [142] created an unconstrained database consisting of 280 subjects and 10 images per individual with slight variations in pose, facial expression and illumination. The capture process was performed in two stages, the first set was taken within 4 hours after birth and the second either after 20 hours if it was a normal birth but after 70 hours in "scissoring cases" [142]. The authors highlight the challenge of waiting for the right moment when the newborn is comfortable, calm but not sleeping (open eyes) and hence not crying or exhibiting too much movement. Sometimes the capture procedure took close to 40–45 minutes just to create a comfortable environment for the newborn. Nevertheless, Tiwari and Singh attained 87.04% identification accuracy for newborns with a neutral facial expression and highlight that "this negates the notion that all newborns look alike" [142]. Nevertheless, the authors do not investigate the persistence over time but only a very short time frame

(24 hours at most). This raises the question regarding the trait's suitability as an infant biometric, especially when one considers a newborn's rapid growth and its dramatic physical changes.

5.1.5 Fingerprint

In contrast to the other biometric traits, the two studies in fingerprints in Table 5.1 deal with infants and adolescents rather than newborns and hence focus on infant growth and how it affects the ridge line pattern. Castellanos [19] states that the ridge line structures of newborns are three to five times smaller when compared to adults; Weingaertner *et al.* [155] show a newborn and an adult fingerprint next to each other (this particular figure can be found in Figure 1.1 on page 8 of this thesis) supporting the original claim visually. They add that the skin ridges "are very fragile, easily deforming upon contact". Furthermore, Weingaertner *et al.* point out that newborns usually have their hands clenched which requires them be opened in order to obtain a fingerprint capture. Hence, the usage of this biometric trait forces a high burden on the sensor technology and on the newborn.

Therefore, researchers have examined the fingerprint modality as an infant biometric rather than a newborn biometric and have focused on growth and how it affects the physical biometric features. In one study, Schneider [127] used digital scans of 308 children aged 1 – 18 years (12 were younger than 4 years and 79% were aged 7 – 10). A second set was collected from 186 of the subjects one year later. He failed to identify any linear affine transformation to approximate fingerprint growth. However, Gottschlich *et al.* [47] state that fingerprint growth can be approximated by isotropic scaling and propose the ratio of the median population height at the different ages of the specific gender as a suitable scaling factor. They used 48 criminal records of juveniles initially aged 6 – 15 and a second set of prints collected between one and many years later. The authors point out that the use of the scaling factor approximately halves the error rates, leading to an EER of 1.8% in a verification scenario for a test set of 462 fingers. It is unknown if the subjects were unwilling to cooperate during the image acquisition.

Recently, in 2014 Jain *et al.* investigated if fingerprints are suitable for toddler and infant recognition [63]. The authors collected two different databases: the East Lansing dataset consisting of 1,600 images of 20 infants aged 0 – 4 years (with the majority being older than 6 months) collected over 5 sessions about one week apart and the Benin dataset which contains 420 images from 70 infants collected in one session and no specified age group apart from the remark that most subjects were younger than 6 months of age.

Jain *et al.* point out *three* major challenges faced during data acquisition: first, the uncooperative nature of the infant, second, the skin conditions (oily or wet fingers), and third, the small size of the fingers. According to the authors, the skin condition of younger children (and hence the image quality and relevant information captured) may be influenced by "the habit of sucking their fingers". They also highlight the importance of a compact and comfortable sensor device to make the capture process as pleasant as possible. They also stress the need for fast capture speed on the one hand to minimise the burden forced upon the infant by holding its finger onto the sensor as well as to avoid smudged images due to movement.

Finally, they performed identification experiments, which lead for the East Lansing data to 83.8% and 98.97% rank-1 identification accuracy with a commercial fingerprint SDK and a state-of-the-art latent fingerprint SDK respectively. For the Benin data 40.0% and 67.14% accuracy were respectively achieved.

5.1.6 Challenges across newborn biometrics

All biometric traits have their own unique challenges but essentially they all break down to the newborn's will to cooperate, the difficulties evolving around capturing its tiny and fragile physical features and the need to please the parents and to take their protectiveness into account. Whether the biometric modality is intended to be carried over from a newborn to an infant biometric or used only for infants, the rapid growth that permanently changes all physical features is an additional challenge. Therefore, the biometric trait has to be chosen in accordance with the application, its focus and expected target group.

5.2 Footprint crease pattern

In the previous sections we have discussed the difficulties one faces when trying to implement a newborn or an infant biometric. We have decided to use the newborn's foot because it can be easily accessed, and is well-established and accepted by parents. We also expect that technology has improved over the last decades and hence less training is required to produce captures of good quality. For example, the inkless paper system requires one to wipe the skin with a special towelette and press the skin onto a chemically enhanced sheet of paper instead of spreading ink across the whole foot with a metallic cylinder and bringing it onto paper. The latter procedure is complex, messy and requires training [155].

Blake [16] pointed out that timing is most important when trying to acquire high-

quality footprint captures because the skin starts to dry, peel and crack, which obscures the ridge pattern. This leads us to ask if there is a specific time frame, a "window of opportunity", which allows for high-quality captures shortly after birth. We visually inspected the footprints and corresponding ballprints collected (for details on the database refer to Section 2.5.1 in Chapter 2 and Section 5.4 below) and reviewed them with respect to the newborn's age (cf. Figure 5.1). We focused on the hallucal area under the big toe.

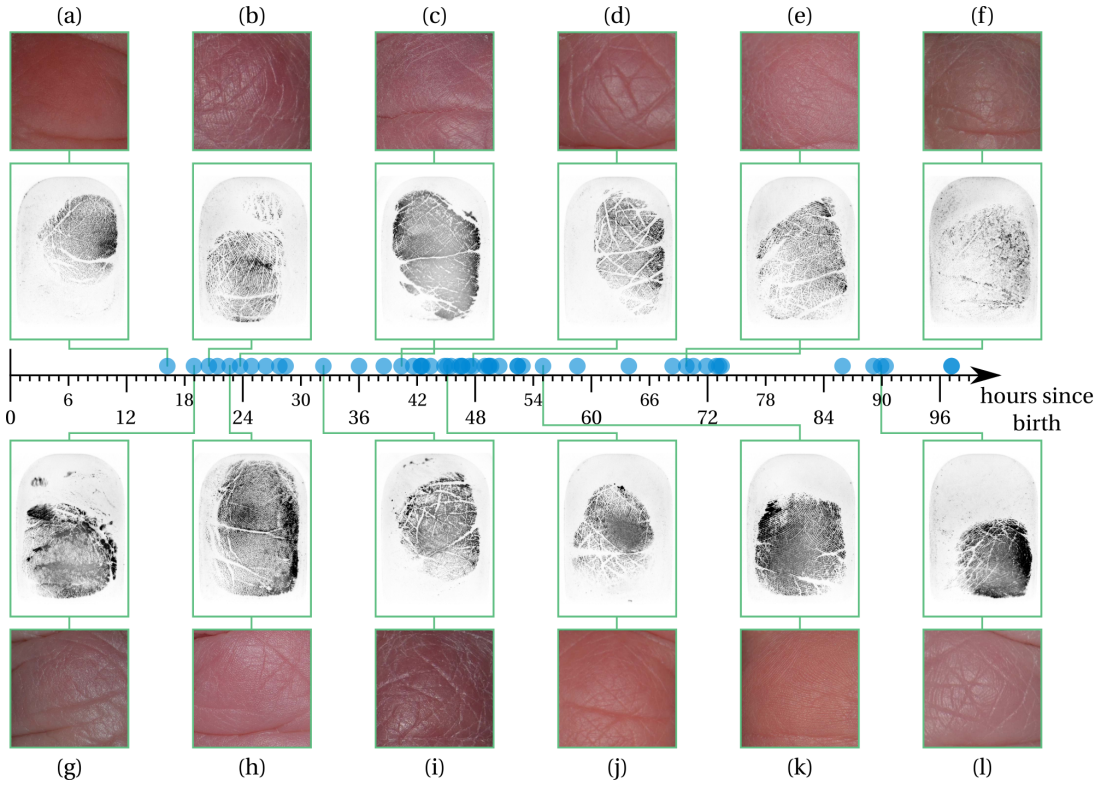


Figure 5.1: Is there a window of opportunity? Visual inspection of the infant's skin condition and its age. The captures shown are taken from the "Happy Feet" database (cf. Section 2.5.1 in Chapter 2). Some captures exhibit reasonable quality and visible ridge pattern (b, c, d, h, k) while others do not (a, e, f, g, i, j, l) for similar capture times after birth.

For the Happy Feet data, we were unable to find a correlation between the age and the image quality of the print captured. Overall, it appears that the skin condition varies greatly and is not subject to the newborn's age but rather to individual circumstances or environmental influences. In Figure 5.1 for example, the friction ridge skin of captures (i) and (j) is strongly affected by the drying skin respectively 32 hours and 45 hours after birth but capture (c) is not 40 hours after birth.

Based on this outcome, we decided first to focus our attention not on the friction ridge

skin (which might be obscured by the condition of the skin) but on the flexion creases. According to Blake, they are not affected by the drying [16] and "remain persistent over time" [99].

5.3 Crease pattern extraction

Literature on automatic identification methods suitable for infants is rare, either for footprints or crease pattern extraction and for this combination in particular. In all cases, police fingerprint experts are used to establishing an infant's identity manually. The first step towards identification is the crease pattern extraction. Uhl and Wild [143] downsample adult footprint images and apply Prewitt kernels in different directions. The final result is a feature vector based on the variance of overlapping blocks. Multi-resolution approaches for adult palmprints use directional filtering and the Radon Transform [65] or analyse a direction map locally [21]. The last technique is not suitable for infants because ridge lines are not picked up adequately at capture. In [28], a method for infant palmprint identification is proposed; it is based on the ridge pattern and uses inter alia the Short Time Fourier Transform, directional field estimation, and Gabor Filtering.

Our general approach is robust against capturing mistakes, both in timing and execution. It is based on the choice of the flexure creases, rather than the ridge pattern as features. A second advantage is that our algorithm works on infant footprints taken at different ages and qualities due to the use of directional filters applied to several resolution levels, morphological processing, and crease line reconstruction. This work is published in [72].

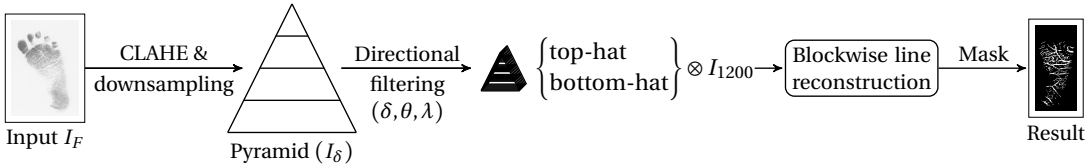


Figure 5.2: Schematic diagram of the algorithm. The input footprint image $\boldsymbol{I}_F$ is first optimized for its contrast using CLAHE and afterwards downsampled to the resolutions $\delta = \{1200, 600, 300, 150\}$ppi. These images can be seen as a pyramid; each level is convolved with a directional filter and rotated by $\theta = \{0°, 22.5°, \ldots, 157.5°\}$ versions. After summation over θ the tophat and bothat transformation are applied to the resulting pyramid. The results are normalized, the bottom-hat outcome is inverted, and they are pixel-wise multiplied. All resolution layers are upsampled to 1200 ppi, added up, and weighted by the image's gray-values. Afterwards, a block-based line reconstruction is performed and finally a mask is applied to the result.

The algorithm consists of 3 main steps (directional filtering, morphological processing, and line reconstruction) and is briefly illustrated in Figure 5.2. The input image $\boldsymbol{I}_F$ (cf. Figure 5.2) is optimized for its contrast using CLAHE [168] (see Section 3.2.2). A mask with toes removed is obtained similarly to [131] and the rotation is corrected using the mask's eigenvectors; this results in the image $\boldsymbol{I}_{2400}$.

To extract crease lines independent of their size, a multi-resolution approach [65] is chosen and $\boldsymbol{I}_{2400}$ is downsampled to $\boldsymbol{I}_\delta$ for $\delta = \{1200, 600, 300, 150\}$ppi. The ridge pattern introduces noise regarding the crease extraction; to suppress it, a 2 dimensional Gaussian lowpass filter $\boldsymbol{G}_s$ is applied to $\boldsymbol{I}_{1200}$. For the other resolution levels $\boldsymbol{G}_w$ is used. By choosing their values, there is a trade-off involved: the larger the filter, the stronger the ridge line suppression, and the higher the likelihood to lose small crease lines.

(a) Visit 1 (b) Visit 2 (c) Visit 3

Figure 5.3: The captured right footprint (rotated and cropped) of infant 001 at different ages: (a) 3 days, (b) 2 months, (c) 6 months. The footprints are to scale.

5.3.1 Directional filtering

All $\boldsymbol{I}_\delta$ are convolved with the Gabor filter h [55] and its versions rotated by $\theta = \{0°, 22.5°, \ldots, 157.5°\}$; the results are summed over θ: $\boldsymbol{I}'_\delta = \sum_\theta \boldsymbol{I}_\delta * h(x, y : \theta, \lambda)$. The filter has been slightly modified to smooth in the y-direction and is generated by the following equations:

$$h(x, y : \theta, \lambda) = \exp\left\{-\frac{1}{2}\left[\frac{x_\theta^2}{\sigma_x^2} + \frac{y_\theta^2}{2\sigma_y^2}\right]\right\}\cos\left(\frac{2\pi}{\lambda}x_\theta\right), \tag{5.1}$$

$$x_\theta = x\cos(\theta) + y\sin(\theta),\ \ y_\theta = -x\sin(\theta) + y\cos(\theta). \tag{5.2}$$

5.3.2 Morphological processing

The tophat and the bothat transformations have been defined in Section 3.2.2 in Chapter 3. Here the tophat uses the *Structure Element* (SE) $\boldsymbol{S}_t$ and the bothat the SE $\boldsymbol{S}_b$. Both transformations are applied to $\boldsymbol{I}'_\delta$, the results are normalised to $[0, 1]$, and the tophat and the inverted bothat outcome are pixel-wise multiplied:

$$\boldsymbol{I}''_\delta = \mathrm{n}(\mathrm{tophat}(\boldsymbol{I}'_\delta, \boldsymbol{S}_t)) \circ [1 - \mathrm{n}(\mathrm{bothat}(\boldsymbol{I}'_\delta, \boldsymbol{S}_b))], \tag{5.3}$$

$$\mathrm{n}(\boldsymbol{I}) = \frac{\boldsymbol{I} - \min(\boldsymbol{I})}{\max(\boldsymbol{I}) - \min(\boldsymbol{I})}. \tag{5.4}$$

Disk-shaped structures with the radii s_t and s_b for respectively $\boldsymbol{S}_t$ and $\boldsymbol{S}_b$ are chosen slightly bigger than the smallest expected crease line (cf. Section 5.4.3). The next step is that all $\boldsymbol{I}''_\delta$ are upsampled to 1200ppi and pixel-wise added up creating $\boldsymbol{I}_C$. The result is weighted by the corresponding gray-value of the enhanced image $\boldsymbol{I}_{1200}$, normalized, and the square root is pixel-wise taken: $\sqrt{\mathrm{n}(\boldsymbol{I}_C \circ \boldsymbol{I}_{1200})}$. Finally, hysteresis thresholding is applied to this image using the thresholds t_1 and t_2. Therefore, pixels are selected (1) if their value is greater than or equal to t_1 and rejected (0) if smaller than t_2. All pixels with a value between t_1 and t_2 are kept (1) if they are connected to a pixel already selected in the first place.

5.3.3 Block-based line reconstruction

The thresholding after the morphological processing breaks weak structures apart. Therefore a line reconstruction is performed for every non-overlapping square block. First, all binary objects are located (connected pixels with the value 1). Then, the following steps are performed for every object: its orientation is estimated by computing its eigenvectors and eigenvalues on the basis of its x and y-coordinates; its shape is broken down to the ratio *ocl* between the smaller and the bigger eigenvalue. No further steps are performed for this object if it is small or has no clearly identifiable line-like shape (cf. Section 5.4.3). Otherwise, a bottom-hat transformation is performed. A line $\boldsymbol{L}$ with the object's orientation and length l is used as the SE to bridge broken crease lines. All resulting structures connected to the object are weighted by the original pixels' gray-values in $\boldsymbol{I}_{1200}$ and the median is calculated. If the median is greater than a threshold t_3, the changes are kept; otherwise they are discarded.

Finally, the result is improved by (i) breaking weak structures apart, (ii) removal of small and mid-sized objects, (iii) closing of small holes, and (iv) applying the mask obtained from the beginning.

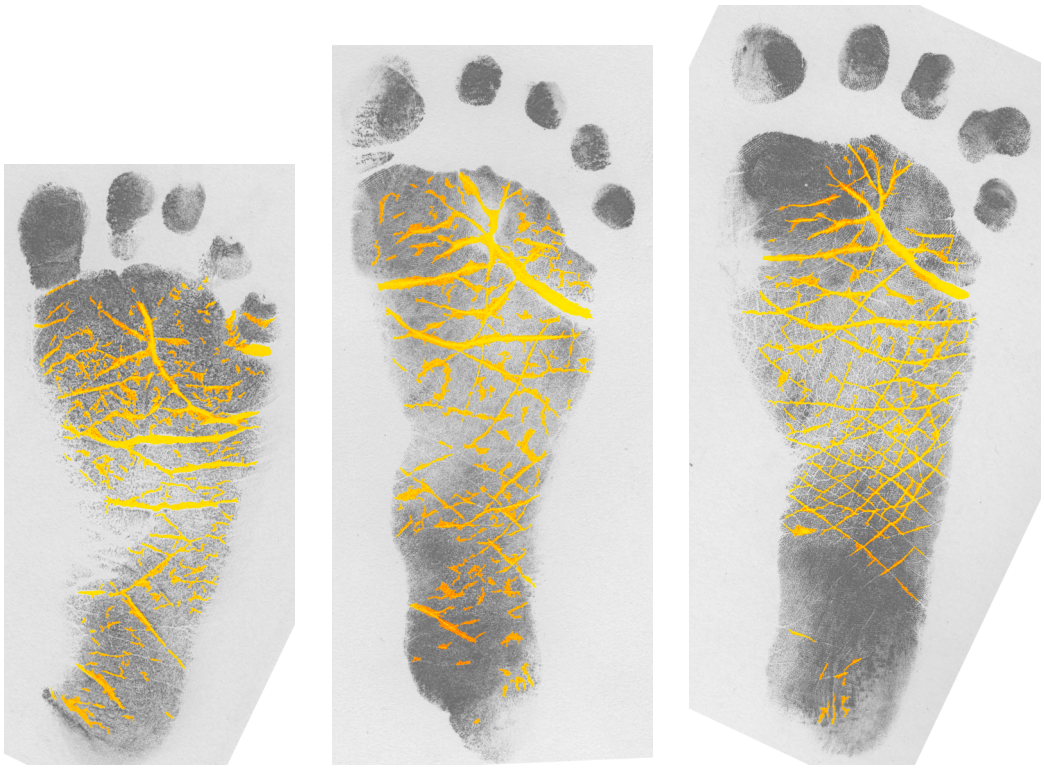
5.4 Happy Feet and algorithm application

The described algorithm has been applied to our longitudinal Happy Feet database of infant footprints captured with the inkless paper system and scanned afterwards at a resolution of 2400 ppi. A detailed description of the database and the capture process can be found in Section 2.5.1. An example of the same foot captured at the ages of 3 days, 2 months and 6 months is shown in Figure 5.3. We perform a manual and an automatic verification experiment.

5.4.1 Manual experiment

In order to test whether a human can match the footprints, a short experiment was conducted wherein 20 pairs of footprints were created. 11 of the 20 pairs came from the same foot (some from different visits) and 9 were from different feet but pairs were chosen that looked similar. 7 non-experts correctly classified at least 55% of the pairs and 2 researchers with considerable experience with ridge-based biometrics correctly classified 19 of the 20 pairs. Lastly, 2 forensic police fingerprint experts correctly classified respectively 15 and 19 of the pairs.

We are unable to eliminate the likelihood that the test participants used other struc-



(a) Visit 1 (b) Visit 2 (c) Visit 3

Figure 5.4: The footprint images after CLAHE (cf. Figure 5.2) overlaid by the extracted crease pattern. Note that the result is weighted by the image gray-values. The result's colour ranges from orange (overlaying dark image parts, e.g. a ridge) to bright yellow (light areas, e.g. a crease line). The same foot as in Figure 5.3 is used.

tures as well as the crease pattern to arrive at their final decision if two prints are from the same individual or not. Hence we are unable to draw a conclusion if the crease pattern is a suitable newborn or infant biometric.

5.4.2 Automatic extraction and matching

We performed an automatic version of the experiment on the same 20 footprint pairs in order to see if it was worthwhile setting up a full automatic verification experiment. We applied our algorithm to the footprints and extracted the crease pattern. In order to compare two different crease patterns, we used the Iterative Closest Point algorithm [11, 22]. It is a standard algorithm to align two point sets and we elaborate on the underlying idea in Section 3.2.1 in Chapter 3.

For this purpose, let us assume we register $\boldsymbol{A}$ onto $\boldsymbol{B}$; they contain n and m points,

respectively. This results in the aligned $\boldsymbol{A}'$. Two points are considered corresponding if their Euclidean distance is smaller than a certain threshold t.

$$\|\boldsymbol{A}_{(\cdot,i)} - \boldsymbol{B}_{(\cdot,j)}\|_2 \leq t, \tag{5.5}$$

where $i = \{1,\ldots,n\}$ and $j = \{1,\ldots,m\}$. If there are several matches or some points are used multiple times, the one with the smallest distance is always chosen; in case of multiple points with the same distance, the greedy algorithm chooses the first occurrence. It is applied to $\boldsymbol{B}$ first, then $\boldsymbol{A}'$. Let the total number of matching points be p. The final matching score is then given by

$$\frac{p}{\sqrt{n \cdot m}}. \tag{5.6}$$

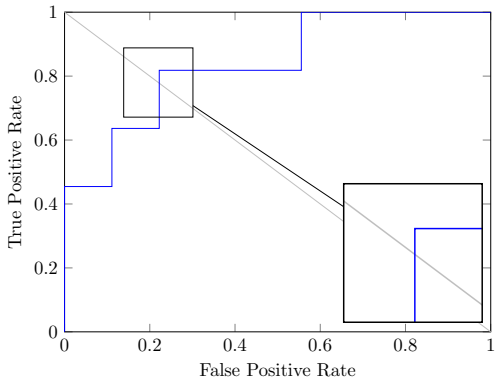
The threshold was experimentally set to $t = 6$ but the choice of t actually does not affect the outcome because of the huge number of points, on average roughly 539,000.

However, the two point sets are formed by the points belonging to the crease pattern. Due to the huge number of points, the crease pattern extraction results were downscaled to 300 ppi. This experiment of 20 comparisons resulted in an EER of 22.22%. The experiment was repeated with 10,000 random genuine comparisons and 10,000 random imposter comparisons. The EER increased to 46.39%. Furthermore, we compare a pair of *dizygotic* twins, infants 003 and 004 (528 genuine and 576 imposter comparisons), resulting in an EER of 50.07%. The corresponding ROCs and the score distributions are shown in Figure 5.5.

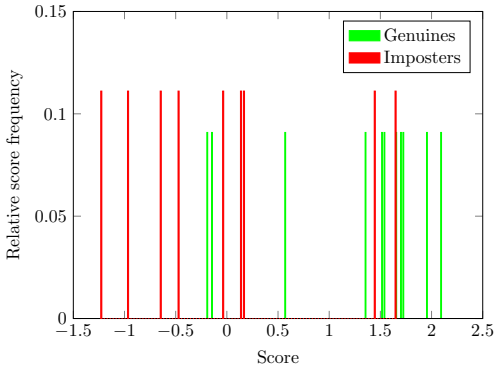
A visual inspection of the prints used in the manual experiment and their extraction results brought to light that the actual reason for the high EER is not the crease pattern extraction algorithm (except for *one* case) but the biometric captures themselves. Most of the footprints suffer from serious quality issues and exhibit great variations: even large flexion creases seem to appear and disappear when comparing captures taken right after each other; the same may even apply to the captures of adult footprints collected under controlled conditions [99]. The main reason is the infant's uncooperative (or playful) nature together with the capture process. In order to obtain high-quality prints, it is required that the foot does not move during the procedure. Unfortunately, most infants were playful, kept kicking or flexing their feet which lead to suboptimal captures. Furthermore, inkless paper possesses a very fine structure which is picked up in the scans as well and interferes with the friction ridge pattern, making the collected footprints difficult to use for the examination of the friction ridge skin.

The matching of all possible genuine and imposter comparisons for one pair of dizy-

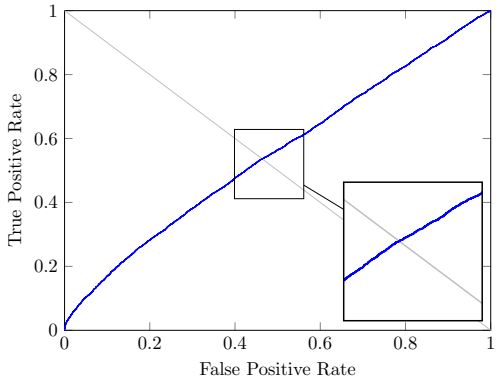
gotic twins lead to an EER similar to the one obtained by the automatic experiment.



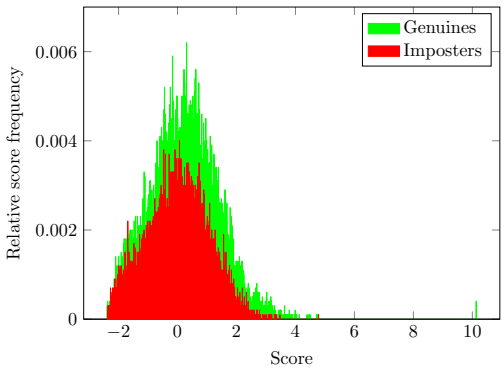
(a) ROC curve



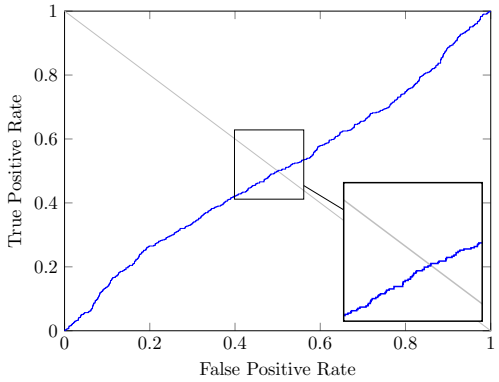
(b) Normalised score distribution



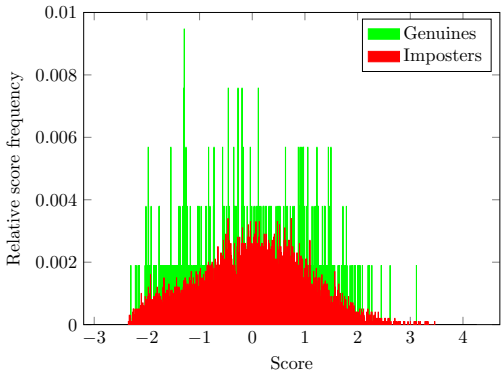
(c) ROC



(d) Normalised score distribution



(e) ROC curve



(f) Normalised score distribution

Figure 5.5: The ROCs (a) for the crease pattern manual experiment (20 comparisons), (c) performed automatically with a random sample of 10,000 genuines and imposters each and (e) all possible genuine (528) and imposter (576) comparisons for a set of (dizygotic) twins. EERs of 22.22%, 46.39% and 50.07% have been achieved, respectively. The crease pattern extracted was matched using ICP. The score respective distribution (b, d, f) is normalised according to the imposter distribution.

5.4.3 Extraction parameters

All parameters have been obtained experimentally: The Gaussian filters $\boldsymbol{G}_s$ and $\boldsymbol{G}_w$ are set to (width = 25, σ = 8) and (width = 15, σ = 6), respectively. These values were found to provide the right balance between ridge line suppression and keeping small crease lines. The directional filtering parameters $\sigma_x = \sigma_y = 5$ and $\lambda = 5\pi$ have been chosen. The trade-off that occurs here is discussed in [55]. The SEs used during the morphological processing are chosen to be slightly bigger than the smallest expected crease line: $s_t = s_b = 6$. The hysteresis thresholding uses $t_1 = 0.5$ and $t_2 = 0.3$; these are chosen to return only strong crease lines and ignore weak ones. The block-based line reconstruction uses the block edge length $m = 128$, the length of line $\boldsymbol{L}$ is set to $l = 20$. An object is considered to be small (mid-sized) if it consists of less than 200 (500) pixels. Values of *ocl* smaller than 0.25 indicate a line-like shape. The threshold to keep reconstructed areas is $t_3 = 0.7 \cdot 255$. An example result of the algorithm using these parameter settings is shown in Fig. 5.4.

5.4.4 The algorithm

The algorithm itself is not the main reason for the high EER in the automatic verification experiment and we would like to stress the fact that an EER of 22.22% is actually in line with other published footprint experiments (see Table 5.1). Nevertheless, we consider that the current algorithm has four weaknesses: (i) contrast optimization amplifies both the actual footprint pattern and the image noise and therefore a choice is forced upon the user to lose creases in weak contrast areas or introduce false crease lines, (ii) if the foot's structure has not been completely picked up by the inkless paper due to skin dryness or too little pressure (areas with a few distant dots instead of ridge pattern), then most likely the whole area is selected as a crease, (iii) there is an unfortunate trade-off between suppressing the ridge pattern but preserving the crease lines, (iv) unusual foot shapes might prevent the cropping of all toes from the image.

5.5 Conclusion

In this chapter we have reviewed the literature for potential newborn and infant biometrics and pointed out challenges being faced. Footprints have been captured in hospitals for many years and the idea is therefore familiar and well-established. So, we have proposed a novel algorithm for newborn and infant crease pattern extraction as a first step towards a biometric for newborns and infants. It has been applied to a set of footprints taken at 3 different ages.

Despite the use of modern capture equipment, the process remains at best a challenging one and does not guarantee high-quality images. This does not infer anything about the underlying biometric and its suitability per se but it does confirm that measurability is an important factor to take into account when choosing a suitable infant biometric. Our automatic verification attempts demonstrated only moderate performance on a small database and underscore the importance of reliable features. In our experience, flexion creases show great variability; they appear and disappear and hence are not a suitable newborn biometric.

We were unable to confirm or deny the existence of a window of opportunity which would limit the capture time frame to specific times; the skin condition depends on the individual. There is no need to exclude the friction ridge skin in general.

Another disadvantage of the approach here is the equipment required: inkless paper, towelettes and a flatbed scanner. The inkless paper and towelettes are costly, heavy and the paper's chemical compound can be accidentally activated by humidity or sweat. Furthermore, the fingerprints of the adult who is capturing the infant footprints are also easily captured, unless they wear disposable gloves. These drawbacks make the inkless paper system less suitable for use in developing countries where lack of resources or environmental conditions play a role.

In the chapters to come we do not pursue the crease pattern but concentrate on a part of the infant's friction ridge skin, the hallucal area under the big toe, which is rarely affected by flexion creases. It can be captured without consumable materials and relies instead on standard (fingerprint) technology. The ballprint exhibits similar properties to a fingerprint. We will discuss and investigate the ballprint, its features and suitability as infant biometric in the next two chapters.

6 Ballprints for infants with known age

This chapter continues to investigate the features of the friction ridge skin located under the big toe, on the ball of the foot. When one looks at the ball for the very first time, it shows a striking similarity to the friction ridge skin present at the fingertip. There are two main questions: First, how does growth affect an infant ballprint captured at different ages? Second, is it possible to compensate for physical growth so that the biometric ballprint can be used successfully in either verification or identification scenarios? These questions are driven by the search for a reliable infant biometric. In this chapter we also stress the technical challenges associated with the ballprint such as the problem to acquire high quality captures (cf. Section 5.1.1).

First, we elaborate on the "age problem", the change of a biometric modality over time (Section 6.1) and compare the biometric traits ballprint and fingerprint side by side (Section 6.2). We present the methodology applied to estimate ballprint growth (Section 6.3) before we set up the experiments and present the results for ballprint growth (Section 6.4). We next use our findings to lay out verification experiments and explain the results obtained and the implications for newborns (Section 6.5) and for newborns (Section 6.6). We then establish external correspondence between the ballprint IRS and physical length (Section 6.7), before we discuss our findings (Section 6.8).

The preliminary results of this chapter have been published in [71]. The chapter has been submitted in its extended form to a journal [73].

6.1 The age problem

Biometrics are built on two main principles. Firstly, that certain intrinsic characteristics are distinct to one person and secondly that their properties remain constant for this person's entire life [35]. For fingerprints inter alia this refers to the minutiae type, distribution or connectivity.

Fingerprints have been used by law enforcement agencies for over a century, are well established and have proven to be a valuable and reliable tool [54]. The biometric's distinctiveness has really been questioned, in regard to its template distinctiveness when solely based on minutiae [103].

However there are also projects such as India's "Aadhaar" program. It aims to enroll every Indian resident, making it the largest project of its kind. The system designers chose to use fingerprint (and iris) because of its easiness to capture, its high acceptance, low failure to enroll and its stability over time. In a preliminary study however, they found that the false rejection rate increases for children (5-15 years) and seniors (60+ years) [148].

Several other studies question the biometric's stability over time [145] too. Modi *et al.* raise the concern that ageing affects the skin, but also the ability to use the fingers to e.g. press them onto a scanner due to arthritis [104].

Uhl and Wild investigated for fingerprint how ageing affects recognition performance [145]. They separate the issue of ageing into two concerns: (i) age groups and (ii) template age. The first one deals with the feature extraction's sensitivity to age-induced variation of the biometric; the second category looks at extracted templates and their change over time. Uhl and Wild conclude that recognition performance for children is severely subject to ageing. Furthermore, they state that template age increases a system's Equal Error Rate by a factor of two to four and that genuine comparison scores decrease by approximately one third.

Although these studies are not dealing with the rapid growth of very young children, it becomes evident that both categories (age groups and template age) have to be addressed in order to (i) find a reliable infant biometric and (ii) to develop a verification or identification system.

6.2 Similarity between ballprint and fingerprint

There are several studies that consider different biometrics such as ear, face, fingerprint or footprint for infants and newborns. Unfortunately only very few consider longitudinal datasets (spanning several months or years) to address physical growth of children (cf. Section 5.1.6 and Table 5.1). In the following sections, we briefly compare the physical features of friction ridge skin located within the hallucal area under the big toe and on the fingertip. A longer and more general discussion of this topic can be found in Section 2.2.

6.2.1 The ballprint

The biometric ballprint refers to the print of the hallucal area under the big toe. It is covered by friction ridges similar to those on a fingertip. Extensive research has shown that all friction ridges start to form around the gestational age of 10.5 – 16 weeks creating distinct features "due to developmental noise"; this process is finalised by the 24th week [54].

Usually, fingerprint features are categorised as belonging to three different levels of coarseness. There are (i) the global level (e.g. ridge flow), (ii) the local level (e.g. minutiae such as ridge endings and bifurcations) and (iii) the detailed level (e.g. pores). We already have described these for ridge-based biometrics in Chapter 2.2 but present ballprint and fingerprint features side by side in order to present a convincing argument regarding their similarity.

Level 1

Most authors distinguish between three patterns formed by the ridge flow in the hallucal area. They are (i) arch, (ii) loop and (iii) whorl [26, 42, 107, 157, 158]. The FBI's footprint classification scheme is built on these three classes [37].

The most frequently found feature is the loop [26, 42, 107] but this can depend on the investigated population [157]. Montgomery observed some correlation between the feature type of an individual's right and left ballprint [107]. Okajima [119], and Wertheim and Maceo [156] offer an explanation for these findings based on friction ridge skin morphogenesis, its development process.

The FBI's AFIS distinguishes between four patterns for fingerprint: (i) arch, (ii) left-slant loop, (iii) right-slant loop and (iv) whorl [54]. Examples of the different patterns are shown in Figures 2.3 and 2.1 for ballprint and fingerprint, respectively.

One additional Level 1 feature is the *inter-ridge spacing* (IRS); the spacing between individual ridges varies depending on personal attributes such as gender, ethnicity (such as African-Americans and Caucasians [1]) or age but also due to the area it has been captured at and the spatial closeness to singular points [1, 49, 116, 136].

When one compares the average spacing of fingerprints and ballprints, it becomes evident that the ridge structures of the ballprint are significantly larger. We have measured the median IRS for the adult ballprint database (Section 2.5.2 in Chapter 2) and for the FVC2002 DB1, and we observed 10.375 px (mean 10.398 px, SD 1.744 px) and 8.31 px (mean 8.42 px, SD 0.84 px) respectively. Hence the physical ridge structures of an adult ballprint are *roughly* 25% *larger* when compared to an adult fingerprint. The different capture resolutions of the two databases have been taken into account. We chose to work with the median over the mean to increase robustness against potential outliers.

For more details about the whole procedure of IRS estimation refer to Section 6.4.3.

Level 2

Uhl and Wild measure 300 to 400 minutiae in high quality images of the adult ballprint database (flatbed scanner) [144]. Okajima usually counted 100 to 130 minutiae [118]. He used an ink-based method to obtain the prints from 345 Japanese teenagers. The difference between the ballprint studies is most likely due to different capture techniques and the exact definition of which part of the hallucal area belongs to the ball.

Okajima also analysed rolled fingerprints of 77 males and 82 females. He counted an upper limit of 112 and 91 minutiae for males and females, respectively. The highest averaged minutia count was found for both sexes and both hands on the thumb and the lowest count in general on the little finger [117]. He only counted minutiae in the centre area of the print. The average count per finger is presented in Table 2.2 in Section 2.

Maltoni *et al.* assume the hypothetical number of 36 minutiae on average per fingerprint when estimating its individuality for different models reported throughout the literature [97, p. 351].

The higher number of minutiae found within a ballprint compared to a fingerprint can be explained by the size of the physical area; it spans an area at least twice as large as a fingerprint. This statement is based on the assumption that the fingerprint is not a rolled print which captures the sides of the finger, too. Additionally, larger

physical structure allows more detail to be captured and extracted. The adult ballprint competes with or even exceeds the average number of minutiae encountered on a thumb and exceeds that on a (non-rolled) fingerprint.

Level 3

Ballprints also clearly possess very fine details such as pores. However, they cannot always be captured due to the method used or to insufficient image quality; the same issue applies to fingerprints. Nevertheless, we were able to spot pores for some (high quality) ballprint captures taken from older infants (see Figure 2.3 in Chapter 2 and Figure 6.3 in this chapter).

6.2.2 Ballprint as an infant biometric

Based on this evidence, we conclude that the biometric ballprint possesses similar properties to fingerprint. This may allow us to apply existing technology to ballprint that has been developed for fingerprint.

Uhl and Wild did so by applying a standard publicly available fingerprint extractor and matcher to adult ballprints (the adult ballprint database, Section 2.5.2 in Chapter 2). They achieved a False Match Rate of 1.35% at a False Non-match Rate of 4.06% [144].

For infants, the ballprint has several advantages over fingerprint. It is easily accessible and can be captured from infants without use of force because they do not clench their toes in the way they curl their fingers into a fist [155]. Later in life the print is protected by footwear in developed countries, making it difficult for a ballprint to be collected without consent or accidentally left behind. Hence the biometric is more protected against usage for an unintended purpose.

In the case of a mass accident, the ballprint is again protected by footwear such that law enforcement would possibly individualise victims very quickly; DNA analysis can take several days to be processed and requires specialised laboratory equipment [134, 155]. Lastly the ridge lines on the ball are significantly larger and more widely spaced than their finger counterpart [155] (also see Section 6.2.1). This means that they can be collected from even very young children with standard fingerprint equipment whereas the capture of a fingerprint at the same age would require special (and expensive) equipment, or not be possible.

6.3 Methodology

The problem that rapid infant growth affects all physical features and hence challenges an infant biometric's permanence has been discussed in Sections 1.1 and 2.4 of Chapters 1 and 2, respectively. Recent studies show that normalising the biometric modality to counteract growth increases automatic identification or verification performance greatly [47, 63]. We confirm these findings in Section 6.5.2.

Researchers have examined fingerprint growth with varying results. On the one hand, Schneider [127] failed to identify a linear affine transformation to approximate the fingerprint growth. He used digital scans of 308 children aged 1 – 18 years and only 12 were younger than 4 years, with 79% aged 7 – 10. A second set was collected from 186 subjects one year later. On the other hand, Gottschlich *et al.* [47] state that fingerprint growth can be approximated by isotropic scaling (the print is scaled by the same factor k for the x and y direction, basically a similarity transformation without rotation and translation) and propose the ratio of the median population height at the different ages of the specific gender as a suitable scaling factor. They used 48 criminal records of juveniles initially aged 6 – 15 and a second set of prints collected between one and many years later.

In order to compensate for potential growth effects, one first needs to estimate the growth. Here we use minutiae, the discontinuities in the ridge line pattern of the ballprint and limit them to ridge bifurcations and ridge endings. These form a set of spatial features that are typically used when comparing ridge-based biometrics (cf. Section 2.2 in Chapter 2). We use the minutiae in the ballprints to align two captures of the same ballprint (even if they are captured at different ages).

A ballprint with n minutiae is represented by its $(3 \times n)$ coordinate matrix $\boldsymbol{M}$ using homogeneous coordinates. So every minutia corresponds to one column and can be written as $\begin{bmatrix} x_i & y_i & 1 \end{bmatrix}^T$ where $i = 1, \ldots, n$.

Now consider 2 different ballprints $\boldsymbol{A} = \boldsymbol{M}_{t_1}$ and $\boldsymbol{B} = \boldsymbol{M}_{t_2}$ from the same foot but taken at ages t_1 and t_2 where the minutiae (columns) correspond precisely to each other. In this case the dimensions of $\boldsymbol{A}$ and $\boldsymbol{B}$ are the same. We assume $t_1 < t_2$.

If there is no growth it would be possible to transform $\boldsymbol{A}$ onto $\boldsymbol{B}$ using only translation and rotation. However as the capture process introduces noise by e.g. stretching the skin, an error term $\boldsymbol{E}$ has to be added (also see Section 3.2.1 in Chapter 3) even when there is no growth.

On introducing growth, the question is whether the physical change can be approxi-

mated by a similarity transform (Section 3.2.1 in Chapter 4) using translation, rotation and scaling. Mathematically, a matrix $\boldsymbol{T}$ has to be chosen to deal with the trade-off between generalisation and low estimation error:

$$\boldsymbol{B} + \boldsymbol{E} = \boldsymbol{T}\boldsymbol{A}, \text{ with } \boldsymbol{T} = \begin{bmatrix} k\cos\theta & \sin\theta & t_x \\ -\sin\theta & k\cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix}. \tag{6.1}$$

This section is split as follows: firstly, we present approaches to align two ballprints and to estimate the transformation required (Section 6.3.1), secondly, we discuss transformation-independent methods to estimate an isotropic scaling factor in order to compensate for infant growth (Section 6.3.2), and thirdly, we introduce the methods and measures used to analyse the error introduced by the scaling factors (Section 6.3.3).

6.3.1 Ballprint alignment – transformation estimation

In order to estimate the growth, we need to align two corresponding ballprints. The minutiae alignment error will give some insight into the growth process with respect to the transforms and transformation estimation methods. We use four different approaches: (i & ii) the full and the partial Procrustes analysis and two RANSAC approaches estimating either (iii) a similar or (iv) an affine transformation. All of these techniques have been described in Section 3.2.1 of Chapter 3. Nevertheless, we briefly recapitulate and adjust them to the notation used here.

(i) The full Procrustes analysis finds the similarity transform that minimises the squared error between two shape configurations [32].

(ii) The partial Procrustes analysis ignores scaling ($k = 1$). In our case it can be used to find the transformation matrix $\boldsymbol{T}$ that projects $\boldsymbol{A}$ onto $\boldsymbol{B}$.

$$\boldsymbol{T} = \underset{\boldsymbol{T}^\star}{\operatorname{argmin}} \sum_{i=1}^{n} \left\| \boldsymbol{T}^\star \boldsymbol{A}_{(\cdot,i)} - \boldsymbol{B}_{(\cdot,i)} \right\|_2^2. \tag{6.2}$$

The error $\tilde{e}$ is defined as the median of the Euclidean distance between the transformed ballprint coordinates and their corresponding ones (column-wise Euclidean distance). Hence let $e_i = \|\boldsymbol{E}_{(\cdot,i)}\|_2$ (cf. Equation 6.1) and reorder the e_i so that

$e_{i-1} \leq e_i \leq e_{i+1}$. Then

$$\tilde{e} = \begin{cases} e_{\frac{n+1}{2}} & n \text{ odd} \\ \frac{1}{2}\left(e_{\frac{n}{2}} + e_{\frac{n}{2}+1}\right) & n \text{ even} \end{cases}. \tag{6.3}$$

(iii) Next, a similarity transformation matrix projecting $\boldsymbol{A}$ onto $\boldsymbol{B}$ is estimated by employing a RANSAC approach.

(iv) Similarly, an affine transformation matrix is estimated using a RANSAC approach.

RANSAC's advantage is its robustness against potential outliers. It chooses a random subset of points (as many as needed to estimate the transformation), performs the estimation, applies it to the whole dataset, and counts the number of corresponding points closer to each other than a given threshold. This procedure is executed a certain number of times and finally the transformation with the highest count is chosen.

6.3.2 Scaling factor estimation assuming isotropic growth

RANSAC is robust against outliers but alignment using the full Procrustes analysis could be biased due to individual outliers or an uneven minutiae distribution on the ballprint, because the Procrustes analysis minimises the total error. Therefore we estimate the scaling factor k (cf. Equation 6.1) in four transform independent ways: (i) analysis of the IRS, (ii & iii) two measures of the ratio of distances between corresponding minutiae and (iv) the median population height ratio as suggested by Gottschlich *et al.* [47].

We will use all of these scaling factors to estimate infant growth isotropically between different ages. We will show that they arrive at very similar results, despite their different approaches (cf. Section 6.4.3), which supports the correctness of the assumption (of isotropic growth).

We assume now for all formulations in this Section 6.3.2 that the underlying growth process is isotropic. In case this simplifying assumption has been falsely made, the error analysis will bring it to light.

Inter-ridge spacing

The IRS is one important property of the ballprint. As the ball grows, the IRS of the ballprint mirrors the growth. The IRS ratio between two points in time for one specific

ball resembles an individual scaling factor for this ball's growth.

$$k^*_{irs} = \frac{irs_A}{irs_B}. \tag{6.4}$$

The IRS is a rough indicator and can be potentially misleading as it varies across the ball. The spacing significantly depends on the area it is measured on because the area's properties such as the presence of singular points (core, delta) influence the measurement to some degree (see Section 2.2 in Chapter 2). The median of all k^*_{irs} (over the Happy Feet growth set for each time period, cf. Section 6.4.1) counteracts this effect and is denoted k_{irs}.

Distance similarity

With isotropic growth the distance between any two minutiae will be similar.

$$k_{(i,j)} = \frac{b_{(i,j)}}{a_{(i,j)}}, \tag{6.5}$$

$$a_{(i,j)} = \|\boldsymbol{A}_{(\cdot,i)} - \boldsymbol{A}_{(\cdot,j)}\|_2, \tag{6.6}$$

$$b_{(i,j)} = \|\boldsymbol{B}_{(\cdot,i)} - \boldsymbol{B}_{(\cdot,j)}\|_2. \tag{6.7}$$

The scaling factor for a particular ballprint pair is the median of all ratios $k_{i,j}$ and is denoted $\tilde{k}$. As infants grow at different rates a global scaling factor usually differs from an individual one. Two of our four global scaling factors are derived from the individual $\tilde{k}$ as the median (k_{median}) and the average (k_{mean}) of all $\tilde{k}$.

Population height

The fourth scaling factor (k_{who}) follows the suggestion of Gottschlich *et al.* [47]. They propose for fingerprints to calculate the scaling factor as the ratio between the median population height at the two given ages. Hemy *et al.* [51] confirm that there is a indeed a connection between physical height and the left foot length. Therefore the final value is the average for boys and girls of the height ratio at different ages, taken from the World Health Organization (WHO) growth charts [159] as recommended by the *Centers for Disease Control and Prevention* (CDC) to health care providers to monitor "growth for infants and children ages 0 to 2 years of age" [20]. Figure 6.1 illustrates the infant growth process as a plot of median length versus age.

We chose to average between boys and girls because "the sex differences in growth are much smaller than the normal variations in growth during infancy [...]" [50, p. 211],

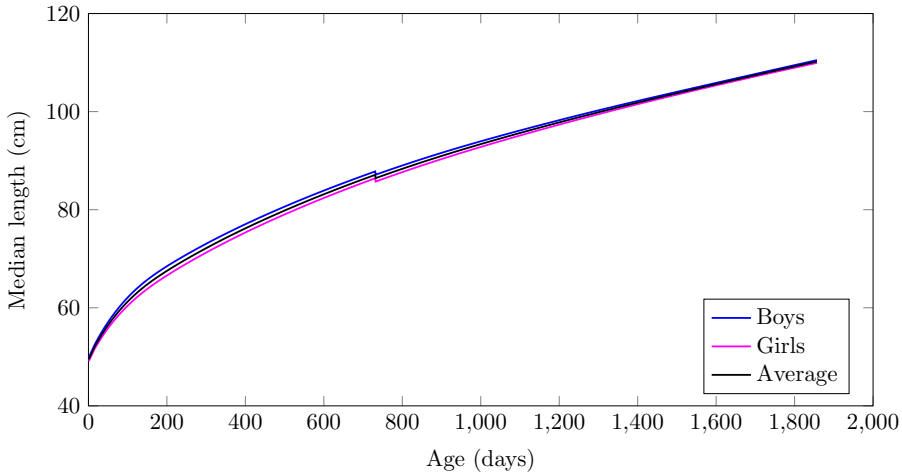


Figure 6.1: A growth chart displaying the median length (in cm) versus age (in days) spanning the time from birth to the age of 60 months. It is based on data collected by the WHO [159]. The discontinuity between 730 and 731 days is due to a change in methodology, the switch from recumbent length to standing height.

as can be seen in Figure 6.1.

Furthermore, for simplicity we are trying to estimate a single scaling factor which can be applied in the field regardless of the ethnic origin or received nutrition (to some degree). Hence we are investigating global scaling factors which do not rely on individual measurements, eliminating the need for potentially error-prone measures like the IRS. We define k_{who} as the ratio of median population height at the two different ages; this is the slope in Figure 6.1 between the two different points on the black curve.

6.3.3 Error analysis

In order to investigate the error introduced by choosing a global scaling factor over an individual one, the $n \times n$ error matrix $\boldsymbol{D}$ is calculated element-wise with respect to a fixed $k' \in \{k_{irs}, k_{median}, k_{mean}, k_{who}\}$ by

$$d_{(i,j)} = k_{(i,j)} - k'. \tag{6.8}$$

Hence there is an error matrix $\boldsymbol{D}$ for every comparison of the same ballprint captured at different ages where n denotes the number of corresponding minutiae. The subtraction of a fixed k' represents only a global shift. Hence the actual value of k' does not affect the (spatial) error correlation.

The error of corresponding minutiae ($\boldsymbol{E}$) or their pairwise distances ($\boldsymbol{D}$) between the isotropically scaled ballprint $\boldsymbol{A}^\star = \boldsymbol{TA}$ and the observed print $\boldsymbol{B}$ is expected to be randomly distributed. This is assumed because isotropic growth should be compensated for by the alignment process. Hence, only the minutia selection and the capture process should influence the error. If minutiae close to each other suffer from similar errors, there is spatial error correlation and a systemic error can be assumed (cf. Fig. 6.2). The following methods are employed to test for correlation: (i) eigenvalue difference, (ii) Moran's I [109] and (iii) Geary's C [45].

Geary's C measures auto-correlation locally whereas Moran's I does this more globally; they are loosely inversely related [48].

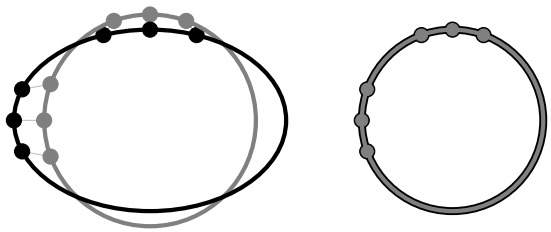


Figure 6.2: The dots representing minutiae which are close to each other suffer from similar errors; thus they are spatially correlated. The black ellipse (left) and circle (right, coincides with the grey one) are the actual surface the minutiae are placed on and the grey circles represent isotropically scaled estimates.

Eigenvalue difference

The eigenvalue difference is given by $evd = |\lambda_1 - \lambda_2|$; it is computed from the alignment error of corresponding minutiae ($\boldsymbol{E}$), by the following covariance matrix

$$\Sigma = E\left[(\boldsymbol{X} - E\{\boldsymbol{X}\})(\boldsymbol{X} - E\{\boldsymbol{X}\})^T\right], \tag{6.9}$$

where $\boldsymbol{X} = \boldsymbol{E}_{(1:2,\cdot)}$ and E stands for the expectation. Again, we assume the minutiae alignment error $\boldsymbol{E}$ to be randomly distributed. The growth is accounted for by isotropic scaling. Therefore, the evd is expected to be close to 0 as there is no major error direction. So, the unit eigenvectors adjusted by the corresponding eigenvalue would span rather a circle than an ellipse. We chose the difference of eigenvalues over the ratio to avoid penalising small eigenvalues.

Statistical measures

The other techniques (Moran's I and Geary's C) require a weight matrix $\boldsymbol{W}$ containing the pairwise reciprocal distances between pairs of minutiae (see Equation 6.7):

$$w_{(i,j)} = \begin{cases} 0 & \text{if } i = j \\ b_{(i,j)}^{-1} & \text{else} \end{cases}. \tag{6.10}$$

The value to be tested for spatial correlation is the (directed) summed error for a particular minutia (cf. Equation 6.8). The column vector $\boldsymbol{v}$ represents row-wise summation of $\boldsymbol{D}$:

$$v_i = \sum_{j=1}^{n} d_{(i,j)}. \tag{6.11}$$

Set $w = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{(i,j)}$ and $\bar{\boldsymbol{v}} = \boldsymbol{v} - E[\boldsymbol{v}]$. Moran's I is defined as follows

$$I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{(i,j)} \bar{v}_i \bar{v}_j}{w \bar{\boldsymbol{v}}^T \bar{\boldsymbol{v}}}. \tag{6.12}$$

Geary's C is defined as

$$C = \frac{(n-1) \sum_{i=1}^{n} \sum_{j=1}^{n} w_{(i,j)} (\boldsymbol{v}_i - \boldsymbol{v}_j)^2}{2 w \bar{\boldsymbol{v}}^T \bar{\boldsymbol{v}}}. \tag{6.13}$$

6.4 Growth estimation and databases

This section focuses on the estimation of infant ballprint growth and aims to derive a model so that we can compensate for its effects on the ballprint. We are going to apply our findings in Section 6.5.

Here, we firstly specify more clearly what databases and subsets we are going to use for any further experiments and justify the need to scale the ballprint captures (Section 6.4.1). Secondly, we present the steps taken to analyse and approximate the growth (Section 6.4.2), before we, thirdly, show the characteristics measured for the datasets employed (Section 6.4.3), fourthly we illustrate our approximation for ballprint growth and the scaling factors it infers (Section 6.4.3), and fifthly we discuss our findings (Section 6.4.4).

6.4.1 Databases and subsets

Our longitudinal ballprint database consists of 54 infants who were enrolled into the study "Happy Feet". It has been described in Section 2.5.1 in Chapter 2 in detail. The print quality ranges from very poor to exceptional and hence the distinctiveness from almost nil to exceptional (cf. Figure 2.10 in Chapter 2). This is mostly due to babies being uncooperative subjects, the ridges' tininess and dry and cracking skin.

The experiments are conducted on a subset of the collected prints. We chose only captures from the second, third and fourth visit of good quality and where by visual inspection we knew there was reasonable overlap between corresponding prints. Captures from the first visit were disregarded as the overall image quality is too low due to the dry and cracking skin (see Figure 2.10e in Chapter 2 and Figure 5.1 in Chapter 5).

Test set

The test set $\boldsymbol{S}$ contains ballprint specimens which have been manually selected to have high quality and reasonable overlap between the different specimens; the resulting numbers of prints and different feet for the second (V2), third (V3) and fourth visit (V4), respectively appear in Table 6.1. The total number of genuine and imposter comparisons is shown in Table 6.7.

Set		V2	V3	V4
Test set	images	94	98	178
	feet	72	72	62
Tuning set	images	13	47	47
	feet	7	21	23
Growth set	images	65	65	63
	feet	65	65	63

Table 6.1: Number of feet and images for the different sets per visit.

Tuning set

Our tuning set does not overlap with the test set in any way. If the test set includes a certain individual's specimen of the right ballprint at a particular age, the tuning set includes at most (if at all) its left ballprint at that age. We consider left and right ballprints of the same individual to be different despite there being historical evidence

of some correlation between the Level 1 ridge flow patterns [107]. In our experience, when this occurs, these features tend to be mirrored and the Level 2 features remain distinct. Thus it is fair to assume that the right and the left ballprint are indeed distinct. We chose high quality images that represent the test set as well as possible. This lead to 13 (V2), 47 (V3), and 47 (V4) images (see Table 6.1).

Growth set

Finally, there is the growth set which is a subset of the test set. We employ it to estimate the ballprint growth by manually marking minutiae. We chose at most one pair per foot (cf. Figure 6.3). There are 65 and 63 pairs in total for Visit 2 (V2) & Visit 3 (V3) and for Visit 3 (V3) & Visit 4 (V4), respectively. Some infants were unavailable for a fourth capture session; hence the smaller number of pairs (cf. Table 6.1).

All prints were chosen to have sufficient image quality and reasonable overlap to ensure matching was possible. V3-V2 pairs have been selected first, the closest visually matching V4s have been added afterwards.

Adult set

The adult ballprint database (cf. Section 2.5.2 in Chapter 2) was used within the verification experiment as a reference set. The set stands out because these ballprints have been captured from cooperative subjects. Additionally, these subjects are adults, hence their physical structures are well developed (the median IRS is greater than the one of a standard fingerprint database – cf. Section 6.2.1 earlier in this chapter) and can be easily picked up with standard technology.

Image scaling

There are two effects one needs to take into account when trying to use standard fingerprint technology for infant ballprints. Firstly, the infant ballprints are similar to adult fingerprints but do not exhibit the same inter-ridge spacing. This would lead to mediocre extraction and subsequently poor comparison results because fingerprint software is optimised for a capture resolution of 500 ppi and makes assumptions regarding the IRS (cf. Section 4.3 in Chapter 4). One needs to scale ballprint captures to make them compatible with the latest fingerprint technology available. Secondly, as the infant grows, the physical size of the biometric changes.

We have introduced four different scaling factors $\{k_{irs}, k_{median}, k_{mean}, k_{who}\}$ in Section 6.3.2 and we will present their results with respect to the scaling between the different

visits in Section 6.4.3.

In order to adapt the captures to fingerprint extraction and matching software, we used our growth set and the manually marked minutiae as ground truth. We established the greatest correlation for ballprint images captured at the age of 2 months with the minutiae picked up by the software (here: Neurotechnology Verifinger) by varying the resolution suggested to the SDK. We performed this over the range 175 ppi to 350 ppi in steps of 25 ppi. Then we compared the number of minutiae found that were present in both the manual mark-up and the automatic extraction. The resolution with the highest number of corresponding minutiae was chosen. This is the only time we had an SDK perform the image *scaling internally* (i.e. we just supply the image as captured and suggest a capture resolution).

In this thesis, we consistently used the software Matlab to perform *external scaling* with nearest-neighbour interpolation. We made this choice so that performance is based on the feature extraction and matching but not the choice of image scaling algorithms.

6.4.2 Growth analysis

In order to obtain a minutiae ground truth, one researcher marked only corresponding minutiae between the captures from different visits in the growth set; minutiae without correspondence were disregarded. The number of marked corresponding minutiae can be found in Table 6.2 and a visual example is shown in Figure 6.3.

	V2, V3	V4, V3	V4, V2
Minimum	5	1	0
Maximum	52	66	33
Median	17	28	9

Table 6.2: Number of corresponding minutiae between visits, which have been manually selected.

The selection process was split. First, we selected corresponding minutiae between the prints of the third and the second visit (V3, V2), then corresponding minutiae for the fourth and third visit (V4, V3). The correspondence between the fourth and second visit has been inferred via Visit 3. Hence, it is the subset of minutiae occurring in both of the marked sets (V3, V2) & (V4, V3). These minutiae are used for the growth analysis but they are not employed for any verification experiments.

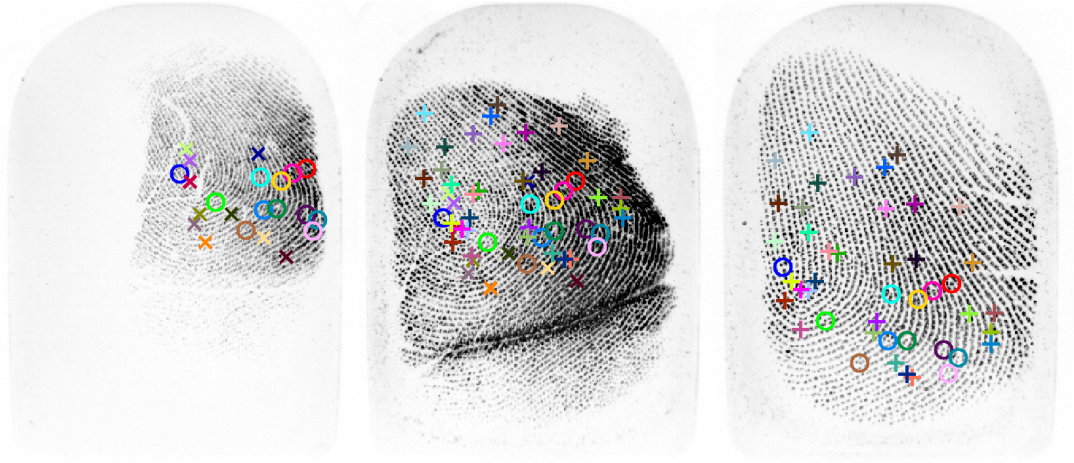


Figure 6.3: Comparison of three ballprints from infant 019's left foot captured at 2 months, 6 months and 2 years and their manually marked minutiae; corresponding ones share the same colour. The different markers indicate for which visits the correspondence could be established: ∘: $\{V2, V3, V4\}$, ×: $\{V2, V3\}$, +: $\{V3, V4\}$.

6.4.3 Results of growth estimation

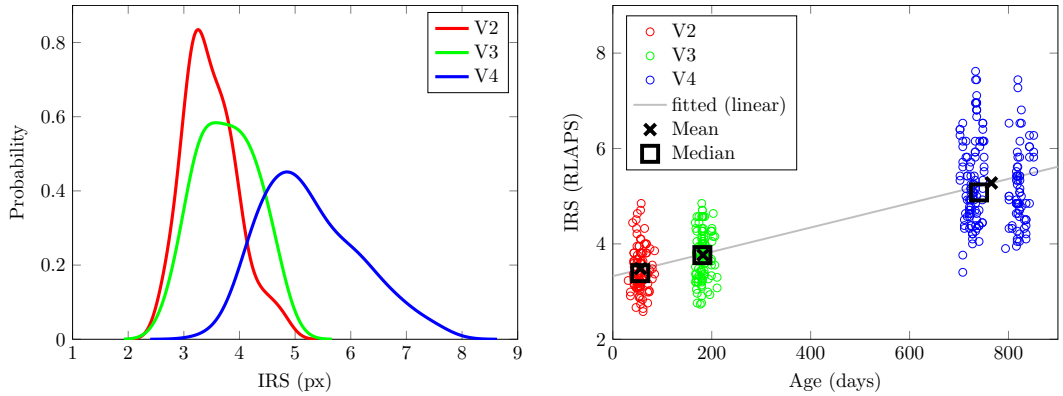
This section presents the results from the analysis of the IRS for our test set (Section 6.4.3), the growth estimation on the growth set (Section 6.4.3), and the scaling factors (Section 6.4.3).

Database analysis

We employed our RLAPS algorithm (cf. Section 4.3 in Chapter 4) to measure the median inter-ridge spacing and found values of 3.4 px (2 months), 3.8 px (6 months) and 5.1 px (2 years); the parameters $\gamma = 0.4$, $\alpha = 15°$ have been used for V2 and V3 but $\gamma = 0.2$ for V4. The IRS median, mean and standard deviation are shown in Table 6.3. The children's were on average 55.74 days (V2), 182.4 days (V3), and 765.4 days (V4) old.

The analysis of the inter-ridge spacing for our test set shows that the IRS as a measure exhibits great variance with respect to to the infant's age as shown in Figure 6.4. The two clusters for the fourth visit are because the research assistant needed to spend some time recovering after an accident. The IRS is subject to the infant's physical features, its nutrition and individual growth. Because of the huge overlap across the different age groups, it is not possible to derive an infant's age inversely from its IRS.

We subsequently decided to base the scaling factor k on the inter-ridge spacing (so



(a) IRS probability distribution

(b) IRS vs age

Figure 6.4: Measurement of the inter-ridge spacing of the infant ballprints with respect to the infant's exact age (in days). Results are presented as (a) an IRS-*Probability Density Function* (PDF) and (b) a scatter plot with highlighted average and median per visit. We were able to fit a line through the average age and IRS.

k_{irs}) to derive a function with respect to the child's age for multiple reasons. First, the IRS is a physical feature that can be measured easily, especially when the ballprint has to be re-captured in a verification scenario. Thus it allows to verify the scaling factor in a real world scenario, for potential application when the child's age is unknown, or when the child has suffered from malnutrition and subsequently from stunted growth. Hence the child does not exhibit the same physical-features-to-age relationship as properly nourished children do. Second, fingerprint SDKs are heavily tuned towards a population average of the fingerprint's physical features. Therefore it is sensible to base a scaling function on those features considering that the ultimate goal is to apply standard fingerprint technology to the ballprint. Thus k_{irs} seems to be the most feasible and practical solution to use.

	V2	V3	V4
Median	3.3864 px	3.7647 px	5.0794 px
Mean	3.4818 px	3.7674 px	5.2784 px
SD	0.4762 px	0.5242 px	0.8740 px

Table 6.3: Median, mean and standard deviation of the IRS for the test set.

The fitted line in Figure 6.4b has the following equation:

$$\mathrm{irs}(t) = 0.00255 \cdot t + 3.323, \tag{6.14}$$

where t is the infant's age in days.

We estimate the inter-ridge spacing Neurotechnology Verifinger is optimised for via the correlation between our manual mark-up and its automatic extraction results as described in Section 6.4.1.

We found the greatest consensus is achieved for the global scaling factor $k = 2$. In other words, the inter-ridge spacing of ballprints at the age of 2 months exhibits approximately half the width of the average adult fingerprint IRS Verifinger is optimised for. Hence, we suggested a capture resolution of 250 ppi = $500/k$ ppi to the SDK, to use *internal scaling* by the fingerprint extractor instead of performing it externally.

There might be other internal assumptions which lead to a general lower value of IRS the SDK is optimised for when comparing ballprint and fingerprint (see Section 4.3 in Chapter 4 and Section 2.5.2 in Chapter 2). Nonetheless, we conclude that Verifinger is optimised for an inter-ridge spacing of 6.8 px (2-times the median IRS for V2).

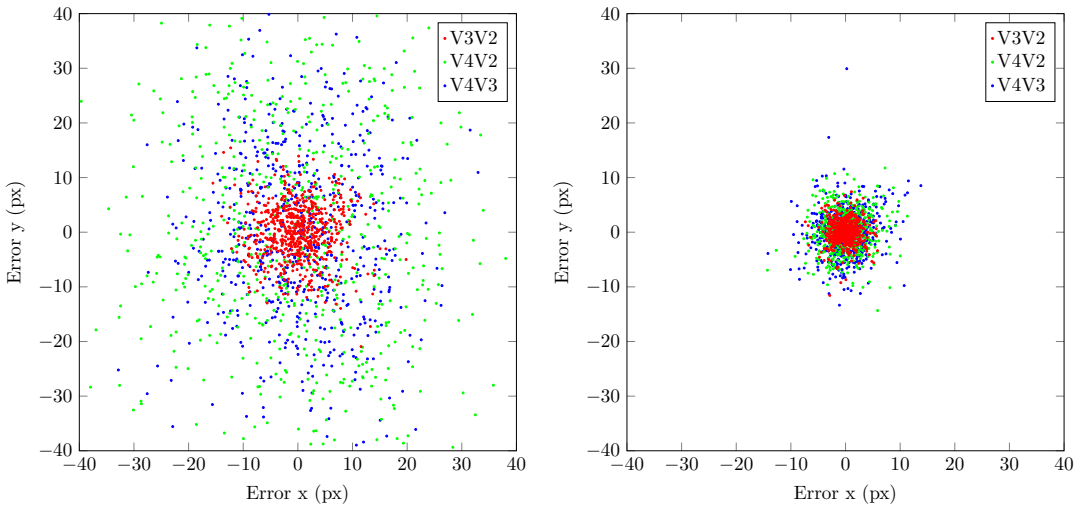
$$irs_{VF} = 6.8 \tag{6.15}$$

We chose the median over the mean for two reasons: (i) it is more robust against outliers and (ii) Gottschlich *et al.* chose the ratio of the median population height as their scaling factor which allowed for a dramatic improvement in their performance.

Growth analysis

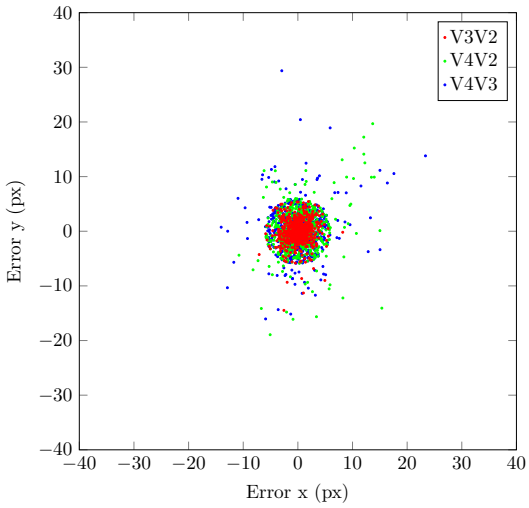
When the corresponding minutiae (of ballprints from the same infant but captured at different ages) were aligned using Procrustes without scaling (compensates only for rotation and translation) and with scaling (affine transformation) or RANSAC approaches estimating a similarity or affine transformation, we of course observe error. This is illustrated in Figure 6.5.

The similarity transform compensates for the error sufficiently when trying to align corresponding minutiae in prints at different ages. The affine transform reduces the error further (cf. Figure 6.5, Table 6.4) because it compensates not only for translation, rotation and isotropic scaling but also shearing and hence scaling with different factors for each direction. Nevertheless, it remains unclear if this error is actually due to growth or to the image capture process. The cumulative probability distribution over the minutiae error with respect to the chosen transform for different age comparisons illustrates this behaviour. The distributions are shown in Figure 6.6 and the median errors are stated in Table 6.4.

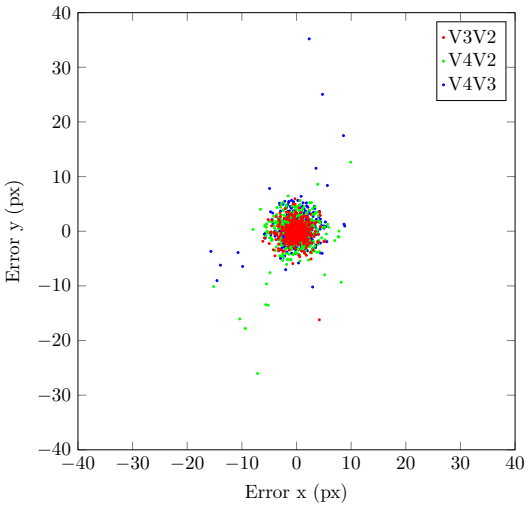


(a) Procrustes, No Scaling

(b) Procrustes Scaling



(c) RANSAC Similarity T.

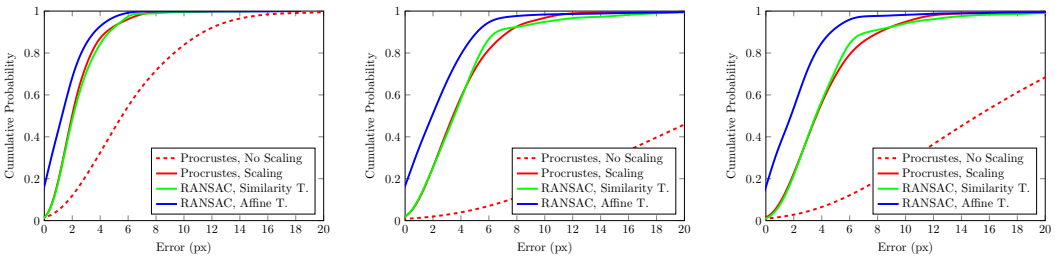


(d) RANSAC Affine T.

Figure 6.5: Visualisation of the error in x and y direction for corresponding minutiae after alignment. The greater the point concentration around the centre $(0,0)$, the smaller the alignment error and hence better the growth (and measurement error) compensation. The averaged minutiae error is stated in Table 6.4.

This evidence confirms that one *must* compensate for growth in one way or the other.

The resulting follow-up question is what technique or approach compensates for the growth most efficiently? Hence we use statistical measures to investigate if there is any spatial correlation between the minutiae alignment errors as it may highlight a systematic error.

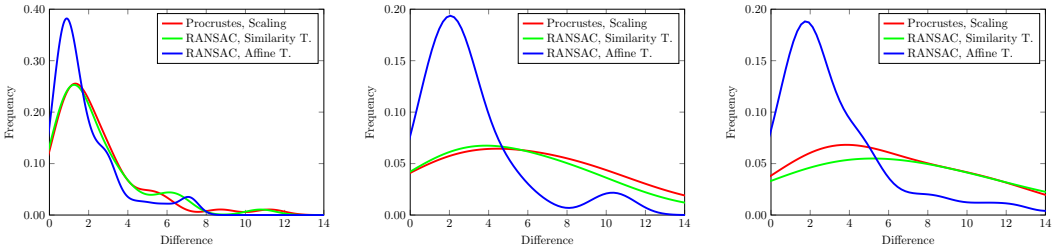


(a) V3V2 (b) V4V2 (c) V4V3

Figure 6.6: Cumulative probability distribution over the minutiae error w.r.t. the chosen transform for different age comparisons.

	Procrustes Analysis		RANSAC	
	No scaling	Similarity T.	Similarity T.	Affine T.
V3V2	5.45 px	1.95 px	2.03 px	1.02 px
V4V2	20.67 px	3.37 px	3.25 px	1.48 px
V4V3	15.36 px	3.57 px	3.44 px	1.29 px

Table 6.4: The averaged median alignment error ($\tilde{e}$, cf. Equation 6.3) over all ballprint pairs of the growth set resulting from different transforms and estimation methods.

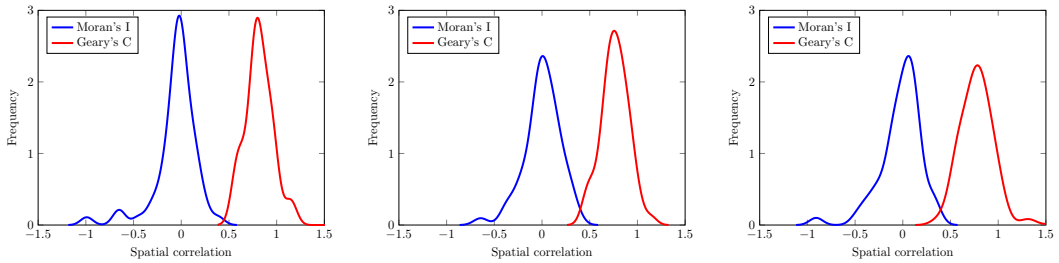


(a) V3V2 (b) V4V2 (c) V4V3

Figure 6.7: Estimated error eigenvalue difference distribution. The PDF is based on the *evd*s from all 63 manually marked minutiae pairs (cf. Equation 6.9, Section 6.3.3).

The eigenvalue difference distribution of the minutiae alignment error indicates that the affine transformation compensates for growth well and does not encounter a major error direction. However, the techniques which are limited to the similarity transformation suffer from the effect that one direction is not as well compensated for as the other. This effect might not be due to the actual growth process but the minutiae mark-up procedure (because correspondence for V4-V2 is not directly established but via the third visit). This procedure required all prints to have significant overlap but it turned out to be difficult to satisfy this due to the change in size (cf. Figure 2.11 in Chapter 2). Hence many corresponding minutiae are located close to the edge of

the ballprint area captured. This might favour a certain error direction as the skin is stretched in those areas during the capture process. The affine transformation can handle this type of error more easily than the similarity transformation due to its higher degree of freedom.



(a) V3V2 (b) V4V2 (c) V4V3

Figure 6.8: Spatial error correlation distribution indicated by Moran's I and Geary's C for the directed minutia ratio between visits compared to a global scaling factor k'.

This explains the slight positive spatial auto-correlation indicated by Geary's C which is more sensitive to local effects, while Moran's I measures more globally a slight negative spatial auto-correlation. This is illustrated in Figure 6.8 and Table 6.5.

	Moran's I		Geary's C	
	$\bar{I}$	SD	$\bar{C}$	SD
V3V2	−0.0561	0.2238	0.8322	0.1848
V4V2	0.0026	0.1872	0.7644	0.1423
V4V3	−0.0201	0.2159	0.7885	0.2006

Table 6.5: Spatial error correlation distribution indicated by Moran's I and Geary's C for the directed minutia ratio between visits compared to a global scaling factor k'.

We argue that if the minutiae alignment error were to be due to growth that the effect (and hence the error) should increase as the infant grows because so does its ball. Hence the effect would become more prevalent, would outweigh the capture error and would not be compensated for using the similarity transform. This is not the case when we compare the median minutiae error 3.37 px and 3.57 px (see Table 6.4) for an age gap of 18 months (V4V3) and 22 months (V4V2), respectively. Furthermore, the statistical measures Moran's I and Geary's C indicate that there exists local rather than global spatial error correlation (cf. Table 6.5), which could be due to the skin being stretched at the edges of the scanner.

Therefore, we conclude that the infant growth process can be sufficiently well approximated by isotropic scaling.

Scaling factor

We have argued that isotropic scaling may be sufficient to compensate for the growth effects encountered between the different age gaps investigated, and hence we now need to derive a function depending on the infant's age. Following our argument in Section 6.4.1, we described to use the median instead of the mean to increase robustness against potential outliers.

We define the function $res(t) \to \mathbb{R}$ which returns the "capture resolution" of a ballprint when its IRS is compared to an adult fingerprint that has been captured at 500 ppi. We decided to quantise the resolution in steps of $s = 25$ px to account for some measurement artefacts and to increase its generality and robustness. The function is based on the fitted line described in Equation 6.14, the approximate median IRS measured for the V2 ballprints $irs_{V2} = 3.4$, Verifinger's estimated internal IRS optimisation $irs_{VF} = 6.8$, and of course the standard resolution of 500 ppi:

$$\mathrm{res}(t) = \mathrm{round}\left(\frac{500 \cdot \mathrm{irs}(t)}{s \cdot irs_{VF}}\right) \cdot s = \mathrm{round}\left(\frac{500 \cdot \mathrm{irs}(t)}{25 \cdot 6.8}\right) \cdot 25, \text{ with } t \in \{40, \ldots, 800\}. \quad (6.16)$$

This can be expressed as a scaling factor as well:

$$\mathrm{k}(t) = 500 \cdot \mathrm{res}(t)^{-1}. \quad (6.17)$$

	$\boldsymbol{k_{irs}}$	k_{median}	k_{mean}	k_{who}
V3V2	**1.10**	1.15	1.16	1.17
V4V2	**1.50**	1.53	1.55	1.52
V4V3	**1.36**	1.34	1.34	1.30

Table 6.6: The different scaling factors $k' \in \{k_{irs}, k_{median}, k_{mean}, k_{who}\}$ for the different inter-age comparisons. The factors have been calculated on the growth set (if applicable) and k_{who} is the average of boys and girls.

For the verification experiments, we used one scaling factor per visit based on the test subjects' median age at that specific visit to allow for some individual size differences. This results in the final capture resolutions of 250 ppi (V2), 275 ppi (V3) and 375 ppi (cf. Equation 6.16) based on our choice to use the IRS function (cf. Section 6.4.3). The corresponding scaling factors are shown in Table 6.6.

Adult ballprints The adult ballprints need to be scaled to resemble adult fingerprints as well. Those prints have been captured at a resolution of 600 ppi and measurement

using RLAPS shows the database possesses an (unadjusted) median IRS of 12.45 px (RLAPS, $\gamma = 0.2$, $\alpha = 15°$). Verifinger is optimised for a capture resolution of 500 ppi and an inter-ridge spacing of 6.8 px when dealing with ballprints (cf. Section 6.4.3). Hence we derive the global scaling factor k_{abp} for adult ballprints as follows:

$$k_{abp} = \frac{2}{3} \approx \frac{6.8 \cdot 600}{12.45 \cdot 500}. \tag{6.18}$$

6.4.4 Discussion of growth estimation

All of our scaling factors arrive at similar results despite the use of (entirely) different approaches. This indicates that growth can indeed be approximated by isotropic factors. So one can choose the most practical or sensible factor in accordance with the application it is going to be used for (in our case k_{irs}). The reason for this decision is that it allows us to verify the results obtained and to check them for plausibility easily. The inter-ridge spacing can usually be measured even if the print captured suffers from quality issues.

We are aware the function $\text{res}(t)$ (Equation 6.16) is not applicable to newborns because of their rapid growth. For children past infancy, growth will slow down. This will be the case shortly after the age of 2 years (or 730 days) as the WHO growth chart highlights (Figure 6.1). Hence our linear function is a first approximation applying only to the period of infancy. Nevertheless, it is remarkable that there is a linear relationship between the inter-ridge spacing and an infant's age for the investigated age gap.

We note that the use of the median population height (k_{who}) may become a reasonable alternative to our inter-ridge spacing-based growth function (Equation 6.16) if there is no necessity to verify or adjust to a child's physical features.

6.5 Ballprint verification

Now we are going to employ the set of scaling factors k_{irs} we found in the previous section. First, we describe the experimental setup (Section 6.5.1), second, we present the results (Section 6.5.2), and third, we discuss our findings (Section 6.5.3).

6.5.1 Verification experiments

We perform two different verification experiments. The first one is performed on all specimens included within the test set and on the adult ballprint database for

reference purposes as cooperative subjects without any growth effects (Section 6.5.1). The second experiment is limited to high-quality infant ballprint images. We use EVA (cf. Figure 4.1 in Chapter 4) to derive a quality estimate for infant ballprints. The test set is then limited to high-quality specimens and the verification experiment is performed again (Section 6.5.1).

The visit order in the bracket notation indicates first the visit from which the reference image comes, and second the visit from which the query image is taken. In all Tables and Figures V2 refers to Visit 2 (child was 2 months of age), V3 refers to Visit 3 (child was 6 months of age) and V4 refers to Visit 4 (child was 2 years of age).

Complete test set

We applied two different feature extractors (Verifinger and SourceAFIS) and three matchers, SourceAFIS (sAFIS), M3gl and Verifinger (VF), to the Happy Feet test set and the adult ballprint database.

We have already described Neurotechnology Verifinger in Section 3.3.2 in Chapter 3. *SourceAFIS* is an open-source extractor and matcher that has been developed by Važan [153]. It competed in the FVC-onGoing [93] and demonstrated solid performance but cannot compete with commercial products. Nevertheless, it is used for many other projects because the source code is freely available.

One of these projects is M3gl by Medina-Pérez [101]. This fingerprint matcher uses SourceAFIS for its feature extraction and then performs the comparison itself. M3gl's code is disclosed as well (apart from the latest version); it achieved competitive results at the FVC-onGoing. To the best of our knowledge, it is the state-of-the-art open-source fingerprint matcher.

It is worth noting that specimen self-comparisons (intra-visit, $\{(V2, V2), (V3, V3), (V4, V4)\}$) have been excluded and that symmetric comparisons of two specimens are considered to be different because they do not lead necessarily to identical scores. The fraction of genuine comparisons is given in Table 6.7.

	V2	V3	V4
V2	$\frac{44}{94\times93}$	$\frac{131}{94\times98}$	$\frac{246}{94\times178}$
V3		$\frac{52}{98\times97}$	$\frac{245}{98\times178}$
V4			$\frac{474}{178\times177}$

Table 6.7: Fraction of the number of genuine comparisons on the total number of comparisons across different visits. Self comparisons for intra-visit experiments are excluded but symmetric comparisons are considered to be different.

High quality specimens

The uncooperative nature of the infant test subjects means that even our test set contains prints with questionable quality (and hence questionable distinctiveness). The question is how the verification scenario result changes if these images are removed from the set. We used the forensic experts' decisions for adult fingermarks (cf. Section 4.2 in Chapter 4) to train a classifier to classify a ballprint's quality according to its image features. Figure 6.9 shows this procedure in detail.

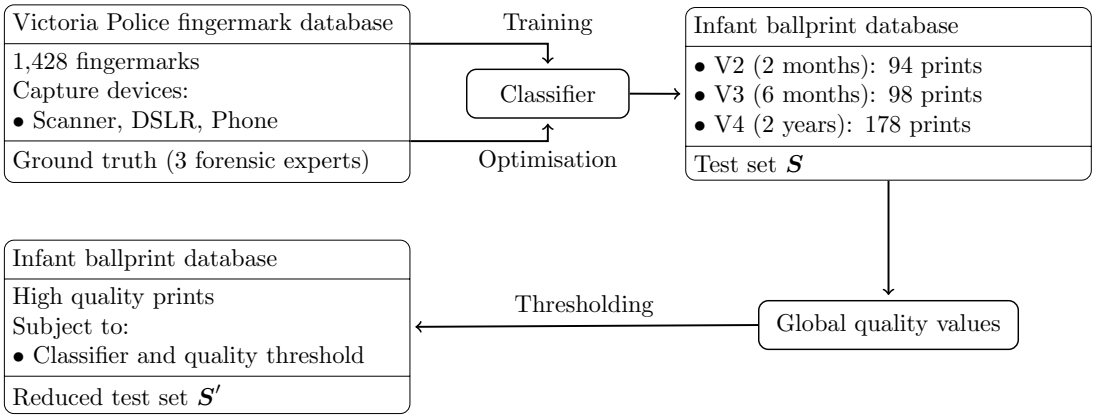


Figure 6.9: Diagram showing how the image quality is derived from adult pseudo fingermarks and applied to infant ballprints. We use the examiners' ground truth and extracted image features (Fusion, cf. Sections 4.2 and 4.5.2 in Chapter 4) to train a classifier. The underlying assumption is that fingermarks, or in this particular case images of ballprints of sufficient evidential value possess sufficient quality for further analysis and comparison. Therefore, we apply the (on fingermarks) trained classifier to the infant ballprints (after adjusting their resolution to compensate for any growth). However, instead of the binary evidential value, we use the classifier's raw score to produce a smaller ballprint subset of high quality images via thresholding.

This approach is based on the assumption that sufficient evidential value and image quality are correlated. This hypotheses is reasonable because we have encountered

that for pseudo fingermarks the evidential value can be derived from a set of image features. Therefore, the argumentum e contrario should hold true as well and the evidential value allows for a basic image quality estimation.

We used the entire fingermark database to train and optimise a classifier for the evidential value estimation according to the experts' ground truth.

The Fusion image feature set (see Sections 2.5.3 and 4.5.2 in Chapters 2 and 4, respectively) is extracted from each infant ballprint image, after it is scaled according to its age in order to compensate for growth. The trained classifier is then applied to the image features and the raw score $q \in [0,1]$ (not a binary evidential value decision) is used.

Now every ballprint image of the original test set $\boldsymbol{S}$ possesses a quality value which we threshold on to create a reduced test set $\boldsymbol{S}'$ of high-quality ballprints. The EER is recomputed for $\boldsymbol{S}'$ with respect to the threshold and the resulting number of images. Finally, the intersection between the EER curve and the percentage of removed images is computed. Hence, some EER values are hypothetical (because they can be interpolated values) rather than observed values.

6.5.2 Verification results

We have performed three different verification experiments. The first does not perform any adjustment to derive a baseline performance, the second compensates for the ballprint images for growth and adjusts them to match the characteristics of adult fingerprints. The third experiment additionally removes low-quality images and repeats the previous experiment.

Baseline performance

We used Verifinger to derive a baseline because it demonstrated the strongest performance in the FVC-onGoing fingerprint competition of all matchers used [93]. Hence no isotropic image scaling is performed and the ballprint captures are used as captured without any further processing.

Verifinger (see Table 6.8) achieves sufficient EERs only for intra-visit comparisons, results obtained for prints from different visits are not competitive as shown in Figure 6.10. This clearly stresses the necessity to take a child's growth into account when specimens of different ages are compared.

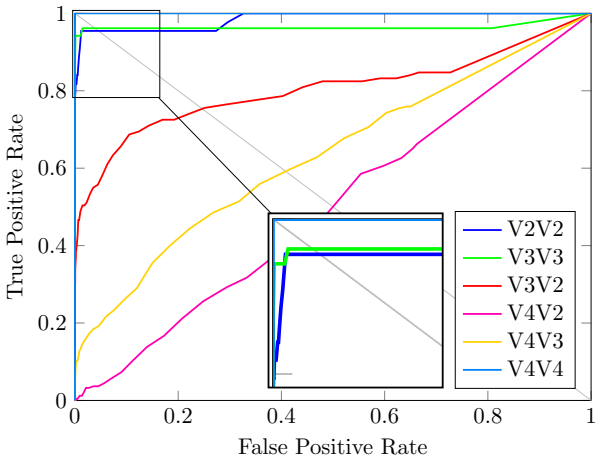


Figure 6.10: The ROC for Verifinger if *no* image capture scaling is performed. The ROCs for all intra-visit identification experiments demonstrate that Verifinger is able to match ballprints reasonably well while there is no change in scale (due to e.g. growth). As soon as the scale changes, the performance decreases significantly.

	Equal Error Rate						
Method	V2, V2	V3, V3	V4, V4	V3, V2	V4, V2	V4, V3	Adult
VF*	4.55%	3.85%	0.00%	24.66%	49.98%	40.82%	2.17%

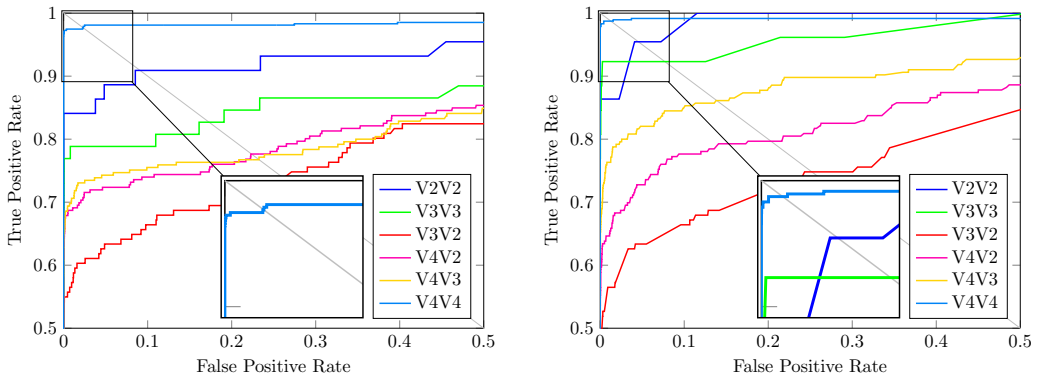
Table 6.8: EER scores achieved by VF for all possible ballprint age gaps without growth compensation or any adjustment (VF*).

Performance with isotropic scaling

When isotropic scaling accounts for physical growth, the different fingerprint matchers achieve adequate EERs when specimens within the same visits are compared. As soon as the specimens have been captured at different ages (the images have been externally scaled), the performance drops. This behaviour is illustrated in Figure 6.11 and numerical values appear in Table 6.9.

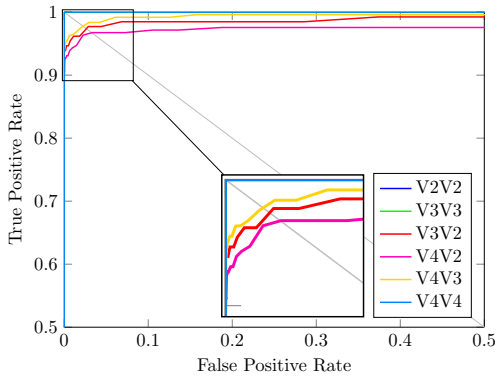
The infant ballprints have been scaled with k_{irs} depending on the visit the capture has been taken at (cf. Table 6.6). The adult ballprints have been scaled by $k_{abp} = 2/3$ as we have explained in Section 6.4.3.

Verifinger achieves remarkable performance when adjusted appropriately (by isotropic scaling) to the infant ballprint. The genuine and imposter score distributions, as shown in Figure 6.12, exhibit complete separation for all intra-age comparisons and slight overlap otherwise.



(a) Source AFIS

(b) M3gl



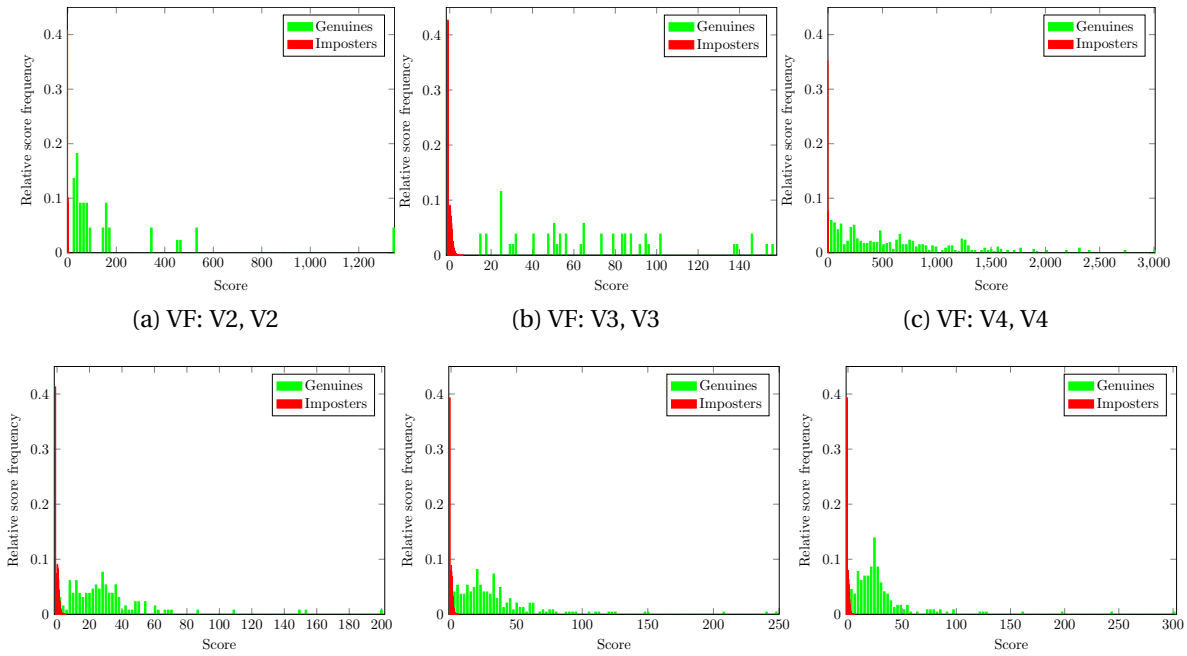
(c) Verifinger

Figure 6.11: The top left corner of the ROCs for all fingerprint matchers after tuning towards ballprints using external image scaling with k_{irs} (V2: 250 ppi, V3: 275 ppi, V4: 375 ppi, cf. Sections 6.4.3 and 6.4.3).

	Equal Error Rate						
Method	V2, V2	V3, V3	V4, V4	V3, V2	V4, V2	V4, V3	Adult
sAFIS	9.09%	17.31%	2.32%	26.18%	22.36%	23.25%	11.68%
M3gl	4.55%	7.69%	1.27%	25.19%	20.33%	13.88%	3.42%
VF	0.00%	0.00%	0.00%	2.63%	3.27%	2.24%	0.31%

Table 6.9: EER scores achieved by all fingerprint matchers for all possible ballprint age gap comparisons using k_{irs}. The EER resulting when applied to the adult ballprint database is shown as reference.

The verification experiment run on the adult ballprints allows the reader to compare results of infant with adult ballprints. On the one hand, the infant captures are obtained from uncooperative subjects at different ages and show partially questionable quality and tiny physical structures which are difficult to capture with standard fin-



(a) VF: V2, V2 (b) VF: V3, V3 (c) VF: V4, V4

(d) VF: V3, V2 (e) VF: V4, V2 (f) VF: V4, V3

Figure 6.12: Visualisation of the VF verification scores for (a) intra-visit 2, (b) intra-visit 3 and (c) intra-visit 4; as well as (d) the inter-visit 3&2, (e) the inter visit 4&2 and (f) the inter visit 4&3. All scores have been normalised according to the mean of the imposter distribution.

gerprint scanners. On the other hand, the adult ballprints have been obtained from cooperative subjects, show large physical structures and do not suffer as much from low-quality captures. Nevertheless, no matcher is able to achieve complete separation on the adult ballprints as Figure 6.13 illustrates. No optimisation or parameter tuning has been performed apart from scaling the images to adjust them to adult fingerprints in order to “trick” the extractors and matchers. A more detailed explanation can be found in Section 6.4.1.

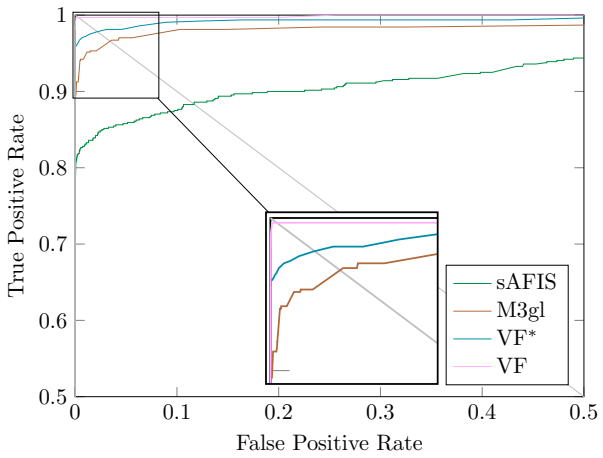


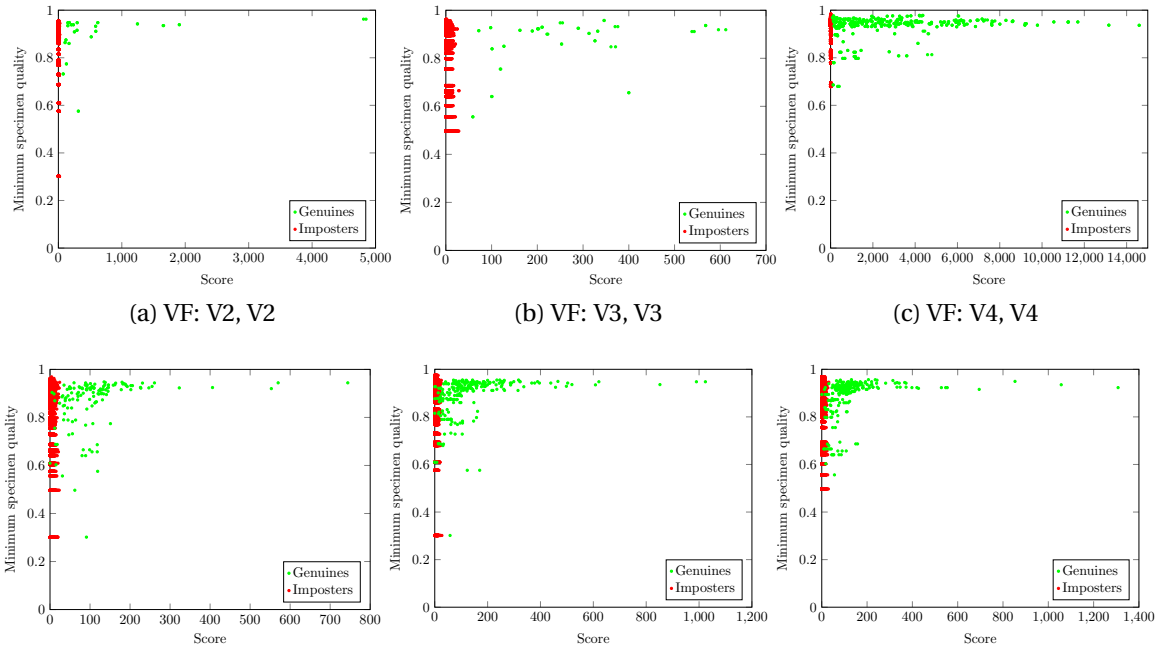
Figure 6.13: The top left corner of the ROCs for all fingerprint matchers after tuning towards adult ballprints using external image scaling ($k_{abp} = 2/3$, cf. Section 6.4.3).

High-quality specimens

Finally, we investigate the removal of low-quality infant ballprint specimens. We limit this experiment to the most promising matcher, Neurotechnology Verifinger. Figure 6.14 illustrates the individual (raw) comparison scores and the minimum quality of the two images compared. It is noticeable that most genuine comparisons tend to have a higher score than imposter comparisons but a few do not. Those genuine comparisons suffer from the fact that at least one of the templates has been extracted from a low-quality capture. Hence, we are basically looking for a horizontal line (graphically speaking), a quality threshold, that permits us to remove genuine comparisons with low scores. Subsequently, this results in a minimised EER.

If one restricts the test set to high-quality captures only, nearly complete separation can be achieved across all different age comparisons. The quality is derived from adult fingermarks examined by Victoria Police fingerprint experts regarding their evidential value. This process is explained and illustrated in Section 6.5.1 and Figure 6.9, respectively. In this case, the classifier (kNN) has been trained on the fingermarks from the Scanner and optimised for the DSLR and Phone images at FMR10. Figure 6.15 illustrates how the EER decreases while the percentage of images removed is increased.

We threshold on the quality score in order to remove low-quality images and to create a test set of high-quality prints. Nearly complete separation can be achieved with VF, if a third (33.3%) of the low-quality images contained in the original test set is removed. Only V4V3 retains a non-zero EER: 0.01%.



(a) VF: V2, V2 (b) VF: V3, V3 (c) VF: V4, V4

(d) VF: V3, V2 (e) VF: V4, V2 (f) VF: V4, V3

Figure 6.14: Visualisation of the VF verification scores for (a) intra-visit 2, (b) intra-visit 3 and (c) intra-visit 4; as well as (d) the inter-visit 3&2, (e) the inter visit 4&2 and (f) the inter visit 4&3 versus the minimum quality of the two ballprints compared. The matching score is *not* normalised.

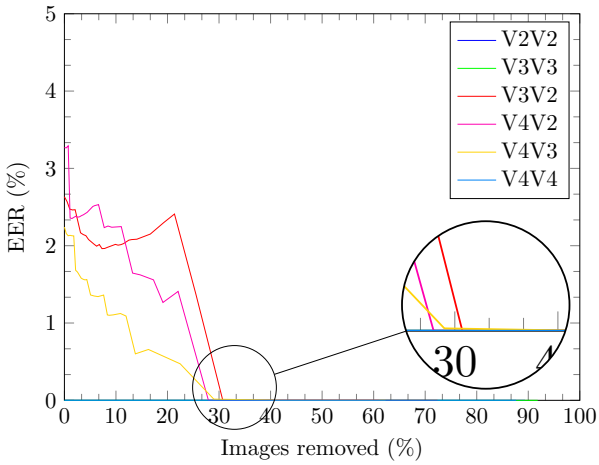


Figure 6.15: This diagram illustrates how the performance, in terms of the Equal Error Rate, improves when low-quality images are removed from the test set. The image quality of the infant ballprints is derived from adult fingermarks (cf. Figure 6.9).

6.5.3 Verification discussion

The three experiments underline that isotropic scaling using scaling factors derived from the inter-ridge spacing to compensate for growth leads to remarkably good performance. The reasons are that the scaling compensates for the growth effects reasonably well and, the ballprints become more similar to adult fingerprints. This allows the fingerprint extractors and matchers to use their highly specialised and optimised algorithms.

When the test set is limited to high-quality images, the commercial Verifinger benefits and nearly complete separation across all different age comparisons can be achieved if images with questionable quality are removed.

6.6 Verification of newborns

Originally we had excluded all newborn prints from these investigations. Due to their quality and lack of visible ridge detail, we were unable to find Level 2 minutiae in order to use those prints for growth estimation as well. Encouraged by the overall results of modern fingerprint SDKs and the success of scaling based on the Level 1 feature IRS we return to the newborn ballprints (V1). These prints were collected between 11 hours and 7 days after birth (except for one at 28 days). We selected in total the best 53 prints (of 20 individual feet) leading to 108 (V1,V1), 51 (V2,V1), 53 (V3,V1) and 63 (V4,V1) genuine comparisons for the different age gaps.

6.6.1 Newborn experiment

The median newborn age when these prints were captured is 3.97 days (mean 3.01 days, SD 5.03 days). Regardless of the quantity of ridge detail, we proceeded to measure the IRS for these 53 prints (RLAPS, $\gamma = 0.55$ and $\alpha = 15°$). The median was 2.96 px (mean 3.02 px, SD 0.31 px).

We consider the scaling adjustment for Visit 2 captures to be our baseline because we measured Neurotechnology Verifinger's behaviour for this dataset and established the scaling ratio ($k = 2$) for this particular point in time (cf. Section 6.4.3). The scaling factor was determined to be $k_{irs} = 0.87 = \frac{2.96\,\text{px}}{3.4\,\text{px}}$ for V1, which translates to a suggested resolution of approximately 220 ppi.

We perform newborn verification for all possible age gaps using VF as extractor and matcher after scaling the images with k_{irs}.

6.6.2 Newborn results

The ROCs resulting from newborn verification experiments are relatively poor as shown in Figure 6.16 compared with the results achieved for infants. The intra-visit verification, the most common case when preventing mix-ups at hospitals exhibits performance (see Table 6.10) comparable with what the literature reports for other newborn biometrics (cf. Section 5.1 in Chapter 5).

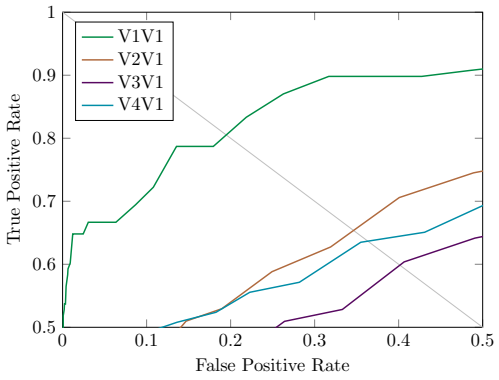


Figure 6.16: ROCs for ballprints captured from newborns. VF is employed as extractor and matcher after the prints have been scaled by k_{irs}. The corresponding EERs are stated in Table 6.12.

When the specimens compared are from different visits, the performance decreases. Nevertheless, the EERs for these comparisons are still in line with what VF without any image adjustment, or other (non-commercial) matchers, achieve for the later inter-visit comparisons (see Table 6.9).

	Equal Error Rate			
Method	V1, V1	V2, V1	V3, V1	V4, V1
VF, k_{irs}	19.48%	34.63%	40.14%	36.33%

Table 6.10: EER scores achieved by VF for all ballprint ages compared to the newborn captures.

6.6.3 Discussion of newborn results

The scores underline how challenging newborn verification or identification is and the personal burden to handle the uncooperative nature of the newborn. The EER of 19.49% is similar to the scores of other newborn and infant biometrics reported in the literature.

Shepard *et al.* [129] report the correct manual matching of only 10 out of 51 newborn footprints; Balameenakshi and Sumanthi [10] state that they matched 130 of 200 footprint images correctly.

Weingaertner *et al.* [155] achieved on average 65% correct manual pair matching for palmprints; Lemes *et al.* [28] reached 78% correct recognition rate for palmprints as well. Tiwari *et al.* [141] attained 83.67% identification accuracy for the ear.

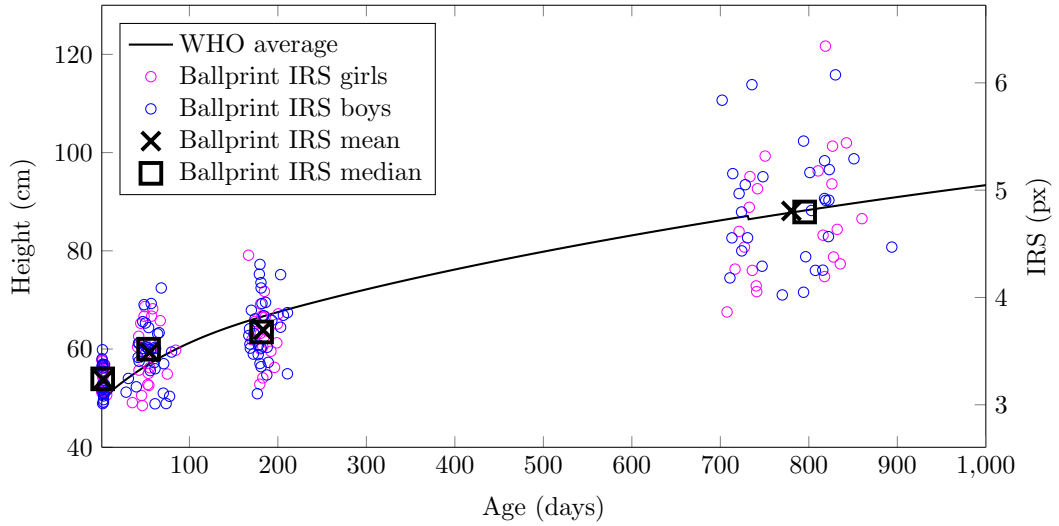
Our result of 19.49% EER has been achieved despite four drawbacks: (i) the drying and cracking of the skin that obscures the ridge lines, (ii) the tininess of the ridge lines which are difficult to capture and in other studies have required the use of high-resolution sensors (1400 ppi [155], 1500 ppi [28], 2400 ppi [29]), (iii) the need for adjustment to suit adult fingerprint characteristics, and (iv) the uncooperative nature of the newborn.

Despite these challenges, we achieved a good inter-age verification result for newborns. We remind the reader that our good results might be due to the small number of usable ballprint captures taken shortly after birth. Nevertheless, our result underlines the usefulness and potential of ballprint as a newborn and infant biometric. It may be used in difficult conditions and may still allow for reliable newborn verification.

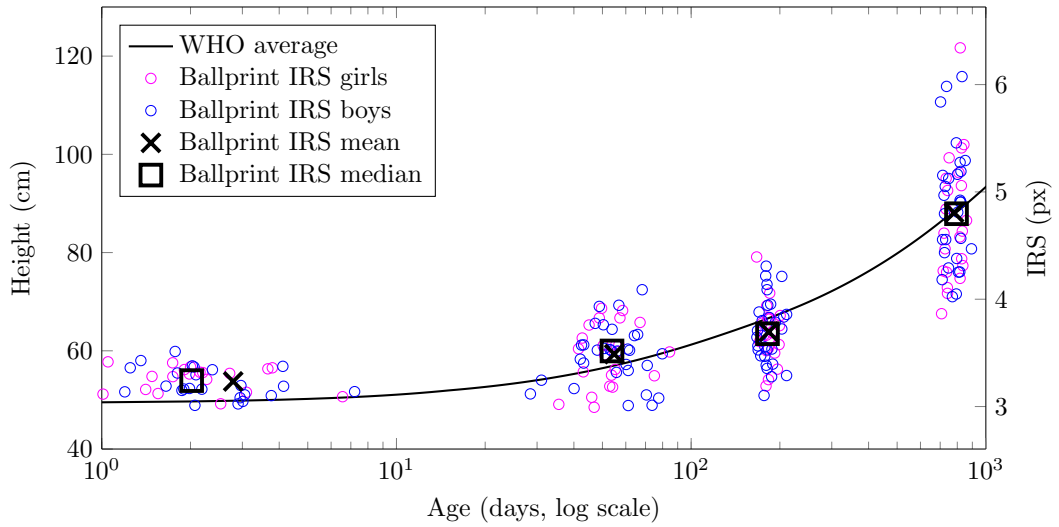
6.7 An external correspondence: IRS and length

We have shown that growth of the ballprint during infancy can be approximated by isotropic scaling. The scaling factor possesses a linear relationship with infant's age for the ages investigated (cf. Sections 6.4.3 and 6.5.1). Clearly, this can only be a first approximation. Therefore the question arises if the inter-ridge spacing is correlated with other physical features which mirror growth as well. Gottschlich *et al.* [47] found this to be the case for median population height and the growth of the fingertip, and Hemy *et al.* [51] found that there is a relationship between physical height and the left foot length. Hence it is worthwhile to investigate the relationship between physical length and IRS. We follow Gottschlich *et al.* and use the growth chart from the World Health Organization (cf. Figure 6.1).

We are limited to a first approximation to growth in IRS due to the size of our database. There are no such limitations associated with the WHO growth chart. Its data show the relationship between an infant's median height or length from shortly after birth up to the age of 5 years in steps of single days (cf. Section 6.3.2). The median is also derived from a much larger sample study involving several countries and ethnic groups [159]. This allows for much greater generality.



(a) Linear scale



(b) Log scale

Figure 6.17: WHO growth chart on (a) a linear age scale and (b) a log age scale overlaid by IRSs measured from the Happy Feet database. The IRS is not limited to the test set but the entire database has been used. We have measured the IRS for each individual at each visit for all images available (regardless of their quality) and the median has been taken per individual and visit.

We overlay the WHO growth chart with our IRS specimens (see Figure 6.17) to investigate the relationship between the two length measures (body length/height and IRS). One can see that the median and mean IRS fit tightly to the curve and that the individual IRS specimens follow the curve well. The correlation coefficient of the median IRS for all four visits and the length obtained from the curve at the median

age per visit is 0.9874. The correlation coefficient of the mean IRS and the lengths at the average age is 0.9855.

Staheli [133, p. 8] states that infancy ends at the age of 2 years and that at this point the infant is on average half the height of its final height during adulthood. The ratio between the median IRSs of the infants aged 2 years and the adults (adjusted for the difference in capture resolution (cf. Section 6.4.3)) follows precisely being $\approx 1/2$:

$$\frac{1}{2} \approx 0.4916 = \frac{5.1\,\text{px} \cdot 600}{12.45\,\text{px} \cdot 500}. \tag{6.19}$$

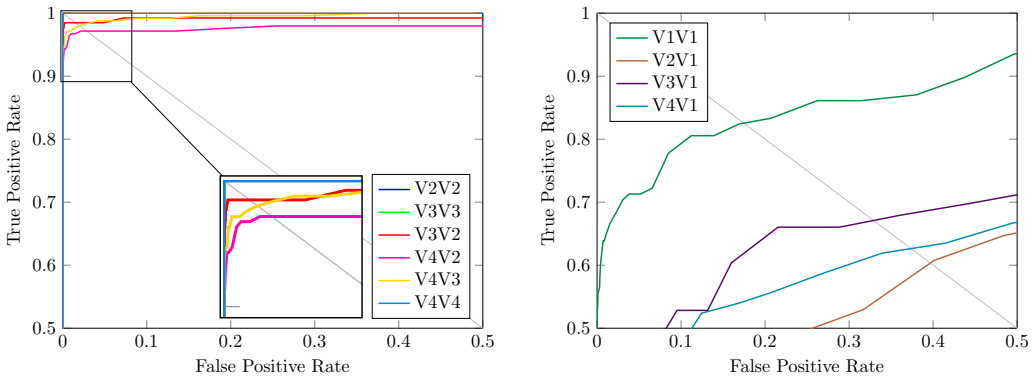
6.7.1 Experiments

Now we perform two verification experiments. Firstly, we repeat the standard verification experiment for the complete test set (see Section 6.5.1) using VF but this time with k_{who} instead of k_{irs}. Secondly, we perform newborn verification for all possible age gaps. We employ VF as extractor and matcher after scaling the images with k_{who}. The prints need to be adjusted to adult fingerprints, so the actual scaling factor is used in reference to the second visit and its resolution correspondence to adult fingerprint we have established in Section 6.4.3.

The scaling factor based on the World Health Organization growth chart for the first visit is $k_{who} \approx 0.87$.

6.7.2 Results

The ROCs obtained by using k_{who} are very similar to the ones we achieved with k_{irs} (see Figures 6.11c and 6.16). Again, complete separation (0 EER) is achieved for all intra-visit comparisons (see Table 6.11). The inter-visit EERs improved slightly with the use of k_{who}. Both scaling factors k_{irs} and k_{who} favour different inter-visit comparisons for newborn verification. The scaling based on the IRS does better when the newborn print is compared to the ones captured at 2 months and 2 years of age. The scaling based on the WHO growth chart favours the newborn verification when the prints of newborns are compared to prints of the same age (V1, V1) or to infants 6 months old (V3, V1).



(a) Infant verification

(b) Newborn verification

Figure 6.18: ROCs for ballprints captured of newborns and infants. VF is employed as extractor and matcher after the prints have been scaled by k_{who}. The corresponding EERs are stated in Tables 6.11 and 6.12.

Method	Equal Error Rate					
	V2, V2	V3, V3	V4, V4	V3, V2	V4, V2	V4, V3
VF, k_{irs}	0.00%	0.00%	0.00%	2.63%	3.27%	2.24%
VF, k_{who}	0.00%	0.00%	0.00%	1.53%	2.85%	2.01%

Table 6.11: EER scores achieved by VF for all possible ballprint age gap comparisons when k_{who} is used for scaling. Neurotechnology Verifinger with k_{irs} is repeated from Table 6.9 to allow for an easier comparison.

Method	Equal Error Rate			
	V1, V1	V2, V1	V3, V1	V4, V1
VF, k_{irs}	19.48%	34.63%	40.14%	36.33%
VF, k_{who}	17.46%	39.67%	32.90%	37.37%

Table 6.12: EER scores achieved by VF for all ballprint ages compared to the newborn captures.

6.7.3 Discussion of IRS-length correspondence

We have established earlier on that the inter-ridge spacing is easily measured for ballprint even on low-quality prints and that its growth mirrors infant age (Figure 6.4). We overlaid the WHO growth curve with our IRS data. The mean and median per visit fit tightly onto the curve and the individual IRSs follow the curve well. The physical length and the IRS are highly correlated.

This is not overly surprising because both features are length measures and hence one dimensional. Nevertheless, we did not expect them to be this highly correlated and thereby being affected by growth so similarly. Physical length is a well-established measure that has been shown to mirror growth in the past. Hence this applies to the ballprint IRS as well, justifies its use, and makes it a valuable feature.

The relationship between the length and the inter-ridge spacing carries over into the scaling factors k_{irs} and k_{who}. They are very similar for infants and even differ only after the second decimal point for newborns. When employed in verification scenarios, both lead to impressive results: they achieve complete separation for all infant intra-visit comparisons and considerably low EERs for inter-visit comparisons. In the most important newborn verification scenario, that is the newborn intra-visit, k_{who} has slightly better chances to prevent the mix-up of newborns.

The reader has to keep in mind that our function to infer the IRS from the age has been tuned for robustness and therefore we applied rounding. The use of exact IRS values could favour k_{irs} in terms of the EER.

This also shows that ballprint adjustment for the use with fingerprint technology is clearly necessary, that impressive results are possible and the scaling factor can be chosen which suits the application the most. A linear approximation might be easier to work with in the field than a large table to look up IRS to age or length/height to age correspondences.

6.8 Discussion

In this chapter we have investigated the ballprint's potential as an infant and even newborn biometric. We have shown that the ballprint and its physical feature structure are similar to the features of the biometric fingerprint modality. Because we were looking at the ballprint in the context of an infant biometric, there are multiple challenges being faced. First, the physical features are much smaller than the features of an adult which represents a technical burden for the capture technology used. Second, infants are uncooperative subjects which makes the capture process difficult in itself but the acquisition of high-quality images becomes even more challenging. Third, the infant's rapid growth affects its physical features drastically.

The physical features, the ridge lines on the ball, are smaller than an adult's ridge lines. They are, however larger and more featured than infant fingerprint ridge lines, though they are also smaller than the ridge lines of an adult fingerprint.

Nevertheless, the latest generation of commercial fingerprint scanners possesses a sensor resolution high enough to capture infant ballprints and their ridge details to some degree. In order to use standard fingerprint matching technology, we found that the infant ballprints need to be adjusted to become more similar to adult fingerprints. The spacing between the ridge lines, the *inter-ridge spacing* turned out to be an excellent indicator for the adjustment necessary. The IRS mirrors the infant growth process which can be approximated as linear for the ages investigated. Furthermore, we have shown that the physical change of the ball due to the infant's growth can be approximated by an isotropic scaling factor. This approximation (with respect to the infant's age) is more than sufficient for a real world verification scenario.

We have demonstrated that standard fingerprint technology can be used to match infant ballprints across different ages up to an age gap of 22 months without any errors. In order to achieve these results, we used isotropic scaling to account for the infant's growth and to adjust the print's inter-ridge spacing to that of an adult fingerprint. We also relied on high-quality ballprint captures with reasonable overlap. Remarkably, we were able to derive a quality estimate for the ballprints based on algorithms built for *adult fingermarks*.

We were able to perform a newborn verification experiment. This scenario is sometimes encountered at hospitals when potential newborn mix-ups need to be prevented or solved. The results for ballprint are in line with what the literature reports for other newborn biometrics.

Last, we demonstrated there is a relationship between an infant's inter-ridge spacing and its length/height, both length measures are highly correlated. The length/height data are obtained from an external growth chart of the WHO and the IRS was measured on our ballprint database. It is clearly observed that physical height mirrors growth which again justifies the employment of the IRS as a feature from which to estimate ballprint and infant growth.

Therefore we conclude that ballprint is a suitable infant, even newborn, biometric that can be acquired and matched with standard fingerprint technology if adjusted accordingly to the infant's age. One has to keep in mind that a proper capture protocol needs to be in place that ensures the repetitive capture of the same ballprint area (especially as the child grows) and high-quality captures (re-capture if image is of insufficient quality) to enable matching.

This is a very constrained scenario. Unfortunately, an infant's age is not always known. In the context of a vaccination program in a developing country, it is possible that infants or children without any known (vaccination) history present and are unable to

identify themselves. Hence their age remains unknown. Another issue is malnutrition because it slows down growth and thus the age may not correspond with the infant's physique. We are going to investigate these challenges in the next chapter.

6.9 Work added since first submission

In the process of preparing this chapter for publication, we re-evaluated our choice of 250 ppi as base resolution for V2 (cf. Section 6.4.1). We used a fine grid search around 250 ppi in steps of 1 ppi. We chose the resolution that lead to high correlation between the spatial position of the minutiae marked manually (see Section 6.4.2) and those extracted automatically by Neurotechnology Verifinger, and that showed the lowest variance over all specimens of the percentage of correctly extracted minutiae. This is 261 ppi.

6.9.1 Experiments

Afterwards we repeated the infant and newborn verification experiments for the scaling factors k_{irs} and k_{who}. The scaling factors remain unchanged and are stated in Table 6.13. The corresponding suggested capture resolutions are mentioned in Table 6.14. We performed the experiments for the most promising extractor and matcher only and repeated the experiment on the high-quality test set for k_{irs} only as we did before.

Scaling factor	V3V2	V4V2	V2V1
k_{irs}	1.10	1.50	0.87
k_{who}	1.17	1.52	0.87

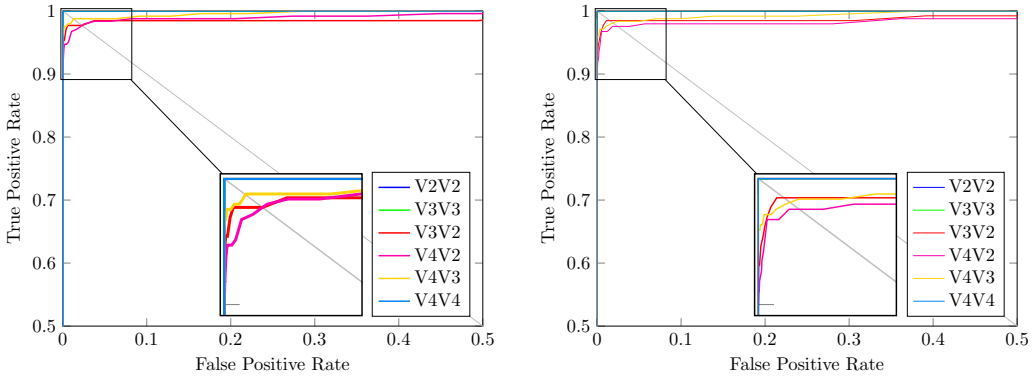
Table 6.13: The scaling factors k_{irs} and k_{who} as stated in Table 6.6 and Section 6.6.1. The scaling factors to scale the prints from the later visit down to the size of the earlier one differ after the second decimal point which results in different suggested capture resolutions (cf. Table 6.14).

Scaling factor	V1	V2	V3	V4
k_{irs}	228.1 ppi	261.0 ppi	287.1 ppi	391.5 ppi
k_{who}	227.1 ppi	261.0 ppi	305.4 ppi	396.7 ppi

Table 6.14: Suggested capture resolutions for the scaling factors k_{irs} and k_{who} based on Table 6.13. The resolutions for V1 are different because the scaling factors differ after the second decimal point (see Table 6.13).

6.9.2 Results

The new base resolution leads to slightly improved results. When used with the complete test set, the inter-visit verification EERs are as good or better in every case.



(a) k_{irs} (b) k_{who}

Figure 6.19: ROCs for ballprints captured of infants after fine-tuning the base resolution. VF is employed as extractor and matcher after the prints have been scaled by (a) k_{irs} and (b) k_{who}. The corresponding EERs are stated in Table 6.15.

The ROCs are shown in Figure 6.19; the previous and the new EERs are documented in Table 6.15.

	Equal Error Rate					
Method	V2, V2	V3, V3	V4, V4	V3, V2	V4, V2	V4, V3
VF, k_{irs}	0.00%	0.00%	0.00%	2.63%	3.27%	2.24%
VF$^+$, k_{irs}	0.00%	0.00%	0.00%	2.29%	2.30%	1.29%
VF, k_{who}	0.00%	0.00%	0.00%	1.53%	2.85%	2.01%
VF$^+$, k_{who}	0.00%	0.00%	0.00%	1.53%	2.44%	1.90%

Table 6.15: EER scores achieved by VF before (VF) and after (VF$^+$) the resolution fine-tuning for all possible ballprint age gaps. The previous results (VF) are taken from Table 6.11.

When a third of the images classified as having the lowest quality is removed, error-free verification is achieved in all cases. This is demonstrated in Figure 6.20.

We also repeated our newborn experiments. All inter-visit comparisons for k_{irs} improved; one by nearly 29% to become as low as 28.53% due to the new base resolution. Most of the inter-visit comparisons for k_{who} improved as well. This is not the case for the intra-visit comparison, where the EERs increased. We present the ROCs in Figure 6.21 and the corresponding EERs in Table 6.16.

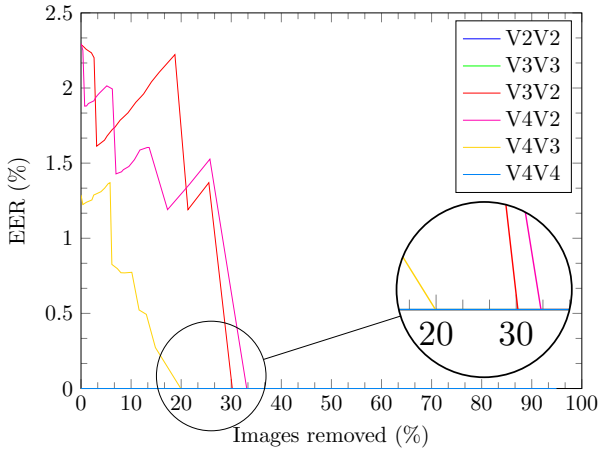
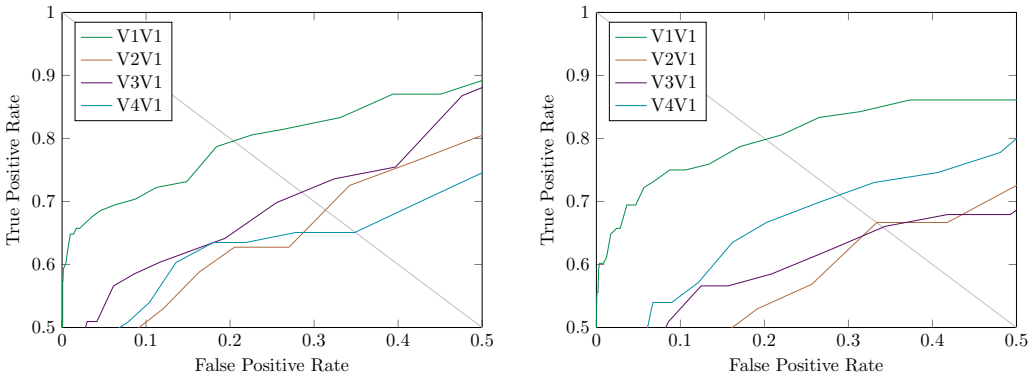


Figure 6.20: This diagram illustrates how the performance, in terms of the Equal Error Rate, improves when low-quality images are removed from the test set. The image quality of the infant ballprints is derived from adult fingermarks (cf. Figure 6.9).



(a) k_{irs} (b) k_{who}

Figure 6.21: ROCs for ballprints captured of newborns after fine-tuning the base resolution. VF is employed as extractor and matcher after the prints have been scaled by (a) k_{irs} and (b) k_{who}. The corresponding EERs are presented in Table 6.16.

Method	Equal Error Rate			
	V1, V1	V2, V1	V3, V1	V4, V1
VF, k_{irs}	19.48%	34.63%	40.14%	36.33%
VF$^+$, k_{irs}	20.41%	31.36%	28.53%	34.89%
VF, k_{who}	17.46%	39.67%	32.90%	37.37%
VF$^+$, k_{who}	20.15%	33.36%	34.10%	28.96%

Table 6.16: EER scores achieved by VF before (VF) and after (VF$^+$) the resolution fine-tuning for all possible ballprint age gaps. The previous results (VF) are taken from Table 6.12.

6.9.3 Discussion

The fine-tuning of the base resolution lead to improved verification results for newborns and infants. There are two main points. Firstly, these results emphasise our previous argument about the importance of proper and thorough adjustment of adult fingerprint technology when used for infant ballprints. Secondly, the case for ballprint being a suitable infant biometric is strengthened because we demonstrated that scaling is effective using the ballprint IRS and that error-free verification is possible up to an age gap of 22 months.

7 Ballprints for infants of unknown age

This chapter continues to investigate the ballprint as an infant biometric. In the previous chapter we assumed that an infant's age is known and that it is well nourished. So, there is a correspondence between the size or growth of its physical features and its age. We extensively relied on this relationship between the *inter-ridge spacing* (IRS) and age in order to resize the images captured and to increase their similarity to adult fingerprints.

This is however a constrained scenario. Infants without any form of (vaccination) documentation or national ID are unable to identify themselves [14]. Hence their age and identity is unknown. This causes many problems, for instance in the context of vaccine delivery for the developing world. In order to increase the vaccination coverage without being wasteful, it becomes valuable to link the infant's or child's vaccination record to a physical trait.

Another potential problem that may seen in a developing country is that the child might suffer from malnutrition which can induce stunted growth when experienced over a long timespan [160]. The consequence would be that an individual's growth slows down and its physical features are smaller when compared to a well nourished child at the same age. Jain *et al.* work towards the improvement of vaccination coverage using fingerprint as an infant biometric [63]. Nevertheless, stunted growth and dealing with children of unknown age represent new territory that has not been covered in the literature.

A manuscript on BGM and sBGM has been drafted and is planned to be submitted as a journal paper. It involves the Sections 7.1.1, 7.1.2, 7.2, 7.3.1 and 7.3.2.

We attempt to solve this extended scenario in three ways. Firstly, we investigate if the infant ballprint can be represented by novel features which are scale-invariant and to derive novel templates. Secondly, we research the effects of image scaling with respect to a dimensionless unit such as the normalisation of the *inter-ridge spacing* (IRS) by a specific value. Thirdly, we describe an approach that transforms the unknown-age problem into a known-age problem.

The chapter is organised as follows. Firstly, in the Methodology section we introduce *Biometric Graph Matching* (BGM) for ballprints and extend it to use novel scale-invariant features (Sections 7.1.1 and 7.1.2), secondly, we describe the process to normalise infant ballprint captures with respect to the inter-ridge spacing (Section 7.1.3), and thirdly, we explain the approach to break the unknown-age problem down to a known-age problem (Section 7.1.4).

We continue to describe our experiments and their assumed scenarios (Section 7.2), before we present the results obtained (Section 7.3) and discuss them (Section 7.4), and lastly, we explain the challenges faced, the limitations encountered and present our recommendations (Section 7.5).

7.1 Methodology

7.1.1 Biometric graph matching

Biometric Graph Matching (BGM) [56] is an attractive approach here because (i) it demonstrated competitive performance on other biometrics, (ii) it uses minutiae connectivity as a feature which may help in dealing with growth effects, and (iii) it is completely disclosed. BGM builds upon the basic definition stated in Section 3.2.1 in Chapter 3 and has been applied successfully to fingerprint [56], retina [84], hand vein [85], and palm vein [7].

Following the definition in [84], the graph $\boldsymbol{G}$ is defined by $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E}, \mu, \upsilon)$, where the vertex set $\boldsymbol{V}$ is the set of feature points extracted from the ballprint image or its skeleton and the edge set $\boldsymbol{E} \subset \boldsymbol{V} \times \boldsymbol{V}$ is the set of feature points which are directly connected by a ridge line without any (real or false) minutiae in between. The vertex labelling function $\mu : \boldsymbol{V} \to \mathbb{R}^2$ maps each vertex υ onto its Cartesian coordinates (x, y) in the captured ballprint image and the edge labelling function $\upsilon : \boldsymbol{E} \to \mathbb{R}^2$ maps each edge e onto the value pair (l, θ), where l is the Euclidean distance between the two feature points it connects (i.e. the length of a direct line between those two points) and θ represents the line's slope. The graph is undirected.

The general idea of graph matching has been introduced in Section 3.2.1 in Chapter 3. To compute a matching score between two ballprints, their graphs $\boldsymbol{G}$ and $\boldsymbol{G}'$ are extracted, registered and matched as described below. We follow the approach introduced by Lajevardi *et al.* [84].

Graph extraction

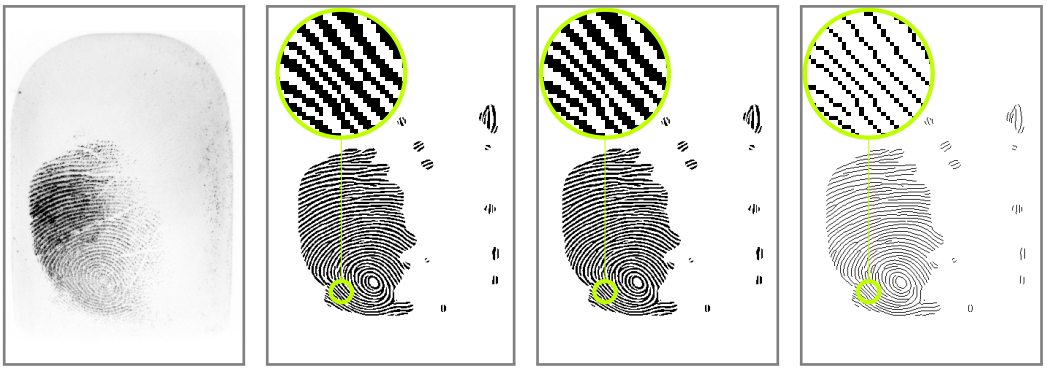
Extracted features are the minutia's spatial coordinates (x, y), type, orientation and quality. Additionally, connectivity is recorded by edges and the number of ridges crossed by a straight line drawn between any two minutiae (see Section 7.1.2 below). Extracted graphs projected onto the corresponding ballprint captures are illustrated in Figure 7.6.

A *Ridge Line Tracer* (RLT) was developed to extract graphs from ridge-based images. This research's focus is neither the actual graph extraction or the image processing involved in this. Therefore Neurotechnology Verifinger has been used to obtain a binarised image, the minutiae and their properties (Cartesian coordinates, type, orientation, and quality). Verifinger comes with disadvantages when applied to ballprints: the program is optimised to deal with adult fingerprints captured at a resolution of 500 ppi. The fingerprint scanner NEC PU900-10 captures images with 1000 ppi but the resulting output is a 320×480 image with 508 ppi. The problem is that the fine ridge line structures in the ballprint are not picked up by the software if applied to it at the image's actual resolution. The solution is to dilate the image using nearest neighbour interpolation; this increases the distance between ridges and Verifinger can be used as usual. The scaling factor used to do this was the subject of Chapter 6.

The binarised image $\boldsymbol{I}$ has n columns and m rows. Because of the scanner's downsampling process and the physical tininess of infant ridge structures, $\boldsymbol{I}$ tends to have false connections between ridge lines.

These are introduced by too much pressure during the capture and (perhaps) by Verifinger's algorithm. These false connections consist of one diagonal pixel that is perpendicular to actual ridges (cf. Figures 7.1 and 7.3), i.e. there are always two pixels involved. In order to remove these connections a two step process has been developed:

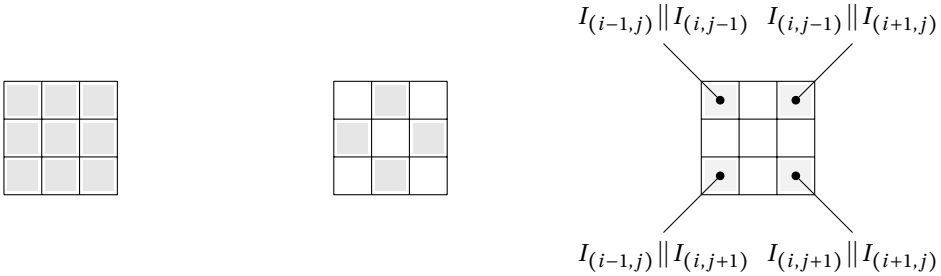
1. Find all candidate pixels.
2. Remove at least one of the pixels without breaking the actual ridge apart.



(a) Captured image (b) Verifinger output I (c) Removal of false connections (d) Thinned image I'

Figure 7.1: Visualisation of the different enhancement stages. The first specimen of baby 009's left ballprint at the second visit (`BS_009_V2_L_1`) is used. Firstly, the captured image is processed by Verifinger at 225 ppi. Secondly, false connections are removed from the result using the described technique. Thirdly, thinning is performed. The border indicates the actual image size as the fingerprint scanner outputs it.

Candidate ridge line pixels A pixel $I_{(i,j)}$ is considered to be a real part of a ridge line if it is connected only by 8-Neighbourhood (N_8, cf. Figure 7.2a) but not by 4-Neighbourhood (N_4, cf. Figure 7.2b). To evaluate the connectivity, the matrix $\boldsymbol{I}_{(i-1:i+1,j-1:j+1)}$ holds the 3×3 area around pixel $I_{(i,j)}$.



(a) N_8: 8-Neighbourhood (b) N_4: 4-Neighbourhood (c) N_c: Actual connectivity

Figure 7.2: Schematic representation of pixel neighbourhoods: (a) 8-neighbourhood, (b) 4-neighbourhood and (c) actual connectivity. The corner pixels of the actual connectivity are `true` as soon as they are connected via N_4 to the centre pixel.

As soon as Equation 7.1 is true for at least one matrix element the pixel under investigation is a candidate:

$$\left[\sum_{k=1}^{3}\sum_{l=1}^{3}\left(I_{(i-1:i+1,j-1:j+1)} \circ N_8 \circ !N_4 \circ !N_c\right)_{k,l}\right] > 0 \tag{7.1}$$

The $\circ$ operator represents the Hadamard product, where ! stands for a binary inversion, and || is the logical *or*. Every candidate is further investigated if it can be considered as stable. This is the case if the pixel is connected to at least 2 other pixels via 4-Neighbourhood. Hence its removal would not break a thin ridge apart under any circumstances.

$$\left[\sum_{k=1}^{3}\sum_{l=1}^{3}\left(I_{(i-1:i+1,j-1:j+1)} \circ N_4\right)_{k,l}\right] \geq 2 \tag{7.2}$$

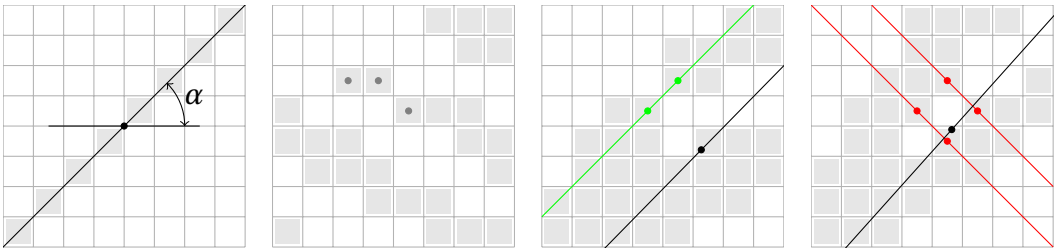
Removal of false ridge line connectivity Usually, only two candidates are connected by 8-Neighbourhood but there might be more if the ridge spacing is small due to internal scaling (cf. Section 7.1.1). However, all diagonally connected candidate pairs e.g. $I_{(i,j)}$ and $I_{(i',j')}$ containing at least one stable pixel are further investigated. The actual structure containing the problem pixels within a window of size $p \times p,\ p = 10$ centred around the rounded candidate pixels' centroid coordinates and the potential ridge breaking direction are taken into account. Therefore the main structure's orientation $\alpha \in [-90, 90]$ is estimated by computing its eigenvectors and eigenvalues on the basis of its x and y-coordinates. The candidate pixels' orientation β to each other is calculated as well (cf. Equation 7.4).

$$\beta' = \arctan\left(\frac{j-j'}{-i+i'}\right) \cdot \frac{180}{\pi},\ \beta' \in \{-135, -45, 45, 135\} \tag{7.3}$$

$$\beta = \begin{cases} \beta' & |\beta'| \leq 90 \\ \beta' - 180\,\mathrm{sgn}(\beta') & \text{otherwise} \end{cases} \tag{7.4}$$

If and only if the difference between the two pixels' and the structure's orientation is closer to being perpendicular than being parallel i.e. fulfils $|\beta - \alpha| > 45$, the pixels are considered a false connection; all stable pixels are removed by being set to `false`. After this has been performed for all candidate pixels the outcome is thinned and the resulting image is named $\boldsymbol{I}'$.

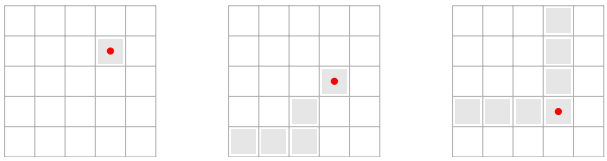
Finally, three post-processing steps are performed to enhance the tracing process and minimise the number of false minutiae. The binary image is (i) thinned using morphological operations until the output stops changing [86], (ii) pixels without any further connections and spurious pixels are removed and (iii) the centre-pixel of "L"-like structures is removed to ensure N_8 and prevent N_4 connectivity. Examples of



(a) Example structure and its computed angle (b) Unstable ridge line pixels (border pixels ignored) (c) Preserving "correct" connectivity (d) Removing "false" connectivity

Figure 7.3: Examples of the binary enhancement process; Figures 7.3b to 7.3d are cropped parts of the Figure 7.1b. The ridge structure's centre of mass is highlighted by a black circle whereas the black line indicates its orientation. Unstable candidate pixels are indicated by grey circles and are going to be kept to preserve ridges and to avoid breaking them apart (cf. Figure 7.3b). The green or red coloured circles and lines show stable candidate pixels and their orientation. Green symbolises that the structure is preserved and the pixels are going to be kept (Figure 7.3c: $|\beta - \alpha| = |45.5407 - 45| = 0.5407$); red emphasises their future removal (Figure 7.3d: $|\beta - \alpha| = |48.1201 + 45| = 93.1201$).

binary structures and their post processing step are shown in Figure 7.4.



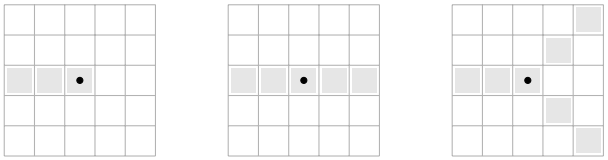
(a) Lonely pixel (b) Spurious pixel (c) "L"-like structure

Figure 7.4: Examples of the binary structures and their post processing step; the pixel marked by a red dot is going to be removed.

Finding candidate minutiae The candidate minutiae are chosen from the skeleton via their connectivity to other pixels, the following three cases are possible:

1. One other pixel: Ridge ending (Figure 7.5a),
2. Two pixels: Part of a ridge line – no minutia (Figure 7.5b),
3. More than two pixels: Bifurcation (Figure 7.5c) – only the centre is kept as a candidate point.

This set of candidate points is combined with Verifinger's output. Connected minutiae are fused as these little inaccuracies are due to the thinning process.



(a) Ridge ending (b) Ridge line (c) Bifurcation

Figure 7.5: Examples of pixel connectivity and the inferred minutia type for the pixel in question (black dot).

Local quality rejection Some minutiae are rejected based on Verifinger's quality value for this particular minutia. If it is smaller than threshold ϵ, it is rejected and removed from the candidate set.

The tracing The tracing process starts from a randomly chosen minutia and continues to follow all connected ridge lines separately until another minutia is encountered close by (Euclidean distance less or equal than 2). The start and the end minutiae are recorded as being connected in an adjacency matrix. The tracing assigns a unique but consecutive ID to every ridge line pixel. This ensures all ridge line pixels have been visited and simplifies further feature extraction and computation.

Minutiae located next to each other are due to artefacts and the skeleton's structure. They are joined after the tracing step and their individual features such as the ridge line length are combined and correspondence between the minutiae present in the skeleton structure (cf. Figure 7.5) and Verifinger is established. All minutiae derived from the skeleton or found by Verifinger without correspondence between each other are removed from the minutia list and the adjacency matrix. A minutiae has to be picked up in the skeleton and by Verifinger to remain in the minutia list and the adjacency matrix. Therefore, there exists only *one* edge between any two minutiae if and only if they are connected by a ridge line and there has not been a "false" minutia on the ridge in between, which has been removed and thus there is no edge. Extracted graphs projected onto the corresponding ballprint captures are shown in Figure 7.6.

Graph registration

In order to perform the comparison of two different ballprint graphs $\boldsymbol{G}$ and $\boldsymbol{G}'$, they need to be roughly aligned. We use edges to do so.

All edges from the edge sets $\boldsymbol{E}$ and $\boldsymbol{E}'$ are assessed pairwise regarding their similarity. We use the Euclidean distance of the difference of the edge length and slope (normalised by the average length) to determine their similarity. There is no assumption of rough alignment, because we are relying on the slope difference between the two edges. This results in the dissimilarity matrix $\boldsymbol{D}^{(|\boldsymbol{E}|\times|\boldsymbol{E}'|)}$.

$$d_{i,j} = \frac{2}{l_i + l'_j}\sqrt{(l_i - l'_j)^2 + (\theta_i - \theta'_j)^2}, \tag{7.5}$$

where $e_i \in \boldsymbol{E}$ and $e'_j \in \boldsymbol{E}'$, l_i and l'_j is the edge length, and θ_i and θ'_j is the edge slope (in degrees).

Now the τ edge pairs with the lowest dissimilarity are chosen for a RANSAC approach to determine the best alignment. Let us say that the edge pair consists of e_i and e'_j with their corresponding labels and that the edge e_i connects the two vertices u and v. Now all vertices in $\boldsymbol{V}$ are translated by x_u and y_u so that u becomes the origin. Furthermore, we rotate all nodes *clockwise* by θ_i so that both vertices u and v are now aligned on the x-axis. This is performed for both graphs and their respective edge of the edge pair.

Afterwards, the alignment of both graphs is scored by the total number of corresponding vertices. Two vertices are considered to correspond if their Euclidean distance is smaller than a given threshold t. Each vertex can only correspond to one other vertex of the other graph. This is ensured using a greedy algorithm. The alignment responsible for the highest count is assumed to be correct and applied to the respective graphs.

Graph matching

As the two ballprint graphs $\boldsymbol{G}$ and $\boldsymbol{G}'$ are now aligned, we can perform error-tolerant graph matching. We follow Riesen und Bunke who demonstrated [123] how to use Munkres' optimisation algorithm [111] for graph matching. This technique has been successfully applied to biometric graphs [7, 56, 84, 85].

The idea is to use a smart matrix representation, the cost matrix $\boldsymbol{C}$, to represent competing graph edits so that Munkres' algorithm can be applied. Let us assume that $\boldsymbol{G}$ and $\boldsymbol{G}'$ contain $m = |\boldsymbol{V}|$ and $m' = |\boldsymbol{V}'|$ vertices, respectively, then $\boldsymbol{C}^{((m+m')\times(m+m'))}$ is square (cf. Equation 7.7).

Every matrix element $c_{(i,j)}$ represents the cost for a certain graph edit (substitution, deletion, insertion) to transform $\boldsymbol{G}$ into $\boldsymbol{G}'$. More specifically, the cost matrix $\boldsymbol{C}$ is split

into four quadrants:

$$\left[\begin{array}{c|c} \boldsymbol{C}_1 & \boldsymbol{C}_2 \\ \hline \boldsymbol{C}_3 & \boldsymbol{C}_4 \end{array}\right]. \tag{7.6}$$

The first quadrant $\boldsymbol{C}_1$ represents the cost for each vertex v_i being substituted by v'_j, where $i \in \{1,\dots,m\}$ and $j \in \{1,\dots,m'\}$. Both sub-matrices $\boldsymbol{C}_2$ and $\boldsymbol{C}_3$ are square and contain the cost associated with the deletion ($\boldsymbol{C}_2$) or insertion ($\boldsymbol{C}_3$) of a specific vertex. All elements along the diagonal contain the actual cost, all others are ∞. The fourth quadrant $\boldsymbol{C}_4$ is all zero.

$$\left[\begin{array}{ccc|ccc} c_{(1,1)} & \cdots & c_{(1,m')} & c_{(1,m'+1)} & \cdots & \infty \\ & & & \infty & & \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ & & & & & \infty \\ c_{(m,1)} & \cdots & c_{(m,m')} & \infty & \cdots & c_{(m,m'+m)} \\ \hline c_{(m+1,1)} & \cdots & \infty & 0 & \cdots & 0 \\ \infty & & & & & \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ & & \infty & & & \\ \infty & \cdots & c_{(m+m',m')} & 0 & \cdots & 0 \end{array}\right] \tag{7.7}$$

Cost function The choice of cost function affects BGM's efficiency greatly. Here we use the Euclidean distance $\|v_i, v'_j\| = \sqrt{(x_i - x'_j)^2 + (y_i - y'_j)^2}$ between two vertices v_i and v'_j and incorporate their difference in connectivity. For this purpose, let n_i be the degree of vertex v_i and l^*_i the average length of all edges it is connected to. Then the cost function for substitution is given by:

$$c_{(i,j)} = \|v_i, v'_j\| + \min(c_{id}, \Upsilon_{(i,j)}), \text{ with } c_{id} \in \mathbb{R}^+. \tag{7.8}$$

$$\Upsilon_{(i,j)} = \chi c_{id} |n_i - n'_j| \sqrt{|l^*_i - l'^*_j|} \tag{7.9}$$

So we use the Euclidean distance between two vertices and add extra cost (never more than c_{id}) being the difference in degree scaled by the factor $\chi \in [0,1]$, c_{id}, and the square root of the difference of the average edge length connected to this particular vertex.

The cost for deletion and substitution can but does not have to take edge connectivity into account. We chose to do so, so then the deletion cost is given by

$$c_{(i,m'+i)} = c_{id}(1+\chi n_i). \tag{7.10}$$

Similarly, the insertion cost is

$$c_{(m+j,j)} = c_{id}(1+\chi n'_j). \tag{7.11}$$

The scaling parameters c_{id} and χ have to be chosen carefully because if they are too low the algorithm inserts and deletes vertices instead of substituting them. This behaviour would lead to low matching scores, even for genuine comparisons, as the ballprint capture and the graph extraction process are biased by quality distortions (cf. Section 2.5.1 in Chapter 2).

After the creation of the cost matrix $\boldsymbol{C}$, we apply Munkres' optimisation algorithm, which results in a modified binary cost matrix $\boldsymbol{C}'$. The only interesting part of this matrix is the first quadrant $\boldsymbol{C}'_1 = \boldsymbol{C}'_{(1:m,1:m')}$ as it indicates which vertices have been substituted. Hence one can count the non-zero entries to find the number of corresponding vertices.

Maximum common subgraph The *Maximum Common Subgraph* (MCS) is defined as the subgraph that the graphs $\boldsymbol{G}$ and $\boldsymbol{G}'$ have in common. We refer to it as $\text{mcs}(\boldsymbol{G},\boldsymbol{G}')$. It is given by the substitutions performed in $\boldsymbol{C}'$. Only the vertices that have been substituted are part of the MCS. An edge is only included in the MCS if both vertices it connects in $\boldsymbol{G}$ have been substituted and if there is an edge between these two vertices in $\boldsymbol{G}'$ as well. The total number of vertices and edges included in the MCS is given by $|\text{mcs}(\boldsymbol{G},\boldsymbol{G}')|$ and $|\text{mcs}(\boldsymbol{G},\boldsymbol{G}')|_e$, respectively.

Matching score The final score s between the two ballprints is given by a combination of their similarity in terms of vertices (d_v) and in terms of edges (d_e).

$$d_v = \frac{|\text{mcs}(\boldsymbol{G},\boldsymbol{G}')|}{\sqrt{|\boldsymbol{V}|\cdot|\boldsymbol{V}'|}} \tag{7.12}$$

$$d_e = \frac{|\text{mcs}(\boldsymbol{G},\boldsymbol{G}')|_e}{\sqrt{|\boldsymbol{E}|\cdot|\boldsymbol{E}'|}} \tag{7.13}$$

Both similarity measure scores are fused via a weighted mean to the final matching score:

$$s = \gamma d_v + (1-\gamma) d_e. \tag{7.14}$$

7.1.2 Scale-invariant biometric graph matching

We refer to Biometric Graph Matching that only uses scale-invariant features as *scale-invariant BGM* (sBGM). The use of scale-invariant features should allow us to skip the growth estimation or scaling process because the features are the same regardless of their scale.

Scale-invariant features

To deal with the unknown age of a child we investigate the use of (physical) features of the ballprint which remain persistent over time and are not affected by growth and the change of scale. These features assume that the pattern of ridge lines does not change (apart from its scale) and remains persistent over time.

The ballprint of an infant is challenging to capture due to its uncooperative nature. In particular, it is difficult to capture the same area of the ball reliably and accurately over time, especially when the child is upset or playful and moves its feet constantly. An additional challenge is imposed if a fingerprint scanner is used to capture the ballprint. Their sensor area is optimised for the size of an adult fingertip and not an infant ballprint. This means that while the child grows, its physical features undergo the same process and eventually outgrow the sensor area.

Therefore, scale-invariant features should not rely heavily on any particular point as it is possible that it is not included in the print captured. Figure 2.10 in Chapter 2 and Figure 6.3 in Chapter 6 illustrate this problem.

Considering the similarity between ballprint and fingerprint (cf. Section 6.2 in Chapter 6), we decided to use two features that are often used for fingerprint: (i) the minutia orientation and (ii) the minutia type (ridge ending or bifurcation). Additionally (iii), we chose the *ridge line connectivity c*, a binary measure which states if two minutiae are directly connected by a ridge line. Lastly (iv) we selected the *Ridge Line Count* (RLC) to measure the distance between minutiae.

Ridge Line Count The *Ridge Line Count* (RLC) is a growth independent distance measure. The underlying idea is that the physical structure of an individual's friction ridge pattern does not change over time and remains persistent. Let us choose two minutiae M_i and M_j randomly. The RLC between these two minutiae M_i and M_j is described by the number of other ridge lines intersected when these two points are connected by a straight line. The distance is stored in D at $D_{i,j}$, $i < j$. It is a symmetric and non-negative measure; hence there are only entries above the main diagonal in D during the computation.

In order to keep the computational complexity low, Xiaolin Wu's line drawing algorithm [163] has been chosen as a fast approximation for drawing the line from M_i to M_j, instead of working with the line's slope at a high resolution. The thinned ridge lines are quantized approximations of their original continuous locations. As described in Section 7.1.1 every pixel on a ridge line has a unique number unless it is a minutia, then it is a negative value stating the number of joining or bifurcating lines.

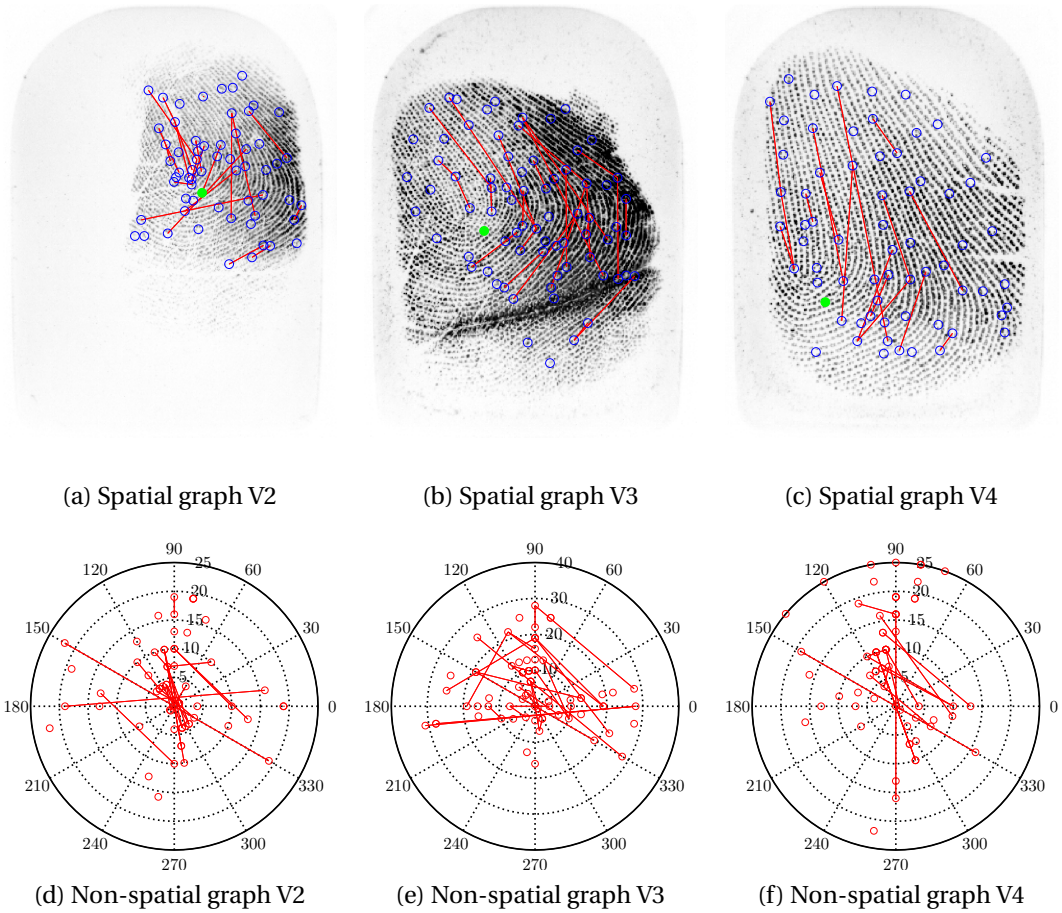
However Wu's algorithm uses the fact that a line almost always interacts with two pixels perpendicular to its direction. It calculates a normalized gray value $g \in [0,1]$ for both of them (their sum equals 1). We consider the pixel with the higher value to be the actual line where the other pixel contains additional information. It is used to confirm that zero-entries are not caused by the skeleton quantisation but because the ridge line has been crossed.

The RLC is obtained by counting the number of local peaks on the drawn line and negative values if they are *not* connected to other values greater than zero. These peaks and negative values are respectively ridge lines and minutiae which have been assigned individual values during the tracing process (cf. Section 7.1.1).

Graph definition

Let $\boldsymbol{G}^* = (\boldsymbol{V}, \boldsymbol{E}^*, \mu^*, \upsilon^*)$, where the vertex set $\boldsymbol{V}$ is the set of scale-invariant feature points extracted from the ballprint image or its skeleton and the edge set $\boldsymbol{E}^* \subset \boldsymbol{V} \times \boldsymbol{V}$ is the set of all possible distinct pairs of feature points. The function $\mu^* : \boldsymbol{V} \to (q, \alpha, t) \in \mathbb{R}^3$ maps each vertex υ to its image quality measure q, the minutia's orientation α and type t it is based on. The edge labelling function $\upsilon^* : \boldsymbol{E}^* \to (c, r) \in \mathbb{R}^2$ maps each edge e between two vertices to the binary value c which is 1 if and only if the minutiae the vertices correspond to are directly connected in the ballprint skeleton and there are no false minutiae in between (cf. Section 7.1.1); it also maps to the RLC r between these two minutiae (cf. Section 7.1.2). The graph is undirected. Examples of the spatial and the non-spatial graphs across different ages for the same ballprint are shown in Figure

7.6.



(a) Spatial graph V2 (b) Spatial graph V3 (c) Spatial graph V4

(d) Non-spatial graph V2 (e) Non-spatial graph V3 (f) Non-spatial graph V4

Figure 7.6: Extracted spatial graphs (a, b, c) projected onto the ballprint captures of infant 019's left foot and their non-spatial representation (d, e, f). The blue circles (and one green circle) represent the vertices, the red lines the connecting edges. All non-spatial representations use the same minutia (highlighted in green) as their centre vertex to align on. Furthermore only edges with $c = 1$ are represented.

Let us assume that there is a graph $\boldsymbol{G}^*$ with m vertices, which satisfies the definition above. Because of the way the edges between vertices are defined, the graph is complete. Hence, there are $n = {}^{(m^2-m)}/_2$ edges. We will reference vertex specific attributes by one index and edge specific attributes with two indices indicating the two corresponding vertices. Hence α_i is the minutia orientation associated with v_i via μ^* and $r_{(i,j)}$ is the Ridge Line Count between the vertices v_i and v_j using v^*, with $i, j \in 1, \ldots, m$, $i \neq j$.

The main difference from spatial graphs (cf. Section 7.1.1) is that the scale-invariant variation uses non-spatial features and a polar coordinate representation (θ, ρ). There-

fore, they heavily depend on the centre vertex chosen. This is indicated by an index a: $\boldsymbol{G}_a^* = (\boldsymbol{V}, \boldsymbol{E}, \mu_a, \upsilon_a)$. Hence, the angle α needs to be re-defined in the context of the centre vertex:

$$\theta_i = \alpha_i^a = \min\{\alpha_i - \alpha_a, 360 + \alpha_i + \alpha_a, 180 + \alpha_i - \alpha_a, 180 + \alpha_i + \alpha_a\}. \tag{7.15}$$

The distance ρ is given by the Ridge Line Count between the two vertices adjusted so that only the centre vertex remains at the origin.

$$\rho_i = r_{(a,i)} + 1 \tag{7.16}$$

Graph registration

The graph registration process is a RANSAC-like procedure performed between the graphs $\boldsymbol{G}_a^*$ and $\boldsymbol{G}_{a'}^{*\prime}$ for every pairwise vertex combination possible between the two graphs (the choice of a and a'). It results in a list L of length l. It contains the alignments, which achieved the highest *Quick Score* (QS) (see below) and hence genuine comparisons are more likely to register on a truly common structure. This is essential for the graph matching performed afterwards. We assume that graphs $\boldsymbol{G}_a^*$ and $\boldsymbol{G}_{a'}^{*\prime}$ have m and m' vertices and n and n' edges, respectively.

Every alignment (pairwise vertex combination possible) is scored via QS, a quick and robust scoring method, in order to obtain the best alignment possible between $\boldsymbol{G}_a^*$ and $\boldsymbol{G}_{a'}^{*\prime}$ in terms of what centre vertex to choose (a and a'). The alignment is based on the minutia orientation and RLC of the centre vertex chosen.

Quick Score – variable nearest neighbours-based alignment This procedure is entirely node based and needs to be computed for all possible vertex pairs υ_a and $\upsilon'_{a'}$ between the two graphs $\boldsymbol{G}_a^*$ and $\boldsymbol{G}_{a'}^{*\prime}$. It focuses on the local neighbourhood structure of the centre vertex. The number of vertices included in both neighbourhoods is not required to be identical.

Let us assume that there are the vertices $\boldsymbol{V}_t = \{\upsilon_b, \upsilon_c, \upsilon_d\} \subseteq \boldsymbol{V}$ and $\boldsymbol{V'}_t = \{\upsilon'_{b'}, \upsilon'_{c'}, \upsilon'_{d'}, \upsilon'_{e'}\} \subseteq \boldsymbol{V'}$.

The ridge line count between any vertex included in $\boldsymbol{V}_t$ and the centre vertex υ_a is smaller or equal to $t_{rlc}+1$; the same holds for a' and $\boldsymbol{V}'_t$. The threshold t_{rlc} is iteratively increased if and only if one of the subsets remains empty to ensure that at least one

vertex is included.

$$\rho_i \leq t_{rlc} + 1 \text{ and } \rho'_{i'} \leq t_{rlc} + 1, \text{ with } i \in \{b, c, d\} \text{ and } i' \in \{b', c', d', e'\}. \tag{7.17}$$

Now the pairwise Euclidean distances between all vertices between the subsets are computed and stored in the distance matrix $D'^{|\boldsymbol{V}_t| \times |\boldsymbol{V}'_t|}$; the distance is normalised (by $s_\rho \in \mathbb{R}^+$) in order to be on a similar scale as the angle.

$$D_{(i,i')} = \sqrt{(\theta_i - \theta'_{i'})^2 + s_\rho^{-2}(\rho_i - \rho'_{i'})^2} \tag{7.18}$$

The next step is to establish correspondence between the two vertex subsets $\boldsymbol{V}_t$ and $\boldsymbol{V}'_t$. It is stored in the initially empty matrix $F^{|\boldsymbol{V}_t| \times |\boldsymbol{V}'_t|}$. This is performed by a greedy approach. The global minimum of $\boldsymbol{D}'$, say $D'_{(i,i')}$ is found and correspondence between the vertices v_i and $v'_{i'}$ is assumed. The matrix entry $F_{(i,i')}$ is set to $D'_{(i,i')}$ and $D'_{(i,i')}$ afterwards to infinity, respectively. Subsequently both vertices are removed from their individual subsets. This iterative step is repeated until one subset is empty.

A distance score d is computed by the summation of all non-zero entries in $\boldsymbol{F}$.

$$d' = \sum_{i=1}^{|\boldsymbol{V}_t|} \sum_{i'=1}^{|\boldsymbol{V}_t|} F_{(i,i')} \tag{7.19}$$

Finally, similar local edge connectivity is rewarded by counting the number of edges, which occur in both neighbourhoods from the centre vertex to another vertex close by (and hence are included in $\boldsymbol{V}_t$ or $\boldsymbol{V'}_t$). This is only performed for the vertices v_i and $v'_{i'}$, which correspond according to the greedy algorithm.

$$d^{(a,a')} = {}^{d'}\!/_{d''}, \text{ where } d'' = 1 + \sum \delta^{(a,a')}_{(i,i')}, \text{ with } \delta^{(a,a')}_{(i,i')} = c^a_i c'^{a'}_{i'}. \tag{7.20}$$

This step is performed pairwise between all vertices $a \in \boldsymbol{V}$ and $a' \in \boldsymbol{V}'$, resulting in the distance matrix $D^{(m \times m')}$, where $D_{(a,a')} = d^{(a,a')}$. These distance measures describe the similarity of the local structure within a ridge line count of t_{rlc} between the two vertices $a \in \boldsymbol{G}^*$ and $a' \in \boldsymbol{G}^{*\prime}$. The assumption is that the smaller the distance is, the more similar is the local structure and hence it is very likely that these two vertices actually correspond.

The global similarity between the two graphs (with respect to to their centre vertex) is measured as the number of vertices which are closer than a given threshold, say there

are p corresponding vertices. Then the global similarity is given by

$$d_{QS} = \frac{p}{\sqrt{|\boldsymbol{V}| \cdot |\boldsymbol{V}'|}}. \tag{7.21}$$

The computation of the global similarity is only performed for the most similar local alignments (the τ lowest values of D) due to computational burdens.

The graph matching process is applied to the top τ' results.

Graph matching

The graph matching remains mostly unchanged from BGM (cf. Section 7.1.1) but after registration (cf. Section 7.1.1) a different cost function is used due to the use of scale-invariant features. Afterwards we compute similarity scores and derive a matching score. This final score is entirely based on scale-invariant features. It is performed for the best τ' results returned.

Cost function The cost function for sBGM relies on vertex attributes and the RLC of the two vertices v_i $v_i \in \boldsymbol{G}_a^*$ and $v'_j \in \boldsymbol{G}_{a'}^{*\prime}$ to their respective centre vertex. This is computed on the two scale-independent graphs aligned on the centre nodes a and a', respectively. The sub-matrix $\boldsymbol{C}_1$ states the cost for substituting vertex v_i for v'_j as follows:

$$c_{(i,j)} = |\rho_i - \rho'_j| + \varpi_\theta \theta_{i,j'} + |t_i - t'_j|, \tag{7.22}$$

where $\theta_{i,j'}$ stands for the relative angle difference (in terms of θ) between v_i and v'_j. It is the acute angle (in *radians*, not in degrees) between the two vertices and it is scaled in order to match the RLCs magnitude. The minutiae type difference of the two nodes becomes 1 if the type recognised differs, otherwise it is zero.

This time, the cost for deletion and insertion do not take edge connectivity into account (cf. Section 7.1.1); the main diagonal of the sub-matrices $\boldsymbol{C}_2$ and $\boldsymbol{C}_3$) are simply set to c_{id} where all other values are set to ∞. All values of $\boldsymbol{C}_4$ remain zero. Then the deletion cost is given by

$$c_{(i,m'+i)} = c_{id}. \tag{7.23}$$

Similarly, the insertion cost is

$$c_{(m+j,j)} = c_{id}. \tag{7.24}$$

Now Munkres' optimisation algorithm is applied once again to the cost matrix $\boldsymbol{C}$, resulting in the binary cost matrix $\boldsymbol{C}'$. Again, we focus the first quadrant $\boldsymbol{C}'_1 = \boldsymbol{C}'_{(1:m,1:m')}$ as it indicates which vertices have been substituted. Hence one can count the non-zero entries to find the number of corresponding vertices. The result is applied to the *spatial* representations of $\boldsymbol{G}^*_a$ and $\boldsymbol{G}^{*\prime}_{a'}$ to find the MCS.

Maximum common subgraph The MCS $\text{mcs}(\boldsymbol{G}^*_a, \boldsymbol{G}^{*\prime}_{a'})$ is derived from the *spatial* graph representations $\boldsymbol{G}$ and $\boldsymbol{G}'$ of $\boldsymbol{G}^*_a$ and $\boldsymbol{G}^{*\prime}_{a'}$, respectively. The procedure is identical for BGM and sBGM (cf. Section 7.1.1). We assume that the number of vertices and edges in the MCS are $|\text{mcs}(\boldsymbol{G}^*_a, \boldsymbol{G}^{*\prime}_{a'})|$ and $|\text{mcs}(\boldsymbol{G}^*_a, \boldsymbol{G}^{*\prime}_{a'})|_e$, respectively.

Matching score The matching score between the two ballprints is given by the combination of their similarity in terms of vertices (d_v), edges (d_e) and Quick Score (d_{QS}).

$$d_v = \frac{|\text{mcs}(\boldsymbol{G}^*_a, \boldsymbol{G}^{*\prime}_{a'})|}{\sqrt{|\boldsymbol{V}| \cdot |\boldsymbol{V}'|}} \tag{7.25}$$

$$d_v = \frac{|\text{mcs}(\boldsymbol{G}^*_a, \boldsymbol{G}^{*\prime}_{a'})|_e}{\sqrt{|\boldsymbol{V}| \cdot |\boldsymbol{V}'|}} \tag{7.26}$$

As d_v and d_e have been computed for the best τ' alignments, we choose the scores to be the final result for which the sum of d_v and d_e is maximum: d'_v and d'_e with the corresponding d'_{QS}. Then the final matching score s is given by the score fusion of all similarity measures using a weighted mean:

$$s = \gamma d_v + \iota d_e + (1 - \gamma - \iota) d_{QS}. \tag{7.27}$$

7.1.3 Inter-ridge spacing normalisation

Another approach is to use a physical feature of the biometric to normalise the feature to a given value and therefore to scaling the image obtained. One needs to be able to capture this feature reliably and accurately. Hence all Level 2 or Level 3 features (cf. Sections 2.2 and 2.2 in Chapter 2) are out of the question to avoid any major quality and

reliability issues. This leaves us with Level 1 measurements (see Section 2.2 in Chapter 2). Singular points such as cores and deltas are not feasible candidates because it is uncertain if those points are included in a certain capture and some prints do not contain them. The sensor area, especially when using a standard single fingerprint scanner, is too small to capture all of the ball area of older children. Hence only the general properties of the friction ridge skin such as its flow and inter-ridge spacing are left. We choose the inter-ridge spacing because it is a global and robust measure that mirrors physical the growth in terms of length/height as shown in Section 6.4.3 in the previous chapter.

An additional advantage is that since we are using adult fingerprint technology, the IRS needs to be similar to that of an adult fingerprint and so the ballprint capture needs to be scaled. We suggest *Radially Limited Averaged Power Spectrum* (RLAPS) which we have introduced in Section 4.3 in Chapter 4. It has been developed for adult fingermarks but it can be applied to infant ballprints (or any other friction ridge skin-based biometric) as well.

The difference from the previous approach in Chapter 6 is that there we estimated the IRS for a set of prints and used the median of the IRS to increase robustness and generality.

Here we replace global scaling factors by the IRS measured for an individual ballprint image $\boldsymbol{I}$. This gives us the scaling factor $k_{RLAPS}^{(\boldsymbol{I})}$ which is used to scale the entire image isotropically:

$$k_{RLAPS}^{(\boldsymbol{I})} = \frac{irs_{VF}}{\text{irs}(\boldsymbol{I})}. \tag{7.28}$$

There are two potential problems. Firstly, the measurement of the IRS may introduce inaccuracies which subsequently make the matching even more challenging. Secondly, it normalises a potentially valuable feature as it removes all individuality in inter-ridge spacing. These individual differences were conserved when we used a more general scaling factor. We refer to the scaling of an image with respect to to the individually measured IRS as *scaling via RLAPS*. We remind the readers that the IRS does not allow us to estimate an infant's age due to the overlap between the different age groups (Figure 6.4 in Chapter 6).

To compensate for measurement inaccuracies, we extend RLAPS-scaling by using bins and hence perform an indirect plausibility check. We call this RLAPS*. The different bins have been selected to span the area between the intersections of the PDF curves (see Figure 6.4 in Chapter 6) for the IRS with respect to age estimated. The values

falling into a certain bin are assigned the mean inter-ridge spacing of the dominant age for this bin. The bin borders and means are derived from our data; the actual values are stated in Equation 7.29 in Section 7.2.

7.1.4 Unknown-age problem break-down

An infant without any background information or ID presents a particular problem. The vaccination scenario and the trade-off between maximising the vaccination coverage while not being wasteful was mentioned at the beginning of the chapter. One can argue that the *worst case scenario* is that not only the query print but all prints stored in the database are of unknown age. This implies that all comparisons have to deal with this issue regardless of if it is for an identification or verification process. This case is very unlikely.

A more *realistic scenario* is that the capture date is stored alongside its corresponding print in the database. So the age of all database prints is known and only the query print's age is unknown. Firstly, one can normalise the print in accordance with a physical feature (e.g. the inter-ridge spacing via RLAPS or RLAPS*) or estimate the age from a physical feature such as the child's height or its inter-ridge spacing and adjust the print via this estimated age. Now the adjusted query print is run against the database and the top n prints (n-rank) are returned. Because the age difference between those n prints and the query print is known, an age can be inferred for the query capture and it can be adjusted accordingly. Then the comparisons for all n database prints and the adjusted query (based on the age difference for each particular comparison) is performed again.

Basically, the first adjustment of the query image is a quick way to reduce the computational burdens. This step could be skipped of course and an individual adjustment via the inferred age for all prints in the database can be performed as well, but at significant computational expense. The cost might be reduced if there are predefined capture ages with little variation, such as the recommended ages for specific vaccination visits, because it eliminates the need to scale the query print for each comparison. Nevertheless, regardless of what modus operandi might be the most suitable for the application, we can simplify the unknown age problem to a series of "regular" known age identification or verification tasks.

7.2 Experiments

All experiments are repeats of the verification experiment with the complete test set as explained and performed in Section 6.5 in Chapter 6. This time we restrict the fingerprint extractor and matcher to Neurotechnology Verifinger because it has outperformed its competitors. We also test Biometric Graph Matching and scale-invariant BGM.

In Section 7.3.1, we evaluate BGM and sBGM in the context of known age and compare the results achieved to commercial fingerprint technology, Neurotechnology Verifinger. We use exactly the same image scaling factors and techniques as before. Both approaches, BGM and sBGM, use novel ideas and ballprint graph representations, where sBGM relies entirely on a novel, scale-invariant (or according to our definition: growth-independent) set of features.

For this purpose we selected the following parameters experimentally for BGM: The Quick Score is computed for the $\tau = 980$ most similar edge pairs, where we set $t = 5$ (the threshold used to establish correspondence between edges) The cost function uses $c_{id} = 3$ and $\chi = 0.8$. The final score fusion is calculated with $\gamma = 0.58$.

We have experimentally obtained the following parameters for the use of sBGM. The variable nearest neighbours-based alignment (Section 7.1.2) uses $t_{rlc} = 5$ as neighbourhood radius to begin with, computes the global Quick Score for the $\tau = 100$ best alignments. The graph matching procedure is then performed for the $\tau' = 3$ best alignments according to the global QS. The local QS uses $s_\rho = {}^{10}/_{pi}$ to scale the angles. The cost function of the graph matching procedure scales the angle by $\varpi_\theta = 10\sqrt{2}$ and uses $c_{id} = 1.5$ as cost for deletion and substitution. The final score fusion is calculated with $\gamma = 0.6175$ and $\iota = 0.3325$. These values are not highly optimised due to the computational burden of sBGM (cf. Section 7.3.1).

In Section 7.3.2, we assume that the capture age of all prints (query and database) is unknown. Therefore we normalise each ballprint capture based on its inter-ridge spacing individually (RLAPS), and individually but within some constraints (RLAPS*). These constraints are given by the following equation:

$$k^{(\boldsymbol{I})}_{RLAPS*} = \begin{cases} 5.3 & k^{(\boldsymbol{I})}_{RLAPS} > 4.46, \\ 3.8 & 3.8 < k^{(\boldsymbol{I})}_{RLAPS} \leq 4.46, \\ 3.5 & k^{(\boldsymbol{I})}_{RLAPS} \leq 3.8. \end{cases} \tag{7.29}$$

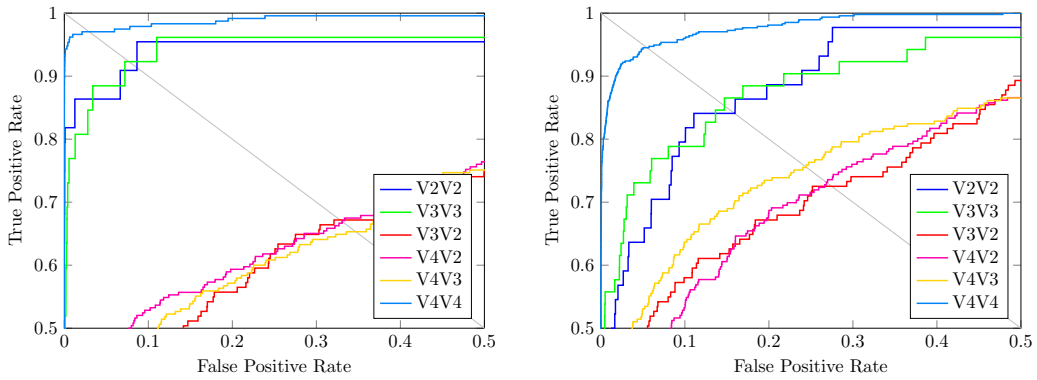
In Section 7.3.3, we evaluate the performance for the scenario if a query ballprint's capture age is unknown and has to be normalised via RLAPS or RLAPS* but the capture age for each print in the entire database is known.

7.3 Results

The results of our verification experiments are organised in three sections, (i) the performance achieved by BGM and sBGM with known age (Section 7.3.1), (ii) the results obtained when the capture age of all prints (query and database reference) is unknown (Section 7.3.2) and (iii) when only the query prints capture age remains unknown but the capture age for all prints stored in the database is available (Section 7.3.3).

7.3.1 BGM and sBGM

The Biometric Graph Matching achieves competitive results for all intra-visit comparisons when compared with disclosed fingerprint SDKs (cf. Table 6.9 in Chapter 6). This does not apply to the inter-visit comparisons, the performance drastically decreases and becomes similar to the ones Neurotechnology Verifinger obtains without any adjustment. The main difference is that BGM operates on the scaled images (V2: 250 ppi, V3: 275 ppi, V4: 375 ppi). The scaling process and the scaling factors have been discussed in Section 6.4.3 in Chapter 6. The corresponding ROCs are shown in Figure 7.7 and the respective EERs are stated in Table 7.1.



(a) BGM (b) sBGM

Figure 7.7: The top left corner of the ROCs for (a) BGM and (b) sBGM after tuning towards ballprints using external image scaling (V2: 250 ppi, V3: 275 ppi, V4: 375 ppi, cf. Section 6.4.3 in Chapter 6).

The use of scale-invariant features in the form of sBGM comes at some expense. The intra-visit EERs increase roughly by a factor of two when compared to BGM but the inter-visit results improve noticeably.

	Equal Error Rate					
Method	V2, V2	V3, V3	V4, V4	V3, V2	V4, V2	V4, V3
BGM	8.65%	7.69%	2.95%	32.82%	33.08%	34.69%
sBGM	15.91%	14.70%	5.49%	27.48%	26.85%	24.43%
VF*	4.55%	3.85%	0.00%	24.66%	49.98%	40.82%
VF	0.00%	0.00%	0.00%	2.63%	3.27%	2.24%

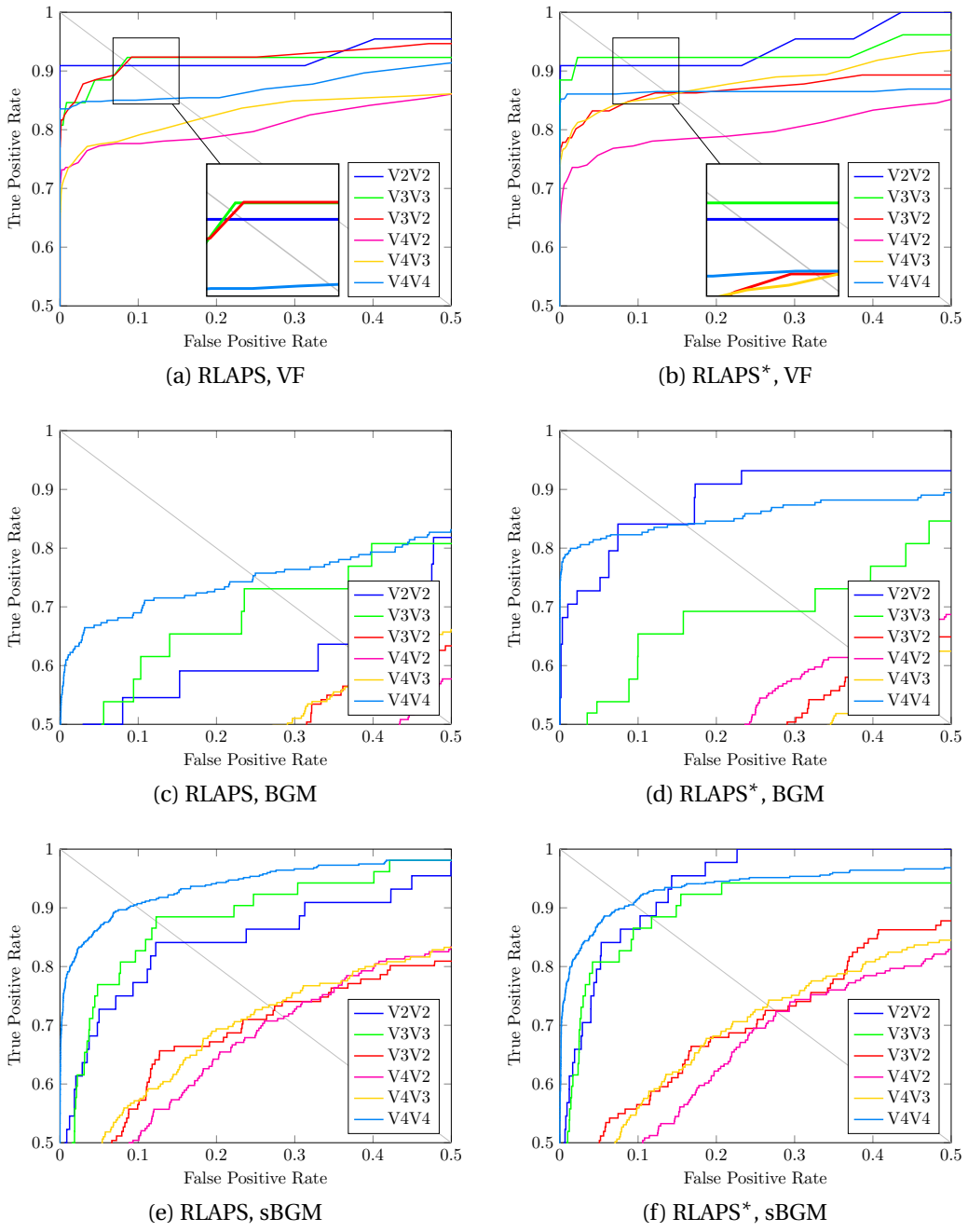
Table 7.1: EER scores achieved by the novel approaches BGM and sBGM for all possible ballprint age gap comparisons. The results for Neurotechnology Verifinger with scaling (VF) and without (VF*) are repeated from Tables 6.9 and 6.8 in Chapter 6 to allow for an easier comparison.

The use of graph matching comes with great computational burden. A single comparison between two biometric graphs takes on average 0.3844 s for BGM, 2.4372 s for sBGM and 0.0126 s for VF. These timings have been measured on an Intel i7-4700MQ @2.4 GHz with 32 GB of RAM; all data required had been stored in the RAM. These timings measure solely the comparison but not the template extraction process. BGM and sBGM are implemented in Matlab and VF is implemented in C++.

7.3.2 All prints normalised with RLAPS or RLAPS*

This is the verification scenario where all capture ages are unknown and all prints are normalised by their inter-ridge spacing to match the one of an adult fingerprint (RLAPS). The ROCs in Figure 7.8 show that the performance of VF and BGM suffers drastically, where sBGM is hardly affected at all.

If the normalisation process is performed within some constraints to minimise potential distortions introduced by the measurement of the inter-ridge spacing (RLAPS*), mostly the EERs corresponding to intra-visit comparisons benefit. This can be seen, when one compares the results obtained by RLAPS and RLAPS* in the respective Tables 7.2 and 7.3.



(a) RLAPS, VF

(b) RLAPS*, VF

(c) RLAPS, BGM

(d) RLAPS*, BGM

(e) RLAPS, sBGM

(f) RLAPS*, sBGM

Figure 7.8: The top left corner of the ROCs for VF (a, b), BGM (c, d) and sBGM (e, f) after RLAPS (a, c, e) and RLAPS* (b, d, f), respectively.

Again, the scale-invariant BGM is hardly affected by the change in the way scaling is done. It shows consistent performance across all different scaling modes.

Method	Equal Error Rate					
	V2, V2	V3, V3	V4, V4	V3, V2	V4, V2	V4, V3
BGM	36.36%	26.92%	24.68%	40.63%	45.94%	40.82%
sBGM	15.91%	22.29%	9.49%	27.43%	28.61%	26.72%
VF	9.09%	7.69%	13.50%	13.74%	21.14%	14.09%

Table 7.2: EER scores achieved by VF, BGM and sBGM for all possible ballprint age gap comparisons. We assume that all capture ages are unknown and normalise the IRS using RLAPS.

Method	Equal Error Rate					
	V2, V2	V3, V3	V4, V4	V3, V2	V4, V2	V4, V3
BGM	15.91%	30.77%	16.03%	39.71%	37.49%	43.17%
sBGM	11.36%	11.72%	9.44%	27.48%	27.64%	26.53%
VF	9.09%	8.28%	14.66%	8.52%	21.01%	18.09%

Table 7.3: EER scores achieved by VF, BGM and sBGM for all possible ballprint age gap comparisons. We assume that all capture ages are unknown and normalise the IRS using RLAPS*.

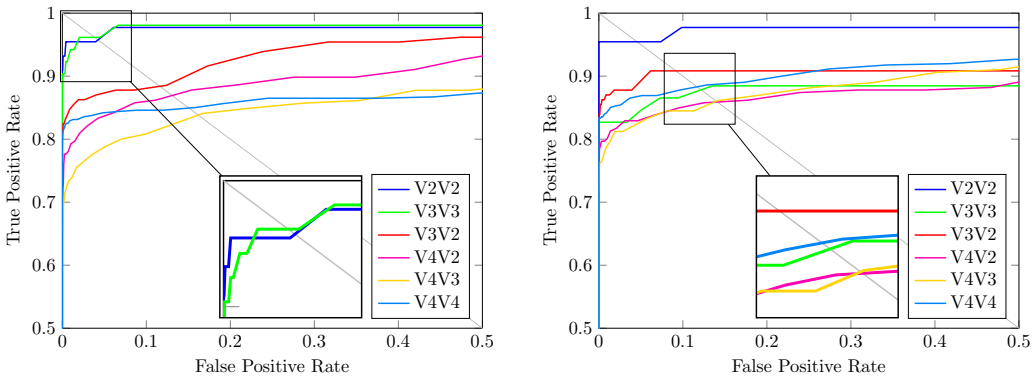
7.3.3 Query print normalised with RLAPS or RLAPS*

Our last verification experiment is restricted to VF because there is no added benefit to include BGM and sBGM due to their respective weak or stable performance. The scenario assumes that only the query print's capture age is unknown but the capture age for all prints stored in the database is available. The query is normalised either by RLAPS or RLAPS*. The ROCs obtained are illustrated in Figure 7.9.

The Equal Error Rates are presented in Table 7.4.

	Equal Error Rate					
	V2, V2	V3, V3	V4, V4	V3, V2	V4, V2	V4, V3
VF, RLAPS	4.25%	3.85%	15.06%	11.61%	13.07%	16.18%
VF, RLAPS*	4.55%	12.17%	11.69%	9.16%	14.10%	13,99%

Table 7.4: EER scores achieved by VF for all possible ballprint age gap comparisons. We assume that the capture ages for all ballprints stored in the database are known and only the query print's capture age remains unknown. We normalise its IRS using respectively RLAPS and RLAPS*.



(a) VF, RLAPS (b) VF, RLAPS*

Figure 7.9: The top left corner of the ROCs achieved by VF for all possible ballprint age gap comparisons. We assume that the capture ages for all ballprints stored in the database are known and only the query print's capture age remains unknown. We normalise its IRS using respectively (a) RLAPS and (b) RLAPS*.

7.4 Discussion

We briefly discuss the experiments in the order they have been performed: (i) BGM and sBGM with k_{irs}, (ii) the normalisation of all prints with RLAPS or RLAPS*, and (iii) individual normalisation of the query print only.

The EERs achieved by BGM and sBGM when the ballprint images are scaled by k_{irs} are comparable to the results attained by SourceAFIS. BGM's weak inter-visit EERs are likely due to its alignment process and the relatively small overlap between the ballprints from different visits (cf. Section 6.4.2 in Chapter 6). It uses edge pairs to achieve rough alignment. Now if there are hardly any edges in the overlapping area it becomes increasingly likely that the algorithm aligns the two ballprint graphs suboptimally. This again results in low genuine verification scores. This explanation is supported by the fact that sBGM performs better in the inter-visit comparisons because it does not necessarily require the edge information that BGM does.

When all prints are normalised with RLAPS or RLAPS*, the performance of VF and BGM suffers noticeably, where sBGM is hardly affected at all.

The main question in this context is why does Neurotechnology Verifinger's performance decrease so much? One could argue that scaling on an individual basis should benefit the result and lead to improved EERs. We can think of two reasons. The first potential reason is due to the measured IRS and its variance. It may be caused by either shortcomings in the measurement algorithm, or to small overlap between the

ballprints so that one image contains for example, singular points, where the other does not. This would lead to the measurement of different IRSs. The second reason we can think of is that Neurotechnology Verifinger uses the inter-ridge spacing and its deviation from the fingerprint median as a feature and relies heavily on it to distinguish between prints. Of course, a combination of both reasons is possible as well. Nevertheless, we are unable to confirm or exclude any potential causes because VF is a commercial fingerprint SDK and hence fully undisclosed (a black box). So our reasoning remains speculative.

When only the query print is normalised, the results for VF are mixed. The use of RLAPS* groups most of the ROCs more closely together (and achieves overall slightly better results) than RLAPS but at the expense of the intra-visit comparison for the third visit. This behaviour is most likely due to the large overlap of the IRS for V3 with the other two visits and its rather small bin (cf. Section 7.1.3, see also Figure 6.4 in Chapter 6).

At first glance, more surprising is that the EERs do not always show a clear improvement or any improvement at all over the scenario where all prints are normalised. This might be due to the fact that the scaling processes via the infant's age or the ballprints inter-ridge spacing may lead to very different results in terms of scale. This may be the case for an older child possessing ballprints with rather small IRS when compared to similarly aged infants. The underlying cause could be malnutrition and thus stunted growth. That does not apply to the Happy Feet database. Here the reason is more likely the individual variance during the estimation of the IRS combined with the necessity to adjust the ballprints for Neurotechnology Verifinger which reacts strongly to even slight scaling adjustments (cf. Section 4.6.2 in Chapter 4).

7.5 Challenges, limitations and recommendations

We have presented novel approaches and scenarios to handle infant growth without having to model the growth explicitly in order to compensate for its effects. We have introduced Biometric Graph Matching for ballprints and scale-invariant BGM, a variant that relies solely on scale-independent features. We also have shown the effects of normalising ballprints individually with respect to their measured inter-ridge spacing.

Both graph matching algorithms show reasonable performance but exhibit different advantages. For intra-visit comparisons BGM demonstrates competitive performance in line with adult fingerprint SDKs but struggles with inter-visit comparisons (similar to the non-commercial SDKs as well). This is most likely due to the small overlap of

the ballprints from different visits.

sBGM performed very consistently across all different scaling modes regardless of whether it was a global factor per visit or individual scaling. It was hardly affected by it. This robustness is one of its strengths. Additionally there is no need to derive a scaling factor, it can handle the changes in scale applied to the ballprint due to the infant growth process effortlessly. However, its weakness is it is not competitive, its Equal Error Rates are not as good as those of fingerprint SDKs. It demonstrates the potential of scale-independent features in that they can be used successfully to sidestep growth.

One has to keep in mind that both graph matching procedures rely on graphs which have been extracted using Neurotechnology Verifinger. Hence the input images had to be adjusted (via k_{irs} per visit, RLAPS or RLAPS*) in order to enable VF to extract a ballprint's binarised image and its features. This means that whenever its graph extraction fails, the graph matching cannot work reliably. This issue could be resolved by developing a native extractor for ballprints which can be adjusted to either the infant's age or its inter-ridge spacing.

A major challenge we have faced is the ballprint overlap across different visits. The infant's growth is mirrored by its ballprint so that at some point it outgrows the sensor area of the small capture device. Therefore the person responsible for capturing the data has to take great care to capture the same area of print across different visits, especially when one considers the infant's uncooperative nature. A proper protocol and a fitting to align the capture device reliably to the same part of the ball could potentially overcome this challenge and ensure greater areal consistency. So would the use of a larger sensor.

8 Conclusion and future research

John F. Kennedy recognised that "*Children are the world's most valuable resource and its best hope for the future*" [68]. An infant biometric would allow authorities and public health agencies to link an infant's personal information and vaccination history to a physical feature instead of a piece of paper that can be easily lost, damaged or stolen. It is an intriguing idea that faces many hurdles. These include the tininess and fragility of an infant's features, their uncooperative nature and their rapid growth. The latter is a particular challenge since the basic principles of a biometric trait are permanence and stability over time, individuality, robustness and ease to be captured. Finding a physical trait that is feasible is often referred to as the "*infant biometric problem*".

In this thesis we researched various aspects of the "ballprint" as an infant biometric, which refers to the friction ridge skin area of the pad under the big toe. We demonstrated that this area of skin exhibits similar properties to fingerprint which has been used by law enforcement agencies for over a century.

We followed 54 infants from birth, capturing ballprints within 3 days of birth, at two months old, at 6 months and at 2 years. We found that the skin of a newborn's foot dries and cracks so the ridge lines are often not visible to the naked eye and an adult fingerprint scanner cannot capture them.

We showed that the ballprint grows isotropically during infancy and that it can be well approximated by a linear function of the infant's age. Additionally, mature fingerprint technology that has been originally developed for adult fingerprints can match ballprints if they are adjusted by a physical feature (the inter-ridge spacing) to be the same size as adult fingerprints. We demonstrated that when the growth is compensated for by isotropic scaling, impressive verification scores are achievable even for captures

taken 22 months apart. The scores improve even further if low-quality prints are rejected.

We ranked infant ballprints by quality using a classification algorithm, trained on expert decisions, for whether adult latent fingerprints had Evidential Value. When the worst third were removed the Equal Error Rates dropped from 1 – 2% to 0%.

We propose that the ballprint is a solution to the infant biometric problem and it can serve to improve vaccination uptake and delivery, and to stem the worldwide traffic in children. This is the main contribution of the thesis.

We now discuss the individual results and contributions of each chapter and show that all *research questions* (RQs) have been answered, before we offer future research directions.

8.1 Individual contributions

Chapters 1, 2 and 3 set the background of this thesis. They give some insight into the infant biometric problem, introduce two novel databases (fingermark, infant ballprint) and describe important concepts related to biometrics.

Chapter 4 introduces a novel algorithm EVA to estimate if a certain fingermark may be of evidential value in an ongoing forensic investigation. It uses image quality features of the SDKs NFIQ2 and VF, and their respective fusion and is trained on a ground truth supplied by forensic police examiners. In order to be able to use these SDKs, the digital captures of the fingermarks need to be suited to the SDKs' resolution requirements. For this purpose we developed an algorithm RLAPS to estimate a mark's capture resolution based on the inter-ridge spacing. All results are evaluated for different capture devices and types of distortions such as pressure variations (too little or too much) or partial marks. Additionally, we have shown that our algorithm and the police examiners agree on most marks and that most of the marks which can be compared successfully against a reference database have been classified by either the experts or by EVA as being of Evidential Value. This answers RQ 1.

Chapter 5 presents an extensive literature review on newborn and infant biometrics and models for how growth affects a biometric trait. We investigate the infant footprint crease pattern as to whether it is a suitable newborn biometric. We were able to confirm that drying and cracking of the skin occurs and that this (mostly) obscures the ridge pattern. We were however unable to confirm that there is a time gap shortly after birth which allows for reasonable skin conditions to capture good-quality prints. The

crease pattern in infants changes constantly and we were unable to achieve reasonable verification scores in an automatic experiment. A manual experiment proceeded to be more effective but the examiners did not necessarily limit themselves to the crease pattern in their comparisons. The literature review answers RQ 2a for fingerprint, and our crease pattern investigation satisfies RQ 2b.

Chapter 6 investigates the infant ballprint as a physical trait changing as a baby grows. It points out the similarity to fingerprint but also its larger physical features. We demonstrate that the growth can be approximated by isotropic scaling (RQ 2a) and highlight that the inter-ridge spacing (employing our algorithm from Chapter 4) and the infant's age have a linear relationship as a first approximation. We next show that standard fingerprint technology can be used if the prints are adjusted to match the physical features of adult fingerprints (RQ 2c)). The already impressive Equal Error Rates (2 – 3% for an age gap of up to 22 months) improve even further (to 0%) if one limits the specimens to high-quality prints. Their quality is inferred via our Evidential Value Algorithm from Chapter 4 which has been developed and trained for adult fingermarks. We find some evidence that ballprints can be used as a newborn biometric and succeed in newborn verification. Finally, we exhibit that the two length measures: the IRS measured on our ballprints and physical height data obtained externally, correlate with respect to the infant's age. This chapter answers RQs 2, 2a, 2b and 2c.

Chapter 7 describes the consequences if an infant's age is unknown and shows approaches and strategies to overcome this issue. This is done by normalisation of a physical feature (inter-ridge spacing) and, smart protocols that allow the age to be inferred or scale-invariant features to be extracted such as a minutia's type and orientation, its connectivity and the number of ridge lines between two minutiae. These features are built into Biometric Graph Matching resulting in a scale-invariant version of BGM. sBGM exhibits widely consistent performance across different scenarios (at great computational expense) but does not achieve the performance of commercial fingerprint products (RQs 2d).

8.2 Future research

In the future we would like to investigate the ballprint as an infant and newborn biometric on a much larger scale. For this purpose we would like to conduct a large sample study regarding the ballprint and its IRS distribution for infants with respect to age, gender, ethnicity, nutrition and other environmental influences.

We are also interested in the ridge line pattern of the foot and how it is affected when

one does not wear shoes but walks barefoot the entire day over a great period of time.

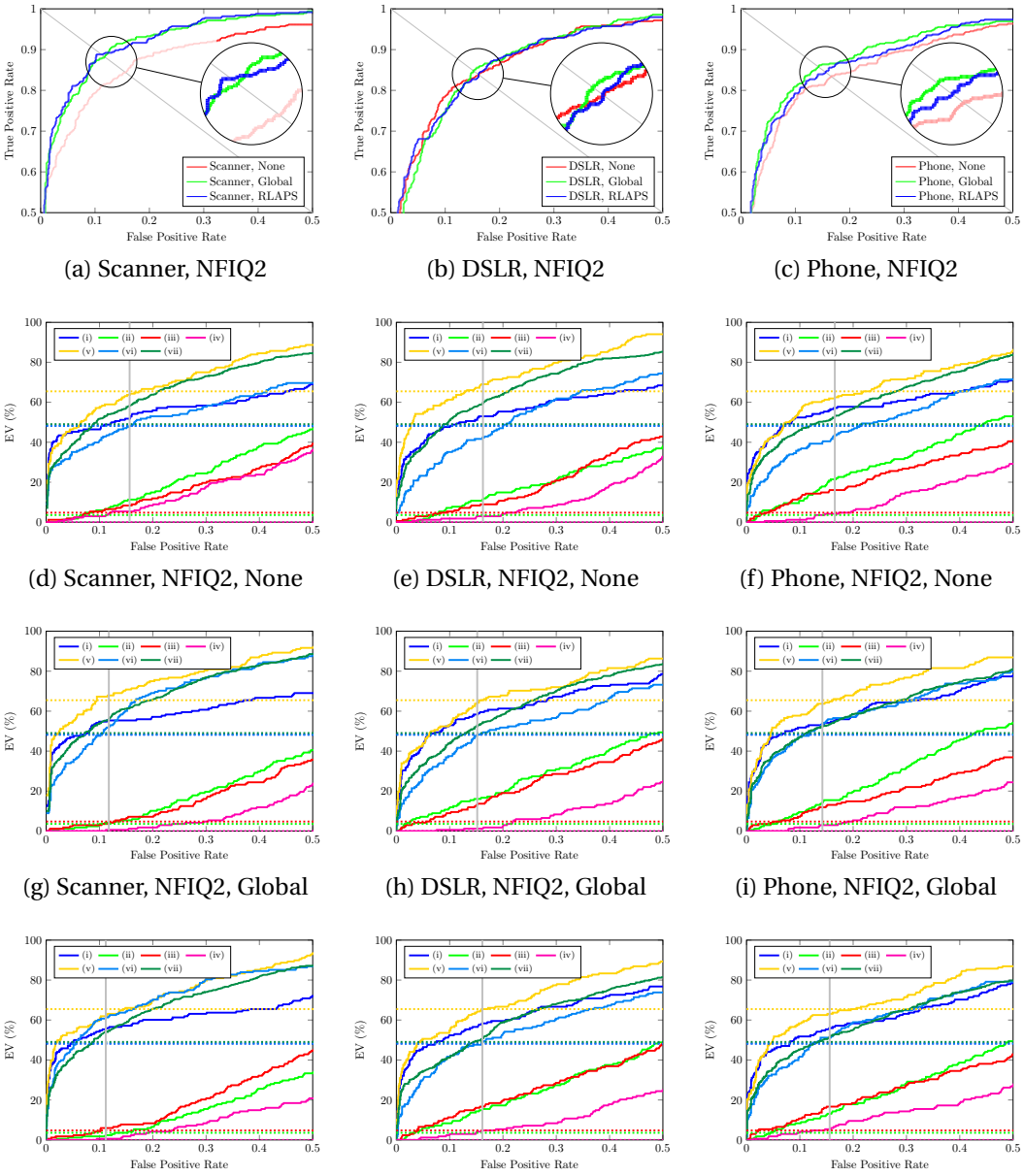
We have struggled with the low resolution of our fingerprint scanner. Hence we are curious if modern mobile phone cameras can be employed to achieve similar capture results at a much higher resolution. It may even be possible to improve the phone's usability by adding attachments which help to handle a rebellious newborn or infant.

Another valuable idea is to develop a completely disclosed SDK to extract the ballprint ridge pattern and so to exclude commercial products from the process. This might allow us to improve the reliability because it removes the immediate need to estimate base scaling factors or adjust ballprint captures to suit adult fingerprints.

Lastly, we suggest the improvement of graph matching. The general idea has proved to work but suffers from unreliable extraction. So a better capture device and a ballprint SDK would allow to optimise its parameters and to continue the idea of scale-invariant features. Scale-invariant features, which are directly based on the physical trait, paired with effective comparison techniques, remove the necessity to deal with growth at all.

The final goal is to combine all these ideas and approaches into one application to be used with smartphones. A health worker could then retrieve a child's identity or vaccination data without any knowledge of age or ID and thereby see if a vaccination is required or not. Law enforcement agencies could fight the trafficking of children more effectively because a child's identity can be established regardless of its age, origin or presence of ID.

A Adult fingermarks



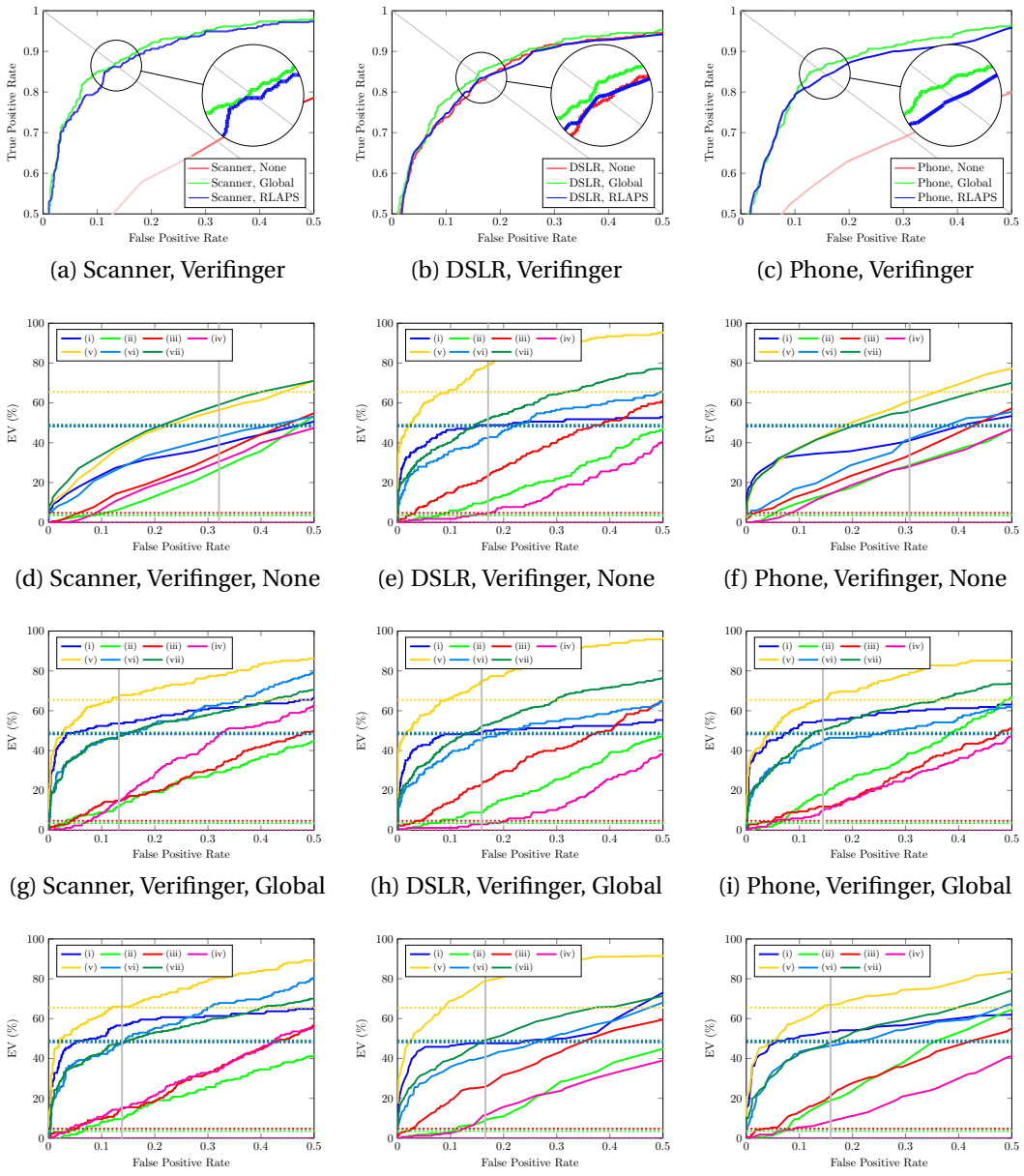
(a) Scanner, NFIQ2 (b) DSLR, NFIQ2 (c) Phone, NFIQ2

(d) Scanner, NFIQ2, None (e) DSLR, NFIQ2, None (f) Phone, NFIQ2, None

(g) Scanner, NFIQ2, Global (h) DSLR, NFIQ2, Global (i) Phone, NFIQ2, Global

(j) Scanner, NFIQ2, RLAPS (k) DSLR, NFIQ2, RLAPS (l) Phone, NFIQ2, RLAPS

Figure A.1: The first row (Figure A.1a to A.1c) shows the top left corner of the ROCs for all capture devices calculated on the NFIQ2 image quality feature set. The colour varies according to the fraction of $^{\text{EV}}\,\&\,^{\text{ccID}}/_{\text{ccID}}$ as the classification threshold moves along the ROC; the smaller the fraction, the lighter the colour. The other rows (Figure A.1d to A.1l) show the EV distribution according to the mark's distortion class which has been computed by EVA (solid) and the experts (dashed). The gray vertical line is the threshold corresponding to the EER when the operating point moves along the ROC.



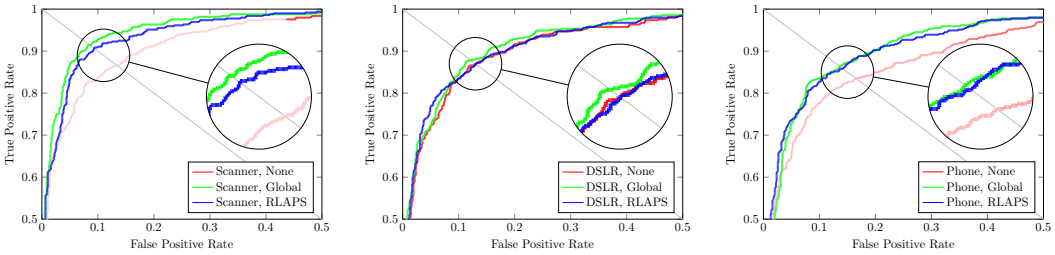
(a) Scanner, Verifinger (b) DSLR, Verifinger (c) Phone, Verifinger

(d) Scanner, Verifinger, None (e) DSLR, Verifinger, None (f) Phone, Verifinger, None

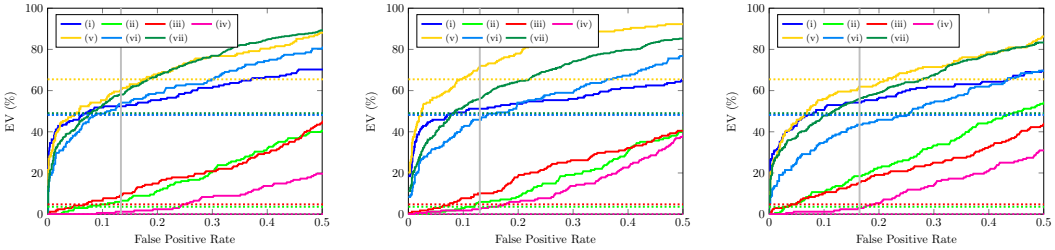
(g) Scanner, Verifinger, Global (h) DSLR, Verifinger, Global (i) Phone, Verifinger, Global

(j) Scanner, Verifinger, RLAPS (k) DSLR, Verifinger, RLAPS (l) Phone, Verifinger, RLAPS

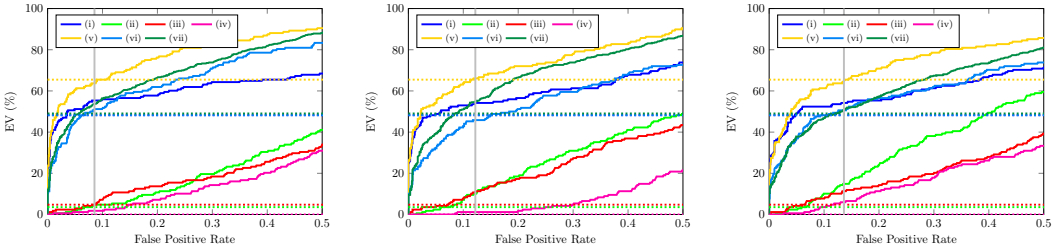
Figure A.2: The first row (Figure A.2a to A.2c) shows the top left corner of the ROCs for all capture devices calculated on the Verifinger image quality feature set. The colour varies according to the fraction of $^{\text{EV \& ccID}}/_{\text{ccID}}$ as the classification threshold moves along the ROC; the smaller the fraction, the lighter the colour. The other rows (Figure A.2d to A.2l) show the EV distribution according to the mark's distortion class which has been computed by EVA (solid) and the experts (dashed). The gray vertical line is the threshold corresponding to the EER when the operating point moves along the ROC.



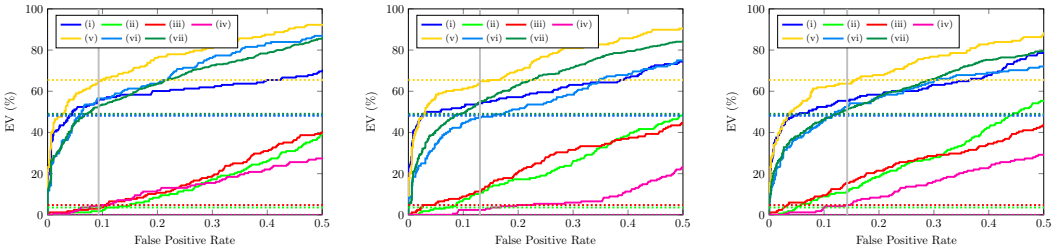
(a) Scanner, Fusion (b) DSLR, Fusion (c) Phone, Fusion



(d) Scanner, Fusion, None (e) DSLR, Fusion, None (f) Phone, Fusion, None

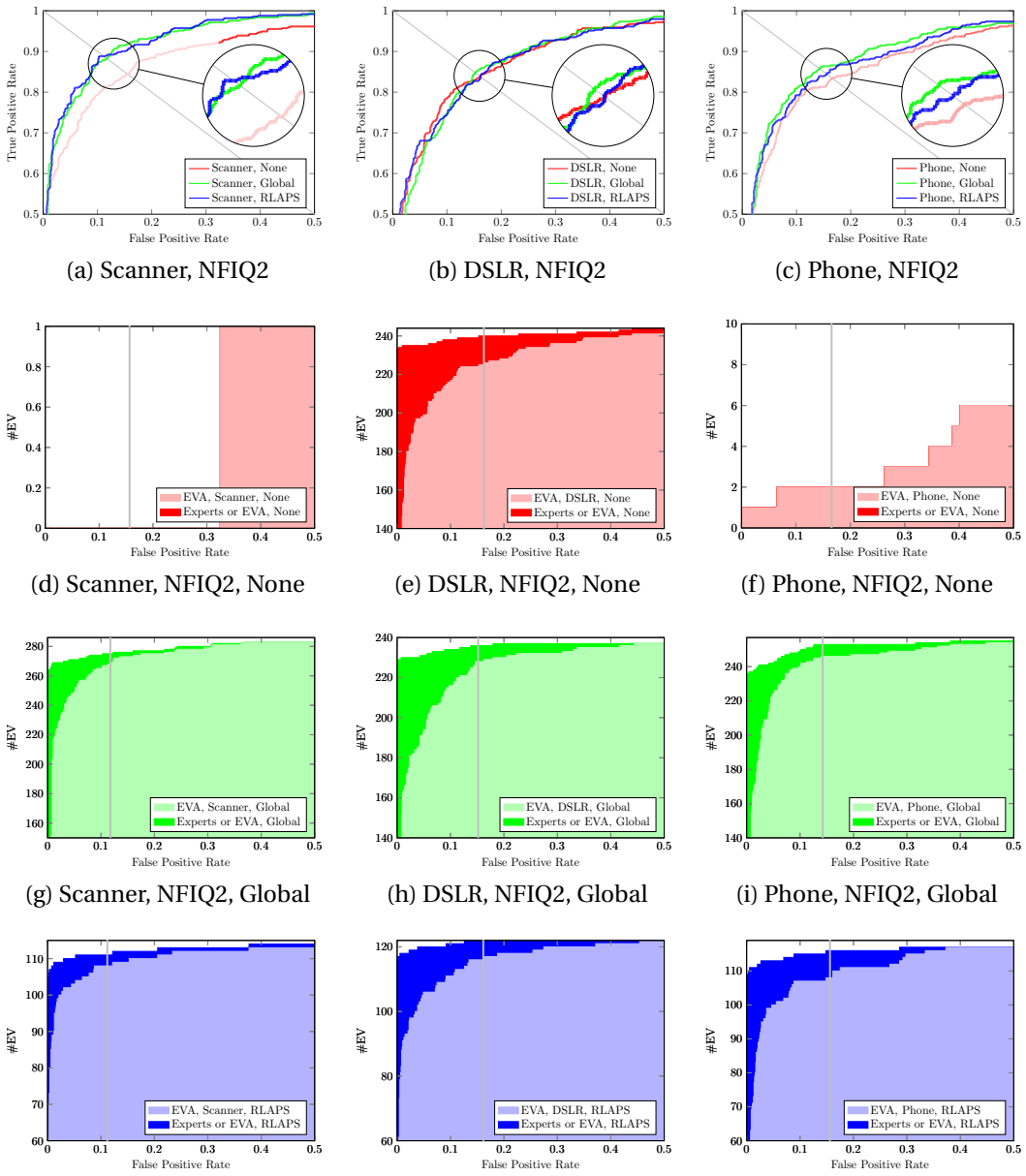


(g) Scanner, Fusion, Global (h) DSLR, Fusion, Global (i) Phone, Fusion, Global



(j) Scanner, Fusion, RLAPS (k) DSLR, Fusion, RLAPS (l) Phone, Fusion, RLAPS

Figure A.3: The first row (Figure A.3a to A.3c) shows the top left corner of the ROCs for all capture devices calculated on the Fusion image quality feature set. The colour varies according to the fraction of $^{\text{EV}\,\&\,\text{ccID}}/_{\text{ccID}}$ as the classification threshold moves along the ROC; the smaller the fraction, the lighter the colour. The other rows (Figure A.3d to A.3l) show the EV distribution according to the mark's distortion class which has been computed by EVA (solid) and the experts (dashed). The gray vertical line is the threshold corresponding to the EER when the operating point moves along the ROC.



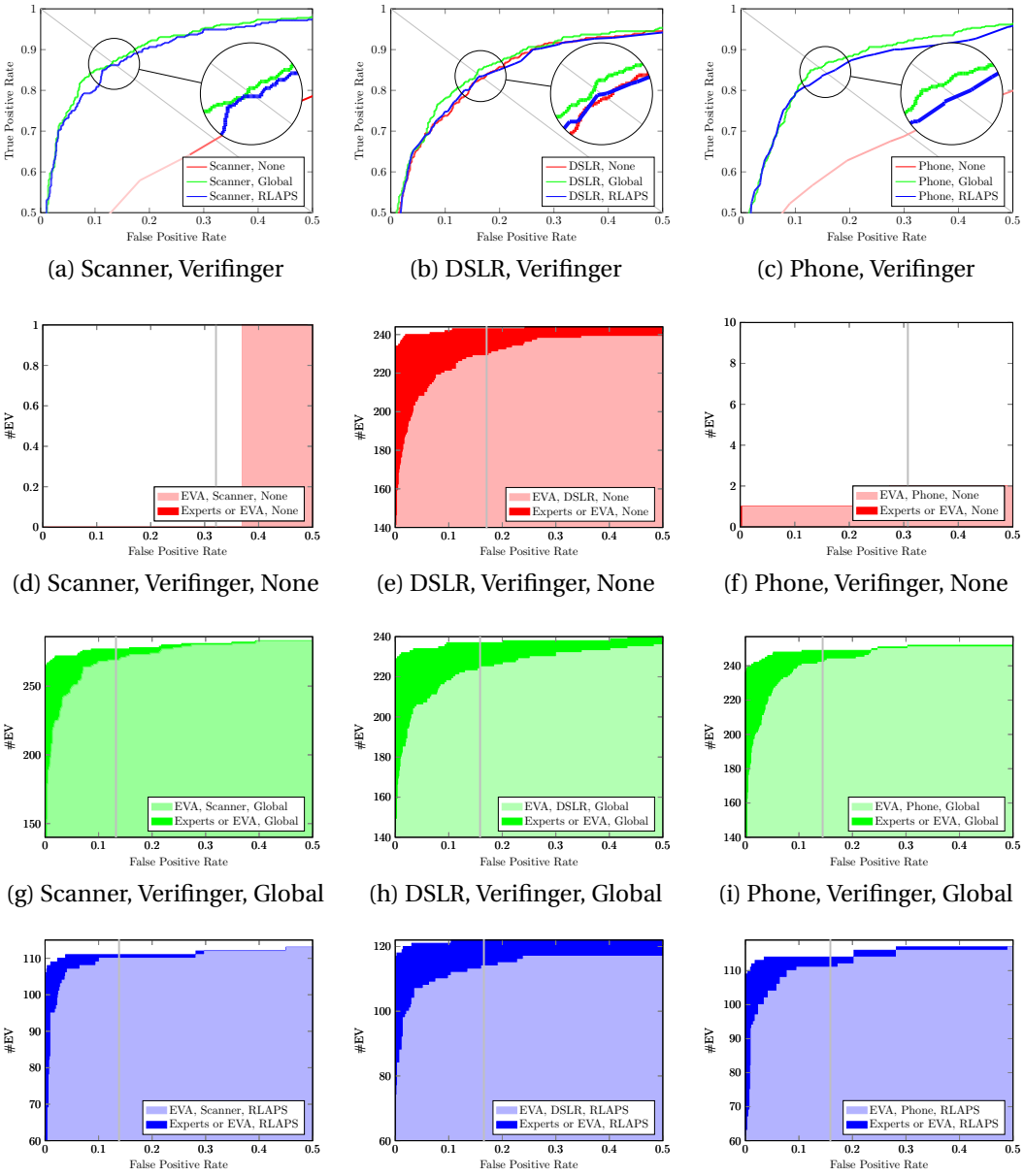
(a) Scanner, NFIQ2 (b) DSLR, NFIQ2 (c) Phone, NFIQ2

(d) Scanner, NFIQ2, None (e) DSLR, NFIQ2, None (f) Phone, NFIQ2, None

(g) Scanner, NFIQ2, Global (h) DSLR, NFIQ2, Global (i) Phone, NFIQ2, Global

(j) Scanner, NFIQ2, RLAPS (k) DSLR, NFIQ2, RLAPS (l) Phone, NFIQ2, RLAPS

Figure A.4: The first row (Figure A.4a to A.4c) shows the top left corner of the ROCs for all capture devices calculated on the NFIQ2 image quality feature set. The colour varies according to the fraction of $^{\text{EV \& ccID}}/_{\text{ccID}}$ as the classification threshold moves along the ROC; the smaller the fraction, the lighter the colour. The other rows (Figure A.4d to A.4l) show the number of ccIDs classified as EV by EVA (light colour) and the additional ones by the experts (dark colour). The gray line is the threshold corresponding to the EER when the operating point moves along the ROC.



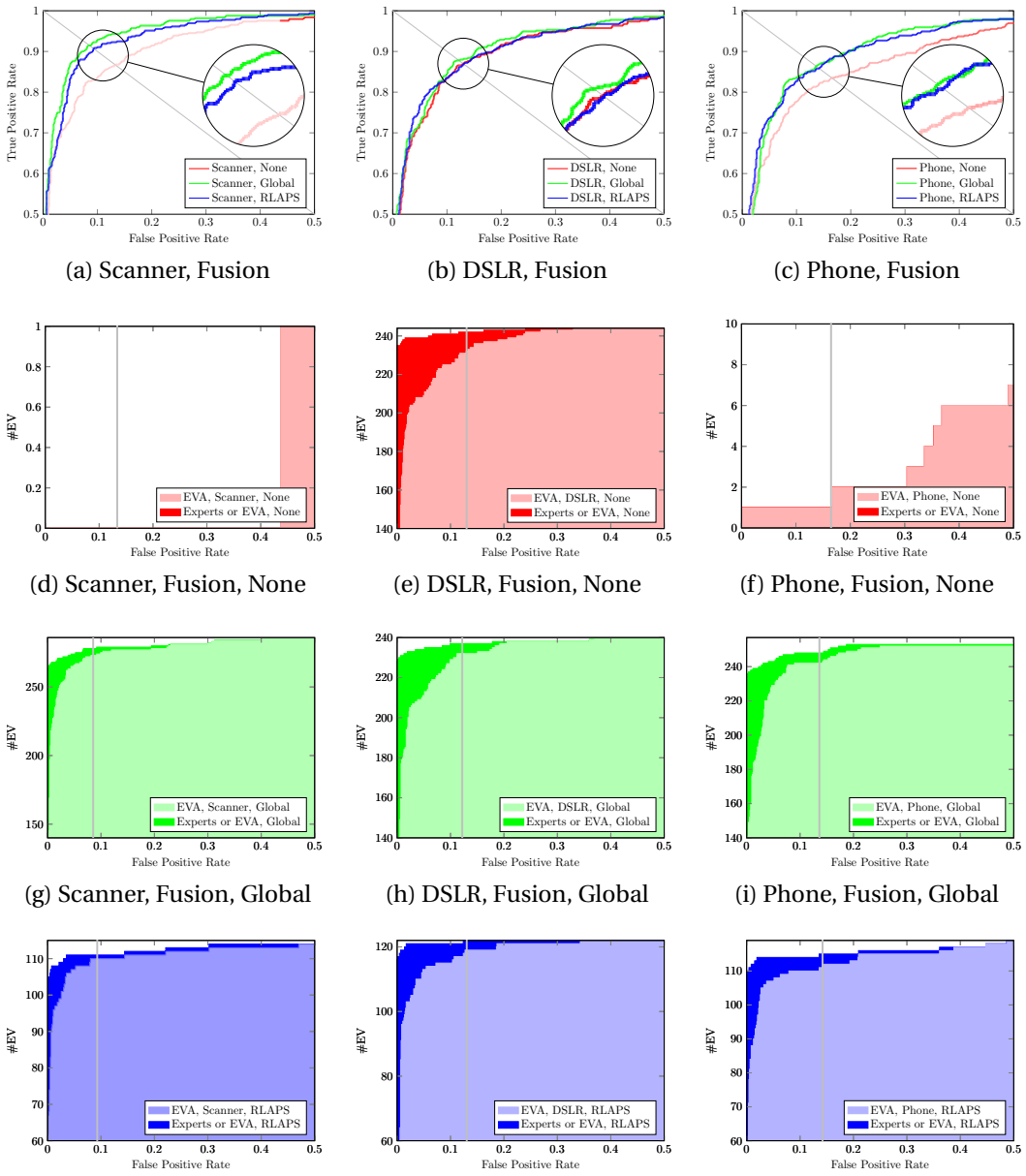
(a) Scanner, Verifinger (b) DSLR, Verifinger (c) Phone, Verifinger

(d) Scanner, Verifinger, None (e) DSLR, Verifinger, None (f) Phone, Verifinger, None

(g) Scanner, Verifinger, Global (h) DSLR, Verifinger, Global (i) Phone, Verifinger, Global

(j) Scanner, Verifinger, RLAPS (k) DSLR, Verifinger, RLAPS (l) Phone, Verifinger, RLAPS

Figure A.5: The first row (Figure A.5a to A.5c) shows the top left corner of the ROCs for all capture devices calculated on the Verifinger image quality feature set. The colour varies according to the fraction of $^{\text{EV \& ccID}}/_{\text{ccID}}$ as the classification threshold moves along the ROC; the smaller the fraction, the lighter the colour. The other rows (Figure A.5d to A.5l) show the number of ccIDs classified as EV by EVA (light colour) and the additional ones by the experts (dark colour). The gray line is the threshold corresponding to the EER when the operating point moves along the ROC.



(a) Scanner, Fusion (b) DSLR, Fusion (c) Phone, Fusion

(d) Scanner, Fusion, None (e) DSLR, Fusion, None (f) Phone, Fusion, None

(g) Scanner, Fusion, Global (h) DSLR, Fusion, Global (i) Phone, Fusion, Global

(j) Scanner, Fusion, RLAPS (k) DSLR, Fusion, RLAPS (l) Phone, Fusion, RLAPS

Figure A.6: The first row (Figure A.6a to A.6c) shows the top left corner of the ROCs for all capture devices calculated on the Fusion image quality feature set. The colour varies according to the fraction of $^{\text{EV \& ccID}}/_{\text{ccID}}$ as the classification threshold moves along the ROC; the smaller the fraction, the lighter the colour. The other rows (Figure A.6d to A.6l) show the number of ccIDs classified as EV by EVA (light colour) and the additional ones by the experts (dark colour). The gray line is the threshold corresponding to the EER when the operating point moves along the ROC.

Bibliography

[1] M. A. Acree, "Is there a gender difference in fingerprint ridge density?" *Forensic science international*, vol. 102, no. 1, pp. 35–44, 1999.

[2] A. Alexander, O. Forth, J. Nash, and N. Yager, "Zooplots for speaker recognition with tall and fat animals."

[3] F. Alonso-Fernandez, J. Fierrez, J. Ortega-Garcia, J. Gonzalez-Rodriguez, H. Fronthaler, K. Kollreider, and J. Bigun, "A comparative study of fingerprint image-quality estimation methods," *Information Forensics and Security, IEEE Transactions on*, vol. 2, no. 4, pp. 734–743, 2007.

[4] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *The American Statistician*, vol. 46, no. 3, pp. 175–185, 1992.

[5] A. Anthonioz and C. Champod, "Integration of pore features into the evaluation of fingerprint evidence," *Journal of forensic sciences*, vol. 59, no. 1, pp. 82–93, 2014.

[6] Apple. (2015, August) iPhone 6 - Touch ID. https://www.apple.com/au/iphone-6/touch-id/.

[7] A. Arakala, H. Hao, S. Davis, and K. Horadam, "The palm vein graph: Feature extraction and matching," in *ICISSP 2015*. SCITEPRESS, 2015, pp. 56–64.

[8] D. R. Ashbaugh, *Quantitative-qualitative friction ridge analysis: an introduction to basic and advanced ridgeology*. CRC press Boca Raton, 1999.

[9] W. Babler *et al.*, "Embryologic development of epidermal ridges and their configurations," *Dermatoglyphics: science in transition. New York: Wiley-Liss*, pp. 95–112, 1991.

[10] S. Balameenakshi and S. Sumathi, "Biometric recognition of newborns: Identification using footprints," in *Information Communication Technologies (ICT), 2013 IEEE Conference on*, Apr 2013, pp. 496–501.

[11] P. J. Besl and N. D. McKay, "A method for registration of 3-d shapes," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 14, no. 2, pp. 239–256, Feb 1992.

[12] S. Bharadwaj, H. S. Bhatt, R. Singh, M. Vatsa, and S. K. Singh, "Face recognition for newborns: A preliminary study," in *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, Sept 2010, pp. 1–6.

[13] J. Bigun, "Fingerprint features," in *Encyclopedia of Biometrics*, S. Z. Li and A. K. Jain, Eds. Springer US, 2014, pp. 1–13. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-27733-7_50-3

[14] Bill & Melinda Gates Foundation. (2011) Create low-cost cell phone-based solutions for improved uptake and coverage of childhood vaccinations. http://www.grandchallenges.org/MeasureHealthStatus/Topics/CellPhoneApps/Pages/Round7.aspx.

[15] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

[16] J. Blake, "Identification of the new-born by flexure creases," *Ident. News*, vol. 9, no. 9, pp. 3–5, 1959.

[17] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*. ACM, 1992, pp. 144–152.

[18] R. Cappelli, D. Maio, and D. Maltoni, "Sfinge: an approach to synthetic fingerprint generation," in *International Workshop on Biometric Technologies (BT2004)*, 2004, pp. 147–154.

[19] I. Castellanos, *Dermopapiloscopia clínica*. Imp. P. Fernández y Cia., 1953. [Online]. Available: https://books.google.com.au/books?id=SgrsoQEACAAJ

[20] Centers for Disease Control and Prevention. (2014) Growth charts. http://www.cdc.gov/growthcharts/.

[21] J. Chen, C. Zhang, and G. Rong, "Palmprint recognition using crease," in *Image Processing, 2001. Proceedings. 2001 International Conference on*, vol. 3. IEEE, 2001, pp. 234–237.

[22] Y. Chen and G. Medioni, "Object modelling by registration of multiple range images," *Image and vision computing*, vol. 10, no. 3, pp. 145–155, 1992.

[23] Y. Chen, S. C. Dass, and A. K. Jain, "Fingerprint quality indices for predicting authentication performance," in *Audio-and Video-Based Biometric Person Authentication*. Springer, 2005, pp. 160–170.

[24] S. A. Cole, "Forensics without uniqueness, conclusions without individualization: the new epistemology of forensic identification," *Law, Probability and Risk*, vol. 8, no. 3, pp. 233–255, 2009.

[25] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[26] H. Cummins and C. Midlo, *Finger prints, palms and soles: An introduction to dermatoglyphics*. Dover Publications New York, 1961, vol. 319.

[27] S. C. Dass, "Individuality of fingerprints: A review," in *Encyclopedia of Biometrics*, S. Z. Li and A. K. Jain, Eds. Springer US, 2014, pp. 1–19. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-27733-7_58-2

[28] R. de Paula Lemes, O. R. P. Bellon, L. Silva, and A. K. Jain, "Biometric recognition of newborns: Identification using palmprints," in *Biometrics (IJCB), 2011 International Joint Conference on*, Oct 2011, pp. 1–6.

[29] R. de Paula Lemes, M. Pamplona Segundo, O. Bellon, and L. Silva, "Dynamic pore filtering for keypoint detection applied to newborn authentication," in *Pattern Recognition (ICPR), 2014 22nd International Conference on*, Aug 2014, pp. 1698–1703.

[30] G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds, "Sheep, goats, lambs and wolves a statistical analysis of speaker performance in the nist 1998 speaker recognition evaluation," in *International Conference on Spoken Language Processing*, 1998.

[31] E. Dougherty and R. Lotufo, *Hands-on morphological image processing*. Society of Photo Optical, 2003, vol. 59.

[32] I. L. Dryden and K. V. Mardia, *Statistical Shape Analysis*, 1st ed. Wiley, Sept 1998.

[33] K.-B. Duan and S. Keerthi, "Which is the best multiclass svm method? an empirical study," in *Multiple Classifier Systems*, ser. Lecture Notes in Computer Science, N. Oza, R. Polikar, J. Kittler, and F. Roli, Eds. Springer

Berlin Heidelberg, 2005, vol. 3541, pp. 278–285. [Online]. Available: http://dx.doi.org/10.1007/11494683_28

[34] T. Dunstone and N. Yager, *Biometric system and data analysis: Design, evaluation, and data mining*. Springer Science & Business Media, 2008.

[35] M. Fairhurst, Ed., *Age Factors in Biometric Processing*, ser. Professional Applications of Computing. Institution of Engineering and Technology, 2013. [Online]. Available: http://digital-library.theiet.org/content/books/pc/pbsp010e

[36] Federal Bureau of Investigation. (2012, March) Integrated Automated Fingerprint Identification System. http://www.fbi.gov/about-us/cjis/fingerprints_biometrics/iafis/iafis.

[37] Federal Bureau of Investigation and US Dept of Justice, United States of America, "Classification of footprints," *FBI LAW ENFORCEMENT BULLETIN*, vol. 40, no. 9, pp. 18–22, 1971.

[38] J. Feng, Y. Shi, and J. Zhou, "Robust and efficient algorithms for separating latent overlapped fingerprints," *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 5, pp. 1498–1510, 2012.

[39] C. Fields, H. C. Falls, C. P. Warren, and M. Zimberoff, "The ear of the newborn as an identification constant," *Obstetrics & Gynecology*, vol. 16, no. 1, pp. 98–101, 1960.

[40] M. Fischler and R. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[41] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[42] K. M. Fox and C. C. Plato, "Toe and plantar dermatoglyphics in adult american caucasians," *American journal of physical anthropology*, vol. 74, no. 1, pp. 55–64, 1987.

[43] 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*. IEEE, 2006, pp. 30–30.

[44] F. Galton, *Finger prints*. Macmillan and Company, 1892.

[45] R. C. Geary, "The contiguity ratio and statistical mapping," *The incorporated statistician*, pp. 115–146, 1954.

[46] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (3rd Edition)*. Prentice Hall, 2007.

[47] C. Gottschlich, T. Hotz, R. Lorenz, S. Bernhardt, M. Hantschel, and A. Munk, "Modeling the growth of fingerprints improves matching for adolescents," *Information Forensics and Security, IEEE Transactions on*, vol. 6, no. 3, pp. 1165–1169, Sept 2011.

[48] D. Griffith, "Spatial autocorrelation," in *International Encyclopedia of Human Geography*, R. K. Thrift, Ed. Oxford: Elsevier, 2009, pp. 308 – 316. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780080449104005228

[49] E. Gutiérrez-Redomero, C. Alonso, E. Romero, and V. Galera, "Variability of fingerprint ridge density in a sample of spanish caucasians and its application to sex determination," *Forensic science international*, vol. 180, no. 1, pp. 17–22, 2008.

[50] R. C. Hauspie, N. Cameron, and L. Molinari, Eds., *Methods in Human Growth Research (Cambridge Studies in Biological and Evolutionary Anthropology)*. Cambridge University Press, Feb 2011.

[51] N. Hemy, A. Flavel, N. Ishak, and D. Franklin, "Estimation of stature using anthropometry of feet and footprints in a western australian population," *Journal of Forensic and Legal Medicine*, vol. 20, no. 5, pp. 435 – 441, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1752928X13000036

[52] E. R. Henry, *Classification and uses of finger prints*. London: George Routledge and Sons, 1900.

[53] R. A. Hicklin, "Anatomy of friction ridge skin," in *Encyclopedia of Biometrics*, S. Z. Li and A. K. Jain, Eds. Springer US, 2014, pp. 1–6. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-27733-7_48-3

[54] 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.

[55] L. Hong, Y. Wan, and A. K. Jain, "Fingerprint image enhancement: algorithm and performance evaluation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 8, pp. 777–789, 1998.

[56] K. J. Horadam, S. A. Davis, A. Arakala, and J. Jeffers, "Fingerprints as spatial graphs: Nodes and edges," in *Digital Image Computing Techniques and Applications (DICTA), 2011 International Conference on*. IEEE, 2011, pp. 400–405.

[57] R. Hummel, "Image enhancement by histogram transformation," *Computer graphics and image processing*, vol. 6, no. 2, pp. 184–195, 1977.

[58] IBM and Institute of Culinary Education, *Cognitive cooking with Chef Watson: recipes for innovation from IBM & the Institute of Culinary Education*. Naperville, Illinois: Sourcebooks, Inc, 2015.

[59] IBM Research. (2015, September) Watson and the Jeopardy! Challenge. https://www.youtube.com/watch?v=P18EdAKuC1U.

[60] IBM Research. (2015, September) What is Watson? https://chefwatson.zendesk.com/hc/en-us/articles/206296267.

[61] K. Inman and N. Rudin, *Principles and practice of criminalistics: the profession of forensic science*. CRC Press, 2000.

[62] International Labour Office, *Every child counts: new global estimates on child labour*. Geneva, Switzerland: International Labour Organization, April 2002.

[63] A. K. Jain, K. Cao, and S. S. Arora, "Recognizing infants and toddlers using fingerprints: Increasing the vaccination coverage," in *Biometrics (IJCB), 2014 International Joint Conference on*, October 2014.

[64] A. K. Jain, S. C. Dass, and K. Nandakumar, "Soft biometric traits for personal recognition systems," in *Biometric Authentication*. Springer, 2004, pp. 731–738.

[65] A. K. Jain and M. Demirkus, "On latent palmprint matching," MSU Technical Report, http://biometrics.cse.msu.edu/Publications/Palmprints/ JainDemirkusOnLatentPalmprintMatching08.pdf [Jan. 23, 2013], MSU-CSE-08-8, 2008.

[66] A. K. Jain, P. Flynn, and A. A. Ross, *Handbook of biometrics*. Springer US, 2008.

[67] W. Jia, R. Hu, J. Gui, and Y. Lei, "Newborn footprint recognition using band-limited phase-only correlation," in *Medical Biometrics*. Springer, 2010, pp. 83–93.

[68] John F. Kennedy Presidential Library and Museum. (2015) John F. Kennedy Quotations. http://www.jfklibrary.org/Research/Research-Aids/Ready-Reference/JFK-Quotations.aspx.

[69] M. Kücken and A. Newell, "A model for fingerprint formation," *EPL (Europhysics Letters)*, vol. 68, no. 1, p. 141, 2004.

[70] P. J. Kellman, J. L. Mnookin, G. Erlikhman, P. Garrigan, T. Ghose, E. Mettler, D. Charlton, and I. E. Dror, "Forensic comparison and matching of fingerprints: Using quantitative image measures for estimating error rates through understanding and predicting difficulty," *PloS one*, vol. 9, no. 5, p. e94617, 2014.

[71] J. Kotzerke, A. Arakala, S. Davis, K. Horadam, and J. McVernon, "Ballprints as an infant biometric: A first approach," in *Biometric Measurements and Systems for Security and Medical Applications (BIOMS) Proceedings, 2014 IEEE Workshop on*, Oct 2014, pp. 36–43.

[72] J. Kotzerke, S. Davis, K. Horadam, and J. McVernon, "Newborn and infant footprint crease pattern extraction," in *Proceedings of 2013 IEEE 20th International Conference on Image Processing (ICIP 2013)*, IEEE, Ed., 2013.

[73] J. Kotzerke, S. Davis, J. McVernon, and K. Horadam, "A solution to the infant biometric problem," *Nature*, submitted for publication.

[74] J. Kotzerke, S. A. Davis, R. Hayes, L. J. Spreeuwers, R. N. J. Veldhuis, and K. J. Horadam, "Discriminating fingermarks with evidential value for forensic comparison," in *Biometrics and Forensics (IWBF), 2015 International Workshop on*, March 2015.

[75] J. Kotzerke, S. Davis, H. Hao, R. Hayes, L. Spreeuwers, R. Veldhuis, and K. Horadam, "Identification performance of evidential value estimation for ridge-based biometrics," *EURASIP Journal on Information Security (JIS) on "Advances in Biometrics 2015"*, submitted for publication.

[76] J. Kotzerke, S. Davis, R. Hayes, J. McVernon, and K. Horadam, "Newborn and infant identification: revisiting footprints," *Forensic Science International*, submitted for publication.

[77] J. Kotzerke, S. Davis, R. Hayes, L. Spreeuwers, R. Veldhuis, and K. Horadam, "Identification performance of evidential value estimation for fingermarks," in *Biometrics Special Interest Group (BIOSIG), 2015 International Conference of the*, Sept 2015, pp. 1–6.

[78] B. Kraal and D. Dean. (2015, August) You're the voice – the science behind speaker recognition tech. http://theconversation.com/youre-the-voice-the-science-behind-speaker-recognition-tech-31579.

[79] D. Kroon. (2014) Finite Iterative Closest Point. http://www.mathworks.com/matlabcentral/fileexchange/24301.

[80] M. Kumar, A. Insan, N. Stoll, K. Thurow, and R. Stoll, "Stochastic fuzzy modeling for ear imaging based child identification," *Systems, Man, and Cybernetics: Systems, IEEE Transactions on*, 2015, to appear.

[81] H. Kumbnani, "Dermatoglyphics: a review," *Anthropology today: trends, scope and applications. Delhi, India: Kamla-Raj Enterprises*, pp. 285–95, 2007.

[82] Q. Y. Kwan, "Inference of identity of source," Ph.D. dissertation, University of California, Berkeley, 1977.

[83] M. Lahiri, C. Tantipathananandh, R. Warungu, D. I. Rubenstein, and T. Y. Berger-Wolf, "Biometric animal databases from field photographs: Identification of individual zebra in the wild," in *Proceedings of the 1st ACM International Conference on Multimedia Retrieval*, ser. ICMR '11. New York, NY, USA: ACM, 2011, pp. 6:1–6:8. [Online]. Available: http://doi.acm.org/10.1145/1991996.1992002

[84] S. M. Lajevardi, A. Arakala, S. A. Davis, and K. J. Horadam, "Retina verification system based on biometric graph matching," *Image Processing, IEEE Transactions on*, vol. 22, no. 9, pp. 3625–3635, Sept 2013.

[85] S. Lajevardi, A. Arakala, S. Davis, and K. Horadam, "Hand vein authentication using biometric graph matching," *Biometrics, IET*, vol. 3, no. 4, pp. 302–313, 2014.

[86] L. Lam, S.-W. Lee, and C. Y. Suen, "Thinning methodologies-a comprehensive survey," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 14, no. 9, pp. 869–885, 1992.

[87] S. Lee, H. Choi, K. Choi, and J. Kim, "Fingerprint-quality index using gradient components," *Information Forensics and Security, IEEE Transactions on*, vol. 3, no. 4, pp. 792–800, 2008.

[88] C. T. Lewis and C. Short, *A Latin Dictionary*, 1879.

[89] S. Z. Li and A. K. Jain, *Encyclopedia of Biometrics*. Springer US, 2014.

[90] K.-L. Low, "Linear least-squares optimization for point-to-plane icp surface registration," Department of Computer Science, University of North Carolina, Chapel Hill, North Carolina, Tech. Rep. TR04-004, 2004.

[91] D. G. Lowe, "Object recognition from local scale-invariant features," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, vol. 2, 1999, pp. 1150–1157.

[92] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[93] D. Maio, D. Maltoni, R. Capelli, A. Franco, M. Ferrara, and F. Turroni. Fvc-ongoing: on-line evaluation of fingerprint recognition algorithms. https://biolab.csr.unibo.it/fvcongoing/UI/Form/Home.aspx.

[94] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain, "Fvc2000: Fingerprint verification competition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 3, pp. 402–412, 2002.

[95] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain, "Fvc2002: Second fingerprint verification competition," in *Pattern recognition, 2002. Proceedings. 16th international conference on*, vol. 3. IEEE, 2002, pp. 811–814.

[96] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain, "Fvc2004: Third fingerprint verification competition," in *Biometric Authentication*. Springer, 2004, pp. 1–7.

[97] D. Maltoni, *Handbook of fingerprint recognition*. Springer, 2009.

[98] A. M. Martínez and A. C. Kak, "PCA versus LDA," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 2, pp. 228–233, 2001.

[99] S. Massey, "Persistence of creases of the foot and their value for forensic identification purposes," *Journal of Forensic Identification*, vol. 54, no. 3, pp. 296–315, 2004.

[100] G. McLachlan, K.-A. Do, and C. Ambroise, *Analyzing microarray gene expression data*. John Wiley & Sons, 2005, vol. 422.

[101] M. A. Medina-Pérez, O. Loyola-González, A. E. Gutierrez-Rodríguez, M. García-Borroto, and L. Altamirano-Robles, "Introducing an experimental framework in c# for fingerprint recognition," in *Pattern Recognition*. Springer, 2014, pp. 132–141.

[102] D. Meuwly, "Forensic individualisation from biometric data," *Science & Justice*, vol. 46, no. 4, pp. 205–213, 2006.

[103] D. Meuwly, "Forensic use of fingerprints and fingermarks," in *Encyclopedia of Biometrics*, S. Z. Li and A. K. Jain, Eds. Springer US, 2014, pp. 1–15. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-27733-7_181-3

[104] S. K. Modi, S. J. Elliott, J. Whetsone, and H. Kim, "Impact of age groups on fingerprint recognition performance," in *Automatic Identification Advanced Technologies, 2007 IEEE Workshop on*. IEEE, 2007, pp. 19–23.

[105] R. Montgomery, "Sole prints of newborn babies," *The American Journal of the Medical Sciences*, vol. 169, no. 6, pp. 830–837, 1925.

[106] R. Montgomery, "Classification of foot-prints," *Journal of the American Institute of Criminal Law and Criminology*, vol. 18, no. 1, pp. 105–110, 1927.

[107] R. B. Montgomery, "Sole patterns. a study of the footprints of two thousand individuals," *The Anatomical Record*, vol. 33, no. 2, pp. 107–114, 1926.

[108] R. B. Montgomery, "Sole patterns of twins," *Biological Bulletin*, vol. 50, no. 4, pp. 293–300, 1926.

[109] P. A. P. Moran, "Notes on continuous stochastic phenomena," *Biometrika*, vol. 37, no. 1-2, pp. 17–23, 1950.

[110] A. Z. Mundorff, E. J. Bartelink, and T. A. Murad, "Sexual dimorphism in finger ridge breadth measurements: A tool for sex estimation from fingerprints," *Journal of Forensic Sciences*, vol. 59, no. 4, pp. 891–897, 2014. [Online]. Available: http://dx.doi.org/10.1111/1556-4029.12449

[111] J. Munkres, "Algorithms for the assignment and transportation problems," *Journal of the Society for Industrial and Applied Mathematics*, vol. 5, no. 1, pp. 32–38, 1957.

[112] National Institute of Standards and Technology. (2015, October) Biometric Quality Homepage. http://www.nist.gov/itl/iad/ig/bio_quality.cfm.

[113] National Institute of Standards and Technology. (2015, October) Mission, Vision, Core Competencies, and Core Values. http://www.nist.gov/public_affairs/mission.cfm.

[114] Neurotechnology. (2015) VeriFinger SDK. http://www.neurotechnology.com/verifinger.html.

[115] Neurotechnology. (2015) VeriFinger SDK Documentation. http://download.neurotechnology.com/Neurotec_Biometric_SDK_Documentation.pdf.

[116] M. Nithin, B. Manjunatha, D. Preethi, and B. Balaraj, "Gender differentiation by finger ridge count among south indian population," *Journal of forensic and legal medicine*, vol. 18, no. 2, pp. 79–81, 2011.

[117] M. Okajima, "Frequency of forks in epidermal-ridge minutiae in the finger print," *American Journal of Physical Anthropology*, vol. 32, no. 1, pp. 41–48, 1970. [Online]. Available: http://dx.doi.org/10.1002/ajpa.1330320106

[118] M. Okajima, "Epidermal-ridge minutia in the hallucal area," *Mitteilungen der Anthropogischen Gesellschaftin in Wien*, vol. 107, pp. 135–139, 1977.

[119] M. Okajima, "Dermal and epidermal structures of the volar skin." *Birth defects original article series*, vol. 15, no. 6, pp. 179–198, 1979.

[120] E. Pekalska and R. Duin, *The dissimilarity representation for pattern recognition: foundations and applications*. World Scientific Pub Co Inc, 2005, vol. 64.

[121] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," *Computer vision, graphics, and image processing*, vol. 39, no. 3, pp. 355–368, 1987.

[122] N. Poh, S. Bengio, and A. Ross, "Revisiting doddington's zoo: A systematic method to assess user-dependent variabilities," in *Proc. of Second Workshop on Multimodal User Authentication (MMUA)*, 2006, p. 119.

[123] K. Riesen and H. Bunke, "Approximate graph edit distance computation by means of bipartite graph matching," *Image and Vision Computing*, vol. 27, no. 7, pp. 950–959, 2009.

[124] K. Riesen and H. Bunke, "Graph classification based on vector space embedding," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 23, no. 06, pp. 1053–1081, 2009.

[125] C. M. Rodriguez, A. de Jongh, and D. Meuwly, "Introducing a semi-automatic method to simulate large numbers of forensic fingermarks for research on fingerprint identification," *Journal of Forensic Sciences*, vol. 57, no. 2, pp. 334–342, 2012. [Online]. Available: http://dx.doi.org/10.1111/j.1556-4029.2011.01950.x

[126] S. Rusinkiewicz and M. Levoy, "Efficient variants of the icp algorithm," in *3-D Digital Imaging and Modeling, 2001. Proceedings. Third International Conference on*. IEEE, 2001, pp. 145–152.

[127] J. K. Schneider, "Quantifying the dermatoglyphic growth patterns in children through adolescence," https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=254839, US Department of Justice, Tech. Rep. FR00A178000-1, Dec 2010.

[128] J. Serra and L. Vincent, "An overview of morphological filtering," *Circuits, Systems, and Signal Processing*, vol. 11, no. 1, pp. 47–108, 1992.

[129] K. Shepard, T. Erickson, and H. Fromm, "Limitations of footprinting as a means of infant identification," *Pediatrics*, vol. 37, no. 1, pp. 107–108, 1966.

[130] N. J. Short, M. Hsiao, A. Abbott, and E. Fox, "Latent fingerprint segmentation using ridge template correlation," in *Imaging for Crime Detection and Prevention 2011 (ICDP 2011), 4th International Conference on*, 2011, pp. 1–6.

[131] L. Silva, O. Bellon, R. de Paula Lemes, J. Meira, and M. Cat, "An image processing tool to support gestational age determination," in *Computer-Based Medical Systems, 2006. CBMS 2006. 19th IEEE International Symposium on*. IEEE, 2006, pp. 867–874.

[132] R. Sinclair and C. Fox, "Infant-to-adult footprint identification," *Journal of Forensic Identification*, vol. 57, no. 4, p. 485, 2007.

[133] L. T. Staheli, *Fundamentals of pediatric orthopedics*. Lippincott Williams & Wilkins, 2008.

[134] M. E. Stapleton, "Best foot forward: Infant footprints for personal identification," *FBI LAW ENFORCEMENT BULLETIN*, vol. 63, no. 11, pp. 14–17, 1994.

[135] D. A. Stoney, "What made us ever think we could individualize using statistics?" *Journal of the Forensic Science Society*, vol. 31, no. 2, pp. 197–199, 1991.

[136] M. Sudesh Gungadin, "Sex determination from fingerprint ridge density," *Internet Journal of Medical Update*, vol. 2, no. 2, 2007.

[137] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, no. 3, pp. 293–300, 1999.

[138] E. Tabassi, C. Wilson, and C. Watson, "NIST Fingerprint Image Quality," http://www.nist.gov/customcf/get_pdf.cfm?pub_id=905710, National Institute of Standards and Technology, Gaithersburg, Maryland, Tech. Rep. NISTIR7151, 2004.

[139] The National Institute of Standards and Technology. (2013, April) NFIQ2 feature definitions document (v0.5). http://biometrics.nist.gov/cs_links/quality/NFIQ_2/NFIQ-2_Quality_Feature_Defin-Ver05.pdf.

[140] J. Thompson, D. Clark, B. Salisbury, and J. Cahill, "Footprinting the newborn infant: Not cost effective," *Journal of Pediatrics*, vol. 99, no. 5, pp. 797–798, 1981.

[141] S. Tiwari, A. Singh, and S. K. Singh, "Newborn's ear recognition: Can it be done?" in *Image Information Processing (ICIIP), 2011 International Conference on*, Nov 2011, pp. 1–6.

[142] S. Tiwari and S. K. Singh, "Face recognition for newborns," *Biometrics, IET*, vol. 1, no. 4, pp. 200–208, Dec 2012.

[143] A. Uhl and P. Wild, "Footprint-based biometric verification," *Journal of Electronic Imaging*, vol. 17, no. 1, pp. 1–10, 2008.

[144] A. Uhl and P. Wild, "Personal identification using eigenfeet, ballprint and foot geometry biometrics," in *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on.* IEEE, 2007, pp. 1–6.

[145] A. Uhl and P. Wild, "Ageing effects in fingerprint biometrics," in *Age Factors in Biometric Processing*, ser. Professional Applications of Computing, M. Fairhurst, Ed. Institution of Engineering and Technology, 2013, pp. 153–170. [Online]. Available: http://digital-library.theiet.org/content/books/10.1049/pbsp010e_ch8

[146] B. T. Ulery, R. A. Hicklin, J. Buscaglia, and M. A. Roberts, "Accuracy and reliability of forensic latent fingerprint decisions," *Proceedings of the National Academy of Sciences*, vol. 108, no. 19, pp. 7733–7738, 2011.

[147] B. T. Ulery, R. A. Hicklin, J. Buscaglia, and M. A. Roberts, "Repeatability and reproducibility of decisions by latent fingerprint examiners," *PloS one*, vol. 7, no. 3, p. e32800, 2012.

[148] Unique Identification Authority of India. (2014) Aadhaar. http://uidai.gov.in/.

[149] United Nations Children's Emergency Fund. (2013, December) Every child's birth right: Inequities and trends in birth registration. http://www.unicef.org/publications/index_71514.html.

[150] United Nations Office on Drugs and Crime, "Estimating illicit financial flows resulting from drug trafficking and other transnational organized crimes," http://www.unodc.org/documents/data-and-analysis/Studies/Illicit_

financial_flows_2011_web.pdf, United Nations Office on Drugs and Crime, Vienna, Austria, Tech. Rep., October 2011.

[151] U.S. Department of Justice. (2013, May) A review of the fbi's handling of the brandon mayfield case. http://www.justice.gov/oig/special7s0601/exec.pdf.

[152] U.S. Department of Justice. (2013, May) A review of the fbi's progress in responding to the recommendations in the office of the inspector general report on the fingerprint misidentification in the brandon mayfield case. http://www.justice.gov/oig/special/s1105.pdf.

[153] R. Važan. (2015, August) SourceAFIS. http://www.sourceafis.org/blog/.

[154] V. Vapnik and A. Lerner, "Generalized portrait method for pattern recognition," *Automation and Remote Control*, vol. 24, no. 6, pp. 774–780, 1963.

[155] D. Weingaertner, O. R. P. Bellon, L. Silva, and M. N. L. Cat, "Newborn's biometric identification: Can it be done?" in *Proceedings of the Third International Conference on Computer Vision Theory and Applications*, 2008, pp. 200–205.

[156] K. Wertheim and A. Maceo, "The critical stage of friction ridge and pattern formation," *Journal of Forensic Identification*, vol. 52, no. 1, p. 35, 2002.

[157] H. H. Wilder, "Racial differences in palm and sole configuration," *American Anthropologist*, vol. 6, no. 2, pp. 244–293, 1904.

[158] H. H. Wilder and B. Wentworth, *Personal identification: methods for the identification of individuals, living or dead*. Richard G. Badger, 1918.

[159] World Health Organization. (2006) Child growth standards. http://www.who.int/childgrowth/standards/en/.

[160] World Health Organization. (2015) Child malnutrition – underweight, stunting, wasting and overweight. http://apps.who.int/nutrition/landscape/help.aspx?menu=0&helpid=391&lang=EN.

[161] World Health Organization. (2015) Infant, newborn. http://www.who.int/topics/infant_newborn/en/.

[162] C. Wu, S. Tulyakov, and V. Govindaraju, "Image quality measures for fingerprint image enhancement," in *Multimedia Content Representation, Classification and Security*. Springer, 2006, pp. 215–222.

[163] X. Wu, "An efficient antialiasing technique," in *ACM SIGGRAPH Computer Graphics*, vol. 25, no. 4. ACM, 1991, pp. 143–152.

[164] S. Yoon, J. Feng, and A. K. Jain, "Altered fingerprints: Analysis and detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 3, pp. 451–464, 2012.

[165] S. Yoon and A. K. Jain. (2013, April) Quality assessment of latent fingerprints. http://biometrics.nist.gov/cs_links/quality/NFIQ_2/presentations_4-26/nfiq2_yoon-2013-04-26.pdf.

[166] S. Yoon and A. K. Jain, "Longitudinal study of fingerprint recognition," *Proceedings of the National Academy of Sciences*, vol. 112, no. 28, pp. 8555–8560, 2015. [Online]. Available: http://www.pnas.org/content/112/28/8555.abstract

[167] Y. Zhu, S. C. Dass, and A. K. Jain, "Statistical models for assessing the individuality of fingerprints," *Information Forensics and Security, IEEE Transactions on*, vol. 2, no. 3, pp. 391–401, 2007.

[168] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics gems IV*. Academic Press Professional, Inc., 1994, pp. 474–485.