UNIVERSIDAD AUTÓNOMA DE MADRID

ESCUELA POLITÉCNICA SUPERIOR

DEPARTAMENTO DE TECNOLOGÍA ELECTRÓNICA Y DE LAS COMUNICACIONES

FINGERPRINT RECOGNITION FOR FORENSIC APPLICATIONS

–TESIS DOCTORAL–

RECONOCIMIENTO DE HUELLAS DACTILARES PARA APLICACIONES FORENSES

Author: Ram Prasad Krishnamoorthy
(Master's in Computer Science,
Chennai Mathematical Institute, India)

A Thesis submitted for the degree of:

Doctor of Philosophy

Madrid, May 2015

Colophon

This book was typeset by the author using $\LaTeX$2e. The main body of the text was set using a 11-points Computer Modern Roman font. All graphics and images were included formatted as Encapsuled Postscript (^{TM} Adobe Systems Incorporated). The final postscript output was converted to Portable Document Format (PDF) and printed.

Department:	Tecnología Electrónica y de las Comunicaciones Escuela Politécnica Superior Universidad Autónoma de Madrid (UAM), SPAIN
PhD Thesis:	Fingerprint Recognition for Forensic Applications
Author:	**Ram Prasad Krishnamoorthy** Master's in Computer Science Chennai Mathematical Institute, INDIA
Advisor:	**Julián Fiérrez Aguilar** Doctor Ingeniero de Telecomunicación (Universidad Politécnica de Madrid) Universidad Autónoma de Madrid, SPAIN
Co-advisor:	**Daniel Ramos Castro** Doctor Ingeniero de Telecomunicación (Universidad Autónoma de Madrid) Universidad Autónoma de Madrid, SPAIN
Year:	2015
Committee:	President: **Javier Ortega-Garcia** Universidad Autónoma de Madrid, SPAIN Secretary: **Roberto Paredes Palacios** Universidad Politécnica de Valencia, SPAIN Vocal 1: **Juan Carlos San Miguel Avedillo** Universidad Autónoma de Madrid, SPAIN Vocal 2: **Raymond Nicolaas Johan Veldhuis** University of Twente, THE NETHERLANDS Vocal 3: **Josef Bigun** Halmstad University, SWEDEN

The research described in this Thesis was carried out within the Biometric Recognition Group – ATVS at the Departamento de Tecnología Electrónica y de las Comunicaciones, Escuela Politécnica Superior, Universidad Autónoma de Madrid (from 2011 to 2015). The project was funded by European Commission Marie Curie Fellowship Project - Bayesian Biometrics for Forensics (BBfor2) and the European Union Project - Biometrics Evaluation and Testing (BEAT).

The author was awarded with a European Commission Marie Curie Fellowship under the Innovative Training Networks (ITN) in the project Bayesian Biometrics for Forensics (BBfor2, FP7-PEOPLE-ITN-2008) under Grant Agreement number 238803 between 2011 and 2013. The author was also funded through the European Union Project - Biometrics Evaluation and Testing (BEAT) for 2014 and 2015 which supported the research summarized in this Dissertation.

The author's paper was selected as one of the best papers in IEEE 2^{nd} International Workshop on Biometrics and Forensics, for one publication originated from this Dissertation: Ram P. Krish, Julian Fierrez, Daniel Ramos, Javier Ortega-Garcia and Josef Bigun, "Partial Fingerprint Registration for Forensics using Minutiae-Generated Orientation Fields", in IEEE 2^{nd} International Workshop on Biometrics and Forensics (IWBF), Valletta, Malta, March 2014.

Abstract

FORENSIC LATENT FINGERPRINT RECOGNITION has taken a significant transition in recent decades from a fully manual identification procedure to the incorporation of Automated Fingerprint Identification Systems (AFIS) by law enforcement agencies to identify suspects. Latent fingerprints are those fingerprints that are revealed using chemical or optical processing and collected from a crime scene by specialists trained in forensic sciences. These latent fingerprints are then compared to reference fingerprints stored in the AFIS database called tenprints or exemplars. However, the latent fingerprints thus obtained from crime scenes are generally of very poor quality. They can be highly partial in nature, non-linearly deformed, smudgy, and with overlapped fingerprints. These characteristics of latent fingerprints introduce many challenges in employing a fully automatic latent fingerprint matching.

This PhD Thesis is focused on solving some of the major challenges in automated latent fingerprint identification. In particular, the Thesis focuses on improving the performance accuracies of minutiae-based fingerprint matchers for latent fingerprints by exploring the problem of partial fingerprint to full fingerprint matching. Further, we also develop an evidence evaluation model for the matched fingerprints based on likelihood ratio from the similarity scores generated by the minutiae-based matchers. More specifically, the main contributions of the Thesis can be divided in three blocks.

First, most fingerprint matching algorithms in general assume approximately the same size of the minutiae set, the set of key features of a fingerprint based on which comparisons can be made, between the query and the reference minutiae for good identification accuracy. In practice, however, it is frequent that the size of latent minutiae set is very small compared to the size of tenprint minutiae sets, as a result of which matching performance is hampered. To make the best use of minutiae-based matchers, it is advantageous to know the location of partial fingerprint minutiae pattern in the full fingerprint minutiae pattern, thereby reducing the minutiae search space to improve the matching performance. However, existing minutiae based alignment techniques are not well adapted to use in partial fingerprint alignment. An image-based registration is also not feasible due to the poor quality of latent fingerprints. In the first part of this Thesis, we focus on the problem of aligning a partial fingerprint against a full fingerprint, especially of poor quality latents. Instead of minutiae, we used orientation fields (OF) to perform the alignment. We reduce fingerprint images to orientation images, and we look at the alignment problem as registering the partial fingerprint orientation image into the full fingerprint orientation image. To achieve this task, we develop a new correlation-based hierarchical registration method for orientation images to register a partial fingerprint in a full fingerprint. The OF representing the flow of ridges is a relatively stable global feature of fingerprint images. We experimentally demonstrate a significant improvement in the rank identification accuracies for minutiae-based matchers by incorporating our registration algorithm

to reduce the search space of minutiae in full fingerprints. We also demonstrate the usefulness of our proposed method as a fully automatic tool.

Second, AFIS use only a limited subset of the types of discriminatory features which can be automatically extracted from the fingerprint images using a feature extraction algorithm. On the other hand, forensic examiners use a richer set of features during manual comparison as compared to AFIS comparisons. This could be a possible reason why manual comparisons outperform AFIS comparisons. The features not currently used by commercial AFIS are generally termed as Extended Feature Sets (EFS). Many commercial minutiae-based matchers do not use EFS. They mostly use only two prominent ridge characteristics namely ridge-endings and bifurcations. To use EFS in automated systems, reliable feature extraction algorithms are mandatory. In the second part of Dissertation, we focus on the problem of using EFS in a typical minutiae-based matcher. A realistic database from forensic fingerprint casework consisting of rare minutiae features was obtained from the Spanish law enforcement agency, Guardia Civil. We propose a method to improve the identification accuracy of minutiae-based matchers for partial latent fingerprints by incorporating reliably extracted rare minutiae features. Our proposed algorithm modifies the similarity scores of minutiae-based matchers based on the presence of rare minutia features like fragments, enclosures, dots, interruptions, etc. These rare features are used to automatically estimate an affine function that transforms the latent minutiae set to the tenprint minutiae set, generating a fitting error which is then used to adjust the baseline minutiae-based matching score. We experimentally demonstrate significant improvement in the rank identification accuracies of minutiae-based matchers when their similarity scores are modified in this way.

Third, the uniqueness of a fingerprint is not an established fact but only an empirical observation. There is a widespread concern about the scientific basis underlying the individuality of fingerprints, especially when using them in the court of law. Many individualization models for fingerprints have been proposed in the research literature. However, there is no scientific framework in use at the criminal justice system to characterize the uncertainty involved in the friction ridge analysis methodology, as well as to express the strength of opinion of the forensic examiner quantitatively. Such a requirement has been articulated in several influential reports like the National Research Council 2009 report "Strengthening Forensic Science in the United States: A Path Forward". The new paradigm coming forward in this regard avoids hard identification decisions by considering evidence reporting methods that incorporate uncertainty and statistics. Among all the methods for evidence evaluation, the likelihood ratio has shown much promise and is receiving greater attention. Using the technique we developed for improving the rank identification accuracies of minutiae-based matchers by incorporating rare minutiae features outlined above, we build a robust likelihood ratio evidence evaluation model for individualization.

In summary, in this Disseration we address three key challenges for automated latent fingerprint matching: 1) partial fingerprint registration using Orientation Fields, 2) use of Extended Feature Sets, and 3) development of a robust evidence evaluation tool.

Resumen

El reconocimiento forense de huellas dactilares latentes ha evolucionado considerablemente en las últimas décadas mediante la incorporación de sistemas automáticos (*Automated Fingerprint Identification Systems*, AFIS) para identificar sospechosos por parte de las fuerzas de seguridad. Cualquier impresión producida por el contacto de dedos humanos se conoce generalmente como huella dactilar. Las imágenes de huellas dactilares obtenidas mediante el procesado químico u óptico de escenas criminales por especialistas forenses se denominan huellas dactilares latentes. Dichas huellas dactilares recuperadas de escenas criminales tienen frecuentemente una calidad muy baja. Pueden ser parciales, deformadas de modo no lineal, borrosas, o puede haber varias huellas superpuestas, etc. Estas características de las huellas dactilares latentes introducen numerosos desafíos en el uso de comparadores completamente automáticos de huellas dactilares latentes. Las huellas dactilars latentes se comparan usando sistemas AFIS con huellas dactilares de referencia previamente almacenadas en una base de datos.

Esta Tesis Doctoral se centra en resolver algunos de los mayores desafíos en la identificación automática de huellas dactilares latentes. En particular, la Tesis de centra en la mejora de la precisión de la identificación de comparadores de huellas dactilares basados en minucias mediante la exploración del problema de la comparación de huellas dactilares parciales, y desarrolla un modelo de valoración de las evidencias basado en ratios de verosimilitud de las puntuaciones de similitud generadas por comparadores basados en minucias. Las tres líneas principales de investigación cubiertas en esta Tesis se resumen a continuación.

En primer lugar, para obtener una identificación precisa, la mayor parte de los algoritmos de comparación de huellas dactilares asumen que el tamaño del conjunto de minucias es aproximadamente igual entre las minucias de referencia y las de entrada. Sin embargo, el tamaño de la huella latente de entrada es frecuentemente muy pequeño en comparación con el tamaño de la referencia. Con el objetivo de aprovechar al máximo los comparadores basados en minucias, sería ventajoso conocer la posición del patrón de minucias de la huella parcial con respecto al patrón completo, reduciendo de este modo la búsqueda en el espacio de minucias para mejorar el rendimiento. Las técnicas de alineado de minucias existentes no se adaptan bien al uso de huellas dactilares parciales. Un registro basado en imagen tampoco es factible dada la baja calidad de las huellas latentes. En la primera parte de esta Tesis, nos centramos en el problema del alineamiento de huellas dactilares parciales respecto a huellas completas, especialmente en el caso de huellas latentes de baja calidad. En lugar de minucias, usamos campos de orientación (*Orientation Fields*, OF) para el alineamiento. Reducimos las imágenes de las huellas a imágenes de orientación, y tratamos el problema del alineamiento como el registro de imágenes de orientación de huellas dactilares parciales dentro de la imagen completa. El OF que representa el flujo de las crestas es una representación relativamente estable de características globales de huellas dactilares. El dispositivo de captura, las variaciones de contraste y otros efectos de calidad no afectan mucho a la representación OF de la huella dactilar comparada con la imagen de entrada o las

minucias. Se ha desarrollado un nuevo método jerárquico de registro de huellas parciales para imágenes de orientación basado correlación, y demostramos experimentalmente la significativa mejora en la precisión de la identificación para comparadores basados en minucias al incorporar nuestro algoritmo de registro para reducir el espacio de búsqueda de minucias en huellas completas. También probamos la utilidad del nuestro método como herramienta completamente automática.

En segundo lugar, los sistemas AFIS usan normalmente tipos limitados de características extraídos de las huellas dactilares. Por otra parte, los examinadores forenses utilizan conjuntos más ricos de características durante la comparación manual, con respecto a las comparaciones de los AFIS. Ésta podría ser la razón por la que la comparación manual funciona mejor que la realizada por los AFIS. Las características que actualmente no son usadas por AFIS comerciales se denominan generalmente Conjuntos de Características Extendidas (*Extended Feature Sets*, EFS). En su mayor parte, los sistemas AFIS utilizan sólo dos características distintivas de las crestas llamadas final de cresta y bifurcación. Para usar EFS en sistemas automáticos se necesitan algoritmos de extracción de características fiables. En la segunda parte de la Disertación, nos centramos en la incorporación de EFS a comparadores típicos basados únicamente en minucias. Se ha utilizado para ello una nueva base de datos de huellas dactilares forenses obtenida de casos reales gracias a la Guardia Civil. Proponemos un método para mejorar la precisión de los comparadores basados en minucias para huellas latentes parciales al incorporar minucias atípicas extraídas de forma fiable. El algoritmo propuesto modifica las puntuaciones de similitud de los comparadores basados en minucias teniendo en cuenta la información proporcionada por dichas minucias atípicas. Dichas minucias atípicas o de baja frecuencia de aparición se usan para estimar una función afín que transforma el conjunto de minucias latentes al conjunto de minucias de la referencia, y cuyo error de ajuste se usa para adaptar la puntuación de similitud. Demostramos con ello experimentalmente una mejora significativa en la precisión de la identificación de comparadores basados en minucias.

En tercer lugar, la unicidad de la huella dactilar no es un hecho establecido sino solamente una observación empírica. Hay una preocupación generalizada acerca de la base científica que soporta la individualidad de las huellas dactilares, especialmente cuando se usan en un juicio. Se han propuesto numerosos modelos de individualización de huellas dactilares en la literatura. No hay un marco científico en uso en el sistema de justicia que caracterice la incertidumbre en la metodología de análisis y comparación de huellas dactilares, así como la expresión cuantitativa de la certeza de la opinión del examinador forense. Dicha necesidad se ha manifestado en diversos informes influyentes como el informe del Consejo de Investigación Nacional Americano de 2009 "*Strengthening Forensic Science in the United States: A Path Forward*". El nuevo paradigma que está surgiendo a este respecto evita decisiones rígidas al considerar métodos para presentar evidencias que incorporan incertidumbre y estadísticas poblacionales. Entre todos los métodos de evaluación de evidencias, el ratio de verosimilitud está recibiendo una gran atención. Usando la técnica desarrollada para mejorar la precisión de la identificación de ranking de comparadores basados en minucias con nuestro método de modificación de puntuaciones basado en

EFS, hemos construido un modelo robusto de ratios de verosimilitud, que sirve para cuantificar estadísticamente el peso de la evidencia en la comparación forense de huellas dactilares.

En resumen, en esta Tesis se abordan tres problemas fundamentales para el uso de la huella dactilar como evidencia forense: 1) registro de huellas parciales basado en campos de orientación (OF), 2) uso de conjuntos extendidos fiables de características (EFS), y 3) desarrollo de metodología estadística para la valoración estadística de la evidencia.

To my Parents.
To my Sister.

Science is what we understand well
enough to explain to a computer.
Art is everything else we do.

– Donald E. Knuth

Acknowledgements

THIS PHD THESIS summarizes the work carried out during my Ph.D. studies with the ATVS - Biometric Recognition Group since 2011. This research group was established in 1994 at the Dept. of Ingeniería Audiovisual y Comunicaciones (DIAC) of the Universidad Politécnica de Madrid (UPM) and since 2004 is affiliated to the Escuela Politécnica Superior of the Universidad Autónoma de Madrid (UAM). The financial support for the Ph.D. studies came from European Commission Marie Curie Fellowship under the Innovative Training Networks (*ITN*) in the project Bayesian Biometrics for Forensics (*BBfor2*), and through the European Union Project - Biometrics Evaluation and Testing (*BEAT*).

I owe my foremost gratitude to my advisor, Dr. Julian Fierrez for providing tremendous guidance and support for the last four years. I was always been amazed at his ability in focusing on high quality research work, and the way he inspires others through his dedicated hardwork. He has excellent insight at suggesting research problems that are well motivated and challenging. My interactions with him have greatly shaped my perspectives about research. I am especially thankful for his support and encouragement during many difficult situations encountered while progressing with research, and for the confidence he had in me. He opened up many opportunity to work on problems which are not directly related to this thesis. This has helped me to improve my work experience, and publish papers in fields which are not my main stream PhD research. I consider this a very valuable asset for which I owe my deep gratitude to him.

I am grateful to my co-advisor, Dr. Daniel Ramos for being a wonderful mentor, providing valuable inputs and suggestions on various topics, and his patience and ability to make sure whatever he communicated has been understood properly by me. I am also very thankful to him for many discussions that in general focused on the philosophy of life, about diverse cultures and the moral goodness present in them. Working together with both Julian and Daniel was a great learning experience for me, and I strongly believe that the qualities which I have learned from them will help me become a great mentor in my future endeavors in teaching and research. Thanks to you Julian and Daniel for the kind of people you are.

During my doctoral studies, I had the opportunity to work with Prof. Josef Bigun at Halmstad University, Sweden and with Prof. Raymond Veldhuis at University of Twente, The Netherlands, for four months each. It was a great experience visiting another research lab and collaborate with them on my research problems which helped in progressing the research into another higher levels. During my stay at Halmstad, I also had the opportunity to finish a graduate level course on Image Analysis by Prof. Josef Bigun, which helped me obtain more theoretical insight into many image analysis algorithms. This has helped me to adapt some algorithms into my research which directly helped in solving the problem. Working with Prof. Raymond was a great experience, and his insight in thinking about possible difficulties that may come up when dealing with a problem was of great help. My regular interactions with them have helped me shape my perspectives about doing research, and how to work under time constraints.

I am grateful to Prof. Javier Ortega-García and Prof. Joaquin González-Rodríguez of ATVS group for their support extended by them, and enquiring and keeping track about the progress and status of my doctoral work. I thank Prof. Dorreteo Torre-Toledano for his help with administrative issues. I especially thank Dr. Aythami Morales, a post doctoral research fellow at ATVS who helped me with the discussions on many of the practical issues related with my research, and provided helpful suggestions in improving some experiments. I thank all my work mates at ATVS (in no particular order): Ruben Vera, Ruben Tolosana, Ruben Zazo, Javier Franco-Pedroso, Alicia Lozano, Ester Gonzalez-Sosa, Marta Gomez-Barrero, Fatima García-Donday.

I thank Henk van den Heuvel and David van Leeuwen, the organizers of the BBfor2 project, and other Marie Curie research fellows (in no particular order): Ruifang Wang, Anna Mikaelyan, Abhishek Dutta, Tauseef Ali, Rudolf Haraksim , Mohamed Hasan Bahari, Nathalie Fecher, Erica Gold, Manuel Gunther, Laurent El Shafey, Roy Wallace, Vasileos Vasilakakis, Rahim Saeidi, Mira Indar Mandassari and Mitchell McLaren through whom I really had the great opportunity of establishing contacts with an international team of researchers. This was really a great learning experience. I also extend my special thanks to other principal investigators involved in the BBfor2 project (in no particular order): Didier Meuwly, Hugo Van Hamme, Dominic Watt, Sebastien Marcel, Pietro Laface.

Above all, I thank my parents and sister who have always been very understanding and encouraging, showering unconditional love and affection upon me all the time. My gratitude to them is beyond expression.

Ram Prasad Krishnamoorthy
Madrid, May 2015

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Forensic sciences play a critical role in the criminal justice system by providing the crime investigators with scientifically based information through the analysis of physical evidences obtained from the crime scenes. A crime investigation typically involves collecting forensic evidences from the crime scenes, analyze the evidences in laboratory and then present the conclusions in the court. Different crime cases presents different types of challenges. Some cases might involve collecting and analyzing large amount of evidences. Conclusions derived from multiple evidence will then be combined to produce the objective results to the case under investigation.

In this chapter, we briefly describe various types of evidences generally involved in the forensic sciences, an overview of latent fingerprints, types of forensic testimony standards to be followed for general acceptance in courts, an overview of current practices in friction ridge analysis, the challenges faced towards individualization of forensic fingerprints, and a brief overview of the use of newer technologies to reduce human-errors in the examination process. We will discuss the main recommendations put forward by law enforcement agencies towards improving the friction ridge analysis, and briefly describe an overview of this Dissertation contributing in solving some of the challenges addressed.

1.1. Types of evidences in forensic sciences

Forensic evidences constitute all the means by which any alleged matter of fact whose truth is investigated at judicial trial is established or disproved. Admissible evidence are those which a court admits for judges and juries to consider for the deciding of a particular case. Forensic evidences can originate from diverse sources - from genetic material or trace chemicals to dental history or fingerprints. Evidence can serve many roles in an investigation, such as to trace an illicit substance, identify remains or reconstruct a crime [U.S.Department, 2013].

Various types of forensic evidences usually collected with regards to a crime are as follows [U.S.Department, 2013]:

1. *Controlled Substances*: The presence of illegal drugs such as heroin, marijuana or regulated prescription medications are analyzed in the crime scene or related scenarios by the forensic examiners. Those chemicals that have a legally recognized potential for abuse are termed as controlled substances. A major step taken by law enforcement agencies to fight against drug-related crime and violence is by detecting and identifying controlled substances.

2. *Digital Evidence*: Any evidences which in the form of information stored or transmitted in binary forms which may be trusted by court are termed as digital evidence. Such evidences may be retrieved from storage medias such as computer hard drive, memory cards of mobile phone, a CD, floppy disc, flash cards of camera, online data storage etc.

3. *Forensic Anthropology*: The examination of the human skeletons or the remains in an advanced stage of decomposition is called forensic anthropology. It involves estimating the time of death and the means of injury, or assessing the age, gender, height and ancestry of the victim. These sort of examinations helps the crime investigators to identify the victims.

4. *Forensic Dentistry*: Forensic dentistry is the application of dental knowledge towards assisting investigators in identifying the human remains. This involves examining the development, anatomy and any restorative dental corrections of the teeth, such as fillings, to make a comparative identification of a person.

5. *DNA Evidence*: DNA testing or DNA profiling is the technique followed by forensic examiners to identify individuals by examining the characteristics of their DNA. For unrelated individuals, the DNA profile shows small variations among their DNA. Only 0.1% of the DNA differs from one person to another except for monozygotic twins. Scientists can use these variable regions to generate a DNA profile of an individual, using samples from blood, bone, hair, and other body tissues and products.

6. *Forensic Pathology*: It is the science by which the cause of death is determined by examining the corpse. Forensic pathologist is a medical doctor who has specialized in forensic pathology. An autopsy is conducted by the medical examiner focusing to determine the cause and manner of the death that is violent, unusual or untimely. Also, the medical examiner may identify a wound pattern that can be matched to a weapon, or can determine entry and exit wounds in deaths involving firearms and other projectiles.

7. *Forensic Toxicology*: The chemical analysis of biological samples for the presence of poison or drugs is termed as toxicology. When such a technique is used to determine a substance's contribution towards an individual's death, illness, or physical or mental impairment is called forensic toxicology. The toxicology report can provide key information as to the type

of substances present in an individual and if the amount of those substances is consistent with a therapeutic dosage or is above a harmful level.

8. *Questioned Documents*: Questioned documents are those documents which needs to verify the authenticity that will aid an investigation or serve the purpose of evidence in court. Through visual examination or advanced chemical analysis of inks and paper, forensic investigators can determine information relating to a questioned document's authentication, authorship or creation date.

9. *Trace Evidence*: Trace evidences are those that are transferred between people, objects or the environment during the crime. They can be used to reconstruct crimes, and to describe the people, places and things involved in them. Fibers, hair, soil, wood, gunshot residue and pollen are few examples of trace evidence.

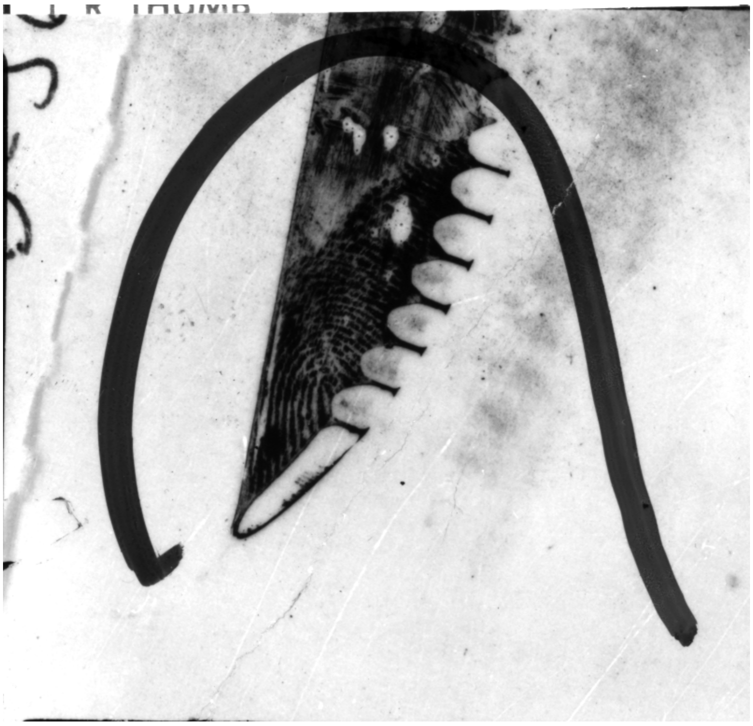
10. *Impression and Pattern Evidence*: Pattern and impression evidence include any markings produced when one object comes into contact with another object, such as latent fingerprints, shoe-prints, toolmarks, and tire treads. It also includes pattern analysis, such as is used when evaluating handwriting, typewriting, and writing instruments.

Among these forensic evidences, *latent fingerprints* are often crucial piece of evidence that can link a suspect to a crime [Holder *et al.*, 2011].

1.2. Latent fingerprints

Any impression made by the friction ridge skin of the human finger is generally termed as *fingerprint*. Those fingerprints which are revealed using chemical or optical processing and collected from a crime scene by specialists trained in forensic sciences are called *latent fingerprints* (Figure 1.1(a)). These are unintentionally left fingerprints found in the crime scenes. The latent fingerprints are then photographed, marked up for discriminatory features by forensic fingerprint examiners, and are used to search by Automated Fingerprint Identification Systems (AFIS). In the realm of forensic analysis, the use of latent fingerprints is a routine procedure to identify suspects.

Law enforcement agencies maintain a huge database of the fingerprints of individuals who are arrested or imprisoned. The forensic fingerprint database are typically collected by obtaining a rolled fingerprints from each finger. Such fingerprints in the database are called *tenprints* or *exemplars* fingerprints (Figure 1.1(b)). When a latent fingerprint is found, the criminal investigators first search for the suspect in an AFIS database to establish the identity of the individual to link with a particular criminal record. If there is a match, then the individual is linked to the crime under investigation.



(a) Latent fingerprint

(b) Tenprint card

Figure 1.1: *An example latent and tenprint fingerprints used in forensic analysis.*

1.3. Forensic testimony standards

Testimonies provided by forensic examiners based on the analysis of evidence is routinely collected and presented in court. The testimonies based on fingerprints carried substantial credibility and weight compared among various sources of evidences. A reasonably high degree of match between the latent and the exemplar fingerprint features leads the forensic examiner to testify irrefutably with high confidence about the decision. The testimonies provided by the latent examiners were never questioned for decades. Central to the idea of establishing such fingerprint based identity is the assumption that each fingerprint is unique. This assumption allows the forensic examiner to provide unquestionable conclusions even though this assumption lacks sound theoretical and empirical foundations [Maltoni *et al.*, 2009]. Also, such assumption precludes opportunities to establish any scientific error rates.

Some high profile cases challenged the scientific methodology that is followed to arrive at conclusions by forensic examiners. These led to establish a standard for forensic expert testimonies to be produced in courts. *Frye standard* and *Daubert standard* are two such standards followed by courts in United States towards accepting the expert testimonies.

1.3.1. Frye standard

Frye standard is a test to determine the admissibility of scientific evidence in court. It is also known as "Frye test" or "general acceptance test". Frye standard states that [Frye, 1923]:

> *The evidence could be admitted in court only if the thing from which the deduction is made is sufficiently established to have gained general acceptance in the particular field in which it belongs.*

When a scientific evidence is widely disputed, then the application of Frye test is by providing a number of experts to speak for the validity of the science behind the issue in question. The court then examines the scientific papers, books and judicial precedents on the subject under issue to make determinations as to the reliability and general acceptance of the evidence.

Frye standard originates from a 1923 case of Frye v. United States (293F. 1013, D.C. Cir 1923) discussing about the admissibility of polygraph test as evidence. The Frye standard was eventually superseded by Daubert standard.

1.3.2. Daubert standard

The 1993 case of Daubert v. Marrell Dow Pharmaceuticals (113 S. Ct. 2786) again made a significant break on the scientific acceptance test followed by Frye standard. Daubert standard states that [Maltoni *et al.*, 2009]:

> *For expert forensic testimony to be accepted, the particular tool or methodology in question should be subjected to three main criteria of scientific validation:*
>
> 1. *has been tested,*
> 2. *has been subjected to peer-review, and*
> 3. *possess known error rates.*

In forensic fingerprint examination methodology, a systematic way to establish error rates could be as follows:

a) Collecting population sample.

b) Analyze inherent discriminant feature variability.

c) Report probability of two different people sharing common features.

The Daubert v. Marrell Dow Pharmaceuticals was a case where the parents of Jason Daubert and Eric Schuller sued the pharmaceutical company claiming that the drug Bendectin manufactured by the company caused birth defects to their children. The complainants evidence were based on methodology which did not gained the general acceptance of scientific community. This led the court to establish new standards which superseded Frye standard.

Following the Daubert case, the fingerprint individualization was first challenged in the 1999 case of United States v. Byron Mitchell case [ByronCase, 1999]. The challenge was based on the premise that the uniqueness of fingerprint has not been objectively tested and the matching errors are unknown. A list of 22 known exposed cases of erroneous fingerprint identification made by forensic experts are reported in [Cole, 985-1078].

1.4. Methodology of ACE-V

The most common type of forensic evidence used in criminal investigations is latent fingerprint. Despite latent fingerprints being a crucial evidence in individualization, latent fingerprint comparison is not an easy task. This is mainly attributed to the poor quality of the latent fingerprints taken from the crime scenes. In order to improve the matching efficiency, the concept of “Lights-Out System” was introduced for latent matching [Dvornychenko and Garrism. M, 2006]. A Lights-Out System is a fully automatic (no human intervention) identification process. Here, the system should automatically extract the features from latent fingerprint and match it against the exemplars stored in the AFIS database to obtain a shortlist of possible suspects based on some ranking strategy.

In general, the latent fingerprints obtained from crime scenes are partial in nature, and are mostly distorted, smudgy, blurred etc. These factors will lead to high number of unreliable extracted fingerprint features which degrades the AFIS performance. AFIS do not use all kind of discriminatory features that could be derived from a fingerprint, mainly due to the limitations of automatic and reliable extraction of all types of features. The performance of feature extraction and matching algorithms of AFIS in forensic scenario is of great importance to avoid erroneous individualization, as well as in saving the time. To evaluate the commercial AFIS performance in Lights-Out mode, NIST has conducted multi-phase open project called Evaluation of Latent Fingerprint Technologies (ELFT) [NIST-ELFT, 2013]. The best performing latent matching system achieved Rank-1 identification of only 63.4%.

In [Indovina *et al.*, 2011a], it is concluded that only a limited class of latent fingerprints (possibly of good quality latents) benefits from automated procedures, but the procedures of marking the minutiae, determining the subjective quality of latents, etc still need to be carried out manually. Due to these limitations, “Semi Lights-Out Systems” are only feasible currently in the forensic scenario. In Semi Lights-Out System, some human intervention is necessary during the feature extraction from latent fingerprints. This will involve marking the region of interest, extracting the minutiae features, aligning the latent fingerprint etc. Once the latent fingerprint features are extracted and encoded manually, it is compared automatically against exemplars stored in database. The AFIS provides a list of 10 or 20 candidates with the highest matching scores. The fingerprint examiner will then analyze these high scoring exemplar fingerprints manually following the friction ridge examination methodology known as Analysis, Comparison, Evaluation and Verification (ACE-V).

Friction ridge examination

The methodology that a forensic examiner should follow for a particular type of forensic evidence has been well documented by *Scientific Working Groups* (SWG) which consists of experts developing the standard. For forensic fingerprints, until 2014, the group was known as *Scientific Working Group on Friction Ridge Analysis, Study and Technology* (SWGFAST), and they prepared the document titled *Standards for Examining Friction Ridge Impressions and Resulting Conclusions* [SWGFAST.v01, 2011]. From 2015, SWGFAST is replaced with the Organization of Scientific Area Committees (OSAC) Friction Ridge Subcommittee (FRS) (OSAC-FRS) [OSAC, a] [OSAC, b].

Bifurcation	Dot	Ridge Ending
		

Figure 1.2: *Level 2 details (minutiae) of the fingerprints [NIST-EWG, 2012].*

The SWGFAST standard specifies the following rules to be considered during friction ridge examination [SWGFAST.v01, 2011] [Srihari, 2013]: 1) the fundamental principles by which examinations are conducted, 2) features to be used for friction ridge examination and 3) specific steps for friction ridge examination.

- *Fundamental principles* considered for friction ridge examinations by a forensic examiner is as follows:

  1. The morphology of friction ridge skin is unique.
  2. The friction ridge pattern is permanent.
  3. The impression of friction ridge pattern is transferred during contact with surface.
  4. An impression that contains sufficient quality and quantity of friction ridge details can be uniquely identified or excluded.
  5. Sufficiency of the friction ridge details are determined by the forensic examiner.

- *Fingerprint features* : There are three levels of details and "other" features described to be used for friction ridge examination.

  1. *Level One Details* of fingerprint constitutes the general overall direction of ridge flow. They can be used for pattern interpretation and to determine the anatomical source (i.e, finger, palm, foot, toe) and orientation. These features are not unique to each fingerprints. So they cannot be used for individualization but can be used for fingerprint type classification and indexing.
  2. *Level Two Details* describes the path of specific ridge. The actual ridge path includes the starting position of the ridge, the path the ridge takes, the length of the ridge path and where the ridge path stops. They principally define the typical minutiae points such as ridge ending, bifurcations and dots (Figure 1.2) of the fingerprint. Level two features are generally believed to be discriminative, stable and robust.
  3. *Level Three Details* are the shapes of the ridge structures. This level of detail encompasses the morphology (edges, textures, and relative pore positions) of the ridge. They can also include secondary creases, ridge breaks, etc.

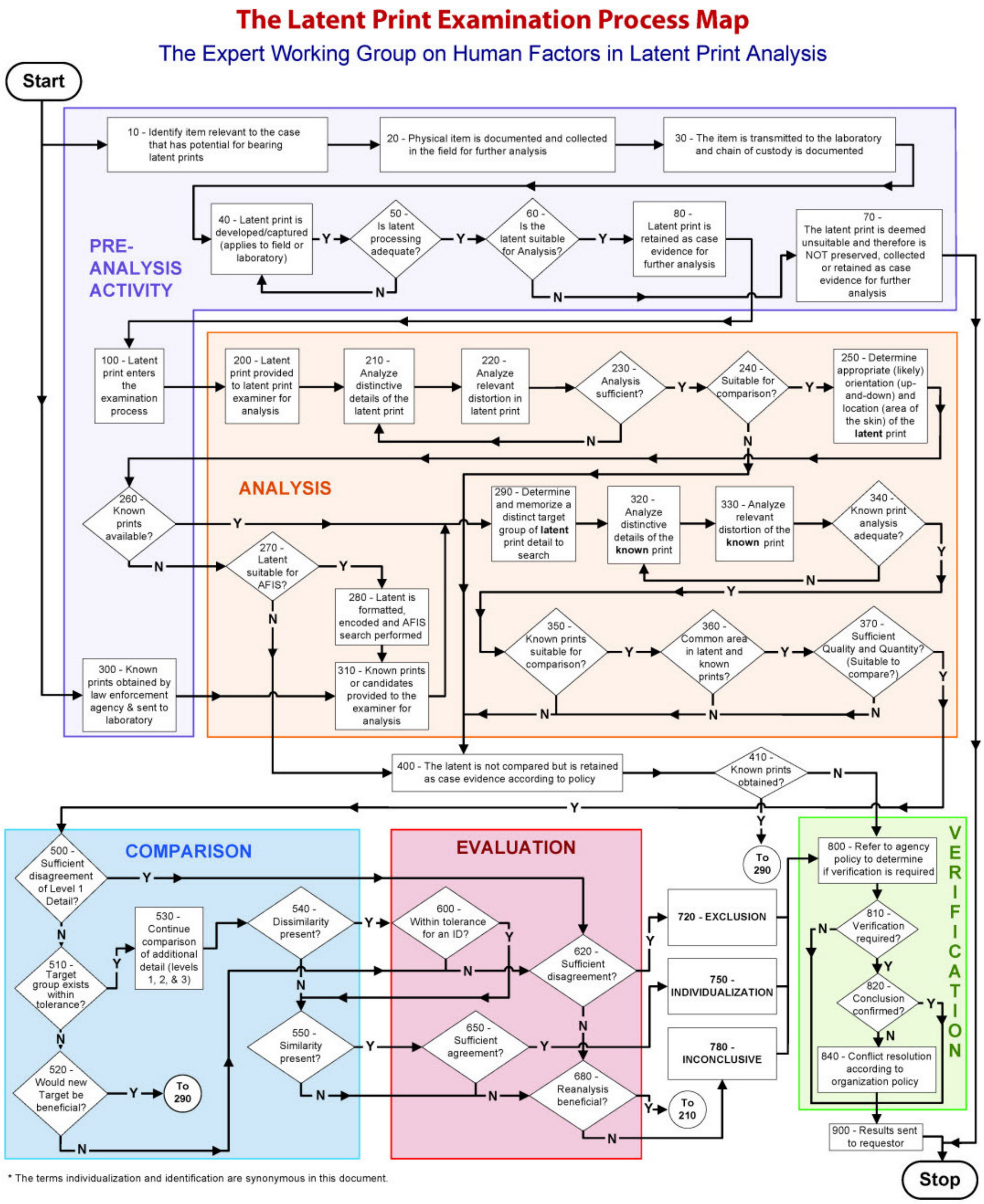


Figure 1.3: *Flow chart of various steps involved in Analysis, Comparison, Evaluation and Verification (ACE-V) [NIST-EWG, 2012].*

4. *Other features* describe the features like creases, scars, warts, cuts, blisters and other imperfections. They may be permanent or temporary. Additionally they may exist as level one, two or three details.

- *ACE-V* methodology: The steps involved in examining a latent fingerprint is described by the ACE-V (Analysis, Comparison, Evaluation and Verification) process. A flowchart comprising various steps in ACE-V is shown in Figure 1.3.

  ACE-V comprises of the following four phases:

  1. *Analysis:* The forensic examiner looks for sufficiency of the details present in the given latent fingerprint. This comprises of checking for ridge clarity, quantity of Level One, Level Two and Level Three details.

  2. *Comparison:* Once the latent fingerprint passes the analysis phase, many useful friction ridge details are extracted manually by the forensic examiner and are compared against one or more exemplar/reference fingerprints shortlisted by an AFIS to determine whether they are in agreement.

  3. *Evaluation:* Based on the conclusions derived from the analysis and comparison phases, the forensic examiner yields a decision as *individualization (identification or match), exclusion (non-match) or inconclusive* for the given latent and impression fingerprint image pair.

  4. *Verification:* In this phase, another qualified forensic forensic examiner reexamines the decision made by the previous examiner by following the above three phases once again.

1.5. Challenges in individualization of forensic fingerprints

The major challenges that are encountered in the process of individualization of latent fingerprints are summarized in this section.

1.5.1. Characteristics of fingerprints

The use of latent fingerprint in forensic analysis is based on two fundamental premises:

1. *Persistent*: the fingerprint pattern retains its ridge pattern over time.

2. *Distinctiveness*: the fingerprint pattern of an individual is unique.

Of these two premises, the characteristic of fingerprints being persistent is widely accepted, but the second premise for the fingerprint being unique is often challenged [Pankanti *et al.*, 2002] [Srihari and Srinivasan, 2007].

Persistence of fingerprints

The friction ridge patterns (ridge flow and minutiae) are often described to be permanent. However, the cellular surface of the friction ridge skin is not permanent but they are persistent over time. The surface cells are replaced on a regular basis. The regeneration of the skin naturally makes the effort to reproduce in a manner to maintain the previous form. There will be microscopic variations but the overall form and features will be remarkably persistent over time so that effective comparisons can be made.

Distinctiveness of fingerprints

The distinctiveness characteristic of fingerprint is often challenged. The uniqueness of fingerprint is not an established fact but is only an empirical observation. Another assumption related to this uniqueness is the fact that pattern formations in nature are never repeated in their morphological structures (or as the saying goes, "nature never repeats itself"). There are also some biological explanations for the friction ridge pattern to be unique [Holder *et al.*, 2011].

Basic fingerprint minutiae are used in traditional fingerprint individualization, statistical modeling and in AFIS. All of these do allow some variations within a threshold to accept the uniqueness. AFIS do not use all discriminatory features that are possible in fingerprints, but uses only a limited types of features automatically extracted by a feature extraction algorithm. In spite of these limitations, no model and application has provided evidence that fingerprints are not unique. The cases of erroneous identifications has arisen due to the partial nature of the fingerprint as well as poor quality of latent fingerprints.

1.5.2. Effect of human factors

The examination of latent fingerprint consists of a series of steps (ACE-V) involving the comparison of latent against exemplars. During this process, the latent examiner must reach correct conclusion. The perception and the decision making ability among forensic examiners vary, e.g, the decision made by a novice examiner is not always consistent with the decision made by an experienced examiner for the same casework [Vanderkolk, 2011]. It is not guaranteed that same conclusions can be reproduced by two different examiners.

There is no scientific framework in use at the criminal justice system to characterize the uncertainty involved in the ACE-V procedure, as well as to express the strength of opinion of the forensic examiner quantitatively [Srihari, 2013]. Such a requirement has been articulated in several influential reports [Srihari, 2013] like the NRC 2009 report [NAS-NRC, 2009] and the NIST Human Factors report [NIST-EWG, 2012]. The new paradigm coming forward in this regard [Saks and Koehler, 2005] avoids hard identification decisions by considering evidence reporting methods that incorporate uncertainty and statistics.

1.5.3. Partial latent fingerprints

The state of the art fingerprint matching algorithms usually assume approximately the same size of the minutiae set between the query and the reference minutiae template for good identification accuracy [Jea and Govindaraju, 2005]. The latent fingerprints that are obtained from crime scenes tend to be mostly partial in nature.

The performance of the existing partial fingerprint identification systems mainly depends on the image quality, the number of minutiae available and other derived and extended features that can be obtained from the partial fingerprint region.

Various approaches in state of the art partial fingerprint identification [Wang and Hu, 2011] include:

- the use of localized secondary features derived from relative minutia information [Jea and Govindaraju, 2005].
- using representative points along ridge lines in addition to minutiae [Fang *et al.*, 2007].
- use of Level Three features such as dots and incipients [Jain *et al.*, 2007a].

Comparing partial fingerprint against a full fingerprint with only limited number of discriminatory features is a challenging problem.

1.6. Emerging and improving technology

Latent fingerprint examiners use AFIS, online database, digital enhancement software and other types of technologies to assist with the Analysis, Comparison, Evaluation and Verification process [NIST-EWG, 2012]. Combining these tools with the examiner's own experience and expertise make investigations more reliable and easier to explain to juries. Of these emerging technologies, AFIS has got more importance as it helps in shortlisting the candidate or possible suspects.

The Report of the Expert Working Group on Human Factors in Latent Print Analysis [NIST-EWG, 2012] has made the recommendation for some of the challenges for AFIS as follows:

> *Recommendation 4.1: The federal government should support research programs to improve automated fingerprint identification systems. Such programs could address the following issues:*
>
> 1. *Expanding the algorithms used to match prints to account for the fact that the diagnostic value of minutiae depends on the region in which they are located;*
> 2. *Making fingerprint and palm print databases interoperable among local, state, and federal automated identification systems; and*
> 3. *Increasing compatibility between automated identification systems and other latent print software tools, including digital enhancement programs, probability calculation programs, and automated quality assessment programs.*

1.7. Pre-registration for latent fingerprints

The alignment between the input and the reference fingerprint is a crucial step. This is because the fingerprint images captured in different instances might have different rotation, translation or non-linear deformation between them. The main objective of fingerprint alignment is to estimate the transformation parameters between input and reference fingerprints. One such methodology is based on the generalized Hough transform [Ratha *et al.*, 1996]. The main disadvantage for such technique is the inaccuracy in the transformation estimation due to discretization of the parameters space. A sufficiently small discretization step will lead to exponential runtime thereby incresing the computational cost. Other approaches could be to use brute force to check for all possible correspondences between minutiae pairs. There exists some alignment techniques that augment minutiae with other supplementary features such as ridge information, orientation fields around a small neighborhood of minutiae, geometric relationships between minutiae and its neighbors, etc.

Alignment of full fingerprints is a well studied problem. But these methods are limited in alignment accuracy due to quantization of transformation parameters, or are not adapted for the partial fingerprint scenario. Most fingerprint matching algorithms in general assume approximately the same size of the minutiae set between the query and the reference minutiae for good identification accuracy [Jea and Govindaraju, 2005]. Trying to align a partial fingerprint to a full fingerprint only based on minutiae features could lead to errors.

In the first part of Dissertation, we focus on the problem of aligning a partial fingerprint against a full fingerprint, especially of poor quality latents. Instead of minutiae, we used Orientation Fields (OF) to perform the alignment. We reduce fingerprint images to orientation images, and we look at the alignment problem as registering the partial fingerprint orientation image into the full fingerprint orientation image. The OF representing the flow of ridges is a relatively stable global feature of fingerprint images, and it represents the intrinsic nature of the fingerprint. The representative OF of a fingerprint is very less affected by the type of capture device, contrast variations, and other quality effects compared to the input image or the minutiae.

1.8. Extended Feature Sets

Even though the friction ridge analysis methodology follows a pre-defined set of rules (as detailed in Section 1.4), there are some problems reported by the forensic community in these practices. Two major problems in friction ridge analysis as reported in [Hicklin, 2007] are as follows:

1. Latent AFIS searches are limited by over simplified feature set.
2. During the latent examiner comparison, there are no standard format to document the features used in comparison decision. This leads to difficulty with future reference or interchange with other forensic examiners.

AFIS uses only a limited types of features automatically extracted from the fingerprints using a feature extraction algorithm. On the other hand, forensic examiners use rich set of features during manual comparison as compared against AFIS comparisons. This could be a possible reason why manual comparisons outperforms AFIS comparisons [Jain, 2010]. Any features that are not currently used by commercial AFIS are generally termed as Extended Feature Set (EFS) [Zhao and Jain, 2010]. The use of EFS by forensic examiners in manual comparison decision is much debated, mainly due to non-repeatability by another examiner to validate the previous decision.

SWGFAST (Scientific Working Group on Friction Ridge Analysis, Study, and Technology) drafted a memo to NIST noting that forensic examiners use features that are not currently addressed in fingerprint standards. The ANSI/NIST Standard Workshop II charted the Committee to Define an Extended Fingerprint Feature Set (CDEFFS). The CDEFFS included 45 members from various federal agencies, latent community, AFIS vendors, and academia [Hicklin, 2007]. The purpose of CDEFFS is to define a standard to completely represent the distinctive information in the fingerprint which are quantifiable, repeatable and develop a clear method of characterizing information for: 1) forensic examiner initiated AFIS searches, and 2) forensic examiner markup and exchange of latent fingerprints.

Many commercial minutiae-based matchers mostly use only two prominent ridge characteristics namely *ridge-endings* and *bifurcations*. To use EFS in automated systems, reliable feature extraction algorithms are mandatory. Many law enforcement agencies follow a 500 ppi scanning resolution for fingerprint images to be used in AFIS. With such a resolution, it is difficult to reliably extract many of the extended features automatically. Due to advances in fingerprint scanning technologies, SWGFAST during ANSI/NIST Fingerprint Standard Update Workshop in 2005 proposed 1000 ppi as minimum scanning resolution for fingerprint images.

In the second part of Dissertation, we focus on the problem of using EFS in a typical minutiae-based matcher. We propose a method to improve the identification accuracy of minutiae-based matchers for partial latent fingerprints by incorporating reliably extracted rare minutiae features. Our proposed algorithm will modify the similarity scores of minutiae-based matchers based on the presence of rare minutia features like *fragments*, *enclosures*, *dots*, *interruptions, etc.* The weights that we multiply to modify the similarity scores are obtained by both the derived entropy and probability of occurrence of rare minutia features. The decision for a match or non-match is automatically estimated based on least squares fitting of an affine transformationbetween the latent minutiae set and tenprint minutiae set.

1.9. Evidence evaluation from AFIS similarity scores

In the area of forensic biometrics, interpretation of forensic evidences, and evidence evaluation from the similarity scores generated by a biometric system is a topic of importance. The uniqueness of fingerprint is not an established fact but is only an empirical observation. There is a widespread concern about the scientific basis underlying the individuality of fingerprints, especially when using in the court of law. Evidence evaluation using a Bayesian probabilistic framework has been proposed in recent years as a logical and appropriate way to report evidence to a court of law [Aitken and Taroni, 2004]. In Bayesian approach, a likelihood ratio is computed to represent the value of the evidence, and to be reported to a court of law. This framework clearly complies with the requirements of modern forensic science [Saks and Koehler, 2005]: it is scientifically sound (transparent procedures, testability, formally correct), and clearly separates the competences of the forensic examiner and the court.

The establishment of this Bayesian evaluative framework has motivated the convergence of pattern recognition and machine learning approaches to yield probabilistic outputs in the form of likelihood ratios. A common architecture for this considers two steps: first, the computation of a discriminating score between two evidential materials (e.g., a fingermark in the crime scene and a fingerprint from a known suspect), which can be performed by standard minutiae-based fingerprint matcher; and second, the transformation of the similarity score into a likelihood ratio. This process of transforming scores relating two pieces of evidence into likelihood ratios has been dubbed calibration [Brümmer and du Preez, 2006; Ramos and Gonzalez-Rodriguez, 2013; vanLeeuwen and Brümmer, 2007].

In this Dissertation, we focus on this problem of evidence evaluation, address its importance, and propose a solution to evaluate the forensic evidence in the light of the method proposed to improve the identification accuracies of minutiae-based matcher by incorporating extended fingerprint features. Moreover, we will show how the incorporation of rare minutiae improve the performance of the system, also at the level of forensic interpretation.

1.10. Latent fingerprint evaluation

Fingerprint matching is the process by which the discriminatory features of two fingerprints are compared to determine whether they came from same finger or from different fingers. The extraction of discriminatory features of a given fingerprint can be done manually or using automated algorithms. Similarly, matching can also be manual, automatic or semi-automatic. "Lights-Out" and "Semi Lights-Out" systems are used to generate a shortlist of suspects from a criminal database stored in AFIS. In friction ridge analysis, both feature extraction and matching are manually performed by forensic examiners for individualization. In the present section, we summarize key results from international evaluations on latent fingerprint recognition in both "Lights-Out" and 'Semi Lights-Out" mode.

1.10.1. Performance evaluation : CMC

In latent fingerprint identification, it is common practice to evaluate the performance of different latent AFIS with respect to its rank identification accuracies. The Cumulative Match Characteristics (CMC) curve plots the probability of identification against the returned $1 : N$ candidate list size. *Rank-k* identification accuracy shows the probability that a given user appears in a k-sized candidate list. In the CMC plot, the horizontal axis represents the rank k, and the vertical axis represents the identification probability in a k-sized rank list [Bolle *et al.*, 2005; Moon and Phillips, 2001].

1.10.2. Lights-Out mode

Current practice in latent AFIS technology involves marking the latent fingerprint features manually by forensic examiner, then use the latent fingerprint image and the manually marked features to search in the AFIS for possible list of suspects. To avoid this burden of manual marking and with the hope to automate the latent AFIS in Lights-Out mode, NIST conducted a public evaluation of commercial AFIS performance in Lights-Out mode. This was a multi-phase open project called Evaluation of Latent Fingerprint Technologies (ELFT) [NIST-ELFT, 2013]. The results of various phases of ELFT are summarized in Table 1.1. The reported accuracies from Phase-I and Phase-II of ELFT cannot be directly compared as the database and the quality of the latents were different. In [Indovina *et al.*, 2011a], it is concluded that only a limited class of latents benefits from automated procedures, and still manual intervention

Phase of ELFT	Database size	Best Rank-1 accuracy
Phase-I [NIST-ELFT-1, 2007]	100 latents compared against 10,000 rolled prints	80.0%
Phase-II, Evaluation-1 [Indovina *et al.*, 2009]	835 latents compared against 100,000 rolled prints	97.2%
Phase-II, Evaluation-2 [Indovina *et al.*, 2012a]	1,114 latents compared against 100,000 rolled prints	63.4%

Table 1.1: *Summary of NIST Evaluation of Latent Fingerprint Technologies (ELFT) results.*

ELFT-EFS	Database size	Best Rank-1 accuracy
Evaluation-1 [Indovina *et al.*, 2011b]	1,114 latents compared against 1,000,000 rolled prints and 1,000,000 plain prints	66.7%
Evaluation-2 [Indovina *et al.*, 2012b] [Indovina *et al.*, 2012c]	1,066 latents compared against 1,000,000 rolled prints and 1,000,000 plain prints	71.4%

Table 1.2: *Rank-1 identification accuracy of NIST Evaluation of Latent Fingerprint Technologies - Extended Feature Sets (ELFT-EFS).*

is necessary. The procedures of marking the minutiae, determining the subjective quality of latents, etc still need to be carried out manually.

1.10.3. Semi Lights-Out mode

To determine the effectiveness of forensic examiner marked latent fingerprint features on the latent identification accuracy, another public evaluation known as ELFT-EFS was launched. ELFT-EFS was conducted in a “Semi Lights-Out” mode as compared to “Lights-Out” mode for ELFT. NIST conducted two evaluations for ELFT-EFS and the best achieved Rank-1 identification accuracy for each of the evaluations is summarized in Table 1.2. As in ELFT, the results of different evaluations in ELFT-EFS cannot be directly compared because the database used were not exactly the same [Indovina *et al.*, 2012b] [Indovina *et al.*, 2011b]. In [Indovina *et al.*, 2012b], it is reported that though the highest measured accuracy achieved by any individual matcher at Rank-1 was 71.4%, approximately 82% of the latents were matched at Rank-1 by one or more matchers combined. This concludes the potential for additional accuracy improvement through improved algorithms.

1.11. Motivation of the Thesis

The observations and results described before as well as the recommendations and new standards set forth by the forensic community have motivated the research carried out in this Thesis. The research mainly focuses on improving the identification accuracies of existing minutiae-based matchers, and also in developing a robust evidence evaluation based on the likelihood ratio model which incorporates extended fingerprint feature sets. In particular, three major research lines are carried out in this Thesis:

1. *Pre-registration*: The latent fingerprints obtained from the crime scenes are usually partial in nature. Most of the available minutiae-based matchers are well adapted for full-to-full fingerprint comparisons [Jea and Govindaraju, 2005]. Existing partial fingerprint matchers either rely on derived secondary minutiae features such as relative minutiae information, ridge skeleton information or other extended features [Fang *et al.*, 2007; Jain *et al.*, 2007a; Jea and Govindaraju, 2005; Wang and Hu, 2011]. When a partial query fingerprint minutiae pattern needs to be compared against a full reference fingerprint minutiae pattern, it will be advantageous to know the location of the partial minutiae pattern in the full minutiae pattern. This will help to reduce the minutiae search space of full minutiae pattern with respect to the size of partial minutiae pattern, thereby reducing the matching scenario of partial-to-full comparison into full-to-full comparison where both the minutiae patterns are almost of the same size.

2. *Extended Feature Sets*: The public evaluation of latent fingerprint matching technologies in both fully automatic (ELFT) [Indovina *et al.*, 2009, 2012a, 2011a; NIST-ELFT, 2013; NIST-ELFT-1, 2007] and semi-automatic mode (ELFT-EFS) [Indovina *et al.*, 2012b,c, 2011b] conducted by NIST concluded that human intervention is inevitable in case of latent fingerprint matching, and also the use of extended fingerprint feature sets reliably extracted manually have contributed towards improving the latent fingerprint matching performances. More research into the use of EFS towards improving the identification accuracy is needed [Jain, 2010].

3. *Evidence evaluation*: The uniqueness of a fingerprint is not an established fact but only an empirical observation. There is a widespread concern about the scientific basis underlying the individuality of fingerprints, especially when using them in the court of law. Many individualization models for fingerprints have been proposed in the research literature. However, there is no scientific framework in use at the criminal justice system to characterize the uncertainty involved in the friction ridge analysis methodology, as well as to express the strength of opinion of the forensic examiner quantitatively [NIST-EWG, 2012;

Srihari, 2013]. Such a requirement has been articulated in several influential reports like the National Research Council 2009 report [NAS-NRC, 2009]. The new paradigm coming forward in this regard [Saks and Koehler, 2005] avoids hard identification decisions by considering evidence reporting methods that incorporate uncertainty and statistics. Among all the methods for evidence evaluation, the likelihood ratio is receiving greater attention [Aitken *et al.*, 2011; Srihari, 2013].

1.12. The Thesis and main contributions

The Thesis developed in this Dissertation can be stated as follows:

When comparing a partial fingerprint minutiae pattern against a full fingerprint minutiae pattern, it is advantageous to know in advance where the partial pattern is located in the full pattern so as to reduce the matching error, thereby improving the identification accuracy. Additionally, together with the typical automatically extracted minutiae, the use of reliably extracted Extended Feature Sets for fingerprints also help in improving both the identification accuracy and the computation of Likelihood Ratios for statistical evidence evaluation.

The technical contributions of this work are:

- *Pre-registration using orientation field*: We proposed a new correlation-based hierarchical registration method for orientation images to register a partial fingerprint in a full fingerprint. To register a partial fingerprint against a full fingerprint based on minutiae alone is a hard problem. Most minutiae-based alignment techniques rely on reference points such as core or delta singular points. In partial latent fingerprints, presence of these singular points are not always guaranteed. So, for reliable alignment, we made use of orientation fields of the fingerprint. The orientation field representing the flow of ridges is a relatively stable global feature of fingerprint images, and it represents the intrinsic nature of the fingerprint. The representative orientation field of a fingerprint is very less affected by the type of capture device, contrast variations, and other quality effects compared to the input image or the minutiae.

- *Best representative orientation field*: We concluded the best representative orientation fields for latent fingerprints and tenprint fingerprints. Several methods exist to estimate the orientation fields of a given fingerprint. Depending on the type of fingerprint, i.e, latent or tenprint, different orientation estimation strategies need to be used for better registration accuracy.

- *Affine transform based fitting error*: We developed a method to make use of reliably extracted rare minutiae features to modify the similarity scores of minutiae-based fingerprint matchers which significantly improves the rank identification accuracies. Based on the least square fitting error to transform the partial fingerprint minutiae pattern and the full fingerprint minutiae pattern with the rare minutiae as reference point, the similarity scores of matchers are modified. This leads to significant improvement in the rank identification accuracies.

- *Likelihood ratio model*: We proposed a robust evidence evaluation from AFIS similarity scores based on likelihood ratio. We showed how the incorporation of rare minutiae features improve the performance of the minutiae-based matcher and thereby the forensic interpretations.

1.13. Outline of the Dissertation

The main objectives of this PhD Thesis are as follows: 1) reviewing and studying the problem of partial fingerprint pre-registration, use of extended feature sets in latent fingerprint matching, and likelihood ratio based fingerprint evidence evaluation; 2) developing algorithms to improve the rank identification accuracies of minutiae-based matchers based on pre-registration and incorporating rare minutiae features; 3) experimental demonstration of the developed algorithms to real casework forensic fingerprint databases.

This Dissertation is structured according to a *traditional complex* type wherein each of the major research problems are presented in separate parts consisting of introduction, related works, algorithm and experiments in which the developed methods are applied [Paltridge, 2002]. The chapter structure is as follows:

- Chapter 1 introduces the topic of latent fingerprint recognition in forensic scenario, and the major challenges faced by the state-of-the-art methodologies, and gives the motivation, outline and contributions of this PhD Thesis.

- Chapter 2 introduces about the pre-registration of latent fingerprints against a tenprint fingerprint based on orientation fields, and reviews related works on fingerprint registration followed by a brief description about the forensic fingerprint database used in experiments for pre-registration.

- Chapter 3 describes the algorithm developed for registering the partial fingerprint in a full fingerprint, followed by experimental demonstration of the performance improvement of minutiae-based matchers when incorporating our proposed algorithm.

- Chapter 4 introduces about the use of Extended Feature Sets in fingerprints to improve the identification accuracy of latent AFIS, reviews related works in the use of extended features, followed by a description of the forensic fingerprint database obtained from Guardia Civil which contains rare minutiae features.

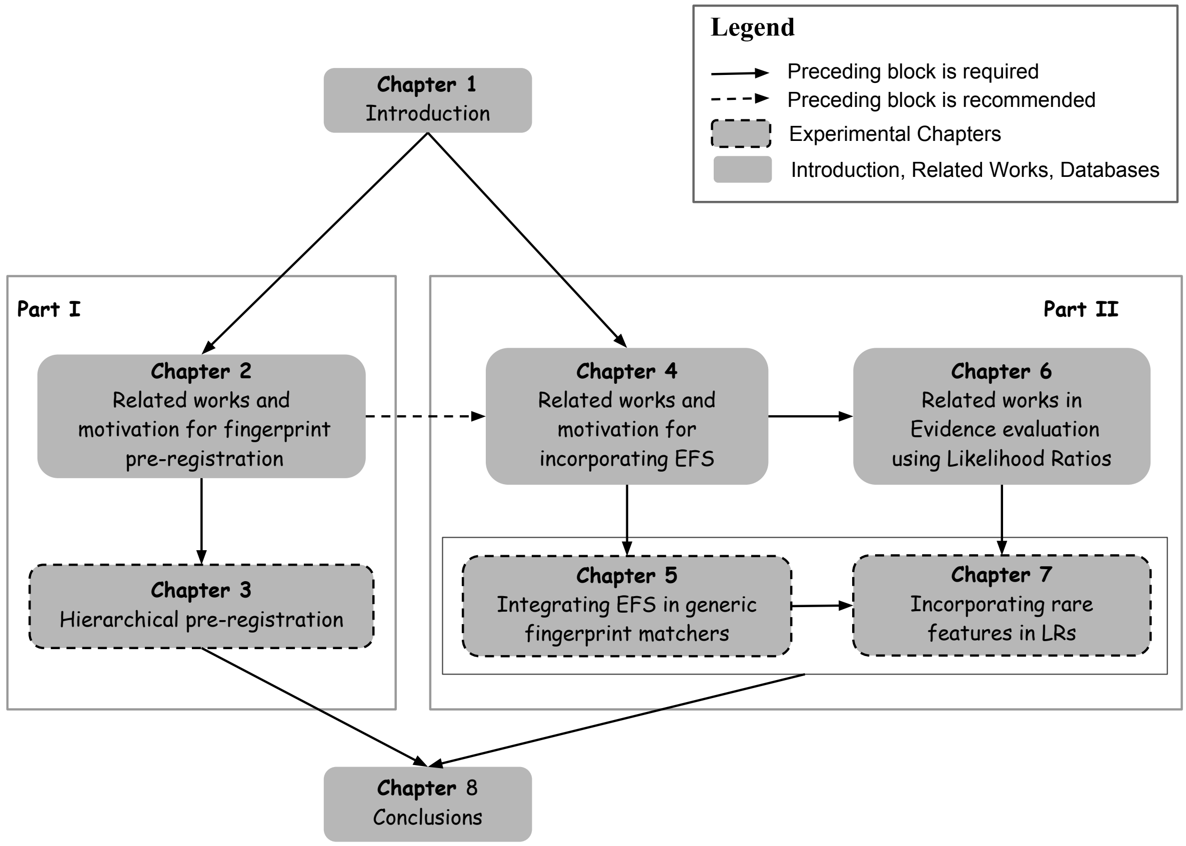


Figure 1.4: *Dependence among different chapters.*

- Chapter 5 describes the algorithm developed based on affine transformation to improve the identification accuracy of minutiae-based matchers by incorporating reliably extracted rare minutiae features. This is supported with experimental demonstration of the improvements in rank identification accuracies of minutiae-based matchers.

- Chapter 6 addresses the issue of the interpretation of forensic evidence from scores computed by a biometric system, and addressing it using the proposed system developed to incorporate rare minutiae features to improve the rank identification accuracies of minutiae-based matchers. Various likelihood ratio computation methods are discussed.

- Chapter 7 describes the proposed solution to evaluate the forensic evidence using a likelihood ratio framework. We used score normalization to rectify the misalignment of the similarity scores computed by the biometric system, abd then experimentally demonstrate the best LR computation method for the proposed rare minutiae-based similarity proposed in Chapter 5.

- Chapter 8 concludes the Dissertation summarizing the main results obtained and outlining future research lines.

The dependence between the chapters is illustrated in Figure 1.4. Part I and Part II of the Dissertation can be read independently as they are almost self contained. The experimental chapters should always be preceded by its introduction chapter.

1.14. Detailed research contributions

A list of the research contributions of this PhD Thesis is provided in this Section. Some publications appear in several items of the list.

1. Literature Review

Pre-registration using orientation field.

- R. P. Krish, J. Fierrez, D. Ramos, J. Ortega-Garcia and J. Bigun, "Partial Fingerprint Registration for Forensics using Minutiae-Generated Orientation Fields", in 2nd International Workshop on Biometrics and Forensics, Valletta, Malta, March 2014 [Krish *et al.*, 2014a].

  (Selected as one of the best papers in IWBF-2014)
- R. P. Krish, J. Fierrez, D. Ramos, J. Ortega-Garcia and J. Bigun, "Pre-Registration for Improved Latent Fingerprint Identification", in Proc. IAPR/IEEE 22nd Int. Conf. on Pattern Recognition, ICPR, pp. 696-701, Stockholm, SWEDEN, August 2014 [Krish *et al.*, 2014b].

Rare-minutiae features.

- R. P. Krish, J. Fierrez, D. Ramos and R. Wang, "On the importance of rare features in AFIS-ranked latent fingerprint matched templates", in Proc. 47th IEEE International Carnahan Conference on Security Technology (ICCST), Medellin, Colombia, October 2013 [Krish *et al.*, 2013c].

Affine transform based fitting error, and likelihood ratio framework for evidence evaluation.

- R. P. Krish, J. Fierrez, D. Ramos, J. Ortega-Garcia, "Improving Automated Latent Fingerprint Identification using Extended Feature Sets", Forensic Science International, 2015, (Submitted and under review) [Krish *et al.*, 2015].
- D. Ramos, J. Fierrez, R. P. Krish, F. J. Gomez-Herrero, "Evidence Evaluation using AFIS scores: Integrating Rare Features", IET Biometrics, 2015 (Under Preparation) [Ramos *et al.*, 2015]

2. Theoretical Framework

Theoretical framework to register the partial fingerprint orientation field and full fingerprint orientation field.

- R. P. Krish, J. Fierrez, D. Ramos, J. Ortega-Garcia and J. Bigun, "Pre-Registration of Latent Fingerprints based on Orientation Field", IET Biometrics, pp. 1-11, January 2015 [Krish. *et al.*, 2015].

Theoretical framework for least square fitting error based on affine transformation.

- R. P. Krish, J. Fierrez, D. Ramos, J. Ortega-Garcia, "Improving Automated Latent Fingerprint Identification using Extended Feature Sets", Forensic Science International, 2015, (Submitted and under review) [Krish *et al.*, 2015].
- R. P. Krish, J. Fierrez, D. Ramos and R. Wang, "On the importance of rare features in AFIS-ranked latent fingerprint matched templates", in Proc. 47th IEEE International Carnahan Conference on Security Technology (ICCST), Medellin, Colombia, October 2013 [Krish *et al.*, 2013c].

Likelihood Ratio based evidence evaluation.

- D. Ramos, J. Fierrez, R. P. Krish, F. J. Gomez-Herrero, "Evidence Evaluation using AFIS scores: Integrating Rare Features", IET Biometrics, 2015 (Under Preparation) [Ramos *et al.*, 2015]

3. Novel Methods

Registration with correlation based similarity measure decided based on both phase and magnitude of the correlated orientation images.

- R. P. Krish, J. Fierrez, D. Ramos, J. Ortega-Garcia and J. Bigun, "Partial Fingerprint Registration for Forensics using Minutiae-Generated Orientation Fields", in 2nd International Workshop on Biometrics and Forensics, Valletta, Malta, March 2014 [Krish *et al.*, 2014a].

Hierarchical registration based on correlation, Manhattan and Euclidean based distance, Orientation consistency.

- R. P. Krish, J. Fierrez, D. Ramos, J. Ortega-Garcia and J. Bigun, "Pre-Registration of Latent Fingerprints based on Orientation Field", IET Biometrics, pp. 1-11, January 2015 [Krish. *et al.*, 2015].

4. Novel Database

Real forensic fingerprint casework database obtained from Guardia Civil.

- R. P. Krish, J. Fierrez, D. Ramos, R. Veldhuis and R. Wang, "Evaluation of AFIS-ranked latent fingerprint matched template", in Proc. 6th Pacific-Aim Symposium on Image and Video Technology, Guanajuato, Mexico, Springer LNCS-8333, pp. 230-241, November 2013 [Krish *et al.*, 2013b].

5. New Experimental Studies

- Experimental demonstration of best representative orientation fields for both latent and tenprint Fingerprints.
- Significant improvement in the rank identification accuracies for minutiae-based matchers when incorporating hierarchical pre-registration.

- R. P. Krish, J. Fierrez, D. Ramos, J. Ortega-Garcia and J. Bigun, "Pre-Registration of Latent Fingerprints based on Orientation Field", IET Biometrics, pp. 1-11, January 2015 [Krish. *et al.*, 2015].

- Importance of rare minutiae in matched template.
- Affine transform based least square fitting error as similarity measure for matched templates.

- R. P. Krish, J. Fierrez, D. Ramos and R. Wang, “On the importance of rare features in AFIS-ranked latent fingerprint matched templates”, in Proc. 47th IEEE International Carnahan Conference on Security Technology (ICCST), Medellin, Colombia, October 2013 [Krish *et al.*, 2013c].
- R. P. Krish, J. Fierrez, D. Ramos, R. Veldhuis and R. Wang, “Evaluation of AFIS-ranked latent fingerprint matched template”, in Proc. 6th Pacific-Aim Symposium on Image and Video Technology, Guanajuato, Mexico, Springer LNCS-8333, pp. 230-241, November 2013 [Krish *et al.*, 2013b].

Other contributions not directly related with this Thesis, and not included in the Dissertation includes:

1. Palmprint Recognition

- R. Wang, D. Ramos, J. Fierrez and R. P. Krish, ”Automatic Region Segmentation for High-Resolution Palmprint Recognition: Towards Forensic Scenarios”, in Proc. 47th IEEE International Carnahan Conference on Security Technology (ICCST), Medellin, Colombia, October 2013 [Wang *et al.*, 2013a].
- R. Wang, D. Ramos, J. Fierrez and R. P. Krish, ”Towards Regional Fusion for High-Resolution Palmprint Recognition”, in Proc. XXVI SIBGRAPI conference on Graphics, Patterns and Images, Arequipa, Peru, August 2013 [Wang *et al.*, 2013b].

2. Online Signature Verification

- R. P. Krish, J. Fierrez, J. Galbally and M. Martinez-Diaz, ”Dynamic Signature Verification on Smart Phones”, in Proc. Workshop on User-Centric Technologies and Applications, PAAMS, Salamanca, SPAIN, May 2013 [Krish *et al.*, 2013a].
- M. Martinez-Diaz, J. Fierrez, R. P. Krish and J. Galbally, ”Mobile Signature Verification: Feature Robustness and Performance Comparison”, IET Biometrics, Vol. 3, n. 4, pp. 267-277, December 2014 [Martinez-Diaz *et al.*, 2014]

Part I

Partial Fingerprint Registration Based on Orientation Field

Chapter 2

Related works and motivation for fingerprint pre-registration

This chapter begins with an overview of Automated Fingerprint Identification Systems (AFIS) development history, the drawbacks of currently existing AFIS in forensic applications, and recent developments in improving latent AFIS through Next Generation Identification systems. We then discuss about the importance of fingerprints as an important piece of evidence in forensic applications, as well as fingerprint's use in commercial applications. We review about different types of fingerprint matching techniques currently employed by automated fingerprint matchers, and also the importance of alignment as an important pre-processing stage in matching for improved performance. The limitations for adapting currently existing alignment methods to partial fingerprint recognition are discussed. The challenges faced by matching algorithms when comparing partial fingerprint against full fingerprint as well as some currently existing partial fingerprint matchers are reviewed. The Evaluation of Latent Fingerprint Technologies (ELFT) conducted by NIST to understand the feasibility of completely automated capability of commercial latent AFIS, and the conclusions derived from these evaluation are discussed.

We discuss about the discriminating ability of orientation fields in fingerprints, and how they are robust against fingerprint image quality. An overview of our orientation field based pre-registration algorithm which helps in significantly improving the identification accuracies of minutiae-based automated fingerprint matchers is discussed. We then review the works related with fingerprint registration already existing in the research literature and their limitations in adapting them for partial fingerprint registration. We conclude this chapter by providing a brief overview of the NIST-SD27 database used in our experiments. The detailed description of our proposed pre-registration algorithm and various experiments supporting their usefulness are provided in Chapter 3.

2.1. Automated Fingerprint Identification Systems

The projects to develop Automated Fingerprint Identification Systems (AFIS) started in early 1960*s*. These were initiated by the FBI in the United States, the Home Office in the United Kingdom, Paris Police in France, and the Japanese National Police. The emerging technology in electronic digital computers fostered the research towards assisting or replacing the labor-intensive process of classifying, searching, and matching tenprint cards used for personal identification [Holder *et al.*, 2011].

Any impression made by the friction ridge skin of the human finger is generally termed as *fingerprint*. Fingerprints which are revealed using some chemical or optical processing from a crime scene are called *latent fingerprints*. These are unintentionally left fingerprints found in the crime scenes. The latent fingerprints are then photographed, marked up for discriminatory features by a forensic fingerprint examiners, and are used to search by AFIS. Law enforcement agencies maintain a huge database of the fingerprints of individuals who are arrested or imprisoned. The forensic fingerprint database are typically collected by obtaining a rolled fingerprints from each finger. Such fingerprints in the database are called *tenprints* or *exemplars* fingerprints. When a latent fingerprint is found, the criminal investigators first search for the suspect in an AFIS to establish the identity of the individual to link with a particular criminal record. If there is a match, then the individual is linked to the crime under investigation.

In the realm of forensic analysis (*criminology*), the use of latent fingerprints is a routine procedure to identify suspects. Such practice has been followed for over a century now, and has most of the time proven to be pertinent in identifying the suspects. Consequently, the identity of an individual established on the basis of fingerprints is accepted by law enforcement agencies [Holder *et al.*, 2011] [Maltoni *et al.*, 2009].

Fingerprints are also widely used in civilian biometric recognition applications such as authentication, passport controls, biometric based digital identity, etc. Since the fingerprint is one of the oldest biometric traits, many techniques have been proposed in the literature for fingerprint recognition. It is comparatively a mature biometric trait compared against face, iris, voice, etc. AFIS are widely used for fingerprint recognition in both forensic as well as commercial domains. Most AFIS currently use two prominent ridge characteristics (called *minutiae*) namely ridge-endings and bifurcations to compare fingerprints. The minutiae-based decision is accepted as a proof of identity legally by courts in almost all countries around the world [Holder *et al.*, 2011] [Maltoni *et al.*, 2009].

There will also be situations where the latent fingerprints remains unidentified, typically referred as an Unsolved Latent File (ULF). As new exemplars are added into the AFIS, criminal investigators match them against ULF with the hope to find a match. It is possible that an

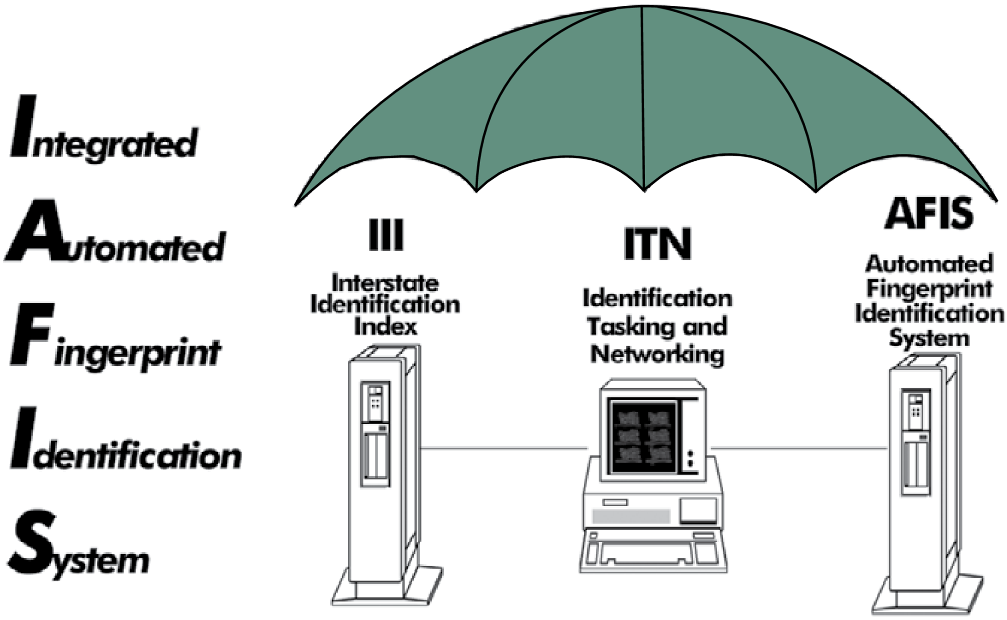


Figure 2.1: *IAFIS segments : An illustration of Integrated Automated Fingerprint Identification Systems [Holder et al., 2011]*

.

ULF from one jurisdiction can match a tenprint record stored in the AFIS database of another jurisdiction. A framework to integrate the AFIS databases from different jurisdictions and a combined search will maximize the chance of making a match. An example for such a framework is the Integrated-AFIS (IAFIS) maintained by Federal Bureau of Investigation (FBI) of United States of America (Figure 2.1).

2.1.1. Integrated-AFIS and Next Generation Identification systems

The Integrated Automated Fingerprint Identification Systems (IAFIS) is the worlds largest collection of criminal history information maintained by FBI. IAFIS provides automated fingerprint search capabilities, latent search capability, electronic image storage, and electronic exchange of fingerprints and responses [FBI-IAFIS]. The IAFIS not only maintains fingerprints, but also the corresponding criminal histories, mug shots, scars and tattoo photos, physical characteristics like height, weight, and hair and eye color, and aliases.

IAFIS consists of three integrated segments: the Identification Tasking and Networking (ITN) segment, the Interstate Identification Index (III), and AFIS (Figure 2.1). The ITN segment provide workflow management for tenprint, latent print and document processing. The III provides subject search, computerized criminal history, and criminal photo storage and retrieval. The AFIS searches the FBI fingerprint repository for matches to tenprint and latent fingerprints.

Major drawbacks to IAFIS are that it cannot store and search palmprints, accept and store 1,000-pixels-per-inch tenprint images. Towards meeting up with this end, FBI has started a project known as the Next Generation Identification Program (NGI). This program will further advance the FBI's biometric identification services, providing an incremental replacement of current IAFIS technical capabilities while introducing new functionality. The NGI system will offer state-of-the-art biometric identification services and provide a flexible framework of core capabilities that will serve as a platform for multimodal functionality [Holder *et al.*, 2011]. In future, IAFIS will be replaced with NGI.

2.2. Minutiae matching and alignment

In general, depending on the nature of the feature used by matching algorithms, fingerprint matching can be broadly classified into *correlation-based matching, minutiae-based matching* and *non-minutiae feature-based matching* [Maltoni *et al.*, 2009].

- In correlation-based matching, gray scale fingerprint images of both input and reference are superimposed and pixel correlations are computed between them.
- In minutiae-based matching, minutiae stored as sets of points are compared using point pattern matching algorithms.
- In non-minutiae feature-based matching, other features of fingerprints such as orientation fields, frequency maps, ridge shapes, texture information etc, are used for matching the input and the reference.

Irrespective of the core methodology used for fingerprint matching, the alignment between the input and the reference fingerprint is a crucial step. This is because the fingerprint images captured in different instances might have different rotation, translation or non-linear deformation between them. The main objective of fingerprint alignment is to estimate the transformation parameters between input and reference fingerprints.

There are two main approaches in pre-alignment, namely: *absolute pre-alignment* and *relative pre-alignment* [Maltoni *et al.*, 2009]. The orientation field based registration in this work falls under the category of relative pre-alignment.

- In *absolute pre-alignment*, the reference fingerprints are pre-aligned independently of the input fingerprint before storing in the database. The input fingerprint is pre-aligned just once before any comparisons are performed with the reference fingerprints. For absolute pre-alignment, the most common technique is to translate the fingerprint according to position of the core point. There are also other techniques which focus on absolute pre-alignment based on the shape of the external fingerprint silhouette, orientation of delta or

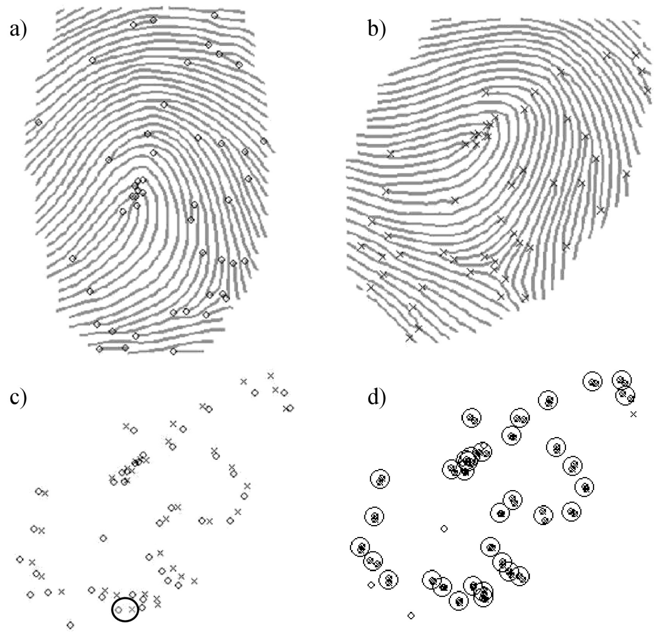


Figure 2.2: *A generalized Hough transform based alignment. a) and b) shows the minutiae extracted from input and reference fingerprint; c) input and reference minutiae superimposed; d) circles denotes minutiae pairs mated using generalized Hough transform technique [Maltoni* et al.*, 2009].*

core points, or average orientations in the neighborhood of cores. Since all these absolute pre-alignment depends on the singular points, and for latent fingerprints singular points are not guaranteed, absolute pre-alignment is not possible for latent scenario.

- In *relative pre-alignment*, the input fingerprint has to be pre-aligned with respect to the reference fingerprints while matching. The most common techniques in relative pre-alignment are performed by superimposing the singular points (core or delta), by comparing ridge features or by correlating the orientation images. Superimposing singular points are not feasible in latent scenario as they are not always guaranteed in latent fingerprint images. The ridge features, i.e, length and orientation of the ridge on which a minutiae resides, seem to be possible feature candidate but a reliable extraction of ridge features from bad or ugly quality latent fingerprints is a challenging problem. Estimation of orientation field is more reliable as compared against ridge feature extraction in latent fingerprints. So, we used the method of correlating the orientation images in this work to register a partial fingerprint in a full fingerprint.

The most widely used alignment method is based on minutiae. The main idea behind minutiae-based alignment is to search in the space of transformation parameters to find an optimal transformation with the maximum number of matched minutiae between the input and the reference fingerprints (Figure 2.2). One such methodology is based on the generalized Hough transform [Ratha *et al.*, 1996]. The main disadvantage for such technique is the inaccuracy in the transformation estimation due to discretization of the parameters space. Other approaches could be to use brute force to check for all possible correspondences between minutiae pairs. There exists some alignment techniques that augment minutiae with other supplementary features such as ridge information, orientation fields around a small neighborhood of minutiae, geometric relationships between minutiae and its neighbors, etc.

Alignment of full fingerprints is a well studied problem. But these methods are limited in alignment accuracy due to quantization of transformation parameters, or are not adapted for the partial fingerprint scenario. Partial fingerprints can arise in a number of situations, for example [Jea and Govindaraju, 2005] [Wang and Hu, 2011]: latent fingerprints lifted from crime scenes, due to small size of the fingerprint capturing devices, or an already enrolled fingerprint has noisy regions and is left only with a partial good/recognizable region for identification. The performance of the existing partial fingerprint identification systems mainly depends on the image quality, the number of minutiae available and other derived and extended features that can be obtained from the partial fingerprint region.

Various approaches in partial fingerprint identification [Wang and Hu, 2011] include:

- the use of localized secondary features derived from relative minutia information [Jea and Govindaraju, 2005].
- using representative points along ridge lines in addition to minutiae [Fang *et al.*, 2007].
- use of Level-3 features such as dots and incipients [Jain *et al.*, 2007a].

Most fingerprint matching algorithms in general assume approximately the same size of the minutiae set between the query and the reference minutiae for good identification accuracy [Jea and Govindaraju, 2005]. It is nevertheless frequent in some scenarios to have very different sizes between query and reference due to the situations discussed above. Trying to align a partial fingerprint to a full fingerprint only based on minutiae features could lead to errors. Law enforcement agencies employ AFIS to shortlist the suspects from its criminal database (exemplar / tenprint fingerprints). In such a scenario, it is crucial that the performance accuracy of AFIS is as good as possible. Latent fingerprints inherently are of poor quality, which leads to poor identification accuracy of AFIS in the latent scenario as compared to full fingerprint identification.

To evaluate the performance of feature extraction and matching techniques of commercial AFIS, NIST has conducted a multi-phase open project called Evaluation of Latent Fingerprint Technologies (ELFT) [NIST-ELFT, 2013].

- In Phase-I of ELFT, the best performing system reported a Rank-1 identification accuracy of 80% in which 100 latents were compared against $10,000$ rolled prints [NIST-ELFT-1, 2007].

- In Phase-II, Evaluation-1, the best performing system reported a Rank-1 identification accuracy of 97.2% in which 835 latents were compared against 100,000 rolled prints [Indovina *et al.*, 2009],
- in Phase-II, Evaluation-2, the best performing system reported a Rank-1 identification accuracy of only 63.4% in which 1,114 latents were compared against 100,000 rolled prints [Indovina *et al.*, 2012a].

The reported accuracies from Phase-I and Phase-II cannot be directly compared as the database and the quality of the latents were different. In [Indovina *et al.*, 2011a], it is concluded that only a limited class of latents benefits from automated procedures, but the procedures of marking the minutiae, determining the subjective quality of latents, etc still need to be carried out manually.

2.3. Motivation and proposed pre-registration technique

In this part of the thesis, we focus on the problem of aligning a partial fingerprint against a full fingerprint, especially of poor quality latents. Instead of minutiae, we used orientation fields (OF) to perform the alignment. We reduce fingerprint images to orientation images, and we look at the alignment problem as registering the partial fingerprint orientation image into the full fingerprint orientation image. Image registration is the process of overlaying (geometrically align) images of the same scene acquired in different time, different viewpoints and from different sensors [Brown, 1992; Lucas and Kanade, 1981].

Image registration is broadly classified into area-based and feature-based registration.

- In area-based registration, no image features are detected and directly focuses on matching stage. The matching strategy involves *correlation-like methods* or *template matching*, *Fourier methods*, *Mutual Information methods* and *optimization methods* [Brown, 1992; Maintz and Viergever, 1998; Zitova and Flusser, 2003].
- In feature-based registration, salient structures from the image are extracted to perform the matching. Feature-based registration is used when enough distinctive features are available. The features are matched using *spatial relations*, *invariant descriptors*, *relaxation methods* etc. [Brown, 1992; Maintz and Viergever, 1998; Zitova and Flusser, 2003]

We used area-based registration in our work. The OF representing the flow of ridges is a relatively stable global feature of fingerprint images, and it represents the intrinsic nature of the fingerprint. The representative OF of a fingerprint is very less affected by the type of capture device, contrast variations, and other quality effects compared to the input image or the minutiae. To improve the rank identification accuracy of minutiae-based matching, we consider only the minutiae around the region where the partial fingerprint orientation image is registered in the full fingerprint. This thereby reduces the search space of minutiae in the full fingerprint to approximately the size of partial fingerprint minutiae set, and consequently improves the performance of the minutiae-based matcher.

The main contributions of this work are as follows:

1. New correlation-based hierarchical registration method for orientation images to register a partial fingerprint in a full fingerprint.

2. Experimental exploration of various types of orientation field generation methods adequate for the registration.

3. Experimental demonstration of the performance improvement of minutiae-based matching by incorporating our registration algorithm to reduce the search space of minutiae in full fingerprints. In particular, our algorithm significantly improves the rank identification accuracy for poor quality latents (Bad and Ugly category) of NIST-SD27 database using NIST-Bozorth3 and MCC-SDK minutiae-based matchers.

In the following sections, we review related works on fingerprint orientation field based registration, describe the database used in our experiments. In the next chapter, we describe the similarity measures used in our algorithm, followed by a detailed description of the proposed algorithm, experiments, results and discussions.

2.4. Related works

In this section, we review the orientation field based fingerprint registration techniques in the literature, and its applicability in registering partial fingerprint images. A basic implementation of orientation-image registration requires computing the similarity between the input orientation image and the reference orientation image for every possible transformation considered between them (e.g., rotation and translation) [Maltoni *et al.*, 2009]. Table 2.1 summarizes various techniques in the literature for orientation field based fingerprint registration together with their limitations for partial fingerprint registration.

Liu *et al.* [2006] uses Normalized Mutual Information (NMI) as the similarity measure between orientation images to perform fingerprint registration. They align fingerprint images by maximizing NMI between the input and reference orientation images under different transformations. This technique is not suitable in aligning a partial fingerprint against full fingerprint

Method	Core technique	Limitations in partial fingerprint registration / latent scenario
[Liu *et al.*, 2006]	Maximize the Normalized Mutual Information between input and reference OF images	1) Needs large area overlaps 2) More reference sample required to correctly estimate OF distribution
[Nilsson and Bigun, 2005]	1) Singular point (SP) detection 2) $1D$ radiograms	1) SP not guaranteed in partial or latent fingerprints 2) Quantized projection angles, and require large area overlaps
[Yager and Amin, 2004] [Yager and Amin, 2006]	1) Distinctive Local Orientations 2) Generalized Hough Transform 3) Steepest Descent	1) SP not guaranteed in partial or latent fingerprints 2) Needs large area overlaps

Table 2.1: *Summary of orientation field based fingerprint registration techniques in the literature together with their limitations to be applied for partial fingerprint registration.*

as reported in [Liu *et al.*, 2006]. In this approach, for good alignment, the size of input and reference orientation images should be almost of similar size. Another drawback in this technique is the necessity of enough samples of reference fingerprints to correctly estimate the distribution of the orientation field, otherwise it leads to incorrect alignment. Both of these scenarios which requires same size between input and reference, as well as training reference samples are not pertinent in forensic fingerprint identification.

Nilsson and Bigun [2005] focus on registering the fingerprints by complex filtering and by $1D$ projections of orientation images. Given the orientation images of the fingerprints represented as complex orientation fields, they first use specific complex filters to locate singular points (core and delta) in the fingerprint. Once these singular points are located in both input and reference orientation images, transformation parameters (rotation and translation) are estimated by superimposing the singular points.

Another technique studied in [Nilsson and Bigun, 2005] is $1D$ projections of orientation images. In this method, the fingerprint image is decomposed into 6 equally spaced directions called orientation images, and a Radon transformation is used to compute $1D$ projections of these orientation images (called radiograms). A translation parameter is estimated between a pair of radiograms from input and reference belonging to the same projection angle by a correlation measure. When utilizing this method, it is already assumed that the rotation alignment between input and reference is negligible or is already corrected. These techniques cannot be adapted to register partial fingerprints because singular points are not always guaranteed in partial fingerprint, and the area of overlap between input and reference is often small.

Yager and Amin [2004, 2006] explore three types of orientation field registration techniques summarized as follows:

1. *Distinctive Local Orientations (DLO):* This approach mainly depends on distinctive patterns in the orientation field called singular points (core and delta). This is similar to the work by Nilsson and Bigun [2005] except for the technique to locate the singular points.

2. *Generalized Hough Transform (GHT):* In this approach, the space of all possible transformation parameters is discretized and analyzed for the best transformation.

3. *Steepest Descent (SD):* Starting with some initial parameters, this algorithm evaluates a cost function. It then evaluates a sample of local neighborhood in the parameter space and selects the parameters that give greatest descent in the cost. This procedure is repeated until a local minimum has been found.

It is reported by Yager and Amin [2004] that both GHT and SD do not perform well when the area of overlap between the input and reference is small, similar to the case using NMI [Liu *et al.*, 2006]. So, both GHT and SD are not suitable for partial fingerprint registration. Moreover, DLO looks for singular points, and it is not assured that a partial fingerprint will have singular point in it. So, all the orientation field registration techniques proposed in the literature are not suitable for partial fingerprint registration, and cannot be quickly adapted to this scenario.

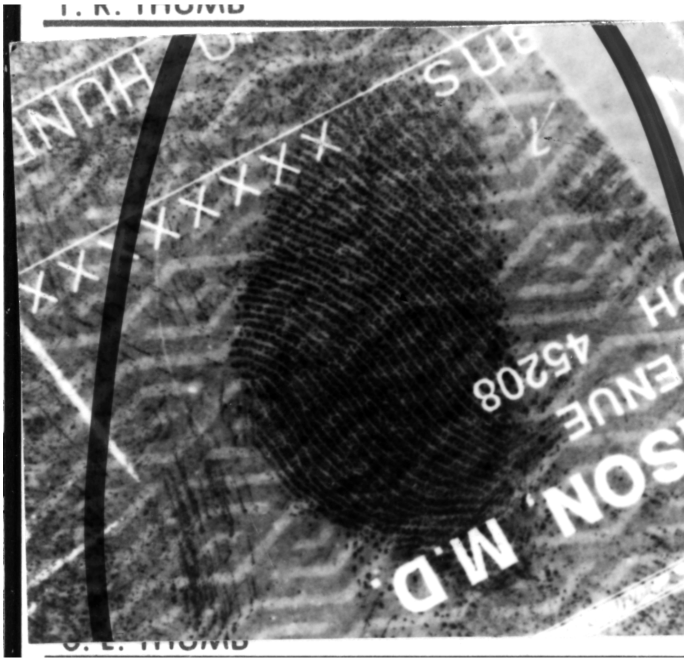
2.5. Database: NIST-SD27

NIST Special Database 27 (NIST-SD27) [Garris and McCabe, 2000] is a publicly available forensic fingerprint database which comprises of 258 latent fingerprint images, its matching 258 tenprint images and their minutiae sets. The NIST-SD27 minutiae set database is classified into two [Garris and McCabe, 2000] [Krish *et al.*, 2013b]: 1) *ideal*, and 2) *matched* minutiae sets.

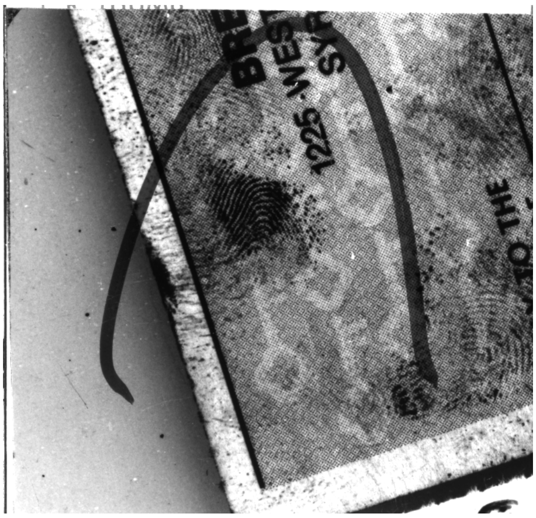
- The *ideal* minutiae set for latents was manually extracted by a forensic examiner without any prior knowledge of its corresponding tenprint image.

- The *ideal* minutiae set for tenprints was initially extracted using an AFIS, and then these minutiae were manually validated by at least two forensic examiners.

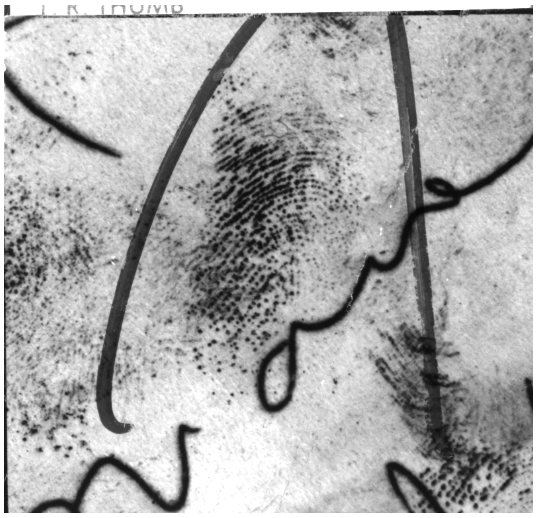
- The *matched* minutiae set contains those minutiae which are in common between the latent and its mated tenprint image. There is a one-to-one correspondence in the minutiae between the latent and its mate in the matched minutia set. This ground truth (matched minutiae set) was established manually by a forensic examiner looking at the images and the *ideal* minutiae.



(a) Good



(b) Bad



(c) Ugly

Figure 2.3: *Subjective quality classification of latent fingerprint images in NIST-SD27 database.*

The matched minutia sets are a subset of ideal minutia set, but the location and orientation information are not exactly the same. There are slight variations in the location and orientation attributes between ideal and its corresponding matched minutia set originated in the annotations by the experts. For example, G028T1I and G028T1M of NIST-SD27 contain 123 and 20 minutiae respectively. G028T1I is the ideal minutia set and G028T1M is its corresponding matched minutia set. The pair $(X, Y, Orientation) = (562, 189, -68)$ of ideal and $(564, 182, -73)$ of matched are supposed to be same minutia in the fingerprint. However there is a slight variation with an euclidean distance of 7.2 pixel units. This variation might be because of the uncertainty introduced by the software used by the examiner while generating the matched minutia set. In general, there is a small non-linear deformation between the ideal and matched minutia sets of the tenprints.

The NIST-SD27 database consists of latent fingerprint images of varying quality. Each image is of 800×768 pixels in size and has been scanned at 500 pixels per inch (ppi) as a gray scale image. It already contains a classification of the latent fingerprints based on the subjective quality of the image into Good, Bad and Ugly, containing 88, 85 and 85 fingerprints respectively determined by the forensic examiner. The average number of minutiae for Good, Bad and Ugly category latents are 32, 18 and 12 respectively. Figure 2.3 shows sample images from the NIST-SD27 database which belong to Good, Bad and Ugly quality categories respectively. In [Jain and Feng, 2011], it is shown that there is a correlation between this subjective quality classification and the matching performance.

Chapter 3

Hierarchical pre-registration: Algorithm and experiments

This chapter describes in detail the proposed hierarchical algorithm to register the orientation field of partial fingerprint to the orientation field of full fingerprint, and experimental results. We discuss various orientation field based similarity measures used in our proposed algorithm such as correlation, Manhattan and Euclidean based distance similarity adapted for orientations represented as complex numbers, and similarity based on orientation consistency. We then explain in detail the proposed algorithm for pre-registration. Our algorithm is hierarchical in nature, and are performed sequentially in two levels. The first level of our proposed algorithm performs a normalized correlation based similarity, and in the second level, utilizes various other similarity measures and performs score level fusions to finalize the partial fingerprint registration on to the full fingerprint. In both levels, we use orientation fields to compute similarity measures.

We made use of various types of orientation field estimation methodologies for fingerprints such as manual extraction, dictionary-based estimation, orientation reconstructed from minutiae, orientation directly estimated from gray-scale fingerprint image, and an average of minutiae reconstructed and image generated orientation field. An overview of all the methods together with an experiment supporting the best representative orientation field for fingerprint is described. This is followed with experiments which shows significant improvement of rank identification accuracies of minutiae-based matchers namely, NIST-Bozorth3 and MCC-SDK when our proposed hierarchical pre-registration is applied. The chapter concludes with runtime analysis and discussions summarizing the usefulness of our proposed algorithm.

3.1. Similarity measures

In this section, we introduce various similarity measures that are used in our hierarchical registration algorithm.

Let $\mathbf{U}$ and $\mathbf{V}$ be discrete images of the same size, represented as a 2D array where the array elements may represent values of gray pixels (*zero-order tensors*), color pixels (*first-order tensors*) or local directions (*second-order tensors*).

The Schwarz inequality:

$$\frac{|\langle \mathbf{U}, \mathbf{V} \rangle|}{\|\mathbf{U}\| \times \|\mathbf{V}\|} \leq 1 \tag{3.1}$$

holds for $\mathbf{U}$ and $\mathbf{V}$ [Bigun, 2005, Chapter 3]. Here, $\langle \mathbf{U}, \mathbf{V} \rangle$ is the inner product between $\mathbf{U}$ and $\mathbf{V}$ calculated as :

$$\langle \mathbf{U}, \mathbf{V} \rangle = \sum_{r,c} \mathbf{U}(r,c)^* \cdot \mathbf{V}(r,c) \tag{3.2}$$

where r, c are the indices, $\mathbf{U}(r,c)^*$ is the complex conjugate of $\mathbf{U}(r,c)$, and $\|\mathbf{U}\|$ and $\|\mathbf{V}\|$ are the L_2 norms of $\mathbf{U}$ and $\mathbf{V}$ respectively.

The L_2 norm $\|\mathbf{U}\|$ is calculated as:

$$\|\mathbf{U}\| = \left[\sum_{r,c} \mathbf{U}(r,c)^* \cdot \mathbf{U}(r,c) \right]^{1/2} \tag{3.3}$$

and similarly for $\|\mathbf{V}\|$.

The normalized correlation between $\mathbf{U}$ and $\mathbf{V}$, referred to as Schwarz Similarity (SS) hereafter is defined as:

$$SS(\mathbf{U}, \mathbf{V}) = \frac{|\langle \mathbf{U}, \mathbf{V} \rangle|}{\|\mathbf{U}\| \times \|\mathbf{V}\|} \tag{3.4}$$

Because of Eq. (3.1), the interval for $SS(\mathbf{U}, \mathbf{V})$ is in the range $[0, 1]$. By calculating SS as a similarity measure, we can locate a given pattern (a small image) in a large image. When $SS(\mathbf{U}, \mathbf{V})$ is 1, then both $\mathbf{U}$ and $\mathbf{V}$ are viewed as most similar patterns, and when $SS(\mathbf{U}, \mathbf{V})$ is 0, they are least similar [Bigun, 2005].

Assuming **U** and **V** represent local directions (*second-order tensors*) in the range $[-90^\circ, +90^\circ)$, we define the Manhattan-based Similarity $MS(\mathbf{U}, \mathbf{V})$ as

$$MS(\mathbf{U}, \mathbf{V}) = \cos\left(\frac{1}{N}\sum_{r,c}\left(\Delta_{r,c}^{U,V}\right)\right) \tag{3.5}$$

and Euclidean-based Similarity $ES(\mathbf{U}, \mathbf{V})$ as

$$ES(\mathbf{U}, \mathbf{V}) = \cos\left(\left[\frac{1}{N}\sum_{r,c}\left(\Delta_{r,c}^{U,V}\right)^2\right]^{1/2}\right) \tag{3.6}$$

where

$$\Delta_{r,c}^{U,V} = \min\left(|\mathbf{U}(r,c) - \mathbf{V}(r,c)|, 180 - |\mathbf{U}(r,c) - \mathbf{V}(r,c)|\right) \tag{3.7}$$

$\Delta_{r,c}^{U,V}$ takes values in the range $[0, +90^\circ)$ and N is the size in pixels of **U** or **V** (**U** and **V** are of same size). Because of Eq. (3.7), the value of MS and ES will be in the range $[0, 1]$.

The Consistency Similarity $CS(\mathbf{U}, \mathbf{V})$ (which was proposed in [Jiang *et al.*, 2006]) between **U** and **V** is defined as:

$$CS(\mathbf{U}, \mathbf{V}) = \frac{1}{N}\left|\sum_{r,c} e^{i2(\mathbf{U}(r,c) - \mathbf{V}(r,c))}\right| \tag{3.8}$$

where i is the complex number $\sqrt{-1}$, and $|z|$ is the magnitude of complex number z. The consistency similarity CS averages the unit vector whose phase is doubled orientation difference, and the value is in the range $[0, 1]$.

All the similarity measures SS, MS, ES and CS are in the normalized range $[0, 1]$ and these measures can be fused directly.

3.2. Algorithm

The algorithm to register the orientation field of the latent fingerprint with that of the tenprint fingerprint is achieved in two hierarchical levels. In the first level, we perform the normalized correlation between the OF of latent and tenprint for various rotation alignments in the range $[-45^\circ, +45^\circ]$ with 1° increments. We then shortlist the correlation peaks for each rotation. These peaks are the possible target locations for registration.

We observed that deciding the target location only based on the normalized correlation score does not always yield satisfactory results. Therefore, a second level, on these candidate locations, we calculate MS, ES and CS similarity measures between the latent centered at the peak location in the tenprint. The final registration location is chosen from the candidate locations as the one that maximizes the mean similarity between SS, MS, ES and CS. This gives better registration accuracies than deciding only based on SS. In the following section we describe this approach in more detail.

3.2.1. Level 1: Normalized correlation

Step 1: Generate the orientation field L for the latent fingerprint and T for the tenprint fingerprint as detailed in Section 3.3. The orientations are obtained for 16×16 block sizes, and are in the range $[-90°, +90°)$.

Fig. 3.1(a), Fig. 3.1(b) shows the OF reconstructed from the minutiae set of latent and tenprint respectively. The expected outcome of the registration algorithm is to locate the minutiae region shown in Fig. 3.1(c).

Step 2: Generate the orientation tensors $\bar{L}$ and $\bar{T}$ for the latent L and tenprint T respectively in double angles (i.e, in the range $[-180, +180]$ degrees) using complex numbers, as follows:

$$\begin{aligned} \bar{L} &= \exp(i \times 2 \times \theta_L) \\ \bar{T} &= \exp(i \times 2 \times \theta_T) \end{aligned} \tag{3.9}$$

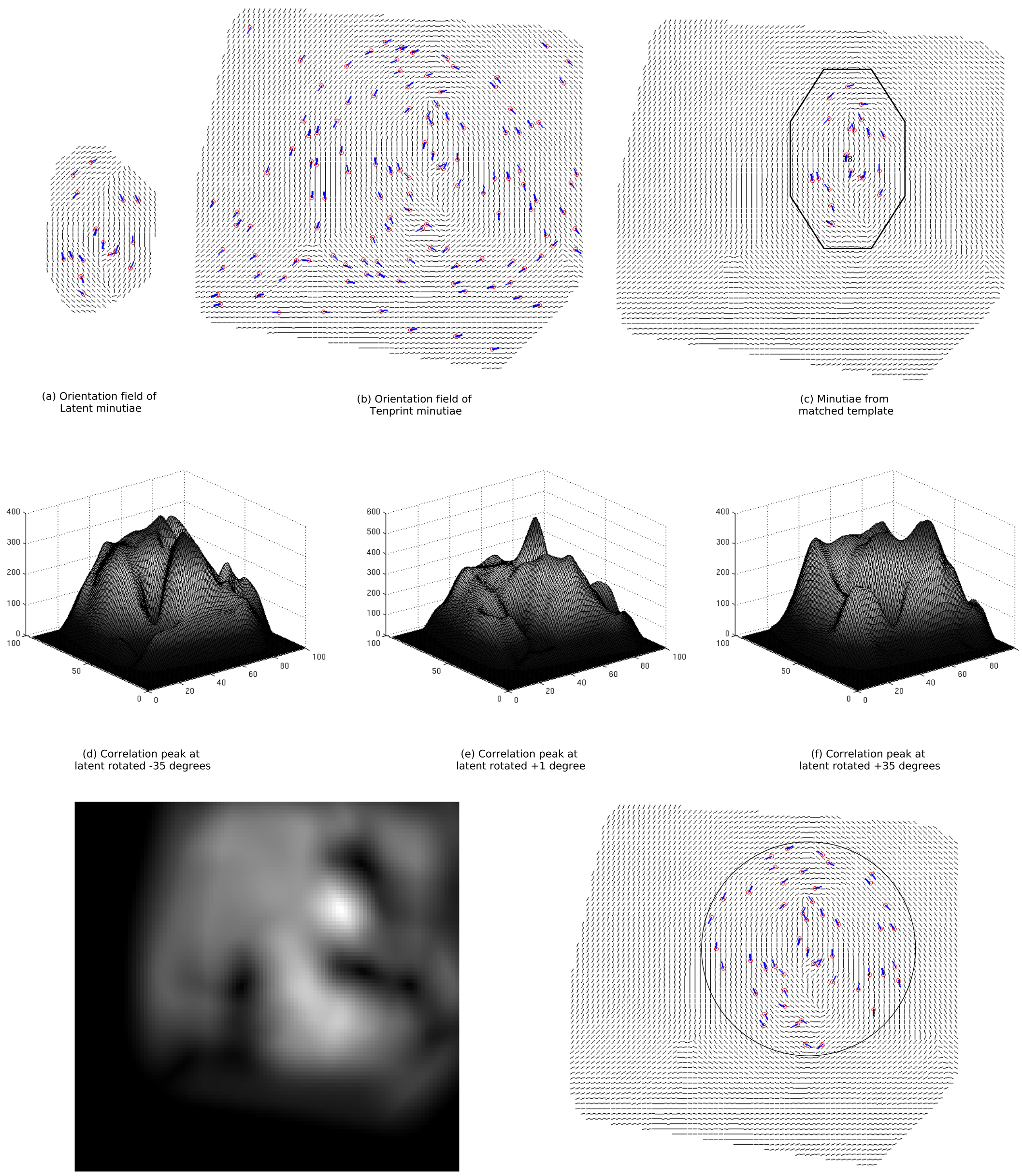
where i is the complex number $\sqrt{-1}$, θ_L and θ_T are the angles of L and T from Step 1. $\bar{L}$ is the smallest rectangular region that covers the latent minutiae.

For each subregion $\bar{T}_s$ of $\bar{T}$ that is of the same size as $\bar{L}$ located at a position indexed by s, we can find the inner product between $\bar{L}$ and $\bar{T}_s$ as follows:

$$\langle \bar{L}, \bar{T}_s \rangle = \sum_{r,c} \bar{L}(r,c)^* \cdot \bar{T}_s(r,c) \tag{3.10}$$

where r, c are the indices, $\bar{L}(r,c)^*$ is the complex conjugate of $\bar{L}(r,c)$.

Step 3: Define the bounding box for the latent orientation tensors $\bar{L}$ by discarding the background. The bounding box can be estimated by the minimum and maximum row and column numbers that correspond to the foreground of latent orientation tensors.



(a) Orientation field of Latent minutiae

(b) Orientation field of Tenprint minutiae

(c) Minutiae from matched template

(d) Correlation peak at latent rotated -35 degrees

(e) Correlation peak at latent rotated +1 degree

(f) Correlation peak at latent rotated +35 degrees

(g) Latent pattern identified in tenprint for rotation alignment of +1 degree

(h) Minutiae subset generated after registration

Figure 3.1: *Various stages in the registration algorithm shown on G028L1 (latent) and G028T1 (tenprint) of NIST-SD27. (a) and (b) are the orientation field (OF) reconstructed from the ideal minutiae set, with the minutiae plotted over the OF. (c) is the region in the tenprint that is to be found after registration of (a) into (b). (d), (e) and (f) are the correlation peaks when the latent is rotated at* $-35°$, $1°$ *and* $+35°$ *respectively and correlated with tenprint. (g) is the region where the latent pattern is identified in the tenprint based on the proposed score fusion for rotation alignment of +1 degree. (h) is the minutiae region selected by our pre-alignment algorithm.*

Step 4: When searching for the pattern $\bar{L}$ in $\bar{T}$, it is possible that $\bar{L}$ is not perfectly aligned with $\bar{T}$, rotation wise. To compensate for the rotation alignment, we need to test the latent $\bar{L}$ against tenprint $\bar{T}$ for various rotations of $\bar{L}$. In our experiments, we rotate $\bar{L}$ in the range $[-45°, +45°]$ with a step size $\Delta\theta$ of $1°$ to compensate for rotation alignment to generate $\bar{L^{\theta}}$. A geometric rotation of $\Delta\theta$ implies a related rotation of the tensor field of $2\Delta\theta$.

Step 5: The correlation is obtained by generating $\langle \bar{L^{\theta}}, \bar{T_s} \rangle$ for all locations s in $\bar{T}$. The result of this operation is a complex image. We then observe the correlation peaks for all θ (magnitude of the complex image).

Fig. 3.1(d), Fig. 3.1(e), Fig. 3.1(f) shows the magnitude of the correlation images of $\bar{L}^{-35°}$, $\bar{L}^{+1°}$ and $\bar{L}^{+35°}$ with $\bar{T}$ respectively.

Step 6: For each θ from the correlated result, find the location of the peak $s^{\theta} = (r_m^{\theta}, c_m^{\theta})$, i.e, the location with maximum magnitude in the correlated image. The peak in the correlated image is where $\bar{L^{\theta}}$ agrees the most in $\bar{T}$. $S = \{(r_m^{\theta}, c_m^{\theta})\}$ is the set containing the coordinates of the correlation peaks for all θ.

Step 7: For all orientations θ, calculate $SS(\bar{L^{\theta}}, \bar{T_s^m})$, where $\bar{T_s^m}$ is the subregion in $\bar{T}$ whose center is $s^{\theta} = (r_m^{\theta}, c_m^{\theta})$. SS is the normalized correlation measure as defined in Eq. (3.4).

The correlation and normalized correlation are essentially equivalent in the scenario where θ_L and θ_T are not estimated from gray pixel gradients but reconstructed from minutiae orientations. Consequently, the orientation tensors $e^{i2\theta_L}$ and $e^{i2\theta_T}$ are complex numbers falling on a unit circle. So, the magnitude of the orientation tensors thus obtained are always 1.

3.2.2. Level 2: Fusion of similarity scores

Step 8: For each $s^{\theta} = (r_m^{\theta}, c_m^{\theta}) \in S$, calculate $MS(\bar{L^{\theta}}, \bar{T_s^m})$, $ES(\bar{L^{\theta}}, \bar{T_s^m})$ and $CS(\bar{L^{\theta}}, \bar{T_s^m})$ as defined in Eq. (3.5), Eq. (3.6) and Eq. (3.8) respectively.

Step 9: SS, MS, ES and CS are all similarity scores in the range $[0, 1]$, where 0 denotes minimum similarity and 1 denotes maximum similarity.

We perform score fusion of SS, MS, ES and CS based on the mean rule, and look for the $s^{\theta} = (r_m^{\theta}, c_m^{\theta}) \in S$ for which the fused similarity score is maximum.

Step 10: The resulting (r_m^θ, c_m^θ) is the location in the tenprint where the latent rotated at θ is registered with best alignment (see Fig. 3.1(g)). The center of the latent L is registered to (r_m^θ, c_m^θ) in tenprint T, and with a radius half the diagonal length of the bounding box of the latent orientation field, a subset of minutiae which falls inside this circular region is chosen (see Fig. 3.1(h)).

3.3. Types of Orientation Field estimation techniques

The orientation field describes the coarse structure, or basic shape of a fingerprint. It is defined as the local orientation of the ridge-valley structure. Orientation fields (or directional fields) falls under the Level-One detail of fingerprint feature categories. A fingerprint image gradually faded into corresponding orientation image is shown in Figure 3.2.

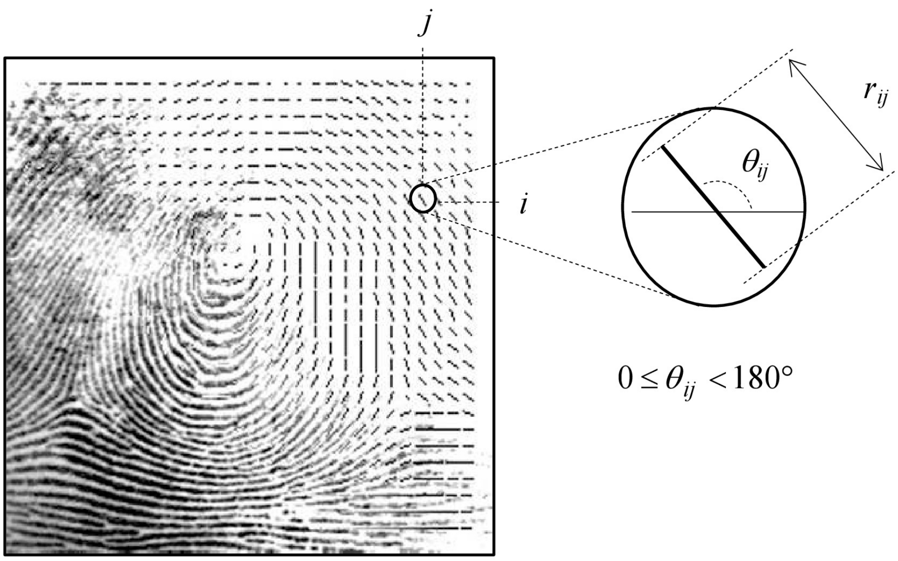


Figure 3.2: *Orientation field of a fingerprint image shown partly [Maltoni et al., 2009].*

In Figure 3.2, the orientation field is computed over a square-meshed grid of size 16×16. Each element θ_{ij} corresponds to the node $[i, j]$ of square-meshed grid located over the pixel $[x_i, y_i]$ denotes the average orientation of the fingerprint ridges in a neighborhood of $[x_i, y_i]$. The value r_{ij} denotes the reliability or consistency of the orientation θ_{ij}. A low value for r_{ij} denotes noisy regions and high value for good quality regions in the fingerprint image.

In this study, we have used five different techniques for computing the orientation field of the fingerprints, and are briefly explained in the following subsections. Various OF estimation techniques are summarized in Table 3.1.

OF Type	Core Technique
MANUAL_OF	Manually marked orientation field from the latent fingerprint image [Feng *et al.*, 2013]
DICT_OF	Orientation field estimated directly from fingerprint image using local Fourier analysis [Jain and Feng, 2009], and then performing context based correction of the OF using dictionary lookup of orientation patches [Feng *et al.*, 2013].
MINU_OF	Orientation field reconstructed from the minutiae [Feng and Jain, 2011]. The work by Feng and Jain [Feng and Jain, 2011] in reconstructing the fingerprint image from minutia sets alone, and successfully launching attacks against fingerprint recognition system indicates that the fidelity of the reconstructed OF to the actual OF is significant.
IMG_OF	Orientation field estimated directly from the fingerprint image using gradient based approach [Alonso-Fernandez *et al.*, 2007]. The gradients are estimated using a gaussian derivative filter. The orientation image thus obtained is a dense OF. The OF is down-scaled using gaussian pyramid approach.
AVG_OF	Orientation field estimated by taking the average of both of *IMG_OF* and *MINU_OF*.

Table 3.1: *Summary of orientation field estimation techniques used in this work.*

3.3.1. Manually extracted

MANUAL_OF: The OF for the latent fingerprints were manually extracted by the authors of [Feng *et al.*, 2013] for NIST-SD27, and is made publicly available. It is a common practice in friction ridge examinations to perform manual tasks for generating relevant discriminatory features useful for individualization.

3.3.2. Dictionary based

DICT_OF: The dictionary-based orientation field estimation consists of an offline dictionary construction stage and an online orientation field estimation stage [Feng *et al.*, 2013]. This procedure is summarized in Figure 3.3.

- In the offline stage, orientation fields of good quality fingerprints consisting of various patterns (arch, loop and whorl) are used to construct a dictionary of orientation patches.
- The online stage is one in which the orientation field is calculated automatically for the given fingerprint and involves following steps:

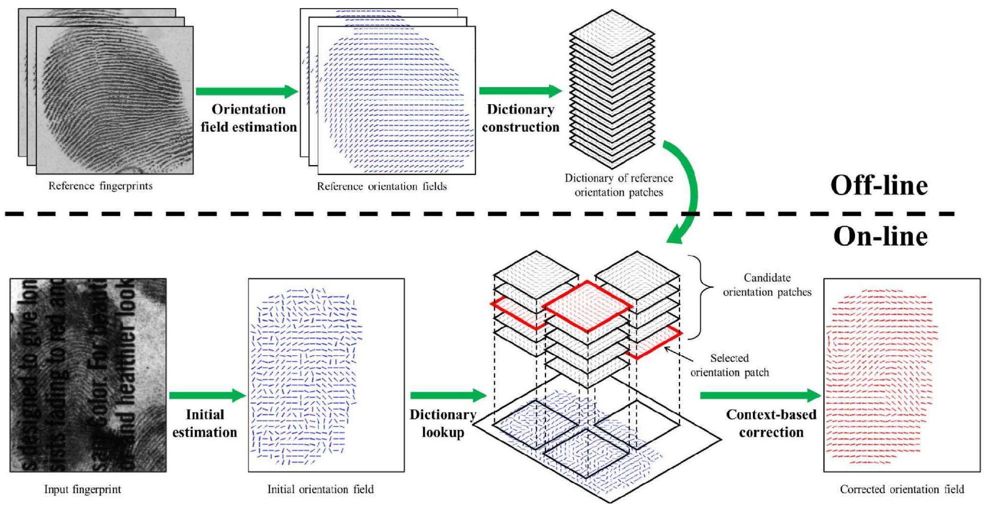


Figure 3.3: *Dictionary based orientation field estimation [Feng et al., 2013].*

1. *Initial estimation*: Orientation field estimated directly from fingerprint image using local Fourier analysis [Jain and Feng, 2009]. The dominant orientation in a 16×16 block is computed by detecting the peak in the magnitude spectrum of the local image. Due to the poor quality of fingerprint in some regions, it is possible that OF thus estimated is noisy. But, these noises are not removed out by any kind of OF smoothing techniques such as Gaussian smoothing or average smoothing. The smoothing is avoided because a correct orientation patch maybe degraded due to noisy neighboring patches.

2. *Dictionary lookup*: The OF thus obtained is divided into overlapping patches. For a given orientation patch belonging to foreground, a list of orientation patches from the dictionary are retrieved which are sorted according to similarity with the patches from foreground.

3. *Context-based correction*: Out of the list of candidate orientation patches retrieved from the dictionary for an orientation patch of foreground, a single candidate patch need to be chosen. To determine this single dictionary candidate, contextual information is used. For each of the patch belonging to foreground, there corresponds a list of candidate dictionary patches. Appropriate dictionary candidate are chosen to correct the foreground patches such that an energy function is minimized.

   The energy functions are designed based on the following two factors:

   a) the similarity between the dictionary orientation patch and the foreground orientation patch.

   b) the compatibility between neighboring dictionary orientation patches.

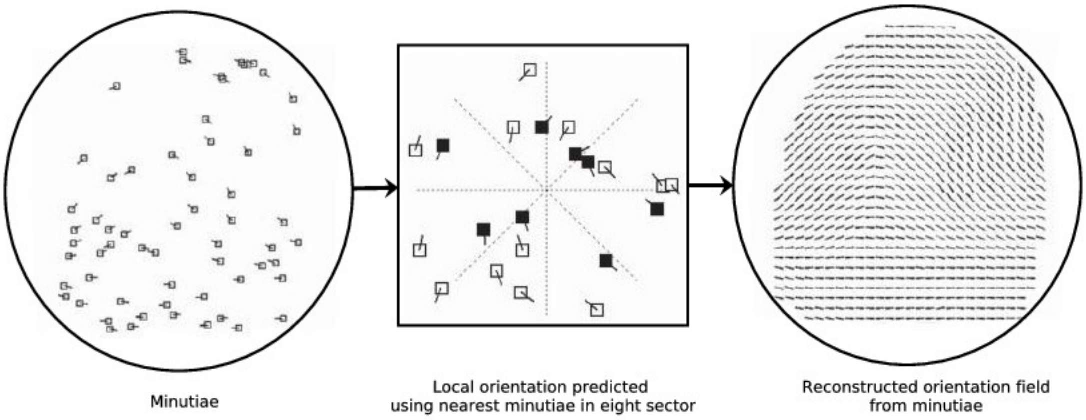


Figure 3.4: *Minutiae based orientation field reconstruction [Feng and Jain, 2011]. The solid black boxes are the nearest minutiae in each sector.*

3.3.3. Reconstructed from minutiae

$MINU\_OF$: In this category, orientation field is reconstructed directly from the minutiae alone [Feng and Jain, 2011]. Assume a blank image with the minutiae plotted on it. The fingerprint image is divided into non-overlapping blocks of 8×8 or 16×16 pixels. The foreground blocks are the ones with minutiae present in it. An orientation values is computed for each of such foreground block. Consider a line passing through the non-overlapping blocks (as shown in Figure 3.4 which divides the image into 8 equally spaced sectors. The local ridge orientation at each block is then predicted by using the nearest minutiae in each of the 8 sectors. The foreground region of interest is the region falling within the convex hull of minutiae.

Let $M = \{x_i, y_i, \alpha_i\}$, $1 \leq i \leq N$ be the fingerprint minutiae set consisting of N minutiae, where (x_i, y_i) is the spatial location and α_i the direction of the i^{th} minutiae. The minutiae direction α_i is doubled to make α_i equivalent to $\alpha_i + \pi$. For K minutiae selected from the eight sectors, the cosine and sine components of $2\alpha_i$ are computed and summed as follows:

$$u = \sum_{i=1}^{K} \cos(2\alpha_i) w_i, \tag{3.11}$$

$$v = \sum_{i=1}^{K} \sin(2\alpha_i) w_i, \tag{3.12}$$

where w_i is a weighting function. w_i is taken as the reciprocal of the euclidean distance between the block center and the i^{th} minutiae. This makes the minutiae direction dominate the ridge orientation of neighboring blocks.

The orientation at block (m, n) is computed as:

$$O(m, n) = \frac{1}{2} \arctan\left(\frac{u}{v}\right) \tag{3.13}$$

A reconstructed orientation field minutiae is depicted in Figure 3.4. The work by Feng and Jain [Feng and Jain, 2011] in reconstructing the fingerprint image from minutia sets alone, and successfully launching attacks against fingerprint recognition system indicates that the fidelity of the reconstructed OF to the actual OF is significant. Also, the performance of the algorithm in reconstructing the OF did not drop much even when only 60% of minutiae are only available for OF reconstruction.

3.3.4. Estimated directly from fingerprint image

IMG_OF: Orientation field estimated directly from the fingerprint image using gradient based approach [Alonso-Fernandez *et al.*, 2007] [Maltoni *et al.*, 2009] [Bazen and Gerez, 2002]. This is the most natural approach for extracting local orientations of the fingerprint image.

The elementary orientations in the image are given by gradient $\nabla I(x, y)$ which is a two-dimensional vector $[G_x, G_y]$ defined as:

$$\nabla I(x, y) = [G_x, G_y] = \left[\frac{\partial I(x, y)}{\partial x}, \frac{\partial I(x, y)}{\partial y}\right] \tag{3.14}$$

where I represent the gray scale fingerprint image, G_x and G_y are the derivatives of I at $[x, y]$ with respect to the x and y direction respectively. The gradient phase angle denotes the direction of the maximum change in pixel intensity. In principle, orientation field is perpendicular to the gradient.

We used gaussian derivative kernel to estimate the components of the gradients. For a gaussian $h_\sigma(x, y)$, the gaussian derivative along x-direction is given by

$$\frac{\partial h_\sigma(x, y)}{\partial x} \tag{3.15}$$

and the gaussian derivative along y-direction is given by

$$\frac{\partial h_\sigma(x, y)}{\partial y} \tag{3.16}$$

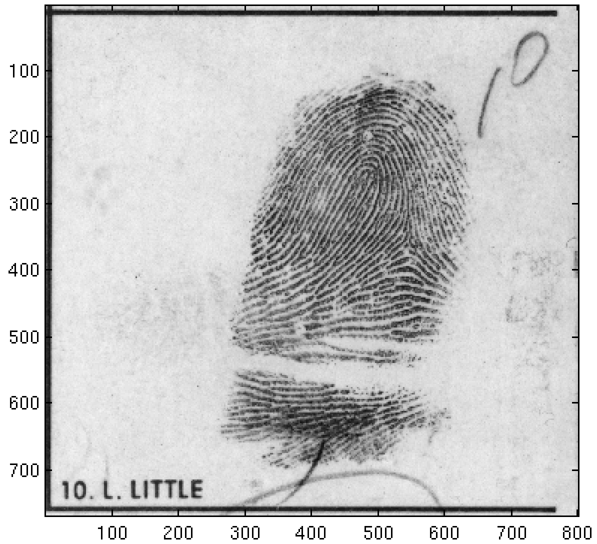
where

$$h_\sigma(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3.17}$$

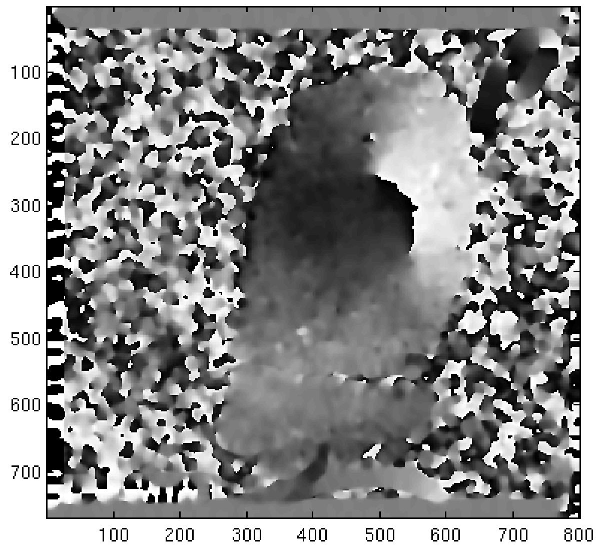
The variances and cross-covariances of gradients G_x and G_y are smoothed using a gaussian kernel around a windows size W as follows:

$$G_{xx} = \sum_W G_x^2 \tag{3.18}$$

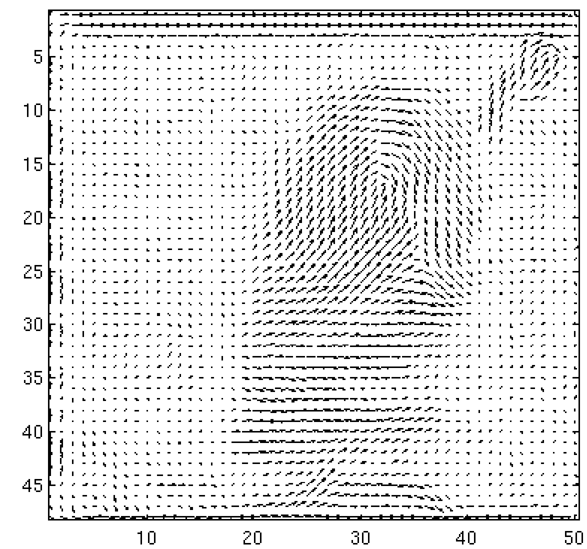
$$G_{yy} = \sum_W G_y^2 \tag{3.19}$$



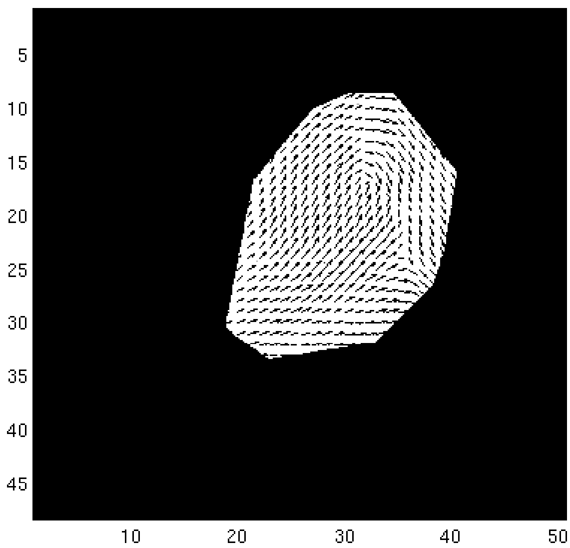
(a) Fingerprint image



(b) Intensity image



(c) Gradient based OF



(d) Convex Hull based region of interest

Figure 3.5: *Gradient based orientation field estimated directly from fingerprint image. (a) is the gray scale fingerprint image. (b) and (c) shows the magnitude and angle of the orientation images obtained. (d) is the region of interest estimated from the convex hull of minutiae.*

$$G_{xy} = \sum_{W} G_x G_y \tag{3.20}$$

where G_{xx} and G_{yy} are the variances, and G_{xy} the cross-covariances of G_x and G_y respectively.

The gradient direction within a window W centered at $[x_i, y_j]$ is given by:

$$\Phi_{ij} = \frac{1}{2}\text{atan2}\left(2G_{xy}, G_{xx} - G_{yy}\right) \tag{3.21}$$

and the orientation of the fingerprint ridge θ_{ij} is given by

$$\theta_{ij} = \frac{\pi}{2} + \Phi_{ij} \tag{3.22}$$

as fingerprint orientation is taken as the perpendicular to the gradient direction. atan2 is defined and used as in the Matlab environment.

The orientation image thus obtained is a dense orientation image, i.e, orientation estimated for each pixel of the fingerprint image. The orientation image is down-scaled using gaussian pyramid approach to obtain orientations for 16×16 blocks as shown in Figure 3.5.

3.3.5. Averaged orientation field

AVG_OF: Orientation field estimated by taking the average of both of IMG_OF and $MINU\_OF$.

AVG_OF is estimated using the technique proposed in [Kass and Witkin, 1987], also detailed in [Maltoni *et al.*, 2009, Chapter 3] to average local gradients.

Let θ_k^i and θ_k^m be the orientation corresponding to k^{th} block of IMG_OF and $MINU\_OF$ respectively. We double the angles to encode them by vectors:

$$\bar{d_k^i} = [\cos(2\theta_k^i), \sin(2\theta_k^i)] \tag{3.23}$$

$$\bar{d_k^m} = [\cos(2\theta_k^m), \sin(2\theta_k^m)] \tag{3.24}$$

where $\bar{d_k^i}$ and $\bar{d_k^m}$ are the vectors corresponding to θ_k^i and θ_k^m.

We then find the average vector $\bar{d_k^a} = [\text{avgCos}_k^a, \text{avgSin}_k^a]$ where

$$\text{avgCos}_k^a = \frac{1}{2}(\cos(2\theta_k^i) + \cos(2\theta_k^m)) \tag{3.25}$$

$$\text{avgSin}_k^a = \frac{1}{2}(\sin(2\theta_k^i) + \sin(2\theta_k^m)) \tag{3.26}$$

From this average vector $\bar{d}_k^a$, find the corresponding orientation of the k^{th} block of AVG_OF as

$$\theta_k^a = \frac{1}{2}\text{atan2}(\text{avgSin}_k^a, \text{avgCos}_k^a) \tag{3.27}$$

The double angle representation avoids any errors due to circularity of angles while averaging. Here, we assume θ_k^i , θ_k^m and θ_k^a are in radians.

Out of these five different techniques, $MANUAL\_OF$ and $DICT\_OF$ were used for latent fingerprints, whereas $DICT\_OF$, IMG_OF, $MINU\_OF$ and AVG_OF were used for tenprints. All the OF estimated were of 16×16 block size. The region of interest for the fingerprint is considered to be the region inside the convex hull of the corresponding ideal minutiae of the fingerprint present in NIST-SD27.

3.4. Experiments

We perform experiments on Good, Bad and Ugly quality classifications of NIST-SD27 to report the accuracy of the proposed registration algorithm. 88 latents of Good category, 85 latents of Bad category and 85 latents of Ugly category were searched in the entire set of 258 tenprints in the NIST-SD27 database. We report the rank identification accuracy for two publicly available minutiae based matchers, namely NIST-Bozorth3 [NIST-NBIS, NBIS-Release v4.2.0] and Minutia Cylinder-Code (MCC) SDK [MCC, MCC-SDK v1.4] [Cappelli *et al.*, 2010] [Cappelli *et al.*, 2011] [Ferrara *et al.*, 2012] before and after incorporating our proposed hierarchical registration algorithm as a pre-registration before the identification.

When reporting the rank identification accuracies, for Good quality, there are 88 match scores and 88×257 non-match scores, for Bad and Ugly qualities, there are 85 match scores and 85×257 non-match scores respectively. When we report the rank identification accuracy for the entire NIST-SD27 database (All category), then there are 258 match scores and 258×257 non-match scores.

NIST-Bozorth3 is a minutiae based fingerprint matcher that is specially developed to deal with latent fingerprints. This matcher is part of the NIST Biometric Image Software (NBIS) [NIST-NBIS, NBIS-Release v4.2.0], developed by NIST. MCC-SDK is a well known minutiae matcher more adapted to good quality fingerprints with reasonable number of minutiae in both query and reference templates. Both NIST-Bozorth3 and MCC-SDK are publicly available. We show the performance accuracy of the matcher using Cumulative Match Characteristic (CMC) curves.

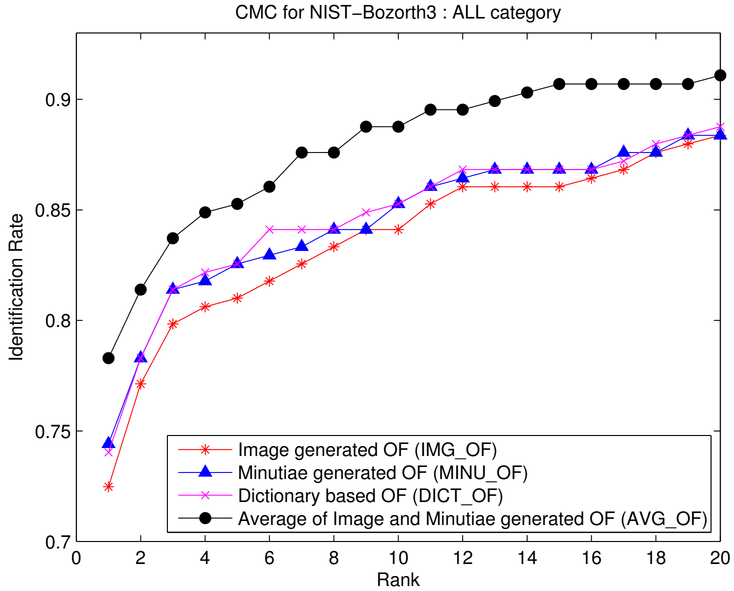


Figure 3.6: *CMC curve showing the rank identification rate of NIST-Bozorth3 for NIST-SD27 when different types of OF estimation techniques were used for the tenprints, and $MANUAL\_OF$ for latents, when applying the proposed OF-based pre-alignment.*

3.4.1. Experiment 1: Choosing the best orientation field for tenprints

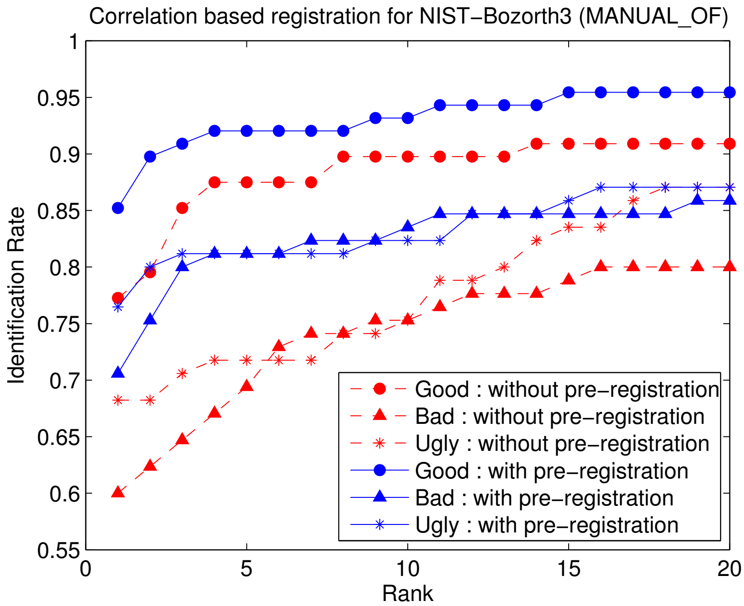
Fig. 3.6 shows the CMC curve of the NIST-Bozorth3 matcher when using $MANUAL\_OF$ for latent against various other OF estimation techniques for tenprints while performing pre-registration using our proposed hierarchical method. We can observe that the rank identification accuracy has a consistent improvement when AVG_OF is used for tenprints. The improvement while using AVG_OF is mainly because the image noise introduced in the estimation of IMG_OF is reduced while averaging with $MINU\_OF$.

Based on this result, we have chosen AVG_OF as the orientation field for tenprints in remaining experiments reported here.

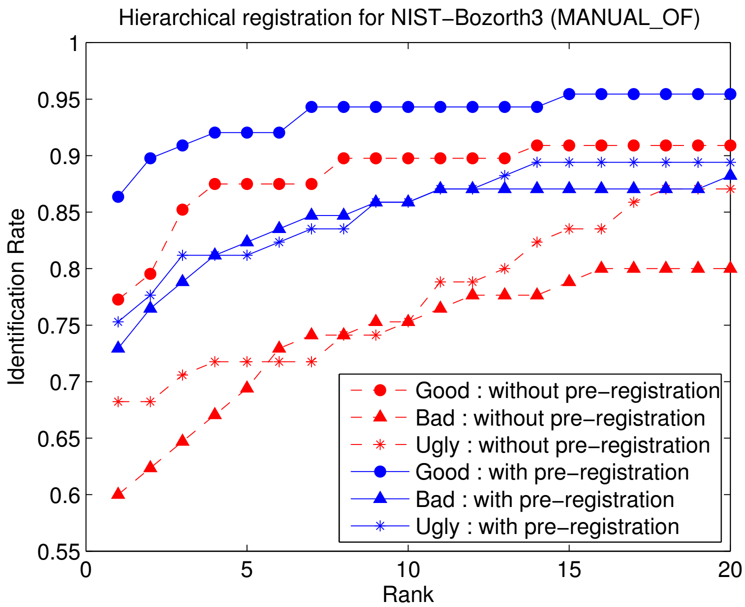
3.4.2. Experiment 2: Pre-Registration

In this experiment, we perform pre-registration using our registration algorithm to reduce the minutiae search space of the tenprint minutiae set, and then use the reduced minutiae set template as the reference template for the matcher. We used NIST-Bozorth3 and MCC-SDK as the minutiae-based matchers.

For latents, $MANUAL\_OF$ and $DICT\_OF$ were used, and for the tenprints, we used AVG_OF to report the rank identification accuracies in this experiment. We also analyze

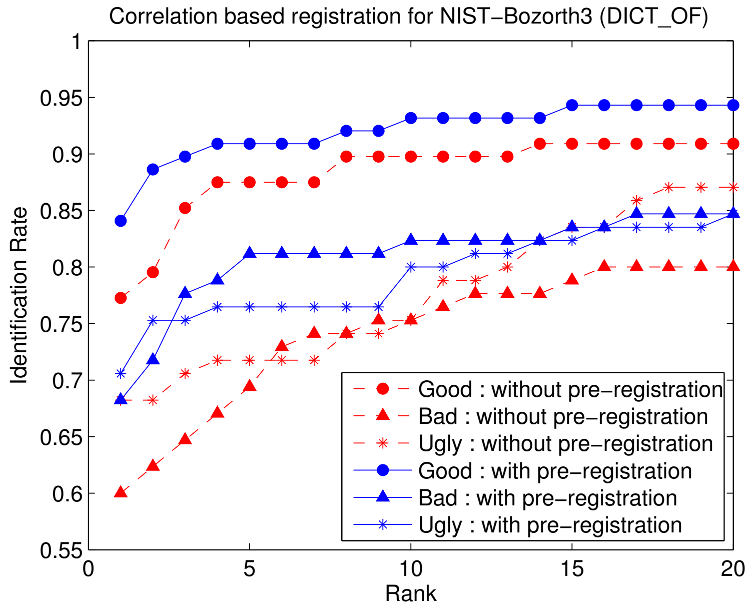


(a) Correlation only based registration (Level 1)

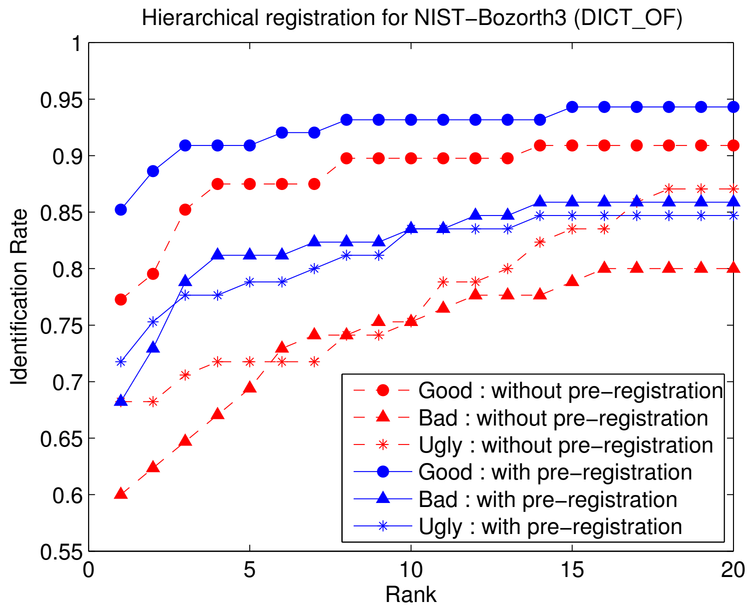


(b) Hierarchical registration (Level 2)

Figure 3.7: *Performance of NIST-Bozorth3 when using MANUAL_OF for latents*



(a) Correlation only based registration (Level 1)



(b) Hierarchical registration (Level 2)

Figure 3.8: *Performance of NIST-Bozorth3 when using DICT_OF for latents*

Quality	Bozorth3 DIRECT(%)	Bozorth3 L1(%)	Bozorth3 L2(%)
All	68.60	77.52	78.29
Good	77.27	85.23	86.36
Bad	60.00	70.59	72.94
Ugly	68.24	76.47	75.29

Table 3.2: *Rank-1 identification for NIST-Bozorth3 with correlation based pre-registration and hierarchical registration when $MANUAL\_OF$ is used for latents.*

Quality	Bozorth3 DIRECT(%)	Bozorth3 L1(%)	Bozorth3 L2(%)
All	68.60	74.42	75.19
Good	77.27	84.09	85.23
Bad	60.00	68.24	68.24
Ugly	68.24	70.59	71.76

Table 3.3: *Rank-1 identification for NIST-Bozorth3 with correlation based pre-registration and hierarchical registration when $DICT\_OF$ is used for latents.*

separately the performance of the matcher using correlation only based registration and using hierarchical registration.

3.4.2.1. NIST-Bozorth3

Fig. 3.7 and Fig. 3.8 show the CMC curve of NIST-Bozorth3 for two different registration levels when $MANUAL\_OF$ and $DICT\_OF$ is used for latents respectively.

Fig. 3.7(a) shows the rank identification accuracy of NIST-Bozorth3 when correlation based registration (Level 1) of our algorithm is used as pre-registration, and also without using pre-registration ($MANUAL\_OF$ for latents). We see a significant and consistent improvement in the rank identification accuracy for all the quality categories when incorporating the proposed pre-registration.

Fig. 3.7(b) shows the rank identification accuracy of NIST-Bozorth3 when hierarchical registration (Level 2) of our algorithm is used as pre-registration with $MANUAL\_OF$ for latents. Here, we notice a consistent improvement in the CMC curve for all subjective quality categories compared to the correlation based registration. Especially there is a significant improvement for both Bad and Ugly quality categories.

Table 3.2 summarizes the Rank-1 identification accuracy of NIST-Bozorth3 for both correlation based registration and hierarchical registration when $MANUAL\_OF$ is used for latents. The column $DIRECT$ represents the Rank-1 identification accuracy of NIST-Bozorth3 when no pre-registration is applied to the minutiae set. Column $L1$ and $L2$ represent the Rank-1 identification accuracy for correlation based registration (Level 1) and hierarchical based registration (Level 2) respectively.

Similarly, Fig. 3.8(a) and Fig. 3.8(b) shows the rank identification accuracy of NIST-Bozorth3 when correlation based pre-registration and hierarchical pre-registration were applied using $DICT\_OF$ for the latents. Table 3.3 summarizes the Rank-1 identification accuracy in this case. Similar results compared to using $MANUAL\_OF$ for the latents are also obtained here when considering $DICT\_OF$. This proves the robustness of the $DICT\_OF$ method for obtaining a reliable OF even with very difficult latents and the feasibility of our method as a fully automatic tool.

3.4.2.2. MCC-SDK

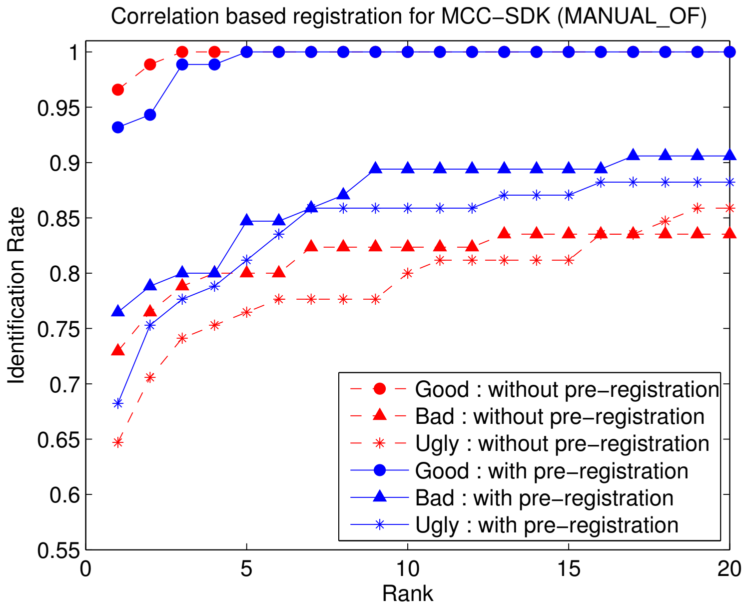
Fig. 3.9 shows the CMC curve of MCC-SDK for the two registration levels considered when $MANUAL\_OF$ is used for latents. Fig. 3.9(a) and Fig. 3.9(b) show the rank identification accuracy of MCC-SDK when correlation based pre-registration and hierarchical pre-registration were applied respectively. Table 3.4 summarizes the Rank-1 identification accuracy in this case. The overall Rank-1 accuracy improved from 78.3% to 79.4% when incorporating Level 1 pre-registration, and to 79.4% when hierarchical based pre-registration (Level 2) is incorporated. Even though the improvement is small, it is consistent and increases for Bad and Ugly quality categories when we look beyond Rank-1.

Quality	MCC-SDK DIRECT(%)	MCC-SDK with L1(%)	MCC-SDK L2(%)
All	78.29	79.46	79.46
Good	96.59	93.18	97.73
Bad	72.94	76.47	75.29
Ugly	64.71	68.24	64.71

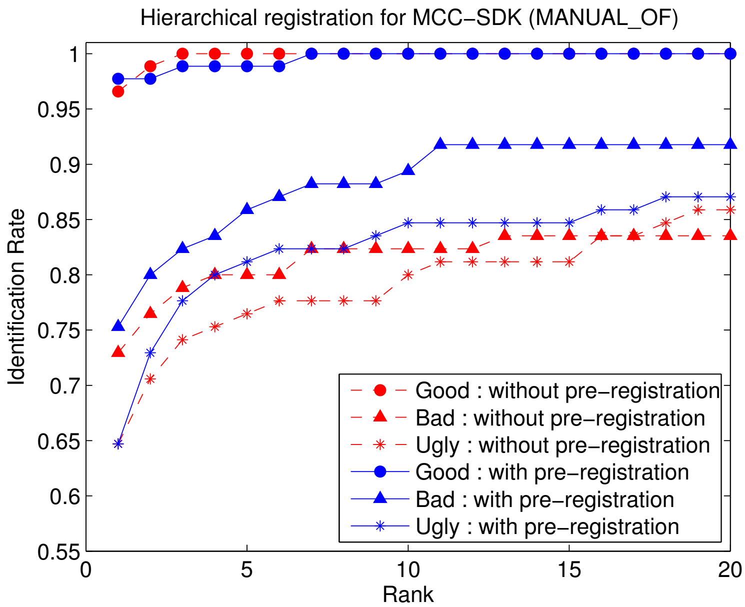
Table 3.4: *Rank-1 identification for MCC-SDK with correlation based pre-registration and hierarchical registration when $MANUAL\_OF$ is used for latents.*

3.4.3. Experiment 3: Parameters - Rotation step size, Radius

In this experiment, we study the quantization step size for rotation alignment (Step 4 in Algorithm) as well as the best radius of the circular region (Step 10 in Algorithm) to generate the subset of minutiae from the tenprint minutiae set. We used $MANUAL\_OF$ for the latents,

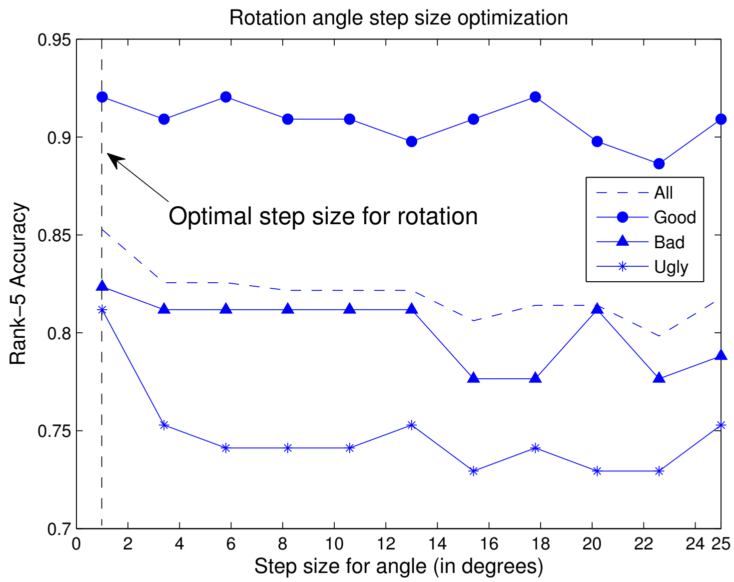


(a) Correlation only based registration (Level 1)

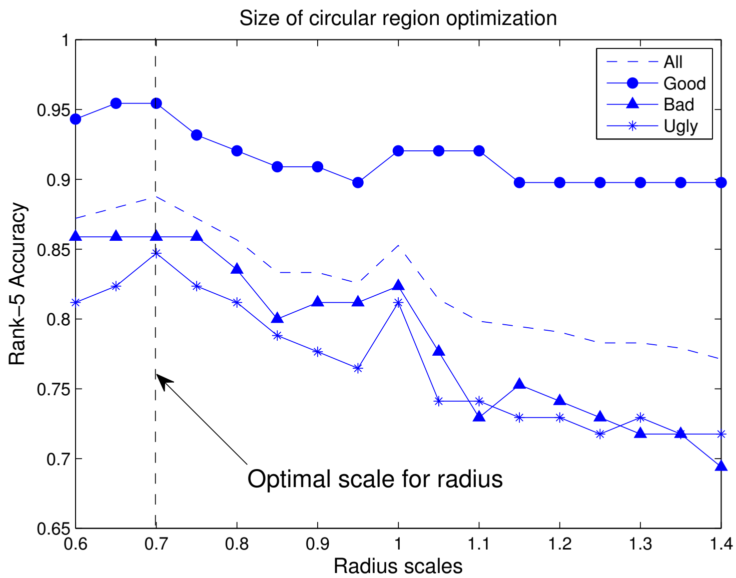


(b) Hierarchical registration (Level 2)

Figure 3.9: *Performance of MCC-SDK when using MANUAL_OF for latents*



(a) Change in Rank-5 accuracies when increasing the step size in the range 1° to 25°



(b) Change in Rank-5 accuracies when changing the scale factor from 0.6 to 1.4

Figure 3.10: *Finding the optimal value for rotation step size and radius scales using NIST-Bozorth3 matcher.*

AVG_OF for tenprints and performed hierarchical registration on NIST-Bozorth3 matcher.

From Fig. 3.10(a) we can observe that when we use a step size (X-axis) for the rotation equal to 1°, we obtain the best performance in terms of rank identification accuracy (Y-axis). We looked at the Rank-5 identified accuracy of the NIST-SD27 database (All category) to evaluate the performance, and looked at the step size varying from 1° to 25°. Also interestingly, the performance is not very much degraded with large steps, which can justify the use of large steps in some scenarios when computation speed is prioritized.

With 1° as the step size, we studied the effect of the radius of the circular region. We observe that the optimal radius is obtained by using a scale factor of 0.7 on half the length of the diagonal of bounding box. Fig. 3.10(b) shows the Rank-5 accuracy for various scales of the radius ranging from 0.6 to 1.4 scale factor in X-axis and the corresponding Rank-5 accuracy in Y-axis.

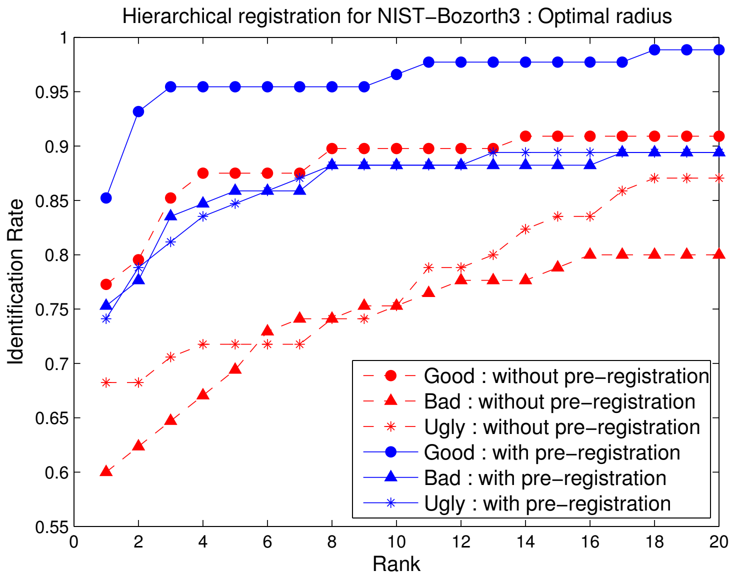
3.4.4. Experiment 4: Best result obtained

With the optimal parameters estimated from our experiments, we have obtained the best performance boost for the matchers when using the hierarchical registration as a pre-registration. Fig. 3.11(a) and Fig. 3.11(b) shows the CMC curve for both NIST-Bozorth3 and MCC-SDK with the optimal parameters for the hierarchical pre-registration. $MANUAL\_OF$ was used for latents and AVG_OF was used for tenprints. Table 3.5 summarizes the Rank-1 identification accuracy of NIST-Bozorth3 and MCC-SDK for the optimal parameters (rotation step size with 1° and radius scale factor of 0.7).

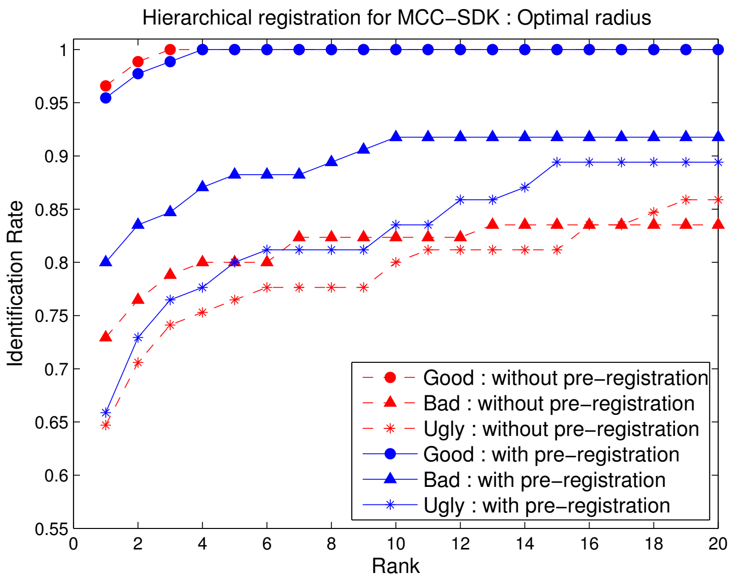
Quality	NIST-Bozorth3 DIRECT(%)	NIST-Bozorth3 with L2(%)	MCC-SDK DIRECT(%)	MCC-SDK with L2(%)
All	68.60	78.29	78.29	80.62
Good	77.27	85.23	96.59	95.45
Bad	60.00	75.29	72.84	80.00
Ugly	68.24	74.12	64.71	65.88

Table 3.5: *Rank-1 identification for NIST-Bozorth3 and MCC-SDK with optimal parameters.*

Using our registration algorithm as a pre-registration, we were able to boost the overall Rank-1 identification accuracy from 68.60% to 78.29% for NIST-Bozorth3, and from 78.29% to 80.62% for MCC-SDK. In other regions of the CMC curve the improvement is even higher.



(a) NIST-Bozorth3 : Hierarchical registration with optimal parameters



(b) MCC-SDK : Hierarchical registration with optimal parameters

Figure 3.11: *CMC curve of NIST-Bozorth3 and MCC-SDK with the optimal parameters.*

3.4.5. Experiment 5: Runtime analysis

We have implemented the proposed hierarchical registration algorithm in MATLAB which is not an optimized version to be directly compared with that of a corresponding C/C++ implementation. Nevertheless, we summarize the average runtime of the MATLAB version for each subjective quality category in Table 3.6.

Quality	Average runtime in milliseconds (ms)
Good	921
Bad	792
Ugly	707

Table 3.6: *Average runtime for each subjective quality category.*

We assume that the minutiae extraction and computation of AVG_OF are pre-computed offline, and they need to be generated only once for the reference fingerprints in the database.

In our MATLAB implementation, we used filter2() function to obtain the correlations mentioned in Step 5 of Algorithm. If the size of the region of interest for the input latent is large, then it will be advantageous to perform the correlation in frequency domain using Fast Fourier Transform (FFT) implementations where correlation reduces to multiplication, and then obtain the inverse FFT to get the equivalent of correlation in spatial domain.

3.5. Discussions

We have proposed an orientation field based registration algorithm for partial fingerprints. When we use our hierarchical registration algorithm as a pre-registration stage and reduce the search space of minutiae in the tenprint minutiae set, we were able to significantly boost the performance of two popular minutiae matchers using challenging and realistic data. The main objective of our research was to improve the rank identification accuracy for poor quality latents. We were able to obtain consistent and significant improvement for both Bad and Ugly quality category of latents from NIST-SD27.

Upon studying various orientation field estimation techniques for fingerprints to be used in our registration, we have noticed that the best representative orientation field for tenprints was obtained by averaging a gradient based orientation field estimated from the fingerprint image and the orientation field reconstructed from the minutiae set. This gave the best performance mainly due to noise reduction while averaging. For latents, we studied two types of orientation fields corresponding to two different scenarios: with manual intervention and fully automated procedure. We obtained the best performance while using manually extracted orientation field for latents, and also a significant improvement with automated dictionary-based orientation field estimation.

We have observed that if the region of interest is very small in the latent fingerprint, especially in Bad and Ugly quality category, the registration accuracy is slightly degraded while using the hierarchical method compared to correlation-based registration. This accounts for a slight variation in the Rank-1 performances between L1 and L2. Since we are not using our own minutiae matcher, but using standard ones, it will be difficult to give a theoretical justification on the behavior for Rank-1 identification between L1 and L2, especially for Bad and Ugly categories. Anyway on an average, we observe that the hierarchical method significantly improves the rank identification accuracy.

We also observed that for a large quantization step in the rotation alignment, we have not degraded the performance very much, and while matching, we have reduced the size of the minutiae search space in the tenprint to good extent which accounts for overall efficiency of our proposed method. Also, we have established the feasibility of our method as a fully automatic tool.

Part II

Extended Feature Sets and Evidence Evaluation from AFIS Scores

Chapter 4

Related works and motivation for incorporating EFS

This chapter discusses about the importance of Extended Feature Set (EFS) towards improving the identification accuracies of automated minutiae-based fingerprint matchers, and a brief overview of our proposed method to use EFS in improving the minutiae-based matchers that are not already adapted to use EFS. We review the limitations of currently available latent-AFIS in Lights-Out mode which do not use all the discriminatory features that can be obtained from fingerprint. Current practices in using latent-AFIS involves human intervention in terms of marking the features manually for the latent fingerprints. This procedure of manual marking of latent fingerprint features and then using the latent-AFIS for identification is termed as Semi Lights-Out mode. We give a brief overview on the results of the public evaluation of commercial latent-AFIS conducted by NIST in both Lights-Out mode (ELFT) and in Semi Lights-Out mode (ELFT-EFS), and the conclusions derived from these evaluations.

We discuss the major problems cited by the forensic community in the friction ridge analysis procedures currently followed. The steps taken by the Scientific Working Group on Friction Ridge Analysis, Study, and Technology (SWGFAST) towards resolving some of the issues noted by the forensic community, and the setting up of a Committee to Define an Extended Fingerprint Feature Set (CDEFFS) by ANSI/NIST are briefly discussed. We then briefly discuss some of the extended features defined by CDEFFS under various fingerprint feature category levels, and the type of extended features that are used in our study. The database used in our experiments to establish the usefulness of our proposed algorithm based on extended features is obtained from Guardia Civil, the Spanish law enforcement agency. We describe in detail the Guardia Civil database (GCDB), the rare minutiae found in GCDB and their statistics in Chapter 5. A review on related works which makes use of EFS to improve the identification accuracies of automated matchers and the conclusions derived are briefly explained here in Chapter 4. The detailed description of our proposed algorithm which incorporates extended features (rare minutiae), and various experiments establishing the usefulness of our proposed algorithm are provided in Chapter 5.

4.1. Lights-Out system evaluation: NIST-ELFT

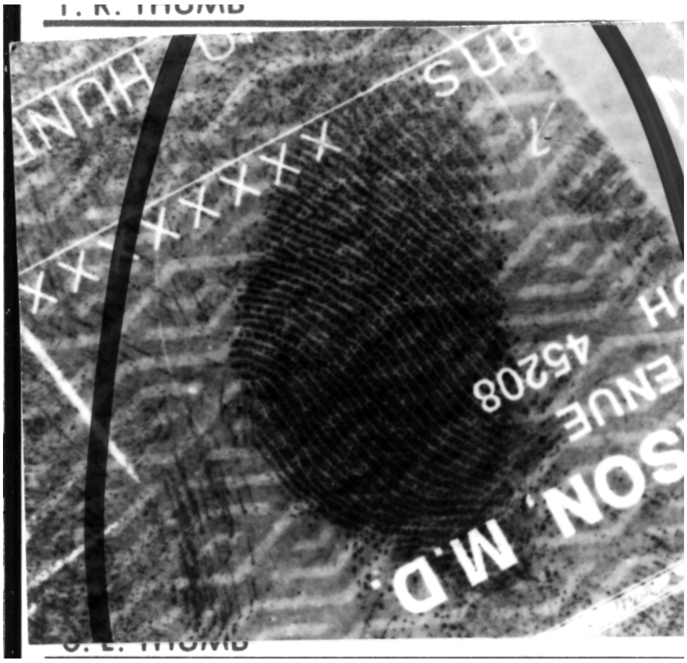
A common forensic evidence used in criminal investigations is latent fingerprint, but comparing latent fingerprints is not an easy task. This is mainly attributed to the poor quality of the latent fingerprints obtained from the crime scenes. When a latent fingerprint is found, the criminal investigators first search for the suspect using an Automated Fingerprint Identification System (AFIS) to narrow down their manual work. If there is a match, then the individual is linked to the crime under investigation. Individualization (*identification or match*) is the decision yielded by a forensic examiner about the latent fingerprint belonging to a particular individual. This is the outcome of the Analysis, Comparison, Evaluation and Verification (ACE-V) [Ashbaugh, 1999] methodology currently followed in friction ridge examination.

In order to improve the matching efficiency, the concept of "Lights-Out System" was introduced for latent matching [Dvornychenko and Garrism. M, 2006]. A Lights-Out System is a fully automatic identification process with no human intervention. Here, the system should automatically extract the features from the latent fingerprint and match it against the tenprints (exemplars) stored in the AFIS database to obtain a set of possible suspects with high degree of confidence. In general, latent fingerprints are partial in nature and are of varying quality (see Figure 4.1), mostly distorted, smudgy, blurred etc. These factors lead to high number of unreliable extracted features in fully automatic mode, and make it difficult for AFIS to perform well.

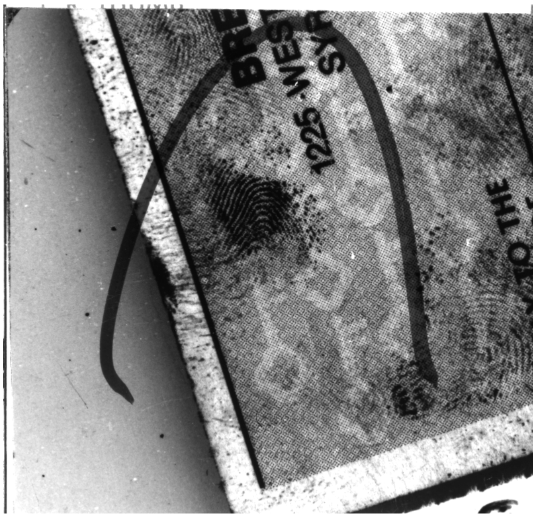
AFIS do not use all the discriminatory features that could be derived from a fingerprint, mainly due to the limitations of automatic and reliable extraction of all types of discriminatory features. The accurate performance of feature extraction and matching algorithms of AFIS in forensic scenario is of great importance to avoid erroneous individualization. NIST conducted a multi-phase public evaluation of latent AFIS called as *Evaluation of Latent Fingerprint Technologies* (ELFT) [NIST-ELFT, 2013]. A discussion about the conclusions derived from these public evaluations are briefed in Section 1.10.2, and the best results achieved for various phases are summarized in Table 1.1.

4.2. Standardizing extended fingerprint features

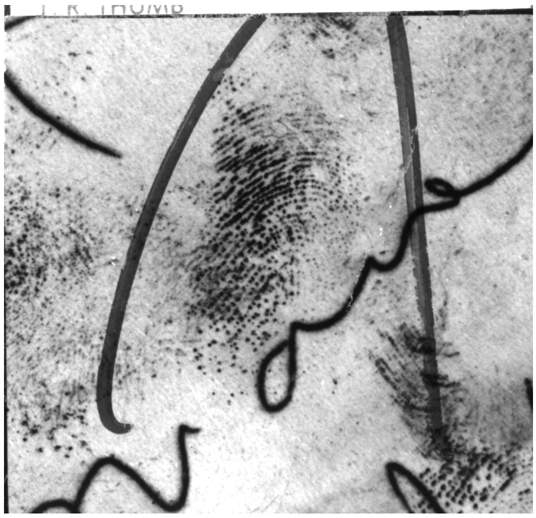
AFIS use only a limited number of features automatically extracted from the fingerprints using a feature extraction algorithm. On the other hand, forensic examiners use a richer set of features during their manual comparisons. This could be a possible reason why manual comparisons outperform AFIS comparisons [Jain, 2010]. Any features that are not currently used by commercial AFIS are generally termed as Extended Feature Set (EFS) [Zhao and Jain, 2010]. The use of EFS by forensic examiners in manual comparisons is much debated, mainly due to non-repeatability by another examiner to validate the previous decision.



(a) Good



(b) Bad



(c) Ugly

Figure 4.1: *Subjective quality classification of latent fingerprint images in NIST Special Database 27 (NIST-SD27).*

Type of category	Extended Feature Set
Level-One Details	Ridge flow map, local ridge quality, pattern classification (whorl, arch, tentarch, left/right loop etc), singular points (core, delta), core-delta ridge count.
Level-Two Details	Minutiae-ridge relationship, ridge curvature, feature relationship, unusual/rare minutiae, scars, creases, incipient ridges, dots.
Level-Three Details	Pores, edge shapes, ridge/furrow width.

Table 4.1: *Extended features defined by CDEFFS categorized into respective fingerprint feature details (not a comprehensive list).*

Two major problems in friction ridge analysis as reported in [Hicklin, 2007] are as follows:

1. Latent AFIS searches are limited by an over simplified feature set.
2. During the latent examiner comparison, there are no standard format to document the features used in comparison decision. This leads to problems with future reference or interchange with other forensic examiners.

The SWGFAST (Scientific Working Group on Friction Ridge Analysis, Study, and Technology) drafted a memo to NIST noting that forensic examiners use features that are not currently addressed in fingerprint standards. The ANSI/NIST Standard Workshop II charted the Committee to Define an Extended Fingerprint Feature Set (CDEFFS). The CDEFFS included 45 members from various federal agencies, the forensic community, AFIS vendors, and academia [Hicklin, 2007]. The purpose of CDEFFS was to define a standard to completely represent the distinctive information in the fingerprint which are quantifiable, repeatable and develop a clear method of characterizing information for: 1) AFIS searches initiated by forensic examiner, and 2) forensic examiner markup and exchange of latent fingerprints.

Fingerprint features are categorized into three levels as well as a feature category called "other" to be used for friction ridge examination. Level-One considers general overall direction of the ridge flow. Level-Two describes the path of specific ridges. Level-Three are the shapes of the ridge structure. "Other" features describe temporary features or imperfections in ridges [Holder *et al.*, 2011]. Some of the extended fingerprint features defined by CDEFFS under each of the three level categories [Hicklin, 2007] [Hicklin, 2005] are summarized in Table 4.1. Figure 4.2 show some general extended features and Figure 4.3 show some rare minutia features and typical

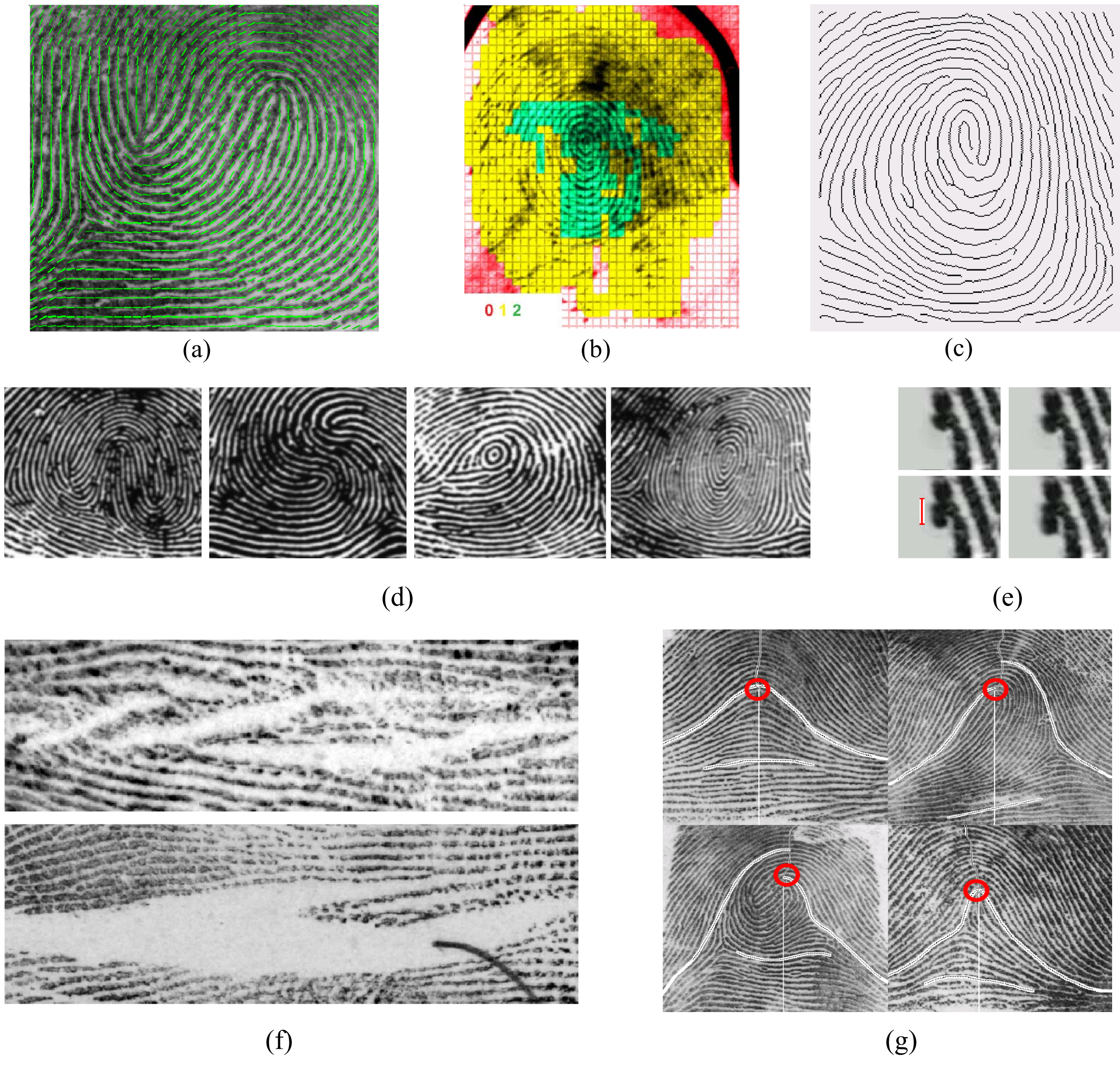


Figure 4.2: *Some example Extended Feature sets [Hicklin, 2007]: (a) ridge flow maps, (b) local ridge quality, (c) ridge path, (d) pattern classifications, (e) protrusions, (f) flexion creases, (g) center point of reference.*

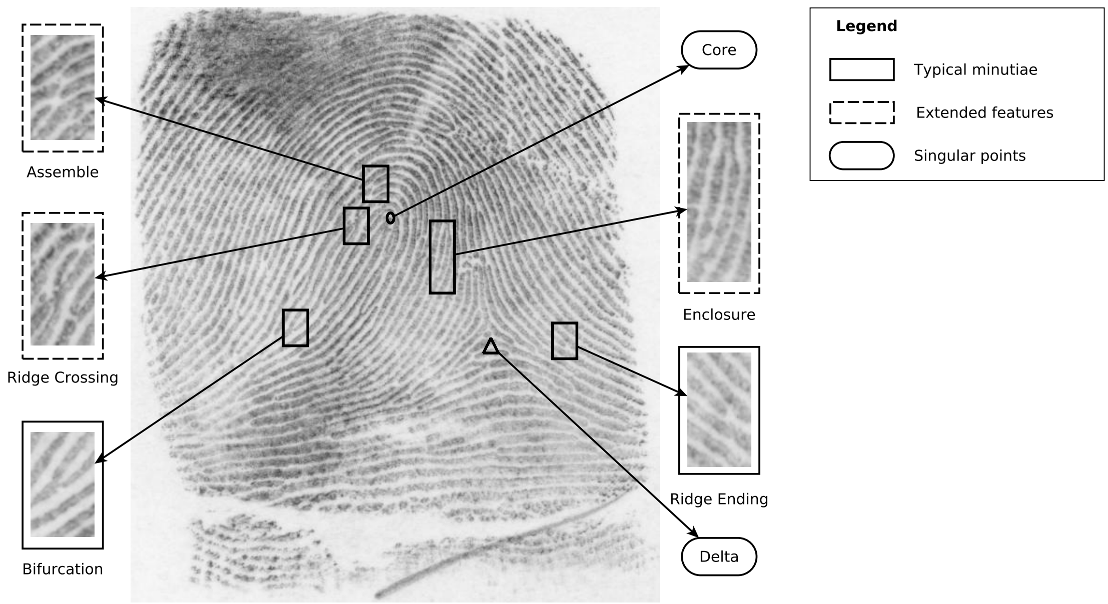


Figure 4.3: *Typical minutiae (ridge-endings, bifurcations) and rare minutia features (assemble, ridge crossing, enclosure) in an exemplar fingerprint from NIST-SD27 database.*

minutiae features (ridge-endings and bifurcations) in an exemplar fingerprint from NIST Special Database 27 (NIST-SD27).

4.3. Semi Lights-Out system evaluation: NIST-ELFT-EFS

To use EFS in automated systems, reliable feature extraction algorithms are mandatory. Many law enforcement agencies follow a 500 ppi scanning resolution for fingerprint images to be used in AFIS. With such a resolution, it is difficult to reliably extract many of the extended features automatically. Due to advances in fingerprint scanning technologies, SWGFAST during the ANSI/NIST Fingerprint Standard Update Workshop in 2005 proposed 1000 ppi as the minimum scanning resolution for fingerprint images. This proposal was hugely supported by the forensic community. To test the feasibility of including EFS in latent AFIS, NIST conducted another multi-phase commercial latent algorithm evaluation called Evaluation of Latent Fingerprint Technologies - Extended Feature Sets (ELFT-EFS) [Indovina *et al.*, 2011b].

ELFT-EFS was conducted in a "Semi Lights-Out" mode as compared to the "Lights-Out" mode for ELFT. The main purpose of ELFT-EFS was to determine the effectiveness of forensic examiner marked latent fingerprint features on the latent identification accuracy. NIST conducted two evaluations for ELFT-EFS and the best achieved Rank-1 identification accuracy for each of the evaluations is summarized in Table 4.2.

ELFT-EFS	Database size	Rank-1 accuracy
Evaluation-1 [Indovina *et al.*, 2011b]	1,114 latents compared against 1,000,000 rolled prints and 1,000,000 plain prints	66.7%
Evaluation-2 [Indovina *et al.*, 2012b] [Indovina *et al.*, 2012c]	1,066 latents compared against 1,000,000 rolled prints and 1,000,000 plain prints	71.4%

Table 4.2: *Rank-1 identification accuracy of NIST Evaluation of Latent Fingerprint Technologies - Extended Feature Sets (ELFT-EFS).*

As in ELFT, the results of different evaluations in ELFT-EFS cannot be directly compared because the database used were not exactly the same [Indovina *et al.*, 2012b] [Indovina *et al.*, 2011b]. In [Indovina *et al.*, 2012b], it is reported that though the highest measured accuracy achieved by a individual matcher at Rank-1 was 71.4%, and approximately 82% of the latents were correctly matched at Rank-1 when more matchers were combined. This corroborates the potential for additional accuracy improvement when combining multiple algorithms [Fierrez-Aguilar *et al.*, 2006].

4.4. Overview and main contributions

In this work, we propose a method to improve the identification accuracy of minutiae-based matchers for partial latent fingerprints by incorporating reliably extracted rare minutiae features. Most minutiae-based fingerprint matchers use only two prominent ridge characteristics namely *ridge-endings* and *bifurcations*. We propose an algorithm that will modify the similarity scores of minutiae-based matchers based on the presence of rare minutia features like *fragments*, *enclosures*, *dots*, *interruptions*, *etc*. The weights that we use to modify the similarity scores are obtained based on the probability of occurrence of such rare minutiae features. The decision for a match or non-match is automatically estimated based on least squares fitting of an affine transformation between the latent minutiae set and the tenprint minutiae set. We show a significant improvement in the overall rank identification accuracies for two minutiae-based matchers (NIST-Bozorth3 and VeriFinger) when their similarity scores are modified using our proposed algorithm which incorporates rare minutiae features.

The main contributions of this work are as follows:

1. A methodology to adapt any minutiae-based matcher by incorporating information from rare features.

2. A specific algorithm to align the latent minutiae pattern and the tenprint minutiae pattern using rare minutiae.

3. Experimental demonstration of the performance improvement of minutiae-based matchers when incorporating information from rare features.

4. We finally present also various population statistics about rare minutiae features present in a realistic forensic casework database obtained from Spanish law enforcement agency (Guardia Civil).

In the following sections, we review related works in the use of extended features of fingerprints towards improving the rank identification of automated matchers. In next chapter, we explain the proposed algorithm to modify the similarity scores of minutiae-based matchers based on rare minutiae features, experiments, results and discussions.

4.5. Related works

An extensive study on extended fingerprint feature sets is reported by Jain [2010]. This includes several extended features from Level-One, Level-Two and Level-Three. It was concluded in [Jain, 2010] that manual intervention is strongly recommended while using EFS, as well as extended features from Level-One and Level-Two are highly recommended to be incorporated in latent AFIS. Extended features such as ridge flow map, ridge wavelength map, ridge quality map, and ridge skeleton have shown significant improvements in latent identification accuracies. Level-One and Level-Two details used in [Jain, 2010] [Jain and Feng, 2011] are insensitive to image quality, and do not rely on high resolution images. To incorporate Level-Three EFS such as pores, dots, incipients, etc, it is essential to improve the quality of enrolled fingerprints.

The use of pores as extended features was studied in high resolution 1000 ppi images by Zhao and Jain [2010] and Jain *et al.* [2007b]. Dots and incipients were studied by Chen and Jain [2007]. Out of pores, dots and incipients, pores resulted in better performance [Zhao and Jain, 2010]. Even though high resolution 1000 ppi images were used, live scan images resulted in easy detection of pores automatically, which was not the case with inked fingerprint images. Pore extraction based on skeletonized and binary images was studied by Stosz and Alyea [1994] and Kryszczuk *et al.* [2004]. These techniques were demonstrated effective only on very good quality high resolution fingerprint images scanned approximately at 2000 ppi [Stosz and Alyea, 1994]. These methods were more sensitive to noise, and the performance degrades for poor quality of fingerprint images and low resolution images.

A local image quality based method applied on extended fingerprint features for high resolution 1000 ppi fingerprint images was reported by Vatsa *et al.* [2008]. A fast curve evolution algorithm was used to quickly extract extended features such as pores, ridge contours, dots and incipient ridges. Together with other Level-One and Level-Two details as proposed in Jain and Feng [2011], these extended features were used to generate a quality-based likelihood ratio to improve the identification performance.

Score level fusion of different algorithms using various extended fingerprint features was report by Fierrez *et al.* [2005]. Features like singular points, ridge skeleton, ridge counts, ridge flow map, ridge wavelength map, texture measures were studied by analyzing the correlation between them using feature subset-selection techniques. Combination of features show significant improvement in the performance of the system.

Chapter 5

Integrating EFS in generic fingerprint matchers: Database, algorithm and experiments

This chapter describes in detail the database acquired and processed (GCDB) within the present PhD research, and the proposed algorithm to improve the identification accuracy of minutiae-based matchers for partial latent fingerprints by incorporating reliably extracted rare minutiae features. The improvement in the identification accuracy for matchers are achieved by modifying the similarity scores of matcher based on the decision yielded by our algorithm. The decision for a match or non-match is automatically estimated based on least squares fitting error of an affine transformation that transforms latent minutiae set onto tenprint minutiae set with the rare minutiae as the reference point. The proposed method is accomplished through two stages. In the first stage, we estimate the fitting error, and in the second stage, the similarity score of the matcher is modified. We show a significant improvement in the rank identification accuracies of two minutiae-based matchers namely, NIST-Bozorth3 and VeriFinger-SDK. The chapter concludes with a discussion summarizing the usefulness of our proposed algorithm.

5.1. Database : Guardia Civil database

The database used in this work was obtained from Guardia Civil, the Spanish law enforcement agency. The Guardia Civil database (GCDB) is a realistic forensic fingerprint casework database acquired and processed within the present PhD research. Apart from having typical minutiae feature types (*ridge-endings, bifurcations*), GCDB also comprises rare minutiae types like *fragments*, *enclosures*, *dots*, *interruptions, etc* [Santamaria, 1955]. A comprehensive list of rare minutiae features used by Guardia Civil are shown in Figure 5.1 and the corresponding minutiae type names are listed in Table 5.1.

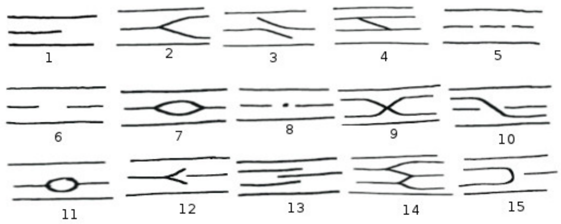


Figure 5.1: *Minutiae types used by Guardia Civil. Names corresponding to individual minutiae type numbers can be found in Table 5.1.*

No	Minutiae type	No	Minutiae type	No	Minutiae type
1	Ridge Ending	6	Interruption	11	Circle
2	Bifurcation	7	Enclosure	12	Delta
3	Deviation	8	Point	13	Assemble
4	Bridge	9	Ridge Crossing	14	M-structure
5	Fragment	10	Transversal	15	Return

Table 5.1: *List of minutiae types used by Guardia Civil. Numbering with respect to Figure 5.1.*

GCDB used in this work consists of 268 latent and tenprint (exemplar) pairs of fingerprint images and minutiae sets. All the minutiae in the latent fingerprint images were manually extracted by forensic examiners of Guardia Civil. The corresponding mated minutiae in the tenprints were also manually established. This includes the typical (ridge-endings and bifurcations) minutiae and the rare minutiae. These are called *matched* minutiae set, i.e, the minutiae sets for which a one-to-one correspondence is established between the latent and the mated tenprint. Here, the number of minutiae in the latent and the corresponding mated tenprint minutiae set are the same.

To generate an *ideal* minutiae set (i.e, all possible minutiae) for the tenprint, we used the minutiae extractor module from VeriFinger SDK [Neurotec-Biometric-4.3]. We performed a Gabor filtering based post-processing to remove any spurious minutiae that are outside the region of interest (ROI). This post-process was needed because the quality of the tenprints in GCDB were not good in most of the cases, and VeriFinger couldn't perform a proper segmentation of the fingerprint region of interest by itself. So, spurious minutiae were generated which lies outside of ROI.

The algorithm to estimate the ROI is outlined as follows:

1. The fingerprint image I (Figure 5.2(a)) is first normalized as proposed by Hong *et al.* [1998]:

$$I'[x,y] = \begin{cases} m_0 + \sqrt{(I[x,y]-m)^2 \cdot v_0/v} & \text{if } I[x,y] > m \\ m_0 - \sqrt{(I[x,y]-m)^2 \cdot v_0/v} & \text{otherwise,} \end{cases} \tag{5.1}$$

   where m and v are the image mean and variance and m_0 and v_0 are desired mean and variance after the normalization. The normalized image I' is shown in Figure 5.2(b).

   In our implementation, we used $m_0 = 100$ and $v_0 = 128$ as the desired mean and variance to obtain I'.

2. Eight different Gabor filter responses for the normalized fingerprint image I' are generated by varying the orientation parameter of Gabor filter. We used the even symmetric two-dimensional Gabor filter defined as follows [Maltoni *et al.*, 2009]:

$$g(x,y:\theta,f) = \exp\left\{-\frac{1}{2}\left[\frac{x_\theta^2}{\sigma_x^2} + \frac{y_\theta^2}{\sigma_y^2}\right]\right\} \cdot \cos(2\pi f \cdot x_\theta) \tag{5.2}$$

   where θ is the orientation of the filter, and $[x_\theta, y_\theta]$ are the coordinates of $[x,y]$ after a clockwise rotation of the Cartesian axes by an angle of $(90° - \theta)$.

$$\begin{bmatrix} x_\theta \\ y_\theta \end{bmatrix} = \begin{bmatrix} \cos(90° - \theta) & \sin(90° - \theta) \\ -\sin(90° - \theta) & \cos(90° - \theta) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \sin\theta & \cos\theta \\ -\cos\theta & \sin\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{5.3}$$

   There are four parameters for Gabor filter. θ, the orientation of filter, f, the frequency of the filter, σ_x and σ_y are the standard deviations of the Gaussian envelope along the x and y axes respectively. In our experiments we fixed the parameters $f = 1/8$, $\sigma_x = \sigma_y = 4$, and discrete orientations $\theta = \{0°, 22.5°, 45°, 67.5°, 90°, 112.5°, 135°, 157.5°\}$.

3. I'_i is the Gabor response of the image I' for the orientation θ_i. Figure 5.2(c)-(j) shows the Gabor responses for θ_i. Generate the mean image I'_{mean} as follows:

$$I'_{mean} = \frac{1}{8}\sum_{i=1}^{8} I'_i \tag{5.4}$$

   Figure 5.2(k) shows the mean image I'_{mean} of all the Gabor responses.

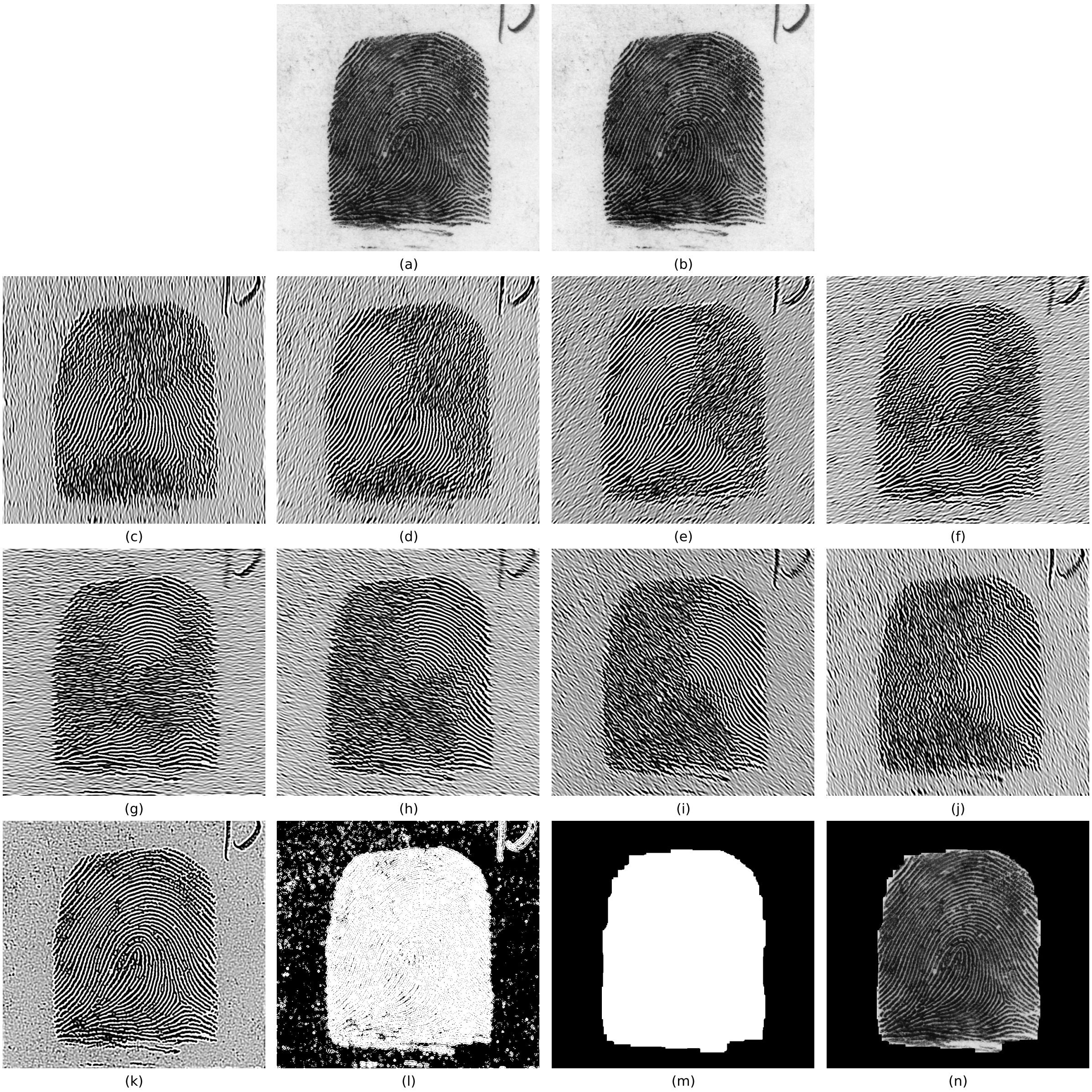


Figure 5.2: *The various stages involved in generating the region of interest using Gabor filter based segmentation. (a) original fingerprint image, (b) histogram normalized, (c)-(j) Gabor responses for eight different orientations (*$0°$, $22.5°$, $45°$, $67.5°$, $90°$, $112.5°$, $135°$, $157.5°$*) respectively on normalized fingerprint image, (k) mean of all 8 different Gabor responses, (l) gradient based thresholding, (m) ROI mask generated after performing erosion and dilation on (l), (n) the segmented fingerprint image.*

4. Generate the gradient along X-axis and Y-axis separately for the mean image I'_{mean}, and threshold each gradient image to generate respective binary images. These thresholded gradient binary images are combined using OR operation to generate a single binary image as shown in Figure 5.2(l).

5. The combined binary image thus obtained to get ROI has lot of noisy pixels due thresholding. To remove such noisy pixels, dilation and erosion operations are performed on these images to obtain a clean ROI, which is essentially the binary mask which represents the ROI as shown in Figure 5.2(m).

6. Based on the binary mask, we perform the segmentation of the given fingerprint image, and ROI is obtained (Figure 5.2(n)).

Once the ROI has been estimated, we consider only those minutiae which lies inside ROI. VeriFinger extracts only the typical minutiae features from the fingerprint image. We then added the rare minutiae from the GCDB tenprint minutiae set into the post-processed VeriFinger generated minutiae set for the tenprints. In this case the number of minutiae between the latent and the corresponding mated tenprint minutiae set are not equal, the latent minutiae set is only a subset of the tenprint minutiae set. The average number of minutiae in the latents was 13 and that of tenprints was 125.

The original latent minutiae sets provided by Guardia Civil and the post-processed VeriFinger generated minutiae sets are used in all our experiments. To represent some rare minutiae, multiple points were needed. For example, to represent a *deviation* two points are needed (see type 3 in Figure 5.1), and to represent an *assemble* three points are needed (see type 13 in Figure 5.1). Whenever multiple points are needed to represent a rare minutiae, we mapped them to a single point representation by taking the average of locations and orientations of all points.

From the 268 latent fingerprint minutiae sets, we estimated the probability of occurrence (p_i) of various minutiae types. The probability (p_i) and the entropy-based weights ($w_i = -\log_{10} p_i$) for each minutiae type present in GCDB are listed in Table 5.2. In the 268 latent fingerprints of GCDB, we noticed only seven types of rare minutiae features. They are listed in Table 5.2. Other rare minutiae types are not found in the current database used in this study.

No	Minutiae Type	Probability (p_i)	Weight ($w_i = -\log_{10} p_i$)
1	Ridge-ending	0.5634	0.2492
2	Bifurcation	0.3620	0.4413
3	Deviation	0.0015	2.8294
4	Bridge	0.0024	2.6253
5	Fragment	0.0444	1.3523
6	Interruption	0.0021	2.6833
7	Enclosure	0.0204	1.6896
8	Point	0.0036	2.4492
10	Transversal	0.0003	3.5284

Table 5.2: *The probability of occurrence and the entropy based weights for the minutiae types present in the* 268 *latent fingerprints of GCDB. The numbers correspond to minutiae types in Figure 5.1*

5.2. Algorithm

The latent fingerprints of GCDB are highly partial in nature, with an average of 13 minutiae per latent. To make an appropriate alignment between the latent minutia points and the ten-print minutia points (with an average of 125 minutia points) requires a reliable reference point. We choose the rare minutia features as reference points to perform the alignment.

Let L and M be the representation of latent and tenprint minutia sets respectively. Each minutia is represented as a quadruple $m = \{x, y, \theta, t\}$ that indicates the (x, y) location as coordinates, the minutia angle θ and the minutia type t:

$$L = [m_1\ m_2\ ...\ m_p]\,,\quad m_i = [x_i\ y_i\ \theta_i\ t_i]^T,\quad i = 1...p$$

$$M = [m'_1\ m'_2\ ...\ m'_q],\quad m'_j = [x'_j\ y'_j\ \theta'_j\ t'_j]^T,\quad j = 1...q,$$

where p and q are the number of minutiae in L and M respectively. If $t > 2$, then the minutia is of rare type (from Table 5.1), and $[\,\cdot\,]^T$ denotes transpose.

The algorithm to generate weighted similarity scores from a minutiae matcher is described in two stages. Similarity scores of minutiae matchers are modified only if they contain rare minutia features.

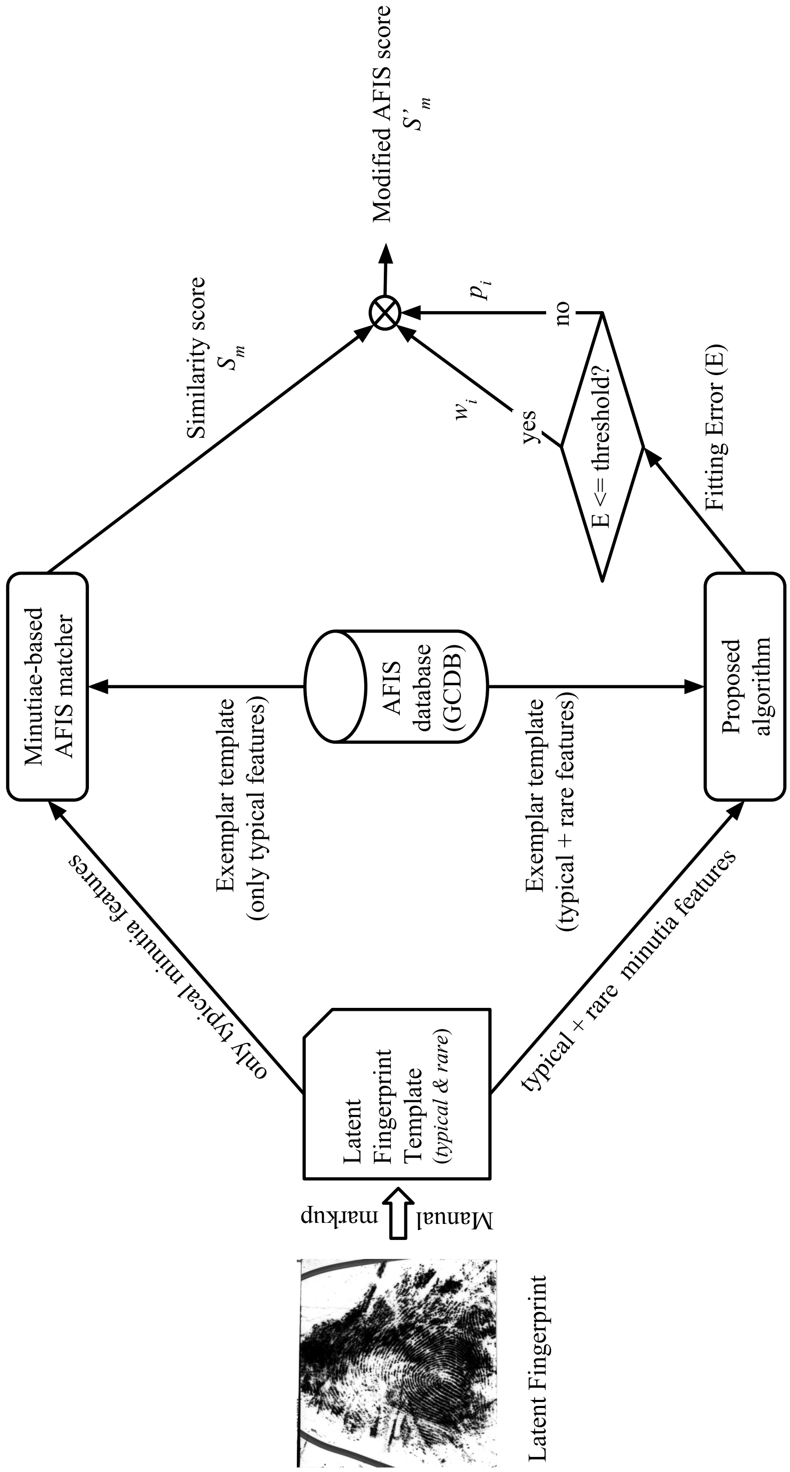


Figure 5.3: *Sequence of steps in estimating the modified similarity score of a reference minutiae-based matcher.*

The first stage of the algorithm estimates the least square fitting error for an affine transformation of the latent minutiae set onto a tenprint minutiae set. The second stage of the algorithm modifies the similarity score generated by the minutiae-based matcher based on the fitting error. Other works related with modifying the similarity score based on pre-alignment are reported in [Krish. *et al.*, 2015; Paulino *et al.*, 2013; Zheng and Yang, 2015]. The sequence of steps involved in generating the modified score of the minutiae matcher using our proposed algorithm is summarized in Figure 5.3.

Stage-1 : Least Square Fitting Error

Step 1: To find the affine transformation between L and M, it is first needed to establish a one-to-one correspondence between minutiae from L and minutiae from M. Let the subset of minutiae from M which establishes correspondence with L be denoted as M_s.

Step 2: Superimpose one rare minutia point of L onto the corresponding rare minutia point of M, only if they both are of the same type (if there are multiple rare minutia points, take any). If the type of the rare minutia between L and M differs, or M does not contain any rare minutiae, then the comparison is assumed to be non-match.

Step 3: To establish the correspondence between latent and tenprint minutia points, we choose the minutia points from M that are close to the minutia points of L. The Euclidean distance is calculated between the minutia pairs to determine whether the pairs are close or not.

Step 4: To compensate for rotation alignment, we rotate the latent in the range $[-45°, +45°]$ with respect to the superimposed rare minutiae, and estimate the Euclidean distance for each rotation step of size $1°$.

Step 5: The optimal rotation is the one for which the average sum of distances between closest pairs is minimum.

Step 6: After the alignment, all those minutia pairs which are within a threshold distance are considered to be mated pairs, and a one-to-one correspondence is established between them. As a result, we obtain a subset M_s of the tenprint minutiae M. After establishing the correspondence, the number of minutiae between L and M_s are the same.

Step 7: Once the correspondence is established, we find the least square fitting error for the affine transformation between the latent minutia points and the subset of tenprint minutiae set.

For $\hat{L}$ and $\hat{M}_s$, which are the modified version of L and M_s with only the (x, y) locations as minutia representation augmented with a value 1, i.e,:

$$\hat{L} = [\hat{m}_1\ \hat{m}_2\ ...\ \hat{m}_p];\quad \hat{m}_i = [x_i\ y_i\ 1]^T;\quad i = 1...p$$
$$\hat{M}_s = [\hat{m'_1}\ \hat{m'_2}\ ...\ \hat{m'_p}];\quad \hat{m'_j} = [x'_j\ y'_j\ 1]^T;\quad j = 1...p,$$

we are looking for some affine transformation matrix

$$A = [a_{jk}]_{j,k=1...3} \tag{5.5}$$

and some translation vector

$$\tau = [\tau_1\ \tau_2\ ...\ \tau_p];\quad \tau_1 = \tau_2 = ... = \tau_p = [\delta_x\ \delta_y\ 1]^T; \tag{5.6}$$

such that

$$\hat{M}_s \approx A\hat{L} + \tau \tag{5.7}$$

where $[\delta_x\ \delta_y]$ is the translation needed to superimpose the rare minutia of L and M.

Step 8: Find the least square fitting error between $\hat{L}$ and $\hat{M}_s$ defined as follows:

$$E^{\hat{L},\hat{M}_s} = \frac{1}{p}\sum_{i=1}^{p} ||\hat{m'_i} - A\hat{m}_i - \tau_i||_2^2 \tag{5.8}$$

where $||\hat{m'_i} - A\hat{m}_i - \tau_i||_2$ is the L_2 norm.

For a match comparison, we expect this fitting error to be small, whereas for a non-match comparison, the fitting error is expected to be large.

If there are multiple matching rare minutiae feature between L and M_s, then $E^{\hat{L},\hat{M}_s}$ is calculated for all such minutiae types. The fitting error for such a comparison is chosen to be the minimum of all the fitting errors calculated.

Stage-2 : Weighted scores

Step 9: Using a standard minutia matcher, generate the similarity score S_m between L and M. The modified similarity score S'_m based on a fitting error threshold E is obtained as follows:

$$S'_m = \begin{cases} S_m \times w_i & \text{if } E^{\hat{L},\hat{M}_s} \leq E \\ S_m \times p_i & \text{if } E^{\hat{L},\hat{M}_s} > E \end{cases} \tag{5.9}$$

where w_i is the derived entropy based weight, and p_i is the probability of occurrence of a particular rare minutia type t_i. The values for w_i and p_i for all minutiae type t_i are listed in Table 5.2.

If $E^{\hat{L},\hat{M}_s} \leq E$, then the comparison is deemed to be a match, and if $E^{\hat{L},\hat{M}_s} > E$, the comparison is deemed to be a non-match.

Thus, we obtain a modified similarity scores S'_m for a particular minutiae matcher by rewarding or penalizing the similarity scores based on the fitting error obtained using our approach.

5.3. Experiments

We performed all our experiments on the minutia sets of 268 latents and corresponding 268 tenprints of GCDB. To generate similarity scores, we used two minutiae matchers namely: NIST-Bozorth3 [NIST-NBIS, NBIS-Release v4.2.0] and VeriFinger SDK [Neurotec-Biometric-4.3]. When reporting the rank identification accuracies in our experiments, there are 268 match comparisons and 268×267 non-match comparisons.

NIST-Bozorth3 is a minutiae based fingerprint matcher that is specially developed to deal with latent fingerprints and is publicly available. This matcher is part of the NIST Biometric Image Software (NBIS) [NIST-NBIS, NBIS-Release v4.2.0], developed by NIST. VeriFinger is a commercial SDK that is widely used in academic research. We report the performance accuracy and improvement of all the matchers using Cumulative Match Characteristic (CMC) curves.

5.3.1. Experiment 1: Fitting Error probability distribution

The least square fitting error probability density estimates for both match and non-match comparisons are shown in Figure 5.4. We can observe that the fitting error itself is discriminatory enough, having separate peaks for both match and non-match distributions. This supports the methodology followed in our algorithm. The following experiments also support this fact.

5.3.2. Experiment 2: Importance of rare minutiae

Four configurations are compared in this experiment to demonstrate the importance of using rare minutia features:

1. *Typical Features*: Only the typical minutia features (ridge-endings and bifurcations) were used, i.e, the fingerprint templates contained only typical features and similarity scores were generated only using the minutiae matchers.

2. *Typical + Rare (all processed as Typical)*: Both the typical minutia features and rare minutiae features were used, considering both as typical minutiae. The original representation of the rare minutia features was maintained in this experiment, i.e, when multiple points are used to represent the same rare minutia feature, they all were used as such without taking any averaging of the minutiae as compared against averaging of rare minutia points

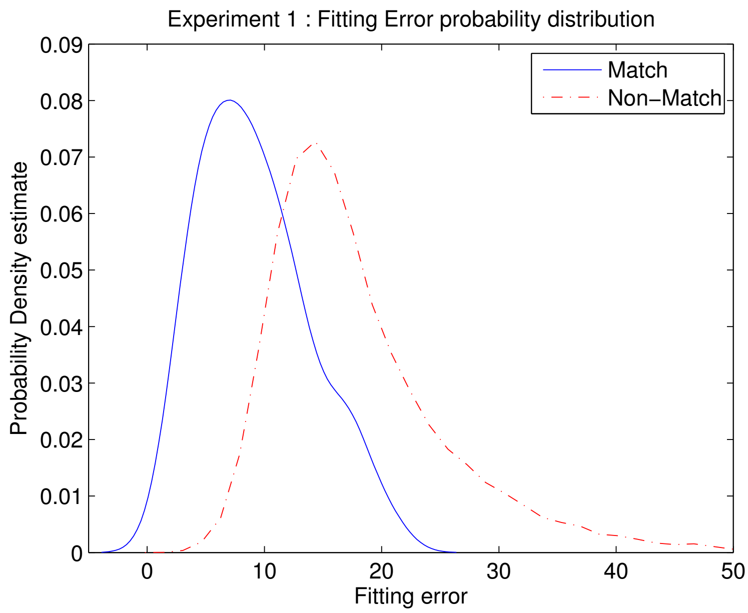


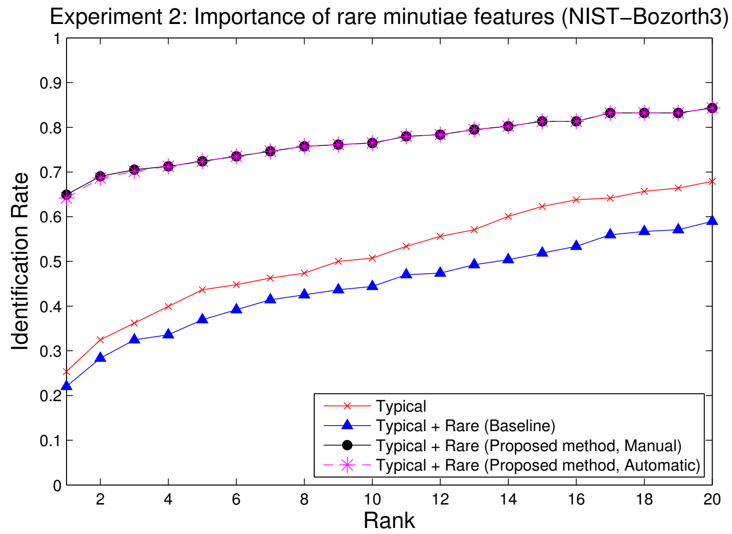
Figure 5.4: *Probability density estimate of the fitting errors for match and non-match comparisons.*

to estimate the least square fitting error.

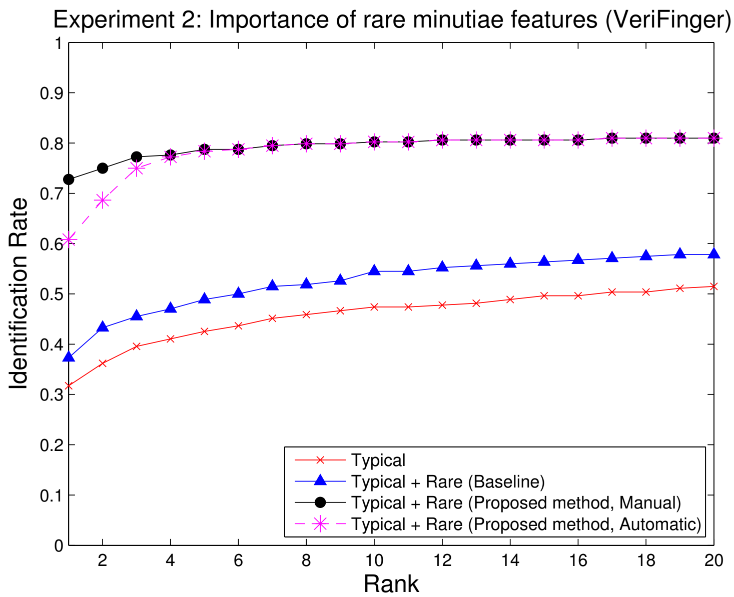
3. *Weighted Scores - Manual*: In this kind of analysis, the match and non-match comparisons were not automatically decided but manually partitioned. The similarity scores were rewarded for match comparisons and penalized for non-match comparisons without any error.

4. *Weight Scores - Automatic*: Here, the similarity scores generated by the minutiae matchers are modified automatically, which is more appropriate in a real-time operational scenario. Without knowing whether a particular comparison is a match or non-match comparison, the similarity scores generated by the minutiae-based matcher are modified based on fitting error alone. If the fitting error was less than or equal to E, then the comparison is deemed to be a match comparison and their similarity score is rewarded as indicated in Eq.(5.9). If the fitting error is more than E (a non-match comparison), then the similarity score is penalized.

Figures 5.5(a) and 5.5(b) show the rank identification accuracy in CMC curve for both NIST-Bozorth3 and VeriFinger separately for the four configurations listed above. *Typical + Rare* did not improve in case of NIST-Bozorth3 but it did slightly improve in the case of VeriFinger. We can also notice that modification of similarity scores based on the fitting error (both *Manual* and *Automatic*) significantly improves the rank identification accuracies of both NIST-



(a) CMC curve for NIST-Bozorth3



(b) CMC curve for VeriFinger SDK

Figure 5.5: *Improvement in rank identification when incorporating rare minutia features.*

Matcher	Typical Features (Rank-1)	Typical + Rare (Rank-1)	Weighted Scores - Manual (Rank-1)	Weight Scores - Automatic (Rank-1)
NIST-Bozorth3	25.37	22.01	64.93	64.18
VeriFinger	31.72	37.31	72.76	60.82

Table 5.3: *Rank-1 identification (in %) for NIST-Bozorth3 and VeriFinger under various categories of analysis.*

Bozorth3 and VeriFinger. For NIST-Bozorth3, the Rank-1 identification improved from 25.37% to 64.18%, and for VeriFinger, the Rank-1 identification improved from 31.72% to 60.82% when rare minutia features were incorporated and the similarity scores are modified based on the fitting error proposed in our algorithm.

Table 5.3 summarizes the Rank-1 accuracy for both NIST-Bozorth3 and VeriFinger under the four configurations considered. The improvement in rank identification accuracy is very similar for *Manual* and *Automatic* modification of similarity scores for NIST-Bozorth3. In case of VeriFinger, the Rank-1 identification for *Automatic* is slightly lower than *Manual*, but beyond Rank-5, the identification accuracy remains the same.

5.3.3. Experiment 3: Parameters - Optimal fitting error threshold

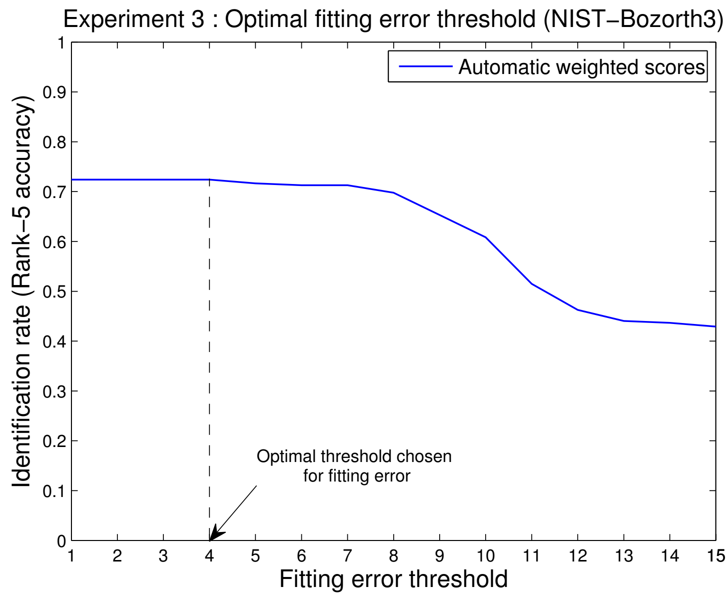
The fitting error threshold E plays a crucial factor in algorithm. So, arriving at an optimal threshold value is of importance. Figures 5.6(a) and 5.6(b) show the performance of both NIST-Bozorth3 and VeriFinger for various fitting error thresholds. The Rank-5 identification accuracies are analyzed for fitting error thresholds ranging from 1 to 15.

We analyzed that beyond the threshold value of 4, the system starts to degrade the performance. So, we chose an optimal threshold value of $E = 4$ for the experiments reported in the rank identification accuracies of both NIST-Bozorth3 and VeriFinger.

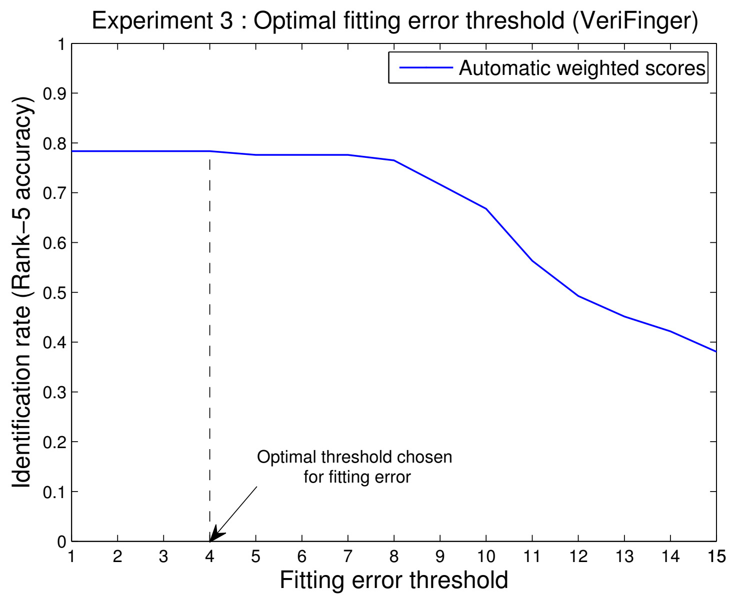
5.4. Discussions

We proposed a method that makes use of reliably extracted rare minutia features to improve the rank identification accuracies for minutiae matchers when dealing with latent fingerprints.

The usefulness of the proposed model is demonstrated on two widely used minutiae-based matchers, NIST-Bozorth3 and VeriFinger. Both matchers showed significant improvements in the rank identification accuracies when their similarity scores were modified based on the fitting error proposed in our methodology. We conclude that even if we have only few number of



(a) CMC curve for NIST-Bozorth3



(b) CMC curve for VeriFinger SDK

Figure 5.6: *Optimal threshold for the fitting error chosen based on Rank-5 identification accuracy.*

minutiae in a partial latent, the presence of reliably extracted rare minutia features makes the comparison more robust. In our experiments, we used the rare minutia features that were manually extracted by forensic examiners. Developing more robust automatic extraction of rare minutiae can significantly improve the current state of the art in AFIS adapted for latent fingerprints.

Chapter 6

Related works in Evidence evaluation using Likelihood Ratios

In this chapter, we address the issue of the interpretation of forensic evidence from scores computed by a biometric system. This is one of the most important topics into the so-called area of forensic biometrics. We will show the importance of the topic, and how to address it using the previously proposed systems to incorporate rare minutiae in the Thesis. The main objective of the Chapter is to propose a solution to evaluate the forensic evidence using the systems proposed in this Thesis. Moreover, we will show how the incorporation of rare minutiae improve the performance of the system, also at the level of forensic interpretation.

6.1. Likelihood Ratio framework for evidence evaluation

The evaluation of the relationship between two pieces of evidence at judicial trials has been the subject of discussion in the past years [Saks and Koehler, 2005]. Here, the problem is to give a value to a comparison of a questioned material or evidence (namely trace, for instance a latent fingerprint in a crime scene or a wire tapping involving an incriminating conversation) with some control material of known origin (for instance, a fingerprint from a suspect, or some recordings of a known individual). From a formal logical perspective [Cook *et al.*, 1998b], the given value should represent the degree of support of the comparison to any of the propositions (or hypotheses) involved in the trial. Examples of simple hypotheses might be "the trace and the control materials were originated from the same source" or "the trace and the control materials were originated from different source", but more complex hypotheses can be considered [Cook *et al.*, 1998b]. In some sense, the value of the evidence represents the weight of the link between the trace and the control material in the context of the propositions considered.

Evidence evaluation using a Bayesian probabilistic framework has been proposed in recent years as a logical and appropriate way to report evidence to a court of law [Aitken and Taroni, 2004]. In Europe, initiatives to foster this approach, some of them in response of notorious miscarriages of justice [Various, 2011], have led to the release of a Guideline [Willis, 2015] by the European Network of Forensic Science Institutes (ENFSI), an organization that includes almost all the main forensic laboratories in Europe. According to this Guideline, a Bayesian framework for forensic evaluative reports is recommended for all disciplines and laboratories within ENFSI. Under this Bayesian approach, a likelihood ratio is computed to represent the value of the evidence, and to be reported to a court of law. This framework clearly complies with the requirements of modern forensic science [Saks and Koehler, 2005]: it is scientifically sound (transparent procedures, testability, formally correct), and clearly separates the competences of the forensic examiner and the court. The establishment of this Bayesian evaluative framework has motivated the convergence of pattern recognition and machine learning approaches to yield probabilistic outputs in the form of likelihood ratios. A common architecture for this considers two steps: first, the computation of a discriminating score between two evidential materials (e.g., a latent fingerprint in the crime scene and an exemplar fingerprint from a known suspect), which can be obtained from a standard biometric system; and second, the transformation of the score into a likelihood ratio. This process of transforming scores relating two pieces of evidence into likelihood ratios has been dubbed calibration [Brümmer and du Preez, 2006; Ramos and Gonzalez-Rodriguez, 2013; vanLeeuwen and Brümmer, 2007].

6.1.1. Challenges in LR based evidence evaluation

Despite its advantages, the computation of likelihood ratios still presents important challenges. We enumerate the most important as follows. First, complex evidence evaluation cases are still problematic. Probabilistic graphical models, particularly Bayesian networks [Taroni *et al.*, 2006], have been proposed to address those situations. However, this emerging field is still an active area of research, and more efforts are needed in order to provide forensic examiners with appropriate models in particular scenarios, especially if those models are to be learned from data.

Second, the typical scenario in forensic science involves data presenting variable and unfavorable conditions, which means that automatic comparisons between traces and control materials will result in a challenging problem. Efforts to model or compensate this variability in likelihood ratio computation should be improved. Some works such as [Zadora and Ramos, 2010] have contributed to evaluate the impact of this problem. Moreover, integration of advanced machine learning algorithms (like in [Dehak *et al.*, 2010; Li *et al.*, 2010]) for variability compensation into forensic evaluation is a solution to this, although it still remains a challenge.

Third, in forensic science the databases are difficult to obtain and to use, even for research purposes. This is because, although there is plenty of forensic data in some disciplines (e.g., large fingerprint databases), there are interoperability, legal and privacy issues that difficult the use of this data. This leads to two opposite situations: either the databases are big or the databases are highly scarce. The use of robust models to data scarcity has been tackled by different techniques as in [Villalba and Brummer, 2011; Zadora *et al.*, 2014]. However, to our knowledge, evidence evaluation models have not been adapted to big-data scenarios to handle big databases when possible, which represents a loss of information in these scenarios.

Fourth, although likelihood ratio computation methods are becoming more and more popular, the validation of those methods is still not standardized. Even if likelihood ratios are computed to evaluate the links between evidential materials, this does not guarantee that they will be able to be integrated into a Bayesian decision framework. Likelihood ratios should present the best possible calibration in order to properly assist decision makers and fact finders in judicial processes [Ramos and Gonzalez-Rodriguez, 2013]. Calibration is a property of a set of likelihood ratios, by which the LR is itself a measure of evidential weight. This leads to the property that 'The LR of the LR is the LR', meaning that the LR is interpreting the evidence with the best probabilistic meaning in terms of Bayes decisions [van Leeuwen and Brummer, 2013]. In that sense, computing likelihood ratios is not enough, they should also be the best calibrated as possible. There are current efforts of the forensic community in order to establish formal frameworks for the validation of likelihood ratio models [Haraksim *et al.*, 2015a; Ramos and Gonzalez-Rodriguez, 2013; Ramos *et al.*, 2013], but research is still needed.

6.1.2. Related works

Evidence evaluation in fingerprints by the use of LR has been recently proposed in remarkable works like in [Neumann *et al.*, 2012] for minutiae configurations extracted manually from forensic examiners. There, distances are computed in order to do the comparison. Other models based on the use of AFIS scores to compute likelihood ratio values can be found in [Egli, 2009], and more recently [Haraksim *et al.*, 2015a]. Models of fingerprint evidence evaluation have been recently reported in [Neumann *et al.*, 2015]. To our knowledge, there have been no models for LR computation including information about rare minutiae as proposed in this Thesis.

6.2. Case assessment and interpretation methodology

As previously mentioned, a milestone in the use of the LR methodology in Europe was the Case Assessment and Interpretation (CAI) methodology developed by the Forensic Science Service (FSS) in the late 90's [Cook *et al.*, 1998b]. This was the result of the efforts of the now closed Forensic Science Service of the United Kingdom, in order to homogenize and make more agile the relationship between court and forensic service providers (*e.g.*, police forces or other public or private forensic laboratories). The ultimate aim is the use of a logical methodology to avoid pitfalls of reasoning and fallacies. The methodology has been described in several papers during the end of the 20^{th} century [Cook *et al.*, 1998a,b; Evett *et al.*, 2000].

There are several characteristics that are typical from this CAI methodology, which we summarize below.

- Full integration of the LR methodology into the forensic evidence evaluation process. In this sense, all the elements typical from LR evidence evaluation are present, namely the evidence, propositions, probabilistic reasoning and assignment, etc.

- A particular emphasis is put in the definition of the propositions in the case, which have to be informed by the circumstances of the case. In that sense, the relationship between the court and the forensic science provider should be essential in order to define the propositions. Issues like the definition of the population used to model the alternative proposition, the specificity of the propositions with respect to the population, the suspect and the trace, or the selection of the most appropriate database to address the propositions [Champod *et al.*, 2004], are of particular importance.

- A hierarchy of propositions [Cook *et al.*, 1998a] is introduced in order to address the forensic casework in the most appropriate manner with respect to the information in the case. In this sense, there are three basic levels in the hierarchy: *source level*, the lowest level of all, where issues about source attribution are considered; *activity level*, where the perpetration of a determined act is under discussion; and *offence level*, the highest level, where the commission of a crime is considered. Depending on the information in the case available to the forensic scientist, it is possible to climb up to higher level, but in most cases the forensic scientist is confined to the source level, especially nowadays when models for activity or offence level are under discussion and research.

- Case pre-assessment is encouraged by the model. Under this concept, a preliminary LR value is reported prior to the case itself, in order to indicate what would be the expected outcome of the forensic analysis by the examiner. This helps to aim the expectations of the client, and has important implications regarding the efficiency of resources in a case.

The CAI methodology is not possible to be fully implemented in this Thesis. However, there are several issues that are possible to address for the proposed fingerprint matchers. First, the LR methodology will be followed. Second, we will try to define our propositions according to the information present in the Guardia Civil database, which will be used to simulate forensic cases. Third, appropriate databases will be selected to model the propositions according to the limitations in the simulated forensic scenarios.

6.3. Evidence evaluation with Likelihood Ratios

The LR framework for interpretation of the evidence represents a mathematical and logical tool in order to aid in the inference process derived from the analysis of the evidence. In this methodology, the objective of the forensic scientist is computing the likelihood ratio (LR) as a degree of support of one proposition versus its opposite [Aitken and Taroni, 2004; Champod and Meuwly, 2000].

The LR framework is stated as follows. Consider the forensic fingerprint evidence E as the materials to compare in a forensic case, namely a recovered latent fingerprint of unknown origin and a control fingerprint (the exemplar) whose origin is known. In a forensic case, the unobserved variable of interest is the true proposition $H = \{H_p, H_d\}$, where H_p and H_d are the prosecution and defense propositions according to the CAI methodology.

Bayes' theorem [Aitken and Taroni, 2004] relates probabilities before and after evidence analysis:

$$P\left(H_p \mid E, I\right) = \frac{p\left(E \mid H_p, I\right) \cdot P\left(H_p \mid I\right)}{p\left(E \mid I\right)} \tag{6.1}$$

where I is the background information available in the case not related to the evidence E, as defined by the CAI methodology. Equation 6.1 then allows the following inference:

$$\frac{P\left(H_p \mid E, I\right)}{P\left(H_d \mid E, I\right)} = LR \cdot \frac{P\left(H_p \mid I\right)}{P\left(H_d \mid I\right)} \tag{6.2}$$

$$LR = \frac{p\left(E \mid H_p, I\right)}{p\left(E \mid H_d, I\right)} \tag{6.3}$$

Equation 6.2 is the so-called *odds form* of Bayes' theorem. In this framework, we can distinguish two values:

1. The prior probabilities $P\left(H_p \mid I\right) = 1 - P\left(H_d \mid I\right)$, which are province of the fact finder and should be stated assuming only the background information (I) in the case [Evett, 1998].
2. The LR (Equation 6.3[1]), computed by the forensic scientist [Aitken and Taroni, 2004].

The LR value (Equation 6.3) is the quotient of two probability densities. On the one hand, the probability density function (pdf) $p\left(E \mid H_p, I\right)$ in the numerator in Equation 6.3 is known as the within-source distribution. Its evaluation in the particular value of the evidence E gives a measure of the probability of observing the evidence under H_p. On the other hand, the pdf $p\left(E \mid H_d, I\right)$ in the denominator is known as the between-source distribution, and its evaluation

[1]Unless explicitly stated, we will use a capital E for referring to the given value of the evidence, according to the literature on $LR-$based analysis of the evidence [Aitken and Taroni, 2004; Champod and Meuwly, 2000]. Thus, the small e will be used as the argument in likelihoods.

in the particular value of the evidence E gives a measure of the probability of observing the evidence under H_d. Both values should be computed in a transparent way by the forensic evaluation system. It is also the duty of the forensic evaluation system, following the background information of the case (I), to select the population of individuals which will be proper for the case at hand[1].

This LR-based framework presents many advantages in a forensic context:

- It allows forensic scientists to evaluate and report a meaningful value for the weight of the evidence to the court [Champod and Meuwly, 2000].

- The role of the examiner is clearly defined, leaving to the court the task of using prior judgments or costs in the decision process.

- Probabilities can be interpreted as degrees of belief [Taroni *et al.*, 2001], allowing the incorporation of subjective opinions as probabilities in the inference process in a clear and scientific way.

The LR value has an interpretation as a *support* to a previously stated opinion, due to the analysis of the evidence E. In other words:

- If the LR> 1 the evidence will support that $H = H_p$, i.e., the prosecutor proposition.
- If the LR< 1 the evidence will support that $H = H_d$, i.e., the defense proposition.

Moreover, the value of the LR represents the *degree of support* of the evidence to the value of H. For instance, LR$= 3$ means that "the evidence supports the odds in favor of $H = H_p$ with a degree of 3". Therefore, a single LR value has a *meaning* by itself.

It is important to note that the LR *supports* an opinion about H, but the LR *is not* an opinion about H. Opinions about H are represented as probabilities or, in our binary case, odds in favor of a given outcome of H. Therefore, it is not possible to make a decision about the value of H based solely on the value of the LR, because decisions will be taken from posterior opinions as it will be shown later in this chapter.

[1]The background information about the case I will be eliminated from the notation for the sake of simplicity hereafter. It will be assumed that all the probabilities defined are conditioned to I.

6.4. Performance measurement of LR methods

A solution to measure the performance of likelihood ratio values has been proposed in [Brümmer and du Preez, 2006] for speaker recognition, and has been dubbed *log-likelihood-ratio cost* (C_{llr}). Later, it has been used in many other fields [Ramos *et al.*, 2013]. C_{llr} is defined as follows:

$$C_{llr} = \frac{1}{2 \cdot N_p} \sum_{i_p} \log_2 \left(1 + \frac{1}{LR_i}\right) + \frac{1}{2 \cdot N_d} \sum_{j_d} \log_2 (1 + LR_j) \tag{6.4}$$

where N_p and N_d are respectively the number of LR values to evaluate where H_p and H_d are respectively true. The indices i_p and j_d respectively denote summing over the LR values where each proposition is respectively true.

An important result is derived in [Brümmer and du Preez, 2006], where it is demonstrated that C_{llr} is the expected decision cost for any value of false alarms and false rejections for every value of decision costs involved in a Bayesian decision [Duda *et al.*, 2001], and assuming $P(H_p) = P(H_d) = 0.5$. This important result means that minimizing the value of C_{llr} also encourages to obtain reduced Bayes decision costs for a wide range of decision costs. When the LR is used to make a decision (as it should happen on trial), the minimization of C_{llr} implies the minimization of the average expected cost of the decisions, assuming a non-informative prior probability. This property has been highlighted as extremely important in forensic science [Ramos and Gonzalez-Rodriguez, 2013].

In [Brümmer and du Preez, 2006], an algorithm known as Pool Adjacent Violators is used in order to decompose C_{llr} as follows:

$$C_{llr} = C_{llr}^{min} + C_{llr}^{cal} \tag{6.5}$$

where:

- C_{llr}^{min} represents the *discrimination loss* of the system under evaluation. It is obtained by the C_{llr} of the LR values obtained after PAV-optimization. C_{llr}^{min} is the lowest C_{llr} value which a LR set can achieve while preserving the discriminating power of the LR set under evaluation. Therefore, the expected cost due to C_{llr}^{min} is due to non-perfect discriminating power.
- C_{llr}^{cal} represents the *calibration loss* of the system under evaluation with respect to the best system preserving discrimination. It is computed as $C_{llr}^{cal} = C_{llr} - C_{llr}^{min}$. If the LR values under evaluation converge to the PAV-calibrated LR values, then C_{llr}^{cal} will be reduced.

C_{llr} is a scalar measure of performance of LR values. Here *Empirical Cross-Entropy* (ECE) plots are presented in order to show the overall performance of the set of LR values in terms of accuracy, which takes into account both the discriminating power of the set of LR values, and also the calibration. Figure 7.9 shows several ECE plots. ECE is defined as a generalization of C_{llr} as the average value of the logarithmic scoring rule, weighted in the following way:

$$
\begin{aligned}
ECE = & -\frac{P(H_p|I)}{N_p}\sum_{i:H_p \text{ is true}} log_2 P(H_p|E_i, I) \\
& -\frac{P(H_d|I)}{N_d}\sum_{j:H_d \text{ is true}} log_2 P(H_d|E_j, I),
\end{aligned} \tag{6.6}
$$

where E_i and E_j denote the evidence in each of the comparisons (cases) in the validation set where either H_p or H_d is true, N_p and N_d are the numbers of cases, and I is the background information. It is informative to express ECE explicitly in terms of the prior odds, which can be shown to be:

$$
\begin{aligned}
ECE = & \quad \frac{P(H_p|I)}{N_p}\sum_{i:H_p \text{ is true}} log_2\left(1+\frac{1}{LR_i \times O(H_p|I)}\right) \\
& + \frac{P(H_d|I)}{N_d}\sum_{j:H_d \text{ is true}} log_2 (1 + LR_j \times O(H_d|I)),
\end{aligned} \tag{6.7}
$$

where $O(H_p|I) = \frac{P(H_p|I)}{P(H_d|I)}$ are the prior odds in favour of H_p.

As it can be seen in Equations (6.6) and (6.7), the averages in ECE are weighted by the value of the prior probabilities. This weighting allows ECE to be interpreted in an information-theoretical way, but this topic is outwith the scope of this work [Ramos *et al.*, 2013].

Equation (6.7) shows that ECE depends on the validation set of LR values in the experiment (i.e. the LR values and their corresponding ground-truth labels). However, ECE also depends on the value of the prior odds $O(H_p|I)$, since there is dependence on the posterior probabilities. Thus, ECE can be represented in a prior-dependent way. An example of such a representation can be seen in Figure 7.9. Base-10 logarithms are used for the prior odds in the X axis because they are typically used for evidence evaluation. However, base-2 logarithms are used for computation of ECE because of its information-theoretical interpretation.

ECE in Figure 7.9 represents the accuracy for all the possible values of the prior probability, but calibration is not explicitly measured in such a representation. Therefore, an explicit measurement of discriminating power and calibration is given in a so-called ECE plot [Ramos *et al.*, 2013], which shows three comparative performance curves together:

- solid, red curve: accuracy. This curve is the ECE of the LR values in the validation set, as a function of the prior log-odds. The lower this curve, the more accurate the method;

- dashed, blue curve: perfectly calibrated accuracy. This curve is the ECE of the validation set of LR values after being perfectly calibrated, as a function of the prior log-odds. Therefore, this shows the performance of a validation set of optimally-calibrated LR values obtained by a transformation applied to the original validation set of LR values. In order to obtain this curve the value of the ground-truth labels should be known. Therefore, this curve is not possible to obtain in practice, and represents a *ceiling of performance* useful to measure calibration. Details about the procedure of obtaining these calibrated LR values are outwith the scope of this work, and can be found in [Brümmer and du Preez, 2006] and [Ramos *et al.*, 2013]. The transformation is essentially conducted by the *Pool Adjacent Violators* algorithm, as it happened for obtaining C_{llr}^{min}, and therefore this curve in the ECE plots is sometimes referred to as *accuracy after PAV*;

- dotted curve: neutral reference. It represents the comparative performance of a so-called *neutral LR method*, defined as the one which always delivers LR= 1 for each forensic case simulated in the set of LR values. This neutral method is taken as a *floor of performance*: the accuracy should always be better than the neutral reference. Therefore, the solid curve in an ECE plot should be always lower than the dotted curve, for all represented values of the prior log-odds (the names *floor* and *ceiling* are the opposite of the usual physical connotations but are chosen to represent the lowest and highest levels of performance).

Thus, and according to [Ramos and Gonzalez-Rodriguez, 2013; Ramos *et al.*, 2013], the following are represented (in ECE plots):

- accuracy: solid curve. The lower the curve, the better the accuracy;

- discriminating power: dashed curve. The lower the curve, the better the discriminating power. The justification that the ECE after *PAV* represents the discriminating power can be found in [Brümmer and du Preez, 2006] (with a theoretical development) and in [Ramos and Gonzalez-Rodriguez, 2013] (more adapted to forensic science);

- calibration: difference between the solid and dashed curves. The closer the dashed and the solid curves, the better the calibration.

6.5. LR computation methods

In this subsection, some common algorithms for LR computation are described. These are the ones which will be used in this Thesis.

6.5.1. Gaussian Maximum Likelihood (Gaussian-ML)

LR computation in forensic automatic speaker recognition has been classically performed by the use of generative techniques modelling the hypotheses-conditional distribution of the evidence scores E. This is the approach already presented in [Meuwly, 2001], and has been followed in subsequent works in the literature. Under this approach, the objective is assigning the likelihoods $p(E|H_p)$ and $p(E|H_d)$ respectively to the scores in the training set, in order to compute the LR value.

Assigning $p(E|H_p)$ and $p(E|H_d)$ implies the selection of a proper model. The most straightforward choice for biometric scores could be the Gaussian distribution, obtained via Maximum Likelihood from the training set of scores. However, this requires the distributions involved to present a good fitting with Gaussian probability density functions, which is not typically the case. Fortunately, some score normalization techniques such as T-Norm tend to generate Gaussian distributions for scores when H_d is true [Navratil and Ramaswamy, 2003], as will be pointed out in the experiments to follow.

6.5.2. Logistic regression

Logistic regression is a well-known pattern recognition technique widely used for many problems including fusion [Brümmer *et al.*, 2007; Pigeon *et al.*, 2000] and more recently calibration [Brümmer and du Preez, 2006; Gonzalez-Rodriguez *et al.*, 2007]. The aim of logistic regression is obtaining an affine transformation (i.e., shifting and scaling) of an input dataset in order to optimize an objective function. Let $E = \{E_1, E_2, \ldots, E_K\}$ be a set of scores from K different biometric systems. The affine transformation performed by the logistic regression model can be defined as:

$$f_{lr} = \log\left(O\left(H_p|E\right)\right) = a_0 + a_1 \cdot E_1 + a_2 \cdot E_2 + \ldots + a_K \cdot E_K \tag{6.8}$$

Using Bayes theorem in odds form this expression allows the computation of the logarithm of the LR value for a given value of the prior probabilities.

$$\begin{aligned} \log(LR) &= a_0 + a_1 \cdot E_1 + a_2 \cdot E_2 + \ldots + a_K \cdot E_K - \log\left(O\left(\theta_p\right)\right) \\ &= a_0' + a_1 \cdot E_1 + a_2 \cdot E_2 + \ldots + a_K \cdot E_K \end{aligned} \tag{6.9}$$

which leads to the following *logistic regression model*:

$$P\left(H_p|E\right) = \frac{1}{1+e^{-f_{lr}}} = \frac{1}{1+e^{-\log(LR)-\log(O(H_p))}} \tag{6.10}$$

As it can be seen, the logistic regression transformation from score values to posterior log-odds is invertible unless it is a constant function. The weighting terms $\{a_0, a_1, a_2, \ldots, a_K\}$ can be obtained from a set of training data with optimization procedures found in the literature[1]. For a given value of the prior odds $(O(H_p))$ we may define $f_{lr}^t = a_0 + \sum_{j=1}^{K} a_j E_j^t$ as the value

[1]In this Thesis we have used the FoCal toolkit for training logistic regression models (http://niko.brummer.googlepages.com).

obtained for a given set $E^t = \left\{E_1^t, \ldots, E_K^t\right\}$ of target scores from the K biometric systems. On the other hand, let define $f_{lr}^{nt} = a_0 + \sum_{j=1}^{K} a_j E_j^{nt}$ the value obtained for a given set $E^{nt} = \left\{E_1^{nt}, \ldots, E_K^{nt}\right\}$ of non-target scores from the K biometric systems. Logistic regression computes the $\{a_0, a_1, a_2, \ldots, a_K\}$ coefficients by making $P\left(H_p \middle| E\right)$ as close as possible to 1 for target trials and to 0 for non-target trials, constrained to the logistic regression model (Equation 6.10). It can be derived [Brümmer *et al.*, 2007; Pigeon *et al.*, 2000] that such optimization leads to the following objective to minimize:

$$C_{wlr} = P\left(H_p\right) \cdot \frac{1}{N_t} \sum_{i=1}^{N_t} \log_2\left(1 + e^{-f_{lr,i}^t}\right) + P\left(H_d\right) \cdot \frac{1}{N_n} \sum_{i=1}^{N_n} \log_2\left(1 + e^{f_{lr,i}^{nt}}\right) = ECE \qquad (6.11)$$

where N_t is the number of f_{lr}^t values and N_{nt} is the number of f_{lr}^{nt} values, both in the training set. As it is highlighted in Equation 6.11, the optimization objective in logistic regression is precisely the Empirical Cross-Entropy (ECE) of the training score set. For a given value of $P\left(H_p\right)$ and assuming a logistic regression model, ECE is convex with respect to the weighting terms $\{a_0, a_1, a_2, \ldots, a_K\}$, and therefore it has a global minimum [Brümmer *et al.*, 2007].

As logistic regression optimizes ECE, it can be used for calibration as well as for fusion. If the number of systems K is more than one, we will be fusing the input scores and mapping them to a single value of the posterior log-odds. As an additional effect, because of the objective function optimization, the output of such a fusion will tend to be calibrated. On the other hand, if $K = 1$ then the input score is transformed by an affine mapping which will tend to give a calibrated output.

Therefore, for a given value of the prior probabilities, but using Equation 6.9, posterior log-odds are mapped again into $\log\left(LR\right)$ values which will also tend to be calibrated. If the prior probabilities are known, the value of $P\left(H_p\right)$ can be set. If the prior probabilities are unknown, the $\log\left(LR\right)$ value can be obtained for an arbitrary value of the prior, and it will tend to be calibrated for any value of the prior. A typical choice for this prior may be $P\left(\theta_p\right) = 0.5$, and therefore logistic regression will optimize C_{llr} in that case.

6.5.3. Pool Adjacent Violators (PAV) calibration

Another approach to score calibration has been proposed by the use of the Pool Adjacent Violators (PAV) algorithm [Brümmer and du Preez, 2006]. The PAV algorithm transforms a set of scores into a set of calibrated LR values. However, it is only possible to apply an optimal PAV transformation if the ground-truth labels of the propositions for each score in the set are known. Nevertheless, as suggested in [vanLeeuwen and Brümmer, 2007], a PAV transformation can be trained on a set of scores for which the true value of the hypotheses are known and then apply the trained transformation to scores for which the hypothesis value is unknown. Although a straight use of PAV leads to a non-invertible transformation, several *smoothing* techniques can be applied to PAV in order to keep it monotonically rising. For instance, adding an infinitely small slope to PAV will lead to an invertible transformation. Interpolating with linear, quadratic or splines approaches are also possible smoothing schemes.

Chapter 7

Incorporating rare features in LRs : Experiments

7.1. Analysis of score distributions from fingerprint recognition systems

In this section, we analyze the distributions of the scores from the minutiae-based fingerprint recognition systems which incorporates extended feature sets as proposed previously in this Thesis. The reason for this is two-fold. First, we want to detect the suitability of those systems for a LR framework for evidence evaluation. In some cases, this will imply a decision of not using those systems for evidence evaluation unless there is an appropriate model to address those distributions. In other cases, further stages will be needed to properly compute LR values that take advantage of the information contained in the scores. Second, we would like to guess what the best model could be used in order to model the probabilities in the numerator and the denominator of the LR. After this analysis stage, we should be able to propose several models for LR computation for several systems proposed in the Thesis.

7.1.1. Histograms of pooled scores

Here we analyze the histograms of pooled scores, *i.e.*, all the scores from all the queries in the database put together.

Figures 7.1 to 7.3 shows the histograms for the three algorithms (namely MCC, NIST-Bozorth3 and VeriFinger) and the four configurations to include rare minutiae (namely *Typical*, *Typical+Rare*, *Manual* and *Automatic*).

Four configurations are summarized as follows:

1. *Typical Features*: Only the typical minutia features (ridge-endings and bifurcations) were used, i.e, the fingerprint templates contained only typical features and similarity scores were generated only using the minutiae matchers.

2. *Typical + Rare (all processed as Typical)*: Both the typical minutia features and rare minutiae features were used, considering both as typical minutiae.

3. *Weighted Scores - Manual*: The match and non-match comparisons were not automatically decided but manually partitioned. The similarity scores were rewarded for match comparisons and penalized for non-match comparisons without any error.

4. *Weight Scores - Automatic*: The similarity scores generated by the minutiae matchers are modified automatically based on fitting error alone as described in Section 5.2.

From the observation of the histograms, several conclusions can be extracted. First, the amount of zeroes in the histograms of systems using NIST-Bozorth3 and VeriFinger algorithms is very high. This means that the score value of zero accumulates many genuine and impostor scores altogether. This is not the case for the MCC algorithm, where the amount of values equal to zero is much more restricted. It is seen that for MCC-*Automatic* and MCC-*Manual* approaches there are many scores close to zero, but they are not exactly zero, as it can be seen in Figure 7.4. This is because the raw score of the minutiae-based matchers are modified based on the fitting error extimated with respect to the rare minutia feature as described in Equation 5.9. However, this will be corrected by the use of score normalization techniques, which will be analyzed further. Worth noting, score normalization techniques cannot easily correct the behavior of the scores coming from NIST-Bozorth3 or VeriFinger algorithms, because many scores are actually zero, not close to zero as in MCC-*Automatic* and MCC-*Manual*.

This behavior of NIST-Bozorth3 and VeriFinger algorithms represents a problem for standard LR computation algorithms in the continuous domain, because they use to model probability density functions, and there is not a clear density that can properly fit the scores in these systems. Since the objective of this chapter is to propose a LR method to confirm the importance of rare minutiae in forensic evidence evaluation, and not to propose a novel LR method, we have decided to consider only the MCC algorithm for the remainder of the chapter. This will perfectly serve to illustrate the objectives of the chapter. In the future we will explore more flexible algorithms to deal with score distributions partly concentrated in zero values, as they are quite common in AFIS technology[1].

[1]These scores are commonly known as *early-outs*.

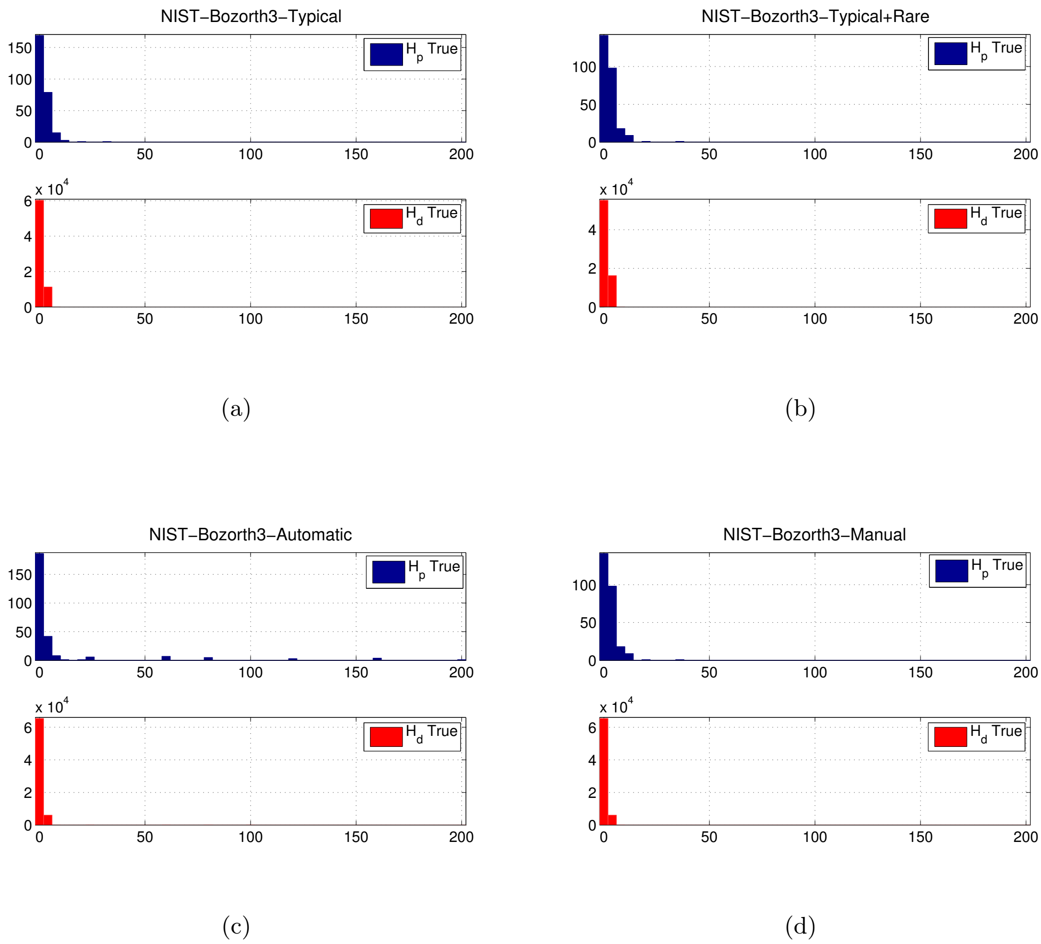
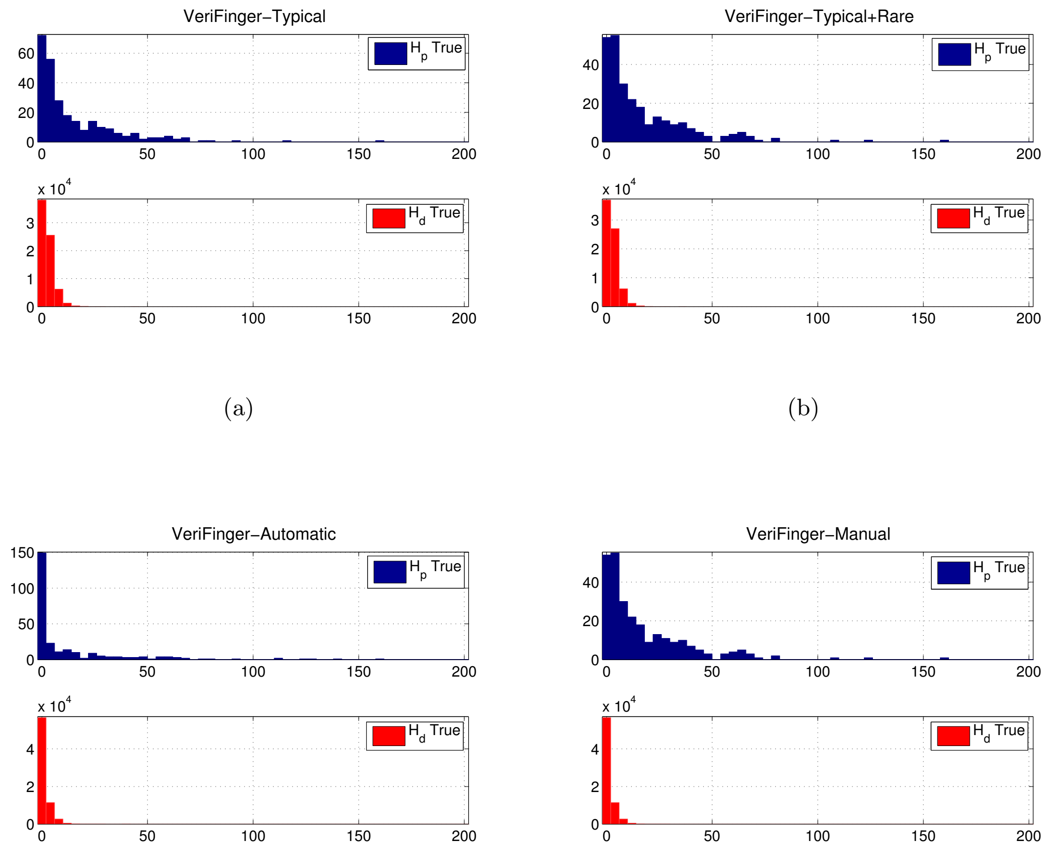
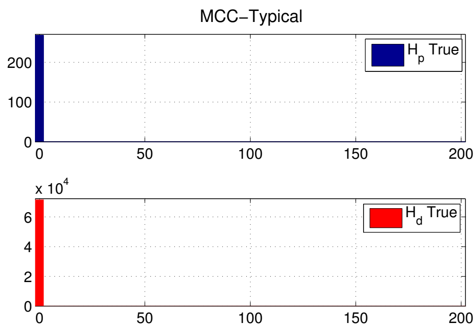


Figure 7.1: *Histograms of pooled scores for (c) and NIST-Bozorth3-*Typical *(a), NIST-Bozorth3-*Typical+Rare *(b), NIST-Bozorth3-*Automatic *(c), NIST-Bozorth3-*Manual *(d).*

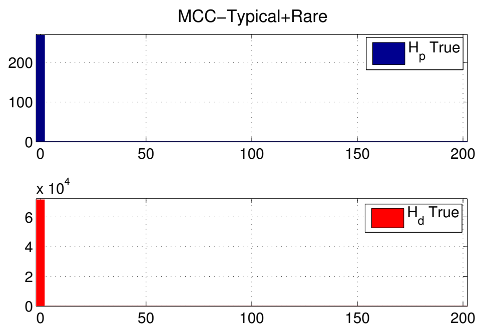


(a) (b) (c) (d)

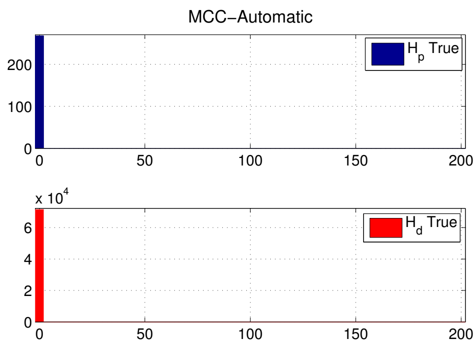
Figure 7.2: *Histograms of pooled scores for (c) and VeriFinger*-Typical *(a), VeriFinger*-Typical+Rare *(b), VeriFinger*-Automatic *(c), VeriFinger*-Manual *(d).*



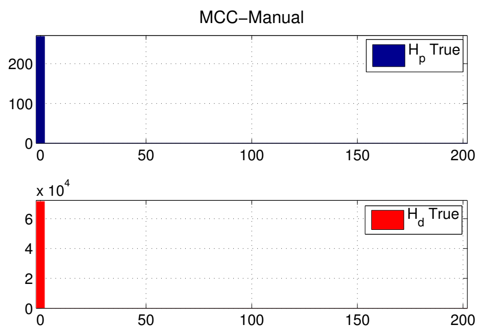
(a)



(b)



(c)



(d)

Figure 7.3: *Histograms of pooled scores for (c) and MCC-*Typical *(a), MCC-*Typical+Rare *(b), MCC-*Automatic *(c), MCC-*Manual *(d).*

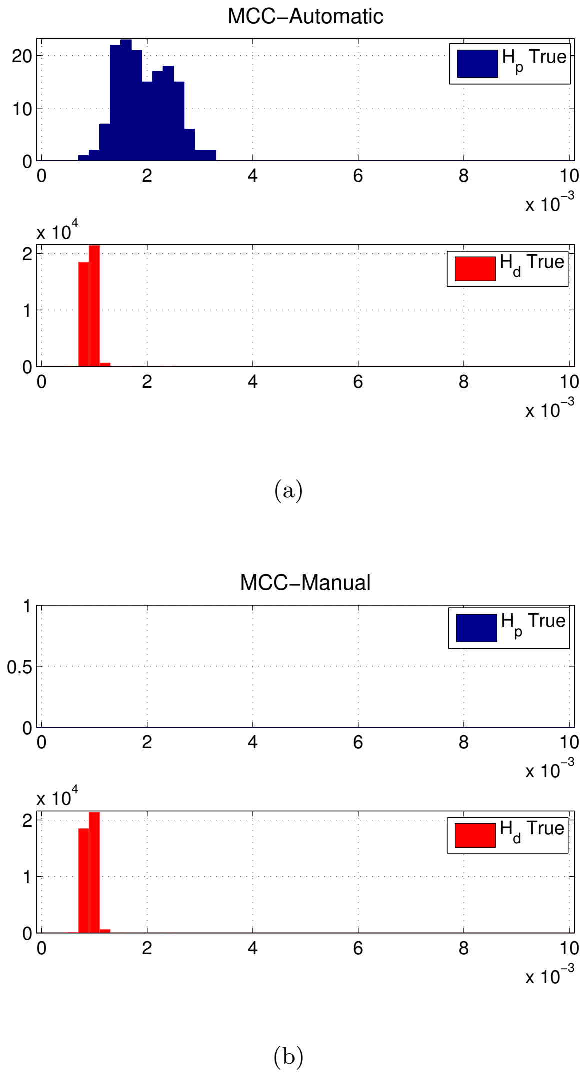


Figure 7.4: *Histograms of pooled scores with detail in scores close to* 0 *for MCC*-Automatic *(a), MCC*-Manual *(b).*

Regarding MCC algorithm scores, another consequence of the distribution of the scores is that it is quite difficult to imagine a parametric model assigned to the scores for *Automatic* and *Manual* approaches. The reason is that there are two clear modes in the impostor distribution mainly, due to the process of altering the scores where rare minutiae are present. As a consequence, parametric models for LR computation like those based on Gaussian distributions cannot be used, because Gaussian pdf is unimodal and must be fit to a multimodal score distribution. Moreover, the use of smooth densities would be problematic, because there are many scores concentrated into a very narrow interval in Automatic and Manual score distributions. This would present problems for approaches such as Gaussian mixture models. This can be solved by the use of Gaussianization techniques like some score normalization methods such as Test-Normalization (T-Norm) [Navratil and Ramaswamy, 2003], as it can be seen below.

7.1.2. Query-by-query analysis of scores

In this section, we analyze the behavior of the score distributions when they are analyzed for each latent fingerprint query separately, for all systems using the MCC minutiae matching algorithm. The aim is to confirm whether the scores are misaligned or not.

Figure 7.5 shows a representation of the scores in a query-by-query basis, for all the four categories to include rare minutiae using our proposed method on the MCC algorithm. It can be seen that the alignment of the impostor scores is moderate for the *Typical* and *Typical+Rare*, although it can be improved. However, the scores for *Automatic* and *Manual* analysis are completely misaligned, due to the procedure used to change the value of the score when rare minutiae are present. This problem of misalignment will be highly problematic for LR models if all scores are pooled together in order to assign probability distributions for the LR. This pooling, as it will be seen in the following section, will be necessary due to the definition of the propositions in the simulated forensic scenario in this Thesis. Therefore, score normalization techniques will be necessary in order to align the impostor distribution of scores, especially for *Automatic* and *Manual* approaches.

7.2. Proposed LR methods

This section proposes several methods for likelihood ratio computation using scores from the MCC algorithm with the Guardia Civil database (GCDB) described in Section 5.1. Several decisions made in this section are justified according to the analysis in Section 7.1.

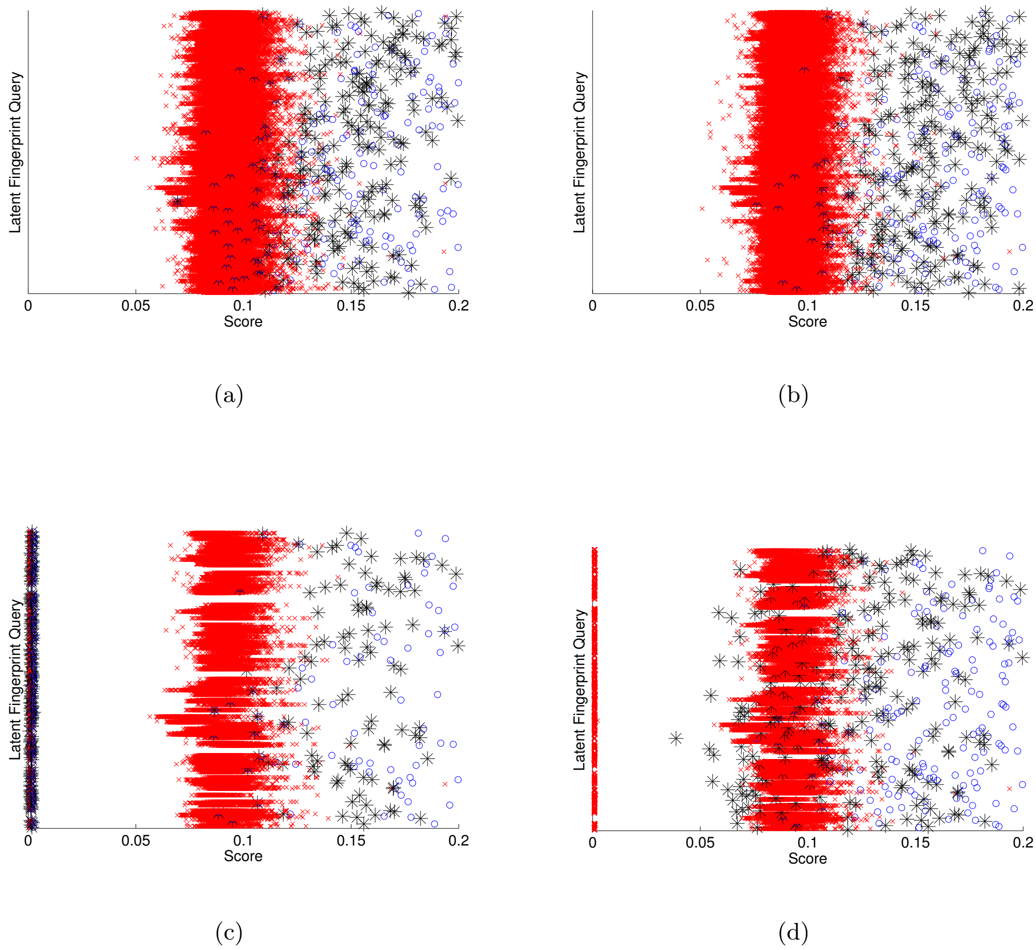


Figure 7.5: *Per-query representation of the scores of the MCC system. Red crosses are always impostor scores and blue circles are genuine scores. X-axis represent the score value, and y axis is discrete and represent the index of each of the latent fingerprint queries. Thus, in each* row *of the graph the genuine and impostor scores of a single query are represented. Black asterisks indicate the threshold of the Equal Error Rate for each query (*row*). Four approaches to include rare minutiae are represented:* Typical *(a),* Typical+Rare *(b),* Automatic *(c) and* Manual *(d).*

7.2.1. Definition of propositions

According to the methodology of CAI, the first step to compute likelihood ratios is to establish the propositions according to the information present in the case.

The *forensic cases* that we are going to simulate here consist of the comparison of one latent fingerprint and one exemplar fingerprint. Both latent and exemplar fingerprint come from GCDB. The scores used to train the models for LR computation are the rest of scores in the GCDB generated from individuals different from the donors of the latent and the exemplar fingerprint. This way, the models are trained with scores not used in the case, and the data handling is honest in the sense of the performance measurement.

According to this set-up, there are several observation that are in order:

- The information in the case is almost non-existent. We only have the images of the latent and the exemplar fingerprint, and therefore no assumption can be done about the donors of latent and exemplar (*e.g.* ethnicity, gender, etc.). This only allow generic proposition about the populations involved.

- The number of the suspect is not known for us. We can only say that there is a latnet and a exemplar fingerprint, but we will not have a name of the donor of the print (due to privacy issues).

- We only have a single genuine comparison for each subject in the database. Therefore, it is impossible for us to focus in models aimed at the suspect, because there is no additional data available to model the particular behavior of their scores in comparison to the whole population of scores.

- There is no information whatsoever about the relevance of the donor of latent and exemplar fingerprint with respect to the action in the crime scene, or even more with respect to any offense. Therefore, only propositions at source level can be addressed.

- Because of the way it was built, we assume that all latent fingerprints in the GCDB dubbed as different in the ground-truth labels are generated by different people. It is assumed also in the corresponding exemplar fingerprint. Therefore, in this database it will be equivalent to talk about donors as about fingers, since different fingerprints will definitely belong to different donors (and not to different fingers of the same donor).

Under this premises, we decide to state source-level, person-generic and general-population propositions for this case. Therefore, we have the following propositions:

H_p: The finger that originated the latent fingerprint is the same finger that originated the exemplar fingerprint.

H_d: The donor that originated the latent fingerprint is not the suspect, but other donor from a wide population of individuals whose characteristics have not been specified.

This definition of propositions implies that, for a forensic case involving the comparison of a latent fingerprint and its corresponding exemplar fingerprint, the scores needed to train the LR model should be generated with comparisons of latents and exemplar fingerprints without the constrain of belonging to a particular individual. This implies that more scores will be typically available to train the models, therefore improving their statistical robustness. On the other hand, the use of person-generic propositions inevitably implies an important loss of information in cases where the identity of the individual is known, as it is typical in court. However, for this Thesis we will consider this person-specific scenario because of the limitations of the GCDB, as explained above.

7.2.2. T-Norm score normalization

Here we describe the score normalization method that is used in this Thesis in order to align the scores in a query-by-query basis.

Score normalization is defined as a transformation to the output scores of a biometric system in order to reduce misalignment in the score ranges due to variations in the conditions of a comparison. We may classify score normalization techniques into: *i*) reference-dependent, when the variability is compensated for reference (exemplar) fingerprint; and *ii*) query-dependent, when the variability is compensated for the query latent fingerprint.

Many score normalization techniques have been presented in the literature. The most popular and widely-used family of normalization techniques is the so-called impostor-centric [Fierrez-Aguilar *et al.*, 2005], where the normalization parameters are estimated from score distributions where H_d is true. In this Section we describe Test-Normalization (T-Norm) as one of the most popular impostor-centric score normalization techniques.

Test-Normalization, or T-Norm [Auckenthaler *et al.*, 2000], exploits the idea that different query latent fingerprint can present different behavior in terms of the range of the scores generated. In this case, impostor scores are generated from a given query, and therefore a distribution of impostor scores for that particular query can be assigned. In order to do that, a set of impostor fingerprints, namely a *cohort* of impostors, is needed. From those so-called T-Norm scores, the mean and the standard deviation μ_{Tnorm} and σ_{Tnorm} are computed. The T-Norm technique is then applied to a particular score computed form that latent fingerprint query as follows:

$$s_{Tnorm} = \frac{s_{raw} - \mu_{Tnorm}}{\sigma_{Tnorm}} \tag{7.1}$$

Thus, T-Norm performs query-dependent score normalization, and the result is the alignment of the query-dependent impostor score distributions for all comparisons in the particular set of scores. Thus, this normalization technique compensates variability in the scores due to the recovered latent fingerprint.

T-Norm presents a key advantage for AFIS technology (like the one developed in this Thesis) with respect to other score normalization methods: the use of T-Norm does not change the CMC characteristic of AFIS technology, and therefore it does not degrades it performance. This is explained as follows. T-Norm applies a linear transformation to all the scores generated from the same query latent fingerprint. Therefore, the query-by-query alignment improves, but the discriminating power of the scores for each query does not change. This means that the CMC curve of a set of scores which uses T-Norm will not change at all, because it is computed considering latent fingerprint queries separately. As a consequence, we can safely apply T-Norm to any AFIS systems without the risk of degrading CMC performance.

Although many advantages of T-Norm are well known in areas like speaker recognition [Auckenthaler *et al.*, 2000; Navratil and Ramaswamy, 2003], it has also some disadvantages. The main one is that it needs a cohort of fingerprints that matches as much as possible the conditions of the fingerprints in the database to search. This has two consequences: on the one hand, some additional data is needed[1]; on the other hand, the more divergence of the conditions of the cohort with respect to those of the reference fingerprints in the cases, the lower the benefits of T-Norm. Considering this, it is in order to warn that the T-Norm cohort in this Thesis has been selected from the same Guardia Civil database that has been used to simulate real forensic latent-exemplar fingerprint comparisons, and therefore the results can be optimistic. Another disadvantage of T-Norm is that it implies the generation of an additional number of comparisons equal to the size of the cohort, and therefore it increases the computational burden of each query.

7.2.3. Likelihood Ratio models

From the analysis performed above, in this Thesis we proposed 4 models for likelihood ratio computation.

- Pool Adjacent Violators (PAV) calibration applied to scores from fingerprint matchers directly (namely *PAV*).

[1] A cohort size of minimum of 50 fingerprints is typically recommended [Auckenthaler *et al.*, 2000].

- Pool Adjacent Violators calibration applied to scores after T-Norm score normalization (namely *T-Norm + PAV*).
- Gaussian-ML density models applied to scores after T-Norm score normalization (namely *T-Norm + Gaussian-ML*).
- Logistic regression applied to scores after T-Norm score normalization (namely *T-Norm + LR*).

7.3. Experimental results

7.3.1. Experimental protocol

The experimental protocol has been designed in order to simulate a real forensic scenario where latent fingerprints are compared with exemplar fingerprints using typical minutia features and also rare minutia features. In our experiments, we used the Guardia Civil database (as described in Section 5.1), because it is the only forensic fingerprint database which contains rare minutiae, as it has been previously described. Since the GCDB is limited in size, a cross-validation strategy has been followed in order to optimally use the data without using the same dataset to train and test the LR models proposed. This cross-validation strategy is described as follows: for each genuine comparison of a latent fingerprint and a mated exemplar fingerprint, the scores to train the LR model for that particular comparison will consist of all the scores generated with the GCDB, except those generated with either the latent or the exemplar fingerprint involved in the case. Therefore, we guarantee separation between the latent and exemplar fingerprint sources and the individuals in the training database.

This cross-validation strategy has many advantages in the sense of the optimal usage of the available database. However, it also presents the disadvantage that the conditions of the training scores matches the conditions of the latent and exemplar fingerprint under comparison to a higher degree than in a potential real case. Thus, the results presented here could be optimistic. However, due to the limitation of the database, and also because the aim of the work is to show how to apply the methodology and to illustrate the improvements due to rare minutiae, we consider it appropriate to use this protocol.

Notice that this cross-validation strategy not only guarantees that the identities in the latent and exemplar fingerprints in the training and testing databases are not the same. Moreover, it also guarantees that the T-Norm scores generated with the cohort are not present in the training database. This is because the T-Norm cohort scores must be generated with the scores of the query latent fingerprint, which will be not present in the training database. Therefore, the situation is realistic in the sense of the handling of the data to normalize the scores and also to train the LR models.

7.3.2. Results on the effects of T-Norm

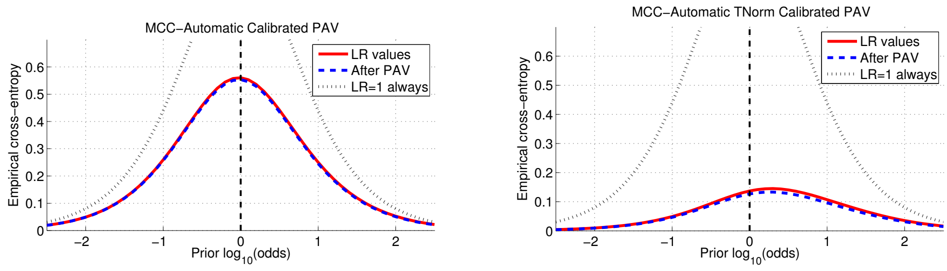
In this section, results are shown in order to confirm the hypothesis that T-Norm is useful for likelihood ratio computation. Thus, in this section the MCC systems without T-Norm will be used as baseline, whereas the MCC systems with T-Norm are the proposed improvements. As only PAV LR computation method makes sense without T-Norm, we will only use this LR method, and we compare it with other methods in the next sections.

Figure 7.6 shows the ECE plots of the LR values computed for the PAV method before and after T-Norm, for all the four categories presented with MCC. It is observed that a significant improvement in the discriminating power (reduction in the blue dashed curve) has been introduced by the use of T-Norm in all cases.

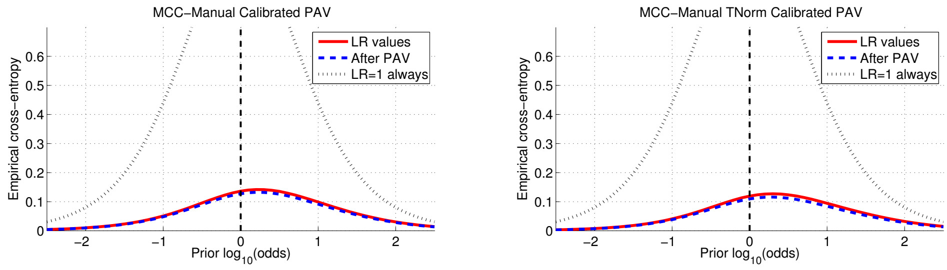
There are two explanations for the nice improvement introduced by T-Norm. The first one related to the misalignment between the different latent fingerprint queries. This can be observed in Figure 7.7, where scores are represented for each query, before and after T-Norm, for the MCC-*Automatic* system. It is clearly shown that the scores are misaligned for different queries if T-Norm is not applied. After T-Norm, the alignment of the impostor scores improves severely. The second explanation is the gaussianization of the histogram of scores when they are pooled together for all queries. Figure 7.8 shows this also for the MCC-*Automatic* system: it is clearly seen that before T-Norm the histogram looks multimodal, and after T-Norm it adopts a form that can be associated much better to a gaussian pdf, especially for the impostor scores. This *Gaussianization* of the scores allows models like Gaussian-ML and logistic regression to better model the scores, and therefore the calibration performance is expected to improve for those models. In fact, without such a T-Norm Gaussianization of scores, Gaussian-ML and logistic regression must present a much worse performance, and we have not considered them.

Interestingly, it is important to notice that T-Norm does not change the CMC characteristic of the matcher by definition, as previously said. Consequently the performance in identification mode is the same with and without T-Norm. Because of this, it can be thought that, because of the nice properties of T-Norm regarding discriminating power and also for likelihood ratio computation, T-Norm should be used as a recommended stage in all AFIS systems. However, if identification mode is to be used, and computational efficiency is critical, then T-Norm should be suppressed, since it means some computational burden for each query comparison with no CMC performance improvement.

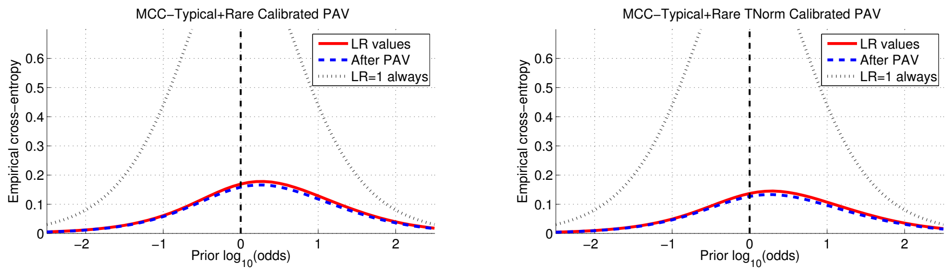
Due to the nice properties of T-Norm for finger-generic likelihood ratio computation, we will use it in all the subsequent experiment in this chapter.



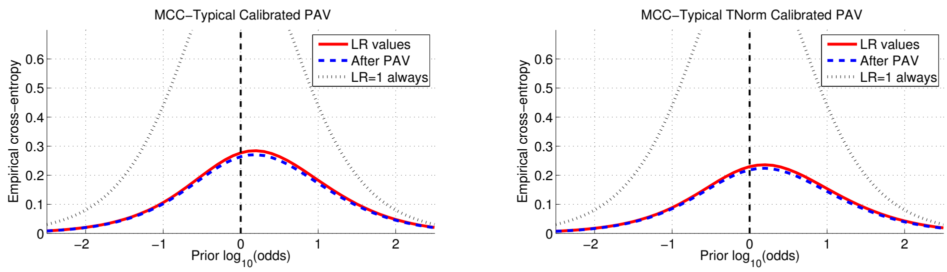
(a)



(b)

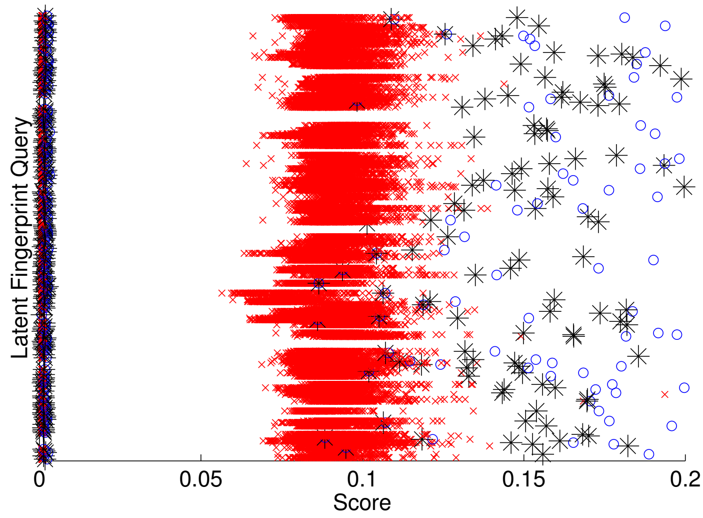


(c)

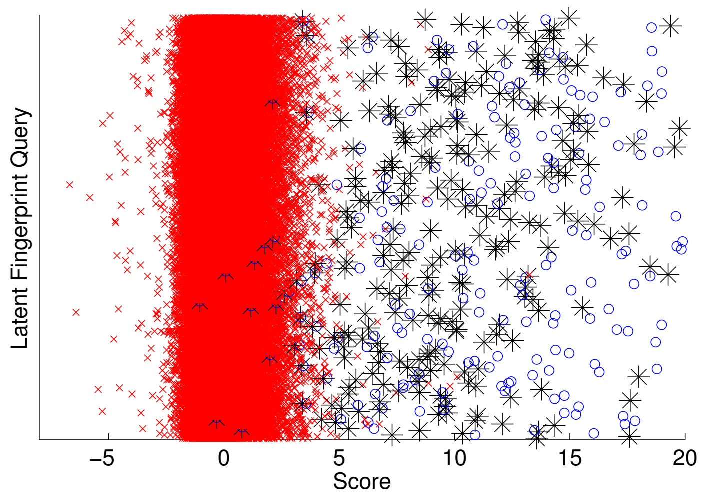


(d)

Figure 7.6: *ECE plots showing performance of PAV calibration method before T-Norm (left) and after T-Norm (right), for MCC*-Automatic *(a), MCC*-Manual *(b), MCC*-Typical+Rare *(c) and MCC*-Typical *(d).*

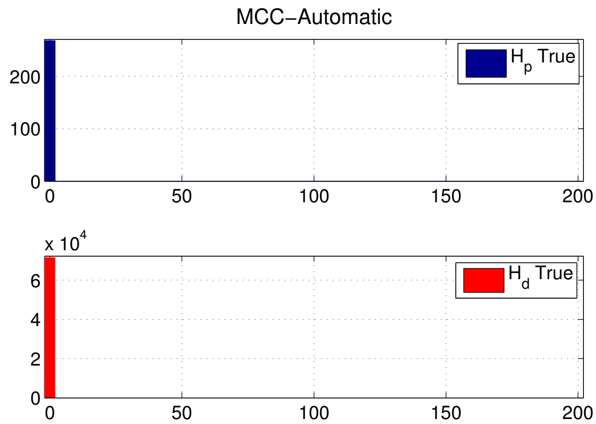


(a)

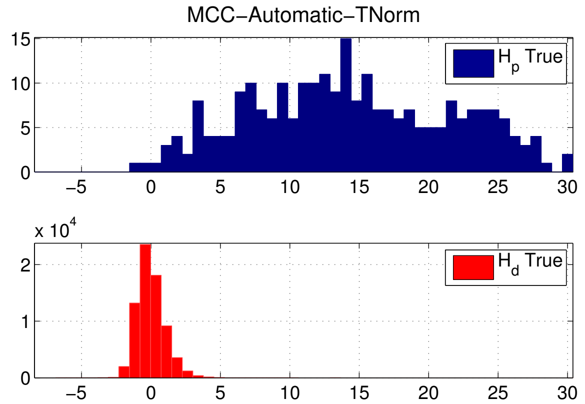


(b)

Figure 7.7: *Per-query representation of the scores of the MCC-*Automatic *system. Red crosses are always impostor scores and blue circles are genuine scores. X-axis represent the score value, and y axis is discrete and represent the index of each of the latent fingerprint queries. Thus, in each* row *of the graph the genuine and impostor scores of a single query are represented. Black asterisks indicate the threshold of the Equal Error Rate for each query (*row*). Scores without T-Norm are in (a), and scores after T-Norm are in (b).*



(a)



(b)

Figure 7.8: *Histograms of scores of the MCC-*Automatic *system. Scores without T-Norm are in (a), and scores after T-Norm are in (b).*

7.3.3. Results on the improvement due to rare minutiae

In this section we compare the improvement in the performance of all the proposed likelihood ratio computation methods with T-Norm, due to the inclusion of rare minutiae. Again, as previously addressed, we will use the MCC-*Typical* system as the baseline, because it does not include rare minutiae information. Also, we will present the MCC-*Manual* as the ceiling of performance when rare minutiae are considered. Finally, MCC-*Automatic* and MCC-*Typical+Rare* are considered here as the proposed systems. Moreover, we will show the improvement introduced by the use of rare minutiae for all the methods proposed with T-Norm, namely PAV, Gaussian-ML and logistic regression.

Figure 7.9 shows the performance in the form of ECE plots for all the LR methods proposed and T-Norm scores. The Figure shows the baseline (without rare minutiae) on the left column, the ceiling (rare minutiae are manually considered) on the right and the proposed systems in the two central columns.

For all the methods, ECE plots show that the discriminating power (blue dashed line) improves by the inclusion of rare minutiae. It is also seen that the two systems automatically including rare minutiae approaches the ceiling of discrimination performance for all LR methods.

As a conclusion for this section, it has been shown that all the methods benefit from rare minutiae in the sense of the discriminating power, confirming the hypothesis throughout this Thesis.

7.3.4. Results on the comparison of LR computation methods

In this section we compare all the proposed LR computation methods not only from the perspective of the discriminating power, but also with respect to the calibration loss. Thus, accuracy as the sum of both performance measure will allow us to select the best choice for LR computation.

From Figure 7.9, the two central columns are the LR computation methods that are realistic to be performed automatically, as previously mentioned. From that Figure, it is therefore seen that the logistic regression model presents the best accuracy (red solid curve) both for MCC-*rare+Typical* and MCC-*Automatic* systems, and therefore this seems to be the best choice.

We now analyze calibration (separation between red and blue curves) more deeply. It is generally seen in Figure 7.9 that the calibration loss is preserved with the inclusion of rare minutiae for PAV and logistic regression methods, making both of them apparent good options for LR computation.

An additional warning is in order here. The cross-validation procedure to train LR models and to select T-Norm scores implies a scenario with low dataset shift between training and testing data. In a forensically realistic set-up, where dataset shift between training and testing data can be severe, the performance of LR methods that excessively fits the training data can seriously degrade. On the other hand, it is known in pattern recognition that models with lower complexity are more robust to this effect to avoid overfitting. Therefore, the much lower complexity of logistic regression with respect to PAV indicates that the former can be potentially more robust to overfitting and dataset shift than the latter in forensically realistic conditions. Due to this reason, logistic regression is more preferable to PAV computation method.

Regarding the Gaussian-ML method, it is seen that the calibration performance is much worse than for PAV or logistic regression. Moreover, it seriously degrades for MCC-*Automatic* scores, being even worse than for the neutral reference. The reason is that, due to the *Automatic* process described before to incorporate rare minutiae information to the raw scores, by means of multiplying a fixed value to the scores, the distribution of scores for each latent fingerprint query strongly deviates from Gaussianity. Moreover, although T-Norm Gaussianizes the impostor distribution of scores when they are pooled among all queries, it is not the case for the genuine scores, and this makes the genuine pooled distribution to seriously diverge from Gaussianity even after T-Norm is applied. As a consequence, it is strongly discouraged to use Gaussian-ML models to compute LR values with MCC-*Automatic* scores, even after T-Norm.

As a conclusion of this section, the calibration loss represents a low percentage of all the loss of accuracy for logistic regression and PAV LR computation methods, in this order. This makes the overall performance of logistic regression superior, which among other reasons makes it the best choice. On the other hand, Gaussian-ML presents higher calibration loss, sometimes presenting worse performance than the neutral reference, which makes it not recommendable for the score computation systems proposed in this Thesis.

7.4. Discussions

In this chapter, we have explored various methods to compute likelihood ratios from the scores generated by the fingerprint recognition systems studied in this Thesis. This has allowed the interpretation of the evidence from the latent-to-exemplar fingerprint comparisons simulating a real forensic case by the use of a cross-validation strategy with the GCDB. First, an analysis of the systems has been done in order to propose the strategies for LR computation. Then, several models have been proposed and compared in terms of discriminative power and calibration performance. Those results clearly show the improvement of rare minutiae in the discrimination performance of the forensic LR methods, where most of them present calibration performance far better than the neutral reference.

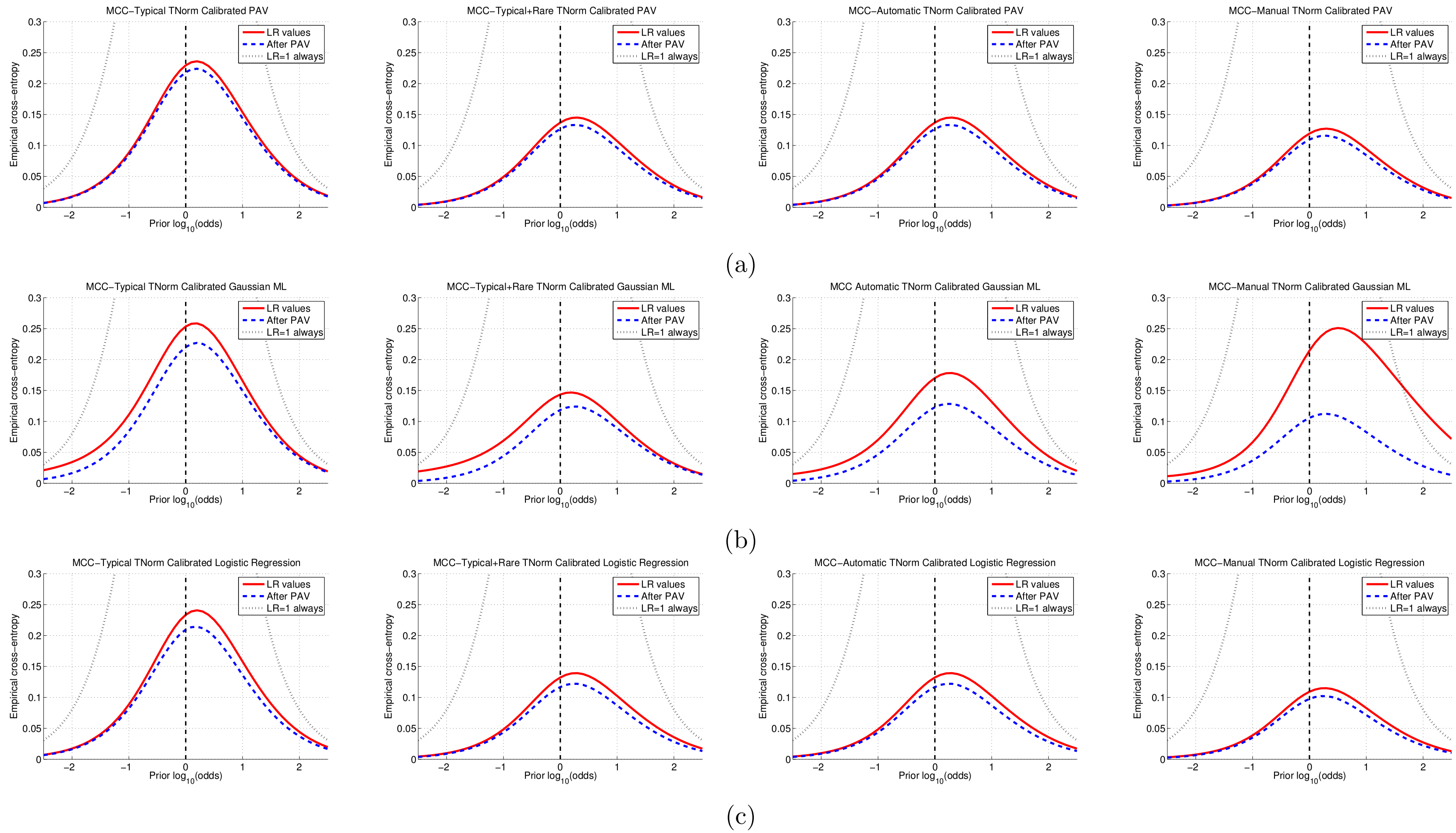


Figure 7.9: *ECE plots showing performance of LR methods with T-Norm scores. From left to right:* Typical *(baseline),* Typical+Rare,Automatic *and* Manual *(Ceiling). The different LR methods are PAV (a), Guassian-ML (b) and Logistic Regression (c).*

The main conclusion of this chapter is that AFIS technology scores are appropriate to generate a candidate list for database search in forensic investigation. However, they cannot be directly used in order to compute likelihood ratios by using standard methods in the continuous score domain. The main reasons for this are as follows:

- Scores from fingerprint recognition systems might present distributions that do not resemble typical parametric probability density functions. Moreover, an important proportion of the scores generated by the systems (both genuine an impostor) might be concentrated into a single value (or a small range of values, like zero, in the case of some of the systems analyzed in this thesis). Therefore, the direct use of standard algorithms like kernel density functions, Gaussian modeling or logistic regression will lead to seriously degraded performance.

- Scores from fingerprint recognition systems can present a very good CMC performance, but the misalignment between scores for each latent fingerprint query can be severe. Therefore, if LR computation is computed from a pool of training scores from different queries (as it is the case in this Thesis), this will lead to a dramatic degradation in the discriminating power.

Some solutions to these problems explored in this Thesis are described below:

- Non-parametric techniques like PAV have been used in order to tackle the problem with the distributions of the scores, and even the concentration of the scores in particular values.

- Score normalization has been used in order to correct the misalignment of the scores from different queries.

- The selected score normalization technique, namely T-Norm, has the property of Gaussianizing the distribution of the scores pooled for all latent fingerprint queries. Therefore, this allows the use of parametric techniques like Gaussian models or logistic regression, the latter outperforming the rest of LR methods proposed.

Chapter 8

Conclusions and Future Work

We have addressed some of the challenges in the current state-of-the-art of automated latent fingerprint matching, and towards individualization of latent fingerprints. After providing a summary of the state-of-the-art in latent fingerprint matching, we present algorithms and models to improve the rank identification accuracies of minutiae-based matchers by proposing an algorithm to perform pre-registration of partial latent fingerprints, and by proposing a model to incorporate reliably extracted rare minutiae features. We experimentally demonstrate the significant improvement in the rank identification accuracies of minutiae-based matchers when using our proposed algorithm, as well as the feasibility of our algorithm as a fully automatic tool. From the model proposed for incorporating rare minutiae features, we also studies various evidence evaluation models based on likelihood ratio.

8.1. Conclusions

Chapter 1 introduces about various types of forensic evidences generally involved in the forensic sciences. This is followed by an overview of the latent fingerprints, types of testimony standards to be followed for general acceptance of the evidence in courts (Frye and Daubert), an overview of the current practices in friction ridge analysis, the challenges faced in the individualization of fingerprints, and about the use of newer technologies to reduce human-errors in the examination process. We discuss the main recommendations put forward by the forensic community to improve the friction ridges analysis, followed by the motivation of this Thesis, and the research contributions originated from this Thesis.

Chapter 2 gives an overview about automated latent fingerprint matching systems, and its drawbacks in forensic applications. We reviewed about different types of fingerprint matching techniques currently employed by automated fingerprint matchers, and also the importance of alignment as an important pre-processing stage in matching for improved performance. We reviewed about various alignment techniques, and their limitations to be used in partial to full fingerprint alignment. We then discussed about the discriminating ability of orientation fields in fingerprints, and how they are robust against fingerprint image quality. An overview of the proposed orientation field based partial fingerprint registration was given together with the database used in the analysis for pre-registration algorithm.

Chapter 3 described in details the hierarchical algorithm to register the orientation field of partial fingerprint into the orientation field of full fingerprint. Through pre-registration, we were able to reduce the minutiae search space of the full fingerprint approximately to the size of the partial fingerprint minutiae set which contributed towards significant improvement in the rank identification accuracies of the minutiae-based matchers. Through these experiments of pre-registration, we were also able to study from various orientation field estimation techniques, the best representative orientation fields for both tenprint and latent fingerprints. We also observed that for a large quantization step in the rotation alignment, we have not degraded the performance very much, and while matching, we have reduced the size of the minutiae search space in the tenprint to good extent which accounts for overall efficiency of our proposed method. Also, we have established the feasibility of our method as a fully automatic tool.

Chapter 4 described the importance of Extended Feature Sets (EFS) towards improving the identification accuracies of minutiae-based fingerprint matchers, and details about the real forensic fingerprint casework database obtained from Guardia Civil which consists of rare minutiae features. Chapter 5 described in details the proposed algorithm to improve the identification accuracy of minutiae-based matchers for partial latent fingerprints by incorporating reliably extracted rare minutiae features. The improvement in the identification accuracy for matchers are achieved by modifying the similarity scores of matcher based on the decision yielded by our algorithm. The decision for a match or non-match was automatically estimated based on least squares fitting error of an affine transformation that transforms latent minutiae set onto tenprint minutiae set with the rare minutiae as the reference point. The usefulness of the proposed model is demonstrated on two widely used minutiae-based matchers, NIST-Bozorth3 and VeriFinger. Both matchers showed significant improvements in the rank identification accuracies when their similarity scores were modified based on the fitting error proposed in our methodology. We conclude that even if we have only few number of minutiae in a partial latent, presence of reliably extracted rare minutiae features makes the comparison more robust.

Chapter 6 addressed the issue of interpretation of forensic evidence from scores generated by biometric systems. We described the importance of evidence evaluation, and addressed it using the method previously developed for generating the modified similarity scores based on rare minutiae features. In Chapter 7, we showed that though AFIS scores can be directly appropriate to generate a ranked list of candidates from the database, but they cannot be used directly for evidence evaluation. Score normalization has been used in order to correct the misalignment of the scores from different queries. We experimentally demonstrated that the selected score normalization technique, namely T-Norm, allows the use of parametric techniques like Gaussian models or logistic regression, the latter outperforming the rest of LR methods proposed.

Summarizing, the main results and contributions obtained from this Thesis are:

- New correlation-based hierarchical registration method for orientation images to register a partial fingerprint in a full fingerprint.
- The best representative orientation fields for both full fingerprint and partial fingerprint have been experimentally demonstrated.
- A methodology to adapt any minutiae-based matcher by incorporating information from rare minutiae features.
- A specific algorithm to align the latent minutiae pattern and the tenprint minutiae pattern using rare minutiae features.
- Presenting population statistics about rare minutiae features present in a realistic forensic fingerprint casework database.
- Selection of normalization techniques to correct the misalignment of the similarity scores generated by fingerprint matchers from different queries.

8.2. Future Work

A number of research lines arise from the work conducted in this Thesis. We consider of special interest the following ones:

- The thesis has considered the use of dictionary based orientation estimation technique in which the orientations are corrected on a global level. A possible improvement to the latent orientation field estimation will be to use localized dictionary based orientation field for latents [Yang *et al.*, 2014], and see the overall improvement in the performance of minutiae-based matchers.
- More robust orientation image comparison methodology which not only emphasis on orientation of fingerprint image, but also takes into account the quality of the fingerprint image for each block of the orientation estimated from the gray scale fingerprint image.
- Developing techniques to automatically estimate rare minutiae features from high resolution fingerprint image will be a useful addition to the system developed for extended feature sets in this Thesis.
- The entropy-based measure estimated from the probabilities of the rare-minutiae features were used to reward the similarity score of the minutiae-based matcher. Other techniques can be used to obtain the weights to penalize the score.

- Generate models more adapted to the casuistic of AFIS scores, in order to obtain improved performance. Some previous work on this topic can be found in [Haraksim *et al.*, 2015b].
- Apply Bayesian methods to evidence evaluation in order to obtain more robust and coherent LR values. Some previous work on this can be found in [Brummer and Swart, 2014].
- Generate or simulate databases in order to address more specific propositions that can more realistically reflect the typical forensic practice.

Appendix A

Conclusiones y Trabajo Futuro

En este trabajo se han abordado retos importantes dentro del estado-del-arte del reconocimiento automático de huellas latentes. Después de resumir el estado-del-arte en reconocimiento de huella latente, se han presentado algoritmos y modelos para mejorar las métricas de rendimiento de sistemas basados en comparación de minucias. Los algoritmos propuestos se basan en modelos de pre-alineamiento de huellas y metodologías para incorporar minucias de baja probabilidad de aparición (*rare minutiae* en inglés). Se ha demostrado experimentalmente la conveniencia de las técnicas propuestas así como su integración en sistemas automáticos. Adicionalmente se han estudiado varios modelos de verosimilitud para incorporar las minucias de baja frecuencia de aparición como evidencia pericial en cotejo de huellas.

A.1. Conclusiones

En el Capítulo 1 se han introducido diferentes tipos de evidencias habitualmente utilizadas en la ciencia forense. A continuación se ha realizado un resumen sobre la utilidad de las huellas latentes, tipos de testimonios más populares y comúnmente aceptados en cortes penales, técnicas más comunes en la comparación de huellas basadas en patrones de cresta, retos en la individualización de huellas y uso de nuevas tecnologías para reducir las tasas de errores humanas. Se han discutido las principales recomendaciones de la comunidad forense para mejorar el cotejo de huellas para finalmente presentar las motivaciones y contribuciones de esta Tesis.

El Capítulo 2 presenta una visión de los sistemas automáticos de comparación de huellas dactilares y su utilidad en escenarios forenses. Se evalúan las técnicas más populares de comparación del estado del arte y se analiza la importancia del alineamiento como fase previa para mejorar las métricas de rendimiento. Seguidamente se evalúan diferentes técnicas de alineamiento.

El Capítulo 3 describe en detalle los algoritmos de alineamiento del mapa de orientación de huellas latentes propuestos. A través del alineamiento se reduce el espacio de búsqueda de coincidencias entre minucias a un tamaño similar al de la huella latente parcial. Esta disminución del espacio de búsqueda permite eliminar gran cantidad de falsos positivos y mejorar así el rendimiento de los sistemas de comparación basados en minucias. A través de estos experimentos de alineamiento de mapas de orientación, también se ha estudiado la mejor forma de representar mapas de orientación de huellas latentes y sus respectivas impresiones. También se ha probado la integración de nuestro método en herramientas de cotejo completamente automáticas.

El Capítulo 4 describe la importancia de los conjuntos extendidos de características (EFS en inglés) para la mejora de algoritmos de comparación de minucias y se presenta la base de datos de casos forenses reales cedida por la Guardia Civil para la evaluación de los modelos propuestos en esta Tesis. El Capítulo 5 describe en detalle la metodología propuesta para incorporar las minucias de baja probabilidad de aparición para la mejora del rendimiento de comparadores automáticos. La mejora en el rendimiento se obtiene a través de la ponderación a través del algoritmo propuesto de los valores de similitud del clasificador. La decisión de coincidencia o no coincidencia se realiza a través de un ajuste automático del error de la transformación afín entre diferentes conjuntos de características. La utilidad de la metodología se demuestra a partir de dos sistemas populares de comparación de minucias, NIST-Bozorth3 y VeriFinger. Ambos sistemas muestran una clara mejora cuando se aplica la metodología propuesta incluso cuando el conjunto de minucias de baja probabilidad de aparición es pequeño.

El Capítulo 6 aborda la interpretación de la evidencia forense a partir de valores de similitud generados con un sistema de reconocimiento biométrico. Se describe la importancia de la evaluación de la evidencia, y se analiza la metodología propuesta anteriormente. En el Capítulo 7 se muestra cómo los valores de similitud de un sistema de reconocimiento automático de huella dactilar pueden ser utilizados para generar una lista de candidatos pero no pueden ser usados directamente como evidencia forense. Se propone la normalización de resultados basados en T-norm para corregir errores de alineamiento de las distribuciones de similitud.

En resumen, los principales resultados y contribuciones obtenidos durante el desarrollo de esta Tesis son:

- Nueva metodología de alineamiento de huellas latentes basada en correlaciones jerárquicas.
- Se ha demostrado experimentalmente su utilidad tanto para huellas latentes completas como parciales.
- Metodología para incorporar minucias de baja probabilidad de aparición en comparadores automáticos de minucias.
- Algoritmos de alineamiento de patrones de minucias usando minucias de baja probabilidad de aparición.
- Estadísticas poblaciones relacionadas con las minucias de baja probabilidad de aparición en bases de datos forenses relativas a casos reales.

- Técnicas de normalización para corregir des-alineamiento en las distribuciones de valores de similitud de comparadores de huellas dactilares.

A.2. Trabajo futuro

Existen múltiples vías para continuar el trabajo presentado en esta Tesis. Consideramos de especial interés las siguientes:

- Mejora de los mapas de orientaciones a través de nuevas técnicas de aprendizaje automático basadas en el uso de diccionarios [Yang *et al.*, 2014]. Estos mapas mejorados podrán ser utilizados para mejorar el pre-alineamiento y por tanto el rendimiento de los sistemas.
- Técnicas más robustas de comparación de mapas de orientación que incorporen características como la calidad de la huella en cada bloque de orientación estimado de la imagen en escala de grises.
- Desarrollo de técnicas de detección automáticas de minucias de baja probabilidad de aparición a partir de imágenes de alta resolución.
- Generación de modelos adaptados a la casuística de los sistemas de comparación automática de huellas que puedan mejorar las métricas de rendimiento. Un trabajo en esta línea es [Haraksim *et al.*, 2015b].
- Aplicaciones de modelos Bayesianos para la evaluación de la evidencia que permitan mejorar la robustez de los valores de verosimilitud. Un trabajo en esta línea es [Brummer and Swart, 2014].
- Generación de bases de datos que reflejen mejor la realidad de las prácticas forenses.

References

C. Aitken, C. E. Berger, J. S. Buckleton, C. Champod, J. Curran, and A. P. Dawid. Expressing evaluative opinions: A position statement. *Science and justice*, 2011. 19

C. G. G. Aitken and F. Taroni. *Statistics and the Evaluation of Evidence for Forensic Scientists*. John Wiley & Sons, Chichester, 2004. 15, 95, 99

F. Alonso-Fernandez, J. Fierrez, and *et al.* A comparative study of fingerprint image-quality estimation methods. *IEEE Trans. on Information Forensics and Security*, 2(4):734–743, 2007. 48, 51

D. R. Ashbaugh. Quantitative-qualitative friction ridge analysis: an introduction to basic and advanced ridgeology. *Boca Raton: CRC press*, 1999. 70

R. Auckenthaler, M. Carey, and H. Lloyd-Tomas. Score normalization for text-independent speaker verification systems. *Digital Signal Processing*, 10:42–54, 2000. 116, 117

A. M. Bazen and S. H. Gerez. Systematic methods for the computation of the directional fields and singular points of fingerprints. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):905–919, 2002. 51

J. Bigun. Vision with Direction: A Systematic Introduction to Image Processing and Computer Vision. *Springer*, 2005. 42

R. M. Bolle, J. H. Connell, S. Pankanti, N. K. Ratha, and A. W. Senior. The relation between the ROC curve and the CMC. *4th IEEE Workshop In Automatic Identification Advanced Technologies*, pages 15–20, 2005. 16

L. Brown. A survey of image registration techniques. *ACM computing surveys*, 24(4):325–376, 1992. 35

N. Brümmer, L. Burget, J. Cernocky, O. Glembek, F. Grezl, M. Karafiat, D. A. van Leeuwen, P. Matejka, P. Scwartz, and A. Strasheim. Fusion of heterogeneous speaker recognition systems in the STBU submission for the NIST speaker recognition evaluation 2006. *IEEE Transactions on Audio, Speech and Signal Processing*, 15(7):2072–2084, 2007. 104, 105

N. Brümmer and J. du Preez. Application independent evaluation of speaker detection. *Computer Speech and Language*, 20(2-3):230–275, 2006. 15, 96, 101, 103, 104, 105

N. Brummer and A. Swart. Bayesian calibration for forensic evidence reporting. *Interspeech*, 2014. 130, 133

ByronCase. United States of America vs Byron Mitchell, Action No. 96-407. *U. S. District Court Eastern District of Pennsylvania*, 1999. 6

R. Cappelli, M. Ferrara, and D. Maltoni. Minutia Cylinder-Code: a new representation and matching technique for fingerprint recognition. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 32(12):2128–2141, 2010. 54

R. Cappelli, M. Ferrara, and D. Maltoni. Fingerprint Indexing based on Minutia Cylinder Code. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5):1051–1057, 2011. 54

C. Champod, I. W. Evett, and G. Jackson. Establishing the most appropriate databases for addressing source level propositions. *Science and Justice*, 44(3):153–164, 2004. 98

C. Champod and D. Meuwly. The inference of identity in forensic speaker recognition. *Speech Communication*, 31:193–203, 2000. 99, 100

Y. Chen and A. Jain. Dots and incipients: extended features for partial fingerprint matching. *In IEEE Biometrics Symposium*, 2007. URL http://fingerprint.nist.gov/STANDARD/cdeffs/Docs/IAI_CDEFFS_2007-07-24.pdf. 76

S. Cole. More than zero: Accounting for error in latent fingerprint identification. *Journal of Criminal Law and Criminology*, 2005, 985-1078. 6

R. Cook, I. W. Evett, G. Jackson, P. J. Jones, and J. A. Lambert. A hierarchy of propositions: deciding which level to address in casework. *Science and Justice*, 38(4):231–239, 1998a. 97, 98

R. Cook, I. W. Evett, G. Jackson, P. J. Jones, and J. A. Lambert. A model for case assessment and interpretation. *Science and Justice*, 38:151–156, 1998b. 95, 97

N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet. Front-End Factor Analysis for Speaker Verification. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(4):788–798, 2010. 96

R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley, 2001. 101

V. N. Dvornychenko and D. Garrism. M. Summary of NIST latent fingerprint testing workshop. NISTIR 7377. 2006. URL http://fingerprint.nist.gov/latent/ir_7377.pdf. 7, 70

N. Egli. *Interpretation of Partial Fingermarks Using an Automated Fingerprint Identification System*. PhD thesis, Institute de Police Scientifique, Ecole de Sciences Criminelles, 2009. 97

I. W. Evett. Towards a uniform framework for reporting opinions in forensic science casework. *Science and Justice*, 38(3):198–202, 1998. 99

I. W. Evett, G. Jackson, and J. A. Lambert. More on the hierarchy of propositions: Exploring the distinction between explanations and propositions. *Science and Justice*, 401(1):3–10, 2000. 97

G. Fang, S. Srihari, H. Srinivasan, and P. Phatak. Use of ridge points in partial fingerprint matching. *Proc. of SPIE: Biometric Technology for Human Identification IV*, 2007. 12, 18, 34

FBI-IAFIS. Integrated Automated Fingerprint Identification System. URL http://www.fbi.gov/about-us/cjis/fingerprints_biometrics/iafis/iafis. 31

J. Feng and A. Jain. Fingerprint reconstruction: from minutiae to phase. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(2):209–223, 2011. xxv, 48, 50, 51

J. Feng, J. Zhou, and A. Jain. Orientation field estimation for latent fingerprint enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(4):925–940, 2013. xxv, 48, 49

M. Ferrara, D. Maltoni, and R. Cappelli. Noninvertible Minutia Cylinder-Code Representation. *IEEE Transactions on Information Forensics and Security*, 7(6):1727–1737, 2012. 54

J. Fierrez, L. Nanni, J. Ortega-Garcia, C. R., and D. Maltoni. Combining multiple matchers for fingerprint verification: a case study in FVC2000. *Proc. 13th IAPR Intl. Conf. on Image Analysis and Processing, ICIAP, Springer LNCS-3617*, pages 035–1042, 2005. 77

J. Fierrez-Aguilar, Y. Chen, J. Ortega-Garcia, and A. K. Jain. Incorporating image quality in multi-algorithm fingerprint verification. In *Proc. IAPR Intl. Conf. on Biometrics, ICB*, volume 3832 of *LNCS*, pages 213–220. Springer, January 2006. 75

J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. Target dependent score normalization techniques and their application to signature verification. *IEEE Trans. on Systems, Man and Cybernetics, part C*, 35(3): 418–425, 2005. 116

Frye, 1923. URL `http://www.law.ufl.edu/faculty/little/topic8.pdf`. Frye Standard. 5

M. Garris and R. McCabe. NIST special database 27: Fingerprint minutiae from latent and matching tenprint images. *Technical Report, NISTIR-6534*, 2000. 38

J. Gonzalez-Rodriguez, P. Rose, D. Ramos, D. T. Toledano, and J. Ortega-Garcia. Emulating DNA: Rigorous quantification of evidential weight in transparent and testable forensic speaker recognition. *IEEE Transactions on Audio, Speech and Language Processing*, 15(7):2072–2084, 2007. 104

R. Haraksim, D. Ramos, D. Meuwly, and C. E. Berger. Measuring coherence of computer-assisted likelihood ratio methods. *Forensic Science International*, 249:123–132, 2015a. 97

R. Haraksim, D. Ramos, D. Meuwly, and C. E. Berger. Measuring coherence of computer-assisted likelihood ratio methods. *Forensic Science International*, 249:123–132, 2015b. 130, 133

A. Hicklin. Extended Fingerprint Feature Set. *ANSI/NIST ITL 1-2000 Standard Workshop*, 2005. URL `http://biometrics.nist.gov/cs_links/standard/archived/workshops/workshop2/\Presentations-docs/Hicklin-Ext-FP-Features.pdf`. 72

A. Hicklin. Standardizing a More Complete Set of Fingerprint Features. *International Association for Identification conference, San Diego California*, 2007. URL `http://fingerprint.nist.gov/STANDARD/cdeffs/Docs/IAI_CDEFFS_2007-07-24.pdf`. xxvi, 14, 72, 73

E. Holder, L. Robinson, and J. Laub. The Fingerprint Sourcebook. *Department of Justice, Office of Justice Programs*, 2011. xxv, 3, 11, 30, 31, 32, 72

L. Hong, Y. Wan, and A. Jain. Fingerprint image enhancement: algorithm and performance evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):777–789, 1998. 81

M. Indovina, V. Dvornychenko, and *et al.* An evaluation of automated latent fingerprint identification technology (Phase II). *NIST Interagency/Internal Report (NISTIR)*, pages 75–77, 2009. 17, 18, 35

M. Indovina, V. Dvornychenko, and *et al.* Evaluation of latent fingerprint technologies: Extended feature sets (evaluation 2). *Technical Report NISTIR 7859*, 2012a. 17, 18, 35

M. Indovina, V. Dvornychenko, R. A. Hicklin, and I. G. Kiebuzinski. ELFT-EFS Evaluation of Latent Fingerprint Technologies: Extended Feature Sets [Evaluation# 2]. *National Institute of Standards and Technology, US Department of Commerce NISTIR 7859*, 2012b. 17, 18, 75

M. Indovina, V. Dvornychenko, R. A. Hicklin, and I. G. Kiebuzinski. ELFT-EFS2 Results. *IAI 97th International Conference, Phoenix, Arizona*, 2012c. URL `http://biometrics.nist.gov/cs_links/latent/elft-efs/IAI_2012/ELFT-EFS2_IAI_2012_Final.pdf`. 17, 18, 75

M. Indovina, R. Hicklin, and *et al.* NIST evaluation of latent fingerprint technologies: Extended feature sets [Evaluation #1]. *NIST Interagency/Internal Report (NISTIR) - 7775, Tech. Rep.*, 2011a. 7, 16, 18, 35

M. Indovina, R. A. Hicklin, and G. I. Kiebuzinski. ELFT-EFS Evaluation of Latent Fingerprint Technologies: Extended Feature Sets [Evaluation# 1]. *National Institute of Standards and Technology, US Department of Commerce NISTIR 7775*, 2011b. 17, 18, 74, 75

A. Jain. Automatic Fingerprint Matching Using Extended Feature Set. *US Department of Justice, Document NNCJ, 235577*, 2010. 14, 18, 70, 76

A. Jain, Y. Chen, and M. Demirkus. Pores and ridges: High-resolution fingerprint matching using level 3 features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 15–27, 2007a. 12, 18, 34

A. Jain, Y. Chen, and M. Demirkus. Pores and ridges: High-resolution fingerprint matching using level 3 features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(1):15–27, 2007b. 76

A. Jain and J. Feng. Latent Palmprint Matching. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 31 (6):1032–1047, 2009. 48, 49

A. Jain and J. Feng. Latent Fingerprint Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(1):88–100, 2011. 40, 76, 77

T. Jea and V. Govindaraju. A minutia-based partial fingerprint recognition system. *Pattern Recognition*, 38(10): 1672–1684, 2005. 12, 13, 18, 34

X. Jiang, M. Liu, and A. Kot. Fingerprint retrieval for identification. *IEEE Transactions on Information Forensics and Security*, 1(4):532–542, 2006. 43

M. Kass and A. Witkin. Analyzing oriented patterns. *Computer vision, graphics, and image processing*, 37(3): 362–385, 1987. 53

R. P. Krish, J. Fierrez, J. Galbally, and M. Martinez-Diaz. Dynamic signature verification on smart phones. In *Proc. Workshop on User-Centric Technologies and Applications, PAAMS*, May 2013a. 25

R. P. Krish, J. Fierrez, D. Ramos, and Ortega-Garcia. Improving Automated Latent Fingerprint Identification using Extended Feature Sets. *Forensic Science International*, 2015. 22, 23

R. P. Krish, J. Fierrez, D. Ramos, J. Ortega-Garcia, and J. Bigun. Partial Fingerprint Registration for Forensics using Minutiae-Generated Orientation Fields. *IEEE 2nd International Workshop on Biometrics and Forensics, Valletta, Malta*, 2014a. 22, 23

R. P. Krish, J. Fierrez, D. Ramos, J. Ortega-Garcia, and J. Bigun. Pre-Registration for Improved Latent Fingerprint Identification. *Proc. IAPR/IEEE 22nd International Conference on Pattern Recognition, ICPR*, pages 696–701, 2014b. 22

R. P. Krish., J. Fierrez, D. Ramos, J. Ortega-Garcia, and J. Bigun. Pre-registration of latent fingerprints based on orientation field. *IET Biometrics*, pages 1–11, January 2015. 23, 24, 86

R. P. Krish, J. Fierrez, D. Ramos, R. Veldhuis, and R. Wang. Evaluation of AFIS-ranked latent fingerprint matched template. *6th Pacific-Aim Symposium on Image and Video Technology, Mexico, Springer LNCS-8333*, pages 230–241, 2013b. 24, 25, 38

R. P. Krish, J. Fierrez, D. Ramos, and R. Wang. On the importance of rare features in AFIS-ranked latent fingerprint matched templates. In *Proc. 47th IEEE International Carnahan Conference on Security Technology (ICCST), Medellin, Colombia*, October 2013c. 22, 23, 25

K. Kryszczuk, A. Drygajlo, and P. Morier. Extraction of Level 2 and Level 3 Features for Fragmentary Fingerprints. *In Proc. COST Action 275 Workshop*, pages 83–88, 2004. 76

P. Li, Y. Fu, U. Mohammed, J. Elder, and S. J. D. Prince. Probabilistic models for inference about identity. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 34(1):144–157, 2010. 96

L. Liu, T. Jiang, J. Yang, and C. Zhu. Fingerprint registration by maximization of mutual information. *IEEE Transactions on Image Processing*, (5):1100–1110, 2006. 36, 37, 38

B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. *Proc. International Joint Conference on Artificial Intelligence*, 81:674–679, 1981. 35

J. A. Maintz and M. A. Viergever. A survey of medical image registration. *Medical image analysis*, 2(1):1–36, 1998. 35

D. Maltoni, D. Maio, A. Jain, and S. Prabhakar. Handbook of Fingerprint Recognition. *Springer Publishing Company, Incorporated*, 2009. xxv, xxv, 5, 6, 30, 32, 33, 36, 47, 51, 53, 81

M. Martinez-Diaz, J. Fierrez, R. P. Krish, and J. Galbally. Mobile signature verification: Feature robustness and performance comparison. *IET Biometrics*, 3(4):267–277, December 2014. 25

MCC. Minutia Cylinder Code. MCC-SDK v1.4. URL `http://biolab.csr.unibo.it`. 54

D. Meuwly. *Reconaissance de Locuteurs en Sciences Forensiques: L'apport d'une Approache Automatique*. Ph.D. thesis, IPSC-Universite de Lausanne, 2001. 103

H. Moon and P. J. Phillips. Computational and performance aspects of PCA-based face-recognition algorithms. *Perception-London*, 30(3):303–322, 2001. 16

NAS-NRC. Strengthening the Forensic Sciences in the United States: A Path Forward. *National Academy of Sciences*, 2009. 12, 19

J. Navratil and G. Ramaswamy. The awe and mystery of t-norm. In *Proc. of ESCA Eur. Conf. on Speech Comm. and Tech., EuroSpeech*, pages 2009–2012, 2003. 104, 113, 117

C. Neumann, C. Champod, M. Yoo, T. Genessay, and G. Langenburg. Quantifying the weight of fingerprint evidence through the spatial relationship, directions and types of minutiae observed on fingermarks. *Forensic Science International*, 248:154–171, 2015. 97

C. Neumann, I. Evett, and J. Skerret. Quantifying the weight of evidence from a forensic fingerprint comparison: a new paradigm. *Journal of the Royal Statistical Society, Series A: Statistics in Society*, 175(2):371–415, 2012. 97

Neurotec-Biometric-4.3. Neurotechnology VeriFinger-SDK. URL `http://www.neurotechnology.com/verifinger.html`. 80, 88

K. Nilsson and J. Bigun. Registration of fingerprints by complex filtering and by 1D projections of orientation images. *Audio-and Video-Based Biometric Person Authentication, Springer Berlin Heidelberg*, pages 171–183, 2005. 37, 38

NIST-ELFT, 2013. URL `http://www.nist.gov/itl/iad/ig/latent.cfm`. Evaluation of latent fingerprint technologies. 7, 16, 18, 34, 70

NIST-ELFT-1, 2007. URL `http://biometrics.nist.gov/cslinks/latent/elft07/phase1aggregate.pdf`. Summary of the results of Phase I ELFT testing. 17, 18, 34

NIST-EWG. Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach. *Expert Working Group on Human Factors in Latent Print Analysis*, 2012. xxv, xxv, 8, 9, 12, 13, 18

NIST-NBIS. NIST Biometric Image Software. NBIS-Release v4.2.0. URL `http://www.nist.gov/itl/iad/ig/nbis.cfm`. 54, 88

OSAC. a. URL `http://www.swgfast.org/index.html`. 7

OSAC. b. URL `http://www.nist.gov/forensics/osac/index.cfm`. 7

B. Paltridge. Thesis and dissertation writing: an examination of published advice and actual practice. *English for Specific Purposes*, 21(2):125–143, 2002. 20

S. Pankanti, S. Prabhakar, and A. Jain. On the individuality of fingerprints. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1010–1025, 2002. 11

A. A. Paulino, J. Feng, and A. K. Jain. Latent fingerprint matching using descriptor-based hough transform. *IEEE Transactions on Information Forensics and Security*, 8(1):31–45, 2013. 86

S. Pigeon, P. Druyts, and P. Verlinde. Applying Logistic Regression to the Fusion of the NIST'99 1-Speaker Submissions. *Digital Signal Processing*, 10(1):237–248, 2000. 104, 105

D. Ramos, J. Fierrez, R. P. Krish, and F. J. Gomez-Herrero. Evidence evaluation using afis scores: Integrating rare features. *IET Biometrics*, 2015. 22, 23

D. Ramos and J. Gonzalez-Rodriguez. Reliable support: Measuring calibration of likelihood ratios. *Forensic Science International*, 230:156–169, May 2013. 15, 96, 97, 101, 103

D. Ramos, J. Gonzalez-Rodriguez, G. Zadora, and C. Aitken. Information-theoretical assessment of the performance of likelihood ratio models. *Journal of Forensic Sciences*, 58(6):1503–1518, November 2013. 97, 101, 102, 103

N. Ratha, K. Karu, S. Chen, and J. A.K. A real-time matching system for large fingerprint databases. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 18(8):799–813, 1996. 13, 33

M. J. Saks and J. J. Koehler. The coming paradigm shift in forensic identification science. *Science*, 309(5736): 892–895, 2005. 12, 15, 19, 95, 96

F. Santamaria. A New Method of Evaluating Ridge Characteristics. *Fingerprint and Identification Magazine*, 1955. 79

S. Srihari. Quantitative Measures in Support of Latent Print Comparison: Final Technical Report: NIJ Award Number: 2009-DN-BX-K208. *University at Buffalo, SUNY*, 2013. 8, 11, 12, 19

S. Srihari and H. Srinivasan. Individuality of fingerprints: comparison of models and measurements. *University at Buffalo, Amherst, New York TR-02-07*, 2007. 11

J. D. Stosz and L. A. Alyea. Automated System for Fingerprint Authentication Using Pores and Ridge Structure. *In Proc. SPIE Conference on Automatic Systems for the Identification and Inspection of Humans*, 2277:210–223, 1994. 76

SWGFAST.v01. SWGFAST : Standards for Examining Friction Ridge Impressions and Resulting Conclusions, ver. 1.0. 2011. 7, 8

F. Taroni, C. Aitken, P. Garbolino, and A. Biedermann. *Bayesian Networks and Probabilistic Inference in Forensic Science*. John Wiley & Sons, 2006. 96

F. Taroni, C. G. G. Aitken, and P. Garbolino. De Finetti's Subjectivism, the Assessment of Probabilities and the Evaluation of Evidence: A Commentary for Forensic Scientists. *Science and Justice*, 41(3):145–150, 2001. 100

o. U.S.Department, 2013. URL http://nij.gov/topics/forensics/evidence/Pages/welcome.aspx. Forensic Sciences: Types of Evidence, National Institute of Justice (NIJ). 1, 2

D. A. van Leeuwen and N. Brummer. The distribution of calibrated likelihood-ratios in speaker recognition. *arXiv preprint*, 2013. URL http://arxiv.org/abs/1304.1199. 97

J. Vanderkolk. Examination Process, The Fingerprint Sourcebook. *U.S Department of Justice*, 2011. 11

D. vanLeeuwen and N. Brümmer. An introduction to application-independent evaluation of speaker recognition systems. In C. Müller, editor, *Speaker Classification*, volume 4343 of *Lecture Notes in Computer Science / Artificial Intelligence*. Springer, Heidelberg - Berlin - New York, 2007. 15, 96, 105

Various. Expressing evaluative opinions: A position statement. *Science and Justice*, 51:1–2, 2011. Several signatories. 96

M. Vatsa, S. Richa, N. Afzel, and S. K. S. Quality induced fingerprint identification using extended feature set. *IEEE International Conference in Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–6, 2008. 77

J. Villalba and N. Brummer. Towards Fully Bayesian Speaker Recognition: Integrating Out the Between-Speaker Covariance. In *Proceedings of the 12th Annual Conference of the International Speech Communication Association, Interspeech 2011*, pages 505–508, Florence, Italy, 2011. 97

R. Wang, D. Ramos, J. Fierrez, and R. P. Krish. Automatic region segmentation for high-resolution palmprint recognition: Towards forensic scenarios. In *Proc. 47th IEEE International Carnahan Conference on Security Technology (ICCST), Medellin, Colombia*, October 2013a. 25

R. Wang, D. Ramos, J. Fierrez, and R. P. Krish. Towards regional fusion for high-resolution palmprint recognition. In *Proc. XXVI SIBGRAPI conference on Graphics, Patterns and Images*, August 2013b. 25

Y. Wang and J. Hu. Global ridge orientation modeling for partial fingerprint identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(1):72–87, 2011. 12, 18, 34

S. Willis. Enfsi guideline for the formulation of evaluative reports in forensic science. monopoly project MP2010: The development and implementation of an enfsi standard for reporting evaluative forensic evidence. Technical report, European Network of Forensic Science Institutes, 2015. 96

N. Yager and A. Amin. Evaluation of Fingerprint Orientation Field Registration Algorithms. *Proc. IAPR/IEEE 17th International Conference on Pattern Recognition, ICPR*, 4:641–644, 2004. 37, 38

N. Yager and A. Amin. Fingerprint alignment using a two stage optimization. *Pattern Recognition Letters*, 27 (5):317–324, 2006. 37, 38

X. Yang, J. Feng, and J. Zhou. Localized dictionaries based orientation field estimation for latent fingerprints. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(5):955–969, 2014. 129, 133

G. Zadora, A. Martyna, D. Ramos, and C. Aitken. *Statistical Analysis in Forensic Science: Evidential Values of Multivariate Physicochemical Data. John Wiley and Sons.* John Wiley and Sons. John Wiley and Sons, January 2014. 97

G. Zadora and D. Ramos. Evaluation of glass samples for forensic purposes – an application of likelihood ratio model and information-theoretical approach. *Chemometrics and Intelligent Laboratory Systems*, 102:62–63, 2010. 96

Q. Zhao and A. Jain. On the utility of extended fingerprint features: A study on pores. *IEEE Computer Society Conference In Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 9–16, 2010. 14, 70, 76

F. Zheng and C. Yang. Latent Fingerprint Match using Minutia Spherical Coordinate Code. 8^{th} *IAPR International Conference on Biometrics*, 2015. 86

B. Zitova and J. Flusser. Image registration methods: a survey. *Image and vision computing*, 21(11):977–1000, 2003. 35