Statistical Modelling of Fingerprints

Stephanie Jane Llewelyn

Submitted for the degree of PhD in Statistics

School of Mathematics and Statistics

June 2014

University of Sheffield

Firstly I would like to thank Dr Nick Fieller who started on this journey with me and set me off in the right direction, and a huge thank you to Prof Paul Blackwell who ended the journey with me and provided enormous support during those complicated moments. Both have been invaluable and I couldn't have made it this far without either. Also I would like to thank Dr Tim Heaton who gave me an extra outlet for help and advice.

I am eternally grateful to the Engineering and Physical Sciences Research Council (EPSRC) and the Forensic Science Service (FSS) who jointly funded this research. Within the Forensic Science Service I would like to thank Roberto Puch-Solis who provided a significant amount of context and background knowledge which was the grounding for this whole piece of research.

Finally I would like to thank my family and friends who have demonstrated unending support over the last few years. To my beloved partner Tom, I could not have stayed sane (or nearly) this whole time without your love and unwavering faith in my ability. You are my rock, giving me somewhere that I can thrash out ideas and talk through problems without any judgment. To my mum the only true constant I have ever known, there are no words as you well know - although maybe now I will talk more when asked how my day was! I can't begin to explain my gratitude so I won't, just know it is there. Very finally, last but of course never least, my partner in crime (and postgraduate research), Emma Jones, who shared the highs and lows with me, we have shared a moment in time that can't be described to anyone who hasn't walked the road we did, I will never have another friend who will understand the depths of me like you.

To the others I have no doubt missed, I apologise and can only blame it on my maths-riddled brain. Thank you all.

Abstract

It is believed that fingerprints are determined in embryonic development. Unlike other personal characteristics the fingerprint appears to be a result of a random process. For example fingerprints of identical twins (whose DNA is identical) are distinct, and extensive studies have found little evidence of a genetic relationship in terms of types of fingerprint, certainly at the small scale. At a larger scale the pattern of ridges on fingerprints can be categorised as belonging to one of five basic forms: loops (left and right), whorls, arches and tented arches. The population frequencies of these types show little variation with ethnicity and a list of the types occurring on the ten digits can be used as an initial basis for identification of individuals. However, such a system would not uniquely identify an individual although the frequency of certain combinations could be extremely small. At a smaller scale various minutiae or singularities can be observed in a fingerprint. These include ridge endings and bifurcations, amongst others. Typical fingerprints have several hundred of these as well as two key points (with the exception of a simple arch) referred to as the core and delta, which are focal points of the overall pattern of ridges. Modern identification systems are based upon ridge endings and bifurcations, not least because they are the easiest to determine automatically from image analysis. The configuration of these minutiae is unique to the individual.

This research explores the relationship between the locations of minutiae to determine if they can be modelled using a statistical process. In addition, since the approach is based on how fingerprints can be examined in a forensic situation an algorithm is created and tested which allows the strength of a match between a fingermark left at a crime and a fingerprint from a known suspect to be calculated. Currently the result of matching a fingermark and fingerprint is expressed as a categorical value of; match, no match or inconclusive. The method in this research allows this to be expressed as a numerical value allowing for a wider and more flexible use of fingerprint evidence.

Contents

1 Introduction **1**

1.1 Motivation 1

1.2 Fingerprint History 3

1.3 Existing Methods for Fingerprint Classification and Comparison 6

1.3.1 Fingerprint Patterns 6

1.3.2 Forensic Methods 9

1.4 Thesis Outline 9

2 Review of Literature **11**

2.1 Introduction 11

2.2 Models for Fingerprint Generation 12

2.2.1 Overview 12

2.2.2 Methods 13

2.2.3 Results 17

2.3 Models To Compute Likelihood Ratios 20

2.3.1 Overview 20

2.3.2 Methods 21

2.3.3 Results 23

2.4 Models to Assess Individuality Using PRC 24

2.4.1 Overview 24
2.4.2 Methods 24
2.4.3 Results 27
2.5 Conclusions 28

3 Data **29**
3.1 Introduction 29
3.2 Data sets from FSS 30
3.3 Dummy Marks 32
3.4 My Prints 35
3.5 Summary 36

4 Assessment of Complete Spatial Randomness **39**
4.1 Introduction 39
4.2 Plotting Programs 40
4.3 Density Plots 43
4.4 Plotting Functions 44
4.4.1 Scaling 44
4.4.2 K-Functions 45
4.4.3 Size Analysis 48
4.5 Summary 49

5 Model Fitting **56**
5.1 Introduction 56
5.2 Investigation of Models 56
5.3 Strauss Model 57
5.4 Isotropic Centred Poisson Process 61

5.5 Summary 62

6 Matching Algorithm 66

6.1 Introduction 66

6.1.1 Method Summary 67

6.1.2 Fingermark Simulation 67

6.1.3 Background to Transformations 68

6.2 Method 69

6.2.1 Candidate Set 69

6.2.2 Transformation 72

6.2.3 The Unique Allocation Problem 75

6.2.4 Hungarian Algorithm 75

6.3 Final Step 78

6.4 Summary 79

7 Simulation Experiments 80

7.1 Introduction 80

7.2 Setup 81

7.3 Theory for Calculating the Likelihood Ratio 82

7.3.1 Prosecutor Hypothesis 83

7.3.2 Defence Hypothesis 85

7.3.3 Likelihood Ratio 86

7.4 Results from Simulation Experiments 87

7.4.1 Results for Changes in Number of Minutiae 87

7.4.2 Results for Changes in σ 91

7.5 Discussion 97

7.6 Summary 97

8 Optimisation **100**

8.1 Introduction 100

8.2 Method of Optimisation 100

8.3 Results from Optimisation 101

8.3.1 Results for Changes in Number of Minutiae 101

8.3.2 Results for Changes in σ 105

8.4 Discussion 111

8.5 Summary 113

9 Expansion with Minutia Type **114**

9.1 Introduction 114

9.2 Method for Inclusion of Minutia Type 114

9.3 Results from the Inclusion of Minutia Type 116

9.3.1 Results for Changes in Number of Minutiae 116

9.3.2 Results for Changes in σ 122

9.4 Discussion and Summary 125

10 Conclusions and Further Work **128**

10.1 Spatial Modelling 128

10.2 Matching Algorithm 129

10.3 Further Work and Discussion 130

References **132**

Appendix A R code - Matching Algorithm **138**

Appendix B R code - Additional Optimisation Steps **150**

Appendix C R code - Extension to Minutia Type Steps **152**

Chapter 1

Introduction

1.1 Motivation

Fingerprints are important since they offer a sound basis for identification; during much of history fingerprints have been widely accepted for this use. Fingerprints are an important aspect of modern society; their benefits are currently utilised for commercial activities (computing, security etc). However a more well known use of fingerprints is in a forensic situation. The two ideas that make fingerprints so relevant are (i) permanence, and (ii) individuality (Alonso et al., 2007); these advantages lend themselves to forensics, where they can be used to establish a suspect's placement at a crime scene. In 2002 alone, in the UK 330,000 crime scenes yielded fingerprint evidence, this recovered data led to the detection of 34,000 suspects (Great Britain Home Office, Forensic Science Pathology Unit, United Kingdom, 2005).

Whilst the validity of the assertion of permanence has been "established by empirical observations as well as based on the anatomy and morphogenesis of friction ridge skin" (Jain et al., 2002), the claim of individuality (uniqueness) has not been formally assessed despite it being commonly accepted as true. Since the Daubert v. Merrell Dow Pharmaceuticals' case in 1993 (Daubert v. Merrell Dow Pharmaceuticals Inc, 1993) forensic evidence has been under increasing scrutiny in court. During this case the Supreme Court ruled that expert forensic testimony must adhere to five criteria:

1. The technique or methodology must have been subjected to statistical hypothesis

testing,

2. Error rates must have been established,
3. Standards controlling the technique's operation exist and have been maintained,
4. It has been peer reviewed and published,
5. It has general widespread acceptance.

Fingerprints were first challenged in court during a case in 1999 (U.S. v. Byron Mitchell, 1999) on the basis that they do not conform to all five criteria from Daubert v. Merrell Dow Pharmaceuticals Inc (1993) (Cole, 2004). In particular, an expert witness for the defense argued that the uniqueness of fingerprints has not been thoroughly tested and potential matching error rates are unknown. The result of this challenge in court was a five day *Daubert* hearing which concluded with the court "declining to decide whether forensic fingerprint identification was properly labelled as scientific evidence" (Cole, 2004). Due to this there has been an increased amount of literature and research into the area of fingerprint uniqueness. This is where this research comes in. The aim is to understand the within and between fingerprint variability. By looking at how properties of local features interact we are hoping to establish the underlying statistical structure of fingerprints so that a strength of a match can be established. This may incorporate different aspects of fingerprints such as level 1 patterns, minutiae types or even finger type. We hope that by using statistics to model fingerprint features a likelihood ratio can be calculated. Currently there is much work being done in presenting likelihood ratios to a court in relation to DNA evidence (which is not given as a categorical decision unlike fingerprints which are), this could be extended to take into account other forms of evidence such as fingerprints, glass fragments, or hair fibres, amongst others.

It is worth noting that the issue of uniqueness of fingerprints can be viewed as a "red herring". We believe that although it is worth knowing about the uniqueness of fingerprints (for intellectual understanding); in a forensic environment it isn't entirely relevant as two fingerprints are never compared. It is the case that a fingermark is compared with a fingerprint. A mark will be of much lower quality and hence the uniqueness of the print it came from is not useful for the comparison, a more interesting question is the uniqueness of fingermarks. In addition when looking at likelihood ratios,

as this research intends to do, as long as the features are very rare the difference between one in seven billion and one in four billion (for example) is negligible.

1.2 Fingerprint History

References to fingerprints exist throughout history. One of the earliest examples of the use of fingerprints is found in a Chinese document entitled "The Volume of Crime Scene Investigation—Burglary" which dates back to the Qin Dynasty (221BC — 206BC). During the Tang Dynasty (617AD — 907AD) in China fingerprints were used as a form of identification. Also in eighth century Japan fingerprints were used instead of a signature on legal documents (Barnes, 2010). In 1823 Jan Purkinje published a thesis (Cummins and Kennedy, 1940) within which he details nine fingerprint patterns. These nine patterns correspond to (i) transverse curves (arch), (ii) central longitudinal stria (tented arch), (iii) oblique stripe (loop left or right), (iv) oblique loop (loop left or right), (v) almond (whorl), (vi) spiral (whorl), (vii) eclipse (whorl), (viii) circle (whorl), and (ix) double whorl (whorl). Purkinje's nine patterns can be seen to fit into the five groups we use today as highlighted in brackets. However no distinction is made between left and right loops but the whorls show more detailed subgroups.

William James Herschel played a key role in the development of the study of fingerprints when he speculated on their uniqueness in 1859. At the time he was using fingerprints instead of a signature on contracts in India (Barnes, 2010). In 1880 the first European published work about fingerprints appeared, this was an article in Nature by Henry Faulds (Faulds, 1880). He wrote about how fingerprints could be used for identification and specifically for identifying criminals. In addition to this he sent his work to Charles Darwin, who, in declining health, forwarded the work to Francis Galton, this undoubtedly led to the publication Finger Prints (Galton, 1892). In fact Stoney (2001) credits Galton with the first proposal of a statistical model for fingerprints when he attempted to specify ridge details.

The Bertillon System was used for many years as a way of identifying criminals and repeat offenders; it used measurements from eleven bony parts of the body as a way of doing this, as well as scars, personality traits and other interesting aspects. This method successfully identified its first criminal in 1883. The details were presented on

record cards which were both time and space consuming to use. The main reason for the abandonment of this system was its lack of uniqueness, in its place fingerprints began being used as the best method of identification. In 1892 there was the first criminal case involving fingerprints; this took place in Argentina and concerned the murder of two boys by their mother (Barnes, 2010). In that same year Galton published Finger Prints (Galton, 1892).

In the late 1800s Edward Henry worked in Bengal. Here with some fellow researchers the Henry system was developed for classifying fingerprints (Barnes, 2010). This system was used from the 1900s until the mid 1990s. It was a logical system for recording fingerprints using a formula which assigns numbers to different patterns and finger allocations. The method requires a full ten-print set, it's basis is the presence of a pattern in the ridge flow called a whorl, these are explained more fully in Section 1.3.1. The method is as follows:

1. Label each finger with a number as in Figure 1.1, starting with the right thumb (1) and ending with the left little finger (10),
2. Assign each finger a value based on whether the finger has a pattern called a whorl, using Table 1.1 (zero otherwise) ,
3. Sum the values for the even and odd labelled fingers separately,
4. Add one to each of these sums,
5. Represent this as a value ranging from 1/1 to 32/32 (even/odd).

Finger Label	Value for a whorl
1, 2	16
3, 4	8
5, 6	4
7, 8	2
9, 10	1

Table 1.1: A table showing the value for the presence of a whorl in the Henry System

A 1/1 is a set of prints with no whorls and a 32/32 is a set of prints with all whorls. For example, if the fingers displayed in red in Figure 1.1 represent a whorl this pattern would

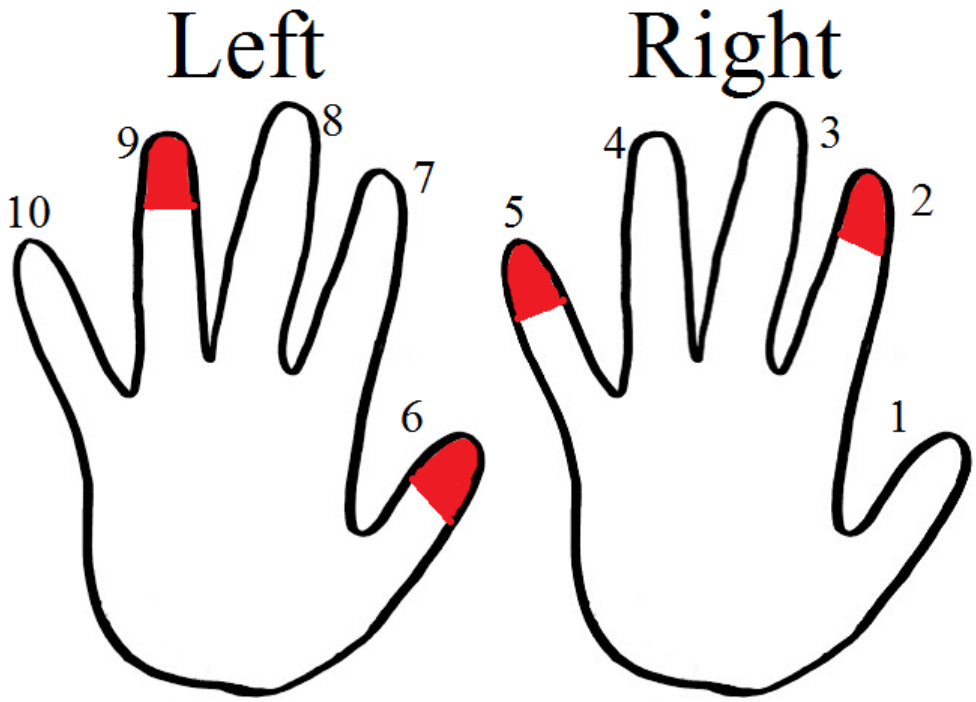

Figure 1.1: Display of the method for the Henry system where the numbers represent the label of the finger and red fingers are whorls

be described as $(16 + 4)/(4 + 1) = 20/5$. This system was not unique but did in fact make it easier to break down the large groups of fingerprint patterns. Since many record cards were being created a better system was needed for searching and storing these and hence automated fingerprint identification systems (AFIS) were created. These are still in use today. The largest repository is in the United States of America (IAFIS) and is governed by the Federal Bureau of Investigation. Today it holds around 100 million sets of prints.

1.3 Existing Methods for Fingerprint Classification and Comparison

1.3.1 Fingerprint Patterns

It is important to investigate the different type of patterns and what is meant by the terminology in this area. It is believed that the ridges and valleys that make up a fingerprint are formed during the 11th and 27th weeks of fetal development (Babler, 1991). Some authors (Maltoni et al., 2003) believe that the specific features of a fingerprint (including overall ridge flow, interruptions in ridge flow inter-ridge distances, pores etc) are a result of changes in the local micro-environment in the amniotic fluid, whilst others (Babler, 1991) think it is a combination of factors; additionally skeletal formation and nerve distribution amongst others. Visually fingerprints are made up of ridges and valleys, the overall ridge flow exhibited on a fingerprint can be categorised; these are described as level one (global patterns). We currently distinguish five main global patterns as determined by the Galton-Henry system of classification; these are (i) right loop, (ii) left loop, (iii) whorl, (iv) arch, and (v) tented arch. These can be seen by merely looking at the finger although in many cases the categorisation can be ambiguous. The most common are the loops and whorls. A study of a sample of people from 2008 gave the proportions to be 32%, 29%, 20%, 14% and 5% respectively (Srihari et al., 2008). The five fingerprint patterns described can be seen in Figure 1.2. Two other important features that can be identified at a global level are referred to as the core and delta (Levi and Sirovich, 1972), these can be seen in Figure 1.2. In Maltoni et al. (2003) these are described as the control points around which the ridge flow is wrapped. A core can be seen in the level one types except arch and tented arch; it

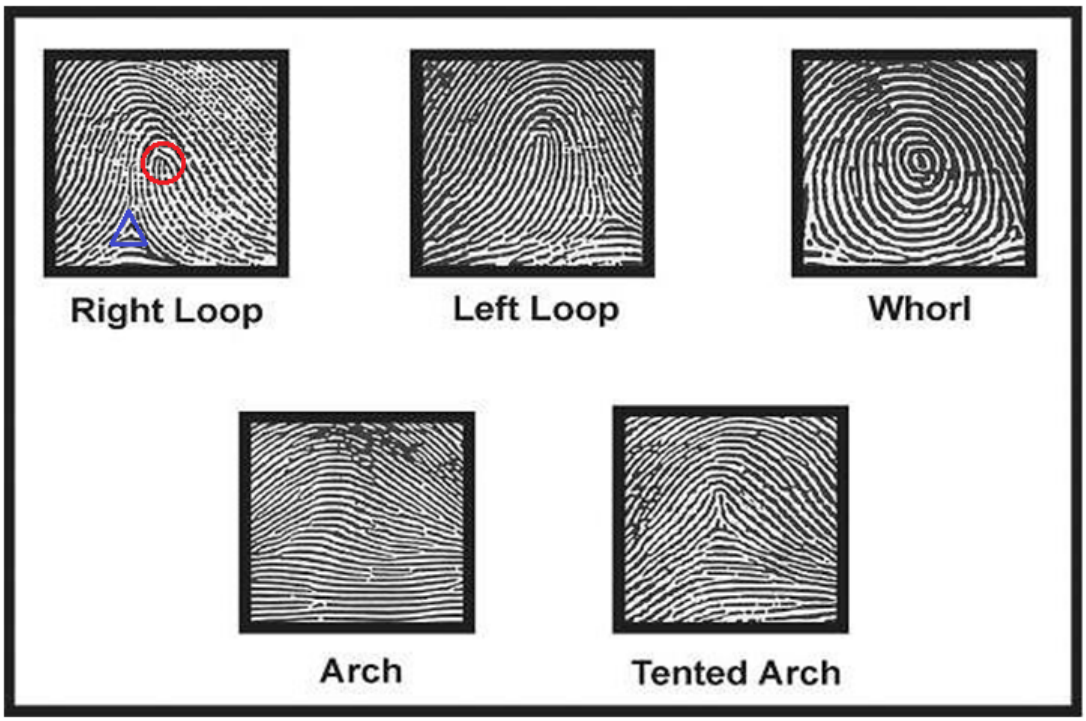

Figure 1.2: Display of the five main global fingerprint patterns, the red circle represents a core and the blue triangle a delta

describes the central location in a fingerprint where the ridge flow changes direction swiftly. A whorl is usually described as having two cores at either side of the changing ridge flow. A delta is seen in all level one patterns (with most whorls having two) and is described as a peak in the ridge flow. These global features are useful for classification.

When observing a fingerprint more closely changes in ridge-flow can be seen, these local ridge characteristics are referred to as minutiae. These level two patterns (local) form the basis of forensic identification since configurations of minutiae are believed to be unique. From the literature (Su and Srihari (2008), Maltoni et al. (2003)) minutiae are features within a fingerprint which can have type and orientation as well as location. In many cases however, location is the key feature used to describe minutiae as this is the only information available; in this situation they can often be referred to as simply points with x and y coordinates. There are many documented types of minutiae (a total of 150 according to Moenssens (1971)) but in a forensic situation only two are used; ridge endings and bifurcations. A ridge ending is described by Maltoni et al.

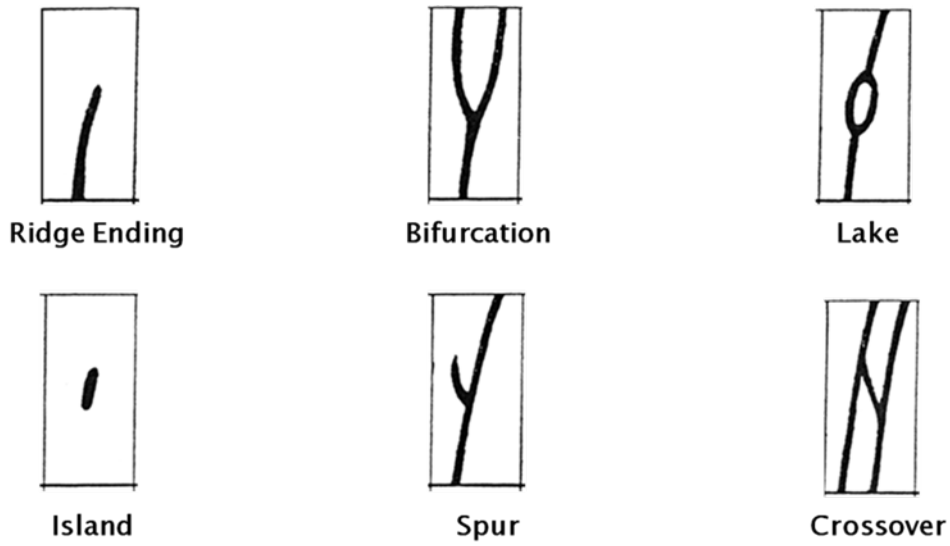

Figure 1.3: Display of several minutia patterns

(2003) as "the ridge point where a ridge ends abruptly" and a bifurcation as "the ridge point where a ridge forks or diverges into branch ridges". The use of only these two types is for several reasons, the first is that they are the most common and easiest to recognise. Secondly other types of minutiae can be described as combinations of ridge endings and bifurcations. For example from the images in Figure 1.3 a lake can be described as two adjoining bifurcations.

There are also level 3 patterns which include features such as pore detail on the ridges, ridge width and inter-ridge distance. However these details are not usually used in forensic identification since a mark left at a crime scene will typically be of poor quality and these details would be very hard to locate with confidence. It is also interesting to note that unlike DNA there is no relationship between family members and the location and types of minutiae. People of the same race or social standing do not display similar fingerprints. There has been a study to test the fingerprints of twins for similarities (Srihari et al., 2008). Despite it concurring that level two patterns are not similar for twins there was some evidence that identical twins are more likely to share the same level one patterns.

1.3.2 Forensic Methods

In a forensic context there are two types of patterns associated with fingers, these are marks and prints. Fingermarks are predominately what is made by a finger and constitute the evidence taken from a crime scene and are typically of poor quality. As well as this they are usually incomplete, distorted or smudged. Fingerprints however, are what is on the finger and can be recorded accurately if they are taken in controlled circumstances from a person. They tend to be the whole finger pad and demonstrate a complete image which is easy to analyse for minutiae. The basis for fingerprint use in forensics is therefore matching a print to a mark.

Forensics is focused on two areas. These are (i) intelligence or investigation and (ii) evaluation or assigning a weight of evidence. Intelligence is associated with findings from a crime scene (e.g. a finger mark) which are then searched through a database (AFIS) to find a matching print. Evaluation concerns a finger mark from a crime scene and a fingerprint taken from a suspect, these are then compared in order to draw a conclusion of match, non match or inconclusive (when there is insufficient evidence to support either result). There is little work available in the area of evaluation. Most of the work is from a computer scientist's point of view and less from methods using a probabilistic model to describe both intra- and inter-print variation. To clarify: intelligence requires a finger mark (of varying quality) whereas evaluation requires both a finger mark and a fingerprint.

1.4 Thesis Outline

Chapter 1 is an introductory chapter. It first outlines what the thesis is about and the aim of the research. Then the structure of the thesis is given with a chapter by chapter summary. Next details about the background to the topic, for example, information about fingerprint patterns, why this research needs to be carried out, some fingerprint history, and details about forensic methods.

Chapter 2 comprises of an in depth review of current literature on the topics of forensic statistics (including the Aitken and Lucy methods), current research in the area of fingerprints (fingerprint generation, likelihood ratios and probability of a random correspondence), as well as any other general statistical knowledge needed for the research

(point pattern analysis, likelihood ratios).

Chapter 3 introduces the data sets used in the research. It gives information about how they were obtained and the information they contain. Also details are given of any dummy data sets created for the purposes of the research, including information I have extracted from my own prints when methods were tested on a single known print.

Chapter 4 contains preliminary work carried out on the data sets. This includes details about plotting programs, density plots and initial plotting of functions (G, F, K).

Chapter 5 includes fitting models to the datasets provided by the Forensic Science Service. The usefulness of built-in models will be assessed, and then they will be fitted to the dataset. Information will be given about how parameters will be estimated from the models and what these are before looking at maximising the likelihood ratios.

Chapter 6 outlines the method for calculating the matching algorithm between a fingermark and a fingerprint, including detailed identification of the candidate set, transformation and the Hungarian Algorithm. This leads into Chapter 7 which details the theory for calculating the likelihood ratio and the results from simulation experiments using the matching algorithm from the previous chapter. This algorithm is optimised in Chapter 8 and again simulation experiments are carried out. The final work on the matching algorithm is done in Chapter 9 where the model is extended to include minutiae types.

Finally in Chapter 10 conclusions from the whole thesis are drawn and some ideas for ways to expand this piece of research are explored.

Chapter 2

Review of Literature

2.1 Introduction

This chapter focuses on reviewing relevant material in the area of fingerprint research. Much of the research previously carried out features techniques based on the ridge pattern created by realisations of a fingerprint. In contrast this research will focus heavily on minutiae locations and the interaction between them. There will be three main research areas outlined in this review, those being:

1. Models for fingerprint generation - these describe current methods for fingerprint generation, most methods focus heavily on ridge flow which differs from the approach in this research which relates to minutiae locations
2. Models to compute likelihood ratios in forensic statistics - current methods used for other types of forensic evidence and how this can be adapted to fingerprints
3. Models to assess individuality using probability of a random correspondence (PRC) - this shows research into creating probabilities for a population of fingerprints which we will need to know in order to evaluate a match.

2.2 Models for Fingerprint Generation

2.2.1 Overview

Fingerprint generation is a general term used to describe techniques which create realistic fingerprints. It has many uses and plays an important role in improving current techniques in both forensic and commercial markets. Many new security systems use fingerprints as a type of person specific recognition to allow valid users access to a particular place or piece of equipment. By improving generation techniques these systems would become more robust against impostor access as the initial fingerprint information is broken down and stored within the system, by researching how this information could be used to attack the system, this can be prevented from happening. In addition developments would aid many forensic avenues by allowing experts to reconstruct realistic fingerprints using the information they have obtained from fingermarks at a crime scene.

There have been several techniques proposed with a view to fingerprint generation. They fall into two main categories; (i) fingerprint synthesis: the generation of a fingerprint image with a view to looking (and performing) like a true fingerprint, and (ii) fingerprint reconstruction: this takes existing knowledge of a fingerprint that has been stored in some way to create an image that is the same as the original. Many steps in reconstruction and synthesis overlap. However, it is the original input that changes the outcome. In this section both of these techniques are discussed since knowledge of both is useful for our purpose.

Reconstruction techniques are currently predominantly used in security systems to see how viable it is to create the original print from a stored minutiae template. When a set of prints are first entered into the system, not all the information is stored. Instead only the main features are stored. This can range from lots of information to only the basics, for example aspects of the print such as: global pattern, minutia location, orientation and type, fingerprint area, frequency image etc. It is widely believed that these systems are one way and it is not possible to recreate the original print from these details however there has been much new work in this area (Cappelli et al., 2007; Jain et al., 2007; Feng and Jain, 2009). The main aim of these approaches are the reconstruction of the ridge lines from the minutiae positions. Reconstructing an

accurate print in this case still poses a difficult challenge. In Maltoni et al. (2003) the aim of fingerprint synthesis is to allow large databases of realistic fingerprints to be created so that fingerprint recognition algorithms can be effectively tested, evaluated and, most importantly, compared with other algorithms. In a forensic scenario more accurate methods for reconstruction could be developed by precisely modelling the variation in fingerprints, specifically in minutia detail. Then the reconstructed image accurately represents the original image, thus aiding the impact of fingerprints as a form of both evidence and identification in cases where only low quality finger marks are available.

2.2.2 Methods

The SFINGE Method

The main source in this area is the Handbook of Fingerprint Recognition (Maltoni et al., 2003). This outlines a procedure for fingerprint synthesis using 'The SFINGE Method'. The fingerprint area is first generated using a model based on four elliptical arcs and a rectangle, Figure 2.1 (Maltoni et al., 2003). This has five parameters; a_1, a_2, b_1, b_2 and c as can be seen below, these determine the shape of the overall fingerprint.

Next the method generates an orientation image. First a level 1 pattern is randomly chosen and positions of the core and delta are randomly selected, this is done according to level 1 specific constraints, for example a whorl has two cores with a delta on either side, these features are described in more detail in Section 1.3.1. Then an algorithm weighted with piecewise linear functions from Vizcaya and Gerhardt (1996) is used to give the general ridge flow. This is a variation on the Sherlock and Monro model (Sherlock and Monro, 1993) and essentially locally corrects the orientation of the ridge flow with respect to the level 1 features (core and delta). Following on from this a frequency image is generated. This sets out the ridge pattern frequency, i.e. the distance between ridges. SFINGE selects a feasible overall frequency from the distribution of those in real prints with an average ridge/valley period of nine pixels. Since this method will give a constant inter-ridge distance it is altered slightly in areas where we know the frequency is lower, for example in the areas above the uppermost core and below the lowest delta. The image is then perturbed and smoothed with the purpose of improving the appearance of the print.

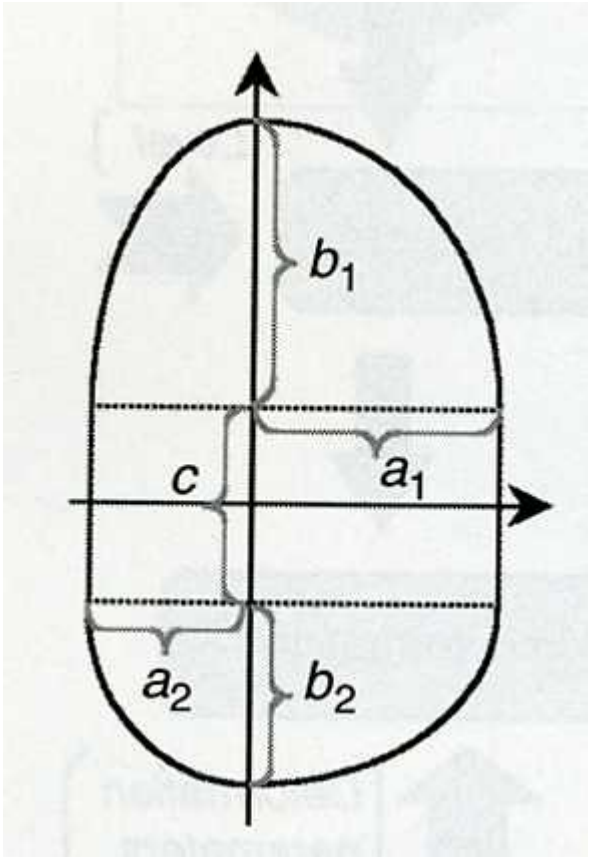

Figure 2.1: Fingerprint area image

Finally the ridge patterns are generated by iteratively enhancing an initial image using Gabor filters, this initial image uses several starting points as Cappelli et al. (2000) found that increasing the number of initial points gives a more realistic pattern of minutiae in the final image. The use of Gabor filters is based on the work by Hong et al. (1998) who proposed an effective method for the enhancement of a fingerprint image by using them to filter out undesired noise. Gabor filters are represented by both frequency and orientation (Daugman, 1985) which makes them ideal for use here as the frequency and orientation of the filter can be determined by the local ridge frequency and orientation. During this process minutiae are automatically generated at random positions as the ridge pattern is being produced.

Basic Reconstruction Method

The reconstruction method starts with the locations, orientations and sometimes types of minutiae and builds the other steps around these. Cappelli et al. (2007) give a basic method for fingerprint reconstruction when you have information about the original fingerprint. The fingerprint area is captured when a fingerprint is entered into the system and this is described in the same way as in the SFINGE method (see Figure 2.1). They also use the same method for determining the orientation image, however in this case the positions of the singularities (core and delta) are not known (since the reconstruction is from minutiae information only). In order to find the locations of the core and delta (or maybe multiple as whorls can have two of each) the orientation of the ridge flow associated with each minutia is taken into account. These are used to create a simple version of the fingerprint's ridge flow which is then compared to the Vizcaya and Gerhardt (1996) algorithm for general ridge flow as in the SFINGE method. The Nelder-Mead simplex algorithm (Nelder and Mead, 1965) then uses minimisation to find the version of the Vizcaya and Gerhardt (1996) which most closely matches the ridge flow created using the minutiae information. This was implemented using the method from Press et al. (1988) Next the minutiae locations, orientations and types (bifurcation and ridge ending) are fixed using a basic representation and these are used as the initial points for the image to be built up as before by iteratively enhancing using Gabor filters. The first step can be seen in Figure 2.2 (Cappelli et al., 2007) for different frequencies T, which refer to ridge frequency, the Gabor filter also uses information about the local ridge orientation. However when it comes to actually creating the ridge pattern

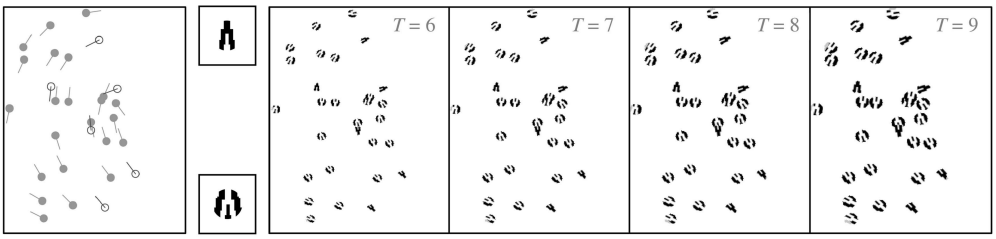

Figure 2.2: Reconstruction of a fingerprint

four different images are produced with varying ridge frequencies (T as in Figure 2.2) according to typical ridge frequencies from real prints. Again as with Maltoni et al. (2003) the image is then rendered to give a realistic final image.

Other Reconstruction Methods

In Jain et al. (2007) a more complicated method is used for generating the orientation image, instead using minutiae triplets to predict orientation in a given area. As with Cappelli et al. (2007) the minutiae locations and orientations are known and fixed (since this paper is concerned with reconstruction) so these are used for the minutiae triplets. Jain et al. (2007) use the orientation image to find the core of the fingerprint. The minutiae around the core are analysed to determine the type of level one pattern of the fingerprint. Streamlines (using minutiae as seed points) and line integral convolution (LIC) are further used to generate the ridge structure and give a textured appearance by making use of the orientation image and pattern class. This is an iterative process similar to using Gabor filters. The use of streamlines and LIC builds on earlier work by Laidlaw et al. (2005) and Cabral and Leedom (1993), see these for more in depth details, however these do not reference the application to fingerprints.

A paper which poses a completely different method is that of Feng and Jain (2009) which uses an FM (frequency modulated) model where the continuous and spiral components are generated separately and then combined, in order to have more control over the minutiae. The key step is the reconstruction of the continuous component. Larkin and Fletcher (2007) previously used a 2D amplitude and frequency modulated (AM-FM) signal with respect to fingerprint structure. Feng and Jain (2009), however, just select the phase component from this work which can then be decomposed into the continuous

and spiral phase. The continuous phase is determined by ridge frequency and the spiral phase corresponds to minutiae locations. Since, in a fingerprint reconstruction, there is usually a record of minutiae locations but also the associated local ridge orientations these are used in the method. After an orientation field has been generated using a novel method; the print is divided into small non-overlapping blocks and then the local minutiae are used to determine the orientation in a given block. It is then used to reconstruct the continuous phase, this is then combined with the spiral phase to give the reconstructed print. Although the results from this method appear successful the authors provide little detail about the actual techniques used and how these could be reproduced. The advantage of this method is that an orientation field can be generated with as little as one minutia unlike the other methods.

2.2.3 Results

The SFINGE Method

Some of the methods described were tested so that their effectiveness could be evaluated. In Maltoni et al. (2003) people at a conference were asked to select the one synthetic image from a group of four, only 23% of people correctly identified the synthetic fingerprint amongst the real images, leading to the conclusion that the synthetic print integrated well into the group. The SFINGE method for generating prints was also tested by comparing its performance to genuine prints with respect to quantities such as FMR (false match rate) and FNMR (false non-match rate), these showed similar behaviour for both the generated and real prints. In this situation an FMR is when a match is declared when in reality the two fingerprints do not come from the same source, and an FNMR is when a non-match is declared when the two fingerprints actually come from the same source.

Basic Reconstruction Method

In Cappelli et al. (2007) the reconstruction method (from a minutiae template) was evaluated by reconstructing prints from known fingerprints taken under controlled conditions. These two images were then compared visually, it was found that the images appear very similar and many of the original minutiae feature in the reconstructed im-

age in both the correct position and orientation. We would expect this to be the case as the original locations and orientations associated with the minutiae template were used as the basis for the reconstruction. The minutiae template doesn't necessarily contain all of the minutiae in a fingerprint though so the method creates more where ridge flow merges, these were located in similar positions and orientations to those in the original fingerprint. Although the orientations associated with the minutiae are used to inform the orientation image, they are not exactly the same as this is smoothed for the whole fingerprint, however it is reassuring that the reconstructed version is similar to the original. The reconstructed image does contain more (spurious) minutiae than the original. Also the orientation model derived from the template does not appear to be entirely accurate. Cappelli et al. (2007) also carried out an experiment to investigate the feasibility of an attack on several commercial fingerprint recognition algorithms. In this context attack means that a reconstructed print using minutiae information from a real print was presented to the system to see whether it was falsely matched against the real print held in the database. The attack is successful if false matches are declared at a high rate and the conclusion is that the reconstruction is very good or the system is poor. It was in fact successful in many of these cases with a percentage of FMR over 90%. This clearly demonstrates that more research into the area of protecting against these types of attacks is warranted.

Other Reconstruction Methods

Jain et al. (2007) assess their reconstructive model firstly by visual inspection as in Cappelli et al. (2007). This gave them a satisfactory conclusion of visual similarity and hence they moved on to investigate their model by matching the reconstructed prints with the originals using a fingerprint identification system as with the other models. However it is noted that the match rate is on average 23% which is lower than the two other methods for fingerprint reconstruction.

In Feng and Jain (2009) the model is tested by attempting to attack a fingerprint recognition system using the reconstructed fingerprint. We describe a type 1 attack as one in which one query print is tested against one specific master print (i.e. one to one) and type 2 as a query print tested against a whole data base of prints (i.e. one to many). The experiment showed that the system was vulnerable to both type 1 and

type 2 attacks with type 1 having a better success rate; the paper also shows that a type 1 attack had an even higher identification rate than the genuine matches.

Feng and Jain (2009) used the same database that was used by Jain et al. (2007), as well as the same minutiae detection and matching algorithm and hence it is easy to compare these two models and their performance directly; as previously stated the match rate for Jain et al. (2007) was 23% compared to Feng and Jain (2009) which had a match rate of 98.1%. Results from Jain et al. (2007) and Feng and Jain (2009) cannot be directly compared with those from Cappelli et al. (2007) since they only consider type 1 attacks in their paper.

Although all the methods described work, in that they produce a fingerprint image, they each have limitations that make them difficult to use in practice. The biggest limitation in Jain et al. (2007) is that the model does not take into account the types of minutiae and can only generate ridge endings. Hence the reconstructed image will never fully represent the original image adequately. In this method the match rates were increased if the class of the print is known (i.e. whorl, arch etc) instead of using the method described to estimate class since the performance of this method is not as good as current state of the art techniques. Jain et al. (2007) also suffer from the same issue as the other methods in that the inter-ridge distance is constant throughout the whole image, except in Maltoni et al. (2003) where it is manually changed in two areas. However in a real print the inter-ridge is not constant and can vary over the entire print and not just in the expected areas (above and below the core and delta respectively). In addition Jain et al. (2007) outline that this method of ridge structure is not as they desire and through just the minutiae information alone their model does not have enough orientation information to perform well in the core and delta regions, this leads to minutiae not matching up well in this region (with the original image) and many missing minutiae in the reconstructed image around these two areas.

Cappelli et al. (2007) highlight that their method of reconstruction does not cope well with low numbers of minutiae in the template since this results in wrong estimation of the orientation image. Furthermore it is assumed that there is no significant rotation of the original image. Also obvious issues with all the models are the structure around the core region, local shape of minutiae and pore detail. Feng and Jain (2009) admit that their model contains spurious minutiae in high curvature areas but propose some methods to solve this. In addition Feng and Jain (2009) proposes the inclusion of

details like ridge frequency and minutiae detail into their model in order to improve the reconstructed image. The method of Cappelli et al. (2007) also suffers from the issue of spurious minutiae which are added during the ridge pattern generation. In contrast Maltoni et al. (2003) does not. In fact it could be argued that too few minutiae are generated, resulting in very long ridges.

2.3 Models To Compute Likelihood Ratios

2.3.1 Overview

As previously discussed, there are two main aims in forensics: (i) intelligence, and (ii) evaluation. Intelligence describes the methods where a trace from a crime scene is searched through a relevant database to give a matching result; this can be thought of as a one-to-many search. Alternatively evaluation is the method by which a trace from a crime scene is compared against a known control piece of evidence taken from a suspect; this describes a one-to-one search. The second method is the concern here. Conventionally an expert will be requested in court to report a meaningful value in order for the court to assess the strength of the forensic evidence in a given context (Champod et al., 2005). In this scenario that means a finger mark will be lifted from a crime scene and compared by a fingerprint examiner to a fingerprint from a known suspect. They will then be requested in court to give testimony as to the match, the results they provide to the court are limited to match, non-match or inconclusive. Since the court is primarily responsible for interpreting the evidence and making a judgement based on this, when using a biometric system we cannot simply return a similarity measure or decision based on some threshold (Champod et al., 2005), hence another measure is required. This is where likelihood ratios become relevant; they are currently used in many areas of forensic evidence to give a score based result. It is determined that the job of the forensic scientist is to submit an objective result so that they comply with the conditions of the judicial procedures, and hence do not draw their own conclusions or give a biased testimony in the court room.

2.3.2 Methods

As with Neumann et al. (2012) (who use notation from Lindley (1977)), we start with the general notation for using likelihood ratios in a forensic scenario which is taken from Bayes Theorem, see also Champod et al. (2005):

$$\frac{p(H_p|E,I)}{p(H_d|E,I)} = \frac{p(H_p|I)}{p(H_d|I)} \times \frac{p(E|H_p,I)}{p(E|H_d,I)} \tag{2.1}$$

where H_p is the hypothesis that the evidence originates from the suspect, H_d the hypothesis that the evidence originates from an unknown individual (these are also referred to as the prosecutor and defence hypothesis respectively), E denotes the forensic information and finally I refers to the background information. A likelihood ratio can then be formed on the basis of this, and hence 'represents the strength of the analysis of the forensic evidence in the inference from prior to posterior odds' (Champod et al., 2007):

$$LR = \frac{p(E|H_p,I)}{p(E|H_d,I)}. \tag{2.2}$$

In Champod et al. (2005) values for both the numerator and denominator are then discussed, it is argued that the numerator in 2.2 is found from information about the within-source (intra-) variability of the specific forensic evidence and that the denominator is obtained from knowledge of the between-source (inter-) variability for the evidence. In our case this directly corresponds to the variability within a given fingerprint and the variability within a whole population of fingerprints respectively. In Fierrez-Aguilar et al. (2005) a table is given relating to the scale of likelihood ratios and how these correspond to a verbal conclusion, for example a value between 1 and 10 is said to have 'limited evidence to support' and a value greater than 10,000 'very strong evidence to support'. This system according to Fierrez-Aguilar et al. (2005) is currently being used in DNA analysis and is being extended to other areas.

In Neumann et al. (2012), the form of the likelihood ratio is represented generally for any item of evidence as:

$$LR = \frac{f_{Y,X|H_p,I}(y,x|H_p,I)}{f_{Y,X|H_d,I}(y,x|H_d,I)} \tag{2.3}$$

where f is the probability density function, Y is the collection of observations on the trace and X is the collection of observations on the control. By making several assumptions they then simplify this in such a way that will be useful for their later analysis. Neumann et al. (2012) then propose a method which defines a configuration of minutiae as a polygon, each minutia in the polygon has a feature array which contains information about the radius, clockwise side length angle, area and type. The method compares a configuration from a mark and print by minimising the distance between a polygon in the mark and a polygon in the print (both of size k). In addition, a weighting function is added to the new LR to ensure that when H_p is true (i.e. the control and trace are known to be from the same source by using simulated sets) the likelihood ratio is greater than one and when H_d is true the likelihood ratio is less than one.

In Fierrez-Aguilar et al. (2005) two methods for estimating the between-source variability are discussed. The first is a non-parametric approach from Meuwly et al. (2003). In its simplest form, it uses so-called histogram estimation; that is, the score axis is divided into bins of length h so that if N samples are provided and k_N of them are in a given bin, then the probability density is estimated as k_N/Nh. This can be generalised to kernel density estimation; see for example Silverman (1986). The second method involves parametric estimation, based on Gonzalez-Rodriguez et al. (2003), using a mixture of Gaussian density functions. Here the EM (expectation-maximisation) algorithm (Dempster et al., 1977) is used to identify the best-fitting model for the given data based on G Gaussian density functions.

2.3.3 Results

It is argued in the papers by Neumann et al. (2012) and Fierrez-Aguilar et al. (2005) that the biggest issue arises with estimating the between-source variability as the underlying probability density functions for the population are not known and the sample data set will always be considerably smaller than the population size. The experiments in Fierrez-Aguilar et al. (2005) showed that better results were obtained for low numbers of mixture components (low G) in the parametric case since these gave better results around a likelihood ratio of 1. It is claimed that with KDF low values of h (where h is the window width) are beneficial however when there is a small sample

size (which is usually the case) the probability density function does not accurately represent the underlying distribution for the population. This is unlike the parametric case which seems to demonstrate much similarity with the population. In Champod et al. (2005) the same experiments were carried out. The paper again showed that the performance of the system degraded when the sample size was small due to the between-source distribution over fitting the data set giving an inflated estimate for the likelihood ratio. This is not desirable in a forensic setting.

In Neumann et al. (2012) their methods were tested using real databases of fingerprints and marks. A small experiment was carried out with positive results which led to a larger experiment. In this large experiment there was obvious definition between LR(H_p) and LR(H_d), in most cases they fell as predicted with LR(H_p)>1 and LR(H_d)<1. What is not obvious in the paper is how the k-configuration is chosen in each fingerprint so that the configuration from the mark can be searched against these. If a typical fingerprint has 120 minutiae, for example, do the authors search every possible 10-point configuration from the 120 minutiae with every print in the database to find the configuration with the smallest distance? If not, how do they restrict the number of searches to be made? Obviously if every search is made this would be very computationally expensive to carry out and would take a large amount of time to process (Llewelyn, 2012).

2.4 Models to Assess Individuality Using PRC

2.4.1 Overview

Fingerprints have been an accepted form of identification for over a century. In 1892 Galton published a book (Galton, 1892) detailing global fingerprint patterns as well as writing about their individuality and permanence. These are the two factors which make fingerprints such an important aspect of modern society; many new devices, for example laptops, have built in fingerprint recognition systems in order to identify a user, as well as the more obvious practice of using fingerprints in forensics. Since the Daubert v. Merrell Dow Pharmaceuticals' case (Daubert v. Merrell Dow Pharmaceuticals Inc, 1993) the Supreme Court ruled that expert forensic testimony must adhere to five criteria. This has led to increased scrutiny of forensic evidence and poses many

problems for the use of fingerprints. Many researchers have tried to establish a measure of uniqueness of fingerprints using a technique called PRC (probability of a random correspondence), in Dass et al. (2007) this is described as 'the chance that an arbitrary impostor fingerprint from a target population will share a sufficiently large number of minutiae with the query print, i.e. to characterise fingerprint individuality. According to several literature sources (Dass et al., 2007), (Dass et al., 2009), small values of the PRC imply that it is unlikely that minutiae in a fingerprint from a given source will match any other prints than those produced by the same source, which in turn indicates low uncertainty and high levels of individualisation of fingerprints. From Dass et al. (2009), we want to test the hypothesis of:

$$h_0 : I_t \neq I_c \text{ vs. } h_1 : I_t = I_c \tag{2.4}$$

where I_t is the print of unknown identity and I_c is the print of known identity.

Here some of the proposed models for describing PRC will be discussed and reviewed, as well as their limitations examined.

2.4.2 Methods

A common theme here is the use of minutiae details. The locations and orientations associated with the minutiae are the primary features in most models. A prominent paper in this field is Jain et al. (2002) where the authors discuss present thoughts on fingerprint individuality before making progress on deriving a method which is based on a hyper-geometric distribution. They discuss how many sources have characterised individuality by looking at the probability of a fingerprint configuration but there seems to be no agreement on the result with the answers ranging greatly. In addition to the investigation of other methods Jain et al. (2002) also propose one of their own. This is based on an input print (which we refer to as a fingermark), I and an arbitrary print (fingerprint) from a known source T. They state that a minutia from I corresponds and hence gives a positive result with a minutiae in H if their locations are within a specified distance r_0. The probability of this occurring by chance is derived as:

$$P(\text{corresponding locations by chance}) = \frac{\text{area of tolerance}}{\text{total area of overlap}} = \frac{\pi r_0^2}{A} = \frac{C}{A}, \tag{2.5}$$

where A is the total area of overlap between I and H which have been aligned using the singularities (core and delta). If there is only one minutia in I then the probability that it finds a corresponding minutia in H by chance is mC/A if there are m minutiae in H. This can be generalised for n minutiae in I as:

$$P(A, C, m, n) = \binom{n}{1}\left(\frac{mC}{A}\right)\left(\frac{A - mC}{A - C}\right). \tag{2.6}$$

Jain et al. (2002) show that by using algebraic manipulation the probability of exactly ρ corresponding minutiae between I and H with n and m minutiae respectively, is given by:

$$p(M, m, n, \rho) = \frac{\binom{m}{\rho}\binom{M-m}{n-\rho}}{\binom{M}{n}} \tag{2.7}$$

which is a hyper-geometric distribution of ρ with parameters M, m and n; with $M = A/C$.

The Jain et al. (2002) paper does highlight some major limitations for the given model. The model is restricted to ridge ending and bifurcation minutiae alone (discounting knowledge from ridge detail). This causes the model to miss some vital information. The x and y coordinates for the locations of the minutiae within the print are assumed to be from a uniform distribution for each, contradicting other literature which identifies some clustering around points of significance (Dass et al., 2007). A further assumption made by many models, including this one, is that of uniform ridge distance over both the whole print and the whole population. This of course cannot be strictly true. However the sort of variation required was too complex for many to include in their research. With further statistical modelling it would be possible to introduce this sort of variation. Unlike Jain et al. (2002), Dass et al. (2007) do not assume that minutiae locations are distributed independently of orientation since it can be seen that minutiae in differing regions are associated with differing orientations; extending this point, minutiae in close proximity tend to have very similar orientations.

The generative model in Su and Srihari (2008) is based on the assumption of non-independence of minutiae locations and orientations which differs from most of the other models discussed. Srihari and Su claim that due to low quality in latent prints ridge features need to be utilised and hence they embed this information into existing

models by developing a distribution for ridge points. The aim of this method is to establish a more precise representation of fingerprint individuality. As in Dass et al. (2005) and Dass et al. (2007) the mixture model used consists of a mixture of components; within each component, a bivariate Gaussian is used to represent minutiae locations and an independent Von-Mises distribution for minutiae orientations. The dependence therefore arises through the differences between the components. This model is then used to give the PRC as from a Poisson distribution with parameter from the model. Su and Srihari (2008) then build on the model proposed by Dass et al. (2005) by incorporating a joint distribution for ridge points based on their location and orientation. Srihari and Su use a Chi-squared test to assess the goodness-of-fit of the mixture model containing ridge information and the one without; they found that the model including ridge information offered a better fit to the observed fingerprints. The PRC taken from the model with ridge information is never greater than the PRC from the model without which implies that ridge information improves the quality of the estimates and strengthens individuality. Another paper calculating individuality is Dass and Zhu (2006) in which a compound stochastic model is used in the calculation. They compare their estimates with that of Jain et al. (2002), despite the estimates being of much higher magnitude. Dass et al. feel that the estimates are more realistic since they closely match the empirical distribution from fingerprint data sets.

A recent innovative paper in this area is by Dass et al. (2009); here they investigate fitting a hierarchical mixture model for the distribution of an observation of a print in the population. Here the top level of the hierarchy corresponds to the groups or individual fingers and the second level are representations of the fingers. Researching available databases Dass and Li found a covariance structure that they feel represents the population of fingerprints. Also in Dass et al. (2009) an approach using RJMCMC (Reversible Jump Markov Chain Monte Carlo) is used to explore the posterior distribution resulting from the hierarchical mixture model.

2.4.3 Results

In Jain et al. (2002) the individual components of the model are derived by obtaining many impressions of fingerprints and creating thousands of matches so as different quantities can be established by averages etc. Dass et al. (2007) carry out an experiment

by finding the best fitting mixture model for each finger in their database and testing this in two ways. The first is by comparing the likelihood ratios based on the mixture model and the uniform distribution for each finger. The second is by doing a goodness of fit test (again for both the mixture model and uniform) with a large p-value (greater than 0.01) corresponding to the adequacy of the tested model. These tests gave strong evidence that the mixture model was preferred for their data set. Using current data Dass et al. (2009) carried out some investigation into their hierarchical method. In a real life scenario the print held in a database (for example AFIS) would have a large number of minutiae and be very clear. However this contrasts to a mark lifted from a crime scene which may only be a partial print, be of poor quality and contain few minutiae. In this case the model by Dass and Li showed that the PRC could be as large as 0.0614 making it more difficult to identify positively a suspect. Furthermore the model proposed in Dass et al. (2009) has the main drawback that it involves a lot of computing time and power, full mixing takes a very long time although this computing would only need to be done once.

The work in Dass et al. (2007) can be improved upon by looking at the spatial dependence between minutiae instead of assuming independence and also by investigating other distributions for the model. Dass et al. (2009) have pointed out that their hierarchical model can be developed in the future by incorporating hierarchical mixture models containing minutiae locations and directions. In addition in Dass et al. (2007) the standard mixture distributions were proposed for each fingerprint individually however a model for the minutiae distribution for a population of prints is needed so that inferences can be made about the PRC for the population.

2.5 Conclusions

This review of current techniques has highlighted areas where the existing research is lacking. Many proposals for fingerprint generation are available however none seem to approach the issue from a statistical background; here this research would help develop a novel technique. All of the methods involve some random elements and algorithms. By representing the variability in fingerprints in a statistical way it will be easier to see how current methods can be improved to produce more reliable results. Although the approach in Champod et al. (2005) is valid for computing likelihood ratios, simply

setting out a formula does not make it usable and hence more problems arise in trying to implement this form of working. Due to this further research into modelling the variability (both within and between sources) is needed in order to establish reasonable results. A benefit of the Bayesian approach of investigating likelihood ratios means that it allows a combination of different types of evidence that may present itself in a criminal case and allows a universal way of representing evidence in court. Finally by using statistical techniques to model the variability in fingerprints a better understanding of the individuality of fingerprints will be gained.

Chapter 3

Data

3.1 Introduction

This chapter will provide details about the data used for this research, which is split into three main types: data provided by the Forensic Science Service (FSS), a dataset of dummy fingermarks and data obtained from my own fingerprints.

The first datasets were provided by the Forensic Science Service and consist of information from roughly 12,000 real fingerprints which were taken under controlled conditions. This dataset provides valuable knowledge about minutia location, orientation and type amongst other details. We intend to use these to investigate the relationships between minutiae locations, orientations and types of patterns. The different fingerprint patterns are categorised by a "level one feature" which visually identifies it in comparison with other fingerprints.

The second dataset used is referred to as dummy fingermarks. Dummy fingermarks are artificially simulated utilising the fingerprints provided by the FSS to give realistic representations of fingermarks which could be obtained from a crime scene i.e. fingermarks that were not obtained under the controlled conditions of the first dataset. In Section 3.3 further information is given about how these are generated from the FSS datasets.

Finally in Section 3.4 I investigate my own fingerprints and obtain information from them about minutia location, orientation and type so that some initial crude matching

can be done to give an idea of how much variation there is between replicates of the same fingerprint recorded at different times.

3.2 Data sets from FSS

Data was provided for this research from the FSS, consisting of a database with details from 12,096 real fingerprints from approximately 12,000 people. The fingerprints were captured using a digital scanner at 1:1 magnification and with a resolution of 500 pixels per inch (ppi) and then the landmarks (minutiae, core and delta) were located using a combination of software (landmark detection) and the human eye. The database holds information about four different finger types, index, middle, ring and thumb, and also three types of level one feature, arch, ulnar loop and whorl. The level one features for arch and whorl are as shown in Figure 1.2, the ulnar loop could be what has previously been described as a left or right loop, depending on which hand it is found. It specifies the direction of the loop based on the ulnar bone in the arm, the opposite being a radial loop (although these weren't featured in our data set). Figure 3.1 demonstrates the difference between radial and ulnar loops, essentially the turning point in the ridge flow is on the side closest to the thumb for ulnar loops. Summary information from the FSS dataset can be seen in Table 3.1.

Each dataset holds a significant amount of information about each type of fingerprint; below in Figure 3.2 is a screen grab of part of the database illustrating a small section of the information held in the database. In Figure 3.2 each row relates to one fingerprint and details can be seen about the x and y co-ordinates for the core and delta for each fingerprint as well as the orientation of these landmarks relative to the axes of the x-y plane. In addition, further information is held in the datasets which cannot be seen in Figure 3.2; this relates to information regarding the minutiae. For each minutia in a fingerprint we have details about the x and y coordinates (for the location at which the ridge flow is interrupted), the angle (given in degrees, where 0 degrees corresponds to a horizontal vector from the origin along the x axis of the x-y plane and moving anticlockwise. Angle is determined by the direction of ridge flow at the point in the co-ordinate system where the ridge flow is interrupted by ending or forking, type (bifurcation or ridge ending, as defined in Section 1.3.1) and an index for the specific minutia. All of the coordinates (for the core, delta and minutiae) are given using the

Figure 3.1: Simple pictorial representation of the difference between the two level 1 one patterns; radial loop and ulnar loop. All are ulnar loops except the thumb which is a radial loop.

Finger	Level 1 Feature	Number of Prints	Minutiae per Print				
			Min	Quartile 1	Median	Quartile 3	Max
Index	Arch	660	42	66.75	76	86.25	134
	Ulnar loop	1996	44	75	85	95	150
	Whorl	660	49	89	97	109	161
Middle	Arch	659	46	75.5	86	98	176
	Ulnar loop	660	87	87	90	100.25	145
	Whorl	659	60	89.5	100	111	163
Ring	Arch	660	30	74	84	97.25	155
	Ulnar loop	659	50	82	94	105	161
	Whorl	659	49	92	102	115	159
Thumb	Arch	830	43	88	102	118	192
	Ulnar loop	1998	48	99	113	129	237
	Whorl	1996	73	114	127	142	227

Table 3.1: A table showing the breakdown by finger type and level one pattern type of the 12,096 fingerprints in the FSS database, where Min=minimum and Max=maximum

Cartesian coordinate system by using the bottom left of the scanned image as the origin.

3.3 Dummy Marks

As the research progresses it becomes vital to use some fingermarks to compare against the fingerprints obtained from the FSS database. However we do not have access to any true fingerprint and fingermark pairings so it was necessary to simulate some from the database of fingerprints that we already have (i.e. the FSS database of approximately 12,000 fingerprints). In order to create a dummy fingermark (FM) associated with a given fingerprint (FP) we start by taking a subset of the FP from an area. To create a FM with N minutiae we select a minutia at random from the FP and then take its $N-1$ nearest neighbours. Nearest neighbours are selected (as opposed to selecting minutiae at random) as this gives the most realistic representation of the kinds of fingermarks that are likely to arise in practice. This gives us a set with N minutiae. The locations of this subset of minutiae need to be perturbed slightly to allow for the distortion that

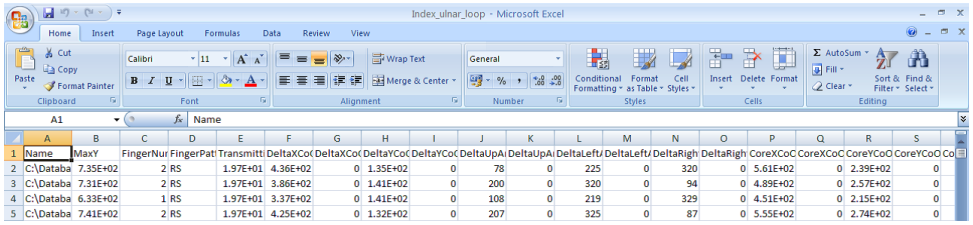

Figure 3.2: Screen grab from FSS database for the subset Index Ulnar Loop showing details of the locations and orientations of the core and delta for each fingerprint, which corresponds to each row in the spreadsheet

would occur when a fingermark is left at a crime scene or retrieved by the crime scene technician. Distortion of fingermarks is a complicated process, the initial distortion being the transfer of a 3D pattern into two dimensions which can be dependent upon the pressure applied by a finger and surface that the finger is pressed against. However this is complicated further by the fact that we are comparing two representations of this process through the fingerprint and fingermark. Neumann et al. (2012) use a method set out in Bookstein (1989) to model distortion in the fingermarks. However in this research we do not use a distortion method which allows directional dependence (for computational reasons), instead opting to use the Rayleigh distribution. The Rayleigh distribution has probability density functions as follows (Cox and Cox, 2010):

$$f(r;\sigma) = \frac{r}{\sigma^2} \exp \frac{-r^2}{2\sigma^2}, \tag{3.1}$$

where r is the length of the vector $S = (x, y)$, it can be shown that this is equivalent to a Bivariate Normal with a mean of zero, zero correlation and a variance of σ^2. This value changes for each experiment and is specified individually in each section.

Since minutia type is not always visible in fingermarks or recorded accurately the type is retained with probability 0.9 and changed to the opposite minutia type (bifurcation or ridge ending) with a probability of 0.1. An example of a FM and its corresponding FP are given in Figure 3.3.

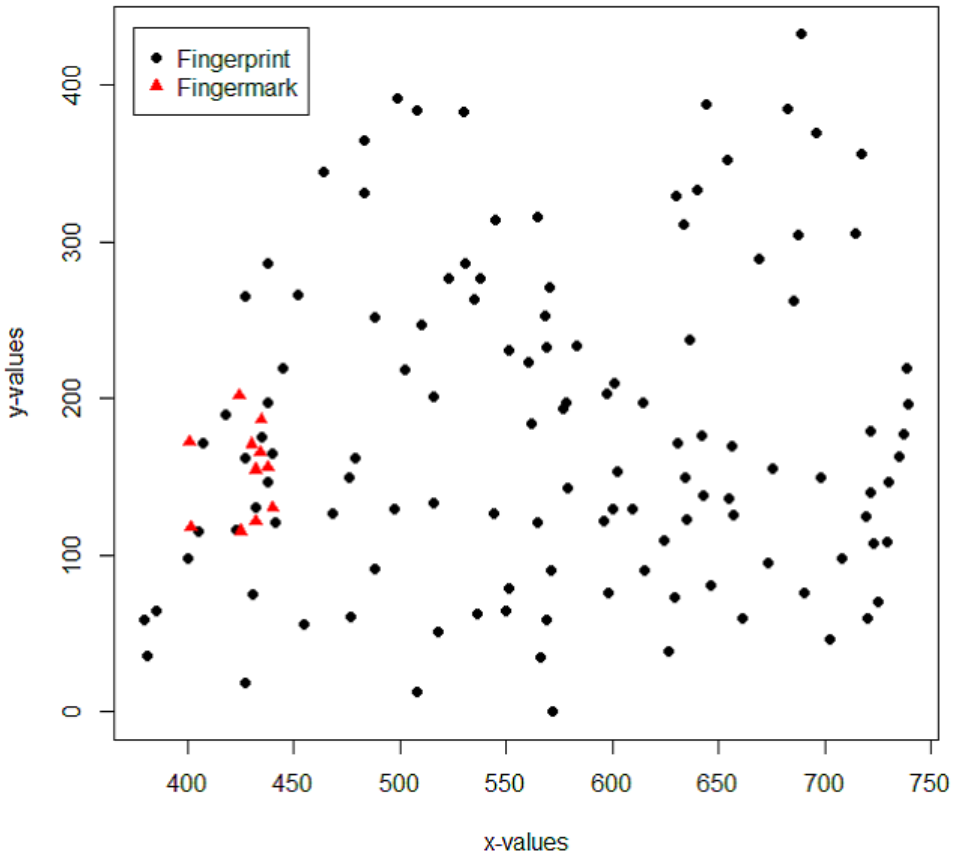

Figure 3.3: Plot of locations of the minutiae in a simulated fingermark (red triangle) and the associated fingerprint (black circle) where the locations are perturbed by sampling from a Bivariate Normal distribution

3.4 My Prints

In order to investigate the practicalities of matching fingerprints and fingermarks I decided to analyse two scans of the same finger taken on separate occasions (these can be seen below in Figure 3.4). I scanned my own fingerprints using a scanner donated by the FSS; this scanner is also used in conjunction with software which processes the scan. Next I needed some method for locating minutiae in both replicates so that we could visually assess the match; I did this by using a crosshair program downloaded from the internet from http://life.bio.sunysb.edu/morph/ called tpsDIG2. This program allows you to locate points by clicking on an image; it then stores the locations of all of these points in a new file. I used this program to compare the minutiae I could visually detect from the two scans taken of the same fingerprint but on separate occasions. I then plotted the file containing the locations of the minutiae obtained using the crosshair program to provide a representation of the two scans; this can be seen below in Figure 3.5. This plot shows that despite both of these scans being taken in controlled conditions using the same scanner it is not a simple task to compare them; one obvious issue is where in the capture window (or co-ordinate system) the finger is scanned. Both times the fingerprint was taken using the scanner and I tried to scan my print in the centre of the window but this cannot be done exactly and hence the representation of the finger is different with each scan. Another element which contributes to different representations of the same fingerprint is the amount of the actual finger scanned; this depends on the angle and pressure used when scanning. All of these issues can explain why the two sets of minutiae do not match up at all; and it is not clear (without seeing Figure 3.4) how to try to "register" them.

Figure 3.6 shows a copy of my full set of prints taken using the scanner from the FSS; a full set similar to this would be taken from a suspect in order to make multiple comparisons from all possible fingermarks left at a crime scene. There are several different types of fingerprints shown in this set including rolled and static prints for the thumbs, and rolled prints for the other fingers individually, palm prints and the writers palm. The most unusual of these is the writers palm; this is the side of the hand running from the tip of the little finger to the wrist; it is called the writers palm as it is typically the impression left by the side of the hand when you apply pressure onto a writing surface.

Figure 3.4: Two scans of the same finger (Left = replication 1, Right = replication 2) taken on separate occasions using an electronic fingerprint scanner

3.5 Summary

This chapter outlined the three types of data used for the analysis in this thesis. The three types of data are the database of real fingerprints provided by the Forensic Science Service, the set of dummy fingermarks we simulated from the FSS database and finally my own set of fingerprints which were obtained using a digital scanner. These various types of data will be used in the analysis carried out in later chapters. We needed to simulate fingermarks in order to evaluate matches between fingermarks and fingerprints to obtain the probability that the fingermark is obtained from the same fingerprint; the goal of this is so that a system can be used in court which relies on a numerical match as opposed to a categorical decision of match, non-match or inconclusive.

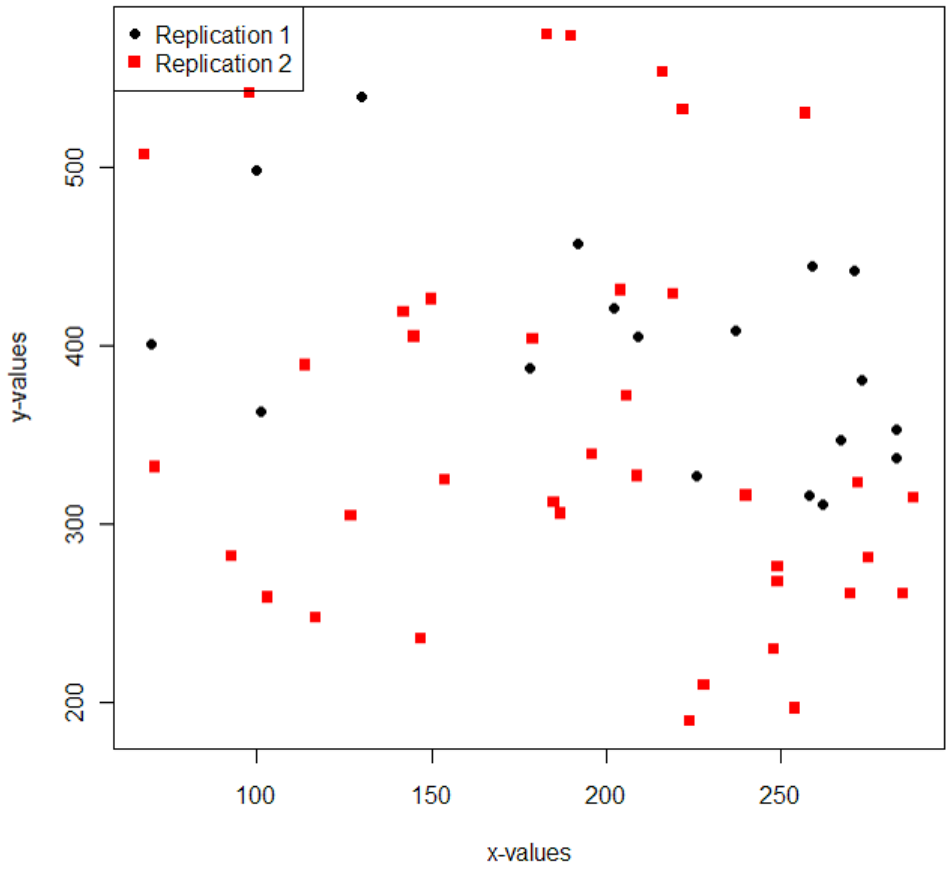

Figure 3.5: Plot of the minutiae from two scans of the same finger taken on different occasions. The minutiae have been captured using a crosshair program. Replication 1 = black circle, Replication 2 = red square

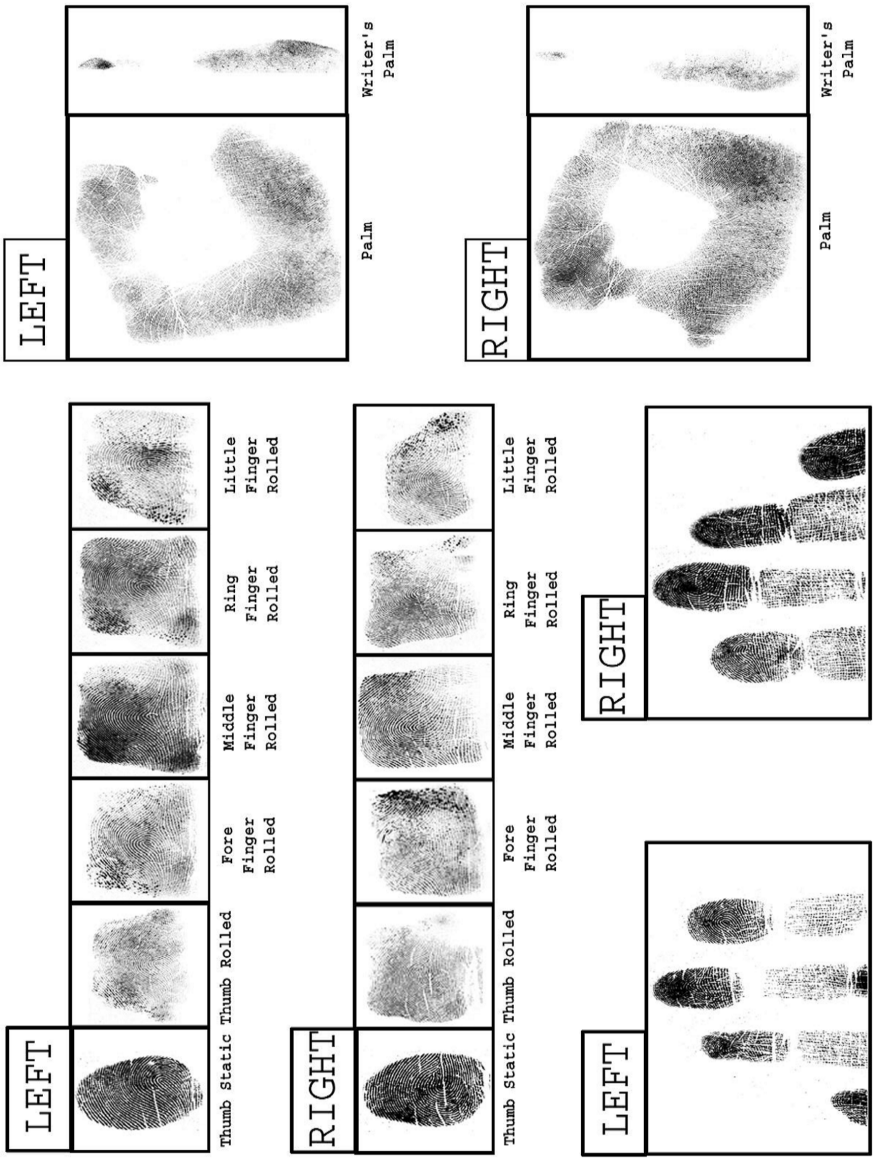

Figure 3.6: Full set of my prints taken using a digital scanner, including rolled and static prints, in addition to the palm and writers palm

Chapter 4

Assessment of Complete Spatial Randomness

4.1 Introduction

We can think of the minutiae in a fingerprint as events from a spatial point pattern; each of these events could have a "mark" corresponding to the minutia type, either ridge ending or bifurcation, or the minutia orientation, or both. In this chapter we consider whether the point pattern of minutiae in a fingerprint adhere to the hypothesis of Complete Spatial Randomness (CSR). In order for a point pattern to demonstrate CSR (Diggle et al., 1983):

1. The number of events in a given region A with area $|A|$ must follow a Poisson distribution with mean $\lambda|A|$, and
2. Given n events x_i in a region A, the x_i are an independent random sample from the uniform distribution on A.

We refer to λ as the intensity of the process; this is also equivalent to the mean number of events per unit area. In order for a fingerprint pattern to be described as showing CSR the intensity of minutiae must be constant over the whole area (or observation window) of the fingerprint and the minutiae must not interact with each other, this meaning that a minutia is not more or less likely to be at any position given the other

minutiae in the pattern. We are testing for CSR as although we know anecdotally that the minutiae tend to be located around the singularities (core and delta) in a fingerprint, this hasn't been formally tested. Here we intend to do this so that we can further our understanding of the distribution of minutiae within each fingerprint, in order to assess whether there is a more complicated process taking place which we can model.

This chapter focuses on the assessment of CSR for the fingerprints in the dataset provided by the Forensic Science Service 3.2. There will be an initial visual assessment of the fingerprints as well as a basic exploration of the location and orientation of the minutiae, including density plots. In addition more sophisticated techniques such as K-functions will be used to aid in the analysis.

4.2 Plotting Programs

Initially the datasets were explored by plotting the information for one fingerprint. The datasets described in Section 3.2 provide information about a variety of aspects of the fingerprint, some of which include; locations of minutiae, types of minutiae and orientations of minutiae. Several stages to producing a plot of the fingerprint were carried out, the first of these being to plot the x and y coordinates of the minutiae. This simple plot can be seen in Figure 4.1; from this it is difficult to visually assess if the locations of the minutiae are independent or if they show any dependent structure between their locations. So next we moved on to add information about minutia orientation; this is defined as the direction the ridge was flowing when the minutia occurred, the locations of the core and delta, and minutiae types. This can be seen in Figure 4.2. From Figure 4.2 it is now obvious that the fingerprint is a loop as the information about orientation gives us an idea of the ridge flow in the fingerprint.

Next we look at whether the relationship between locations and orientations of the minutiae as well as if these variables could be independent. Initially we plot the minutiae locations and orientations using the Uniform distribution, where the locations are Uniform on $[0,1]$x$[0,1]$ and the orientations independently Uniform on $[0, 2\pi]$. It can be seen in Figure 4.3 that when locations and orientations for minutiae are generated using this method the result does not look like a fingerprint and a level one pattern is

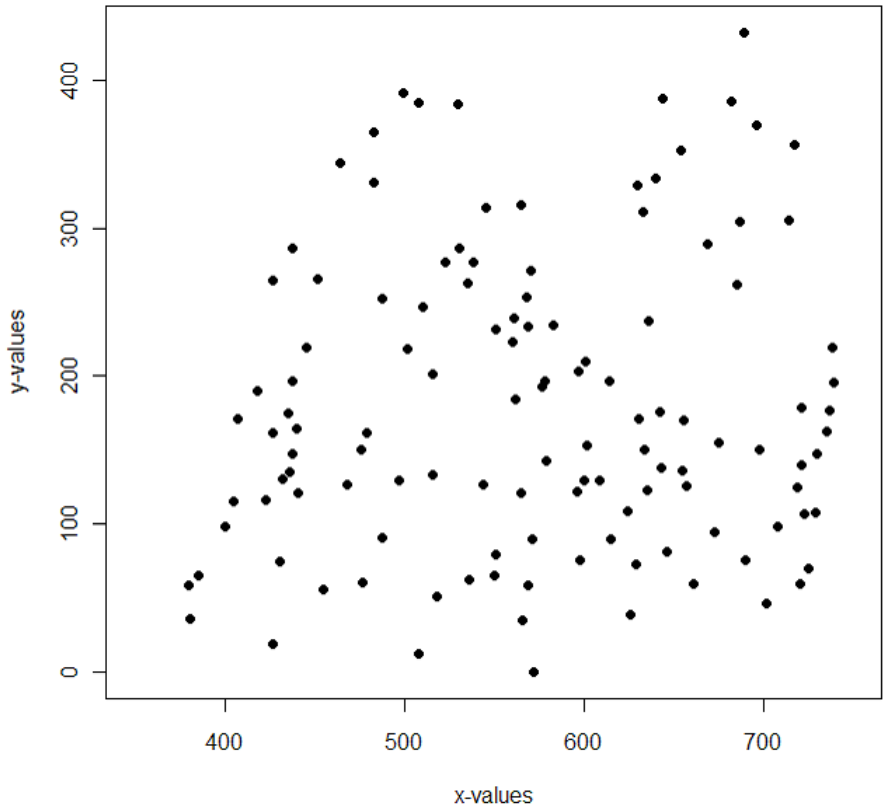

Figure 4.1: Plot of the x and y coordinates for a single fingerprint from the dataset

not identifiable. Figure 4.3 shows that at least the orientations of minutiae show some relationship with the location of the core and delta since the ridge flow is not what would be expected. Sherlock and Monro (1993) produced an algorithm for identifying orientations of minutiae based on their location relative to the core and data. The orientation θ at the point $z = (x, y)$ is:

$$\theta = \frac{1}{2}\left[\sum_{j=1}^{n_d} \arg(z - d_j) - \sum_{i=1}^{n_c} \arg(z - c_i)\right] \tag{4.1}$$

where c_i, $i = 1 \ldots n_c$ and d_j, $j = 1 \ldots n_d$ are the coordinates of the cores and deltas respectively. Usually for loops we have $n_c = n_d = 1$, for whorls $n_c = n_d = 2$ and arches $n_c = 0, n_d = 1$. The result of using this algorithm for producing minutiae orientations can be seen in Figure 4.4. Despite locations of the minutiae still being random,

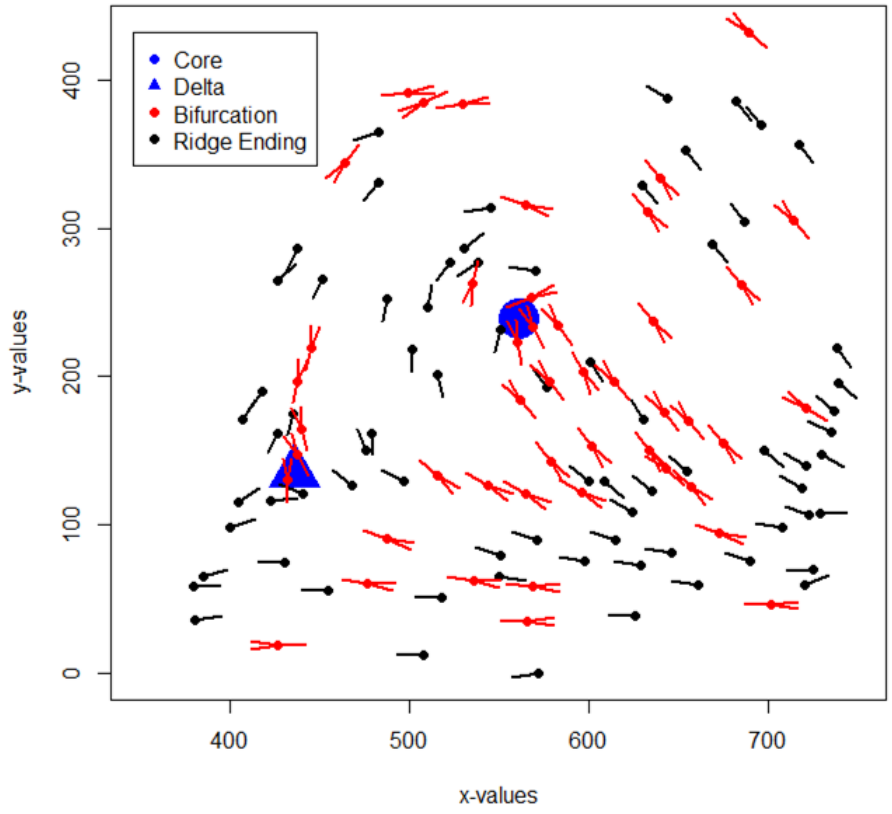

Figure 4.2: Plot of the x and y coordinates for a single fingerprint from the dataset with minutiae types, orientations and the core and delta identified

generated by a Uniform distribution, the plot looks like a realistic representation of a fingerprint. This shows that orientation of the minutia is related to its location on the print in relation to the core and delta and can be determined. In Figure 4.5 a visual comparison of the orientations of minutiae in a real fingerprint and the orientations produced by the Sherlock and Monro algorithm can be seen. The real orientations are identified in red. In most parts of the fingerprint both sets of orientations are very similar with the exception being the top right corner where the orientations differ slightly.

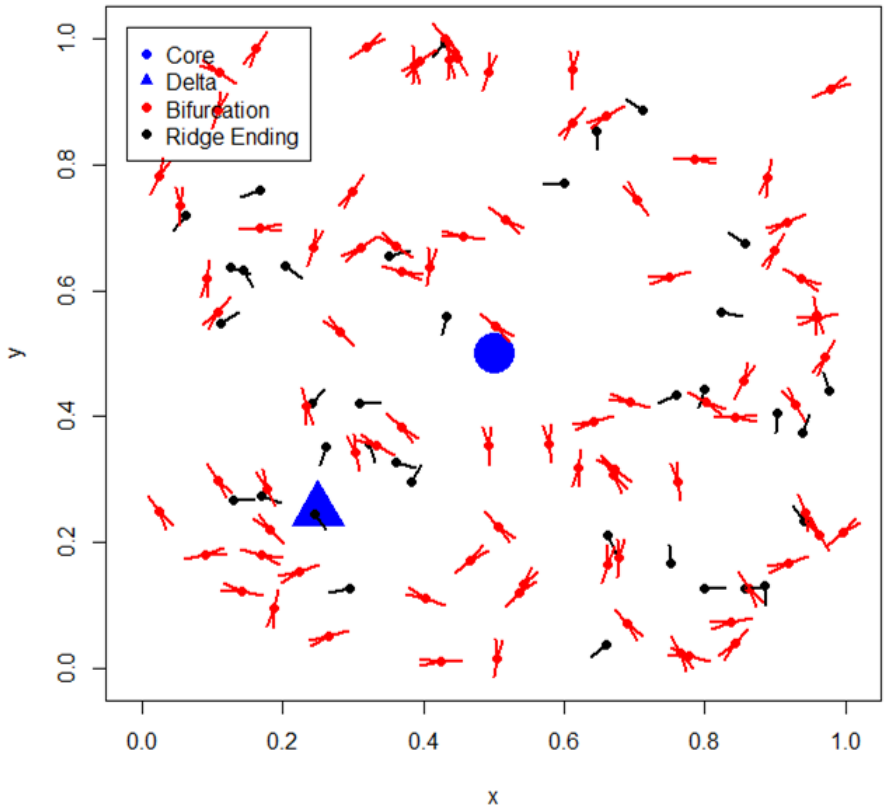

Figure 4.3: Plot showing minutiae locations and orientations as generated by a Uniform distribution

4.3 Density Plots

Since we have established that the orientations of minutiae produced by the Sherlock and Monro algorithm is fairly realistic, we move on to focus on minutiae locations. By looking at density plots of the minutiae we can determine whether they are dense in specific regions of the fingerprint.

Figure 4.6 shows a density plot for all minutiae in a fingerprint. It demonstrates that minutiae are more clustered around the points of significance (i.e. core and delta). This can also be seen in Figures 4.7 and 4.8 which are density plots categorised by minutiae types.

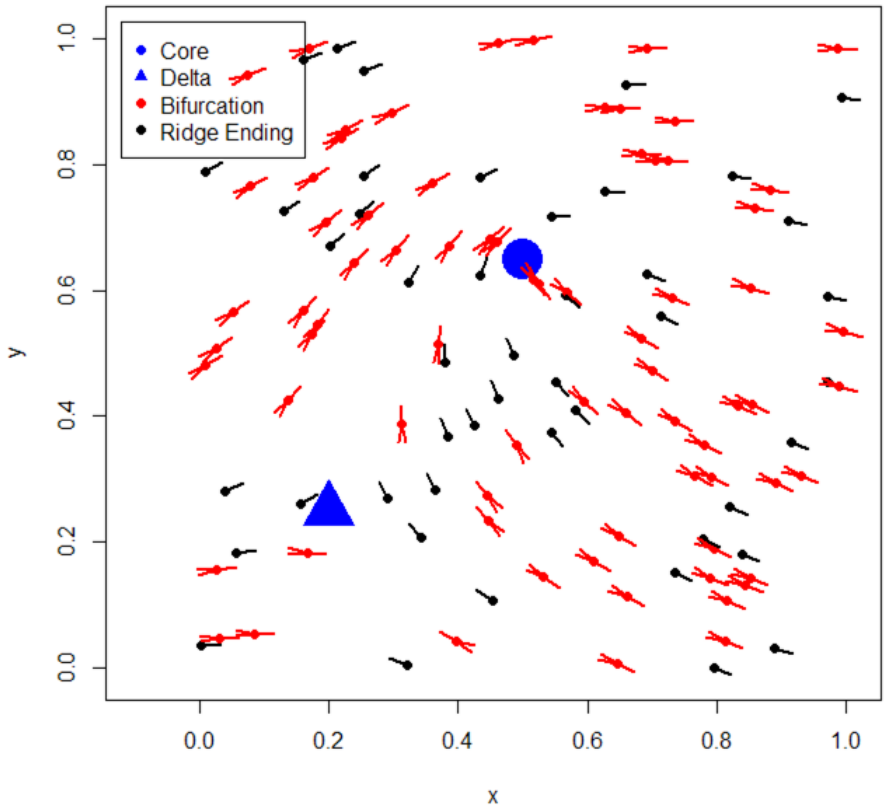

Figure 4.4: Plot showing minutiae with random locations and orientations generated by the Sherlock and Monro algorithm

4.4 Plotting Functions

4.4.1 Scaling

In order to compare the fingerprints more formally using summary functions such as the K-function we need to scale the fingerprints so that they are registered. It is useful to have each fingerprint on the same scale; since fingerprints are not square we keep the aspect ratio the same and scale the largest axis (be it horizontal or vertical) to 0 and 1 and then scale the other axis accordingly.

Next we calculate a convex hull of the points, i.e. the smallest convex region containing them, to ensure that we don't have lots of white space around the edges and edge effects are taken into account when calculating the functions. Figure 4.9 shows a plot with

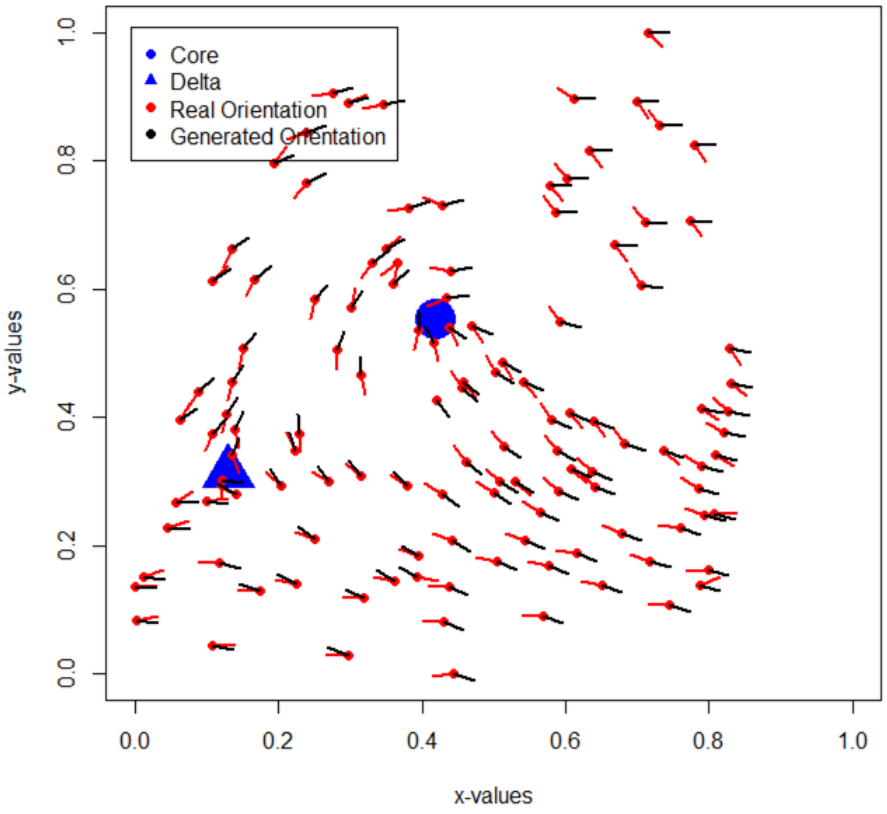

Figure 4.5: Plot showing a comparison of the orientations for a real fingerprint and those orientations generated by the Sherlock and Monro algorithm

minutiae and the convex hull plotted around the points.

4.4.2 K-Functions

In order to look more formally at clustering of minutiae within fingerprints we make use of K-functions. The definition of a K-function is (Ripley, 1977):

$$\mathrm{K(r)} = \lambda^{-1} E\,[\text{number of further events within distance } r \text{ of an arbitrary event}] \quad (4.2)$$

where λ is described as the intensity function; the mean number of events per unit area.

Comparison of the subsections within the database is considered in order to explore the fingerprint data further. The K-function for every fingerprint in a given dataset is

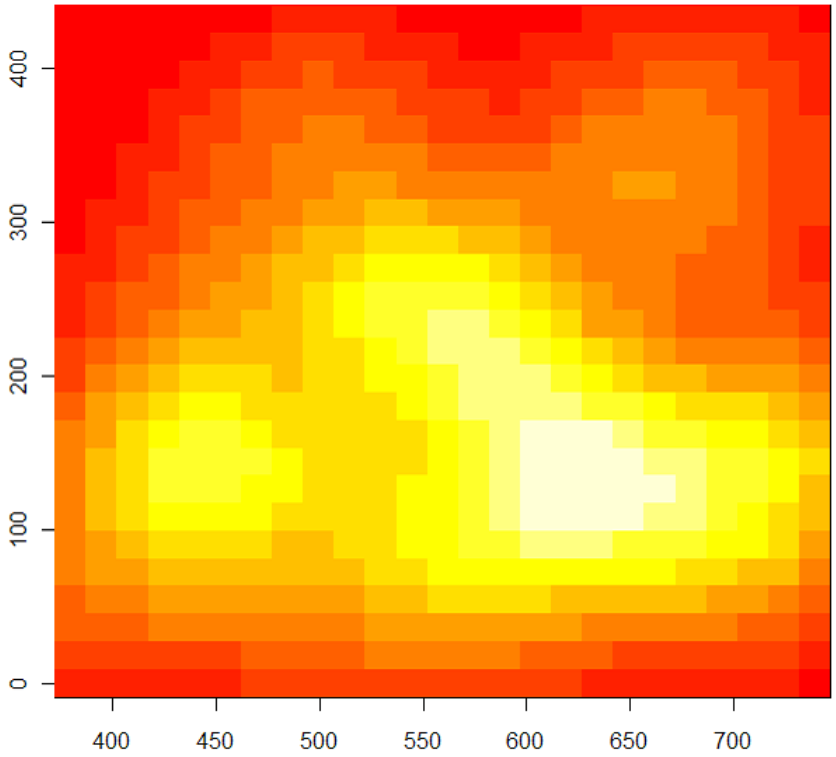

Figure 4.6: Density plot for all minutiae in a loop

found and the mean at each value of r is taken, this gives a "mean K-function". All minutiae from all fingerprints within a dataset are pooled and then a sample taken (say 150 minutiae) and then the K-function calculated, this is referred to as the "sample K-function". A comparison is then made between these two K-functions, this can be seen in Figure 4.10 for the dataset Index Ulnar Loop. If there is simply varying intensity between fingerprints we would expect these two values for the K-function to be very similar and to not show any large differences, however if there is an actual dependence structure between minutiae, it is not so obvious a result. Figure 4.10 shows that the "sample K-function" and "mean K-function" are indeed very similar, although both show a more clustered pattern than the theoretical K-function for a Poisson process (complete spatial randomness) which is defined as πr^2. A more clustered pattern is demonstrated by one whose K-function falls left of the theoretical function since this relates to more points per unit area. However despite both seeming slightly more

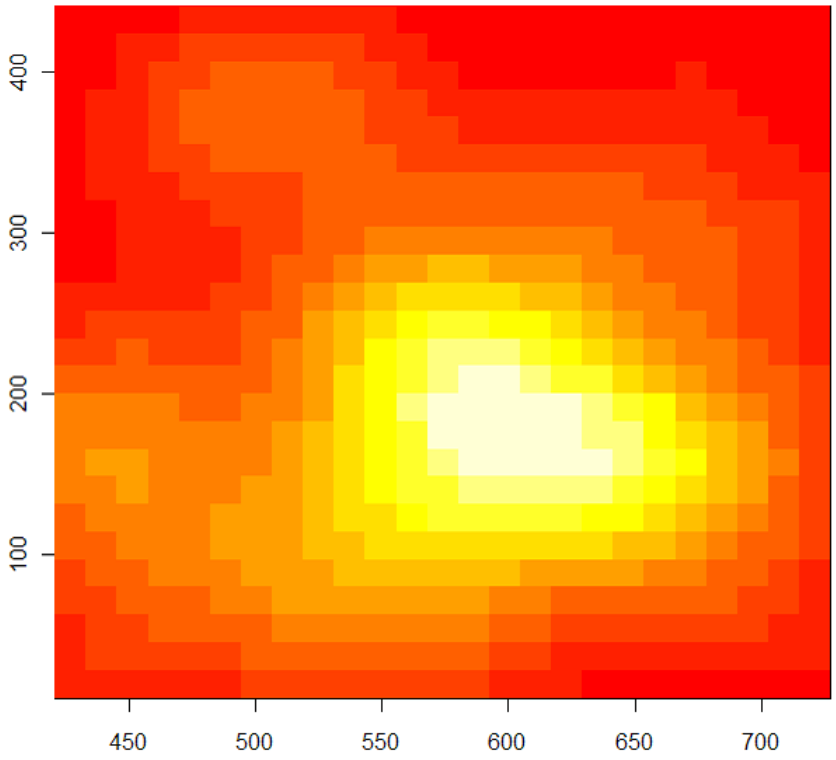

Figure 4.7: Density plot for minutiae of type bifurcation in a loop

clustered they stay within the envelopes for small values of r. The envelopes are calculated in R using the "spatstat" package and are pointwise envelopes.

Next an investigation was carried out into how the K-functions varied for each level one pattern according to which finger they were on. The same test was done comparing K-functions for each finger depending on the level one feature.

From visual inspection it is clear from Figure 4.11 and 4.12 that all four finger types do not behave in the same way. Thumbs seem to be indistinguishable from the theoretical K-function and hence we would conclude that they exhibit CSR. However the other types do vary from this. Index and Middle fingers appear more clustered and Ring fingers less clustered. Figure 4.13 also exhibits variation between patterns with Arch being fairly typical of CSR and Loops and Whorls being clustered.

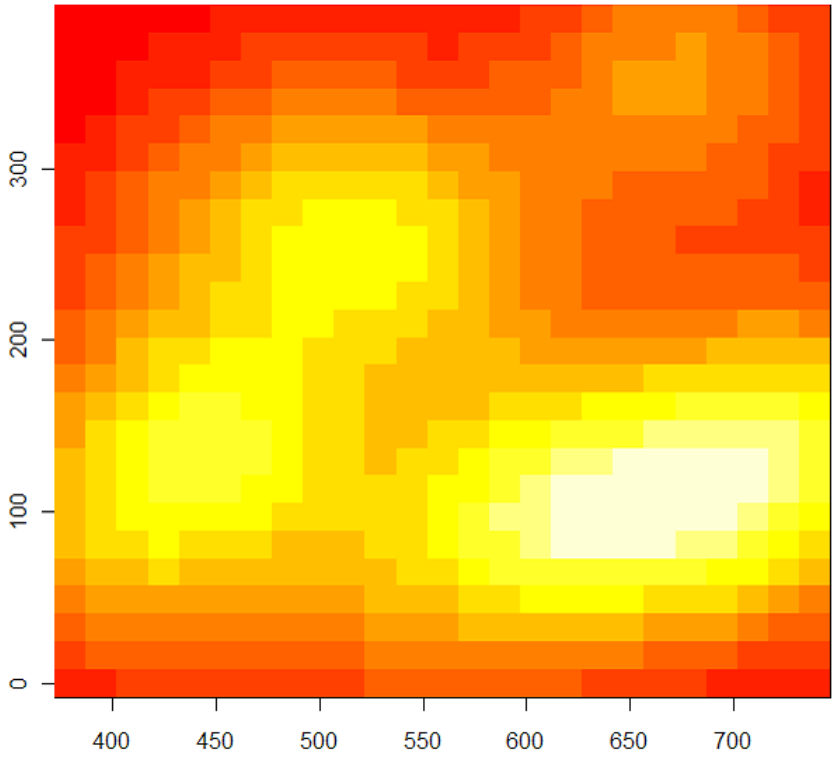

Figure 4.8: Density plot for minutiae of type ridge endings in a loop

4.4.3 Size Analysis

The results in the previous section appear interesting however it is worth carrying out some size analysis on the datasets to establish if the results actually represent an artefact of the data based on different numbers of minutiae within the sets. In order to explore this boxplots were created for the datasets based on level one pattern and finger type. These can be seen in Figures 4.14 and 4.15.

In Figure 4.14 it can be seen that on average Whorls have more minutiae in each fingerprint in the datasets provided by the Forensic Science Service. When considering finger type in Figure 4.15 thumbs have a higher mean number of minutiae. The Middle and Ring fingers have very similar distributions in the box plot however their K-functions are quite different.

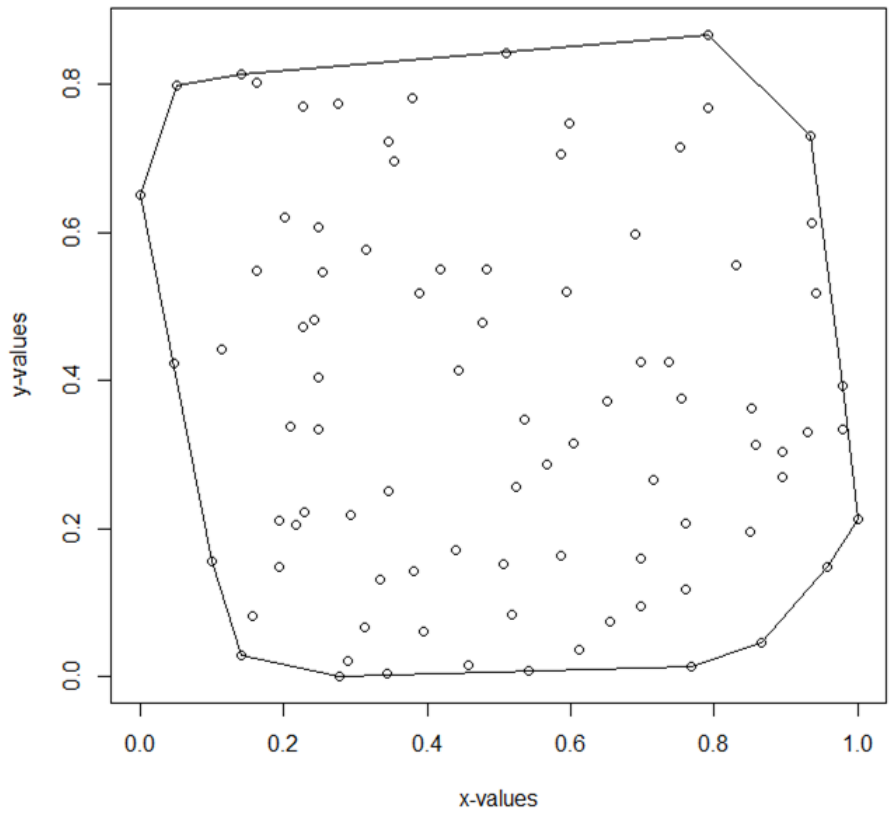

Figure 4.9: Plot showing minutiae locations and the corresponding convex hull

4.5 Summary

This chapter focused on investigating whether the pattern created by minutiae in fingerprints can be described as random. This was done by using the datasets provided by the Forensic Science Service and testing for complete spatial randomness. Initially a visual assessment of the points was made before moving on to look at other techniques. It was observed that minutiae tend to be located around the points of significance in the fingerprint, these being the core and delta. This aligns with anecdotal evidence from fingerprint examiners who observe a similar scenario where minutiae do not tend to be as prevalent towards the edges of the fingerprint. Despite this there was only a little evidence to suggest that the pattern made by minutiae in fingerprints demonstrate some dependence structure.

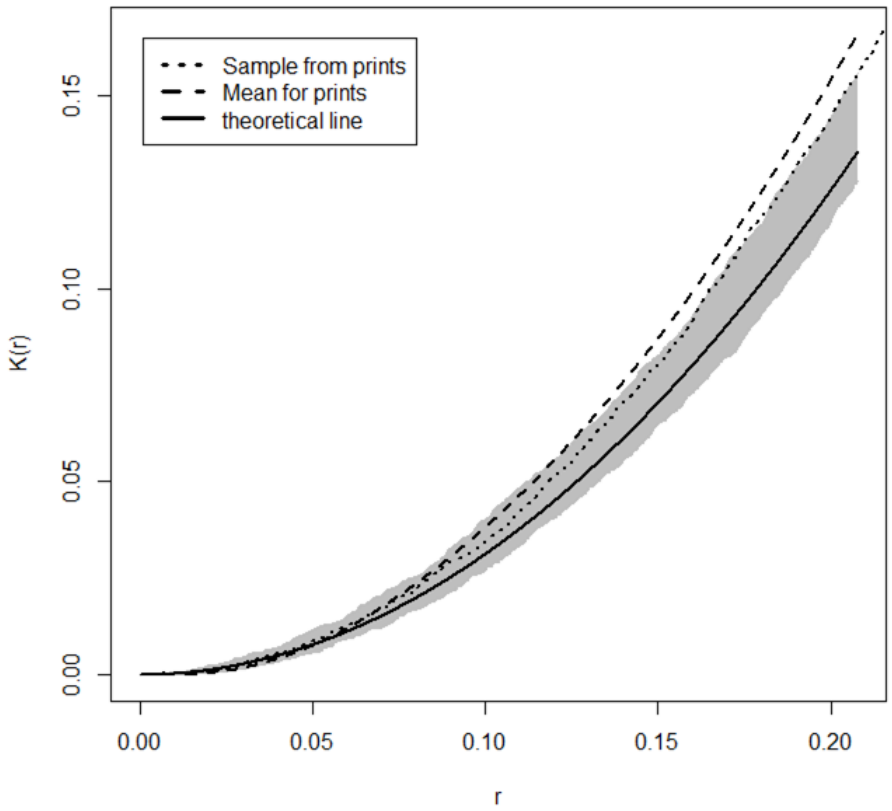

Figure 4.10: Plot showing a comparison between the "mean K-function", "sample K-function" and theoretical function for the dataset "Index Ulnar Loop". The grey band is 5% to 95% pointwise envelopes under CSR.

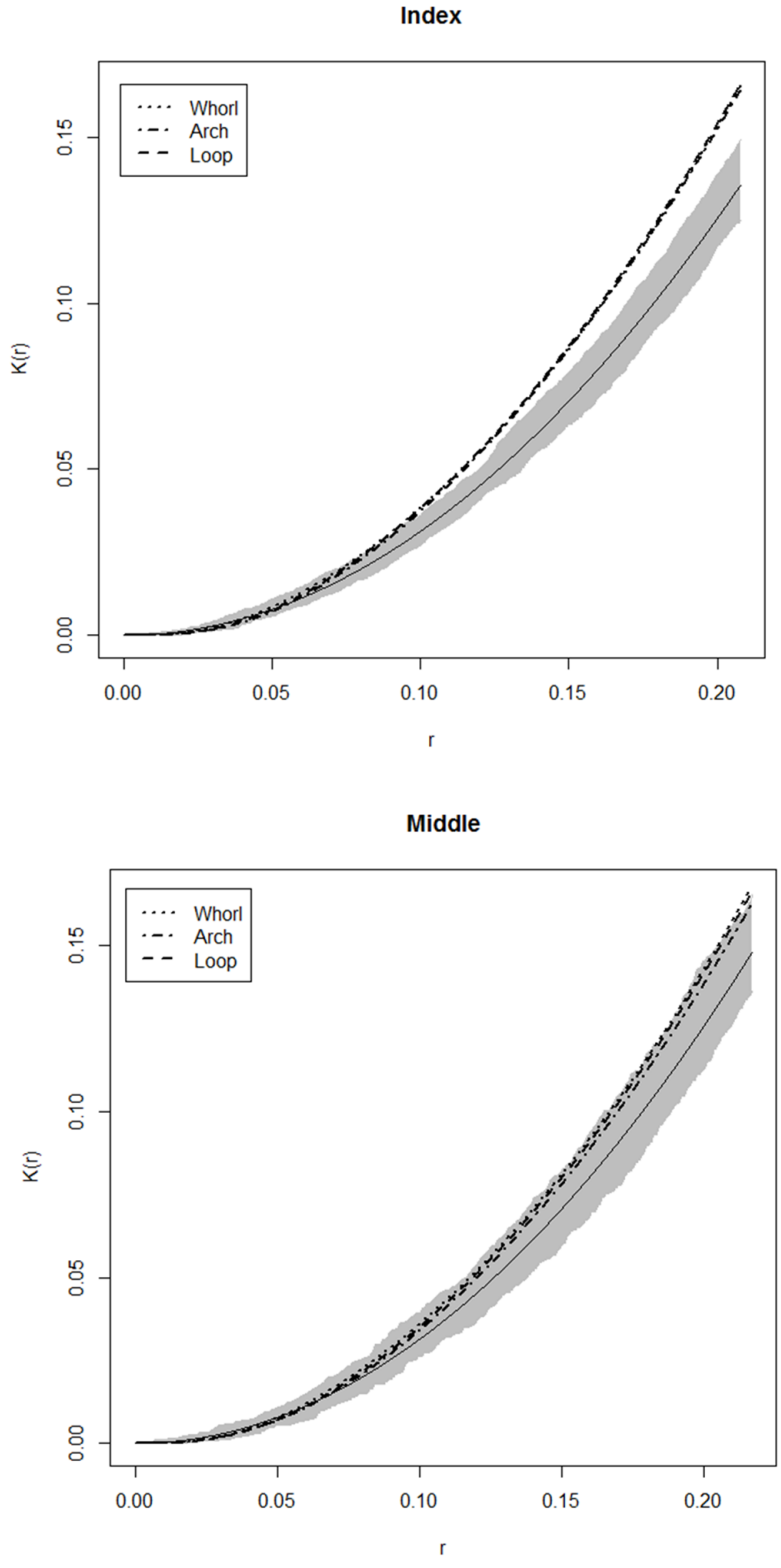

Figure 4.11: Plots showing the K-functions for each level one pattern according to finger type. The thin black line shows a theoretical K-function and the grey band is 5% to 95% pointwise envelopes under CSR.

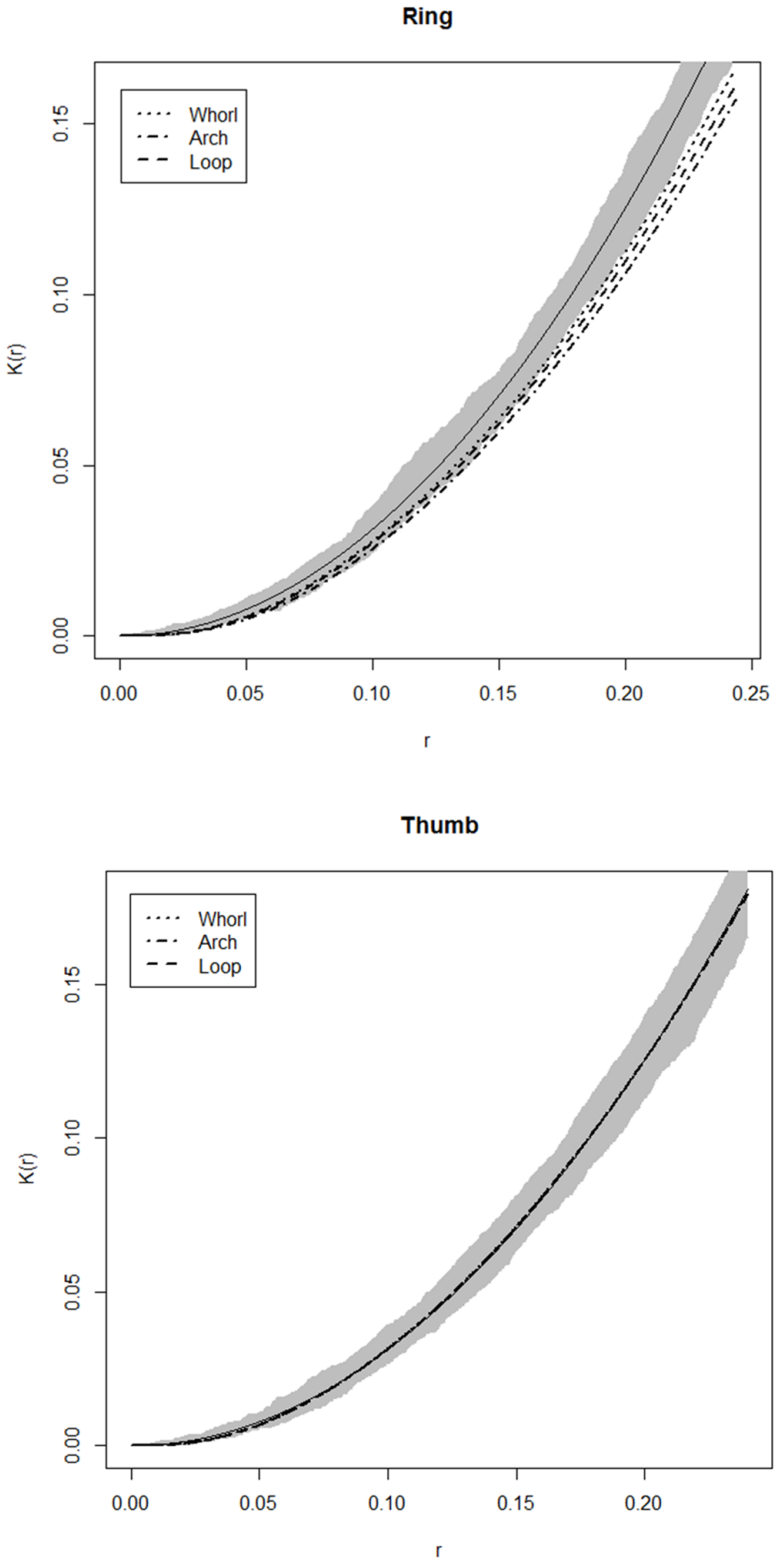

Figure 4.12: Plots showing the K-functions for each level one pattern according to finger type. The thin black line shows a theoretical k-function and the grey band is 5% to 95% pointwise envelopes under CSR.

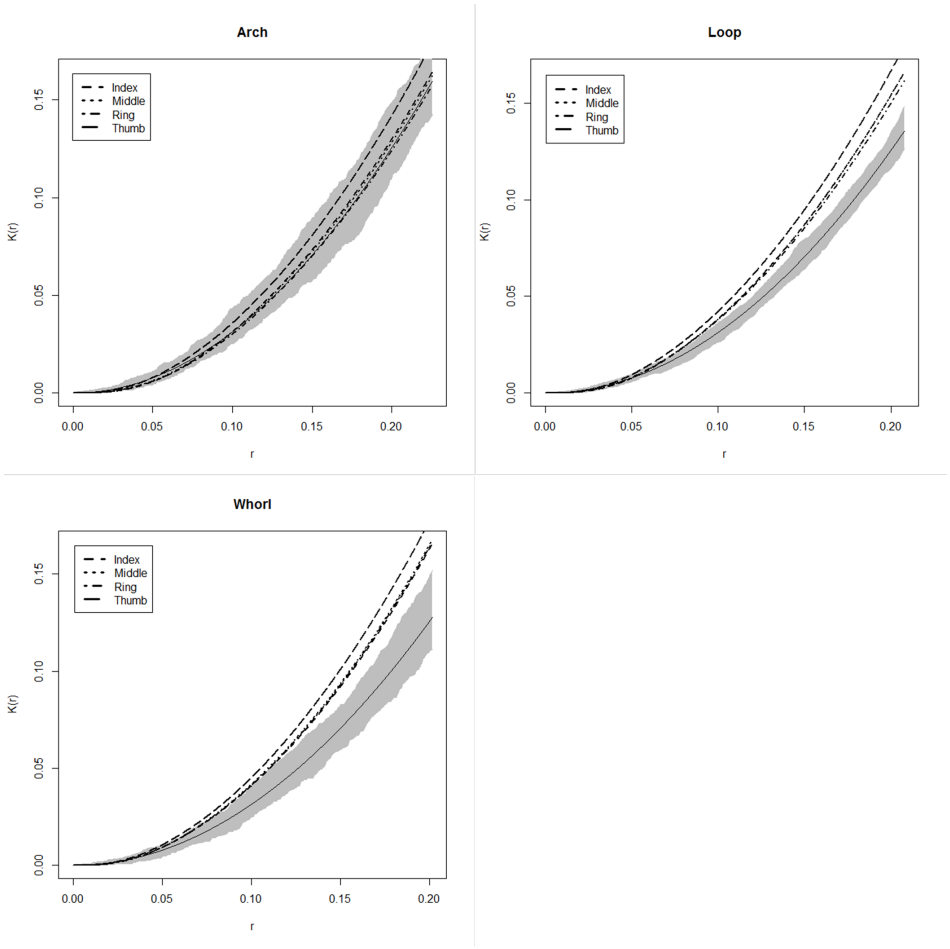

Figure 4.13: Plots showing the K-functions for each finger type according to level one pattern. The thin black line shows a theoretical K-function and the grey band is 5% to 95% pointwise envelopes under CSR.

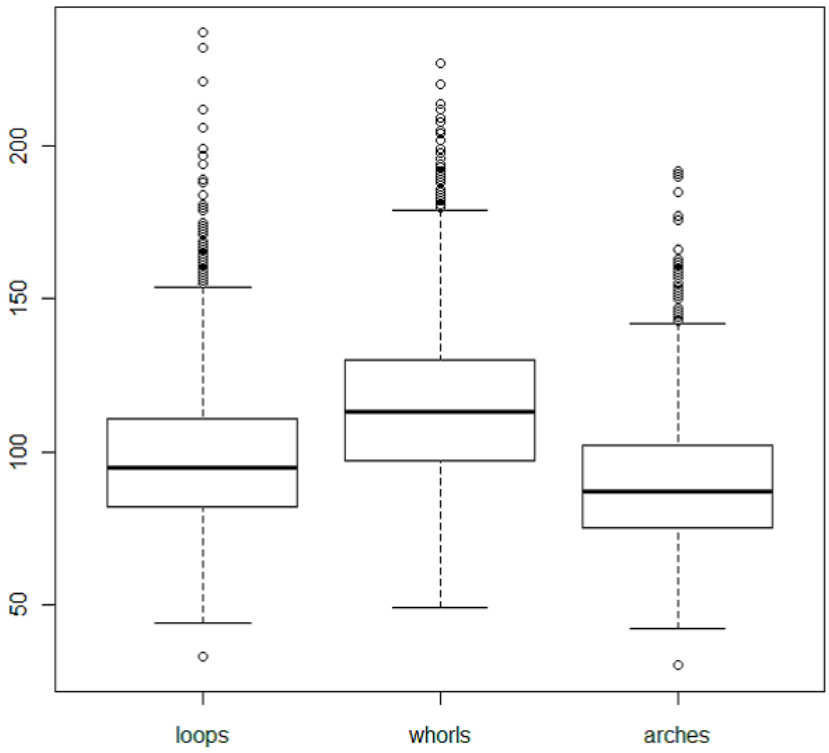

Figure 4.14: Boxplots showing the number of minutiae in each fingerprint according to level one pattern

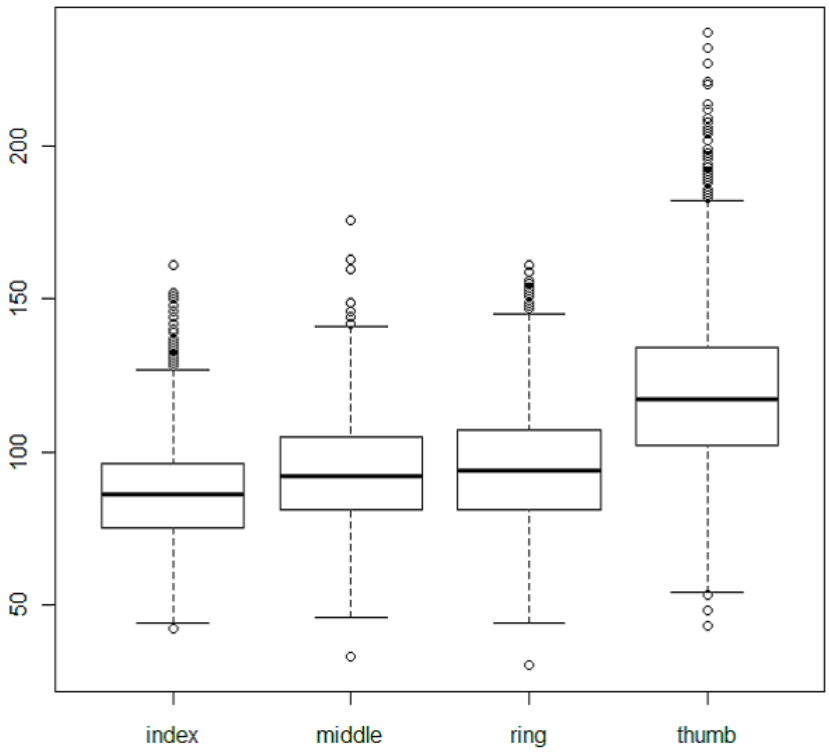

Figure 4.15: Boxplots showing the number of minutiae in each fingerprint according to finger type

Chapter 5

Model Fitting

5.1 Introduction

After tests for complete spatial randomness were carried out it was decided to pursue the thought that minutiae in fingerprints show some dependence structure and follow a pattern as opposed to being random. This was due to common thinking within the forensic community that minutiae are located more densely in specific regions of the fingerprint. In this chapter we will assess this notion more formally by trying to fit point pattern models to the data we have and testing how well these models perform.

5.2 Investigation of Models

Initially research was carried out into different models and the varying features they had. We narrowed down the wide selection of models by looking for one with a small number of parameters as this would be easier to work with. In addition models which are used to describe regular patterns were discounted as our data cannot be described as regular. It was also considered that models with the ability to be extended to multivariate data would be beneficial to this work as minutiae are described by two types, bifurcations and ridge endings.

In addition we limit the models we investigate to finite point processes. This is due to the fact that the fingerprint can be described as the realisation of the whole pattern.

It could be argued that the pattern of minutiae would continue if a larger impression of the finger were taken but this is certainly not infinite in a true sense. Furthermore this makes it easier when thinking about the observation window of the pattern since we assume we have viewed the whole pattern.

5.3 Strauss Model

Since we are looking for a model which takes into account that the points in the pattern are not independent, we need to consider a process which involves interaction amongst the points. There is reason to believe that the locations of minutiae in fingerprints could rely on the locations of other minutiae in the pattern. Hence we need a process in which we allow the points to interact. This leads to Gibbs processes (Cressie, 1993), also referred to as Markov point processes. Gibbs processes are unlike the Poisson process since the points in the pattern can interact with each other. In a Poisson process the points do not interact with each other and are distributed within the given area with intensity λ for a homogeneous process or $\lambda(x)$ for an inhomogeneous process. The simplest form of a Gibbs processes is with a fixed number of points (n) in a given area (commonly called the window). The multivariate probability density for the position of these n points is referred to as the location density function. For a Gibbs process with n points the location density function is (Illian et al., 2008):

$$f_n(x_1, \ldots, x_n) = \exp\Big[-\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\phi(||x_i - x_j||)\Big] \Big/ Z_n \tag{5.1}$$

From this it can be seen that the interaction between points is taken into account by looking at a function of the distance between x_i and x_j; this function $\phi(r)$ is referred to as the pair potential. Z_n is a normalising constant which is calculated so f_n integrates to 1. The value of $\phi(r)$ determines the type of pattern the process describes. The Strauss process (Strauss, 1975) is a type of Gibbs process which allows points within a radius r to interact (Illian et al., 2008) and minutiae with an inter-point distance greater than r to exhibit no interaction. Specifically it is defined as:

$$\phi(r) = \begin{cases} \beta & \text{for } r \leq r_{max} \\ 0 & \text{for } r > r_{max} \end{cases} \tag{5.2}$$

where $\beta > 0$ and r_{max} is the limit at which two points no longer exhibit interaction. A feature of this model is that any two points that fall within a distance r_{max} of each other have the same level of interaction, hence interaction does not vary as points are closer together. This model can also be extended to a Multi-Strauss, which allows for different types of points in the process; this is a useful option as minutiae types could later be incorporated into the model.

We start by using the "Strauss" function which is built into the "Spatstat" package in R. By inputing a value for r_{max} this function outputs the corresponding values for β as described above and γ which is a function of the number of pairs of points that fall within r_{max}. This function can't be used to obtain r_{max} as this is the input value, the values for β and γ are calculated in the function by maximising the pseudo-likelihood. Once these parameters have been obtained the combination of the three are used to simulate a Strauss process with these parameters. This process was carried out 1000 times and the corresponding K-functions were plotted along with the K-function for a random fingerprint. The intention is that the K-function for both patterns, the simulated Strauss process and fingerprint would be similar. The results of using the value $r_{max} = 0.02$ to obtain the Strauss parameters for simulation can be seen in Figure 5.1. Although the K-function for the fingerprint is within those for the Strauss process it does not sit well within the values, and there seems to be a problem with very low values of r. The whole process was repeated using different fingerprints as the test print and values of r_{max} ranging from 0.01 to 0.1 increasing by 0.01 each time. Another case where $r_{max} = 0.05$ can be seen in Figure 5.2. The result in this case was very similar to the previous in that although the fit does seem to be better for low values of r, for higher values it isn't as good.

The Strauss process appears to fit the set of fingerprints better than the K-functions for complete spatial randomness which were investigated in Section 4.4.2. However no values of r_{max} fit the data as well as we would expect and little improvement in the fit was obtained by doing this which led to the conclusion that the choice of model was not correct.

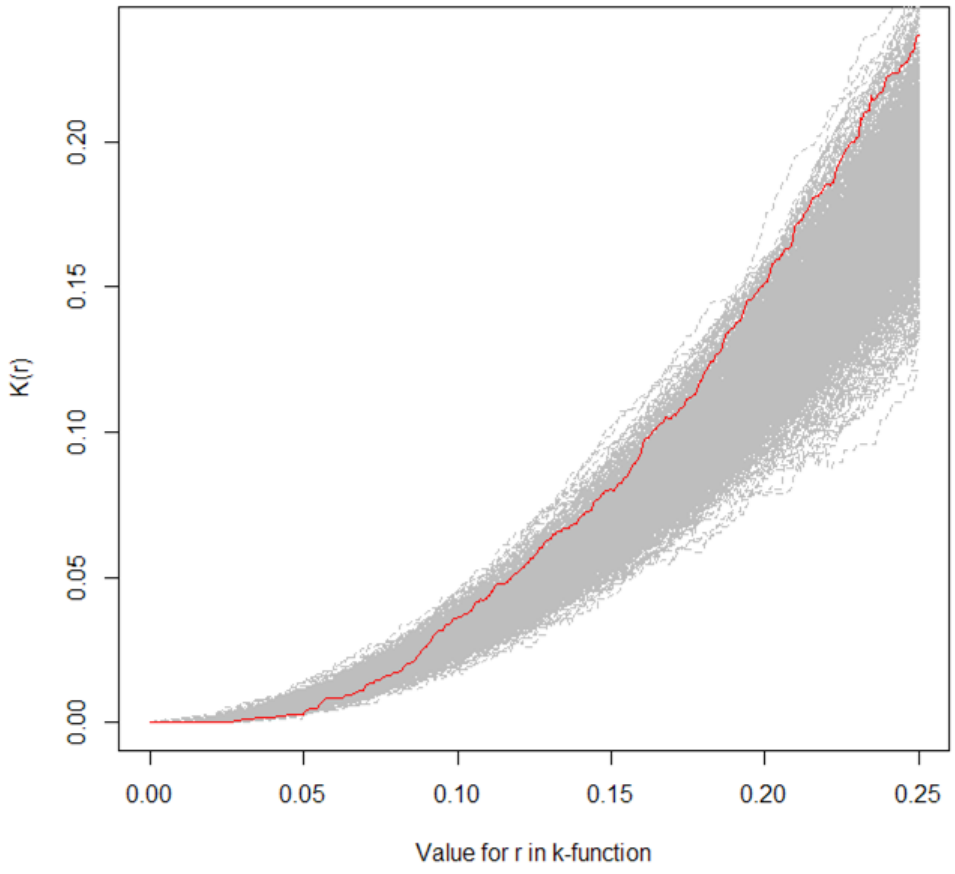

Figure 5.1: Plot comparing the K-functions for a specific fingerprint and simulated Strauss processes with $r = 0.02$

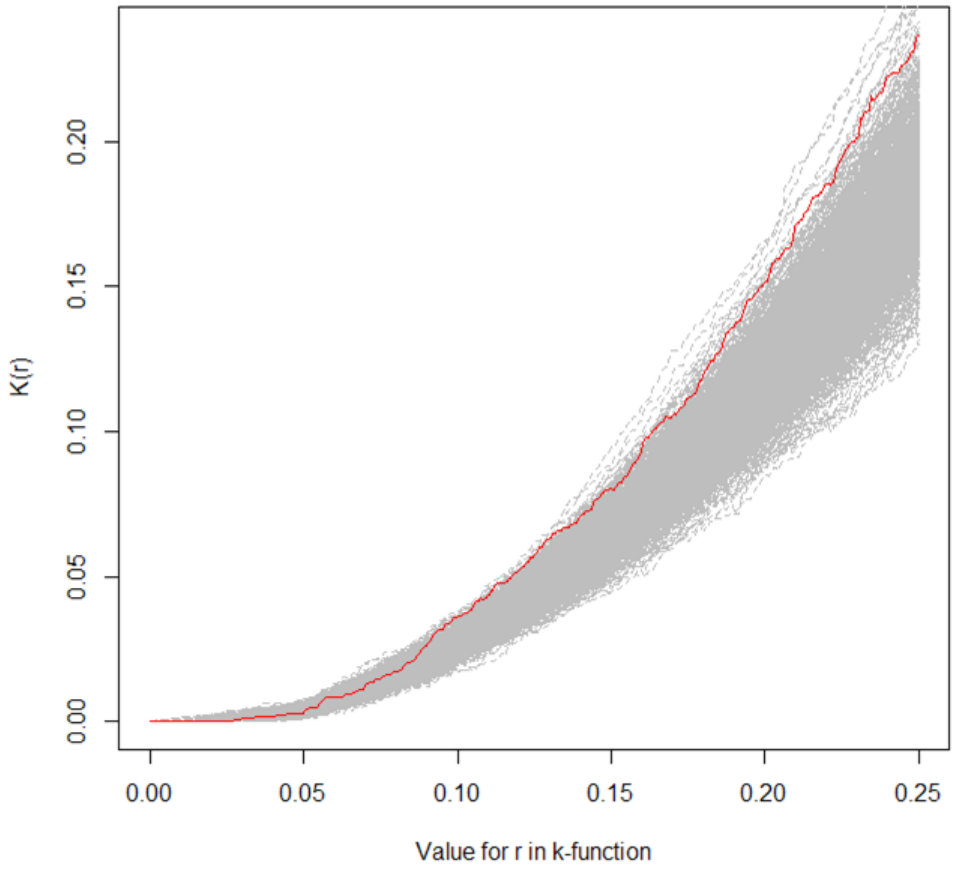

Figure 5.2: Plot comparing the K-functions for a specific fingerprint and simulated Strauss processes with $r = 0.05$

5.4 Isotropic Centred Poisson Process

The previous model used, a Strauss process, did not fit the data as we expected it to. The process refers to a pattern which has clustering throughout whereas a fingerprint can be described as the minutiae being more dense around points of significance (the core and delta) and then the distribution of minutiae reducing as we move away from these locations. For this reason an inhomogeneous Poisson process was examined instead. This refers to a Poisson process where the intensity of points in the pattern is not constant and thus changes depending on location, this is unlike the Gibbs process where the points interact to cause clustering. Here the points are clustered because the intensity varies across the window.

The model allows some level of randomness in the location of the points but also demonstrates an amount of clustering around the centre of the pattern. This is done specifically by creating an isotropic centred Poisson process. In this process the intensity function (λ) which is constant in a homogeneous Poisson process, now varies and for a point x_i is described as (Illian et al., 2008):

$$\lambda(x_i) = \lambda(r_i) = \frac{\alpha m}{\pi B} \exp\left\{\frac{-r_i^2}{B}\right\} \tag{5.3}$$

where r_i is the distance of point i from the centre, α a scaling factor, m is the number of points in the pattern and B is the parameter to be varied. Hence this is isotropic since the intensity depends only on r_i.

Using rejection sampling (see for example Ripley (1987); Lee (2004)) we can simulate an isotropic centred Poisson process as follows Illian et al. (2008). In step 1 a homogeneous Poisson process is simulated (this has constant intensity over the whole pattern). This pattern should contain many more points than are required in the final pattern since we now go on to thin these points using a thinning function which is determined by the model we are fitting (see Equation 5.4), it is this choice of model which reflects the experience of the fingerprint experts. This evidence from experts states that minutiae are more likely to fall around points of significance (the core and delta) so we need points closer to the centre of the pattern to be retained with a higher probability than those further away from the centre. We use a thinning function where a point x_i is retained with probability:

$$p(x_i) = \lambda(x_i)/\lambda^* \tag{5.4}$$

where λ^* is the number of points in the original pattern after step 1.

The parameter B from the intensity function will be estimated informally by trial and error. By changing B the probability of retention changes which means that the intensity of points moving away from the centre of the pattern changes also. We are attempting to estimate B so that the subsequent K-function for the isotropic centred Poisson process is similar to the K-function of any random fingerprint. In each case 1000 simulations are created for a given B. The results for different B can be seen in Figures 5.3 and 5.4.

It can be seen from Figure 5.3 that a low B does not match the K-function for the actual fingerprint very well but by increasing the value of B to 2 the fit is much better. This makes the envelopes sit comfortably around the actual K-function. The fit of this process is far better than both the Strauss process and for complete spatial randomness. It seems realistic to think that the given fingerprint could have come from the process used to create the simulated K-functions in Figure 5.4.

5.5 Summary

In order to represent the variability in the locations of minutiae in fingerprints we attempted to fit a spatial point process to the data. Different models were investigated for their suitability and then investigated by simulating from these models and comparing to the real data we have available to us. In the first case of the Strauss model the variability in the fingerprints could not be accurately captured. However, in the second model, the isotropic centred Poisson process, by changing the parameter B in the intensity function we could recreate the same K-function as the actual fingerprint used to create the simulation. It is useful to know this but in a forensic situation it cannot help us since the K-function can be variable and so the K-function for many different fingerprints could all be described by the same isotropic centred Poisson process. Without a process which is unique and distinct for every fingerprint we cannot successfully use the information to determine whether a fingermark from a crime scene comes from the same process as a known fingerprint as this would not give us a distinct

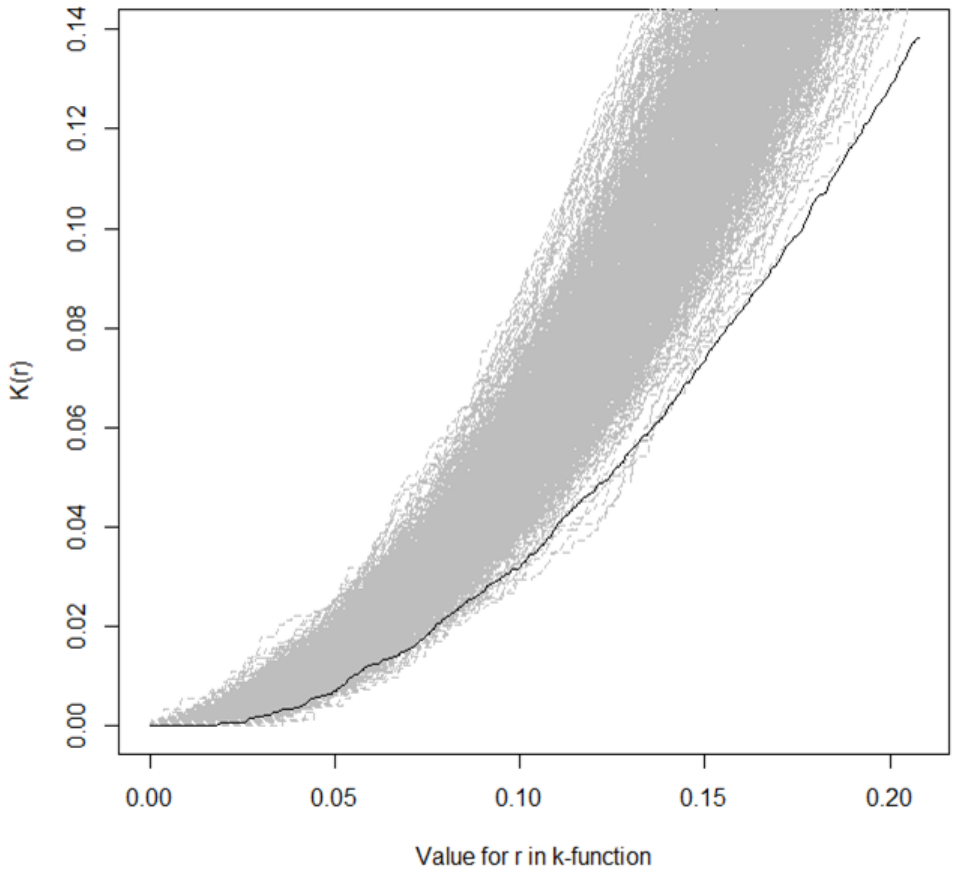

Figure 5.3: Plot comparing the K-functions for a specific fingerprint and simulated isotropic centred Poisson processes with $B = 0.1$

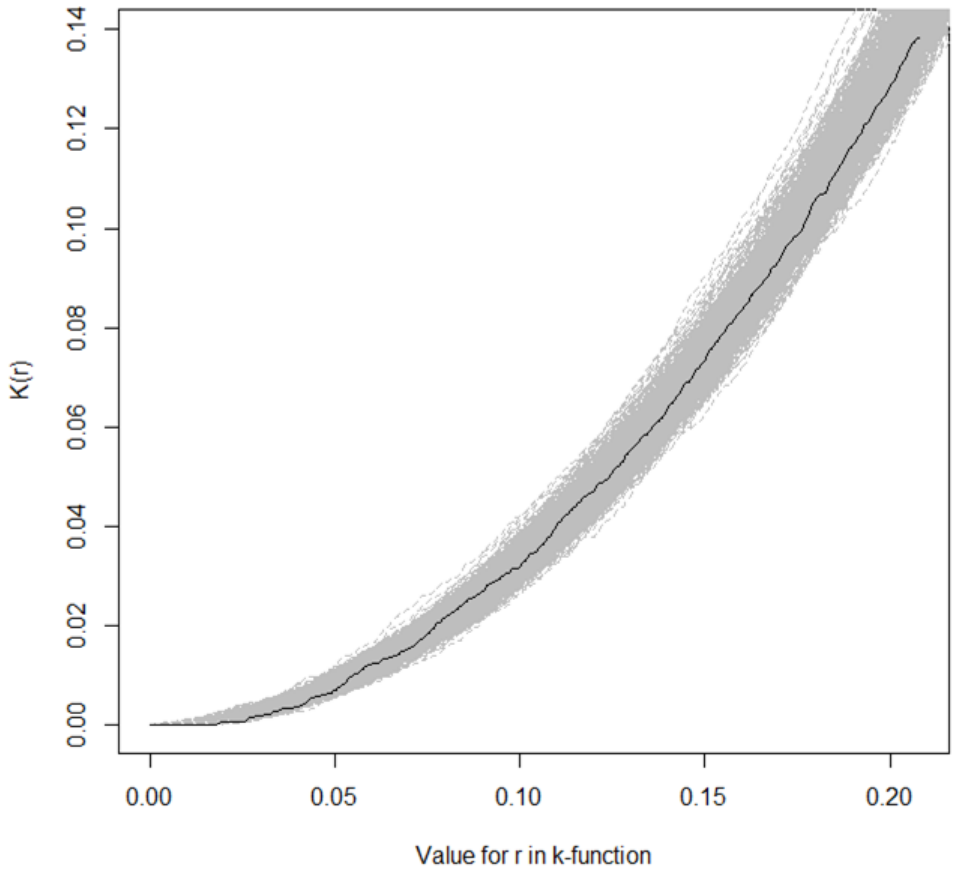

Figure 5.4: Plot comparing the K-functions for a specific fingerprint and simulated isotropic centred Poisson processes with $B = 2$

result. Instead we move on to look at a method which assesses the strength of a match between a fingermark and print.

Chapter 6

Matching Algorithm

6.1 Introduction

In a forensic situation the aim is to assess the strength of a match between the features in a fingermark taken from a crime scene and the features in a fingerprint from a known suspect. This helps to establish a person's presence at a crime scene and adds weight to the other evidence collected; ultimately the preferred goal of investigators is to solve a crime using the evidence from the scene. A fingermark is defined as an impression from a real finger left on a surface unintentionally. The fingermark is usually of low quality and can be distorted or incomplete; it is commonly referred to as a "partial print". Conversely, a fingerprint refers to a complete, undistorted version of the pattern on a finger taken in controlled conditions. This is thought of as a true realisation of the actual features displayed on the fingerpad.

As previously highlighted, currently a categorical system is used for the matching process whereby the fingermark is assigned as either "match", "no-match" or "inconclusive" against the fingerprint. By creating a new algorithm for matching a fingermark and a fingerprint we hope to introduce a numerical representation for the strength of the match which can then be presented in court; this would lead to a fairer system for analyzing evidence. In addition to this it would allow all matches to be presented in court, even those which would previously have been discarded as being "inconclusive". This would allow the layman juror to make a judgment based on clearer, more transparent results.

It is intended to create the matching method by developing a statistical algorithm which assesses the difference between a fingermark and a fingerprint and gives a quantitative measure for the strength of the match. The method used for assessing the difference will be based on the Euclidean distance between the sets of minutiae in the mark and in the print. In real life the fingermark contains far fewer minutiae than the fingerprint, and it can be thought of as being a subset of minutiae in the fingerprint. For this reason it is important to design a method which can take into account differing numbers of minutiae in the fingermark and fingerprint.

6.1.1 Method Summary

Details of the matching method developed, and references, are given in subsequent sections, but in outline it is as follows. Firstly, a triangle is defined within the fingermark by choosing three of its minutiae at random, and a set of approximately matching triangles in the fingerprint is determined and then called the candidate set. For each candidate triangle from the fingerprint, a similarity transformation is found that maps the initial triangle to the candidate triangle; this is then applied to the whole of the fingermark. Within each transformed fingermark, all minutiae are matched to distinct fingerprint minutiae, minimising the resulting sum of squared distances using a Hungarian Algorithm. The smallest such sum of squares, over all candidates, is a measure of how well the mark matches the fingerprint. In order to implement this custom R code can be found in Appendix A, this details the steps carried out in the method.

6.1.2 Fingermark Simulation

As there are no data about real fingermarks related to the fingerprints in the database used, marks were constructed manually. This was done by creating dummy marks as outlined in Section 3.3. To create a dummy fingermark with N minutiae a fingerprint is selected and then a random minutia is selected within the fingerprint. This minutia is used to find its $N-1$ nearest neighbours; this creates a subset from the fingerprint with N points. The $N-1$ nearest neighbours to the original minutia are used since this gives the most realistic version of a fingermark; a fingermark is usually just a small section of the whole fingerprint and so we expect minutiae located close together to be visible. Despite this, the method still holds true if some minutiae from the area of

the fingerprint are not present in the fingermark or if a section of minutiae cannot be determined, e.g. if the middle is smudged.

Some jitter is applied to the points to give a more realistic fingermark (i.e. to allow for distortion, smudging etc). The jitter for each minutia is sampled independently from a bivariate Normal distribution with zero mean, zero correlation and a variance of σ^2 on each coordinate, with a range of values of σ^2 being used in the experiments described below. Using a circular bivariate Normal distribution in this way is equivalent to adding $N(0, \sigma^2)$ jitter to each coordinate independently. Sometimes the minutia type is not recorded correctly since many fingermarks are of poor quality. To take account of this in the fingermark we retain the original minutia type (bifurcation or ridge ending) with probability 0.9 otherwise it is swapped to the other type, this value is based on anecdotal evidence and experience of working with fingerprint examiners.

6.1.3 Background to Transformations

In order to match a fingermark to a fingerprint in this matching algorithm a method of transformation is adopted. According to Cox and Cox (2010) the technique of matching one configuration of points in a Euclidean space with another and producing a measure of the match is called Procrustes analysis. That provides the basis for the comparison of the minutiae in the fingermark with the minutiae in the fingerprint. The theory used in this research is adopted from Cox and Cox (2010), in addition to that we also use Dryden and Mardia (1998) as Cox and Cox (2010) refer to it as a good introduction to Procrustes analysis. The specific transformation carried out here is called a Euclidean similarity transformation and has 3 components: translation, rotation and scale (referred to as dilation in Cox and Cox (2010)). A transformation which does not include scale (simply translation and rotation) is called a rigid-body transformation. From Cox and Cox (2010), the notation for a Euclidean similarity transformation of any point X_i from the configuration of points in X is defined as:

$$X_i' = \rho A^T X_i + b, \tag{6.1}$$

where ρ is the scale, A is the orthogonal rotation matrix, and b is the rigid translation vector. In our case we specifically work with 2-dimensions.

Procrustes analysis is a method of matching two (or more) shapes using a similarity transformation which minimises the Euclidean distance between the two (or more) sets of points. For the method developed to match a fingerprint and a fingermark we assess the strength of the match based on the sum of squared distance, and so choosing a transformation based on minimising this value, is useful. Dryden and Mardia (1998) call this process "Full Ordinary Procrustes Analysis" (OPA); the squared distance between them is known as the Ordinary Procrustes Sum of Squares (OSS).

According to Cox and Cox (2010) the first step is to centre the two configurations around the origin. This is done by subtracting the mean vectors for the configurations from each of the points. We give the notation here for calculating the OSS using the techniques in Dryden and Mardia (1998) since the associated package in R, "shapes", is used to carry out the method. Dryden and Mardia (1998) show that the "global" minimisation over all similarity transformations can be carried out in relatively simple steps. Firstly the shapes X and Y are centred, to give X^c and Y^c. Secondly X^c is scaled by a factor $\hat{\rho}$. Finally $\hat{\rho}X^c$ is rotated by a matrix $\hat{A}$. The OSS is then:

$$D^2_{OPA}(X,Y) = \|Y - \hat{\rho}X\hat{A}\|^2, \tag{6.2}$$

where $\|Y\|$ is defined as the Euclidean norm (or the square root of the sum of squared distance). The process for calculating $\hat{\rho}$ and $\hat{A}$ such that we have the minimal solution can be found in Cox and Cox (2010), here as stated previously we use the package "shapes" in R. To see this process carried out with notation specific to this application see Section 6.2.2.

6.2 Method

6.2.1 Candidate Set

The first step in the method is to select a triangle in the fingermark. This is done by picking three random minutiae and labeling them A, B and C to create triangle ABC. The goal is to then find a corresponding triangle in the fingerprint which we call abc, in Figure 6.1.

Initially, we allow some finite amount of error in the matching and so we come up with

a candidate set of possible triangles. To find triangles which fit into the candidate set we initially look for pairs of points in the fingerprint with distances similar, within some degree of error, to AB, AC and BC. This results in three tables, for example, the first table will contain information about the index of the pairs of points in the fingerprint ab which have a similar distance to that of AB in the fingermark. These three sets of pairs are then cross referenced to find triangles which would satisfy all three distances and then these are added to the candidate set. For example if an index for a appears in both sets for ab and ac then the set for bc is searched to see if the indexes for b and c occur together and so represent a similar distance to BC.

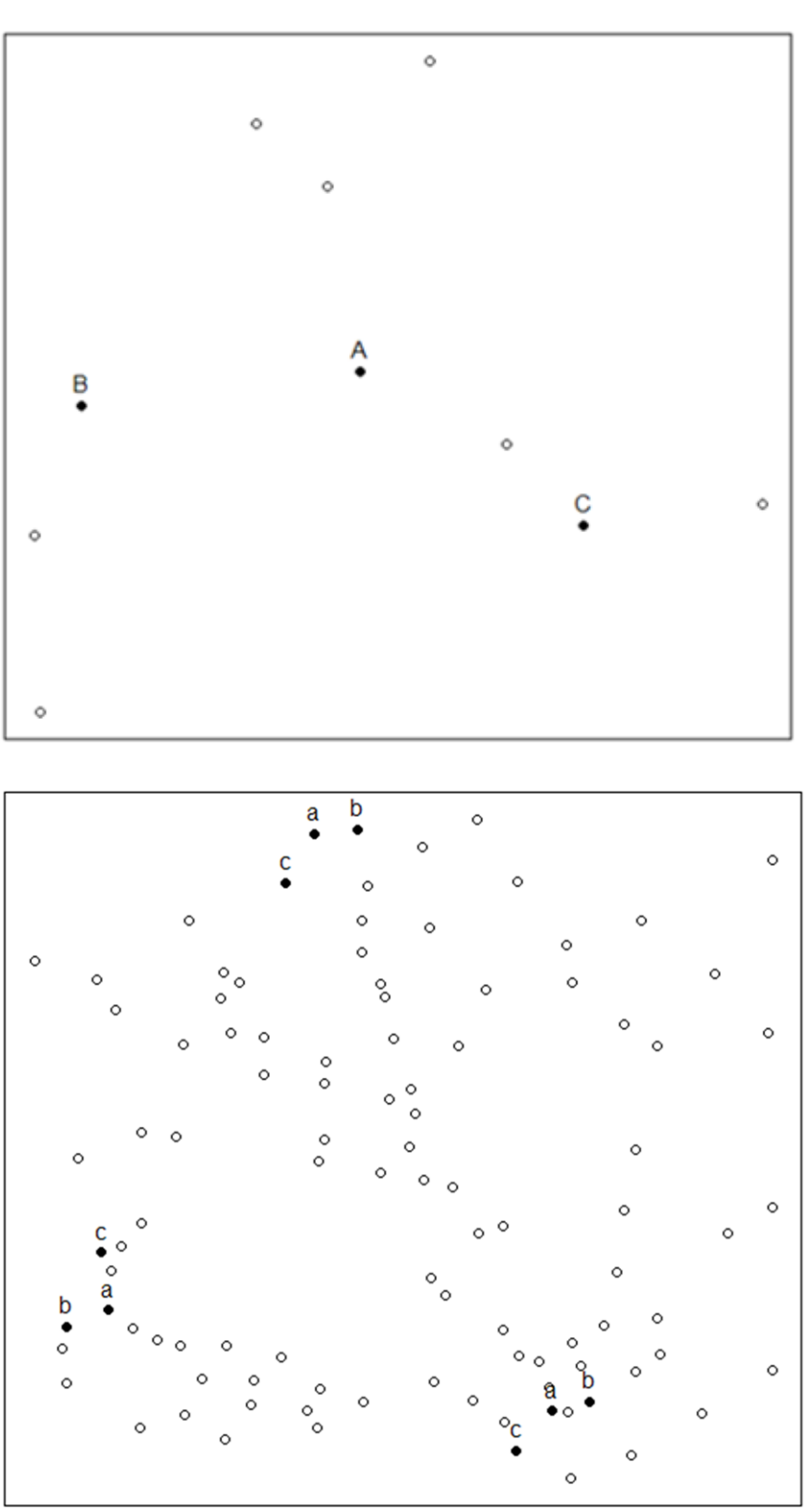

Figure 6.1: Plot showing the minutiae in the fingermark with the triangle ABC labelled (top) and three triangles with similar side lengths in the fingerprint labelled abc (bottom)

6.2.2 Transformation

Next a transformation of the fingermark is required. This is done by taking each *abc* from the candidate set in turn and looking for the similarity transformation of *ABC* that most closely matches it to *abc*. The transformation used is referred to as a Euclidean similarity transformation (Cox and Cox (2010), Dryden and Mardia (1998)) and consists of a translation, rotation and scale. This is described in more detail in Section 6.1.3. To carry out this analysis, elements of the "transformation" function in the "shapes" library of R were used. This function calculates the similarity transformation between two configurations based on the work of Dryden and Mardia (1998).

The transformation provided should match the two triangles of points, *ABC* and *abc*, as closely as possible by minimising the sum of squares (Euclidean) distance between the two sets of points. Cox and Cox (2010) show that this optimal matching can be carried out in the following steps. First the two triangles (*ABC* from the mark and *abc* from the print) are centred at the origin. This is done by calculating the mean of the x and y coordinates in each case and subtracting these means from each minutia. For *ABC* this is defined as:

$$ABC_i^{centered} = ABC_i - \overline{ABC} \tag{6.3}$$

for $i \in ABC$. The value $\overline{ABC}$ is referred to as the translation vector for *ABC*. In terms of scale ideally this could be set to one as we don't want to change the scale of the fingermark at all. However the way we have defined the candidate set forces the scale to be very close to one in practice anyway, and small changes in scale may be realistic for some fingermarks, so we leave the scale unconstrained. Fixing the scale, i.e. assuming congruence, would be a viable alternative. In fact the scale is calculated by finding a scale factor so that the root mean square distance of the points to the origin is 1. We do this using the Euclidean distance as calculated by:

$$|\mathbf{pq}| = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2} \tag{6.4}$$

where $\mathbf{p} = (p_x, p_y)$ and $\mathbf{q} = (q_x, q_y)$. The scale factor (SF) is found by calculating:

$$SF_{ABC} = \sqrt{\sum_i \frac{|ABC_i\overline{ABC}|^2}{3}} \tag{6.5}$$

The scaled and translated ABC can now be written as:

$$\widehat{ABC} = \frac{ABC^{centered}}{SF_{ABC}} \tag{6.6}$$

The final step is to rotate $\widehat{ABC}$ to minimise the Euclidean distance between it and $\widehat{abc}$. To find the rotation matrix, R, we define Q such that

$$abc^T ABC = \|ABC\|\|abc\|Q \tag{6.7}$$

where $\|ABC\| = \{\text{trace}(ABC^T ABC)\}^{1/2}$, then carry out the singular value decomposition of Q to get

$$Q = V\Lambda U^T \tag{6.8}$$

and set

$$R = UV^T. \tag{6.9}$$

Finally to get the new translated, rotated and scaled ABC ($\widetilde{ABC}$) back onto the original scale as the fingerprint we simply carry out:

$$A'B'C' = (\widetilde{ABC} * SF_{abc}) + \overline{abc} \tag{6.10}$$

We apply the transformation, based on mapping ABC to abc, to the whole of the fingermark $A, B, C \ldots, N$ by calculating:

$$X' = \left\{ \left(\frac{X - \overline{ABC}}{SF_{ABC}} \right) \times R \times SF_{abc} \right\} + \overline{abc} \text{ for} X \in \{A, B, \ldots N\}. \tag{6.11}$$

This gives a transformed mark based on the mapping between ABC and abc which is referred to as $A', B', C' \ldots, N'$. An example of this process can be seen in Figure 6.2.

After this has been completed for one value of *abc* in the candidate set it is then repeated for all of the others.

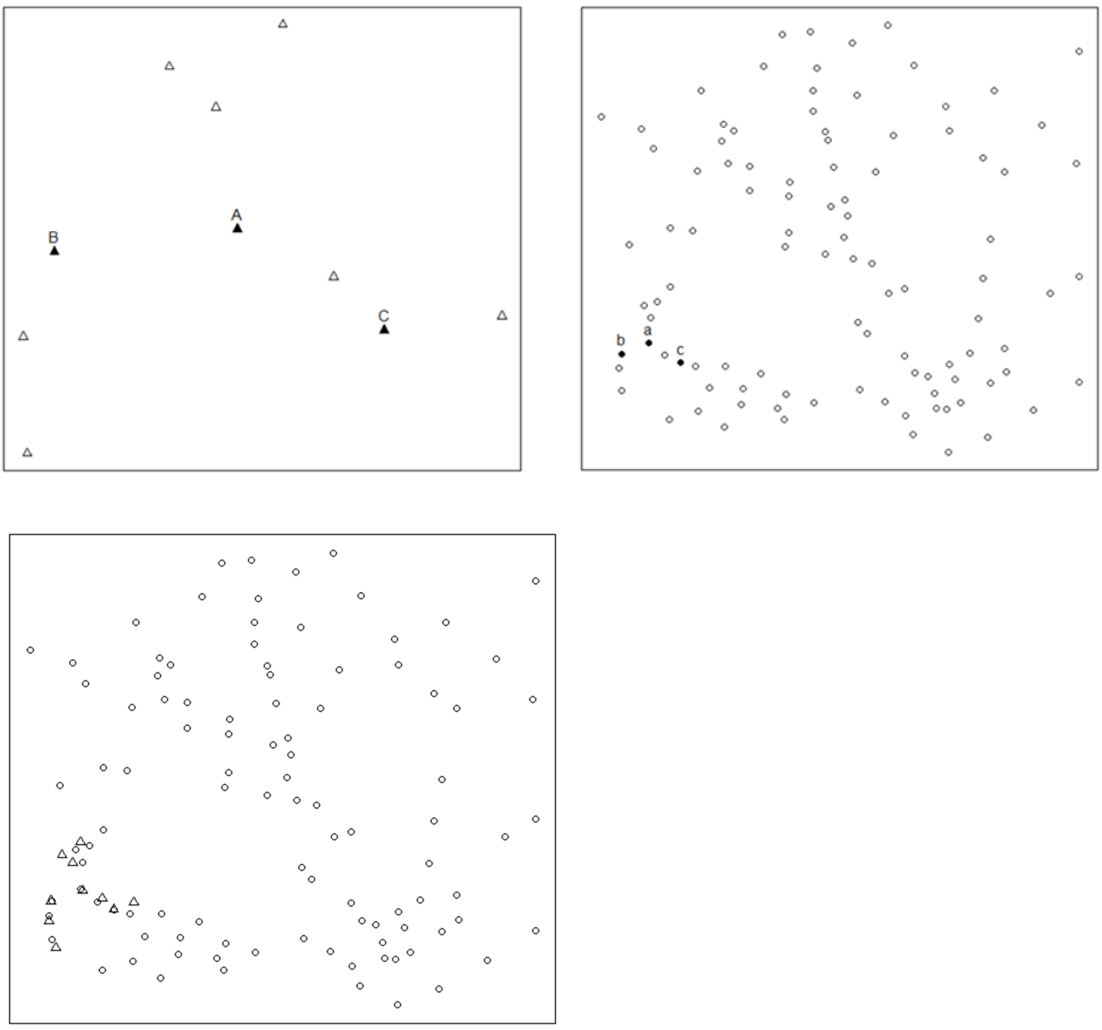

Figure 6.2: Plot showing the minutiae in a fingermark with the triangle ABC labelled (top left), the minutiae in the fingerprint with abc labelled (top right), the transformed fingermark overlayed on the the fingerprint based on the transformation from ABC to abc (bottom)

6.2.3 The Unique Allocation Problem

Based on each candidate triangle abc, we now have a transformation that maps all N minutiae in the fingermark, $A, B, C, \ldots, N$ say, on to points $A', B', C', \ldots, N'$ in the same space as the fingerprint minutiae, with $A'B'C'$ close to abc. If the fingermark really does come from this area of this fingerprint we would expect there to also be minutiae $d, \ldots, n$ close to $D', \ldots, N'$. To assess this more formally, we need to look for a set of distinct fingerprint minutiae, $a^*, \ldots, n^*$ say, such that,

$$|A'a^*|, |B'b^*|, \ldots, |N'n^*| \tag{6.12}$$

are all small, given that

$$|A'a^*| = \sqrt{(A_x{}' - a_x{}^*)^2 + (A_y{}' - a_y{}^*)^2}. \tag{6.13}$$

It is important to note that we do not necessarily have $a^* = a, b^* = b$ and $c^* = c$ since that may not be the "best" (based on minimum Euclidean distance) way to match all $A', B', C', \ldots, N'$ uniquely. Also in order to minimise the overall distance between the minutiae in the fingermark and the minutiae in the fingerprint and to ensure the allocation is unique, d^*, say, may not actually be the closest minutia to d. Finding $a^*, b^*, \ldots, n^*$ is an example of what is called an assignment problem in mathematics. In this application, that refers to assigning each minutia in the transformed fingermark to a unique minutia in the fingerprint, doing this in such a way as to minimise the overall distance between the transformed mark and selected minutiae in the print. The next section describes an algorithm which exactly solves this problem.

6.2.4 Hungarian Algorithm

The Hungarian Algorithm (Kuhn, 1955) is an optimisation algorithm which is designed to solve a linear sum assignment problem. Given a simple example where you want to assign n men to n jobs, this would correspond to a total of $n!$ possible assignments of men to jobs (Munkres, 1957). For small n the options would be easy to list but for larger n this becomes difficult to see all possible assignments. Now say, each of the n men had a preference for each of the n jobs we could produce an $n \times n$ matrix

where each entry corresponds to the preference for each man to job assignment. The Hungarian Algorithm is an algorithm which finds the optimal assignment based on those preferences from the $n!$ possible options. Using the notation in Burkard et al. (2009), given an $n \times n$ matrix C of costs, we want to select n elements of C such that we have one element in each row and column only, in such a way that the sum of the corresponding costs is minimised (Burkard et al., 2009).

Here we give an example using fingerprints and fingermarks, the Hungarian Algorithm is implemented using the actual steps outlined in Flood (1956). In this case we are comparing N minutiae in the fingermark with P minutiae in the fingerprint with $N \leq P$. Table 6.1 shows a small example with $N = 3$ minutiae in the fingermark and $P = 5$ minutiae in the fingerprint.

		Fingerprint				
		a	b	c	d	e
Transformed Fingermark	A'	1.4	0.7	1.2	2.0	0.9
	B'	0.2	0.9	1.9	1.4	1.1
	C'	1.5	0.8	1.2	0.6	1.7

Table 6.1: Small example of data with 3 minutiae in the fingermark and 5 minutiae in the fingerprint, the values in the table correspond to the distances between the minutiae in the mark and the minutiae in the print

Here we have $n = P = 5$ columns and $N = 3$ rows. In order to carry out the Hungarian Algorithm the matrix must be square so we add two dummy rows, the values placed in the dummy rows are equal to the highest value in the table. This is shown in Table 6.2.

	a	b	c	d	e
A'	1.4	0.7	1.2	2.0	0.9
B'	0.2	0.9	1.9	1.4	1.1
C'	1.5	0.8	1.2	0.6	1.7
D_1	2.0	2.0	2.0	2.0	2.0
D_2	2.0	2.0	2.0	2.0	2.0

Table 6.2: Data from Table 6.1 with dummy rows added

The cost assigned to each element is the squared Euclidean distance between each minutia in the transformed mark and each minutia in the fingerprint. The element in

the i-th row and j-th column corresponds to the squared Euclidean distance between i-th minutia in the fingermark and the j-th minutia in the fingerprint. An assignment of minutiae in the print and mark is then found based on minimising the overall sum of the squared Euclidean distances.

Now to carry out the Hungarian Algorithm the minimum value in each row is deducted from that row as shown in Table 6.3. The same is repeated for columns, i.e. the minimum value in each column is deducted from the whole column. With the example we are using there is already a zero in each column so the table will stay the same as shown in Table 6.3.

	a	b	c	d	e
A'	0.7	0	0.5	1.3	0.2
B'	0	0.7	1.7	1.2	0.9
C'	0.9	0.2	0.6	0	1.1
D_1	0	0	0	0	0
D_2	0	0	0	0	0

Table 6.3: Table 6.2 with the minimum from each row deducted

Next the zero elements are covered with as few lines as possible; this can be seen for the example in Table 6.4.

	a	b	c	d	e
A'	0.7	0	0.5	1.3	0.2
B'	0	0.7	1.7	1.2	0.9
C'	0.9	0.2	0.6	0	1.1
~~D_1~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~
~~D_2~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~

Table 6.4: Table 6.3 with the zero elements covered

If the minimum number of lines taken to cover the zeros is equal to the number of columns ($P = 5$) then we can stop and move to the last step. If not add the minimum uncovered element to all of the covered elements (adding twice if covered twice) and repeat the above process of subtracting the minimums for rows and columns and covering the zero elements. Once a stage has been reached where the minimum number of lines taken to cover the zero elements is equal to the number of columns the process is complete; all that is left to do is to choose a set of zeros so that there is only one

selection in each row and column. For our example the process was complete after one round of covering the zero elements and the selections can be seen in Table 6.5. The result after removing the dummy rows is $A' \to b$, $B' \to a$ and $C' \to d$ with the overall cost being $0.7 + 0.2 + 0.6 = 1.5$.

	a	b	c	d	e
A'	0.7	(0)	0.5	1.3	0.2
B'	(0)	0.7	1.7	1.2	0.9
C'	0.9	0.2	0.6	(0)	1.1
D_1	0	0	0	0	(0)
D_2	0	0	(0)	0	0

Table 6.5: Table showing the chosen set of zeros so that there is only one selection in each row and column.

6.3 Final Step

From the Hungarian Algorithm (see Section 6.2.4) we have a unique allocation for the N minutiae in the transformed fingermark to the minutiae in the fingerprint. We also have the overall sum of squared Euclidean distances for this specific allocation, which is:

$$S_{ABC}^{abc} = |A'a^*|^2 + |B'b^*|^2 + \ldots + |N'n^*|^2. \tag{6.14}$$

The whole process is repeated for all triangles in the candidate set which leaves a set of allocations with corresponding sum of squared values for each fit. We take the fit with the smallest overall sum of squares to be the optimal match for this fingerprint and use this in our analysis; this can be described as the minimum S_{ABC}^{abc}. So for a specific fingerprint X, the match to the fingermark used for the analysis is:

$$S(X) = \min(S_{ABC}^{abc}) \text{ for all } abc \text{ in } X \tag{6.15}$$

6.4 Summary

In conclusion the method described above demonstrates how to match a fingermark to a fingerprint in order to give a value for a how closely they match. In each case a candidate set of similar points is found, a transformation for each value in the candidate set is performed and then a Hungarian Algorithm is used to give a unique allocation before calculating the sum of squared distances between the mark minutiae and the print minutiae. The smallest of these values is selected to represent the optimal match between this specific mark and print. The whole process is then repeated many times with the same fingermark but different fingerprints to give us a set of possible matches which can later be analysed, ranked and communicated as required to the necessary parties. This chapter provides a measure of distance between a fingermark and any given fingerprint. To determine the strength of the match requires investigating the distance in relation to the rest of the general population and calculating a likelihood ratio which can be seen in Chapter 7 along with the results from some simulated experiments.

Chapter 7

Simulation Experiments

7.1 Introduction

This chapter focuses on using the matching algorithm from the previous chapter to carry out simulation experiments so that we can assess how well the algorithm performs. In this artificially constructed case where a dummy fingermark is matched to many fingerprints, we would hope the algorithm will find the best match to the fingerprint which it comes from (i.e. the original fingerprint used to construct the fingermark), this is referred to as the "correct" fingerprint. The algorithm will hopefully perform well if the distortion in the fingermark is small enough, and if the fingermark is similar enough in form to what is implicit in the algorithm, that is independent, isotropic, bivariate normal errors. We then rely on the best fitting triangle and the best optimised fit for the whole fingermark based on that triangle to be similar enough to the original fingerprint to be identified as coming from that print. In this situation theoretical results would be difficult to obtain due to the complexity of the geometry and multi-stage nature of the algorithm, hence we use simulation experiments as the appropriate way forward. These simulation experiments can only cover selected cases however we have tried to give a wide range of the key variables; number of minutiae and variance of errors, this should give insight into the usefulness of this approach.

Overall selecting the "correct" fingerprint as the best match to the fingermark relates to the "correct" fingerprint having the lowest sum of squares value amongst the whole set and hence producing the largest likelihood ratio (LR) when making a comparison

over the whole set. A large LR, when comparing a specific fingerprint to the whole set for each experiment, would show that the fingermark matches substantially better to this specific print than the others. Within the correct fingerprint the best match (i.e. the lowest sum of squares value for that fingerprint's candidate set) should yield the correct allocation from the Hungarian Algorithm. Since the fingermark has been constructed artificially from a fingerprint the index of each minutia is known and the index of each point in the fingermark should match to the allocation given by the Hungarian Algorithm.

By matching a specific fingermark to multiple fingerprints and then analysing the results from this it can be seen whether our method picks the correct fingerprint from the set as the best match. Also an observation can be made on whether, within the correct fingerprint, the allocation of minutiae is correct. This should be straightforward to check because, as the data are artificially generated, how well the algorithm is progressing can be tracked in real time.

7.2 Setup

To carry out simulation experiments using the matching algorithm, data is needed. For this study data was provided by the Forensic Science Service. These data contain information about each fingerprint including minutiae locations and types, fingerprint pattern and finger type amongst other information. As described previously, fingermarks are required which are not in the database. For this reason dummy fingermarks were created using the information from real fingerprints. For more information about the origin of the fingerprints and dummy fingermarks see Chapter 3.

Each individual run in an experiment is started by creating the dummy fingermark with standard deviation σ. Then 100 random fingerprints are selected plus the "correct" fingerprint giving 101 fingerprints in the set. Here "correct" fingerprint refers to the original print used to create the dummy fingermark. Then the matching algorithm is carried out between the fingermark and each fingerprint. In each case the best possible match in each fingerprint is selected and the information about this match is stored, for example the sum of squares. This whole process is repeated for 100 runs each with a different fingermark and a random sample of the fingerprints from the whole data set

of 12,096 fingerprints.

In order to make the matching as realistic as possible the initial fingerprint used to create the fingermark is randomly selected and then the 100 other fingerprints to be matched against are also randomly selected. However when selecting 100 fingerprints for the matching set these are sampled from the whole set of fingerprints based on the rough population densities of their level 1 patterns. That is, the sampling is stratified so as to match population proportions rather than proportions in the data set. For example since loops are most prevalent in the population these are sampled with probability of 0.6.

In addition to the above, during each run the finger type and pattern type is tracked in each case as the information may be required later.

7.3 Theory for Calculating the Likelihood Ratio

In this section we show how to use the sum of squared distances (refer to Chapter 6) between the fingermark and the fingerprint to calculate the likelihood ratio. In Section 2.3 we discussed how forensic data is handled in a mathematical way via the use of likelihood ratios. The basic formula we use when working with likelihood ratios is taken from Bayes Theorem, see e.g. Champod et al. (2005):

$$\frac{p(H_p|E,I)}{p(H_d|E,I)} = \frac{p(H_p|I)}{p(H_d|I)} \times \frac{p(E|H_p,I)}{p(E|H_d,I)} \tag{7.1}$$

where H_p is the hypothesis that the evidence originates from the suspect, H_d the hypothesis that the evidence originates from an unknown individual (referred to as the Prosecutor and Defence hypotheses respectively), E is the forensic information and I is background information. On the basis of this we have a likelihood ratio (Champod et al., 2007):

$$LR = \frac{p(E|H_p,I)}{p(E|H_d,I)}. \tag{7.2}$$

For this work we ignore various other effects, such as the uncertainty associated with the similarity transform in Section 6.2.2. We have found a sum of squared distances

for the minutiae in the transformed mark $A', B', C', \ldots, N'$ to their best allocation in the specific fingerprint we are testing against $a^*, \ldots, n^*$, the suspect's fingerprint. We assume that the N minutiae are modelled by $m = 2n$ coordinates having Normal distributions with known means $\mu_0 = (\mu_{0,1}, \mu_{0,2}, \ldots, \mu_{0,m})$ and unknown variance $v = \sigma^2$. We treat μ_0 as known, even though in practice it has been "estimated" using the matching procedure; this ignores the uncertainty associated with the similarity transform and other steps in the matching, and is equivalent to ignoring the contribution of other allocations apart from the best one. On the other hand, we do allow for the uncertainty in the variance v of the error (due to smudging of the fingermark etc.) in the locations of the minutiae, since in practice this will be unknown and estimated from a relatively small number of observations. To allow for the uncertainty in v, rather than making a point estimate, we take a Bayesian approach and put a prior distribution on v. Our likelihood ratio is then based on integrating out v, given the data, rather than plugging in a point estimate for v such as its maximum likelihood estimator. As such, it can be seen as a compromise between the usual likelihood ratio and the Bayes factor which would be obtained if we allowed for all of the uncertainty mentioned above.

7.3.1 Prosecutor Hypothesis

In order to explicitly calculate these values from the sum of squared distances found in Chapter 6 we first consider $p(H_P|E)$. Since the locations of minutiae in fingerprints are not affected by gender, age or other exterior factors, for simplicity we remove I from calculations. We use the known means $\mu_0 = (\mu_{0,1}, \mu_{0,2}, \ldots, \mu_{0,m})$ and unknown variance $v = \sigma^2$ as defined in the previous section and write $y = (y_1, \ldots, y_m)$ for the evidence (E is the notation of Champod et al. (2005)). Then we have:

$$\begin{aligned} p(H_p|y) &= \int_v p(\mu_0, v|y)dv \\ &= \int_v \frac{p(y|\mu_0, v)p(\mu_0)p(v)}{p(y)}dv \\ &= \frac{p(\mu_0)}{p(y)} \int_v p(y|\mu_0, v)p(v)dv. \end{aligned} \tag{7.3}$$

As stated previously, we have a distribution of observations of independent Normal random variables (shown by y) given the true values. An Inverse Gamma prior distri-

bution with hyperparameters α and β, is used for the variance, for tractability. Then following Gelman et al. (2003) we have:

$$\begin{aligned} p(H_p|y) &= \frac{p(\mu_0)}{p(y)} \int_v \left[\prod_{i=1}^{m} \frac{1}{\sqrt{2\pi v}} \exp\left\{ -\frac{1}{2v}(y_i - \mu_{0,i})^2 \right\} \right] \times v^{-(\alpha+1)} \exp(-\beta/v) \frac{\beta^\alpha}{\Gamma(\alpha)} dv \\ &= \frac{p(\mu_0)}{p(y)} \int_v (2\pi)^{-\frac{m}{2}} v^{-(\alpha+\frac{m}{2}+1)} \exp\left\{ -\frac{1}{v} \left[\frac{\sum_{i=1}^{m}(y_i - \mu_{0,i})^2}{2} + \beta \right] \right\} \frac{\beta^\alpha}{\Gamma(\alpha)} dv. \end{aligned} \tag{7.4}$$

By using the probability density function (pdf) for the Inverse Gamma distribution we can represent 7.4 as:

$$p(H_p|y) = \frac{p(\mu_0)}{p(y)} (2\pi)^{-\frac{m}{2}} \frac{\beta^\alpha}{\Gamma(\alpha)} \int_v v^{-(A+1)} \exp(-B/v) dv \tag{7.5}$$

where

$$A = \alpha + \frac{m}{2} \tag{7.6}$$

$$\begin{aligned} B &= \beta + \frac{\sum_{i=1}^{m}(y_i - \mu_{0,i})^2}{2} \\ &= \beta + \frac{\text{SSD}}{2} \end{aligned} \tag{7.7}$$

and SSD denotes Sum of Squared Distances. Since we know that the pdf of the Inverse Gamma distribution integrates to 1 we can simplify this to:

$$p(H_p|y) = \frac{p(\mu_0)}{p(y)} C \left(\beta + \frac{SSD_0}{2} \right)^{-A} \tag{7.8}$$

where

$$C = \frac{(2\pi)^{-\frac{m}{2}} \beta^\alpha}{\Gamma(\alpha)} \Gamma(A). \tag{7.9}$$

7.3.2 Defence Hypothesis

Using the same principles as with the Prosecutor hypothesis we can establish values for the Defence hypothesis. The only difference here is that we are measuring against

a population of fingerprints so we need to average the results over all of these values. In principle we have:

$$p(H_d|y) = \int_{\mu}\int_{v} p(\mu, v|y)dvd\mu. \tag{7.10}$$

However, we must approximate the outer integral using the available sample of k fingerprints, where k is the number of fingerprints in the sample, giving:

$$\begin{aligned} p(H_d|y) &= \sum_{j=1}^{k}\left[\int_{v} p(\mu_j, v|y)dv\right] \\ &= \sum_{j=1}^{k}\left[\int_{v} \frac{p(y|\mu_j, v)p(\mu_j)p(v)}{p(y)}dv\right] \\ &= \frac{1}{p(y)}\sum_{j=1}^{k}\left[p(\mu_j)\int_{v} p(y|\mu_j, v)p(v)dv\right]. \end{aligned} \tag{7.11}$$

Using A, B and C as defined in 7.6, 7.7 and 7.9 respectively we can represent 7.11 as:

$$\begin{aligned} p(H_d|y) &= \frac{1}{p(y)}\sum_{j=1}^{k}\left[p(\mu_j)C\left(\beta + \frac{SSD_j}{2}\right)^{-A}\right] \\ &= \frac{C}{p(y)}\sum_{j=1}^{k}\left[p(\mu_j)\left(\beta + \frac{SSD_j}{2}\right)^{-A}\right]. \end{aligned} \tag{7.12}$$

7.3.3 Likelihood Ratio

If we assume that there is no prior knowledge about the suspect then we can conclude that every member of the population is equally likely to be the best match for the fingermark. Hence:

$$p(\mu_1) = p(\mu_2) = \ldots = p(\mu_k) \tag{7.13}$$

with

$$p(H_p) = p(\mu_0) \tag{7.14}$$

$$p(H_d) = \sum_{j=1}^{k} p(\mu_j) = kp(\mu_1) \tag{7.15}$$

say since we are approximating the population with the sample of size k. We can use 7.13 to represent 7.12 as:

$$\begin{aligned} p(H_d|y) &= \frac{C}{p(y)} \sum_{j=1}^{k} \left[p(\mu_1) \left(\beta + \frac{SSD_j}{2} \right)^{-A} \right] \\ &= \frac{C}{p(y)} p(\mu_1) \sum_{j=1}^{k} \left(\beta + \frac{SSD_j}{2} \right)^{-A} \\ &= \frac{C}{p(y)} \frac{p(H_d)}{x} \sum_{j=1}^{k} \left(\beta + \frac{SSD_j}{2} \right)^{-A}. \end{aligned} \tag{7.16}$$

Now we can replace these values into our original equations to find a useable LR:

$$\frac{p(H_p|y)}{p(H_d|y)} = \frac{p(H_p)}{p(H_d)} \times \frac{\left(\beta + \frac{SSD_0}{2} \right)^{-A}}{\sum_{j=1}^{k} \frac{1}{k} \left(\beta + \frac{SSD_j}{2} \right)^{-A}}. \tag{7.17}$$

Hence

$$\begin{aligned} LR &= \frac{p(y|H_p)}{p(y|H_d)} \\ &= \frac{\left(\beta + \frac{SSD_0}{2} \right)^{-A}}{\sum_{j=1}^{k} \frac{1}{k} \left(\beta + \frac{SSD_j}{2} \right)^{-A}}. \end{aligned} \tag{7.18}$$

Now that we have a method for calculating the LR, we need a way to interpret the results. In Evett (1998) a scale is proposed for interpreting the likelihood ratios and the strength of support for the evidence that this represents. This is done in a way which can be expressed verbally. This was later modified slightly in Fierrez-Aguilar

et al. (2005) and can be seen below in Table 7.1, this is used for expressing the strength of DNA evidence. For example, if we obtained a LR of 50 from the matching algorithm we could say verbally that we have "moderate evidence to support the proposition that the fingermark originated from the specified fingerprint".

LR	Strength of Support
1 to 10	Limited Evidence to Support
10 to 100	Moderate Evidence to Support
100 to 1,000	Moderately Strong Evidence to Support
1,000 to 10,000	Strong Evidence to Support
Over 10,000	Very Strong Evidence to Support

Table 7.1: A table showing a scale for likelihood ratios and the strength of support for the evidence they represent

7.4 Results from Simulation Experiments

7.4.1 Results for Changes in Number of Minutiae

In a true forensic scenario the number of minutiae in a fingermark is not fixed. Sometimes the fingermark may be of particularly good quality and contain a lot of minutiae (upwards of 20). However more often than not they will have far fewer than this present. To make the results of the matching algorithm as realistic as possible the matching programme was carried out when different numbers of minutiae were present in the fingermark. Testing with numbers from 8 to 18 was performed as these numbers are reasonable to expect from a fingermark (based on experience). In each case matching was performed on 100 fingermarks comparing each of these to 100 fingerprints plus the correct print (making 101 in total). The value of σ during these experiments was set to 4, as this value represents a fingermark which looks realistic. The value of σ will be varied during Section 7.4.2 as a sensitivity study. In all experiments the values for the hyperparameters α and β were set to 0.1. In order to give a proper and uninformative prior they were set to values which were small in comparison to m and SSD. Initially some experiments were carried out with a range of values for α and β but as it was insensitive to the choices 0.1 was decided on.

After calculating the likelihood ratio of the correct fingerprint for each fingermark, logs are taken for a better visual assessment as it was difficult to differentiate between the values in the plot. As logs have been taken this will need to be taken into account when the results are interpreted in relation to the values in Table 7.1. In Figure 7.1 it can be seen that as the number of minutiae in the fingermark increase so does the log likelihood ratio. From Table 7.1 it can be seen that in general as the LR (and hence log likelihood ratio) increases the strength of the evidence to support the proposition increases. As these experiments propose that the fingermark comes from the correct fingerprint (defined as the original fingerprint used to generate the fingermark), an increase in the log likelihood ratio would suggest that the evidence supports this as the number of minutiae in the fingermark increases. This is what would be expected intuitively, as an increase in minutiae in the fingermark gives more information for the matching algorithm to use. It is also apparent from Figure 7.1 that in general the number of log likelihood ratios greater than zero increases as the number of minutiae increase. A log likelihood ratio of zero corresponds to a LR of 1 and hence can be described as the minimum value needed to provide limited evidence to support the proposition. If we look at the strongest level of support given in Table 7.1, that being a LR over 10,000, then Table 7.2 shows the number from each of the 100 runs for different minutiae levels that were over this threshold.

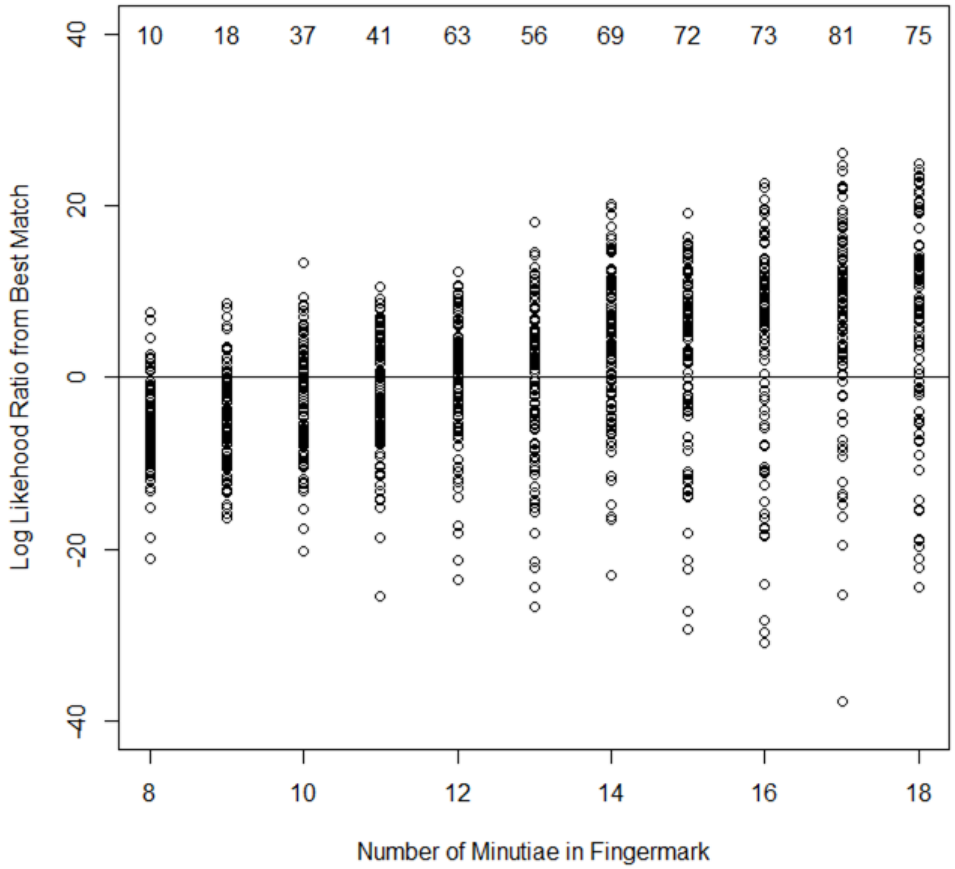

Figure 7.1: Plot showing the log likelihood ratios for the best matches for different numbers of minutiae in the fingermarks. Each is labelled with a number which demonstrates how many from 100 runs have a log likelihood ratio greater than zero

No. of Minutiae in Fingermark	8	9	10	11	12	13	14	15	16	17	18
No. with LR>10,000 for Correct Fingerprint as Best Match	0	0	2	1	5	14	27	28	39	47	46

Table 7.2: A table showing the number of times the matching algorithm identified the correct fingerprint as the best match for differing numbers of minutiae in the fingermark (out of 100) with a likelihood ratio greater than 10,000

In order to assess if the algorithm is matching well by chance we remove the "correct" finger (the fingerprint that is used to simulate the fingermark) and rerun the calculation

for the log likelihood ratio using a randomly selected fingerprint as the one to match against. The results of this can be seen in Figure 7.2 compared with the results from using the "correct" fingerprint. This also shows the number of times the log likehood ratio is greater than zero, where this represents limited evidence to support the proposition that the fingermark comes from the fingerprint tested against. We can conclude from this that the matching algorithm shows no evidence to support a fingerprint which the fingermark doesn't originate from but in many cases gives us at least limited evidence if the fingerprint is the "correct" fingerprint for the associated fingermark.

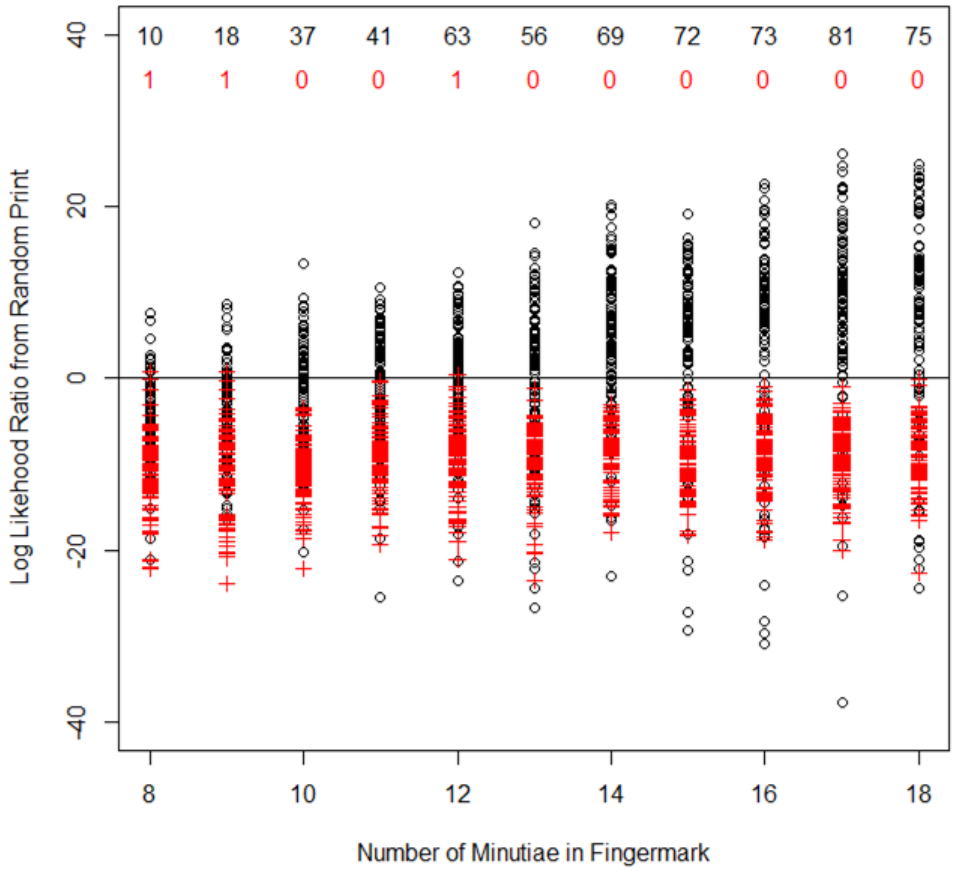

Figure 7.2: Plot showing the log likelihood ratios for the best match in a randomly selected fingerprint for different numbers of minutiae in the fingermarks. Each is labelled with a number which demonstrates how many from 100 runs have a log likelihood ratio greater than zero (black circles = "correct" fingerprint, red crosses = random fingerprint)

Next, we focus on the allocations made by the matching algorithm. The fingerprint that we know to be the correct match to the fingermark can be selected and we can observe how well the algorithm is allocating each minutia in the fingermark to the corresponding minutia in the fingerprint. Ideally all N minutiae would match, however this is not always the case; see Figures 7.3 and 7.4. These plots illustrate that the majority of the time, well over 50%, the algorithm is matching all N minutiae to the right locations with others matching at $N-1$ or $N-2$ minutiae. Between 4% - 29% of the time the matching algorithm gets all N minutiae wrong however the higher percentage values correspond to lower minutiae numbers in the fingermark overall, which can be expected. It is expected that there are relatively high numbers which allocate zero minutiae correctly compared to the middle numbers of three to six (say), as if one or two minutiae in the fingermark are allocated wrongly this may displace the others in the mark as we only allow unique allocation.

Although it is relevant to know how well the algorithm performs when allocating in the correct fingerprint, it is also useful to know how many times the correct fingerprint was identified as the best match to the fingermark. The matching algorithm identifies the fingerprint with the lowest sum of squares as the "best match" as this will give the highest log likelihood ratio when compared to the other fingerprints in the set. The results can be seen in Table 7.3 which shows that as the number of minutiae increase the number of correct fingerprints identified as the best match also generally increases.

No. of Minutiae in Fingermark	8	9	10	11	12	13	14	15	16	17	18
No. with Correct Fingerprint Identified as Best Match	11	20	42	43	64	58	71	72	75	81	75

Table 7.3: A table showing the number of times the matching algorithm identified the correct fingerprint as the best match for differing numbers of minutiae in the fingermark (out of 100)

7.4.2 Results for Changes in σ

Simulation experiments were also carried out where σ is changed. This value for σ refers to the original construction of the fingermark and relates to how similar the fingermark is to the original fingerprint it came from; see Section 3.3. For the experiments σ

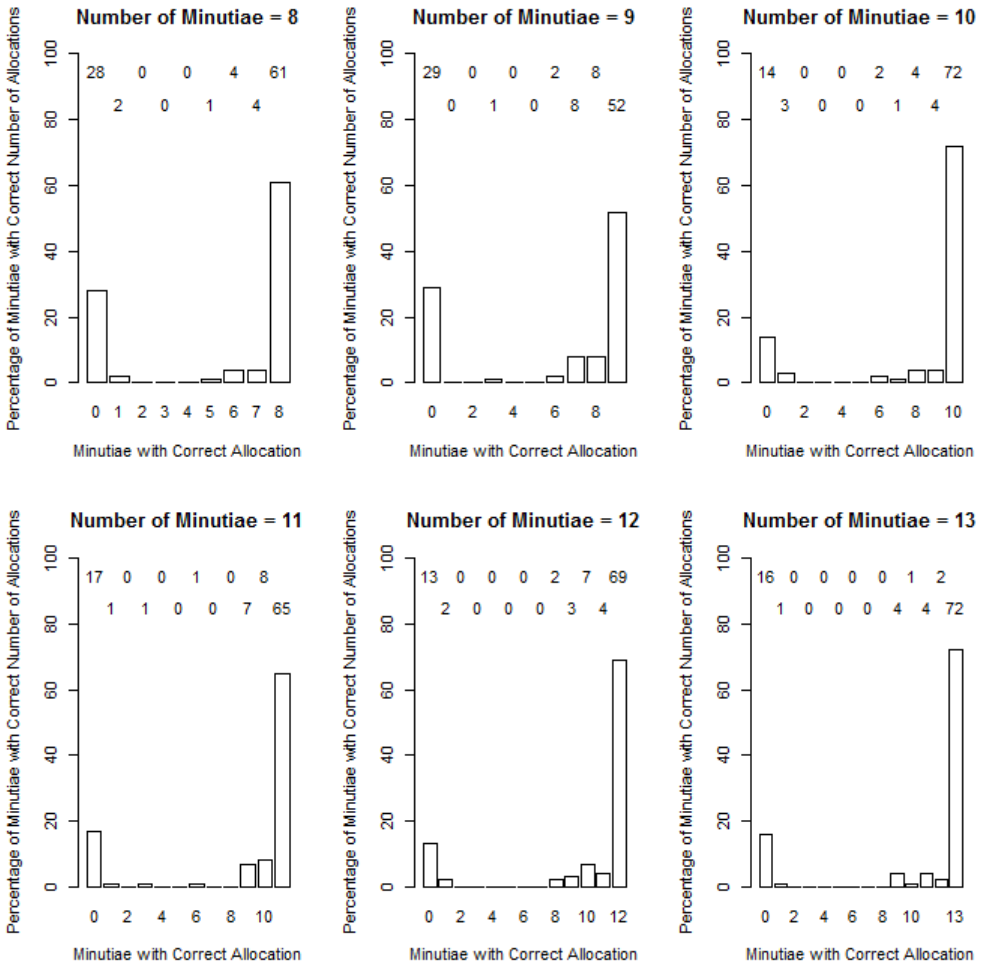

Figure 7.3: Barplot showing how many correct allocations were made in each experiment between the fingermark and the correct fingerprint (minutiae numbers 8-13)

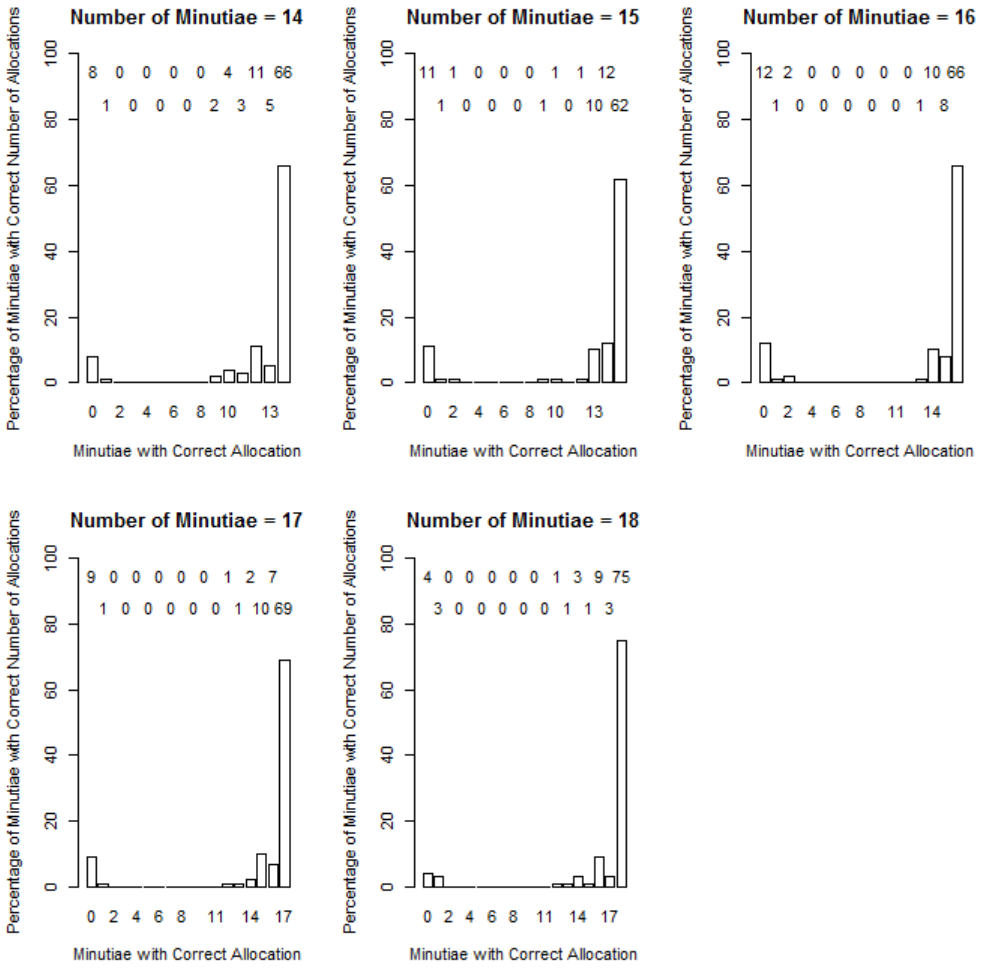

Figure 7.4: Barplot showing how many correct allocations were made in each experiment between the fingermark and the correct fingerprint (minutiae numbers 14-18)

has been varied from 1 to 6. As stated in Section 7.4.1 each test matched a total of 101 fingerprints to each mark, including the correct fingerprint, and was repeated with different fingermarks 100 times. These experiments are carried out with 10 minutiae in the fingermark.

Again, once the log likelihood ratio has been found, it was plotted for the changes in σ to see how well the matching algorithm was performing. Results can be seen in Figure 7.5. From this plot it can be seen that the matching algorithm works better as σ decreases, this is to be expected as a lower σ represents a fingermark which is more similar to the original fingerprint it is simulated from hence we would expect the matching algorithm to perform better in these circumstances. For lower values of σ, the majority of the matches showed distinction from the rest of the population (i.e. the set), this is shown by having a log likelihood ratio greater than zero. Again as in Section 7.4.1, it is interesting to look at the number of cases where the LR shows "very strong evidence to support" the proposition that the fingermark comes from the "correct" fingerprint (taken from Table 7.1). The results of this can be seen in Table 7.4.

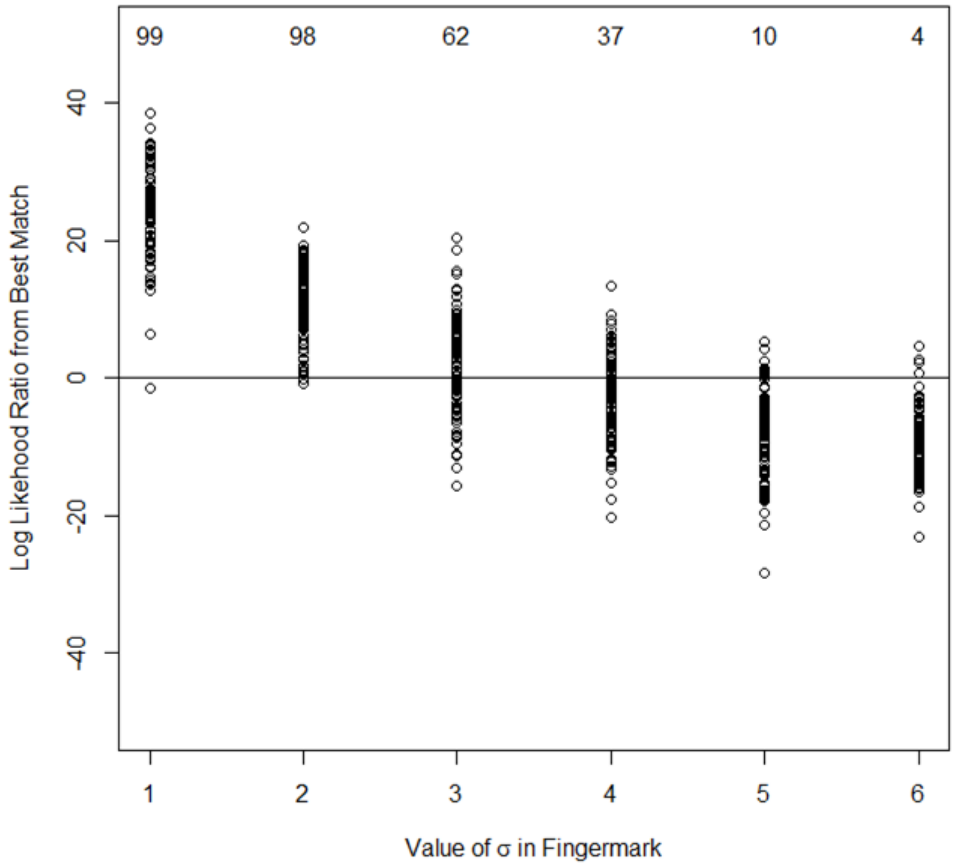

Figure 7.5: Plot showing the log likelihood ratios for the best matches for various values of σ in the construction of the fingermarks. Each is labelled with a number which demonstrates how many from 100 runs have a log likelihood ratio greater than zero. The number of minutiae is 10 throughout

Value of σ	1	2	3	4	5	6
No. with LR>10,000 for Correct Fingerprint as Best Match	98	65	13	2	0	0

Table 7.4: A table showing the number of times the matching algorithm identified the correct fingerprint as the best match for differing values of σ (out of 100) with a likelihood ratio greater than 10,000. The number of minutiae is 10 throughout.

the process of comparing a randomly selected fingerprint (that we know not to be the

"correct" fingerprint) is carried out. The results can be seen in Figure 7.6 which shows that the matching algorithm does not produce a good match if a random fingerprint is used. There are only a limited number of cases where a log likelihood ratio greater than zero is found.

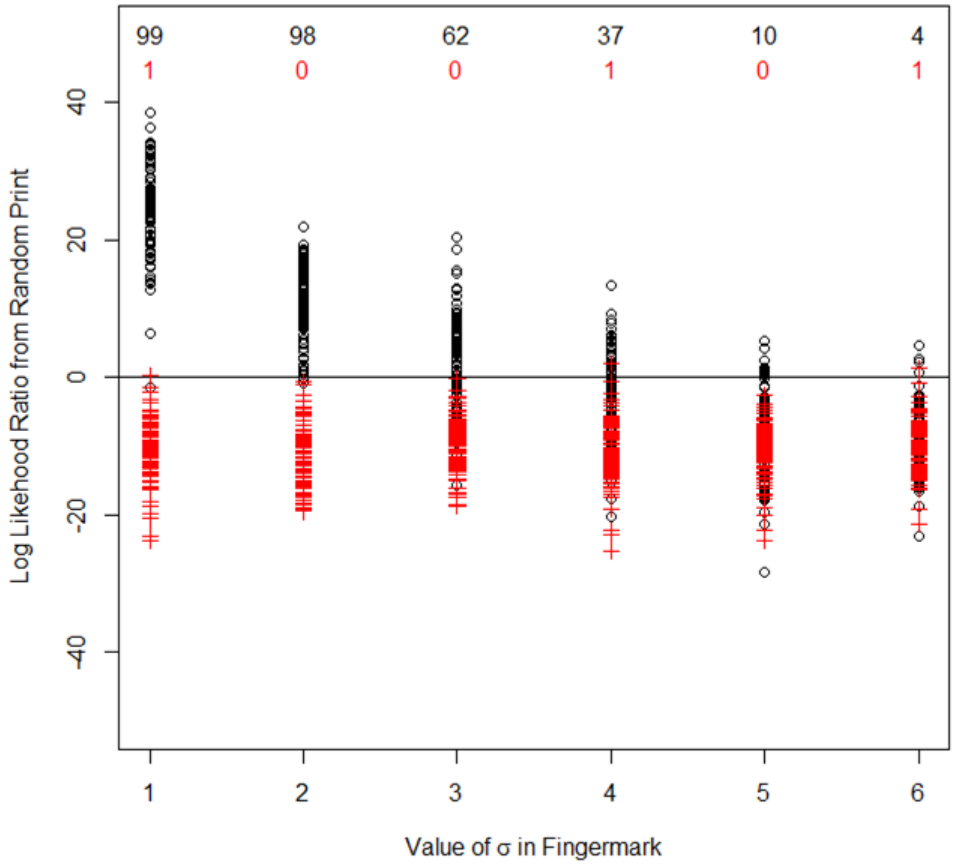

Figure 7.6: Plot showing the log likelihood ratios for the best match in a randomly selected fingerprint for various values of σ in the construction of the fingermarks. Each is labelled with a number which demonstrates how many from 100 runs have a log likelihood ratio greater than zero. The number of minutiae is 10 throughout. (black circles = "correct" fingerprint, red crosses = random fingerprint)

When looking specifically at the correct fingerprint it can be seen from Figure 7.7 that the matching algorithm allocates the minutiae in the fingerprint well for small values of σ. However with values of $\sigma = 5$ and especially $\sigma = 6$ the matching algorithm does

not do a very good job. This is also evident in Table 7.5 which shows that for the lower three values of σ the matching algorithm is identifying the correct fingerprint as the best match most of the time; however, for the higher three values of σ it is very rarely doing this.

Value of σ	1	2	3	4	5	6
No. with Correct Fingerprint Identified as Best Match	99	99	65	42	12	4

Table 7.5: A table showing the number of times the matching algorithm identified the correct fingerprint as the best match for differing values of σ (out of 100). The number of minutiae is 10 throughout.

7.5 Discussion

In both sets of results, changes in number of minutiae and changes in σ, high numbers of minutiae allocated correctly in the right fingerprint also corresponded with the algorithm identifying the correct print as the best match. These cases also correspond with higher values for the log likelihood ratio. This shows that when the algorithm allocates the minutiae in the correct way the method would give a positive result as the correct fingerprint would be identified as the best match from the population. This is further supported by the lack of cases during the random fingerprint testing which showed a large value (greater than zero) for the log likelihood ratio. The results here show that the method works most efficiently (i.e. getting the correct results most of the time) when we have large numbers of minutiae in the fingermark and low values for σ, which corresponds to small amounts of distortion in the fingermark.

7.6 Summary

This chapter has examined how well the matching program performs when using it with real data to assess a match between a fingermark and a fingerprint. A technique has been developed for taking the results from the algorithm into a usable format and then analysed different scenarios. These two experiments looked at different numbers of minutiae in the fingermark and varying the value of σ when constructing the fingermark.

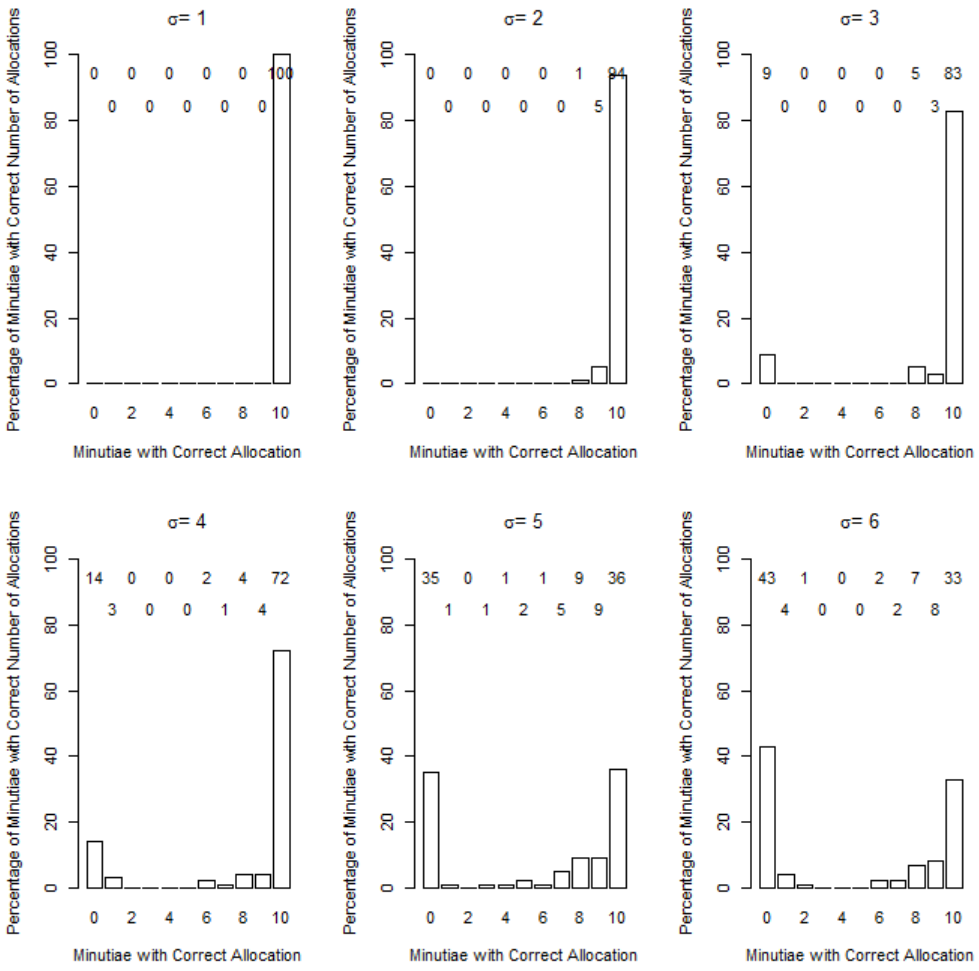

Figure 7.7: Barplot showing how many correct allocations were made in each experiment between the fingermark and the correct fingerprint with varying σ

Although the results from the simulation experiments are promising they do not give the accuracy and reliability being sought. Ideally the algorithm would identify the correct fingerprint as the best match more often (especially with higher values of σ) and also to deliver a higher value for the likelihood ratio. If the value of the likelihood ratio is clearly above zero in most cases then a good result has been attained.

Chapter 8

Optimisation

8.1 Introduction

The results from the simulation experiments were not as successful as would have been desired and hence this chapter will investigate optimising the matching algorithm to produce better results when matching a fingermark to a fingerprint. Then the simulation experiments will be repeated and compared to see how optimisation affects the results.

8.2 Method of Optimisation

The method of optimisation is based on performing a second transformation. The R code which was written to carry out this step of optimisation can be found in Appendix B, this is run additionally to the previous code in Appendix A. In summary the matching algorithm, from Chapter 6, assigns an initial triangle in the fingermark and finds a candidate set of points in the fingerprint which are similar to this triangle. For each case in the candidate set, a transformation is applied based on matching the triangle in the mark to the triangle in the print and then the Hungarian Algorithm is used to find a unique allocation of all N minutiae from the transformed mark. Once this unique allocation has been obtained, the sum of squared Euclidean distances between the minutiae in the fingermark and the minutiae in the fingerprint is found.

To carry out optimisation, another transformation after the Hungarian Algorithm is performed based on all N minutiae instead of 3 minutiae (ABC) which was done previously. The transformed mark $A', B', C', \ldots, N'$ has had every element allocated to a minutia in the fingerprint to give $a^*, \ldots, n^*$ using the Hungarian Algorithm. We now carry out another transformation by centering both sets and calculating the scale parameter and rotation matrix using ordinary Procrustes analysis (Dryden and Mardia, 1998) as done previously in Section 6.2.2. The results of this are then applied to $A', B', C', \ldots, N'$ to give $A'', B'', C'', \ldots, N''$. The sum of the squared Euclidean distances is then calculated using:

$$S_{ABC}^{abc} = |A''a^*|^2 + |B''b^*|^2 + \ldots + |N''n^*|^2. \tag{8.1}$$

This process is repeated for every triangle in the candidate set for that fingerprint. We select the smallest overall sum of squares (the minimum S_{ABC}^{abc}) to be the best match for that fingerprint. Hence for fingerprint X in the analysis we use the best match, which is described as:

$$S(X) = \min(S_{ABC}^{abc}) \text{ for all } abc \text{ in } X. \tag{8.2}$$

The first translation to $A', B', C', \ldots, N'$, can be thought of as getting the mark into the right region, whilst the second to $A'', B'', C'', \ldots, N''$, refines the location to provide a better match. By doing this second transformation it is anticipated that this will reduce the overall sum of the squared Euclidean distances and hence a stronger match can be identified.

8.3 Results from Optimisation

8.3.1 Results for Changes in Number of Minutiae

These results are based on the same experiments as those in Section 7.4.1 where the value for N varies between 8 and 18, and σ is fixed at a value of 4. Before optimisation it was observed that the log likelihood ratio tended to increase as the number of minutiae in the fingermark increased. This is also true post optimisation. However Figure 8.1

illustrates that optimisation in all cases increases the number of log likelihood ratios greater than zero which relates to the best match fingerprint being clearly better than the other fingerprints in the set at matching the fingermark, and hence provides evidence to support the conclusion that the fingermark came form the "correct" fingerprint. It can be seen that for lower numbers of minutiae in the mark, optimisation has greatly increased this distinction, for example for $N = 11$ this number has increased from 41 to 71. The number of cases where the LR is greater than 10,000, corresponding to "very strong evidence to support" the proposition, can be seen in Table 8.1. Except when the number of minutiae in the fingermark is 8 the optimisation step has increased the number of cases where there is very strong evidence that the fingermark came form the "correct" fingerprint.

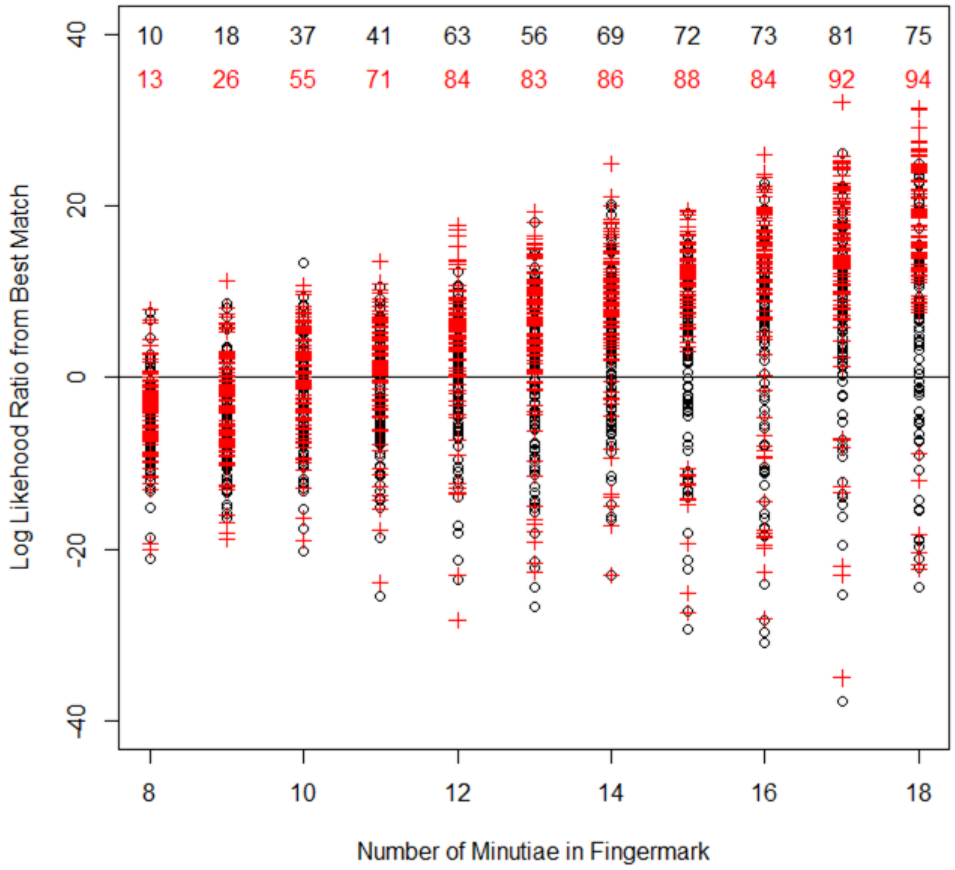

Figure 8.1: Plot showing the log likelihood ratios for the best matches for different numbers of minutiae in the fingermarks, each is labelled with a number which demonstrates how many from 100 runs have a log likelihood ratio greater than zero (black circles = pre-optimisation, red crosses = post-optimisation)

No. of Minutiae in Fingermark	8	9	10	11	12	13	14	15	16	17	18
No. with LR>10,000 for Correct Fingerprint as Best Match	0	1	3	4	15	37	43	66	73	79	87

Table 8.1: A table showing the number of times the matching algorithm identified the correct fingerprint as the best match for differing numbers of minutiae in the fingermark (out of 100) with a likelihood ratio greater than 10,000

As in Section 7.4.1 we repeated the calculations after removing the "correct" fingerprint

and using a randomly selected fingerprint to match again. In this case we know that the fingermark does not originate from the chosen fingerprint. The results of this can be seen in Figure 8.2, it shows again quite clearly that the matching algorithm produces different results if the chosen fingerprint is not the one which the fingermark originates from as there are only a small number of cases with even the lowest level (limited) of evidence to support the match.

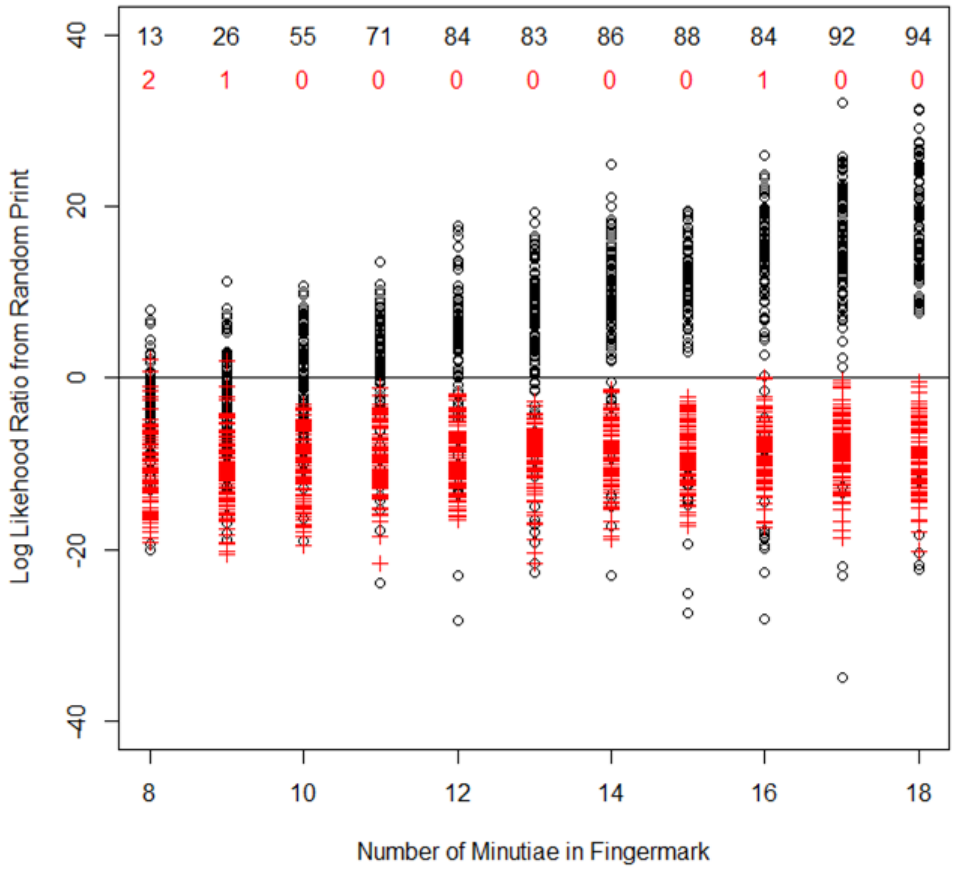

Figure 8.2: Plot showing the log likelihood ratios for the best match in a randomly selected fingerprint for different numbers of minutiae in the fingermarks. Each is labelled with a number which demonstrates how many from 100 runs have a log likelihood ratio greater than zero (black circles = "correct" fingerprint post-optimisation, red crosses = random fingerprint)

Again, as can be seen in Figures 8.3 and 8.4, optimisation has a positive effect on the results as the number of minutiae in the fingermark matching to correct locations in the fingerprint increases, as well as the number where all N match increasing. Table 8.2 shows that the number of allocations remaining unchanged post-optimisation is similar across changes in minutiae numbers, however in all cases between 5% and 22% have some change in minutiae allocation within the correct fingerprint.

No. of Minutiae in Fingermark	8	9	10	11	12	13	14	15	16	17	18
No. of Allocation Changes	20	22	9	18	16	15	13	14	13	11	5

Table 8.2: A table showing the number of times the allocation of minutiae in the fingermark to the correct fingerprint changed post-optimisation

When investigating how often the matching algorithm identifies the correct fingerprint as the best match for the fingermark, we see that optimisation only makes this better. In all cases optimisation increases the proportion of times the correct fingerprint is identified as the best match. It can be seen that for $N = 10$ over half the correct fingerprints are now being identified, unlike pre-optimisation (Table 8.3).

No. of Minutiae in Fingermark	8	9	10	11	12	13	14	15	16	17	18
No. with Correct Fingerprint Identified as Best Match (Pre)	11	20	42	43	64	58	71	72	75	81	75
No. with Correct Fingerprint Identified as Best Match (Post)	17	29	57	71	84	83	86	88	84	92	94

Table 8.3: A table showing the number of times the matching algorithm identified the correct fingerprint as the best match for differing numbers of minutiae in the fingermark (out of 100)

8.3.2 Results for Changes in σ

Optimisation also has an effect when the value of σ is varied and $N = 10$ as in Section 7.4.2. Figure 8.5 shows that as with pre-optimisation the log likelihood ratio decreases as the value of σ increases. In addition to this, optimisation increases the number of values with a log likelihood greater than zero, this value is still however quite low for $\sigma = 5$ and $\sigma = 6$. Similarly for the higher threshold of a LR greater than 10,000

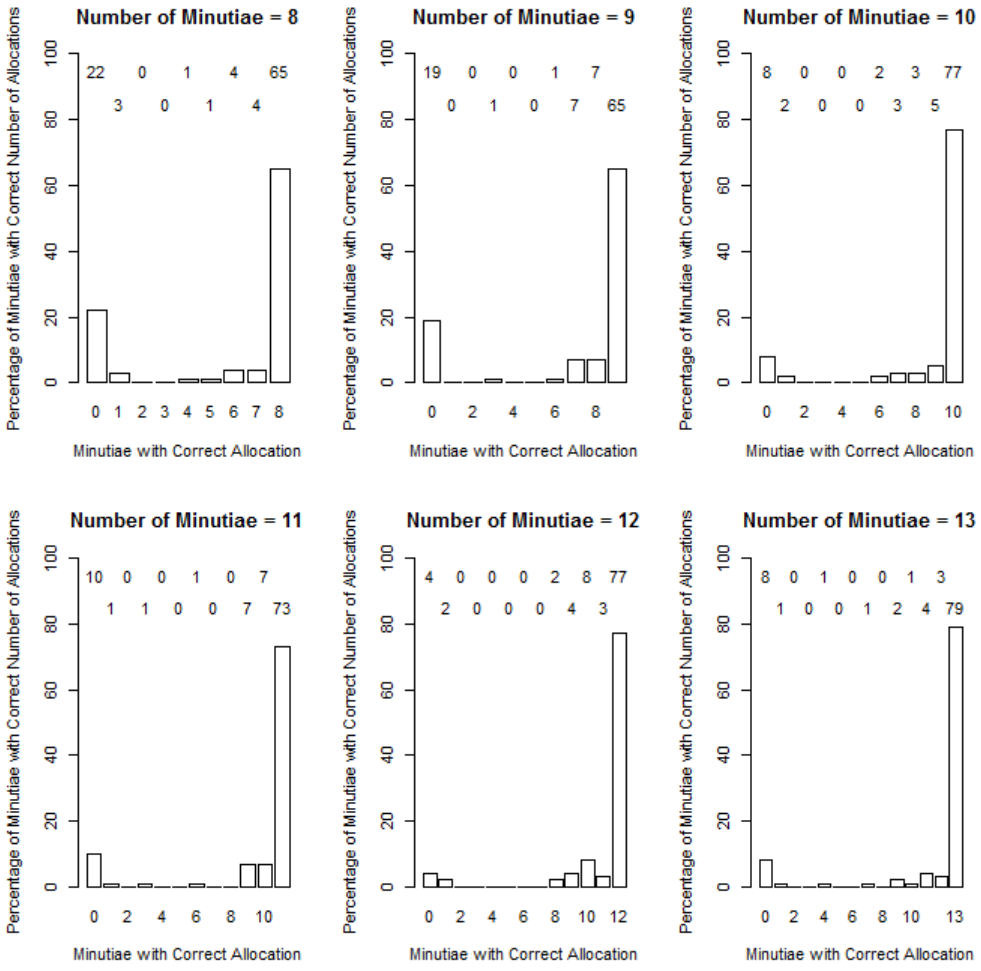

Figure 8.3: Barplot showing how many correct allocations were made in each experiment between the fingermark and the correct fingerprint (minutiae numbers 8-13)

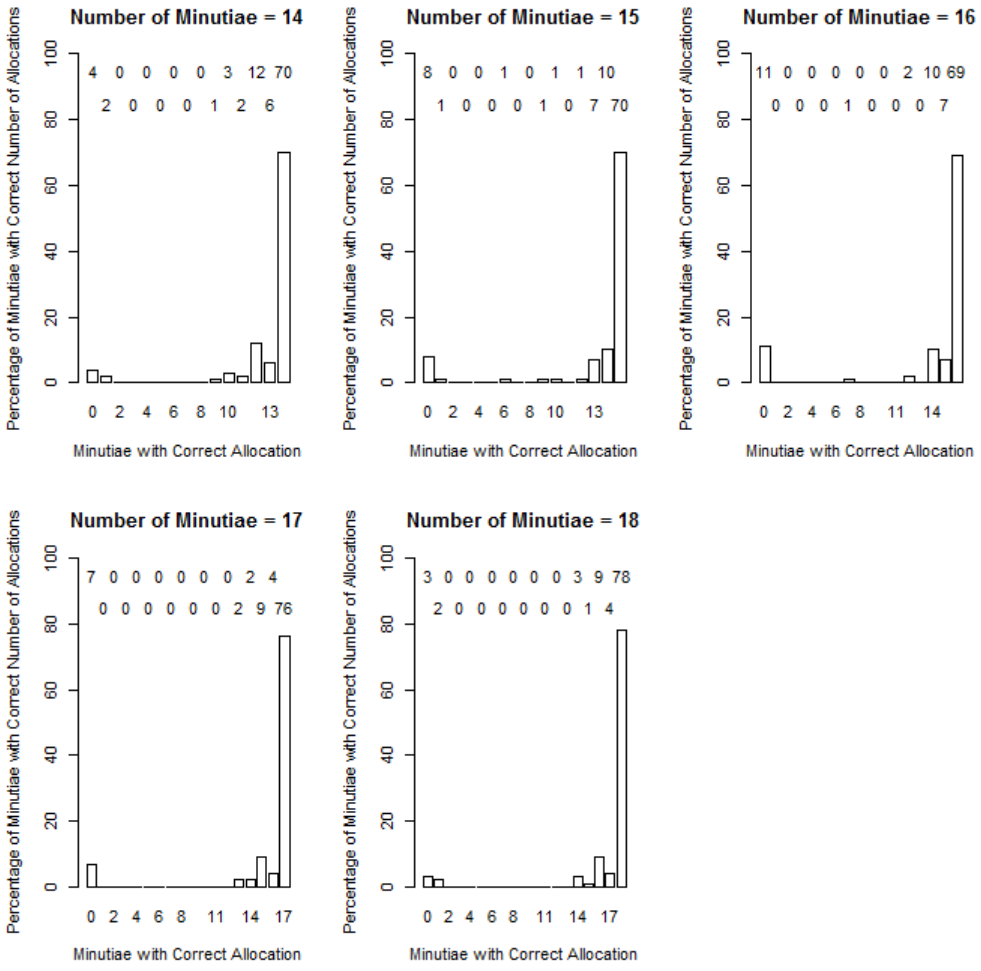

Figure 8.4: Barplot showing how many correct allocations were made in each experiment between the fingermark and the correct fingerprint (minutiae numbers 14-18)

the number of cases passing this decreases as σ increases, the results can be seen in Table 8.4. For the lower values of σ however, we see that in the majority of cases there is very strong evidence to support the proposition that the fingermark came from the "correct" fingerprint.

Value of σ	1	2	3	4	5	6
No. with LR>10,000 for Correct Fingerprint as Best Match	100	84	25	3	0	0

Table 8.4: A table showing the number of times the matching algorithm identified the correct fingerprint as the best match for differing values of σ (out of 100) with a likelihood ratio greater than 10,000. The number of minutiae is 10 throughout.

When randomly selecting a fingerprint we know not to be the one that the fingermark originated from and using this as our comparison print we would expect as in Section 7.4.2 that the results would show no evidence to support the match. The results can be seen in Figure 8.6, here only one case passes the lowest threshold of evidence. We would expect there to be an odd occurrence of this just by chance, however we would want the number of these to be low as is the result presented.

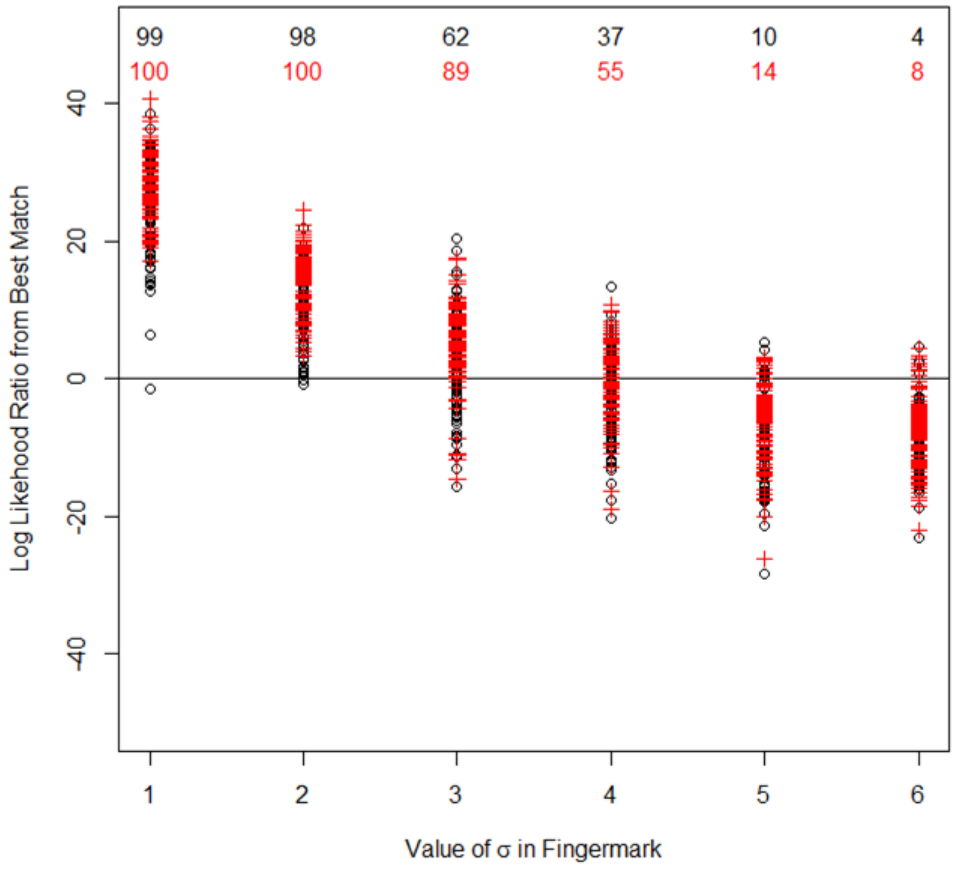

Figure 8.5: Plot showing the log likelihood ratios for the best matches for changes of σ in the construction of the fingermarks, each is labelled with a number which demonstrates how many from 100 runs have a log likelihood ratio greater than zero (black circles = pre-optimisation, red crosses = post-optimisation)

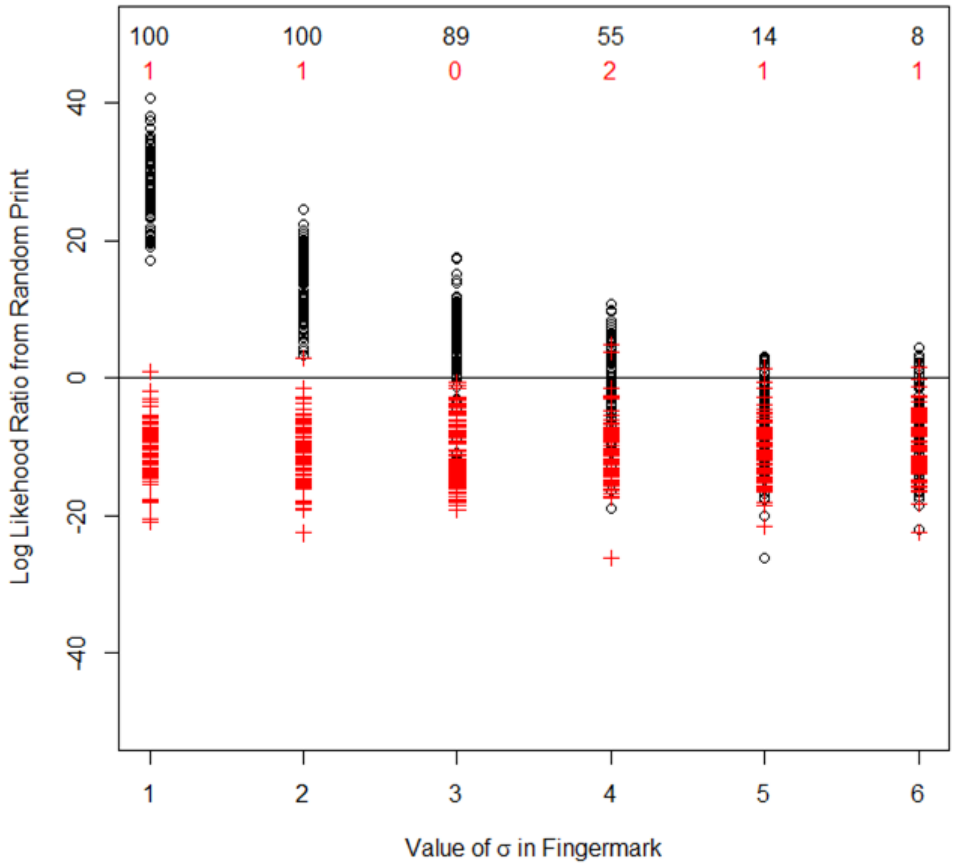

Figure 8.6: Plot showing the log likelihood ratios for the best match in a randomly selected fingerprint for various values of σ in the construction of the fingermarks. Each is labelled with a number which demonstrates how many from 100 runs have a log likelihood ratio greater than zero. The number of minutiae is 10 throughout. (black circles = "correct" fingerprint post-optimisation, red crosses = random fingerprint)

The process of optimistion has a lesser effect on the changes in σ as many of the lower values matched very well to begin with. It can be seen in Table 8.5 that for the values of 1 to 4 optimisation changes none of the allocations however for higher values of σ, which don't fit very well, optimisation changes the allocation in many cases. Figure 8.7 shows how many minutiae in the fingermark are allocated to the correct minutiae in the correct fingerprint. Again it can be seen that for values of 1 to 4 there are no changes, however, for values of $\sigma = 5$ and $\sigma = 6$ the optimisation has a great effect as many

more cases allocate all $N = 10$ minutiae correctly.

Value of σ	1	2	3	4	5	6
No. of Allocation Changes	0	0	0	0	36	37

Table 8.5: A table showing the number of times the allocation of minutiae in the fingermark to the correct fingerprint changed post-optimisation

The number of correct fingerprints identified as the best match from the matching algorithm increases after optimisation. Table 8.6 shows us that the matching algorithm now performs well for values of σ up to 4 and especially well for the values 1 to 3 in which nearly every case is identified properly.

Value of σ	1	2	3	4	5	6
No. with Correct Fingerprint Identified as Best Match (Pre)	99	99	65	42	12	4
No. with Correct Fingerprint Identified as Best Match (Post)	100	100	90	57	16	10

Table 8.6: A table showing the number of times the matching algorithm identified the correct fingerprint as the best match for differing values of σ (out of 100)

8.4 Discussion

The results above show that optimisation has improved the strength of the match between the best matching fingerprint and the fingermark, whilst also maintaining a result of no or limited evidence to support the match when this is not the case. This is shown by an increase in the number of "correct" fingerprints with a log likelihood ratio greater than zero. It can also be seen that for low numbers of minutiae ($N = 8, 9, 10$) the algorithm is allocating the minutiae well in the correct fingerprint but the number of cases where the correct fingerprint is identified as the best match is still low. This could be attributed to the optimisation step improving the strength of the match between the best matching fingerprint and the fingermark even if it is the wrong fingerprint. The same situation occurs with large values of σ.

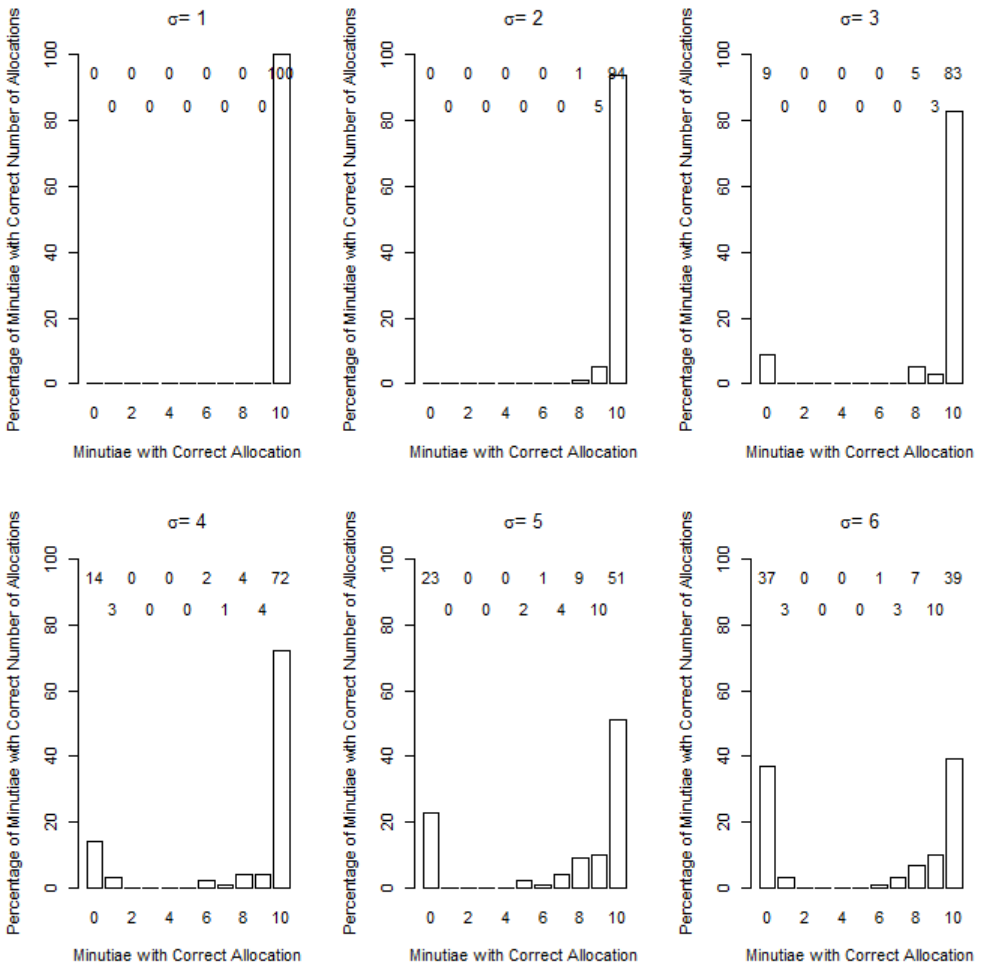

Figure 8.7: Barplot showing how many correct allocations were made in each experiment between the fingermark and the correct fingerprint with varying σ

8.5 Summary

The previous sections have shown that the method of optimisation employed here, namely carrying out a second transformation, works well with this data. It strengthens the matches between the fingermark and the correct fingerprint by increasing the log likelihood ratio and in many cases pushing this above zero. In addition to this, the optimisation step increases the number of correct allocations we see in all cases as well as increasing the number of occasions where the correct fingerprint is identified as the best match by the matching algorithm.

Chapter 9

Expansion with Minutia Type

9.1 Introduction

The optimisation step of the algorithm explored in the previous chapter certainly adds value to the method and increases the performance by increasing the log likelihood ratio for the match between the fingermark and correct fingerprint. In order to improve the method further minutiae types are now incorporated into the algorithm. By extending the method in this way another layer of detail is taken into account which should output more reliable results. The previous simulation experiments from Sections 7.4 and 8.3 will be repeated on the updated algorithm and the results will be compared to those obtained post optimisation.

9.2 Method for Inclusion of Minutia Type

The previous matching algorithm devised in Chapter 6 randomly selects a triangle in the fingermark and then searches the fingerprint for similar triangles to create a candidate set of triangles. For each of the triangles (abc) in this candidate set, a value for the transformation of this triangle in the fingerprint to the triangle (ABC) in the fingermark is calculated which is then applied to the whole of the fingermark to give $A', B', C' \ldots, N'$. The key step is the application of the Hungarian Algorithm which finds a unique allocation of the N minutiae in the transformed fingermark to corresponding points in the fingerprint which minimises the sum of squared Euclidean

distance between the two sets of minutiae. The set of minutiae from the fingerprint relating to this allocation are $a^*, \ldots, n^*$. The optimisation step (see Chapter 8 for more details) then applies another transformation, however this time instead of being based on the triangle of minutiae it is based on all N minutiae in the fingermark, this gives, $A'', B'', C'', \ldots, N''$. Now minutia A'' in the transformed (and optimised) fingermark is allocated to minutia a^* in the fingerprint.

The basis of this extension to the method is to compare the minutiae types of the N minutiae in $A'', B'', C'', \ldots, N''$ with that of $a^*, \ldots, n^*$. If the type of A'' does not match with a^* then we apply a penalty to the sum of squared Euclidean distance. This method of applying a penalty to the minutiae types element of the matching algorithm is not necessarily modelled in a formal way by building on the Likelihood, instead the penalty builds on the sum of squares approach. The method used is both intuitive and computable hence the results are interpretable. The method can be written as:

$$\underline{S^{abc}_{ABC}} = \{|A''a^*|^2 + |B''b^*|^2 + \ldots + |N''n^*|^2\}p^w \tag{9.1}$$

where p is the penalty applied for an incorrect minutia type and w is the number of incorrect minutiae types. A correct minutia type is essentially multiplied by 1 so that the value for the sum of squared distance stays the same. Since a small value for the sum of squared distance between two minutiae represents a good match we want the penalty to increase the value of the sum of squared distance for a wrong match of minutiae types. For the experiments below we set the penalty to be $p = 1.2$. This value for p was based on some initial experiments to obtain a value for p which fit our requirements. The aim of including minutiae types is to strengthen the match between a fingermark and a fingerprint in that the minutiae are similar in both locations and types, this is done by applying a penalty if the type of the minutia in the fingermark is different from the minutia it is allocated to in the fingerprint. However, we know anecdotally that minutiae types are not always labelled correctly due to both human error and poor quality of fingermarks. For this reason it was important that the penalty wasn't too high that it excluded the correct fingerprint if only one or two minutiae were "typed" incorrectly. This led to a value of $p = 1.2$ being selected. We then take the value of the lowest overall sum of squares to be the best match for that fingerprint, this is described as:

$$S(X) = \min(\underline{S^{abc}_{ABC}}) \text{ for all } abc \text{ in } X. \tag{9.2}$$

It is intended that this extension to include minutiae types should increase the strength of the match between the fingermark and the correct fingerprint since the locations of the minutiae will be similar as well as the types of minutiae. The additional custom R code required to implement this step is located in Appendix C.

9.3 Results from the Inclusion of Minutia Type

9.3.1 Results for Changes in Number of Minutiae

The simulation experiments carried out to assess whether optimisation was beneficial were repeated here, where the results of including minutiae types in the matching algorithm were compared with the results post optimisation. We anticipate that by including minutiae types this would enhance the performance of the model further. As with previous experiments in Section 8.3.1 the number of minutiae in the fingermark was varied and the value for σ was set to 4.

The inclusion of minutiae types shows a similar increasing pattern of log likelihood ratio as N increases, this was also seem in previous experiments. In most cases including minutiae types in the algorithm increases the number of cases where the log likelihood ratio exceeds zero, in two cases it stays the same as the number for post optimisation, and in one case ($N = 17$) decreases, these results can be seen in Figure 9.1. For larger numbers of minutiae the extension of the algorithm seems to have little effect since the method already performs well however for lower numbers of minutiae, $N = 8 - 10$ the extended algorithm shows a marked improvement. If we look specifically at how many of these LR are now greater than 10,000 and show "very strong evidence to support" the match between the fingermark and the "correct" fingerprint we see that in all cases including minutiae types increases the number of cases crossing this threshold. The results can be seen in Table 9.1, for the higher numbers of minutiae in the fingermark the majority of cases are matching with the highest level of support from Table 7.1.

With the inclusion of minutiae types in the matching algorithm we compare a random fingerprint to the fingermark and assess how good the match is. The results can be

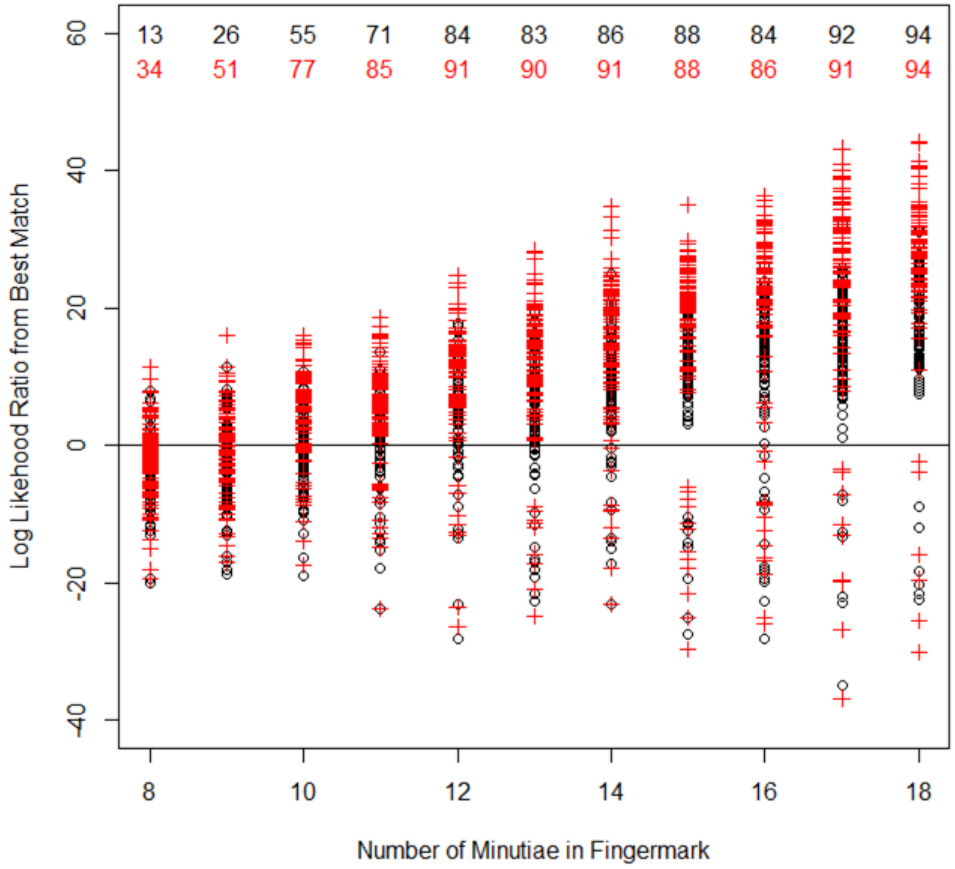

Figure 9.1: Plot showing the log likelihood ratios for the best matches for different numbers of minutiae in the fingermarks, each is labelled with a number which demonstrates how many from 100 runs have a log likelihood ratio greater than zero (black circles = post-optimisation, red crosses = post-minutiae types)

No. of Minutiae in Fingermark	8	9	10	11	12	13	14	15	16	17	18
No. with LR>10,000 for Correct Fingerprint as Best Match	2	7	25	29	60	70	79	86	83	89	94

Table 9.1: A table showing the number of times the matching algorithm identified the correct fingerprint as the best match for differing numbers of minutiae in the fingermark (out of 100) with a likelihood ratio greater than 10,000

seen in Figure 9.2 and show that whilst there are some cases with a log likelihood ratio greater than zero these numbers are low and don't happen very often. The number of these false positives is very infrequent when comparing these cases to the use of the "correct" fingerprint for matching.

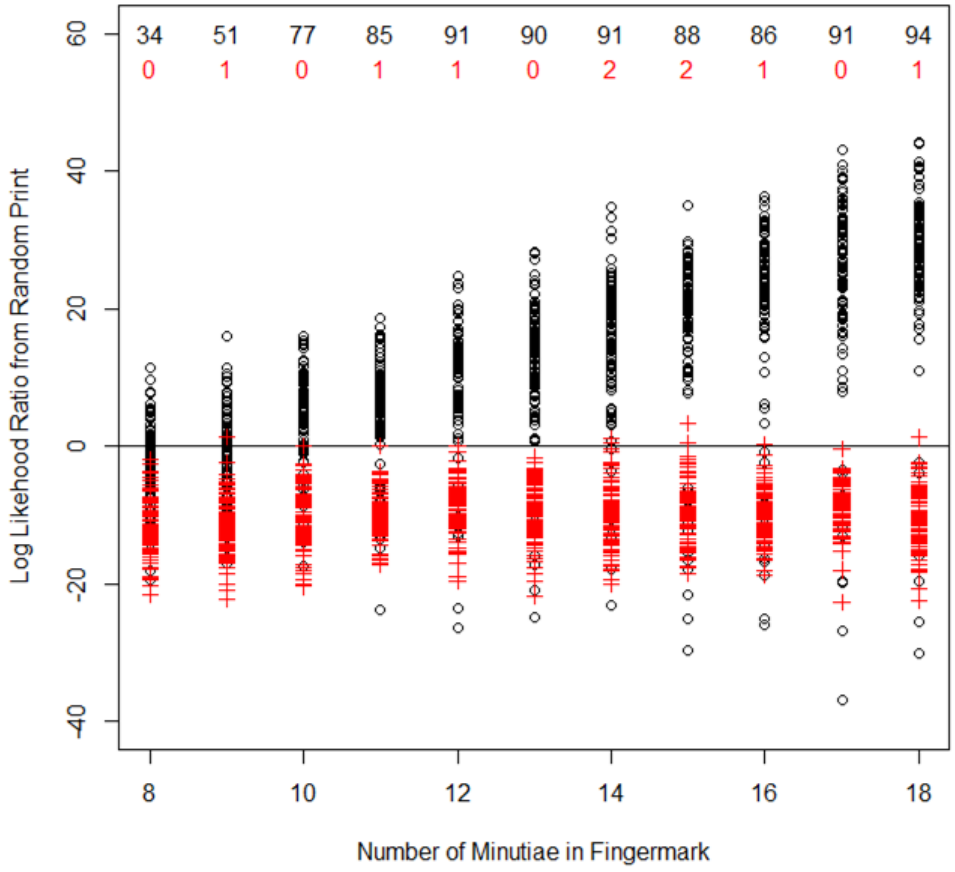

Figure 9.2: Plot showing the log likelihood ratios for the best match in a randomly selected fingerprint for different numbers of minutiae in the fingermarks. Each is labelled with a number which demonstrates how many from 100 runs have a log likelihood ratio greater than zero (black circles = "correct" fingerprint post-minutiae types, red crosses = random fingerprint)

By comparing Figures 9.3 and 9.4 to the corresponding plots in Section 8.3.1 we can see that in most cases the percentage of times when all N minutiae have been allocated correctly has increased or stayed the same (except for $N = 17$). This implies that the

matching algorithm has been improved by extending it to include minutiae types since this is increasing the number of correct allocations of minutiae in the fingermark to minutiae in the original fingerprint it was simulated from. It can also be seen that post optimisation the percentage of cases where zero minutiae were allocated correctly ranged between 3% and 22%, here this has dropped to between 1% and 19%.

Table 9.2 shows the number of times the allocation of minutiae in the fingermark to the correct fingerprint changed after the algorithm was extended. By inspection it can be seen that there were a similar number of allocation changes for all N ranging from 4% to 13%.

No. of Minutiae in Fingermark	8	9	10	11	12	13	14	15	16	17	18
No. of Allocation Changes	13	11	4	9	6	9	7	7	12	8	5

Table 9.2: A table showing the number of times the allocation of minutiae in the fingermark to the correct fingerprint changed after the inclusion of minutiae types

When thinking about the best match that can be made it is hoped that this will be the correct fingerprint which is defined as the fingerprint that the fingermark originates from. In all cases the number of times where the correct fingerprint is the best match has increased or stayed the same as post optimisation except again for $N = 17$ where there seems to have been one case where including minutiae types has changed how that simulation performed. The results can be seen in Table 9.3, it is also worth noting that the algorithm has increased the number of correct matches identified as the best match by a large amount for low numbers of minutiae.

No. of Minutiae in Fingermark	8	9	10	11	12	13	14	15	16	17	18
No. with Correct Fingerprint Identified as Best Match (Post-optimisation)	17	29	57	71	84	83	86	88	84	92	94
No. with Correct Fingerprint Identified as Best Match (Post-minutiae types)	41	54	79	85	91	90	92	88	86	91	94

Table 9.3: A table showing the number of times the matching algorithm identified the correct fingerprint as the best match for differing numbers of minutiae in the fingermark (out of 100)

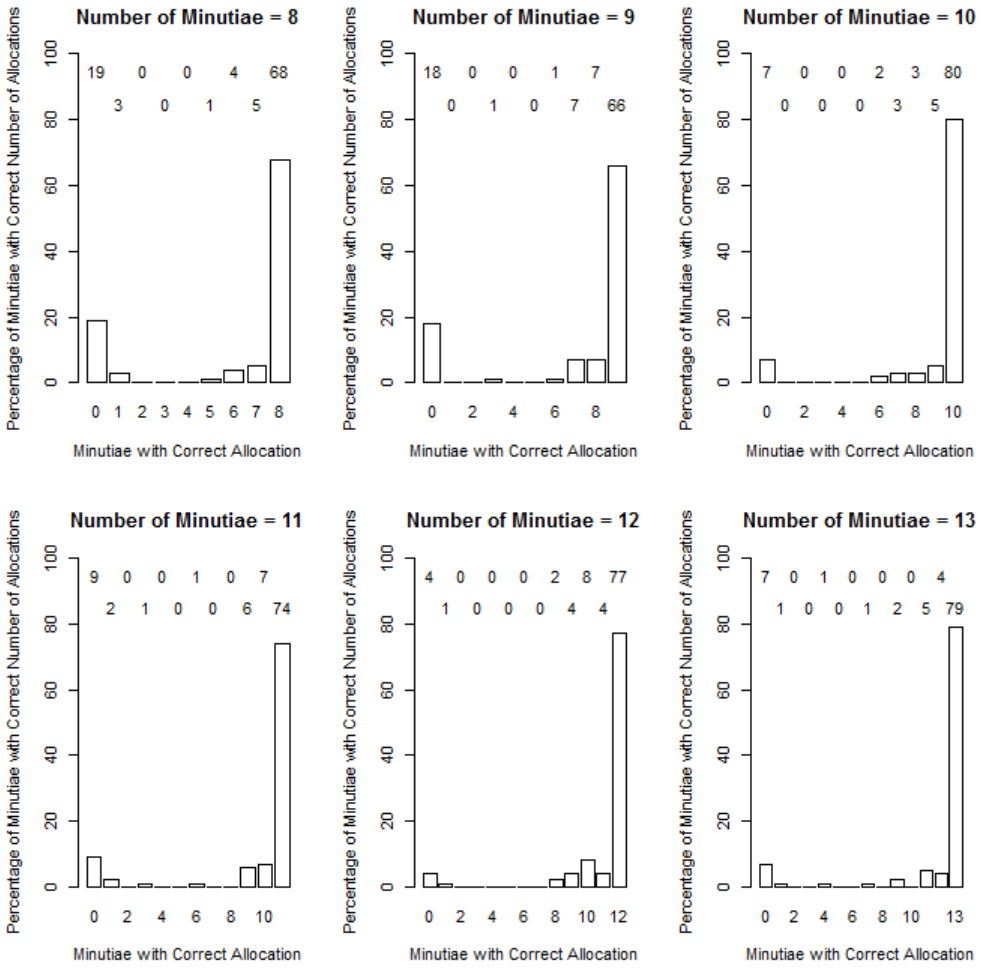

Figure 9.3: Barplot showing how many correct allocations were made in each experiment between the fingermark and the correct fingerprint (minutiae numbers 8-13)

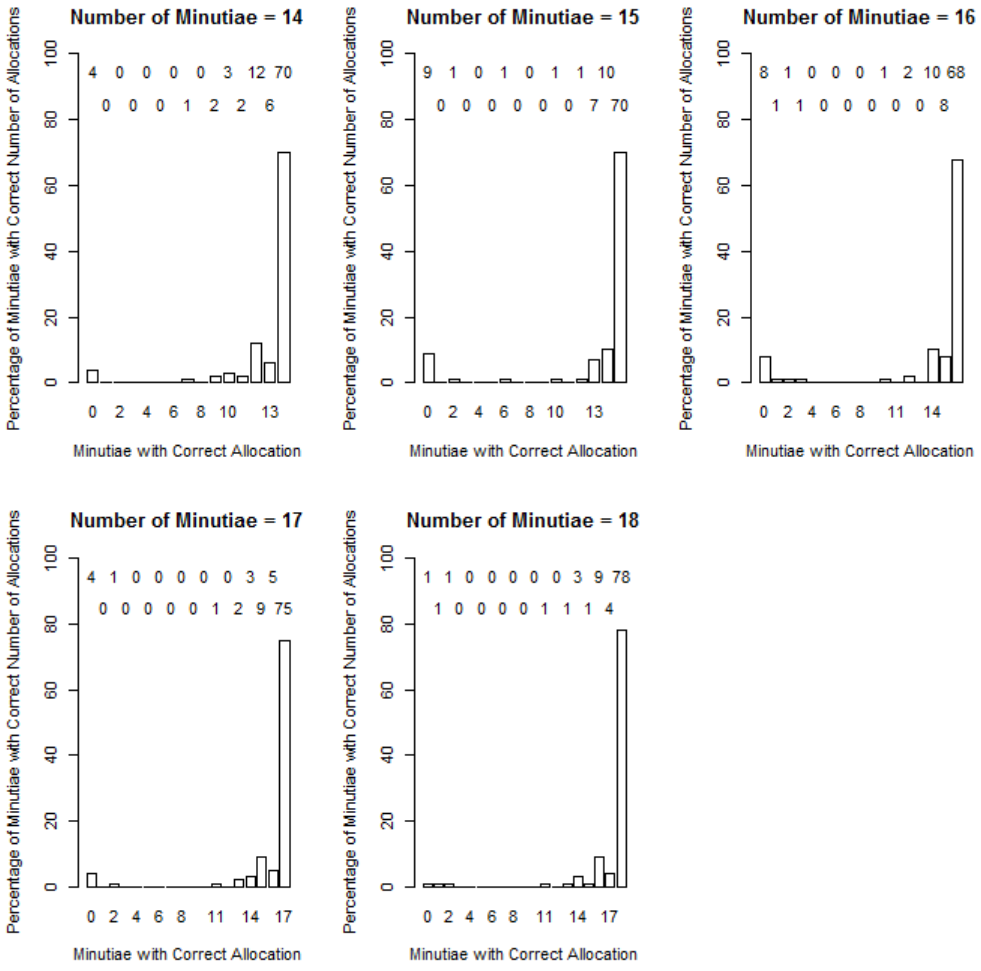

Figure 9.4: Barplot showing how many correct allocations were made in each experiment between the fingermark and the correct fingerprint (minutiae numbers 14-18)

9.3.2 Results for Changes in σ

The inclusion of minutiae types into the matching algorithm also had an effect on the simulation experiments where the value of σ was varied. As in previous chapters the value for N (the number of minutiae) was fixed at 10. As expected the values for the log-likelihood ratio decrease as σ increases, this implies that as the distortion in the fingermark decreases the algorithm provides a stronger match between the best match within a fingerprint and the fingermark. This can be seen in Figure 9.5. The number of cases where the log-likelihood ratio is greater than zero (hence providing some evidence that the fingermark comes from the "correct" fingerprint) has also increased especially for the higher values of σ. If we look at the higher threshold for the LR which relates to very strong evidence to support the proposition that the fingermark originated from the "correct" fingerprint we can see that for low values of σ the matching algorithm produces a positive results and for higher values of σ cases are starting to cross this threshold. Results can be seen in Table 9.4.

Value of σ	1	2	3	4	5	6
No. with LR>10,000 for Correct Fingerprint as Best Match	100	93	57	25	0	1

Table 9.4: A table showing the number of times the matching algorithm identified the correct fingerprint as the best match for differing values of σ (out of 100) with a likelihood ratio greater than 10,000. The number of minutiae is 10 throughout.

By repeating the experiments using a random fingerprint to compare against the fingermark and not the one we know to be "correct" we see in Figure 9.6 that in only a small number of cases (for higher values of σ) does the log likelihood ratio exceed zero and in these cases it is still a low value. This means that in these cases we would say that there is "limited evidence to support" a match between the fingermark and the fingerprint.

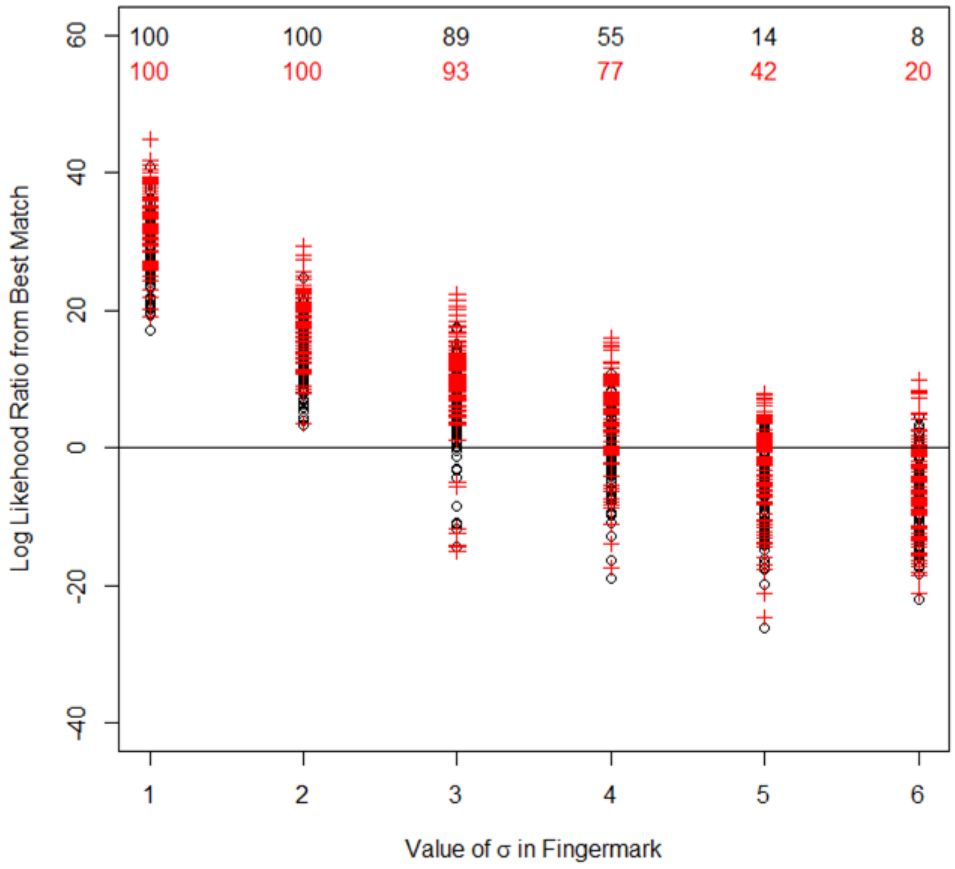

Figure 9.5: Plot showing the log likelihood ratios for the best matches for changes of σ in the construction of the fingermarks, each is labelled with a number which demonstrates how many from 100 runs have a log likelihood ratio greater than zero (black circles = post-optimisation, red crosses = post-minutiae types)

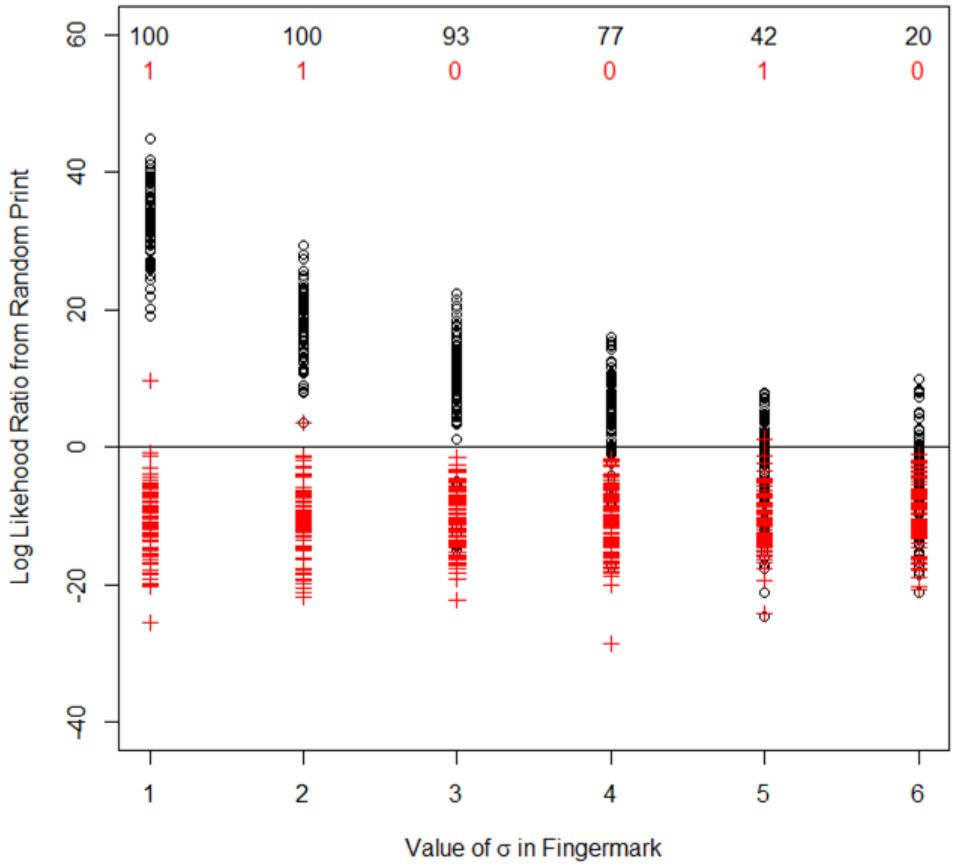

Figure 9.6: Plot showing the log likelihood ratios for the best match in a randomly selected fingerprint for various values of σ in the construction of the fingermarks. Each is labelled with a number which demonstrates how many from 100 runs have a log likelihood ratio greater than zero. The number of minutiae is 10 throughout. (black circles = "correct" fingerprint post-minutiae tpes, red crosses = random fingerprint)

Since many of the lower values of σ already performed well (both before and after optimisation) we would expect there to be few changes in the allocation after including minutiae types, this is reflected in Table 9.5 where the low values of σ show no changes and the higher values have some. When we consider the correct fingerprint (the fingerprint that the fingermark was created from) we look at how many minutiae in the fingermark are allocated to the correct minutiae in the original fingerprint. In Figure 9.7 it can be seen that there are no changes in the number of times the algorithm

matches all $N = 10$ minutiae correction for $\sigma = 1$ and 2, this corresponds with Table 9.5 which also showed no changes for these values. However by taking into account minutiae types in the matching algorithm the results show that for higher values of σ the number of times all ten minutiae are allocated correctly increases. In addition the number of cases where zero minutiae are correctly allocated decreases after taking into account minutiae types.

Value of σ	1	2	3	4	5	6
No. of Allocation Changes	0	0	5	4	15	26

Table 9.5: A table showing the number of times the allocation of minutiae in the fingermark to the correct fingerprint changed after the inclusion of minutiae types

One of the most important factors in the simulation experiments is observing how many times the fingerprint which the fingermark originates from (the "correct fingerprint") is identified as the best match of all the fingerprints tested in that experiment. The result of this from the 100 experiments can be seen in Table 9.6 along with how this compares to the performance of the algorithm after optimisation. For all values of σ incorporating minutiae types into the method has increased the number of cases where the correct fingerprint is determined as the best match.

Value of σ	1	2	3	4	5	6
No. with Correct Fingerprint Identified as Best Match (Post-optimisation)	100	100	90	57	16	10
No. with Correct Fingerprint Identified as Best Match (Post-minutiae types)	100	100	93	79	44	23

Table 9.6: A table showing the number of times the matching algorithm identified the correct fingerprint as the best match for differing values of σ (out of 100)

9.4 Discussion and Summary

Minutiae types were included into the already optimised algorithm by applying a penalty to cases where the minutiae allocations didn't match in type. After including minutiae types in to the matching algorithm its performance increased, particularly when σ was varied. For most values of σ investigated the algorithm worked well, the

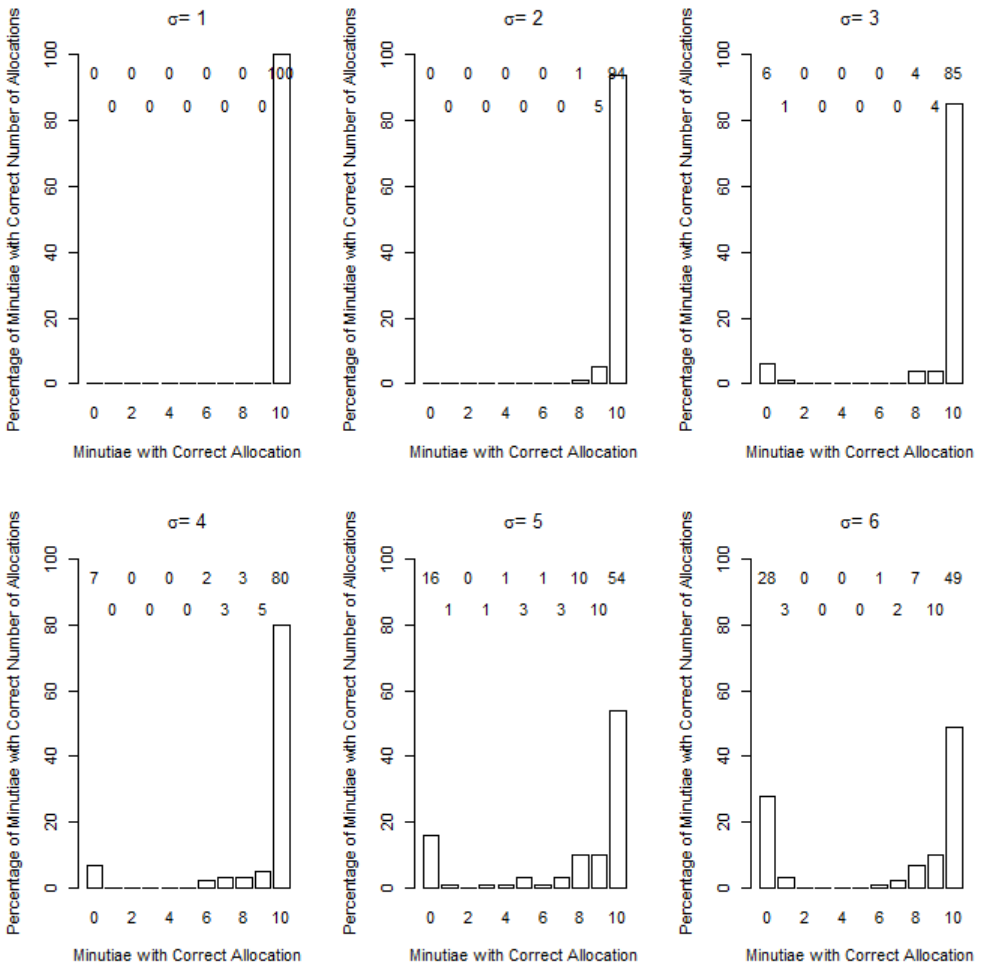

Figure 9.7: Barplot showing how many correct allocations were made in each experiment between the fingermark and the correct fingerprint with varying σ

performance increased as σ decreased. The results are not so clear for changes in N, although the algorithm can be described as working well for larger values of N this pattern is less obvious. For the larger numbers of minutiae the results are fairly similar and show no distinction. Despite the lack of obvious distinction for higher values of N it is true to say that extending the model to include minutiae types has made a positive difference. There was one occasion where this extension didn't help, for one experiment for $N = 17$ including minutiae types had a negative impact. Although this is not ideal it is to be expected that the model will not always perform perfectly but only one occurrence of this is not alarming.

For these experiments the value of the penalty was set to $p = 1.2$. This was based on a compromise between wanting to strengthen the match between a fingermark and a fingerprint where the minutiae types were the same but also not wanting to exclude a correct or good matching fingerprint if minutiae types had been identified wrong in the first case (i.e. if the fingermark is of poor quality). Despite this other values for p were tried. A smaller set of experiments were carried out on a substantially different value set as $p = 2$. This was found to be clearly too high as the algorithm performed much worse than the original simple matching algorithm and in most cases struggled to find the correct fingerprint that the fingermark originated from, even in cases of small σ or large N where the algorithm is known to perform well. Due to this it is obvious that the optimum value for p lies somewhere below this value.

Chapter 10

Conclusions and Further Work

10.1 Spatial Modelling

This research started by carrying out some preliminary investigations into fingerprints. Contrary to many current methods involving fingerprints which focus on simulating ridge flow, and treating minutiae as a secondary feature to be added later, this work places the priority the other way around. By plotting minutiae and investigating how the direction of ridge flow can be determined by the location in relation to the core and delta of the fingerprint it was decided that focusing on minutiae was a more interesting avenue. It is also the case that, for a forensic investigation, minutiae are the main feature in a fingerprint and a fingermark from a crime scene which are compared. Hence any research into minutiae specifically would be novel and valuable.

How minutiae within a fingerprint interacted with each other was investigated by carrying out initial tests for complete spatial randomness to assess whether the pattern of minutiae displayed on real fingerprints could be described as random. By looking visually at density plots it was clear that minutiae are more likely to be located around points of significance (namely the core and delta). Despite this the K-functions only showed weak evidence that the patterns did not adhere to complete spatial randomness.

Since anecdotally from fingerprint examiners, minutiae are considered clustered it was decided to try and extend the work by fitting spatial models to the fingerprint data to find a spatial point process which describes the locations of minutiae. Two models were

considered: the Strauss process and an Isotropic Centred Poisson Process. Whilst the Strauss process was a good starting point and had characteristics which were applicable in the context of fingerprints (for example, points in the pattern interact), simulations of the model did not realistically represent the data from the Forensic Science Service's dataset of approximately $13,000$ fingerprints. However the Isotropic Centred Poisson Process fared better; this didn't allow the minutiae to interact, but had an intensity function which varied with location. This is specified in such a way that a point is more likely to be located close to the centre of the pattern. Parameter values chosen by a trial and error search appeared to fit the data well.

It was concluded that although it was informative to find a spatial point process which reflected real fingerprint data this method still wouldn't be of much use since that same process with the same parameters could be used to describe many fingerprints. So without being unique to a single fingerprint the model can't be used to help match a fingermark from a crime scene to a fingerprint from a suspect.

10.2 Matching Algorithm

The second half of the thesis focuses on creating a method to match a fingermark to a fingerprint. The algorithm essentially starts with a simulated fingermark and locates a triangle within this which it matches to approximately matching triangles in a fingerprint to create a candidate set. For each of these we calculate a similarity transformation mapping the triangle in the fingermark to one in the fingerprint. This transformation is applied to the whole mark and, using a Hungarian Algorithm, a unique allocation for each minutia in the fingermark to a minutia in the fingerprint is obtained. Once we have this allocation a sum of squared distances can be calculated for the minutiae. This is repeated for all instances in the candidate set and the minimum sum of squared distances is determined to be the optimal match for that fingerprint.

In order to assess how well the model performs, this process was repeated many times and for different numbers of minutiae in the mark and values of σ, which corresponds to how close the fingermark is to the original fingerprint it was simulated from. The method performed adequately but we felt it could be optimised by including another transformation after the Hungarian Algorithm, basing this transformation on all points

in the fingermark. This second optimisation can be thought of as refining the match to minimise the sum of squared distances between the whole of the fingermark and the fingerprint. The results from including this step were positive and the performance of the method improved; it increased the amount of times the original fingerprint used to simulate the fingermark was defined as the best match in the experiment and additionally increased the number of times all N minutiae were allocated correctly for this fingerprint.

A final refinement was made to the algorithm by extending it to include minutiae types; bifurcations and ridge endings. When the types of the minutiae in the allocation of the points in the fingermark and fingerprint differed, we applied a penalty increasing the sum of squares so that this was not as favourable a match. A penalty factor value of 1.2 was selected through trial and error; this value is low enough that an initial misidentification of minutia type would not adversely affect the results of the matching algorithm. This was increased depending on how many minutiae types did not match. This extension of the model also improved its performance in similar ways to the optimisation step above although it is probably fair to say that it didn't have as strong an improvement as the initial optimisation.

Overall the matching algorithm performs well and in many cases selects the correct fingerprint as the best match.

10.3 Further Work and Discussion

There are many ways in which this work could be extended. This area has the potential to be of significant benefit to the forensic science community, especially if the methods could be extended to incorporate other types of forensic evidence. Many other types of forensic evidence could be added into the model to enhance its overall power, one of the obvious ones is DNA evidence which already uses a numerical representation for the strength of the weight of the match. Previous work has been carried out using likelihood ratios to explain DNA and combining this with the fingerprint method derived here could be extraordinarily useful.

Another obvious improvement to the method would be to carry out more testing with other combinations of minutiae numbers and changes in σ. By looking at different

combinations of these two parameters this might give a more robust definition of when the method performs well and help to draw significant conclusions about when it could be applied to real scenarios. Since this would give a measure for when the results could be trusted. Along a similar thread a more systematic approach to fingerprint and fingermark distortion could be investigated. This wasn't investigated heavily during this research but could easily be a subject of its own. There is much scope for assessing whether the distortion in real fingermarks is independent (as modelled in this work) or whether there is usually some underlying dependence structure between the variation in the location of points between the mark and known print. For example if a fingermark has been smudged it is reasonable to believe that there may be some directional dependence in the locations of the minutiae in the fingermark. In addition to this the surface that the fingermark is lifted from may introduce some dependence in the distortion of minutiae locations. An example of this is if a fingermark is lifted from a curved surface (like a glass) we would expect that this might introduce a level of distortion which could be modelled specifically.

Another extension to the model would be to further investigate the penalty applied to non-matching minutiae types in Chapter 9. The method used in this research built on the sum of squares approach and was multiplicative; an additive method could also be investigated. In addition a more formal method based on the likelihood could be developed as well as more tests which explicitly look at optimising the value of the penalty. Finally a useful piece of work would be to investigate how the value for the penalty fundamentally interacts with the probability of error in classifying minutiae types in the fingermark.

The most obvious way that this work would benefit from extension is by using real fingermarks. Since we didn't have access to a real set with the corresponding fingerprints, we simulated them. However, undoubtedly there is no replacement for carrying out these tests using real fingermarks that are of poor quality and assessing how well the algorithm matches them to their corresponding fingerprint. This would give more reliability and better trust in the performance of the matching algorithm and subsequent extensions (optimisation, inclusion of minutiae types etc.).

References

Alonso, C., V. Galea, E. Gutierrez, and J. Martinez (2007). Biological variability of the minutiae in the fingerprints of a sample of the spanish population. Volume 172, pp. 98–105. Elsevier.

Babler, W. J. (1991). Embryologic development of epidermal ridges and their configurations. Dermatoglyphics: science in transition, 95–112.

Barnes, J. G. (2010). Chapter 1. History, The Fingerprint Sourcebook, US Department of Justice.

Bookstein, F. L. (1989). Principal warps: Thin-plate splines and the decomposition of deformations. IEEE Transactions on pattern analysis and machine intelligence 11(6), 567–585.

Burkard, R., M. Dell'Amico, and S. Martello (2009). Assignment Problems. SIAM e-books. Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104).

Cabral, B. and L. C. Leedom (1993). Imaging vector fields using line integral convolution. In Proceedings of the 20th annual conference on Computer graphics and interactive techniques, pp. 263–270. ACM.

Cappelli, R., A. Erol, D. Maio, and D. Maltoni (2000). Synthetic fingerprint: image generation. In 15th International Conference on Pattern Recognition, 2000. Proceedings, Volume 3, pp. 471–474. IEEE.

Cappelli, R., A. Lumini, D. Maio, and D. Maltoni (2007). Fingerprint image reconstruction from standard templates. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(9), 1489–1503.

Champod, C., N. M. Egli, and P. Margot (2007). Evidence evaluation in fingerprint comparison and automated fingerprint identification systems, modelling within finger variability. Forensic science international 167(2), 189–195.

Champod, C., J. Fierrez-Aguilar, J. Gonzalez-Rodriguez, J. Ortega-Garcia, and D. Ramos-Castro (2005). Between-source modelling for likelihood ratio computation in forensic biometric recognition. In Audio-and Video-Based Biometric Person Authentication, pp. 1080–1089. Springer.

Cole, S. A. (2004). Grandfathering evidence: Fingerprint admissibility rulings from jennings to llera plaza and back again. American Criminal Literature Review 41, 1189.

Cox, T. F. and M. A. A. Cox (2010). Multidimensional scaling; second edition. CRC Press.

Cressie, N. (1993). Statistics for spatial data, revised edition. J. Wiley.

Cummins, H. and R. W. Kennedy (1940). Purkinje's observations (1823) on finger prints and other skin features. Journal of Criminal Law and Criminology (1931-1951) 31(3), 343–356.

Dass, S., A. Jain, and Y. Zhu (2005). Statistical models for assessing the individuality of fingerprints. In AUTOID '05: Proceedings of the Fourth IEEE Workshop on Automatic Identification Advanced Technologies, pp. 3–9. IEEE Computer Society.

Dass, S. C., A. K. Jain, and Y. Zhu (2007). Statistical models for assessing the individuality of fingerprints. IEEE Transactions on Information Forensics and Security 2(3), 391–401.

Dass, S. C., M. Li, et al. (2009). Hierarchical mixture models for assessing fingerprint individuality. The Annals of Applied Statistics 3(4), 1448–1466.

Dass, S. C. and Y. Zhu (2006). Compound stochastic models for fingerprint individuality. In Pattern Recognition, 2006. ICPR 2006. 18th International Conference, Volume 3, pp. 532–535. IEEE.

Daubert v. Merrell Dow Pharmaceuticals Inc (1993). 509 U.S. 579, 113 S. Ct. 2786.

Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. Journal of the Optical Society of America, A 2(7), 1160–1169.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the em algorithm. Journal of the Royal Statistical Society. Series B (Methodological), 1–38.

Diggle, P. J. et al. (1983). Statistical analysis of spatial point patterns. Academic Press.

Dryden, I. L. and K. Mardia (1998). Statistical shape analysis. Wiley series in probability and statistics: Probability and statistics. J. Wiley.

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

Faulds, H. (1880). On the skin-furrows of the hand. Nature 22.

Feng, J. and A. K. Jain (2009). FM model based fingerprint reconstruction from minutiae template. In Advances in Biometrics, pp. 544–553. Springer.

Fierrez-Aguilar, J., J. Gonzalez-Rodriguez, J. Ortega-Garcia, and D. Ramos-Castro (2005). Bayesian analysis of fingerprint, face and signature evidences with automatic biometric systems. Forensic science international 155(2), 126–140.

Flood, M. (1956). The traveling-salesman problem. Operations Research 4(1), 61–75.

Galton, F. (1892). Finger prints. Macmillan and Company.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2003). Bayesian data analysis. CRC Press.

Gonzalez-Rodriguez, J., J. Fierrez-Aguilar, and J. Ortega-Garcia (2003). Forensic identification reporting using automatic speaker recognition systems. In IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP'03), Volume 2, pp. 93–96.

Great Britain Home Office, Forensic Science Pathology Unit, United Kingdom (2005). DNA expansion programme 2000-2005: Reporting achievement.

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

Illian, J., A. Penttinen, H. Stoyan, and D. Stoyan (2008). Statistical analysis and modelling of spatial point patterns, Volume 70. John Wiley & Sons.

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

Jain, A. K., A. Ross, and J. Shah (2007). From template to image: Reconstructing fingerprints from minutiae points. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(4), 544–560.

Kuhn, H. W. (1955). The hungarian method for the assignment problem. Naval research logistics quarterly 2(1-2), 83–97.

Laidlaw, D., R. Kirby, C. Jackson, J. Davidson, T. Miller, M. Da Silva, W. Warren, and M. Tarr (2005). Comparing 2d vector field visualization methods: A user study. IEEE Transactions on Visualization and Computer Graphics 11(1), 59–70.

Larkin, K. G. and P. A. Fletcher (2007). A coherent framework for fingerprint analysis: are fingerprints holograms? Optics Express 15(14), 8667–8677.

Lee, P. M. (2004). Bayesian statistics: An introduction, third edition.

Levi, G. and F. Sirovich (1972). Structural descriptions of fingerprint images. Information Sciences 4(3), 327–355.

Lindley, D. V. (1977). A problem in forensic science. Biometrika 64(2), 207–213.

Llewelyn, S. (2012). Discussions on the paper by Neumann, C. and Evett, I. W. and Skerrett, J. - 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.

Maltoni, D., D. Maio, A. K. Jain, and S. Prabhakar (2003). Handbook of Fingerprint Recognition. Springer-Verlag New York, Inc.

Meuwly, D., A. Goode, A. Drygajlo, J. Gonzalez-Rodriguez, and J. L. Molina (2003). Validation of forensic automatic speaker recognition systems: Evaluation frameworks for intelligence and evidencial purposes. In Forensic Science International, Volume 136, pp. 364–364.

Moenssens, A. (1971). Fingerprint techniques. Chilton Book Company London.

Munkres, J. (1957). Algorithms for the assignment and transportation problems. Journal of the Society for Industrial and Applied Mathematics 5(1), 32–38.

Nelder, J. A. and R. Mead (1965). A simplex method for function minimization. The Computer Journal 7(4), 308–313.

Neumann, C., I. W. Evett, and J. Skerrett (2012). 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.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling (1988). Numerical recipes: The art of scientific computing. Cambridge university press.

Ripley, B. D. (1977). Modelling spatial patterns. Journal of the Royal Statistical Society. Series B (Methodological), 172–212.

Ripley, B. D. (1987). Stochastic simulation. J. Wiley.

Sherlock, B. G. and D. M. Monro (1993). A model for interpreting fingerprint topology. Pattern recognition 26(7), 1047–1055.

Silverman, B. W. (1986). Density estimation for statistics and data analysis, Volume 26. CRC Press.

Srihari, S. N., H. Srinivasan, and G. Fang (2008). Discriminability of fingerprints of twins. Journal of Forensic Identification 58(1), 109.

Stoney, D. A. (2001). Measurement of fingerprint individuality. In: H. C. Lee, R. E. Gaensslen, Advances in fingerprint technology 2.

Strauss, D. J. (1975). A model for clustering. Biometrika 62(2), 467–475.

Su, C. and S. N. Srihari (2008). Generative models for fingerprint individuality using ridge models. In The 19th International Conference on Pattern Recognition, 2008. ICPR 2008, pp. 1–4. IEEE.

U.S. v. Byron Mitchell (1999). Criminal Action No. 96-407, U.S. District Court for the Eastern District of Pennsylvania.

Vizcaya, P. R. and L. A. Gerhardt (1996). A nonlinear orientation model for global description of fingerprints. Pattern Recognition 29(7), 1221–1231.

Appendix A

R code - Matching Algorithm

```
##########################################
### Author: Stephanie Llewelyn ###
### Date: June 2013 ###
### Description: R code to calculate matching algorithm and then ###
### likelihood ratio for one fingermark (including generation of this ###
### fingermark) against 100 fingerprints plus the original fingerprint ###
### that the mark was simulated from. For the fingermark in this  ###
### script number of minutiae = 10, sigma squared = 4. ###
##########################################



### Step 1: load relevant libraries ###

library(MASS)
library(shapes)
library(clue)
library(gtools)

### Step 2: Load in the 12 separate datasets creating two new columns ###
```

```
### for fingertype (from: index/middle/ring/thumb) and pattern type  ###
### (from: arch/ulnar loop/whorl/arch) ###

a1<-read.csv("Index_arch.csv")
a1[,1]<-"I"
a1[,2]<-"A"
colnames(a1)[1]<- "FingerType"
colnames(a1)[2]<- "PatternType"

a2<-read.csv("Index_Ulnar_loop.csv")
a2[,1]<-"I"
a2[,2]<-"L"
colnames(a2)[1]<- "FingerType"
colnames(a2)[2]<- "PatternType"

a3<-read.csv("Index_whorl.csv")
a3[,1]<-"I"
a3[,2]<-"W"
colnames(a3)[1]<- "FingerType"
colnames(a3)[2]<- "PatternType"

a4<-read.csv("Middle_arch.csv")
a4[,1]<-"M"
a4[,2]<-"A"
colnames(a4)[1]<- "FingerType"
colnames(a4)[2]<- "PatternType"

a5<-read.csv("Middle_Ulnar_loop.csv")
a5[,1]<-"M"
a5[,2]<-"L"
colnames(a5)[1]<- "FingerType"
colnames(a5)[2]<- "PatternType"

a6<-read.csv("Middle_whorl.csv")
```

```
a6[,1]<-"M"
a6[,2]<-"W"
colnames(a6)[1]<- "FingerType"
colnames(a6)[2]<- "PatternType"

a7<-read.csv("Ring_arch.csv")
a7[,1]<-"R"
a7[,2]<-"A"
colnames(a7)[1]<- "FingerType"
colnames(a7)[2]<- "PatternType"

a8<-read.csv("Ring_Ulnar_loop.csv")
a8[,1]<-"R"
a8[,2]<-"L"
colnames(a8)[1]<- "FingerType"
colnames(a8)[2]<- "PatternType"

a9<-read.csv("Ring_whorl.csv")
a9[,1]<-"R"
a9[,2]<-"W"
colnames(a9)[1]<- "FingerType"
colnames(a9)[2]<- "PatternType"

a10<-read.csv("Thumb_arch.csv")
a10[,1]<-"T"
a10[,2]<-"A"
colnames(a10)[1]<- "FingerType"
colnames(a10)[2]<- "PatternType"

a11<-read.csv("Thumb_Ulnar_loop.csv")
a11[,1]<-"T"
a11[,2]<-"L"
colnames(a11)[1]<- "FingerType"
colnames(a11)[2]<- "PatternType"
```

```
a12<-read.csv("Thumb_whorl.csv")
a12[,1]<-"T"
a12[,2]<-"W"
colnames(a12)[1]<- "FingerType"
colnames(a12)[2]<- "PatternType"

### Step 3: Join the 12 datasets together ###

dbp<-smartbind(a1,a2,a3,a4,a5,a6,a7,a8,a9,a10,a11,a12, fill=0)

### Step 4: Randomly select 1 fingerprint to use to simulate the fingermark
###

z1<-seq(22,ncol(dbp)-4,by=5)
mnum<-sample(1:nrow(dbp), 1)
x<-as.numeric(dbp[mnum,z1+1])
y<-as.numeric(dbp[mnum,z1+2])
t<-as.numeric(dbp[mnum,z1+4])
b<-x[(x>0)|(y>0)]
d<-y[(x>0)|(y>0)]
e<-t[(x>0)|(y>0)]
mintype<-as.factor(e)
xdf<-data.frame(b,d)
qxdf<-as.matrix(dist(xdf))

### Step 5: Simulation of fingermark with 10 minutiae and sigma = 4 ###

mpoint<-sample(1:nrow(xdf), 1)
sortnn<-sort(qxdf[mpoint,])
nnset<-sample(sortnn[2:10],9, replace=FALSE)
ss<-rbind(xdf[mpoint,], xdf[which(qxdf[mpoint,]==nnset[1]),],
xdf[which(qxdf[mpoint,]==nnset[2]),], xdf[which(qxdf[mpoint,]==nnset[3]),],
xdf[which(qxdf[mpoint,]==nnset[4]),], xdf[which(qxdf[mpoint,]==nnset[5]),],
```

```
xdf[which(qxdf[mpoint,]==nnset[6]),], xdf[which(qxdf[mpoint,]==nnset[7]),],
xdf[which(qxdf[mpoint,]==nnset[8]),],  xdf[which(qxdf[mpoint,]==nnset[9]),])

mins<-rbind(mpoint, which(qxdf[mpoint,]==nnset[1]),
which(qxdf[mpoint,]==nnset[2]), which(qxdf[mpoint,]==nnset[3]),
which(qxdf[mpoint,]==nnset[4]), which(qxdf[mpoint,]==nnset[5]),
which(qxdf[mpoint,]==nnset[6]), which(qxdf[mpoint,]==nnset[7]),
which(qxdf[mpoint,]==nnset[8]),  which(qxdf[mpoint,]==nnset[9]))
tmarkmin<-mintype[mins]

jittermintype<-0
protyperetained<-rbinom(length(tmarkmin),1,0.9)

for (i in 1:length(tmarkmin)){
if (protyperetained[i]==1) jittermintype[i] = tmarkmin[i] else if
(tmarkmin[i]==1) jittermintype[i]=2 else jittermintype[i]=1
}

jitmark<-ss+ rnorm(2*nrow(ss), 0, 4 )
mark<-data.frame(jitmark)
q2<-as.matrix(dist(mark))

### Step 6: Sample 100 fingerprints from the rest of the fingerprints ###
### with the probability of picking each pattern type set using prior ###
### knowledge  ###

probmark=c(rep(0.1,nrow(a1)),rep(0.6,nrow(a2)),rep(0.3,nrow(a3)),
rep(0.1,nrow(a4)),rep(0.6,nrow(a5)),rep(0.3,nrow(a6)),
rep(0.1,nrow(a7)),rep(0.6,nrow(a8)),rep(0.3,nrow(a9)),
rep(0.1,nrow(a10)),rep(0.6,nrow(a11)),rep(0.3,nrow(a12)))
probmark[mnum]=0
pnum<-sample(1:nrow(dbp), 100, replace = FALSE, prob=probmark)

### Step 7: Set up empty matrices used during the method ###
```

```
newtransss<-matrix(0,1,0)
newtransall<-matrix(0,nrow(mark), 0)
empty20<-matrix(0, 1, 0)
empty21<-matrix(0, nrow(mark), 0)
pmt<-matrix(0, 1, 0)
pmt2<-matrix(0, nrow(mark), 0)

### Step 8: Matching algoirthm repeated separately to find the best ###
### match in all 100 fingerprints selected ###

for (p in 1:100){

### Step 9: Set up the data frame from the fingerprint information ###


z1<-seq(22,ncol(dbp)-4,by=5)
x1<-as.numeric(dbp[pnum[p],z1+1])
y1<-as.numeric(dbp[pnum[p],z1+2])
t1<-as.numeric(dbp[pnum[p],z1+4])
b1<-x1[(x1>0)|(y1>0)]
d1<-y1[(x1>0)|(y1>0)]
e1<-t1[(x1>0)|(y1>0)]
minuttype<-as.factor(e1)
xdf1<-data.frame(b1,d1)

### Step 10: Randomly select 3 minutiae to form a triangle and find a ###
### candidate set of similar triangles in the fingerprint ###

q1<- as.matrix(dist(xdf1))
nn1<-q2[1,2]
indnn1<-which(q2[1,] == nn1, arr.ind=T )
candset<-which( q1 < nn1+10 & q1> nn1-10, arr.ind=T )
nn2<-q2[1,3]
```

```
indnn2<-which(q2[1,] == nn2, arr.ind=T )
As<-candset[,1]
w1<-which(q1[,As]<nn2+10 & q1[,As]>nn2-10,arr.ind=T)
candset2<-cbind(As[w1[,2]], w1[,1])
BC<- q2[indnn1, indnn2]
Bs<-candset[,2]
w2<-which(q1[,Bs]<BC+10 & q1[,Bs]>BC-10,arr.ind=T)
candset3<-cbind(Bs[w2[,2]], w2[,1])
M=outer(candset[,2], candset3[,1], "==")
I<-row(M)[M]
J<-col(M)[M]
AB<-candset[I,]
C<-candset3[J,2]
empty<- cbind(AB,C)
newemp<-empty[,c(1,3)]
colab<-apply(newemp,1, paste, collapse="")
colcd<-apply(candset2,1, paste, collapse="")
newset<-unique(empty[which(colab%in% colcd),])

### Step 11: Find the transformation of the triangle in the mark to  ###
### each individual triangle in the candidate set and apply this ###
### transofrmation to the whole fingermark ###

empty5<-matrix(0,0,1)
for (q in 1:nrow(newset)){
ABC<-array(c(mark[1,1], mark[indnn1,1],mark[indnn2,1],mark[1,2],
mark[indnn1,2],mark[indnn2,2]), c(3,2,1))
orig<-array(c(xdf1[newset[q,1],1], xdf1[newset[q,2],1],
xdf1[newset[q,3],1],xdf1[newset[q,1],2], xdf1[newset[q,2],2],
xdf1[newset[q,3],2]), c(3,2,1))
abcprint<-t(t((((t(t(mark)+transformations(orig, ABC)$translation[,1]))
*transformations(orig, ABC)$scale[1])%*%transformations(orig, ABC)
$rotation[,,1]))+ colMeans(orig[,,1]))
for (r in 1:nrow(mark)){
```

```
empty5<-rbind(empty5, abcprint[r,1])
}
for (k in 1:nrow(mark)){
empty5<-rbind(empty5, abcprint[k,2])
}
}
transmark<-array(empty5 , c(nrow(mark),2,nrow(newset)))

### Step 12: For each transformed mark (for each case in the candidate ###
### set) run the Hungarian Algorithm to find the unique allocation of ###
### minutiae in the mark to minutiae in the print ###

distance<-0
empty11<-matrix(0,nrow(mark),0)
for(s in 1:nrow(newset)){
testmark<-transmark[,,s]
empty10<-outer(testmark[,1], xdf1[,1],"-")^2 + outer(testmark[,2],
 xdf1[,2],"-")^2
solution<-solve_LSAP(empty10)
empty11<- cbind(empty11, solution)
}

### Step 13: Calculate the sum of squared distance between the fingermark ###
### and fingerprint based on this unique allocation for every case in ###
### the candidate set ###

di<-0
empty13<-matrix(0,nrow(mark),0)
for (v in 1:ncol(empty11)){
testmark<-transmark[,,v]
for(w in 1:nrow(testmark)){
di[w]<-(testmark[w,1]-xdf1[empty11[w,v],1])^2+(testmark[w,2]-
xdf1[empty11[w,v],2])^2
}
```

```
empty13<-cbind(empty13, di)
}
sumempty13<-colSums(empty13)

### Step 14: Select the member of the candidate set with the lowest ###
### sum of squared distance since this represents the best match for ###
### that fingerprint - repeat for other 99 fingerprints ###

empty20<- cbind(empty20, min(colSums(empty13)))
empty21<- cbind(empty21, empty11[,which.min(colSums(empty13))])
}

### Step 15: Repeat the above process for the fingerprint that the ###
### fingermark was simulated from ###

z1<-seq(22,ncol(dbp)-4,by=5)
x1<-as.numeric(dbp[mnum,z1+1])
y1<-as.numeric(dbp[mnum,z1+2])
t1<-as.numeric(dbp[mnum,z1+4])
b1<-x1[(x1>0)|(y1>0)]
d1<-y1[(x1>0)|(y1>0)]
e1<-t1[(x1>0)|(y1>0)]
minuttype<-as.factor(e1)
xdf1<-data.frame(b1,d1)
q1<- as.matrix(dist(xdf1))
nn1<-q2[1,2]
indnn1<-which(q2[1,] == nn1, arr.ind=T )
candset<-which( q1 < nn1+10 & q1> nn1-10, arr.ind=T )
nn2<-q2[1,3]
indnn2<-which(q2[1,] == nn2, arr.ind=T )
As<-candset[,1]
w1<-which(q1[,As]<nn2+10 & q1[,As]>nn2-10,arr.ind=T)
candset2<-cbind(As[w1[,2]], w1[,1])
BC<- q2[indnn1, indnn2]
```

```
Bs<-candset[,2]
w2<-which(q1[,Bs]<BC+10 & q1[,Bs]>BC-10,arr.ind=T)
candset3<-cbind(Bs[w2[,2]], w2[,1])
M=outer(candset[,2], candset3[,1], "==")
I<-row(M)[M]
J<-col(M)[M]
AB<-candset[I,]
C<-candset3[J,2]
empty<- cbind(AB,C)
newemp<-empty[,c(1,3)]
colab<-apply(newemp,1, paste, collapse="")
colcd<-apply(candset2,1, paste, collapse="")
newset<-unique(empty[which(colab%in% colcd),])

empty5<-matrix(0,0,1)
system.time(for (q in 1:nrow(newset)){
ABC<-array(c(mark[1,1], mark[indnn1,1],mark[indnn2,1],mark[1,2],
mark[indnn1,2],mark[indnn2,2]), c(3,2,1))
orig<-array(c(xdf1[newset[q,1],1], xdf1[newset[q,2],1],
xdf1[newset[q,3],1],xdf1[newset[q,1],2], xdf1[newset[q,2],2],
xdf1[newset[q,3],2]), c(3,2,1))
abcprint<-t(t((((t(t(mark)+transformations(orig, ABC)
$translation[,1]))*transformations(orig, ABC)$scale[1])%*%
transformations(orig, ABC)$rotation[,,1]))+ colMeans(orig[,,1]))
for (r in 1:nrow(mark)){
empty5<-rbind(empty5, abcprint[r,1])
}
for (k in 1:nrow(mark)){
empty5<-rbind(empty5, abcprint[k,2])
}
})
transmark<-array(empty5 , c(nrow(mark),2,nrow(newset)))

distance<-0
```

```
empty11<-matrix(0,nrow(mark),0)
system.time(for(s in 1:nrow(newset)){
testmark<-transmark[,,s]
empty10<-outer(testmark[,1], xdf1[,1],"-")^2 + outer(testmark[,2],
xdf1[,2],"-")^2
solution<-solve_LSAP(empty10)
empty11<- cbind(empty11, solution)
})

di<-0
empty13<-matrix(0,nrow(mark),0)
system.time(for (v in 1:ncol(empty11)){
testmark<-transmark[,,v]
for(w in 1:nrow(testmark)){
di[w]<-(testmark[w,1]-xdf1[empty11[w,v],1])^2+(testmark[w,2]-
xdf1[empty11[w,v],2])^2
}
empty13<-cbind(empty13, di)
})

empty20<- cbind(empty20, min(colSums(empty13)))
empty21<- cbind(empty21, empty11[,which.min(colSums(empty13))])

### Step 16: Calculate the likelihood ratio that the fingermark came ###
### from the original fingerprint compared to the other 100 ###

m<- nrow(ss)
n<- 2*m
alpha<-0.1
beta<-0.1
A<- alpha + n/2

Phpy<- (beta + (empty20[101])/2)^-A
Phdy<- sum((beta + (empty20[-101])/2)^-A)
```

```
LR<- Phpy / Phdy
```

Appendix B

R code - Additional Optimisation Steps

```
### Step 1: Find the transformation from the whole fingermark to its ###
### unique allocation in the fingerprint and apply this to the whole ###
### mark. Calculate the squared distance for each point in the ###
### transformed mark to the corresponding point in the print ###

indivssoptim<-matrix(0,nrow(mark),0)
for (ij in 1:ncol(empty11)){
allocation<-empty11[,ij]
bestmatch<-array(transmark[,,ij], c(nrow(mark),2,1))
bestmatchmat<-matrix(bestmatch, ncol=2)
trueloc<-array(c(xdf1[allocation,1], xdf1[allocation,2]), c(nrow(mark),2,1))
secondtrans<-t(t((((t(t(bestmatchmat)+transformations(trueloc, bestmatch)
$translation[,1]))*transformations(trueloc, bestmatch)$scale[1])%*%
transformations(trueloc, bestmatch)$rotation[,,1]))+ colMeans(trueloc[,,1]))
newtransssind<-(secondtrans[,1]-xdf1[allocation,1])^2+(secondtrans[,2]
-xdf1[allocation,2])^2
indivssoptim<-cbind(indivssoptim, newtransssind)
}
```

```
### Step 2: Calculate the sum of squared distance for each case in the ###
### candidate set and select the lowest sum of squared distance ###
### between the transformed fingermark and the fingerprint ###

sumindivssoptim<-colSums(indivssoptim)
newtransss<-cbind(newtransss, min(colSums(indivssoptim)))
newtransall<-cbind(newtransall, empty11[,which.min(colSums(indivssoptim))])

### Step 3: Calculate the likelihood ratio that the fingermark came ###
### from the original fingerprint compared to the other 100 ###

NewPhpy<- (beta + (newtransss[101])/2)^-A
NewPhdy<- sum((beta + (newtransss[-101])/2)^-A)
NewLR<- NewPhpy / NewPhdy
```

Appendix C

R code - Extension to Minutia Type Steps

```
### Step 1: Calculate the number of times the n minutiae in the ###
### fingermark have the same type as the corresponding ###
### minutiae in the fingerprint (for every case in the candidate ###
### set). Create a multiplicative factor based on this, here the ###
### penalty for none matching type is set to 1.2 ###

newfact<-0
for (xyz in 1:ncol(empty11)){
allocationmin<-empty11[,xyz]
alloctype<-minuttype[allocationmin]
yes<- sum(jittermintype==alloctype)
no<- nrow(q2)-yes
multfactor<-1^yes*1.2^no
newfact<-cbind(newfact, multfactor)
}
newfact<-newfact[-1]

### Step 2: Multiply the sum of squared distance for each case in the ###
```

```
### candidate set by the corresponding penalty factor. Select the ###
### lowest sum of squared distance between the transformed ###
### fingermark and the fingerprint ###

postmintype<-sumindivssoptim*newfact
pmt<- cbind(pmt, min(postmintype))
pmt2<- cbind(pmt2, empty11[,which.min(postmintype)])

### Step 3: Calculate the likelihood ratio that the fingermark came ###
### from the original fingerprint compared to the other 100 ###

TypePhpy<- (beta + (pmt[101])/2)^-A
TypePhdy<- sum((beta + (pmt[-101])/2)^-A)
TypeLR<- TypePhpy / TypePhdy
```