MACHINE LEARNING IN FINGERPRINT PROBABILITY EVALUATION

by

Chang Su

August 2011

A dissertation submitted to the

Faculty of the Graduate School of

the University at Buffalo, State University of New York

in partial fulfillment of the requirements for the

degree of

Doctor of Philosophy

Department of Computer Science and Enigineering

ABSTRACT

Although fingerprints have been used in forensic identification for over a century, establishing the degree of uniqueness a given fingerprint has remained elusive since it involves relationships and correlations hidden in a large amount of fingerprint data. We develop machine learning approaches to this problem focusing on two aspects: better generative models for fingerprints, and estimating probabilistic metrics for individuality and rarity. Generative approaches are used to evaluate fingerprint Individuality, defined as the probability of random correspondence (PRC) within a tolerance. The evaluation uses Bayesian networks and three fingerprint representations: ridge flow, minutiae, and ridge points. Mixture models of features are used, with parameters estimated using the EM algorithm. Fingerprint PRCs are determined for given numbers of available and matching minutiae. A Bayesian method based on graphical models is introduced to compute fingerprint rarity. The model is able to handle the large number of variables and the complexity of the distributions. Since latent prints are often incomplete and the core point may be missing in the field of view, a machine learning approach based on Gaussian process regression is proposed to predicate the core point. The coordinate system is transformed into standard position based on finding the core point. A graphical model, which takes into account inter-minutia dependencies and minutia

confidences, is used to determine evidence probability from which the specific probability of match among n is evaluated. To improve the accuracy of rarity estimation for low quality latent prints, a fully Bayesian treatment allows considering core point uncertainty. The generative model is validated using a goodness-of-fit test. Rarity evaluation is illustrated using several examples, including simple configurations of minutiae, randomly selected latent fingerprints in a database, and a well-known case of erroneous identification.

Contents

List of Tables

List of Figures

Chapter 1

INTRODUCTION

This dissertation explores the roles of machine learning in fingerprint individuality and rarity evaluation. Although fingerprints have been used in forensic identification for over a hundred years, the uniqueness for each individual fingerprint is being challenged. A scientific basis on individuality and rarity is required for fingerprint evidence presented in courtroom scenarios. Capturing the high complexity of relationships and correlations hidden among large amount of fingerprint data is the main difficulty in fingerprint probability evaluation. To establish the strength of the fingerprint modality or a given piece of fingerprint evidence, this research focuses on using machine learning approaches to formulate and compute probabilistic metrics for measuring fingerprint individuality and rarity. This introduction describes the motivation, contribution, and the structure of this study.

1.1 Background and Motivation

Machine learning provides strong tools to the tasks that cannot be defined well except by examples or the amount of knowledge available might be too large for explicit encoding by humans. This section provides brief background and motivation for using machine learning approaches to evaluate fingerprint probabilities.

1.1.1 Machine Learning for Forensics

Recently, some high-profile cases such as the 1999 lawsuit of USA vs Mitchell and the 2004 Mayfield case grab the headlines and our attention. The erroneous identification in these cases points to an underlying problem with classical forensic methodology. The fact is that many forensic tests have never been exposed to stringent scientific scrutiny. One of the main problems with forensics evidence is that it must be analyzed and interpreted by a person, whose own theory of the crime can introduce a bias in the results. There can also be significant uncertainty in the analyst's conclusions, but oftentimes that uncertainty is never quantified or conveyed to judges and juries. And yet, these traditional forms of forensics evidence can be very helpful, provided they can be looked at objectively and the uncertainty of the results can be measured and properly explained. The relatively new field of computational forensics has sprung up to address those needs.

Computational forensics is an emerging interdisciplinary research domain. It is understood as the hypothesis-driven investigation of a specific forensic problem using computers, with the primary goal of discovery and advancement of forensic knowledge. Computational methods find a place in the forensic sciences in three ways. First, they provide tools for the

human examiner to better analyze evidence by overcoming limitations of human cognitive ability, thus they can support the forensic examiners in their daily casework. Secondly they can be used to provide the scientific basis for a forensic discipline or procedure by providing for the analysis of large volumes of data which are not humanly possible. Thirdly they can ultimately be used to represent human expert knowledge and for implementing recognition and reasoning abilities in machines. It involves modeling and computer simulation and/or computer- based analysis and recognition in studying and solving forensic problems. Therefore, much of computational algorithms used to assistant forensic methods are dominated by statistically based algorithms such as machine learning algorithms. These methods are ideally suited to forensics where there is a need to demonstrate error rates and calculate probabilities.

Machine learning, a branch of artificial intelligence, is a scientific discipline concerned with the design and development of algorithms that allow computers to capturing the patterns or evolve behaviors based on large amount of empirical data. A learner can take advantage of examples (data) to capture characteristics of interest of their unknown underlying probability distribution. Machine learning provides fundamental techniques and have been successfully implemented to supporting forensic investigation. The objective of this research is to explores the roles of machine learning to analyze the evidence, assist in the interpretation of identification results, and reveal previously unknown patterns/links and new rules for one of the most widely used forensic evidences: fingerprints.

1.1.2 Probabilities of Fingerprints

In forensics there are several modalities for identifying an object or individual from evidence. It is useful to establish the strength of a given modality or a given piece of evidence within a given modality. Not least of the reasons for such analysis being court rulings that require a scientific basis for evidence presented [65]. As one of the most wildly used evidences, fingerprints collected at a crime scene or on items of evidence from a crime have been used in forensic science to identify suspects, victims and other persons who touched a surface. The use of fingerprints in human identification has been based on two premises, that (i) they do not change with time, and (ii) they are unique for each individual. Though the first premise has been accepted, the second one on individuality is being challenged under the basis that the premises have not been objectively tested and error rates established. To address this problem, we develop machine learning approaches to compute two important probabilities: fingerprint individuality and fingerprint rarity.

The terms class characterization and individualization are commonly used in forensics. In addition terminology from the biometric domain, such as verification and identification are also present. Thus it is useful to first define these terms. Class-characterization is the narrowing down of the evidence into a sub-class within the forensic modality, e.g., ethnicity. Individualization is sometimes defined as the exclusion of all other sources for the given evidence. Verification is the determination of whether a given evidence is from a given source and is a binary decision. Identification is the determination of the best match of the evidence given a finite set of sources for that evidence. A match here does not necessarily mean an exact match but a match within given tolerance levels. Finally individuality of a

forensic modality or of a piece of forensic evidence is the degree of distinctiveness of that type of evidence in a population.

Fingerprint individuality studies date back to the late 1800s. More than twenty models have been proposed to establish the improbability of two random people (or fingers) have the same fingerprint [53]. These models can be classified into four different categories: grid-based [18, 40], ridge-based [40], fixed probability [20], and relative measurement [63]. All models try to quantify the uniqueness property of the fingerprints. In the modality based on visual patterns such as fingerprints, the set of measurements made from the evidence have an inherent variability even for the same object or individual. Since evidence can be characterized by quantitative measurements [1], a natural metric for individuality is a probability or a probability distribution. Such a metric can be evaluated, for a particular modality, measurement (or feature vector) or a given sample of data, by using one of the classical approaches of machine learning: generative approach[25, 6]. In generative approach, representative samples of evidence are used to construct models during a training phase. Generative models are statistical models that represent the distribution of $\mathbf{x}$ [35, 15, 54]. They are referred to as being generative in that given the distribution, samples can be generated from them. In these models, a distribution of $\mathbf{x}$ is learned through a training data set. Samples can be generated from this distribution to determine the probability of random correspondence. The training set used is immaterial as long as it is representative of the entire population. In contrast with other models, the generative model is a full probability model of all variables. A generative model can be used, for example, to simulate (i.e. generate) values of any variable in the model. If the observed data are truly sampled from the generative model, then fitting the parameters of the generative model to maximize

the data likelihood is a common method.

Fingerprint rarity is determined by the probability of finding a random match for the evidence in a database of n prints. Knowing the rarity of a given fingerprint could be useful to forensic scientists who are trying to determine how valuable the fingerprint is as evidence. We just found something that is unusual, and that makes it an important piece of evidence. Rarity is especially valuable for latent fingerprints. Latent fingerprints are crime scene prints that are usually incomplete patterns taken off doorknobs or glass. The uniqueness of latent prints usually varies dramatically due to the different amounts and quality of features the latent prints contain. Yet methods to compute rarity of features have been elusive due to the large number of variables and the complexity of the distributions. This dissertation, for the first time, presents a systemic approach to calculate the rarity for any given fingerprints.

1.2 Contribution of the Work

We proposed generative models for forensic evidence where the goal is to describe the distributions using graphical models and to use such models to compute probabilistic metrics for measuring the degree of individuality of a forensic modality such as fingerprints. The metric is defined as the probability of random correspondence (PRC) when evidence consists of a set of measurements and correspondence is within a tolerance. Computation of these probabilities is described using Bayesian networks which make all the variables explicit. Three forms of fingerprint representation are considered: ridge flow, minutiae, and ridge points. Mixture models are used to model location and orientation of minutiae. With parameters estimated from standard databases, fingerprint PRCs are determined for given numbers of

available and matching minutiae.

A method for computing the rarity of any given fingerprints represented by minutiae is proposed. It allows determining the probability of finding a random match for the evidence in a database of n prints. First, the coordinate system is transformed into standard position based on finding the core point. Since latent prints are often incomplete and the core point may be missing in the field of view, a machine learning approach based on Gaussian process regression is proposed. Next a generative model, that takes into account inter-minutia dependencies and minutia confidences, is used to determine evidence probability from which the specific probability of match among n is evaluated. The generative model is validated using a goodness-of-fit test. Rarity evaluation is illustrated using several examples, including simple configurations of minutiae, randomly selected latent fingerprints in a database, and a well-known case of erroneous identification.

1.3 Outline of Dissertation

The theme of this dissertation is using machine learning algorithm for fingerprint probability evaulation. We worked on two aspects of this problem: generative models and probabilistic metrics for fingerprint individuality, Graphical models and Bayesian method for fingerprint rarity. Naturally, these two pieces of work are described in Chapter 2 and Chapter 3, one in each chapter.

Chapter 2 discusses the issues on evaluating fingerprint individuality. Probability of random correspondence is defined to measure the individuality. Generative models of fingerprints are presented by modeling ridge flow type, minutiae, and representative ridges. Using

generative approach, we describe the PRC evaluation based on graphical models.

Chapter 3 introduces and solves several problems in fingerprint rarity evaluation. A Gaussian processes based core point prediction algorithm is presented to align coordinates and normalize features. Bayesian networks are applied to capture the relation and correlation among the fingeprints. A fully Baysien method is formulated to compute the specific nPRC, a metrics for rarity.

Due to the independence between these two different aspects, each chapter is self-contained with its own introduction and conclusions. Readers my read the chapters of interest in any order.

Chapter 2

FINGERPRINT INDIVIDUALITY

The use of fingerprints in human identification has been based on two premises, that, (i) they do not change with time and (ii) they are unique for each individual. While in the past, identification based on fingerprints had been accepted by courts, more recently their use has been questioned under the basis that the premises stated above have not been objectively tested and error rates not been scientifically established [64]. Though the first premise has been accepted, the second one on individuality is being challenged. We consider generative model approach for forensic evidence where the goal is to describe the distributions using graphical models and to use such models to compute probabilistic metrics for measuring the degree of individuality of a forensic modality such as fingerprints. The metric is defined as the probability of random correspondence (PRC) when evidence consists of a set of measurements and correspondence is within a tolerance. Computation of these probabilities is described using Bayesian networks which make all the variables explicit. Three forms of fingerprint representation are considered: ridge flow, minutiae, and ridge points. Mixture models are used to model location and orientation of minutiae. With parameters estimated

from standard databases, fingerprint PRCs are determined for given numbers of available and matching minutiae.

2.1 Introduction

The terms class characterization and individualization are commonly used in forensics. In addition terminology from the biometric domain, such as verification and identification are also present. Thus it is useful to first define these terms. Class-characterization is the narrowing down of the evidence into a sub-class within the forensic modality, e.g., ethnicity. Individualization is sometimes defined as the exclusion of all other sources for the given evidence. Verification is the determination of whether a given evidence is from a given source and is a binary decision. Identification is the determination of the best match of the evidence given a finite set of sources for that evidence. Finally individuality of a forensic modality is the degree of distinctiveness of that type of evidence in a population. In this chapter, we look into the issues on fingerprint individuality.

2.1.1 Discriminative and Generative Methods

In many modalities, particularly those based on visual patterns, the set of measurements made from the evidence have an inherent variability even for the same object or individual. Since evidence in many forensic modalities can be characterized by quantitative measurements [1], a natural metric for individuality is a probability or a probability distribution. Such a metric can be evaluated, for a particular modality or measurement(or feature vector),

by using either one of the classical approaches of machine learning: discriminative and generative [25, 6]. In both approaches, representative samples of evidence are used to construct models during a training phase.

In the discriminative approach to measuring individuality, samples are directly used to construct either a two-class or a multi-class classifier [47]. One such approach is that based on determining a similarity (or kernel) function s whose value is high when the input and a template have the same origin and low when they are not. Such a method is used, for instance, in automated fingerprint identification systems that determine the degree of *match* between two fingerprints [68]. By thresholding the value of s into binary classes: *same* and *different*, an accuracy measure can be estimated from known training samples, e.g., average probability of error or risk. The accuracy estimate itself provides a measure of individuality of measurement $\mathbf{x}$. Since the nature of the testing set is crucial, data from cohort groups such as twins are often used to determine error rates [49, 48].

Generative models are statistical models that represent the distribution of $\mathbf{x}$. They are referred to as being generative in that given the distribution, samples can be generated from them. In these models, a distribution of $\mathbf{x}$ is learnt through a training data set. Samples can be generated from this distribution to determine the probability of random correspondence. The training set used is immaterial as long as it is representative of the entire population.

Generative models contrast with discriminative models, in that a generative model is a full probability model of all variables, whereas a discriminative model provides a model only of the target variable(s) conditioned on the observed variables. A generative model can be used, for example, to simulate (i.e. generate) values of any variable in the model, whereas a discriminative model allows only sampling of the target variables conditioned on

the observed quantities. If the observed data are truly sampled from the generative model, then fitting the parameters of the generative model to maximize the data likelihood is a common method.

Both approaches have their advantage and limitations. In the generative approach, the problems with realistic modeling of all the parameters of the measurement may become unsurmountable. But a good model would be a method of gaining insights into the fundamental accuracy bounds of the measurement; more typically, these models may be relatively less effective in predicting performance. In the discriminative approach, definition of decision thresholds based on minimizing risk is an issue [16]. Discriminative methods lead to higher performing automated systems, but do not readily lead to fundamental understanding of underlying issues [8].

2.1.2 Organization of Rest of Chapter

The focus of this chapter is generative models for fingerprints and defines probabilistic metrics that are useful for a given modality or given evidence within a modality. We introduce the related work in Section 2.2. Section 2.3 describes the individuality measurements. The metric is defined as the probability of random correspondence (PRC) when evidence consists of a set of measurements and correspondence is within a tolerance. Computation of these probabilities is described using Bayesian networks which make all the variables explicit. Generative models for fingerprints are discussed in Section 2.4. Three forms of fingerprint representation are considered: ridge flow, minutiae, and ridge points. Mixture models are

used to model location and orientation of minutiae. With parameters estimated from standard databases, fingerprint PRCs are determined for given numbers of available and matching minutiae. In Section 2.6, we present experimental results on standard fingerprint dataset. We conclude in Section 2.7.

2.2 Related Work

Fingerprint individuality studies date back to the late 1800s. Much of it has focused on demonstrating or asserting the mere fact of the absolute non-duplication of complete fingertip-sized areas of friction ridge skin. This is true of both of the two major strands of fingerprint research. Anatomical research focused on detailing the formation of friction ridge skin, while occasionally commenting that this process was sufficiently complex to support an assumption of non-duplication as a working principle [70]. Statistical research focused on estimating the probability that exact duplicate areas of friction ridge skin (usually complete fingertips) exit. All models try to quantify the uniqueness property, eg. the probability of false correspondence. These models can be classified into three categories: fixed probability [20], relative measurement [63], and most recently, generative models. Generative models are used to learn the distribution of fingerprint features such as minutiae. A couple of generative models have been proposed. Pankanti etal. [35] modeled the minutiae as uniformly and independently distributed. Zhu etal. [71] proposed a mixture model to account for the clustering tendency of minutiae. We proposes generative models that incorporate ridge flow types, minutiae, and ridge information.

2.3 Probability of Random Correspondence

Generative approaches to machine learning involve describing the underlying probability distributions and then using such models to compute probabilities. We consider generative models for forensic evidence where the goal is to describe the distributions using graphical models and to use such models to compute probabilistic metrics for measuring the degree of individuality of a forensic modality such as fingerprints. The process involves the following steps:

Step 1: Consider a generative model (proposal) for measurement $\mathbf{x}$

Step 2: Formulate a method for estimating parameters of the model

Step 3: Evaluate the parameters from a data set

Step 4: Use the model to evaluate relevant individuality metrics.

The first four steps are to determine the distribution of the data. The final step involves computing the probability of match or correspondence within some tolerance between two samples. Tolerance specification depends on the modality and type of measurement, e.g., in the case of a continuous scalar it can be specified as $\pm\epsilon$.

As measures of individuality, two probabilities can be defined: PRC (Probability of Random Correspondence), nPRC (Probability of Random Correspondence given n samples). These definitions are further described below.

1. PRC: probability that two randomly chosen samples have the same measured value $\mathbf{x}$ within specified tolerance ϵ.

2. nPRC: the probability that among a set of n samples, some pair have the same value $\mathbf{x}$, within specified tolerance, where $n \geq 2$. Since there are $\binom{n}{2}$ pairs involved this

probability is higher than PRC. Note that when $n = 2$, PRC=nPRC.

We note that PRC and nPRC characterize the forensic modality as described by a set of measurements and furthermore, the second is a function of the first.

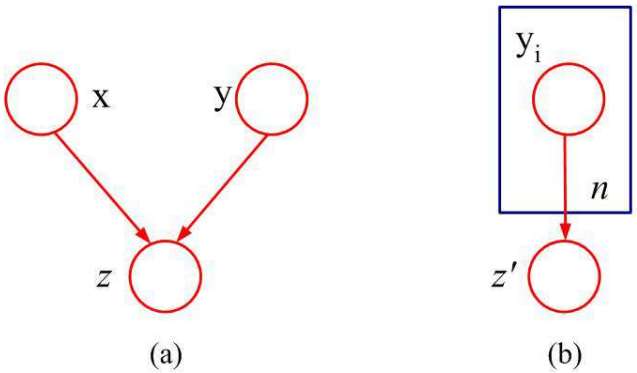


Figure 2.1: Graphical models for computing coincidence probabilities. The models correspond to: (a) PRC, the probability of two samples having the same value, (b) nPRC, the probability of at least two samples among n having the same value. Here $\mathbf{x}$, $\mathbf{y}$, $\mathbf{y}_s$ and $\mathbf{y}_i$ are feature vectors with identical distribution. We are interested in the distributions of z and z' which express probabilities of match/non-match.

Since probability distributions and their computation play a central role in this discussion, it is useful to represent them using graphical models [6]. Directed graphical models for PRC, nPRC and specific nPRC are shown in Figure 2.1. Each node represents a random variable (or a group of random variables), and the links express conditional probabilistic relationships between them. The variables are: $\mathbf{x}$ is a random variable corresponding to the feature vector, $\mathbf{y}$ is another random variable with the same distribution, $\{\mathbf{y}_i\}$ where $i = [1, .., n]$ represents a set of n random variables using the plate representation and $\mathbf{y}_s$ where $s \in [1, .., n]$ is a random variable among $\{\mathbf{y}_i\}$.

A binary-valued random variable z in Figure 2.1(a) indicates that if $\mathbf{x}$ is the same as $\mathbf{y}$ within a tolerance ϵ, in which case $z = 1$ and $z = 0$ otherwise. This leads to the following definition

$$p(z|\mathbf{x},\mathbf{y}) = \begin{cases} p(z=1|\mathbf{x},\mathbf{y}) = 1 \text{ and } p(z=0|\mathbf{x},\mathbf{y}) = 0 & \text{if } \mathbf{x} = \mathbf{y} \\ p(z=1|\mathbf{x},\mathbf{y}) = 0 \text{ and } p(z=0|\mathbf{x},\mathbf{y}) = 1 & \text{if } \mathbf{x} \neq \mathbf{y} \end{cases} \tag{2.1}$$

By marginalizing the joint distribution as

$$p(z) = \sum_{\mathbf{x}} \sum_{\mathbf{y}} p(\mathbf{x},\mathbf{y},z), \tag{2.2}$$

applying Bayes rule to write

$$p(z) = \sum_{\mathbf{x}} \sum_{\mathbf{y}} p(z|\mathbf{x},\mathbf{y}) p(\mathbf{x},\mathbf{y}), \tag{2.3}$$

and noting the independence of $\mathbf{x}$ and $\mathbf{y}$

$$p(z) = \sum_{\mathbf{x}} \sum_{\mathbf{y}} p(z|\mathbf{x},\mathbf{y}) p(\mathbf{x}) p(\mathbf{y}), \tag{2.4}$$

we obtain the PRC as $\rho \equiv p(z=1)$.

Next we consider the case of nPRC where there are n identically distributed random variables $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, .., \mathbf{y}_n]$. This is shown using the plate representation in Figure 2.1(b), where z' is the state indicating if in a set of n random variables at least one value $\mathbf{y}_i$ is the same as another value $\mathbf{y}_j$ in which case $z' = 1$ and $z' = 0$ otherwise. This leads to the following definition

$$p(z'|\mathbf{Y}) = \begin{cases} p(z'=1|\mathbf{Y}) = 1 \text{ and } p(z'=0|\mathbf{Y}) = 0 & \text{if } \exists \mathbf{y}_i, \mathbf{y}_j \in \mathbf{Y} \ni \mathbf{y}_i = \mathbf{y}_j \\ p(z'=1|\mathbf{Y}) = 0 \text{ and } p(z'=0|\mathbf{Y}) = 1 & \text{if } \forall \mathbf{y}_i, \mathbf{y}_j \in \mathbf{Y} \ni \mathbf{y}_i \neq \mathbf{y}_j \end{cases} \tag{2.5}$$

The marginal distribution of z' is obtained as

$$p(z') = \sum_{\mathbf{Y}} p(z'|\mathbf{Y})p(\mathbf{Y}). \tag{2.6}$$

The specific instance of $p(z' = 1)$, which is the nPRC, can be written as

$$p(z' = 1) = 1 - p(z' = 0) = 1 - [1 - \sum_{\mathbf{x}} \sum_{\mathbf{y}} p(z = 1|\mathbf{x}, \mathbf{y})p(\mathbf{x})p(\mathbf{y})]^{\frac{n(n-1)}{2}} \tag{2.7}$$

Denoting the nPRC value as $\rho[n] \equiv p(z' = 1)$ and using (2.4), we have a relationship between PRC and nPRC as

$$\rho[n] = 1 - (1 - \rho)^{\frac{n(n-1)}{2}} \tag{2.8}$$

The rest of this paper concerns the application of the developed methods to three cases: birthdays, human heights and fingerprints represented by minutiae and ridges. In the case of birthdays, which are discrete-valued, the tolerance is zero, i.e., requiring exact match. In the case of heights, which are continuous-valued scalars, a tolerance is specified in terms of the height differences that are considered to be the same. In the case of fingerprints, the measurement is a variable-length vector with each element represented as a triple, and tolerance is specified within the fingerprint matching algorithm.

2.4 Generative Models for Fingerprints

Our goal is to model the distribution of fingerprints based on features. Features for representing fingerprints are classified into three types [3]. Level 1 features provide class-characterization of fingerprints based on ridge flow. They are divided into five primary

classes: whorl, left loop, right loop, arch and tent (Figure 2.2). Some of the primary classes have secondary classes resulting in a total of eight class types.

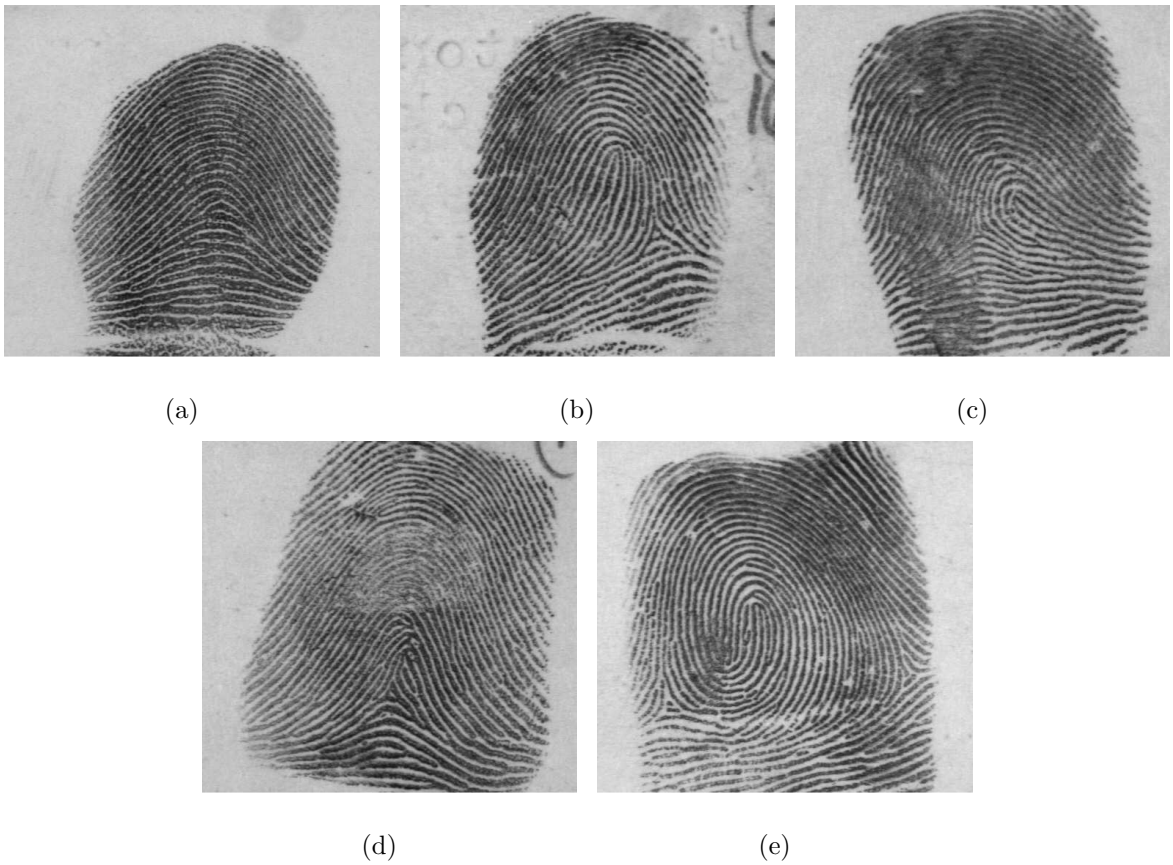
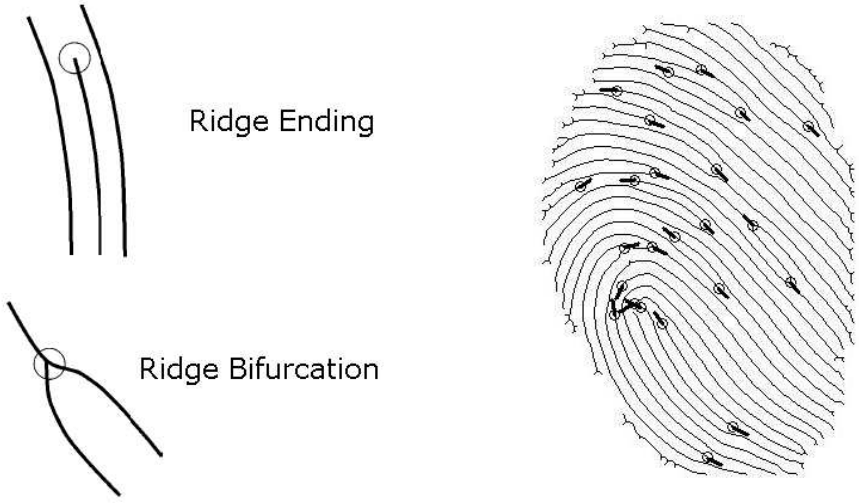


Figure 2.2: Examples of five main types of ridge flow in fingerprints, referred to as Level 1 features: (a) arch, (b) left loop, (c) right loop. (d) tented arch, and (e) whorl. From NIST Special Database 4.

Level 2 features, which are more useful for identification, are also known as minutiae. Fingerprints such as those shown in Figure 2.2 are first aligned. This is done manually where core points are identified and then the image is centered. The minutiae correspond to ridge endings and ridge bifurcations. Automatic fingerprint matching algorithms use minutiae as the salient features, e.g., [68], since they are stable and are reliably extracted. A minutia is represented by its location and direction; direction is determined by the ridge ending at the location (Figure 2.3). The type of minutiae (either bifurcation or ending) is not distinguished



(a) (b)

Figure 2.3: Representation of fingerprints using minutiae: (a) locations of ridge endings and ridge bifurcations are indicated by circles, and (b) minutiae directions are indicated by line segments in a skeletonized fingerprint image.

since this information is not as reliable as the information on location and direction. Level 3 features, such as pores and scars are ancillary features and seldom used in practice.

In this research, we consider the distributions for ridge flow types (Section 2.4.1), minutiae only (Section 2.4.2) and a combination– one based on minutiae and representative ridges (Section 2.4.3). Methods for estimating distribution parameters are discussed in Section 2.4.4.

2.4.1 Distribution of Ridge Flow Type

A simple distribution of the Level 1 ridge flow types is obtained by counting the relative frequency of each of the primary and secondary types in a fingerprint database. In one such evaluation [48] loops account for 64% of the fingers, with the secondary types being: 30% left loops, 27% right loops and 7% double loops. Arches account for 18% of the primary types, with the seondary types being: plain arches (13%) and tented arches (5%). Whorls

Table 2.1: Ridge Flow Types: nPRC

n	2	3	4	5	6	7	8
nPRC	0.2233	0.5314	0.7805	0.9201	0.9774	1	1

account for the remainder of the Level 1 types (19%).

Level 1 features are clearly broad class characteristics which are useful for exclusion of individual fingers but not by themselves useful for the tasks of verification, identification and individualization.

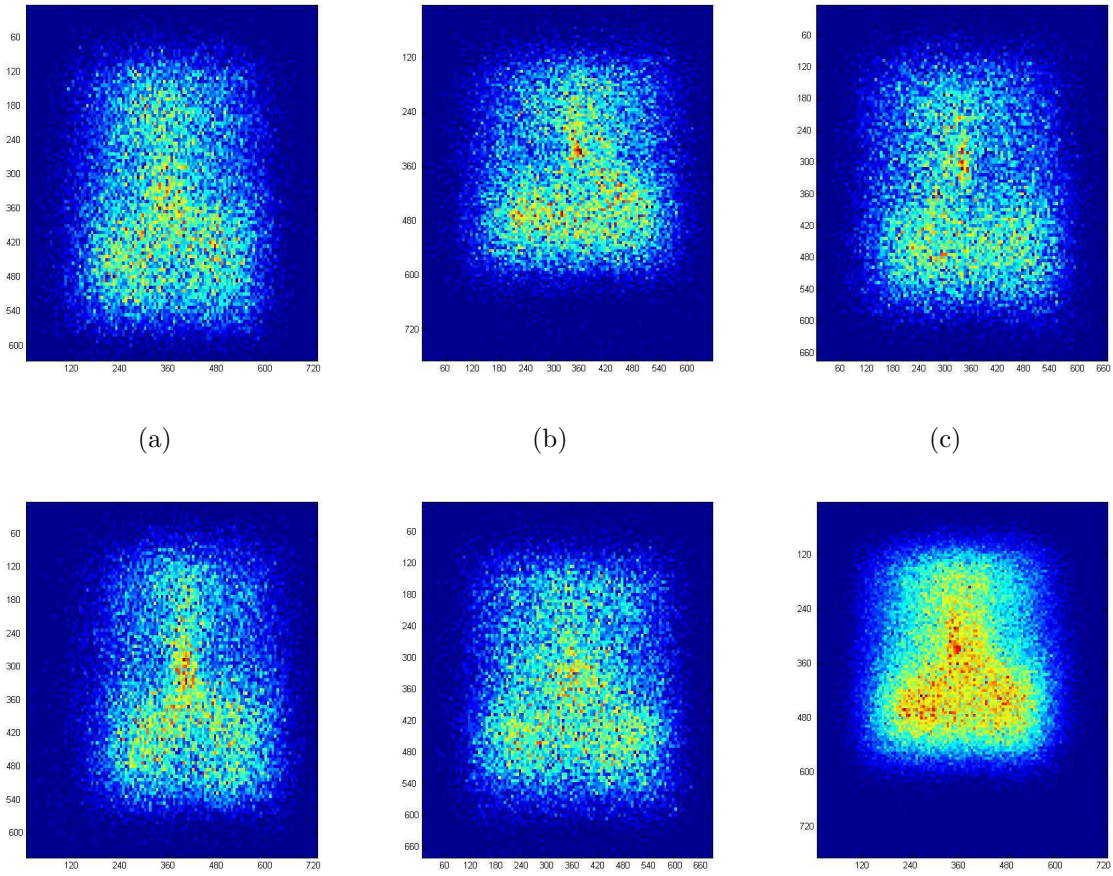
Assuming that fingerprints are distinguished by 6 secondary types, the PRC for ridge flow types is calculated by Eq.2.3. Given above type frequencies, we have PRC value $p_\epsilon = 0.2233$. The nPRC and specific nPRC can be caculated by Eq.2.8 and Eq.**??**. Table 2.1 and table **??** show the nPRC and Specific nPRC with different n.

2.4.2 Distribution of Minutiae

Each minutia is represented as $\mathbf{x} = (s, \theta)$ where $s = (x_1, x_2)$ is its location and θ its direction. The distribution of minutiae location conditioned on ridge flow is shown in Figure 2.4 where there were 400 fingerprints of each type. In the model we develop the combined distribution over all types is used (Figure 2.4(f)).

Since minutia location has a multimodal distribution, a mixture of K Gaussians is a natural approach. For the data set considered a value of $K = 3$ provided a good fit, as validated by a goodness of fit test. Values of $K = 4, 5$ do not fit the data as well. A Gaussian mixture with $k = 3$ is shown in Figure 2.5.

Since minutiae orientation is a periodic variable, it is modeled by a *circular normal* or von Mises distribution which itself is derived from the Gaussian [6, 30]. Such a model is



(a) (b) (c)

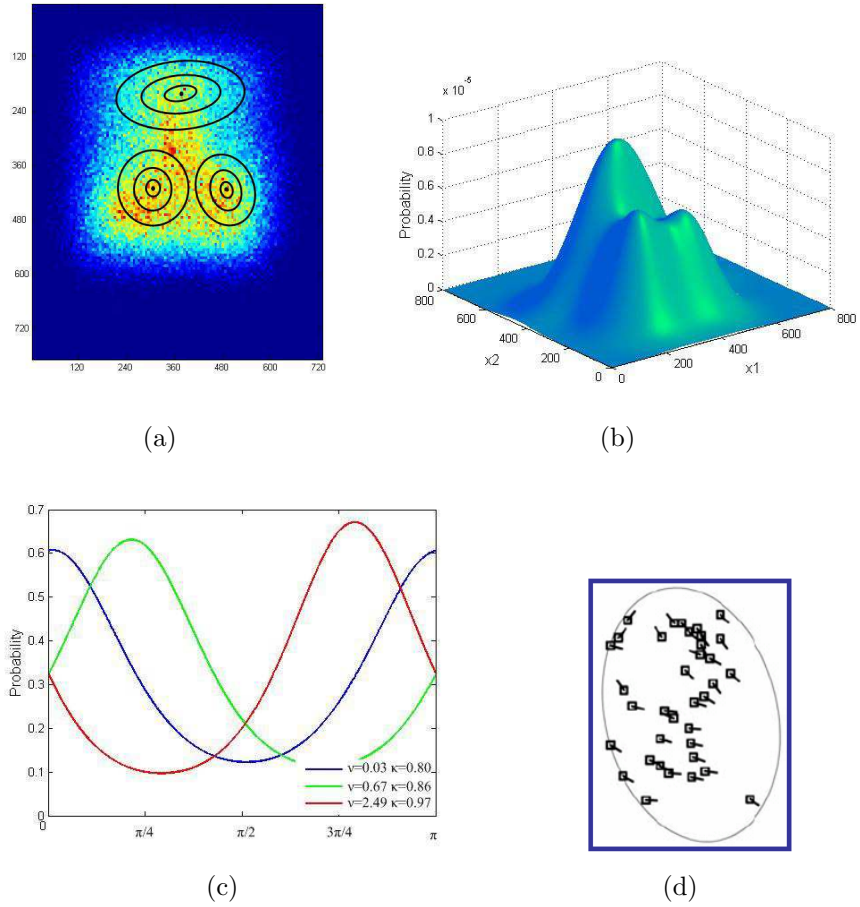
(d) (e) (f)

Figure 2.4: Distribution of minutia location for different types of ridge flow: (a) arch, (b) left loop, (c) right loop, (d) tent, (e) whorl, and (e) all types combined.

better than mixtures of hyper-geometric and binomial distributions [35, 15].

Such a model for minutiae distributions involves a random variable z that represents the particular mixture component from which the minutia is drawn. In this model both minutiae location and orientation depend on the component they belong to. Minutiae location and orientation are conditionally independent given the component. This is represented by

$$p(\mathbf{x}|z) = p(s, \theta|z) = p(s|z)p(\theta|z), \tag{2.9}$$



(a) (b)

(c) (d)

Figure 2.5: Models for minutiae distribution using Gaussian mixture for location and von Mises for direction: (a) Gaussian mixture model for minutia location with three components, (b) three-dimensional plot of mixture model, (c) von Mises distributions of minutiae orientation for each of the three components, where the green curve corresponds to the upper cluster, blue the lower left cluster and red the lower right cluster, and (d) sample generated from model.

whose graphical model is shown in Figure 2.6 from which we have the joint distribution

$$p(\mathbf{x}, z) = p(z)p(\mathbf{x}|z). \tag{2.10}$$

Marginalizing over the components, we have the distribution of minutiae as

$$p(\mathbf{x}) = \sum_{z} p(z)p(\mathbf{x}|z). \tag{2.11}$$

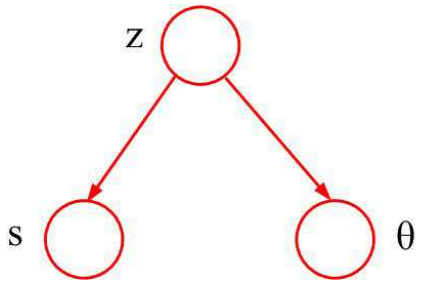


Figure 2.6: Mixture model for distribution of minutiae $\mathbf{x} = (s, \theta)$. The joint distribution of minutia location and orientation is given by $p((s, \theta), z) = p(z)p(s|z)p(\theta|z)$.

Substituting (2.9) in (2.11) we have

$$p(\mathbf{x}) = \sum_{z} p(z)p(s|z)p(\theta|z). \tag{2.12}$$

Since minutiae location within each component is Gaussian and minutiae orientation within each component is von Mises we can write

$$p(\mathbf{x}|\Theta) = \sum_{k=1}^{K} \pi_k \cdot \mathcal{N}(s|\mu_k, \Sigma_k) \cdot \mathcal{V}(\theta|\nu_k, \kappa_k), \tag{2.13}$$

where K is the number of mixture components, π_k are non-negative component weights that sum to one, $\mathcal{N}(s|\mu_k, \Sigma_k)$ is the bivariate Gaussian probability density function of minutia with mean μ_k and covariance matrix Σ_k, $\mathcal{V}(\theta|\nu_k, \kappa_k)$ is the von Mises probability density function of minutiae orientation with mean angle ν_k and precision (inverse variance) κ_k, and $\Theta = \{\pi_k, \mu_k, \Sigma_k, \nu_k, \kappa_k, \rho_k\}$ where $k = 1, 2, .., K$ is the set of all parameters of the k Gaussian and von Mises distributions.

Rather than using the standard form of the von Mises distribution for the range $[0, 2\pi]$, since minutiae orientations are represented as being in the range $[0, \pi)$, we use the alternate form [**?**] as follows

$$\mathcal{V}(\theta|\nu_k, \kappa_k, \rho_k) = \rho_k \upsilon(\theta) \cdot I\{0 \leq \theta < \pi\} + (1 - \rho_k)\upsilon(\theta - \pi) \cdot I\{\pi \leq \theta < 2\pi\} \tag{2.14}$$

where $I\{A\}$ is the indicator function of the condition A,

$$\upsilon(\theta) \equiv \upsilon(\theta|\nu_k, \kappa_k) = \frac{2}{I_0(\kappa_i)} exp[\kappa_i cos2(\theta - \nu_k)], \tag{2.15}$$

minutiae arising from the k^{th} component have directions that are either θ or $\theta + \pi$ and the probabilities associated with these two occurrences are ρ_k and $1 - \rho_k$ respectively.

Since fingerprint ridges flow smoothly with very slow direction changes, direction of neighboring minutiae are strongly correlated, i.e., minutiae that are spatially close tend to have similar directions with each other. However, minutiae in different regions of a fingerprint tend to be associated with different region-specific minutiae directions thereby demonstrating independence [13, 21]. The model allows ridge orientations to be different at different regions (different regions can be denoted by different components) while it makes sure that nearby minutiae have similar orientations (as nearby minutiae will belong to the same component).

Since several minutiae are observed in a finger, a joint distribution model is needed. The minutiae are assumed to be independent of each other, with each minutiae, consisting of an (s, θ) pair, being distributed according to a mixture of component densities. This is discussed further when we consider parameter estimation in Section 2.4.4.

2.4.3 Distribution of Representative Ridges

In the model just discussed, only minutiae was used in the framework of generative models for fingerprints. Models based purely on minutiae may be sufficient to model biometric scenarios where finger-prints are obtained in controlled conditions [35][15], but are insufficient to model forensic scenarios where latent prints are lifted off of surfaces. Due to the poor quality of latent prints, the detected minutiae are of low quantity and quality [61].

In forensics, ridge details in fingerprints provide vital information about fingerprints. Verification systems using minutiae together with ridge information are more accurate than using minutiae alone. Also, any generative model that makes use of ridge details can only be a better representation of the generative model for fingerprints. Ridge features are illustrated in Figure 2.7 where three different fingerprints are shown. In fingerprints 1 and 2, minutiae m_1 and m_2 are similar as well as the associated ridges r_1 and r_2, are similar. In fingerprint 3, minutia m_3 is similar to m_1 and m_2 but the ridge r_3 is dissimilar to r_1 and r_2. Thus two minutiae matching with each other on location and orientation may actually have two very different ridge shapes. Thus, to make more reliable decisions on whether two fingerprints match, we use ridge information as well as minutiae location and orientation information. With this motivation, the distribution of ridge information is embedded into generative models [57].

Since nearby ridges always maintain similar trend, too much redundancy would be involved if all the ridges are modeled. We only consider ridges where the minutiae lie. These ridges are called *representative ridges*. A representative ridge contains two parts, the minutiae and the following ridge. The latter part is represented as a set of ridge points sampled

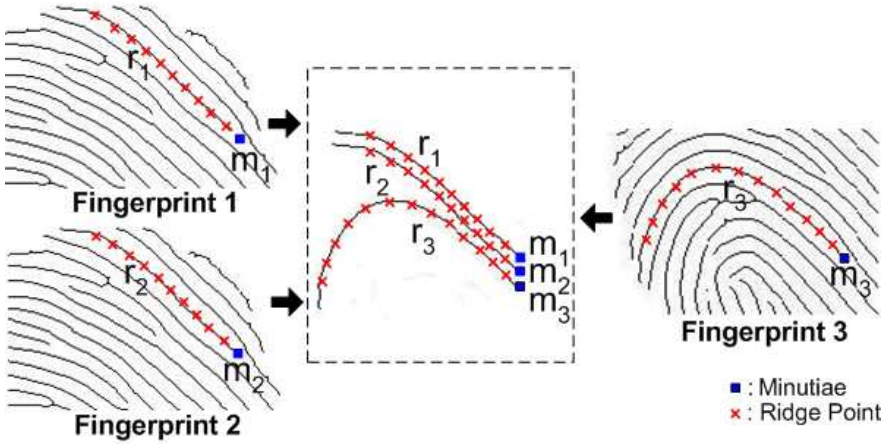


Figure 2.7: Representation of fingerprints using minutiae and ridge information: similarity of ridge shapes allows the matching of corresponding minutiae pair. Three fingerprints are shown, two of which are similar and one dissimilar. Minutiae m_1 and m_2 in fingerprints 1 and 2 are similar since not only their locations are similar but also the associated ridges r_1 and r_2 are similar. However, minutiae m_3 in fingerprint 3 has a location similar to m_1 and m_2 but the associated ridge r_3 is dissimilar to r_1 and r_2.

at equal interval of inter ridge width [17].

The length of the ridge l_r is defined as the number of ridge points that could be sampled on the ridge. Three types of ridges are defined: (i) short: $l_r \leq L/3$, (ii) medium: $L/3 < l_r \leq 2L/3$ and (iii) long: $l_r > 2L/3$, where L is the average ridge length of the top 10% longest representative ridges in the fingerprints database, e.g. in FVC2002, L is 18. By choosing the value of maximum ridge length L as the average of long ridges in the database, unusually long ridges caused by artifacts is avoided. The three possible ridge length types can be associated with any representative ridge. Without loss of generality, we can assume that there exist only three possible ridge length types corresponding to a representative ridge. The distribution of ridge length l_r is modeled as

$$p(l_r) = \frac{c_1}{c} \cdot I\{l_r \leq L/3\} + \frac{c_2}{c} \cdot I\{L/3 < l_r < 2L/3\} + \frac{c_3}{c} \cdot I\{2L/3 \leq l_r \leq L\} \qquad (2.16)$$

where $I\{C\}$ is the indicator function of condition C, c_1, c_2 and c_3 are the numbers of short, medium and long representative ridges, and $c = c_1 + c_2 + c_3$.

For ridges with different lengths, different ridge points are picked as anchors. For medium ridges, $\lfloor L/3 \rfloor^{th}$ ridge point is picked and for long ridges, both $\lfloor L/3 \rfloor^{th}$ and $\lfloor 2L/3 \rfloor^{th}$ are picked. No ridge point is chosen for short ridges. The rationale for choosing these two ridge points is described in [17].

Let $\mathbf{x} = \{\mathbf{x}_m, \mathbf{x}_r\}$ denote the feature vector of a representative ridge, where minutiae feature vector $\mathbf{x}_m$ is given by $\{s_m, \theta_m\}$ and ridge feature vector $\mathbf{x}_r$ is represented by a ridge point set given by $\{\mathbf{x}_{ri} : i \in A \wedge i \leq l_r\}$ where $\mathbf{x}_{ri}$ is the feature vector of the i^{th} ridge point, A is the anchor point index set $\{\lfloor L/3 \rfloor, \lfloor 2L/3 \rfloor\}$ and l_r is the length of the ridge. In contrast to the model for minutiae, ridge points are represented in polar coordinates as shown in Figure 2.8. The location of ridge point s_{ri} is given by $\{r_i, \phi_i\}$ and the direction of ridge point is θ_i. Thus the ridge point $\mathbf{x}_{ri}$ is represented as the combination of location and direction $\{r_i, \phi_i, \theta_i\}$ where r_i is the distance from the i^{th} ridge point to the minutia, ϕ_i is the positive angle required to reach the i^{th} ridge point from the polar axis.

A graphical model that represents the use of ridge information in addition to minutiae is given in Figure 2.9. The model can be represented as

$$p(l_r, \mathbf{x}_m, \mathbf{x}_r) = p(l_r) \cdot p(\mathbf{x}_m | l_r) \cdot p(\mathbf{x}_r | \mathbf{x}_m, l_r) \tag{2.17}$$

where $p(l_r)$ is the distribution of ridge length, $p(\mathbf{x}_m|l_r)$ is the distribution of minutiae given certain ridge length type and $p(\mathbf{x}_r|\mathbf{x}_m, l_r)$ is the distribution of ridge points given corresponding minutiae and ridge length type. For different ridge length, the distribution of ridge

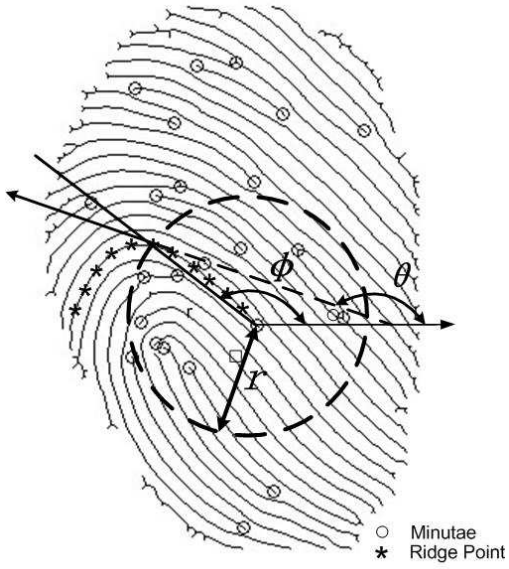


Figure 2.8: Representation of ridge points in polar coordinates. The sixth ridge point from the minutia at the center is represented by $((r, \phi), \theta)$, where r and ϕ are polar coordinates of its location, with the origin at the minutia, and its direction θ is the angle the tangent at the ridge point makes with the horizontal.

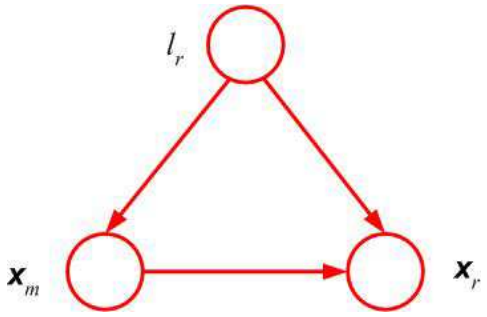


Figure 2.9: Graphical model for representative ridges.

points $\mathbf{x}_r$ is given as

$$p(\mathbf{x}_r|\mathbf{x}_m, l_r) = \begin{cases} 1 & l_r \leq \frac{L}{3} \\ p(\mathbf{x}_{r\lfloor\frac{L}{3}\rfloor}|\mathbf{x}_m, l_r) & \frac{L}{3} > l_r \leq \frac{2L}{3} \\ p(\mathbf{x}_{r\lfloor\frac{2L}{3}\rfloor}|\mathbf{x}_m, l_r) \cdot p(\mathbf{x}_{r\lfloor\frac{L}{3}\rfloor}|\mathbf{x}_m, l_r) & l_r > \frac{2L}{3} \end{cases} \tag{2.18}$$

where $\mathbf{x}_{r\lfloor\frac{L}{3}\rfloor}$ and $\mathbf{x}_{r\lfloor\frac{2L}{3}\rfloor}$ represent the feature vectors of $\lfloor\frac{L}{3}\rfloor^{th}$ and $\lfloor\frac{2L}{3}\rfloor^{th}$ ridge points.

The generative model is based on the distribution of representative ridges. Mixture distributions consisting of K_i components, $i = 1, 2, 3$, is used to model representative ridges of three ridge length types. Each component is distributed according the density of the minutiae and density of ridge points. Assuming that the minutiae and ridge points are independent, representative ridge distribution is given by

$$p(\mathbf{x}|\Theta) = \begin{cases} p(l_r) \cdot \sum_{g=1}^{K_1} \pi_k p_k(s_m, \theta_m|\Theta_k) & l_r \leq \frac{L}{3} \\ p(l_r) \cdot \sum_{k=1}^{K_2} \pi_k p_k(s_m, \theta_m|\Theta_k) & \\ \cdot p_k(r_{\lfloor\frac{L}{3}\rfloor}, \phi_{\lfloor\frac{L}{3}\rfloor}, \theta_{\lfloor\frac{L}{3}\rfloor}|\Theta_k) & \frac{L}{3} < l_r < \frac{2L}{3} \\ p(l_r) \cdot \sum_{k=1}^{K_3} \pi_k p_k(s_m, \theta_m|\Theta_k) \cdot p_k(r_{\lfloor\frac{L}{3}\rfloor}, \phi_{\lfloor\frac{L}{3}\rfloor}, \theta_{\lfloor\frac{L}{3}\rfloor}|\Theta_k) & \\ \cdot p_k(r_{\lfloor\frac{2L}{3}\rfloor}, \phi_{\lfloor\frac{2L}{3}\rfloor}, \theta_{\lfloor\frac{2L}{3}\rfloor}|\Theta_k) & l_r \geq \frac{2L}{3} \end{cases} \tag{2.19}$$

The first condition corresponds to minutiae alone, the second to minutia and one ridge point, and the third to minutia and two ridge points. $p_k(s_m, \theta_m|\Theta_k)$ is the component distribution for minutiae location s_m and direction θ_m and $p_k(r_i, \phi_i, \theta_i|\Theta_k)$ is the component

distribution for the i^{th} ridge point. They are defined as in (2.20) and (2.21) respectively.

$$p_k(s_m, \theta_m|\Theta_k) = \mathcal{N}(s_m|\mu_{mk}, \Sigma_{mk}) \cdot \mathcal{V}(\theta_m|\nu_{mk}, \kappa_{mk}, \rho_{mk}) \tag{2.20}$$

$$p_k(r_i, \phi_i, \theta_i|\Theta_k) = p_k(r_i, \phi_i|\mu_{ik}, \Sigma_{ik}, \nu^{\phi}_{ik}, \kappa^{\phi}_{ik}, \rho^{\phi}_{ik}) \cdot \mathcal{V}(\theta_i|\nu^{\theta}_{ik}, \kappa^{\theta}_{ik}, \rho^{\theta}_{ik}) \tag{2.21}$$

where $p_k(s_m, \theta_m|\Theta_k)$ can be caculated by Eq. 2.13 and $p_k(r_i, \phi_i, \theta_i|\Theta_k)$ is the product of the probabilities of ridge point locations and directions, where $\mathcal{V}(\theta_i|\nu^{\theta}_{ik}, \kappa^{\theta}_{ik}, \rho^{\theta}_{ik})$ presents the distribution of the ridge point direction, θ_i is the direction of the i^{th} ridge point and $p_k(r_i, \phi_i|\mu_{ik}, \Sigma_{ik}, \nu^{\phi}_{ik}, \kappa^{\phi}_{ik}, \rho^{\phi}_{ik})$ is the distribution of ridge point location given by

$$p_k(r_i, \phi_i|\mu_{ik}, \sigma_{ik}, \nu^{\phi}_{ik}, \kappa^{\phi}_{ik}, \rho^{\phi}_{ik}) = \mathcal{N}(r_i|\mu_{ik}, \sigma_{ik}) \cdot \mathcal{V}(\phi_i|\nu^{\phi}_{ik}, \kappa^{\phi}_{ik}, \rho^{\phi}_{ik}) \tag{2.22}$$

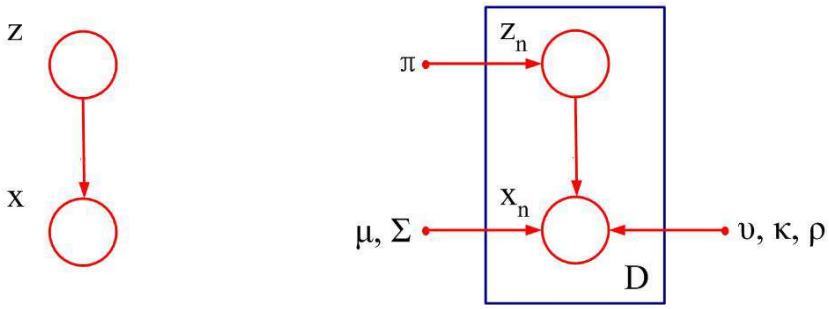
where $\mathcal{N}(r_i|\mu_{ik}, \sigma_{ik})$ is a univariate Gaussian distribution whose mean μ_{ik} and variance σ_{ik} are learnt from a fingerprint database.

2.4.4 Parameter Estimation

We now develop an equivalent formulation of the mixture distribution given in (2.13) by involving an explicit latent variable. This will allow us to formulate the problem of parameter estimation in terms of the expectation maximization (EM) algorithm.

We define the joint distribution $p(\mathbf{x}, \mathbf{z})$ in terms of a marginal distribution p(z) and a conditional distribution $p(\mathbf{x}|\mathbf{z})$, corresponding to the graphical model in Figure 2.10(a).

Given that the total number of minutiae observed in a finger is D, a joint distribution model is needed. The D minutiae are assumed to be independent of each other, with each



(a) (b)

Figure 2.10: Graphical models representing mixture for: (a) single minutia whose distribution is expressed as $p(\mathbf{x}, \mathbf{z}) = p(\mathbf{x})p(\mathbf{x}|\mathbf{z})$, (b) set of D identically distributed minutiae with corresponding latent points $\mathbf{z}_n$, where $n = 1, .., D$.

minutiae, consisting of an $\mathbf{x}(s, \theta)$ pair, being distributed according to a mixture of component densities. This is shown in Figure 2.10(b).

The K-dimensional random variable $\mathbf{z}$ has a 1-of-K representation in which a particular element z_k is equal to 1 and all other elements are equal to 0, we can write

$$p(\mathbf{z}) = \prod_{k=1}^{K} \pi_k^{z_k} \tag{2.23}$$

Similarly the conditional distribution of $\mathbf{x}$ given a particular value for $\mathbf{z}$ is given by

$$p(\mathbf{x}|z_k = 1) = \mathcal{N}(\mathbf{x}|\mu_k, \Sigma_k) \cdot \mathcal{V}(\theta|\nu_k, \kappa_k, \rho_k) \tag{2.24}$$

which can also be written in the form

$$p(\mathbf{x}|\mathbf{z}) = \prod_{k=1}^{K} \mathcal{N}(\mathbf{x}|\mu_k, \Sigma_k)^{z_k} \cdot \mathcal{V}(\theta|\nu_k, \kappa_k, \rho_k)^{z_k}. \tag{2.25}$$

The joint distribution is given by $p(\mathbf{z})p(\mathbf{x}|\mathbf{z})$, and the marginal distribution of $\mathbf{x}$ is obtained by summing the joint distribution over all possible states of $\mathbf{z}$ to give

$$p(\mathbf{x}) = \sum_{\mathbf{z}} p(\mathbf{z})p(\mathbf{x}|\mathbf{z}) = \sum_{k=1}^{K} \pi_k \cdot \mathcal{N}(s|\mu_k, \Sigma_k) \cdot \mathcal{V}(\theta|\nu_k, \kappa_k, \rho_k) \tag{2.26}$$

where we have made use of (2.23) and (2.25). Thus the marginal distribution of $\mathbf{x}$ is a mixture of the form (2.13). If we have several observed minutiae $\mathbf{x}_1, \mathbf{x}_2, ..\mathbf{x}_D$ then, because we have represented the marginal distribution in the form $p(\mathbf{x}) = \sum_{\mathbf{z}} p(\mathbf{z})p(\mathbf{x}|\mathbf{z})$, it follows that for every observed minutia $\mathbf{x}_n$, there is a corresponding latent variable $\mathbf{z}_n$.

We are now able to work with the joint distribution $p(\mathbf{x}, \mathbf{z})$ instead of the marginal distribution $p(\mathbf{x})$. To estimate the unknown parameters using the maximum likelihood approach, we use the EM algorithm. The number of components K for the mixture model was found after validation using k-means clustering.

E-Step: Using γ_{dk} is to denote the responsibility of component k for minutiae $\mathbf{x}_d$, its value can be found using Bayes's theorem

$$\begin{aligned} \gamma_{dk} \equiv p(z_k = 1|\mathbf{x}_d) &= \frac{p(z_k = 1)p(\mathbf{x}_d|z_k = 1)}{\sum_{k=1}^{K} p(z_k = 1)p(\mathbf{x}_d|z_k = 1)} \\ &= \frac{\pi_k \mathcal{N}(s_d|\mu_k, \sigma_k)\mathcal{V}(\theta_d|\nu_k, \kappa_k, \rho_k)}{\sum_{k=1}^{K} \pi_k \mathcal{N}(s_d|\mu_k, \sigma_k)\mathcal{V}(\theta_d|\nu_k, \kappa_k, \rho_k)} \end{aligned} \tag{2.27}$$

M-Step: The estimates of the Gaussian distribution parameters π_k, μ_{mk} and Σ_{mk} at the

$(n+1)th$ iteration, are given by

$$\pi_k^{(n+1)} = \frac{1}{D}\sum_{d=1}^{D}\gamma_{dk}^{(n)} \tag{2.28}$$

$$\mu_{mk}^{(n+1)} = \frac{\sum_{d=1}^{D}\gamma_{dk}^{(n)} s_m}{\sum_{d=1}^{D}\gamma_{dk}^{(n)}} \tag{2.29}$$

$$\Sigma_{mk}^{(n+1)} = \frac{\sum_{d=1}^{D}\gamma_{dk}^{(n)}(s_m - \mu_{mk}^{(n+1)})(s_m - \mu_{mk}^{(n+1)})^T}{\sum_{d=1}^{D}\gamma_{dk}^{(n)}} \tag{2.30}$$

The parameters for orientation distributions are obtained using expectation maximization for the von Mises distribution [4]. The estimates of ν_{mk} and κ_{mk} at the $(n+1)th$ iteration are given by

$$\nu_{mk}^{(n+1)} = \frac{1}{2}tan^{-1}\left(\frac{\sum_{d=1}^{D}\gamma_{dk}^{(n)} sin2\psi_d}{\sum_{d=1}^{D}\gamma_{dk}^{(n)} cos2\psi_d}\right) \tag{2.31}$$

$$\frac{I_0'(\kappa_{mk}^{(n+1)})}{I_0(\kappa_{mk}^{(n+1)})} = \frac{\sum_{d=1}^{D} r_{dk}^{(n)} cos2(\psi_d - \nu_k^{(n+1)})}{\sum_{d=1}^{D} r_{dk}^{(n)}}. \tag{2.32}$$

The solution for (2.32), which involves Bessel functions, obtained using an iterative method gives the estimate for κ_{mk}. The estimate of ρ_{mk} is then obtained as

$$\rho_{mk}^{(n+1)} = \frac{\sum_{d=1}^{D} I\{c_d^{(n+1)} = k, \theta_d \in [0,\pi)\}}{\sum_{d=1}^{D} I\{c_d^{(n+1)} = k\}} \tag{2.33}$$

where $c_d^{(n+1)} = arg\ max_k \gamma_{dk}^{(n+1)}$ is the component label for the observation d at the $(n+1)th$ iteration, ψ_j is the orientation of the minutiae m_j.

The parameters in generative models for minutiae (Eq. 2.13) can be learned by Eq. 2.28, 2.29, 2.30, 2.31, 2.32 and 2.33. In the same way, the parameters for ridge point $\nu_{ik}^{\theta}, \kappa_{ik}^{\theta}, \rho_{ik}^{\theta}$ and $\nu_{ik}^{\phi}, \kappa_{ik}^{\phi}, \rho_{ik}^{\phi}$ can be estimated by Eq. 2.31, 2.32 and 2.33, when ψ_j is set as θ_i and ϕ_i respectively.

2.5 Evaluation of PRCs

To compute the PRCs, we first define correspondence, or match, between two minutiae. Let $\mathbf{x_a} = (s_a, \theta_a)$ and $\mathbf{x}_b = (s_b, \theta_b)$ be a pair of minutiae. The minutiae are said to correspond if for tolerance $\epsilon = [\epsilon_s, \epsilon_\theta]$,

$$|s_a - s_b| \leq \epsilon_s \text{ and } |\theta_a - \theta_b| \leq \epsilon_\theta \tag{2.34}$$

where $|s_a - s_b|$, the Euclidean distance between the minutiae location $s_a = (x_{a1}, x_{a2})$ and $s_b = (x_{b1}, x_{b2})$, is given by

$$|s_a - s_b| = \sqrt{(x_{a1} - x_{b1})^2 + (x_{a2} - x_{b2})^2} \tag{2.35}$$

Then, the probability that a random minutia $\mathbf{x}_a$ would match a random minutia $\mathbf{x}_b$ is given by

$$\begin{aligned} p_\epsilon(\mathbf{x}) &= p(|\mathbf{x}_a - \mathbf{x}_b| \leq \epsilon|\Theta) \\ &= \int_{\mathbf{x}_a} \int_{|\mathbf{x}_a - \mathbf{x}_b| \leq \epsilon} p(\mathbf{x}_a|\Theta) p(\mathbf{x}_b|\Theta) d\mathbf{x}_a d\mathbf{x}_b \end{aligned} \tag{2.36}$$

where Θ is the set of parameters describing the distribution of the minutiae location and

direction.

Finally, the PRC, or the probability of matching at least $\hat{m}$ pairs of minutiae within ϵ between two randomly chosen fingerprint f_1 and f_2 is calculated as

$$p_\epsilon(\hat{m}, m_1, m_2) = \binom{m_1}{\hat{m}}\binom{m_2}{\hat{m}}\hat{m}! \cdot p_\epsilon(\mathbf{x})^{\hat{m}}(1 - p_\epsilon(\mathbf{x}))^{(m_1-\hat{m})\cdot(m_2-\hat{m})} \tag{2.37}$$

where m_1 and m_2 are numbers of minutiae in fingerprints f_1 and f_2, $p_\epsilon(\mathbf{x})^{\hat{m}}$ is the probability of matching $\hat{m}$ specific pairs of minutiae between f_1 and f_2, $(1 - p_\epsilon(\mathbf{x}))^{(m_1-\hat{m})\cdot(m_2-\hat{m})}$ is the probability that none of minutiae pair would match between the rest of minutiae in f_1 and f_2 and $\binom{m_1}{\hat{m}}\binom{m_2}{\hat{m}}\hat{m}!$ is the number of different match sets that can be paired up.

Given n fingerprints and assuming that the number of minutiae in a fingerprint m can be modeled by the distribution $p(m)$, the general PRCs $p(n)$ is given by

$$p(n) = 1 - \bar{p}(n) = 1 - (1 - p_\epsilon)^{\frac{n(n-1)}{2}} \tag{2.38}$$

where p_ϵ is the probability of matching two random fingerprint from n fingerprints. If we set the tolerance in terms of number of matching minutiae to $\hat{m}$, p_ϵ is calculated by

$$p_\epsilon = \sum_{m_1' \in M_1} \sum_{m_2' \in M_2} p(m_1')p(m_2')p_\epsilon(\hat{m}, m_1', m_2') \tag{2.39}$$

where M_1 and M_2 contain all possible numbers of minutiae in one fingerprint among n fingerprints, and $p_\epsilon(\hat{m}, m_1', m_2')$ can be calculated by Eq. 2.37.

2.6 Experiments

Parameters of the fingerprint generative model introduced in Sections 2.4.2 was evaluated using the NIST fingerprint database. The NIST fingerprint database, NIST Special Database 4, contains 8-bit gray scale images of randomly selected fingerprints. Each print is 512 × 512 pixels with 32 rows of white space at the bottom of the print. The entire database contains fingerprints taken from 2000 different fingers with 2 impression of the same finger. Thus, there are a total of 2000 fingerprints using which the model has been developed. The fingerprints are classified into one of five categories (left loop, whirl, right loop, tented arch, and arch) with an equal number of prints from each class (400). The number of components K for the mixture model was found after validation using k-means clustering. In order to justify the generative models, goodness-of-fit test is perform on FVC2002 DB1 dataset. The results are shown in Section 2.6.1. Section 2.6.2 presents the table for PRC values.

2.6.1 Goodness-of-fit Test

Goodness of fit means how well a sample of data agrees with a given distribution as its population. To test the goodness of fit, the chi-square statistical hypothesis test is applied. Chi-square goodness of fit test determines whether observed sample frequencies differ significantly from expected frequencies specified in the null hypothesis. The test is applied to binned data (i.e., data put into classes) and requires a sufficient sample size in each bin in order for the chi-square approximation to be valid [38]. In the case of fingerprint, we partitioned the minutiae location and direction space into 16 × 4 non-overlapping blocks. The blocks are combined with adjacent blocks until both observed and expected numbers

Table 2.2: Results from the Chi-square tests for testing the goodness of fit of the mixture models with and without ridge information.The total number of fingerprints in FVC2002 DB1 is 800.

p-value	**Without Ridge information**	**With Ridge information**
$p - value > 0.01$ (Model Accepted)	574	679
$p - value \leq 0.01$ (Model Rejected)	226	121

of minutiae in the block are greater than or equal to 5. The test statistic used here is a chi-square random variable χ^2 defined by the following equation.

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} \tag{2.40}$$

where O_i is the observed minutiae count for the ith block, and E_i is the expected minutiae count for the ith block.

The p-value, the probability of observing a sample statistic as extreme as the test statistic, associated with each test statistic χ^2 is then calculated based on the chi-square distribution and compared to the significance level. For the FVC2002 DB1, we chose significance level equal to 0.01. The numbers of fingerprints with p-values above (corresponding to accept the model) and below (corresponding to reject the model) the significance level are then computed. The results are given in Table 3.2. Of the 800 fingerprints, 679 are accepted with ridge model which is higher than 574 and 121 are rejected whichis smaller than 226. The results imply that the mixture model with ridge information offers a better fit to the observed fingerprints compared to the model without ridge information.

2.6.2 PRC Evaluation with Fingerprint Datasets

Values of PRC p_ϵ are calculated using the formula introduced in Section 2.5. The tolerance is set at $\epsilon_s = \epsilon_r = 10$ pixels and $\epsilon_\theta = \epsilon_\phi = \pi/8$. For comparison, the empirical PRC $\hat{p}_\epsilon(\mathbf{x})$ was calculated with the same tolerance. To compute $\hat{p}_\epsilon(\mathbf{x})$, the empirical probabilities of matching a minutiae pair or ridge pair between imposter fingerprints are calculated first by

$$\hat{p}_\epsilon(\mathbf{x}) = \frac{1}{I}\sum_{i=1}^{I}\frac{\hat{m}_i}{m_i \times m_i'} \tag{2.41}$$

where I is the number of the impostor fingerprints pairs, $\hat{m}_i$ is the number of matched minutiae or ridge pairs and m_i and m_i' are the numbers of minutiae or pairs in each of the two fingerprints. Then, the empirical PRC $\hat{p}_\epsilon$ can be calculated by Eq.2.37.

Both the theoretical and empirical PRCs are given in Table 2.3. The PRCs are calculated through varying number of minutiae or ridges in two randomly chosen fingerprint f_1 and f_2 and the number of matches between them. We can see that more minutiae or ridges the template and input fingerprint have, higher the PRC is. It should be noted that the PRC values with ridge information model are never greater than PRC values without ridge information, which indicates that ridge information strengthens individuality of fingerprints. Note that the theoretical PRC based on our model are close and have the same trend to empirical PRC. The consistency between the theoretical probabilities and empirical probabilities shows the validation of our generative model. The PRCs for the different m_1 and m_2 with 6, 16, 26 and 36 matches are shown in Figure 2.11. It is obvious to note that, when $\hat{m}$ decreases or m_1 and m_2 increase, the probability of matching two random fingerprints is more.

Based on the PRC value, nPRC can be computed. Table 2.4 shows the nPRCs in $100,000$

Table 2.3: PRC for different fingerprint matches with varying m_1(number of minutiae/ridges in fingerprint f_1),m_2 (number of minutiae/ridges in fingerprint f_2) and $\hat{m}$ (number of matched minutiae/ridges) - With ridge information and without ridge information. p_ϵ is the theoretical PRC for the general population and $\hat{p_\epsilon}$ is the empirical PRC for NIST 4 database.

			Minutiae Only		**Minutiae and Ridge Points**	
m_1	m_2	$\hat{m}$	$p_\epsilon(m_1, m_2, \hat{m})$	$\hat{p_\epsilon}(m_1, m_2, \hat{m})$	$p_\epsilon(m_1, m_2, \hat{m})$	$\hat{p_\epsilon}(m_1, m_2, \hat{m})$
16	16	6	6.3×10^{-9}	1.0×10^{-8}	3.7×10^{-11}	5.9×10^{-10}
		12	3.4×10^{-23}	1.0×10^{-22}	5.1×10^{-28}	9.2×10^{-26}
		16	1.2×10^{-37}	5.4×10^{-37}	6.4×10^{-46}	4.3×10^{-45}
36	36	6	2.1×10^{-4}	3.4×10^{-4}	8.3×10^{-6}	7.0×10^{-5}
		12	1.1×10^{-11}	3.1×10^{-11}	6.0×10^{-14}	1.8×10^{-14}
		24	4.0×10^{-34}	3.6×10^{-33}	5.2×10^{-40}	7.6×10^{-39}
		36	3.6×10^{-72}	9.9×10^{-71}	1.3×10^{-86}	8.7×10^{-84}
56	56	6	1.8×10^{-2}	2.6×10^{-2}	9.1×10^{-3}	2.8×10^{-3}
		12	7.8×10^{-7}	2.1×10^{-6}	5.2×10^{-9}	2.1×10^{-9}
		24	2.6×10^{-21}	2.2×10^{-20}	7.2×10^{-26}	7.7×10^{-25}
		36	1.7×10^{-42}	4.5×10^{-41}	8.9×10^{-51}	4.1×10^{-50}
		48	5.0×10^{-72}	4.1×10^{-70}	3.5×10^{-87}	3.9×10^{-86}
		56	1.1×10^{-101}	2.0×10^{-99}	6.8×10^{-123}	5.3×10^{-120}
76	76	6	1.4×10^{-1}	1.8×10^{-1}	7.5×10^{-2}	2.2×10^{-2}
		12	5.0×10^{-4}	1.1×10^{-3}	5.8×10^{-6}	3.9×10^{-5}
		24	5.6×10^{-14}	4.2×10^{-13}	2.5×10^{-18}	4.7×10^{-17}
		36	4.4×10^{-29}	1.1×10^{-27}	3.3×10^{-36}	8.3×10^{-35}
		48	3.6×10^{-49}	2.8×10^{-47}	8.1×10^{-63}	6.3×10^{-62}
		60	3.6×10^{-75}	8.8×10^{-73}	8.0×10^{-92}	6.3×10^{-90}
		76	4.9×10^{-128}	5.4×10^{-125}	3.4×10^{-155}	6.3×10^{-151}

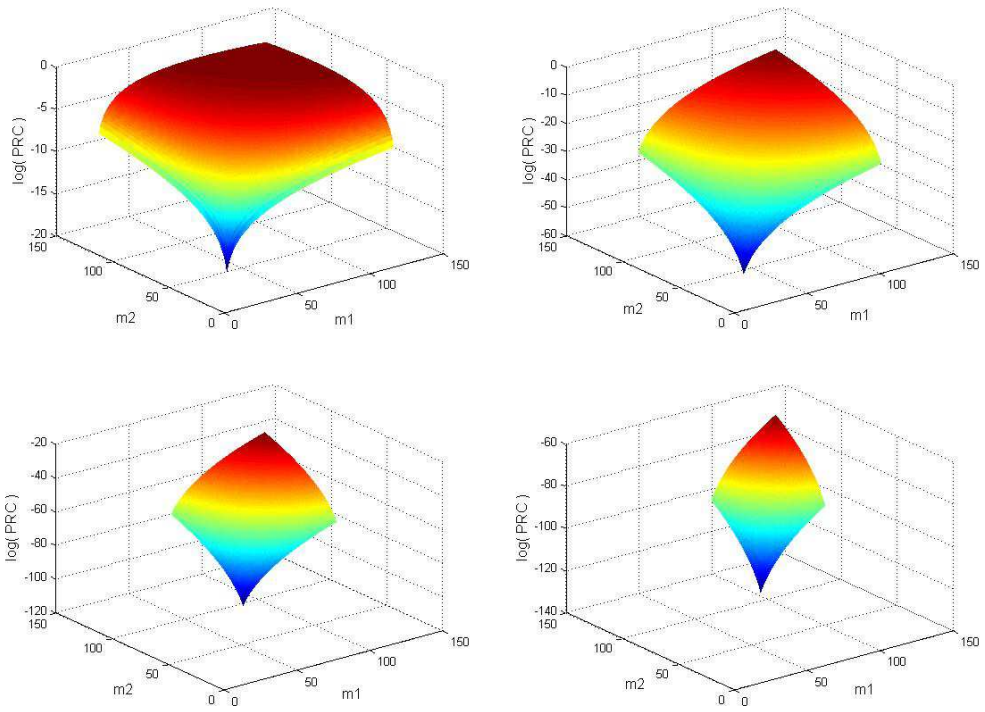


Figure 2.11: PRCs with different numbers of the matched minutiae for (a) $\hat{m} = 6$, (b) $\hat{m} = 26$, (c) $\hat{m} = 56$, and (d) $\hat{m} = 76$.

fingerprints through varying number of minutiae or ridges in each fingerprint m and number of matches $\hat{m}$.

2.7 Conclusions

While forensic evidence of many modalities have long been used in the judicial system, e.g., impression evidence (latent prints, handwriting, shoe-prints), trace evidence (paint flakes, pollen, fibers, glass, hair), etc, characterizing their accuracy in identification is still needed to provide a scientific basis. The degree of individuality of a forensic modality can be established quantitatively by using either discriminative or generative approaches. In the former an error rate is determined after a suitable classifier is constructed. In the latter a

Table 2.4: Fingerprint Probabilities: nPRCs with varying n and $\hat{m}$

No. of fingerprints n	No of matches $\hat{m}$	**Minutiae only** $p_\epsilon(n, \hat{m})$	**Minutiae and Ridge Points** $p_\epsilon(n, \hat{m})$
1000	16	1	2.8790×10^{-1}
	36	4.1731×10^{-5}	8.7502×10^{-6}
	56	1.7219×10^{-18}	7.8466×10^{-20}
	76	4.8546×10^{-38}	1.1363×10^{-42}
	96	4.2369×10^{-64}	5.1036×10^{-75}
100,000	16	1	8.3978×10^{-1}
	36	3.4146×10^{-1}	4.988×10^{-2}
	56	1.7236×10^{-14}	5.3583×10^{-16}
	76	4.8594×10^{-34}	1.4617×10^{-38}
	96	4.2411×10^{-60}	9.3221×10^{-69}
10,000,000	16	1	1
	36	1	7.980×10^{-2}
	56	1.7236×10^{-10}	9.2801×10^{-12}
	76	4.8594×10^{-30}	3.4857×10^{-35}
	96	4.2411×10^{-56}	3.8824×10^{-64}

probability distribution is determined from which different types of probabilities of random correspondence (PRC) are evaluated.

Generative models of individuality attempt to model the distribution of features and then use the models to determine the probability of random correspondence. We have proposed such models of individuality for fingerprints. Individuality is evaluated in terms of two probability measures: probability of random correspondence (PRC) between two individuals, and general probability of random correspondence (nPRC) between two individuals among a group of n individuals.

Models for fingerprint individuality have been proposed for ridge flows, minutiae only and minutiae with ridge points. Mixture distributions are proposed to model minutiae and ridge information. Goodness-of-fit tests are conducted to prove the validation of the proposed generative models. The new generative model is also compared by implementation

and experiments with the empirical results on the NIST 4 dataset. The PRC obtained for a fingerprint template and input with 36 minutiae each with 12 matching minutiae is 1.1×10^{-11}. This probability is very close to the empirical result which is 3.1×10^{-11}. nPRC and specific nPRC of fingerprints are computed also. Considering the case of $100,000$ fingerprints, the nPRC with minutiae information where 56 minutiae pairs are matched is 1.7236×10^{-14}. Given a specific fingerprint with 71 minutiae, the specific nPRC with minutiae information where 55 out of 71 minutiae are matched is 8.5815×10^{-276}. The proposed generative model offers a reasonable and accurate fingerprint representation. The results provide a much stronger argument for the individuality of fingerprints in forensics than previous generative models.

Chapter 3

FINGERPRINT RARITY

Since the earliest days of forensic fingerprint identification, the importance of considering the rarity of features used in the comparison of prints has been known. Yet methods to compute rarity of features have been elusive due to the large number of variables and the complexity of the distributions. When a latent print, typically found in a crime scene, is compared to a known, such as inked or live scan prints in a database, the degree of uncertainty is influenced by rarity as well as similarity. A method for computing the rarity of latent fingerprints represented by minutiae is proposed. It allows determining the probability of finding a random match for the evidence in a database of n prints [55]. First, the coordinate system is transformed into standard position based on finding the core point. Since latent prints are often incomplete and the core point may be missing in the field of view, a machine learning approach based on Gaussian process regression is proposed. Next a generative model, that takes into account inter-minutia dependencies and minutia confidences, is used to determine evidence probability from which the specific probability of match among n is evaluated. The generative model is validated using a goodness-of-fit test. Rarity evaluation

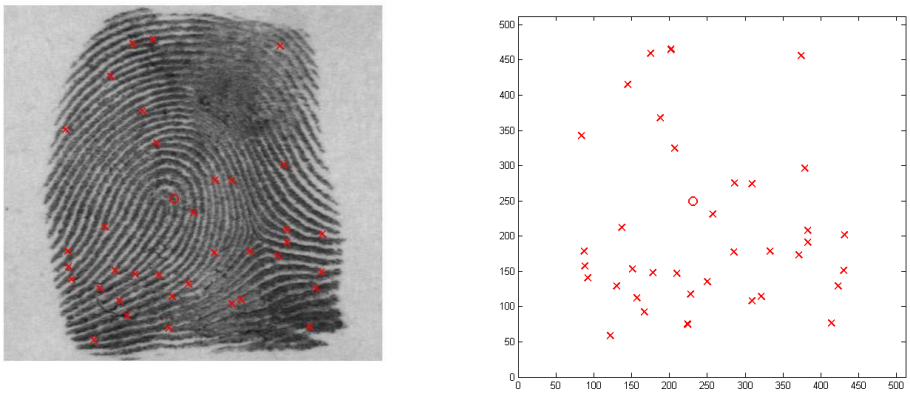
is illustrated using several examples, including simple configurations of minutiae, randomly selected latent fingerprints in a database, and a well-known case of erroneous identification.

3.1 Introduction

In the forensic sciences it has become necessary to quantify the value of evidence. Recent court challenges have highlighted the need for characterizing uncertainty rather than implicitly threshold it into individualization or exclusion [66, 31]. The "gold standard" for the forensic sciences is DNA evidence, where a probability statement can be made after a match has been confirmed between evidence and known, e.g., the chance that a random person on earth would have the same STR (short tandem repeat) pattern is 1 in 24,000,000– which is a description of rarity of the evidence/known [9].

Impression evidence consists of marks created by physical contact between object and surface. In the case of friction ridge impressions, the evidence mark is a latent print which is made visible to the naked eye using powder or ultraviolet illumination. Features commonly used to characterize friction ridge impressions are known as level 1, 2 and 3 detail. Level 1 detail, which characterize ridge flow types such as loops, arch and whorl, is only useful for exclusion. Level 2 detail, which are minutiae that correspond to ridge bifurcations and endings, contain the most discriminative information. Level 3 detail (pores) are not usually available in latent prints. Features most useful to characterize friction ridge impressions are minutiae which correspond to ridge bifurcations. trifurcations and endings (Figure 3.1). Associated with each such point is an attribute of orientation, which is determined by the associated ridge, and the confidence of its presence. Every configuration of points has a

center, called a *core point*, specified both by its location and direction.



(a) (b)

Figure 3.1: Fingerprint as spatial point set: (a) points are ridge endings/bifurcations, and (b) extracted points (x) and inferred center/core (o). Each point has attributes such as orientation, confidence and type.

Practitioners implicitly assign weights to fingerprint features in the their analysis of latent prints and subsequent evaluation of similarity to a known. Since the earliest days of fingerprint- based identification it has been recognized that observed features ought to be assessed as a function of their relative rarity while allowing also for a proper consideration of clarity [29]. This weighing is done at present manually, mainly based on subjective, experience-based judgments of the probability associated with their occurrence. The process is essentially holistic and the subjective probabilities are not explicit.

3.1.1 Likelihood Ratio Methods and Rarity

Latent print evidence used for identification has uncertainty at two levels: rarity of the mark and similarity between evidence and known. These uncertainties can be characterized as probabilities determined by the type(s) of features measured. One measure of the weight of evidence is a likelihood ratio (LR) whose numerator is the joint probability that the evidence

$\mathbf{x}$ and known $\mathbf{y}$ come from the same source (also known as the *prosecution* (p) hypothesis), and the denominator is the joint probability that the two come from two different sources (*defense* (d) hypothesis) [1]. This ratio of the two joint distributions of $\mathbf{x}$ and $\mathbf{y}$, under the prosecution and defense hypotheses, can be written as $\frac{p_p(\mathbf{x},\mathbf{y})}{p_d(\mathbf{x},\mathbf{y})}$. If the underlying distributions are Gaussian (univariate or multivariate) the LR can be simplified as the product of two factors [27]. The first factor is a Gaussian whose exponent of the squared difference of $\mathbf{x}$ and $\mathbf{y}$, which can be interpreted as a probability measure of *similarity* – which is also a significance test of the null hypothesis of identity. The second factor is a reciprocal of a Gaussian whose exponent is the squared difference of $\mathbf{x}$ and the mean of its distribution – which is precisely the reciprocal of the probability of $\mathbf{x}$ and is therefore a measure of *rarity*. While this is an elegant result, evaluation of the joint probability is difficult for fingerprints. This is because the number of possible transformations, e.g., distortions of $\mathbf{x}$ and $\mathbf{y}$, is very large, the number of possible variables, e.g., minutiae, are typically between 6 and 40 and each minutia itself is expressed as a 3-tuple (two spatial co-ordinates and an angle) with associated uncertainties.

The likelihood ratio computation is made tractable by using a similarity (or kernel) function $k(\mathbf{x}, \mathbf{y})$. Ideally this is a tangent distance function that takes into account all possible transformations [43]. Again due to the complexity of fingerprint transformations, the kernel is a much simpler score function such as those commonly used in Automatic Fingerprint Identification Systems (AFIS)[68]. Given the distribution of $k(\mathbf{x}, \mathbf{y})$ under the prosecution and defense hypotheses, the corresponding likelihood ratio is $\frac{p_p[k(\mathbf{x},\mathbf{y})]}{p_d[k(\mathbf{x},\mathbf{y})]}$ [33, 32, 45]. While this allows the joint distribution over many variables to be replaced by the distribution over a single variable, there is clearly a severe loss of information. Moreover the kernel only

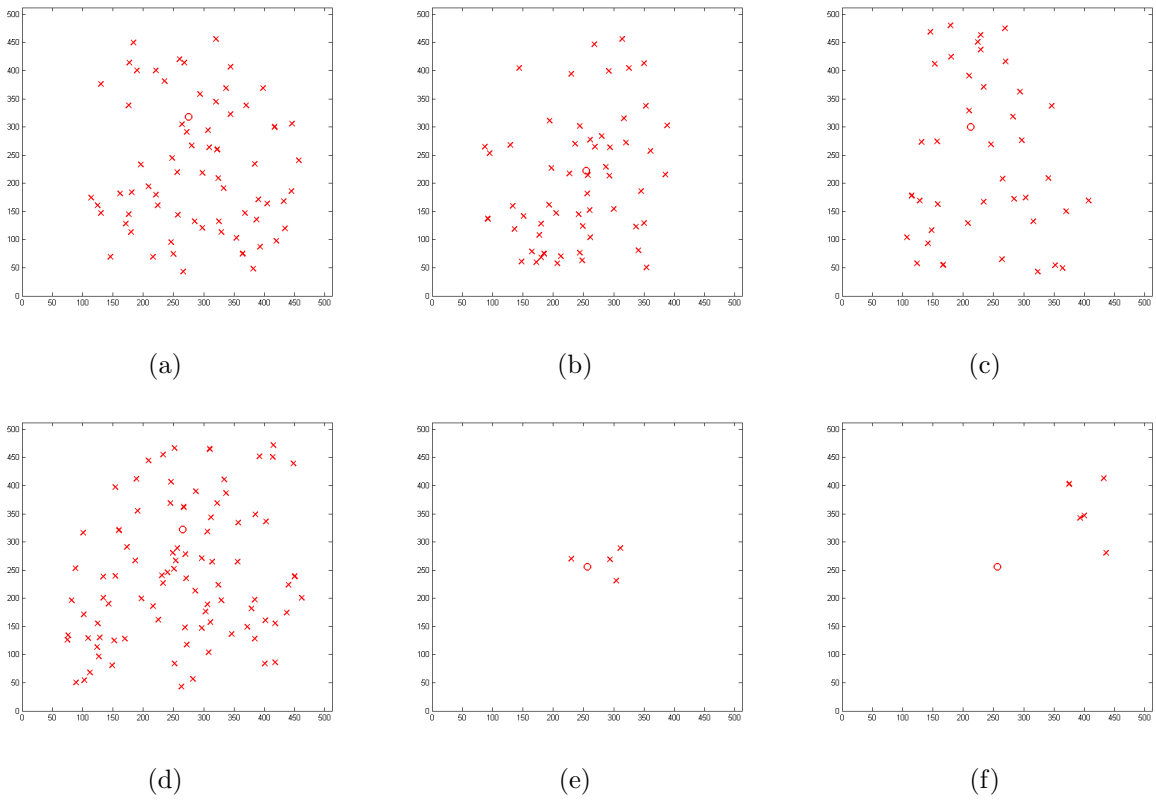
deals with similarity and any consideration of rarity is lost.

Since rarity is the reciprocal of probability, the probability of evidence is itself a measure of rarity. A more useful measure of rarity is the probability that the evidence could be randomly found in a database of a given size, which is low for a rare input and high for a common configuration. This probability can be evaluated from the distribution of the evidence, which can be learnt from a representative training sample, and a tolerance as to how close the evidence and database entry need to be. An example of learning the distribution of points from training samples and the rarity of two point configurations is shown in Figure 3.2, where points are denoted by crosses and the core point, which functions as coordinate origin, is represented by a circle.

The starting point for determining minutia distribution is to find the core point of a minutia set since it functions as the coordinate system origin. However a latent print may or may not contain a core point within it, e.g., a poor quality latent print may only represent a portion of a finger. The area of finger skin that the evidence comes from is determined probabilistically by finding its core point, which in turn is determined using regression.

3.1.2 Organization of Rest of Chapter

Previous work on characterizing minutiae distribution and core point determination are described in Section 3.2. Using the core point to transform the coordinate system into a standard position and determining the core point of a latent print image is described in Section 3.3. Modeling the distribution of minutiae is considered next where a Bayesian network is used to take into account minutia dependencies (Section 3.4); the model is validated using



(a) (b) (c)

(d) (e) (f)

Figure 3.2: Rarity of point sets: (a-d) configurations used as training set, where each has a core point (denoted as a circle), (e) an input configuration that is determined to be commonly occurring, and (f) an input that is rare. Rarity of each input configuration is computed using not only spatial information of points with respect to the core point but also an attribute of orientation associated with each point.

a goodness-of-fit test. As a measure rarity, the probability of random correspondence of a specific print among n samples is defined in Section 3.5. It takes into account uncertainty associated with the core point and the minutiae. Several examples of evaluation of rarity of latent prints and minutia configurations are given in Section 3.6.

The overall process of rarity evaluation is shown in Figure 3.3. Core point prediction and minutia distribution involve learning from sample sets.

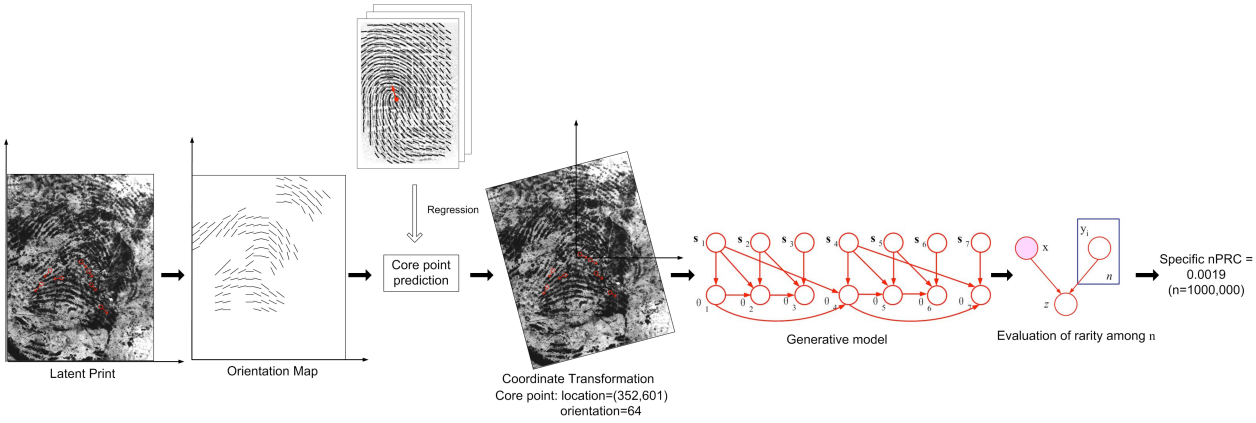


Figure 3.3: Process of rarity evaluation. First, the latent print is processed to get its orientation map. Regression is used to predict the core point. Next the coordinate system is transformed with core point as origin. A generative model is used to provide a distribution for the sequence of minutiae observed. Finally a probability of random correspondence for the specific configuration of minutiae in a database of n entries is evaluated.

3.2 Related Work

3.2.1 Relation to Fingerprint Individuality

Relevant to the task of determining the rarity of a given fingerprint are several classical studies on fingerprint *individuality* [53]. Dating to over a hundred years, the goal of these studies was to determine the degree to which fingerprints are unique. Galton [18] computed the probability of a fingerprint as a product of three factors: factor A is that the configuration is present/absent in each of 24 six-ridge square regions of the fingerprint, factor B is that the region is one of sixteen pattern types and factor C is that the correct number of ridges would enter and exit each region. The three factors are evaluated to have values $(1/2)^{24}$, $(1/16)$ and $(1/256)$. Their product gives the probability of any given fingerprint as 10^{-11}. This model was considered to be a gross underestimate of the variability of fingerprints and improved by Pearson [36] by replacing the $(1/2)$ term by $(1/36)$ to reflect different minutia configurations,

yielding a probability of 1.09×10^{-41}. These evaluations remained untested and subsequently replaced by methods such as those based on minutia types. Stoney and Thornton [50, 51] proposed a set of desired features for a fingerprint model and attempted to meet some conditions using their survey of minutiae and subsequent data analysis. Two limitations to Stoney and Thornton were: the limited scope of the survey and lack of accommodation for prints in poor condition. Champod and Margot [10] presented a statistical model that utilized computer-generated frequencies of minutiae occurrence and minutia densities. A new variable was introduced as compound minutia length. Known weaknesses of the model include: position of the print on the finger must be known, and not allowing connective ambiguities.

3.2.2 Core point determination

A core point corresponds to the center of the north most loop type singularity. For fingerprints that do not contain loop or whorl singularities, the core is usually associated with the point of maximum ridge line curvature[22]. The standard approach for core point detection is the Poincare Index (PI) [26, 5, 23]. Another method, based on a sine map, is realized by multi-resolution analysis [24]. Methods using Fourier expansion[37], fingerprint structures [67] and multi-scale analysis [28] have also been proposed. All of these methods, which are inspired by computational vision, require that the fingerprint is complete and that the core point is present somewhere in it. This assumption does not hold for latent fingerprints, which are usually partial and do not contain core points. So there is no way to detect them by any of the proposed computational vision approaches.

3.2.3 Minutiae Distribution

Several efforts have been made to characterize the distribution of fingerprint features. Since fingerprints can be generated from these models they may be termed as generative. The generative model for level 1 detail simply consists of the probabilities for each ridge flow pattern such as right loop (30%), left loop (27%), double loop (7%), arch (13%), tented arch (5%) and whorl (19%) [46].

Modeling level 2 detail is more important, since fingerprint identification is largely based on features at that level, viz., minutiae. The minutiae correspond to ridge endings and ridge bifurcations. A minutia is represented by its location and direction. The direction is determined by the ridge at the location. Automatic fingerprint matching algorithms use minutiae as the salient features [68], since they are stable and are reliably extracted.

However a generative model becomes much more complex due to the large and variable number of features involved. The simplest model assumes that minutiae locations and orientations are uniformly and independently distributed [35]. An improved model assumes that although minutiae are independent of each other minutiae orientation and location were dependent. Such a model is better than mixtures of hyper-geometric and binomial distributions. A mixture model to account for the clustering tendency of minutiae was proposed [71]. A Markov point process to model minutia location has been attempted [11] with shortcomings, e.g., direction is not incorporated, being based on relative spatial relationships minutia sets from different regions will have the same rarity, and pair potential cannot be expressed in closed form making usability impractical.

There have also been efforts to model distributions beyond minutiae alone. Since latent

print examiners rely on minutiae as well as ridge information, ridges can be represented discretely as ridge points and the model accounts for both minutiae and ridge points [54]. Extending to level 3 detail, a model that incorporates minutiae, ridge and pore features has been recently proposed[14].

Minutiae that are spatially close tend to have similar directions with each other [41]. Moreover, fingerprint ridges flow smoothly with very slow orientation change. The variance of the minutiae directions in different regions of the fingerprint are dependent on both their locations and location variance [51, 12]. These observations on the dependency between minutiae need to be accounted for in eliciting reliable statistical models. The proposed generative model of fingerprint incorporates the distribution of minutiae and the dependency relationship between them.

All existing models for level 2 detail have the drawback of assuming independence of minutiae which leads to inaccurate probability estimates. This paper describes an effort to incorporate minutiae dependencies as well as minutiae uncertainties.

3.3 Core Point Transformation

Latent prints usually correspond to only a small portion of the complete fingerprint. Thus feature sets, such as minutiae, extracted from the print contain only relative spatial relationships. Feature sets with the same relative spatial relationship can lead to different rarity if they come from different areas of the finger. To solve this problem of localization, we first predict the center of the fingerprint, as defined by its core point, and then align the fingerprint by translating the coordinates so that the center is located at the core point and

rotate the coordinates so that the core point orientation points north [56].

3.3.1 Orientation Map

Since ridge flow directions reveal intrinsic features of ridge topologies, they have a critical impact on the core point. A fingerprint field orientation map is defined as a collection of two-dimensional direction fields. It represents the directions of ridge flows in a regularly spaced grid. The gradients of gray intensity of enhanced fingerprints are estimated to obtain reliable ridge orientation [23].

The input fingerprint image is divided into blocks of size $W \times W$ where W is the size in pixels , e.g., $W = 10$. Then compute the gradients G_x and G_y, the gradient magnitudes in the x and y directions, at each pixel in each block. Finally, estimate the local orientation of each block using:

$$\theta_o = \frac{1}{2} tan^{-1}\left(\frac{\sum_{i=1}^{W}\sum_{j=1}^{W} 2G_x(i,j)G_y(i,j)}{\sum_{i=1}^{W}\sum_{j=1}^{W}(G_x^2(i,j) - G_y^2(i,j))}\right). \tag{3.1}$$

An example orientation map is given in Figure 3.4. In an image who field of view includes objects other that the latent print, extraneous regions are manually erased.

3.3.2 Gaussian Process (GP) Regression

Determining the core point of a latent print can be formulated as a regression problem where the independent variable consists of the fingerprint field orientation map and the target variable is the core point. Using a machine learning approach [6], the regression function can be learnt from training data consisting of orientation maps and corresponding core points.

(a)

0 0 0 34 32 33 32 38 42 35 29 32 24 24 22 21 17 10 5 169 143 131 135 131 131 127 139
146 0 0 0 0 0 0 0 32 33 33 30 34 39 45 27 33 31 22 22 29 15 11 7 176 150 141 134 130
136 119 129 131 118 0 0 0 0 0 29 30 31 35 36 35 39 40 33 33 36 24 21 34 25 17 10 1 173
151 140 132 136 122 129 113 110 0 0 0 0 0 34 32 33 31 30 39 41 43 37 32 31 19 23 26
16 9 8 178 161 146 136 125 129 116 134 127 121 0 0 0 0 0 41 38 36 30 33 44 43 43 39
35 31 26 23 26 19 5 1 176 164 156 136 121 129 114 121 110 102 0 0 0 · · · · · ·

(b)

Figure 3.4: Fingerprint orientation map: (a) fingerprint image and (b) vector of gradient values. The corresponding core point has value of $s = (253, 221)$, and $\theta = 85$.

The *Gaussian process (GP)* approach to regression is particularly suitable. GPs dispense with the parametric model and instead define a probability distribution over functions directly. It provides more flexibility and is a better predictor than other approaches such as linear regression, neural networks, etc. The GP model also has the advantage of a Bayesian formulation[39], where instead of a point estimate, the prediction is in the form of a full distribution.

Let the training set $\mathcal{D}$ consist of N fingerprints, $\mathcal{D} = \{(\mathbf{g}_i, y_i)|i = 1, \ldots, N\}$, where $\mathbf{g}_i$ is an orientation map with core point y_i. The regression model with Gaussian noise is given by $y_i = f(\mathbf{g}_i) + \epsilon$ where $f(\mathbf{g}_i)$ is the value of the process or function at $\mathbf{g}_i$ and ϵ is a random noise variable whose value is chosen independent for each observation. Assuming a noise process with Gaussian distribution $p(y_i|f(\mathbf{g}_i)) = \mathcal{N}(f(\mathbf{g}_i), \sigma^2)$, where σ^2 is the variance of

noise, the likelihood function is given by

$$p(\mathbf{y}|\mathbf{f}) = \mathcal{N}(\mathbf{f}, \sigma^2 I) \tag{3.2}$$

where $\mathbf{y} = (y_1, ..y_N)^\top$ is the observed (learning) set of core points and $\mathbf{f} = (f(\mathbf{g}_1), ..f(\mathbf{g}_N))^\top$ contains the corresponding orientation maps $\mathbf{g}_i, i = 1, .., N$.

From the definition of a GP, its prior is given by a Gaussian whose mean is zero and covariance is defined by a covariance function $k(\mathbf{g}, \mathbf{g}')$ so that $f(\mathbf{g}) \sim \mathcal{GP}(0, k(\mathbf{g}, \mathbf{g}'))$. A GP with Gaussian kernel is used to specify the covariance between pairs of variables. $k(\mathbf{g}, \mathbf{g}') = \exp(-||\mathbf{g} - \mathbf{g}'||^2/2)$.

Core Point Distribution

For an input orientation map $\mathbf{g}^*$ the predictive distribution of core point y^* can be evaluated by conditioning the joint Gaussian prior distribution on the observation $(G, \mathbf{y})$, where $G = (\mathbf{g}_1, \ldots, \mathbf{g}_N)^\top$. The predictive distribution is given by

$$p(y^*|\mathbf{g}^*, G, \mathbf{y}) = \mathcal{N}(m(y^*), cov(y^*)) \tag{3.3}$$

where $m(y^*) = \mathbf{k}(\mathbf{g}^*, G)[K + \sigma^2 I]^{-1}\mathbf{y}$,

$$cov(y^*) = k(\mathbf{g}^*, \mathbf{g}^*) + \sigma^2 - \mathbf{k}(\mathbf{g}^*, G)^\top [K + \sigma^2 I]^{-1}\mathbf{k}(G, \mathbf{g}^*),$$

$\mathbf{k}(\mathbf{g}^*, G) = (k(\mathbf{g}^*, \mathbf{g}_1), .., k(\mathbf{g}^*, \mathbf{g}_N))^\top$ and K is the Gram matrix with elements $k(\mathbf{g}_i, \mathbf{g}_j)$.

Point Estimate of Core Point

Rather than work with the distribution of the core point we can work with the maximum a posteriori probability (MAP) solution. Since $\mathbf{g}^*$ may represent the orientation map in one of several possible locations we maximize among all m possible translations and rotations over the set $\{\mathbf{g}_i^* | i = 1, \ldots, m\}$. Using Eq. (3.3), we obtain the predictive distributions $p(y^*|\mathbf{g}_i^*, G, \mathbf{y})$ for all $\mathbf{g}_i^*$. The core point $\hat{y}^*$ should maximize $p(y^*|\mathbf{g}_i^*, G, \mathbf{y})$ with respect to $\mathbf{g}_i^*$. Thus the core point is given by

$$\hat{y}^* = \mathbf{k}(\mathbf{g}_{MAX}^*, G)[K + \sigma^2 I]^{-1}\mathbf{y} \tag{3.4}$$

where $\mathbf{g}_{MAX}^*$, the orientation map corresponding to the most probable core point, is given by

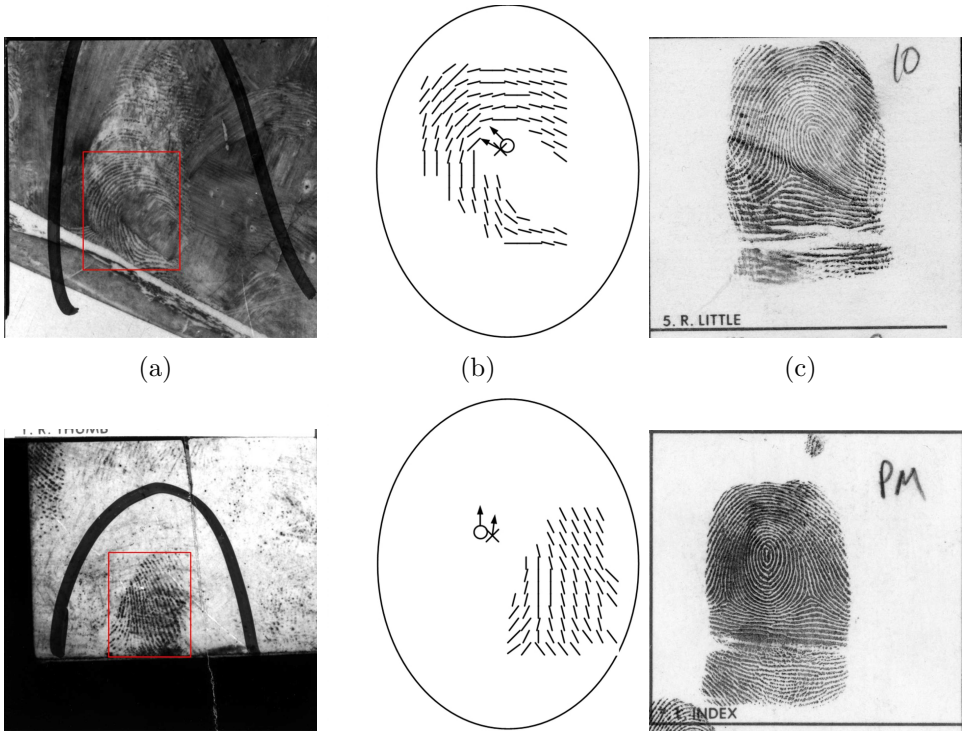
$$\mathbf{g}_{MAX}^* = \underset{\mathbf{g}^*}{\operatorname{argmax}}\ p(m(y^*)|\mathbf{g}^*, G, \mathbf{y}) \tag{3.5}$$

The point estimate is used in evaluating the performance of the GP method in Section 3.3.3 and in the coordinate transformation process described in Section 3.3.4.

3.3.3 Performance of GP Core Point Estimation

The GP model was trained on the *NIST 4* database of fingerprint images [69] and tested on the *NIST 27* set [19] [1]. Ridge orientation maps were extracted from the fingerprint images using the gradient-based approach defined by Eq. 3.1. The images were first divided into

[1] *NIST 4* contains 8-bit gray scale images of randomly selected fingerprints. Each print has 512×512 pixels. The entire database contains fingerprints taken from 2000 different fingers with 2 impression of the same finger. The database is evenly distributed over each of the five classifications with 400 fingerprint pairs from each class. *NIST 27* contains latent fingerprints from crime scenes and their matching rolled fingerprint mates. There are 258 latent cases separated into three quality categories of good, bad, and ugly.



(a) (b) (c)

(d) (e) (f)

Figure 3.5: Core point prediction using GP regression for two latent prints (from *NIST27*): (a) image *g90* which contains a latent print within a rectangle, (b) computed orientation map containing the predicted core point (cross) and true core point (circle), (c) ten-print where the true core point is visible, (d-f) image *g69* with corresponding images. Note that in the second case the predicted core point lies outside of the latent print.

equal-sized blocks of $N \times N$ pixels, where N is the average width of a pair of ridge and valley. The value of N is 8 in *NIST 4* and varies in *NIST 27*. The gradient vectors were calculated by taking the partial derivatives of image intensity at each pixel in Cartesian coordinates. Ridge orientation is perpendicular to the dominant gradient angle in the local block. The training set consisted of orientation maps whose corresponding core points were manually marked.

The learnt GP model was applied to three groups of latent prints in different quality, *good*, *bad* and *ugly*. Figure 3.5 shows the results of core point prediction and subsequent latent print localization given two latent fingerprints from *NIST 27*. For each pixel in the smoothed orientation field the PI at pixel (i,j) is defined with respect to a digital curve which consists of a sequence of pixels that are on or within a distance of one pixel from the curve. It takes the form $PI(i,j) = \frac{1}{2\pi}\sum_{k=0}^{N_\psi - 1} \Delta(k)$ where $\Delta(k)$ are orientation differences between neighboring pixels on the curve with N_ψ pixels. Assign the corresponding pixel a label 1 if its $PI = 1/2$. For each labeled connected component, if its area is larger than 7, a core point is detected at the centroid of the connected component. Further details of the PI algorithm can be found in [23].

Performance of the GP point estimate was compared to that provided by the baseline Poincare Index (PI) method [5] which is based on purely local topological considerations. To evaluate performance of both methods, test latent prints were extracted from the image database with extraneous regions manually erased. The true core point of each latent print was determined from its matching 10-print in the database. Prediction accuracy was determined by comparing the location and direction distances between predicted and true core points with the threshold parameters set at $T_s = 16$ pixels, and $T_\theta = \pi/6$.

Table 3.1: Core point prediction accuracy of standard and proposed methods.

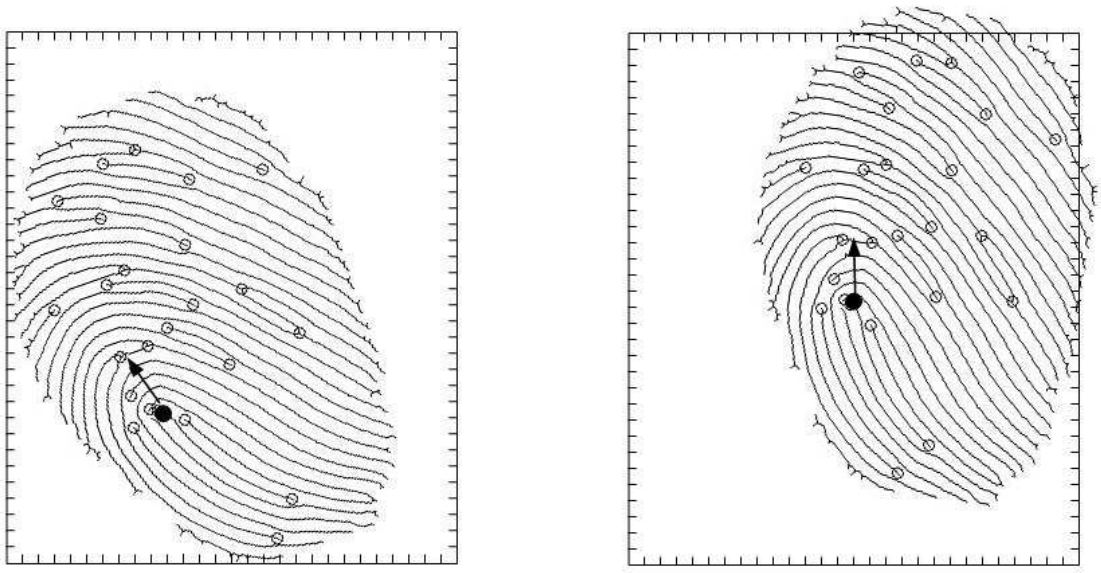
	Poincare Index(PI)	Gaussian Process(GP) Point Estimate
Good	91%	93%
Bad	68%	87%
Ugly	47%	73%
Overall	69%	85%

Prediction accuracies of the PI and GP approaches are given in Table 3.1. The good set has 88 images that mostly contain core points. Both bad and ugly sets contain 85 images of small size that usually do not include core points. For good prints, the two approaches are close. For the bad and ugly prints there is a distinct difference between the methods with GP predicting core points even when it is absent in the latent prints. The GP method also results in higher overall performance.

Since the overall error rate in the point estimate is still as high as 15%, the use of a distribution in further analysis is preferable. A disadvantage of GP core point prediction is its $O(N^3)$ complexity, where N is the number of finger prints in the training set; due to an inversion of the $N \times N$ covariance matrix. However, more efficient GP implementations are available [44, 42].

3.3.4 Coordinate Transformation

After the core point is determined, with a point estimate, the Cartesian coordinate system is transformed such that the origin is the core point and the core point orientation points to $\pi/2$ (Fig. 3.6). Given a minutia $(\mathbf{s}, \theta)$ and predicted core point $(\mathbf{s}^*, \theta^*)$, where $\mathbf{s} = (s_x, s_y)$ represents location and θ the direction, the transformed minutia $(\mathbf{s}', \theta')$ is given by



(a) (b)

Figure 3.6: Fingerprint coordinate transformation based on core point: (a) original fingerprint image with minutiae (represented by circles) and core point (dot and arrow), and (b) fingerprint image after translation and rotation of the core point to the center.

$$s'_x = s_x - s^*_x - sin(\theta - \theta^*) \, \|s - s^*\|$$

$$s'_y = s_y - s^*_y - cos(\theta - \theta^*) \, \|s - s^*\|$$

$$\theta' = \theta - \theta^* + \pi/2$$

Distribution of the extent of the transformation on the $4,000$ *NIST 4* images are given in Appendix 1, for translation δ_{xy} and rotation δ_θ. In subsequent modeling of minutiae distributions, the coordinate system is transformed first.

3.4 A Generative Model for Minutiae

Next is the task of modeling the joint distribution of minutiae. Each minutia is represented as $\mathbf{x} = (\mathbf{s}, \theta)$ where $\mathbf{s} = (s_x, s_y)$ is its location and θ its direction.

We begin with a model for the distribution of individual minutiae $p(\mathbf{x})$ and then consider modeling the joint distribution of a set of minutiae $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ where we do not wish to assume minutiae independence.

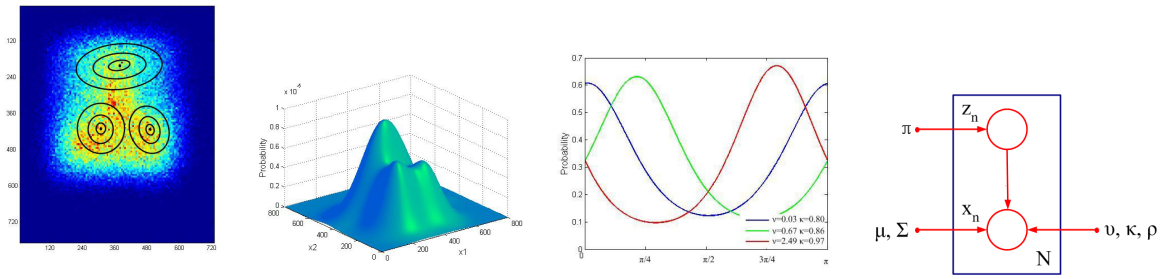
3.4.1 Marginal Distribution of Minutiae

An obvious and effective model for minutiae location is to choose a mixture of Gaussians [71, 58]. For minutiae orientation, which is an angular distribution, the circular normal or von Mises distribution [6, 30] is useful.

The distribution of minutiae location is shown in Fig. 3.7(a); the minutiae data are from 2,000 fingerprints in the *NIST4* database. This multimodal distribution is naturally modeled as a mixture of k Gaussians (Fig. 3.7(b)) with $k = 3$. Minutiae orientation θ is modeled by the von Mises distribution as shown in Fig. 3.7(c). A simple Bayesian network to represent the marginal distribution of individual points as a mixture model is shown in Figure 3.7 (d).

3.4.2 Joint Distribution of Minutiae

It is unsatisfactory to model the joint distribution of minutiae as a product of their marginal distributions particularly when significant dependences exist. In the case of fingerprints, studies indicate that minutiae direction is related to its own location as well as the location of neighboring minutiae [41, 51, 12]. These known dependences can be readily incorporated

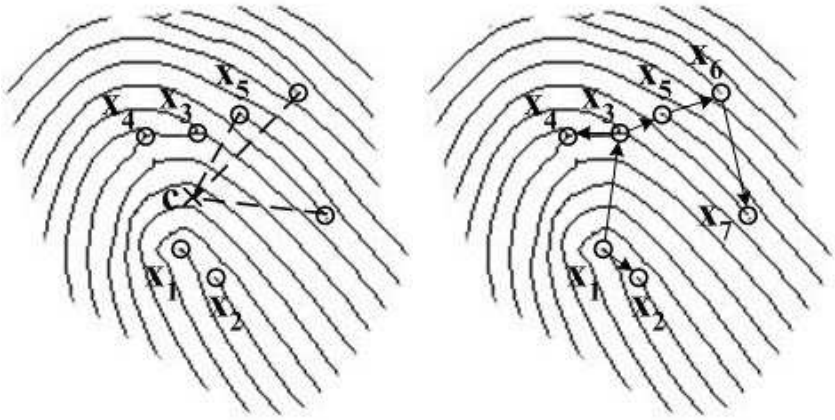


(a) (b) (c) (d)

Figure 3.7: Gaussian mixture model of location and orientation: (a) Distribution of spatial location $s = (x_1, x_2)$ is modeled by a mixture of three bivariate Gaussians whose contours of constant density are shown (b) 3D plot of mixture model for location, (c) von Mises distributions of orientation θ for each of the three components, where the green curve corresponds to the upper cluster, blue the lower left cluster and red the lower right cluster, and (d) graphical model of mixture where $\mathbf{z}_n$ are latent variables corresponding to mixture components and parameters are as in Eq. 3.8.

into a causal Bayesian network. Conditional dependences of minutiae can be incorporated by linearizing them by defining a unique sequence of points. Linearization is used in agglomerative clustering of N points where a point is assigned to a cluster whose mean is nearest; where nearest can be defined in terms of distance to the mean (centroid), closest or furthest point of the cluster[16].

The sequence starts with the point $\mathbf{x}_1$ whose location is closest to the center (core point). Each remaining point $\mathbf{x}_n$ is the spatially closest to the centroid defined by the arithmetic mean of the location coordinates of all the previous points $\mathbf{x}_1, \ldots \mathbf{x}_{n-1}$. Given this sequence, the fingerprint can be represented by a minutiae sequence $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$. The sequence is robust to the variance of the minutiae because the next minutia is decided by the all the previous minutiae. Given the observation that spatially closer minutiae are more strongly related, we only model the dependence between $\mathbf{x}_n$ and its nearest minutiae among $\{\mathbf{x}_1, \ldots, \mathbf{x}_{n-1}\}$. Although not all dependences are taken into account, this is a good trade-off between model



(a) (b)

Figure 3.8: Sequential ordering of minutiae: (a) given minutiae $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4\}$ with centroid c, the next minutia $\mathbf{x}_5$ is selected by comparing the remaining minutia distances to c, thereby providing a sequencing, (b) dependency between the sorted minutiae is represented by arrows.

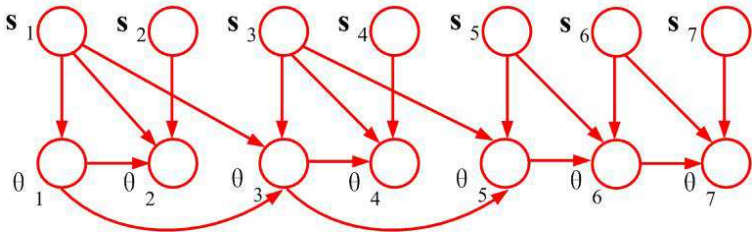


Figure 3.9: Bayesian network to represent the joint distribution of minutiae: this graphical model corresponds to example in Figure 3.8(b). Minutiae locations are represented by nodes labeled $\mathbf{s}_n$ and corresponding orientations are represented by nodes labeled θ_n. This joint distribution can be written directly from the model as $p(\mathbf{s}_1)p(\theta_1|\mathbf{s}_1)p(\mathbf{s}_2)p(\theta_2|\mathbf{s}_1, \mathbf{s}_2, \theta_1)p(\mathbf{s}_3)p(\theta_3|\mathbf{s}_1, \mathbf{s}_3, \theta_1).....$

accuracy and computational complexity. Figure 3.8(a) presents an example where $\mathbf{x}_5$ is determined because its distance to the centroid of $\{\mathbf{x}_1, \ldots, \mathbf{x}_4\}$ is minimal. Figure 3.8(b) shows the minutiae sequence and the minutiae dependency (arrows) for the same configuration of minutiae.

A directed graphical model, or Bayesian network, can be used to represent probabilistic relationships between minutiae. Examples of such relationships are: minutia orientation is dependent on near minutiae (both location and orientation), minutia location is conditionally

dependent on the location of neighboring minutiae given their directions. A graphical model for the minutiae set in Figure 3.8 is given in Figure 3.9, where nodes $\mathbf{s}_n$ and θ_n represent the locations and directions of minutia $\mathbf{x}_n$. For each conditional distribution, a directed link is added from nodes corresponding to the variables on which the distribution is conditioned. A general expression for the joint distribution of any minutiae set $\mathbf{X}$ is

$$\begin{aligned} p(\mathbf{X}) &= p(\mathbf{s}_1)p(\theta_1|\mathbf{s}_1)\prod_{n=2}^{N} p(\mathbf{s}_n)p(\theta_n|\mathbf{s}_n, \mathbf{s}_{\psi(n)}, \theta_{\psi(n)}) \\ &= p(\mathbf{s}_1, \theta_1)\prod_{n=2}^{N} p(\mathbf{s}_n)p(\theta_n|\mathbf{s}_n, \mathbf{s}_{\psi(n)}, \theta_{\psi(n)}) \end{aligned} \tag{3.6}$$

where $\mathbf{s}_{\psi(n)}$ and $\theta_{\psi(n)}$ are the location and direction of minutiae $\mathbf{x}_i$ which have the minimal spatial distance to minutia $\mathbf{x}_n$, and $\psi(n) = \underset{i\in[1,n-1]}{\operatorname{argmin}}\|\mathbf{x}_n - \mathbf{x}_i\|$. To compute the joint probability $p(\mathbf{X})$, three distributions are needed: $p(\mathbf{s})$, distribution of minutia location, $p(\mathbf{s}, \theta)$, the distribution of minutia location and direction, $p(\theta_n|\mathbf{s}_n, \mathbf{s}_{\psi(n)}, \theta_{\psi(n)})$, the conditional distribution of minutiae direction given its location, and the location and direction of the nearest minutiae. Each of these are addressed next.

Minutia location. Since minutiae tend to form clusters [41] a mixture of Gaussians, with K_1 components, is used:

$$p(\mathbf{s}) = \sum_{k_1=1}^{K_1} \pi_{k_1}\mathcal{N}(\mathbf{s}|\mu_{k_1}, \Sigma_{k_1}). \tag{3.7}$$

Minutiae location and direction. Since minutiae in different regions are associated with different region-specific directions, a mixture of joint Gaussian and von-Mises distributions

with K_2 components is used:

$$p(\mathbf{s}, \theta) = \sum_{k_2=1}^{K_2} \pi_{k_2} \mathcal{N}(\mathbf{s}|\mu_{k_2}, \Sigma_{k_2}) \mathcal{V}(\theta|\nu_{k_2}, \kappa_{k_2}). \tag{3.8}$$

While we can get $p(\mathbf{s})$ by marginalizing $p(\mathbf{s}, \theta)$ for each continuous value of $\mathbf{s}$, it is time consuming and unnecessary and thus Eq 3.7 is used.

Conditional minutia direction. Minutiae direction, given its location and the the nearest minutiae among $\mathbf{x}_1$ to $\mathbf{x}_{n-1}$, is a mixture of von-Mises densities with K_3 components:

$$p(\theta_n|\mathbf{s}_n, \mathbf{s}_{\psi(n)}, \theta_{\psi(n)}) = \sum_{k_3=1}^{K_3} \pi_{k_3} \mathcal{V}(\theta_n|\nu_{k_3}, \kappa_{k_3}). \tag{3.9}$$

where $\pi_{k_i}$are non-negative component weights that sum to one, $\mathcal{N}(s|\mu_k, \Sigma_k)$ is the bivariate Gaussian probability density function of minutiae with mean μ_k and covariance matrix Σ_k, and $\mathcal{V}(\theta|\nu_k, \kappa_k)$ is the von-Mises probability density function of minutiae orientation with mean angle ν_k and precision (inverse variance) κ_{k_3} [6]

$$\mathcal{V}(\theta|\nu_k, \kappa_k) = \frac{1}{2\pi I_0(\kappa_k)} exp[\kappa_k cos(\theta - \nu_k)]. \tag{3.10}$$

The number of components K_i in each of the mixture distributions can be determined using the Bayes information criterion (BIC). Other parameters are determined using the EM algorithm (see Appendix (1)).

3.4.3 Goodness-of-fit Test

It is necessary to validate the model, using a goodness of fit test, before using it in rarity evaluation. Goodness of fit means how well a sample of data agrees with a given distribution as its population. To test the goodness of fit, the chi-square statistical hypothesis test was used. The test determines whether observed sample frequencies differ significantly from expected frequencies specified in the null hypothesis. The test, applied to binned data, requires a sufficient sample size in each bin in order for the chi-square approximation to be valid [38]. Three different tests were conducted for each of the distributions in Eqs. 3.7, 3.8, and 3.9. For minutiae location, minutiae location space was partitioned into 16 non-overlapping blocks. For minutiae location and orientation, there were 16×4 non-overlapping blocks. For minutia dependency, the orientation space was divided into 9 non-overlapping blocks. The blocks were combined with adjacent blocks until both observed and expected numbers of minutiae in the block were greater than or equal to 5. The test statistic used was the chi-square random variable $\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$ where O_i is the observed minutiae count for the ith block, and E_i is the expected minutiae count for the ith block. The p-value, the probability of observing a sample statistic as extreme as the test statistic, associated with each test statistic χ^2 is then calculated based on the chi-square distribution and compared to the significance level. For the *NIST 4* dataset, we chose significance level equal to 0.01. The generative models were trained using 4000 fingerprints. BIC yielded $K_1 = K_2 = 3$. For K_3 there are 4096 different values for different condition settings.

To test the models for minutia location, and minutia location and orientation, the numbers of fingerprints with p-values above (corresponding to accept the model) and below

Table 3.2: Chi-square goodness of fit test for generative models.

Generative model	Data size	Accept	Reject
$p(\mathbf{s})$	4000	3387	613
$p(\mathbf{s}, \theta)$	4000	3216	784
$p(\theta_n \vert \mathbf{s}_n, \mathbf{s}_{\psi(n)}, \theta_{\psi(n)})$	4096	3558	538

(corresponding to reject the model) the significance level were computed. The results are shown in Table 3.2.

Of the $4,000$ fingerprints, $3,271$ were accepted and 729 rejected for minutia location model, and $3,103$ were accepted and 897 rejected for minutia location and orientation model. To test the model for minutia dependency, we first collected all the linked minutia pairs in the minutia sequences produced from $4,000$ fingerprints. Then these minutia pairs were separated by the binned locations of both minutiae (32×32) and orientation of leading minutiae (4). Finally, the minutia dependency models were tested on corresponding minutia pair sets. Of the $4,096$ dependency models, $3,558$ were accepted and 538 rejected. The results imply that both generative models have a reasonable and accurate fit for fingerprints.

3.5 Rarity Evaluation

The probability that an input $\mathbf{x}$ coincides with one of n samples, within specified tolerance, is defined as the *specific* nPRC [58]. Since it is conditional to the known input it can also be referred to as the *conditional* nPRC. Specific nPRC is different from the probability of random correspondence (PRC), which is the probability that a random pair of samples will have the same value. It is also different from nPRC which is the probability that some pair of samples among n have the same value.

We make this idea precise using the graphical model of Figure 3.10, where $\mathbf{x}$ represents the input in a feature space, and $\mathbf{Y} = \{\mathbf{y}_1, ..., \mathbf{y}_n\}$, in plate notation, represents the set of n database entries. The binary-valued indicator random variable z takes one of two values $\{z^0, z^1\}$. It has value z^0 when at least one database element $\mathbf{y}_i$ has the value $\mathbf{x}$, and z^1 otherwise. The conditional probabilities can be written as $P(z^0|\mathbf{x} \in \mathbf{Y}) = 1, P(z^0|\mathbf{x} \notin \mathbf{Y}) = 0, P(z^1|\mathbf{x} \in \mathbf{Y}) = 0, P(z^1|\mathbf{x} \notin \mathbf{Y}) = 1$. By marginalizing over $\mathbf{Y}$, the specific or conditional nPRC is given by

$$P(z = z^0|\mathbf{x}) = \sum_{\mathbf{Y}} P(z = z^0|\mathbf{x}, \mathbf{Y})P(\mathbf{Y}) \tag{3.11}$$

where $P(\mathbf{Y})$ is the joint probability, or likelihood, of $\mathbf{Y}$. Since the space of $\mathbf{Y}$ consists of all possible entries in the database, which is huge, we use an alternative method of determining the conditional nPRC. Assume that database entries are independent and identically distributed. Since the probability of $\mathbf{x}$ not matching each element in the database is $1 - P(\mathbf{x})$ we have

$$P(z = z^0|\mathbf{x}) = 1 - (1 - P(\mathbf{x}))^n \tag{3.12}$$

If $\mathbf{x}$ is continuous with distribution $p(\mathbf{x})$ the conditional nPRC is the probability of finding an element among n within tolerance ϵ is

$$P(\mathbf{x}, \mathbf{Y} : \epsilon, n) = 1 - (1 - p_\epsilon(\mathbf{x}))^n \tag{3.13}$$

where p_ϵ is the probability of $\mathbf{x}$ in the interval $\mathbf{x} \pm \epsilon$.

Minutia tolerance is defined as follows. Minutia pair $\mathbf{x}_a = (\mathbf{s}_a, \theta_a)$ and $\mathbf{x}_b = (\mathbf{s}_b, \theta_b)$

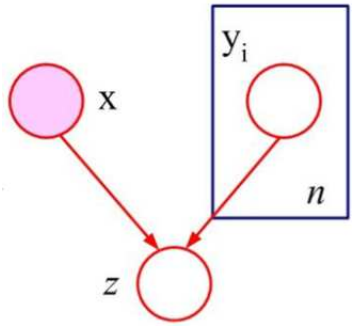


Figure 3.10: Graphical model for rarity. Specific nPRC is the conditional probability that a known $\mathbf{x}$ is found among at least one of $\mathbf{y_1}, .., \mathbf{y_n}$, where z is an indicator variable for a match.

correspond if for tolerance $\epsilon = [\epsilon_s, \epsilon_\theta]$

$$\| \mathbf{s}_a - \mathbf{s}_b \| \leq \epsilon_s \wedge |\theta_a - \theta_b| \leq \epsilon_\theta \tag{3.14}$$

where $\|\mathbf{s}_a - \mathbf{s}_b\|$ is the Euclidean distance between their locations. A match implies the existence of at least $\hat{m}$ corresponding pairs for specified ϵ and $\hat{m}$.

Since $\mathbf{s}$ and θ are continuous-valued, and tolerance is considered, the evaluation of specific nPRC involves integrating over the relevant distributions. Also, since minutiae quality varies greatly in latent images, it is useful to take into account minutia confidence in specific nPRC evaluation. We describe next these evaluations with and without consideration of uncertainty in minutiae and core point.

3.5.1 Minutiae and Core Point with Point Estimates

Here we assume that the core point and minutiae have no associated uncertainty, i.e., they are point estimates. For the core point the MAP estimate, as provided by the GP in Eq. 3.4, is used and for detected minutiae the confidence values are ignored.

Let $\mathbf{X}$ be a fingerprint with N minutiae. Let $\mathbf{X}'$ be another fingerprint with M minutiae. Let $\widetilde{\mathbf{X}}$ be the set of $\hat{m}$ minutiae common to both $\mathbf{X}$ and $\mathbf{X}'$ where the matching minutiae are within ϵ of each other. The distribution of the common points can be written from Eq 3.6 as

$$p_\epsilon(\widetilde{\mathbf{X}}) = p_\epsilon(\mathbf{s}_1, \theta_1) \prod_{n=2}^{\hat{m}} p_\epsilon(\mathbf{s}_n) p_\epsilon(\theta_n | \mathbf{s}_n, \mathbf{s}_{\psi(n)}, \theta_{\psi(n)}) \tag{3.15}$$

where the three product terms are expanded as

$$p_\epsilon(\mathbf{s}_n, \theta_n) = \int\limits_{|\mathbf{s}-\mathbf{s}'_n| \le \epsilon_\mathbf{s}} \int\limits_{|\theta-\theta'_n| \le \epsilon_\theta} p(\mathbf{s}, \theta) d\mathbf{s} d\theta,$$

$$p_\epsilon(\mathbf{s}_n) = \int\limits_{|\mathbf{s}-\mathbf{s}'_n| \le \epsilon_\mathbf{s}} p(\mathbf{s}) d\mathbf{s}, \text{ and}$$

$$p_\epsilon(\theta_n | \mathbf{s}_n, \mathbf{s}_{\psi(n)}, \theta_{\psi(n)}) = \int\limits_{|\theta-\theta'_n| \le \epsilon_\theta} p(\theta | \mathbf{s}_n, \mathbf{s}_{\psi(n)}, \theta_{\psi(n)}) d\theta$$

The specific nPRC is computed using Eq. (3.13) by

$$p_\epsilon(\mathbf{X}, \hat{m}, n) = 1 - (1 - p_\epsilon(\mathbf{X}, \hat{m}))^n \tag{3.16}$$

where $p_\epsilon(\mathbf{X}, \hat{m})$, the probability that $\hat{m}$ pairs of minutiae correspond, is given by

$$p_\epsilon(\mathbf{X}, \hat{m}) = \sum_{m' \in M} p(m') \binom{m'}{\hat{m}} p_\epsilon(\widetilde{\mathbf{X}}) \tag{3.17}$$

where M contains all possible numbers of minutiae in one fingerprint among n fingerprints, $p(m')$ is the probability of a random fingerprint having m' minutiae, minutiae set $\widetilde{\mathbf{X}} =$

$(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{\hat{m}})$ is the subset of $\mathbf{X}$ and $p_\epsilon(\widetilde{\mathbf{X}})$ is the joint probability of minutiae set $\widetilde{\mathbf{X}}$ given by Eq. (3.15).

3.5.2 Minutiae Uncertainty

Next we consider a model for uncertainty associated with minutiae. These arise from confidence values assigned either by a human examiner or by an AFIS system.

Assume that confidence of minutia $\mathbf{x}_n$ is given by (d_{s_n}, d_{θ_n}), where d_{s_n} is location confidence and d_{θ_n} is direction confidence. Given minutiae $\mathbf{x}_n = (s_n, \theta_n)$ and their confidences, the distributions of location s' and direction θ' can be modeled by Gaussian and von-Mises distributions

$$c(\mathbf{s}'|\mathbf{s}_n, d_{\mathbf{s}_n}) \sim \mathcal{N}(\mathbf{s}'|\mathbf{s}_n, d_{\mathbf{s}_n}^{-1}) \tag{3.18}$$

$$c(\theta'|\theta_n, d_{\theta_n}) \sim \mathcal{V}(\theta'|\theta_n, d_{\theta_n}) \tag{3.19}$$

where the precision (inverse variance) of location distribution d_{s_n} represents location confidence and the concentration parameter of direction distribution d_{θ_n} represents direction confidence.

Ranges for these values can be assigned based on image resolution, e.g., since ridges are ten pixels wide in the *NIST* dataset, location confidence d_{s_n} is in the interval $[0.01, 1]$ and orientation confidence d_{θ_n} is in the interval $[1, 10]$, where a high confidence value implies a high quality minutia.

By application of the sum rule of uncertainty of minutiae, the conditional distributions involved in Eq. (3.15) are given by

$$p_\epsilon(\mathbf{s}_n, \theta_n)$$
$$= \int_{s'} \int_{\theta'} \iint_{|\mathbf{x}-\mathbf{x}'|\leq\epsilon} c(\mathbf{s}'|\mathbf{s}_n, d_{s_n}) c(\theta'|\theta_n, d_{\theta_n}) p(\mathbf{s}, \theta) d\mathbf{s}' d\theta' d\mathbf{s} d\theta$$

$$p_\epsilon(\mathbf{s}_n) = \int_{s'} \int_{|\mathbf{s}-\mathbf{s}'|\leq\epsilon_\mathbf{s}} c(\mathbf{s}'|\mathbf{s}_n, d_{\mathbf{s}_n}) p(\mathbf{s}) d\mathbf{s}' d\mathbf{s} \tag{3.20}$$

$$p_\epsilon(\theta_n|\mathbf{s}_n, \mathbf{s}_{\psi(n)}, \theta_{\psi(n)})$$
$$= \int_{\theta'} \int_{|\theta-\theta'|\leq\epsilon_\theta} c(\theta'|\theta_n, d_{\theta_n}) p(\theta|\mathbf{s}_n, \mathbf{s}_{\psi(n)}, \theta_{\psi(n)}) d\theta' d\theta$$

Specific nPRCs can again be computed by Eq. (3.16) and Eq. (3.17). Since confidence distributions are sharply peaked and the definite integral intervals are small, numerical integration can be used in probability calculation.

The complexity of joint probability of a print with $\hat{m}$ matching minutiae is $O(\hat{m})$. The computational cost of specific nPRC for a certain minutia set is $O(M\hat{m})$, where M is the maximum number of minutiae in a fingerprint.

3.5.3 Core Point Uncertainty

Here we take the Bayesian approach where the core point does not have a fixed value but has a distribution instead. Let y be the core point whose distribution is $p(y)$ as given in

Eq. 3.3. Thus the distribution of $\widetilde{\mathbf{X}}$, which is the set of minutiae common to $\mathbf{X}$ and $\mathbf{X}'$, is obtained by integrating out the core point parameter as

$$p_\epsilon(\widetilde{\mathbf{X}}) = \int p_\epsilon(\widetilde{\mathbf{X}}|y)p(y)dy \tag{3.21}$$

where $p_\epsilon(\widetilde{\mathbf{X}}|y)$ is given by Eq.3.15 and the minutia set are subjected to coordinate transformation specified by core point y.

3.6 Examples of Rarity Evaluation

Evaluation of fingerprint rarity is demonstrated with three sets of examples: (i) a few simple minutia configurations where we assume point estimates of core point and minutiae, (ii) the Madrid train bombing case where we use a point estimate of the core point and minutiae with/without confidence values, and (iii) latent prints from a standard data set where we use a fixed core point and an uncertain core point.

As in any machine learning scenario, the evaluation of rarity depends on the data set from which the parameters are determined. The database should be representative enough. We used the largest publicly available database for the learning phase, viz., *NIST special database 4*, which contains $2,000$ 8-bit gray scale fingerprint image pairs.

3.6.1 Simple Minutia Configurations

Determining the rarity of configurations of few minutiae is useful to the latent print examiner who needs to decide whether to proceed further. A minutia configuration involving only three

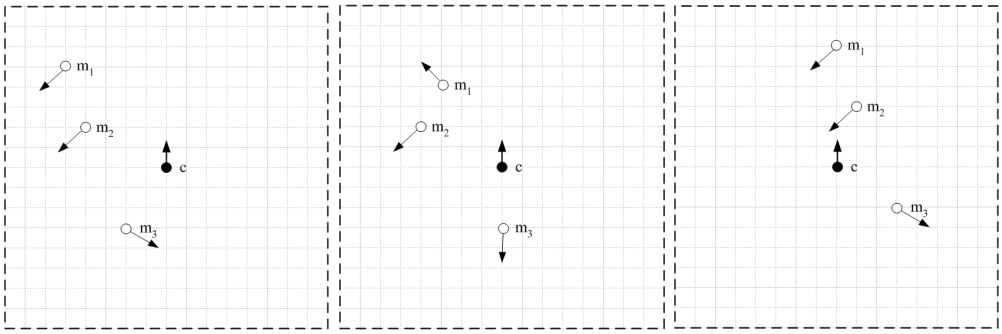


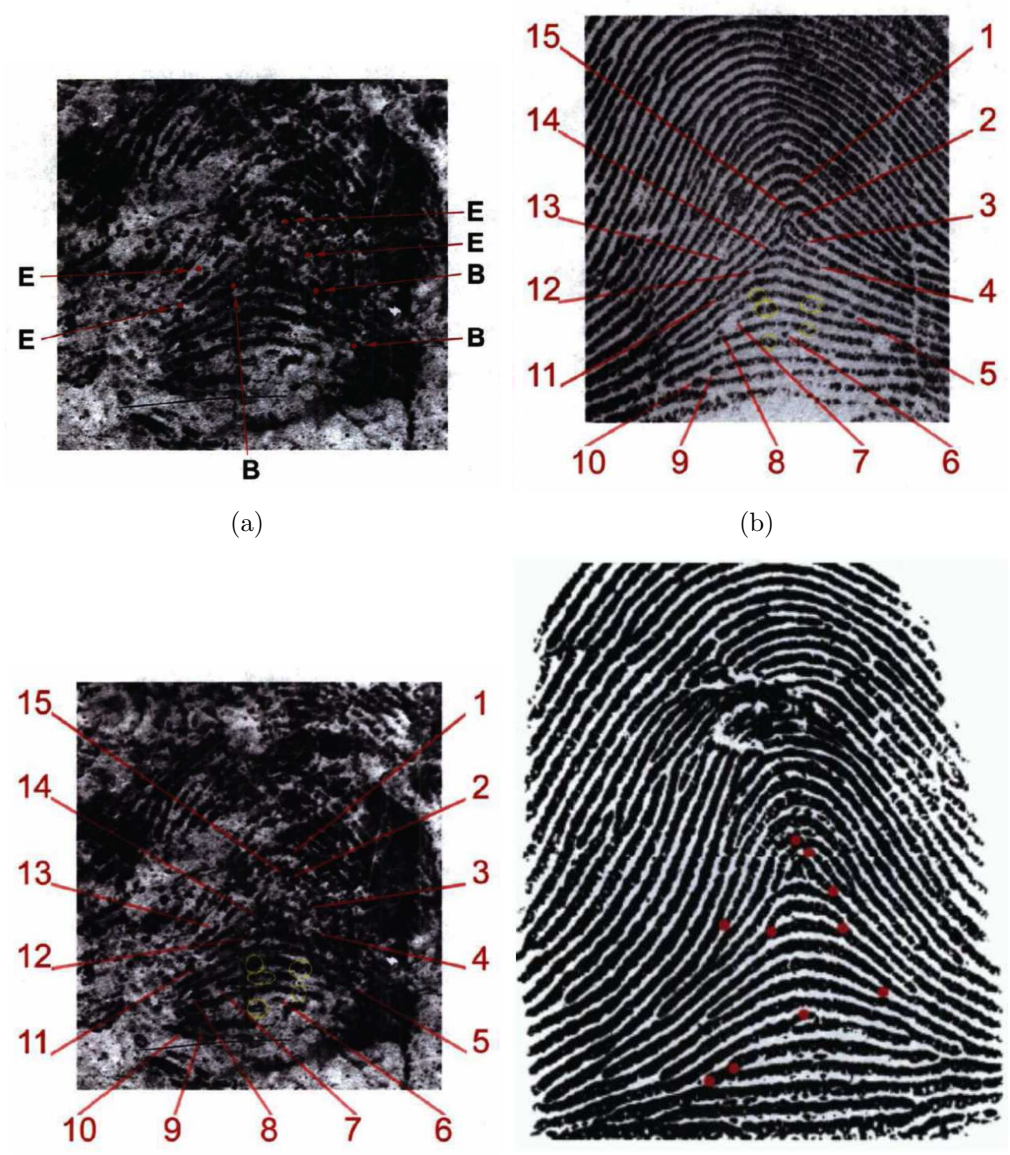
Figure 3.11: Simple configurations of minutia and core points: (a) a common configuration with three minutiae m_1, m_2 and m_3 and core point c, (b) an uncommon configuration obtained by changing the orientations of m_1 and m_3 and (c) an uncommon configuration obtained by translating the three minutiae with respect to the core point. For $n = 1000$, their specific nPRC values are: (a) 1.2×10^{-2}, (b) 7.97×10^{-4} and (c) 2.3×10^{-6} respectively.

minutiae, with no uncertainty in either minutiae or core point, is shown in Figure 3.11. It has a high specific nPRC of 0.012 in a database of $1,000$ entries. When the common minutia structure is perturbed, by changing a few minutiae orientations, a much lower probability of finding a match is observed.

3.6.2 Madrid Bombing Case

The Madrid train bombing case provides a useful test scenario since both the latent print and minutiae annotations of the image are available. Particulars of the case are relevant since they illustrate the value of rarity evaluation.

In 2004 there was a terrorist bombing in the Madrid train system leading to the death of nearly 200 individuals and injury of 2,000 more people. A latent print, tagged as *LFP17*, was found on a plastic bag of detonators in a nearby van (Figure 3.12(a)). This was erroneously identified as corresponding to the inked ten-print of Brandon Mayfield, an attorney in Oregon, in the FBI database of 470 million fingerprints (Figure 3.12(b)). The latent print



(a) (b)

(c) (d)

Figure 3.12: Madrid bombing case latent prints and ten-prints: (a) latent print *LFP17* found at crime scene with seven marked minutiae (initial annotation), (b) matching ten-print of Mayfield found in FBI database with 15 charted minutiae, (c) *LFP17* re-annotated with the same 15 minutiae as in (b), and (d) ten-print of Daoud with ten matching minutiae(from [62]).

Table 3.3: Probability of randomly matching latent print in Madrid bomber case with entry in FBI database.

Minutiae Uncertainty Assumption	**7 Min. in Fig 3.12(a)**	**15 Min. in Fig 3.12(c)**
1. Point Estimates	0.78	1.2×10^{-6}
2. Uncertain Minutiae	0.93	7.8×10^{-7}

was initially marked as having seven identifiable minutiae. Upon observing the Mayfield ten-print (Fig. 3.12(b)), *LFP17* was re-annotated as having 15 minutiae that correspond to the Mayfield ten-print (Figure 3.12(c)). Subsequently Spanish National police identified the true perpetrator as an Algerian national, Ouhnane Daoud, whose ten-print is shown in Fig. 3.12(d). Rarities of the minutiae sets in this problem, evaluated under two scenarios: full confidence of minutiae and minutiae with uncertainty, are given in Table 3.3. In both cases a point estimate for the core point was assumed.

Case 1: Point estimates of minutiae: With $n = 4.7 \times 10^8$ the specific nPRC of the seven minutiae evaluates to 0.78, which is a high probability. For the 15 minutiae in the re-annotated print the specific nPRC is 1.2×10^{-6} which indicates high rarity.

Case 2: Uncertain minutiae: We manually assigned a confidence to each minutia in each set as described in Section 3.5.2. For the seven minutiae set in Figure 3.12(a), confidences (d_s, d_θ) were assigned clockwise starting from 1 o'clock, as follows : (0.1, 7), (0.6, 8), (0.2, 8), (0.7, 6), (0.9, 9), (0.6, 8), (0.9, 9). Specific nPRC evaluates to 0.93, which is a high probability.

For the 15 minutiae set in Figure 3.12(c) the confidences for (d_s, d_θ) were: (0.3, 7), (0.1, 7), (0.6, 8), (0.2, 8), (0.6, 7), (0.8, 6), (0.9, 9), (0.9, 9), (0.6, 5), (0.5, 5), (0.6, 8), (0.8, 9), (0.9, 9), (0.7, 5), (0.6, 5). The specific nPRC is 7.8×10^{-7}, which is high rarity.

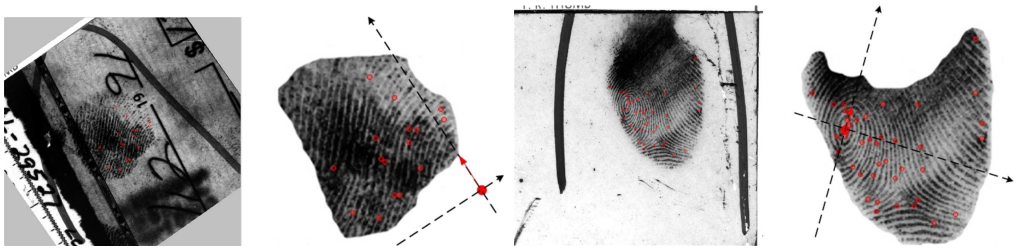
Thus the probability of random match of the latent print for the initial minutia annotation is high (0.78). It is much lower for the re-annotated print (one in a million). Incorporating minutia uncertainties does not have much effect on the conclusions since the core point is clearly discerned.

It is interesting to consider the rarity of the ten minutiae that are common between the re-annotated latent print and the ten-print of Daoud shown in Fig. 3.12(d). The specific nPRC of the ten minutiae, assuming point estimates, is 0.014, again a high probability. In general, configurations of ten minutiae are quite rare, e.g., when 200 ten minutiae sets were randomly chosen from fingerprints in the *NIST4* dataset, which contains $4,000$ images, with $n = 470 \times 10^6$, the average specific nPRC was 7.1×10^{-8}; this average value is by definition known as nPRC [59]. We conclude that while ten minutiae have significant discriminatory power, the particular configuration of ten minutiae common to the latent print and the Daoud print were those that are more common [60].

3.6.3 Latent Prints from Standard Dataset

Next we consider the effect of core point uncertainty on the evaluation of rarity. Two latent prints from the *NIST 27* data set are shown in Figure 3.13: print *b115* is from the *bad* quality set and print *g73* is from the *good* quality set. Latent print *b115*, contains $N = 16$ minutiae and *g73* contains $N = 39$ minutiae. Minutiae confidences were manually assigned by visual inspection; the values assigned for the 55 minutiae are analogous to the smaller minutiae set of minutiae in the example of Section 3.6.2.

Specific nPRCs for the two prints, computed for varying numbers of matching minutiae



(a) (b)

Figure 3.13: Two latent prints from *NIST27*: (a) *b115* is from the bad dataset and (b) *g73* is from the good dataset. In each case the left image is the print and the right its aligned version with predicted core point. Corresponding rarity values for $n = 100,000$ are given in Table 3.4. Rarity values of (a) for different values of n and different numbers of matching minutiae are given in Figure 3.14.

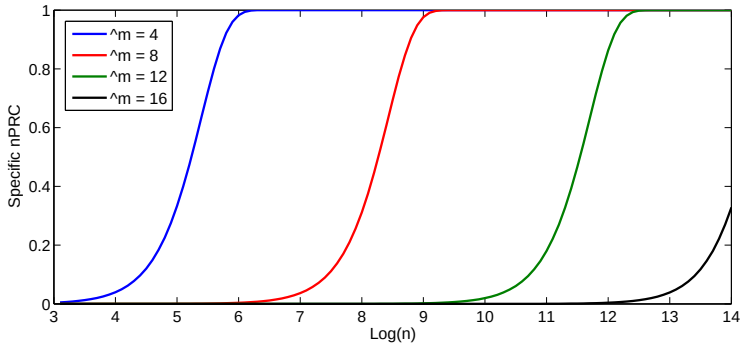


Figure 3.14: Dependence of specific nPRC of latent print *b115* on database size n parameterized by number of corresponding minutiae $\hat{m} = 4, 8, 12, 16$ and n varying from 10^3 to 10^{14}.

pairs $\hat{m}$, assuming fingerprint database size $n = 100,000$, are given in Table 3.4. The tolerance is set at $\epsilon_s = 10$ pixels and $\epsilon_\theta = \pi/8$. In each case the first column shows the specific nPRC with point estimates for the core point and the second column shows the specific nPRC with core point uncertainty as determined by GP.

The cases show that specific nPRC depends on the given latent fingerprint. As the number of minutiae to be matched, $\hat{m}$, increases the probability of finding a random match decreases. When all the minutiae are considered an extremely low value of specific nPRC is observed. Thus the values of specific nPRC provide a measure for the strength of latent

Table 3.4: Probability of randomly finding two *NIST27* latent prints (Figure 3.13) in a database of size $n = 100,000$, tolerance: $\epsilon_s = 10$ pixels and $\epsilon_\theta = \pi/8$, N is the no. of minutiae in input $\mathbf{X}$ and $\hat{m}$ is the no. of matching minutiae.

X = Bad Latent Print *b115*			
N	**No. of matching minutiae** $\hat{m}$	Specific nPRC $= p_\epsilon(\mathbf{X}, \hat{m}, n)$	
		Core is Point Estimated	**Core has a distribution**
16	2	0.73	0.79
	4	9.04×10^{-6}	1.81×10^{-5}
	8	2.46×10^{-19}	4.54×10^{-17}
	12	6.13×10^{-31}	6.05×10^{-28}
	16	1.82×10^{-46}	2.93×10^{-41}
X = Good Latent Print *g73*			
N	**No. of matching minutiae** $\hat{m}$	Specific nPRC $= p_\epsilon(\mathbf{X}, \hat{m}, n)$	
		Core is Point Estimated	**Core has a distribution**
39	4	1	1
	8	3.11×10^{-14}	1.50×10^{-15}
	12	2.56×10^{-25}	1.07×10^{-26}
	24	3.10×10^{-52}	9.93×10^{-55}
	39	7.51×10^{-79}	6.16×10^{-82}

fingerprint evidence.

Whether to use for the core point, a point estimate or a distribution of it, in rarity evaluation is an interesting question. As was seen in Table 3.1 the fixed estimate of the core point has a 15% error rate. Thus it is better to use the distribution of the core point unless the core point is precisely determinable.

It is also interesting to observe how rarity varies with database size. As can be seen in Figure 3.14 for latent print *b115*, finding a match for four minutiae, or $\hat{m} = 4$, in a database of a million entries, i.e., $log\ n = 6$, is guaranteed, whereas for matching twelve minutiae a trillion entries are needed.

3.7 Conclusions

Evaluation of the rarity of minutia configurations is routinely used by latent print examiners in performing comparisons. This evaluation is largely experience-based and the rarity of a given configuration of minutiae is not articulated. The proposed approach to explicitly compute the rarity of a given configuration of minutiae involves determining the core point distribution and a learnt generative model of minutiae to evaluate a probability random correspondence in a database of n entries.

Transforming the latent print to a standard coordinate system is performed using the fingerprint core point. Gaussian process regression is used to determine the core point. This machine learning approach has several advantages over traditional approaches: (i) a core point distribution is obtained which can be factored into subsequent probability evaluations, (ii) the point estimate (MAP) is more accurate than the results of a popular method (Poincare method), and (iii) unlike topology-based approaches it can predict a core point outside the field of view. In the rare case where the complete fingerprint itself has no core point, or there are multiple core points, the predicted value simply serves as the origin for subsequent evaluations.

The distribution of minutiae is modeled using a Bayesian network which takes into account dependencies between minutiae. Directed edges in the network are determined starting with the core point and proceeding to nearby minutiae. The model was validated using a goodness-of-fit test.

The probability of finding a given input latent print in a database of given size can be computed using the models. This evaluation can be done using either point estimates of

core point/minutiae or with confidence values associated with them. Use of the models was illustrated in several cases: (i) simple configurations involving four minutiae, (ii) the Madrid bombing case with the initially annotated latent print and a re-annotated latent print after finding a database match, and (iii) two latent prints from the *NIST27* dataset.

As in any machine learning approach, the quality of results, i.e., rarity values computed, are dependent on the learning set employed. Here we used $4,000$ latent prints in the $NIST4$ dataset to determine various parameters needed. Using the procedures described, rarity evaluations can be improved, or tailored to, any large dataset of representative prints.

The proposed method takes into account only level 2 detail in the form of minutiae. Ridge information can be easily incorporated by using ridge point information represented in a manner similar to minutiae. The types of minutiae such as ridge endings, bifurcations and trifurcations can be incorporated into the statistical model.

Chapter 4

CONCLUSIONS

This dissertation studies the role of machine learning in fingerprint probability evaluation. Two major problems, fingerprint individuality and fingerprint rarity, have been investigated. The contribution of this dissertation is listed as follows.

1. First, we proposed generative models for fingerprints. Three forms of fingerprint representation are considered: ridge flow, minutiae, and ridge points. Mixture models are used to model location and orientation of minutiae. EM algorithms are used to estimate parameters from standard fingerprint databases. Goodness-of-fit test was performed to prove that the proposed generative model offers a reasonable and more accurate fingerprint representation than previous models.

2. Second, we defined probabilistic metrics to measure the degree of individuality of a forensic modality such as fingerprints. The metric is defined as the probability of random correspondence (PRC) when evidence consists of a set of measurements and correspondence is within a tolerance. Computation of these probabilities is described

using Bayesian networks which make all the variables explicit. The experimental results provide a much stronger argument for the individuality of fingerprints in forensics than previous generative models.

3. Third, in order to predict core points, a machine learning approach based on Gaussian process regression is proposed. Since latent prints are often incomplete and the core point may be missing in the field of view, classic computational vision based core point detection algorithms usually fails in latent cases. By using Gaussian processes, the distribution of core points can be discovered. The coordinate system is transformed into standard position based on finding the core point. The experimental results proven that this approach can predict core points whether the prints contain the core points or not.

4. Fourth, a method for computing the rarity of latent fingerprints represented by minutiae is proposed. It allows determining the probability of finding a random match for the evidence in a database of n prints. A generative model, that takes into account inter-minutia dependencies and minutia confidences, is used to determine evidence probability from which the specific probability of match among n is evaluated. The generative model is validated using a goodness-of-fit test. We extend the rarity calculation with a fully Bayesian treatment by considering uncertainty of core points. Fingerprint rarity evaluation is illustrated using several examples, including simple configurations of minutiae, randomly selected latent fingerprints in a database, and a well-known case of erroneous identification.

Bibliography

[1] C. Aitken and F. Taroni. *Statistics and the Evaluation of Evidence for Forensic Scientists.* Wiley, 2004.

[2] D. R. Ashbaugh. *Quantitative-Qualitative Friction Ridge Analysis: An Introduction to Basic and Advanced Ridgeology.* CRC Press, Boca Raton, FL, 1999.

[3] A. Banerjee, I. Dhillon, J. Ghosh, and S. Sra. Generative model-based clustering of directional data. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2003.

[4] A. M. Bazen and S. H. Gerez. Systematic methods for the computation of the directional fields and singular points of fingerprints. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(7):905–919, 2002.

[5] C. Bishop. *Pattern Recognition and Machine Learning.* Springer, New York, 2006.

[6] R. M. Bolle. *Guide to Biometrics.* Springer-Verlag New York, LLC, 2003.

[7] R. Chakraborty. Statistical interpretation of DNA typing data. *American Journal of Human Genetics*, 49(4):895–897, 1991.

[8] C. Champod and P. Margot. Computer assisted analysis of minutiae occurrences on fingerprints. In *Proceeding of Int. Symp. Fingerprint Detection and Identification*, August 1996.

[9] J. Chen and Y. Moon. The statistical modelling of fingerprint minutiae distribution with implications for fingerprint individuality studies. In *Proc. Computer Vision and Pattern Recognition*, 2008.

[10] J. Chen and Y.-S. Moon. A statistical study on the fingerprint minutiae distribution. In *ICASSP 2006 Proceedings.*, volume 2, pages II–II, 2006.

[11] J. Chen and Y.-S. Moon. A minutiae-based fingerprint individuality model. *Computer Vision and Pattern Recognition, 2007. CVPR '07.*, 2007.

[12] Y. Chen and A. K. Jain. Beyond minutiae: A fingerprint individuality model with pattern, ridge and pore features. In *ICB '09 Proceedings*, pages 523–533, Berlin, Heidelberg, 2009. Springer-Verlag.

[13] S. Dass, Y. Zhu, and A. K. Jain. Statistical models for assessing the individuality of fingerprints. *Fourth IEEE Workshop on Automatic Identification Advanced Technologies*, pages 3–9, 2005.

[14] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification, Second Edition.* New York: John Wiley and Sons, 2001.

[15] G. Fang, S. N. Srihari, H. Srinivasan, and P. Phatak. Use of ridge points in partial fingerprint matching. In *Biometric Technology for Human Identification IV*, pages 65390D1–65390D9. SPIE, 2007.

[16] F. Galton. *Finger Prints.* McMillan, London, 1892.

[17] M. Garris and R. McCabe. NIST special database 27: Fingerprint minutiae from latent and matching tenprint images. http://www.nist.gov/srd/nistsd27.htm, June 2000.

[18] E. R. Henry. *Classification and Uses of Fingerprints.* London, 1900.

[19] R. V. Hsu and B. Martin. An analysis of minutiae neighborhood probabilities. *Biometrics: Theory, Applications and Systems, 2008.*, 2008.

[20] A. K. Jain and D. Maltoni. *Handbook of Fingerprint Recognition*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2003.

[21] A. K. Jain, S. Prabhakar, and L. Hong. A multichannel approach to fingerprint classification. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21(4):348–359, 1999.

[22] A. K. Jain, S. Prabhakar, L. Hong, and S. Pankanti. Filterbank-based fingerprint matching. *IEEE Transactions on Image Processing*, 9:846–859, 2000.

[23] T. Jebarra. *Machine Learning: Discriminative and Generative*. Kluwer Academic, 2004.

[24] M. Kawagoe and A. Tojo. Fingerprint pattern classification. *Pattern Recogn.*, 17(3):295–303, 1984.

[25] D. Lindley. A problem in forensic science. *Biometrika*, 64(2):207–213, 1977.

[26] M. Liu, X. Jiang, and A. C. Kot. Fingerprint reference-point detection. *EURASIP J. Appl. Signal Process.*, 2005:498–509, 2005.

[27] E. Locard. La preuve judiciaire par les empreintes digitales. *Archives d'anthropologie criminelle, de mdecine lgale et de psychologie normale et pathologique*, 29(245):321–348, 1914.

[28] K. V. Mardia and P. E. Jupp. *Directional Statistics*. Wiley, 2000.

[29] National Academy of Sciences. *Strengthening the Forensic Sciences in the United States: A Path Forward*. National Academies Press, 2009.

[30] C. Neumann, C. Champod, R. Puch-Solis, N. Egli, A. Anthonioz, and A. Bromage-Griffiths. Computation of likelihood ratios in fingerprint identification for configurations of any number of minutiae. *J Forensic Sci.*, 52:54–64, 2007.

[31] C. Neumann, C. Champod, R. Puch-Solis, D. Meuwly, N. Egli, and A. Anthonioz. Computation of likelihood ratios in fingerprint identification for configurations of three minutiae. *J. Forensic Sci.*, 51:1255, 2006.

[32] S. Pankanti, S. Prabhakar, and A. K. Jain. On the individuality of fingerprints. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(8):1010–1025, 2002.

[33] K. Pearson. *The Life and Letters of Francis Galton*, volume IIIA. University Press, Cambridge, 1930.

[34] D. Phillips. A fingerprint orientation model based on 2d fourier expansion (fomfe) and its application to singular-point detection and fingerprint indexing. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(4):573–585, 2007.

[35] M. A. S. Ralph B. D'Agostino. *Goodness-of-fit Techniques*. CRC Press, 1986.

[36] C. Rasmussen and C. Williams. *Gaussian Processes for Machine Learning*. the MIT Press, 2006.

[37] T. Roxburgh. Galton's work on the evidential value of fingerprints. *Indian Journal of Statistics*, 1:62, 1933.

[38] S. C. Scolve. The occurence of fingerprint characteristics as a two dimensional process. *Journal of the American Statistical Association*, 367(74):588–595, 1979.

[39] Y. Shen, A. Ng, and M. Seeger. Fast gaussian process regression using kd-trees. In Y. Weiss, B. Schölkopf, and J. Platt, editors, *Advances in Neural Information Processing Systems 18*, pages 1225–1232. MIT Press, Cambridge, MA, 2006.

[40] P. Simard, Y. LeCun, and J. Denker. Efficient pattern recognition using a new transformation distance. In S. J. Hanson, J. D. Cowan, and C. L. Giles, editors, *Advances in Neural Information Processing Systems*, volume 5, pages 50–58. Morgan Kaufmann, 2003.

[41] E. Snelson and Z. Ghahramani. Sparse gaussian processes using pseudo-inputs. In Y. Weiss, B. Schölkopf, and J. Platt, editors, *Advances in Neural Information Processing Systems 18*, pages 1257–1264. MIT Press, Cambridge, MA, 2006.

[42] S. Srihari and H. Srinivasan. Comparison of ROC and Likelihood Decision Methods in Automatic Fingerprint Verification. *International J. Pattern Recognition and Artificial Intelligence*, 22(1):535–553, 2008.

[43] S. Srihari, H. Srinivasan, and G. Fang. Discriminability of fingerprints of twins. *Journal of Forensic Identification*, 58:109–127, 2008.

[44] S. N. Srihari, S. Cha, H. Arora, and S. J. Lee. Individuality of handwriting. *Journal of Forensic Sciences*, 47(4):856–872, 2002.

[45] S. N. Srihari, S. Cha, H. Arora, and S. J. Lee. Discriminability of fingerprints of twins. *Journal of Forensic Identification*, 58(1):109–127, 2008.

[46] S. N. Srihari, C. Huang, and H. Srinivasan. On the discriminability of the handwriting of twins. *Journal of Forensic Sciences*, 53(2):430–446, 2008.

[47] D. Stoney and J. Thornton. A method for the description of minutia pairs in epidermal ridge patterns. *Journal of Forensic Sciences*, 31:1217, 1986.

[48] D. A. Stoney. Distribution of epidermal ridge minutiae. *American Journal of Physical Anthropology*, 77:367–376, 1988.

[49] D. A. Stoney. Measurement of fingerprint individuality. In H. Lee and R. Gaensslen, editors, *Advances in Fingerprint Technology*. CRC Press, 2001.

[50] C. Su and S. Srihari. Generative models for fingerprint individuality using ridge models. In *Proceeding of International Conference on Pattern Recognition*, pages 1–4. IEEE Computer Society Press, 2008.

[51] C. Su and S. Srihari. Evaluation of rarity of fingerprints in forensics. In J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems 23, NIPS'2010,*, pages 1207–1215. 2010.

[52] C. Su and S. Srihari. Latent fingerprint core point prediction based on gaussian processes. In *Proceedings of the 2010 20th International Conference on Pattern Recognition*, ICPR '10, pages 1634–1637, Washington, DC, USA, 2010. IEEE Computer Society.

[53] C. Su and S. N. Srihari. Generative models for fingerprint individuality using ridge models. In *Proceedings of International Conference on Pattern Recognition*. IEEE Computer Society Press, 2008.

[54] C. Su and S. N. Srihari. Probability of random correspondence for fingerprints. In *IWCF '09 Proceedings*, pages 55–66, Berlin, Heidelberg, 2009. Springer-Verlag.

[55] C. Su and S. N. Srihari. Generative models and probability evaluation for forensic evidence. In P. Wang, editor, *Pattern Recognition, Machine Intelligence and Biometrics*. Springer, 2011.

[56] C. Su and S. N. Srihari. Latent fingerprint rarity analysis in madrid bombing case. In *Proceedings of the 4th international conference on Computational forensics*, IWCF'10, pages 173–184, Berlin, Heidelberg, 2011. Springer-Verlag.

[57] E. Tabassi, C. L. Wilson, and C. I. Watson. Fingerprint image quality. *NISTIR 7151, National Institute of Standards and Technology*, August 2004.

[58] H. Templeman. Fingerprint identification based on likelihood ratio, match probability and relevant fingerprint population.

[59] M. Trauring. Automatic comparison of finger-ridge patterns. *Nature*, 197:938–940, 1963.

[60] United States Court of Appeals for the Third Circuit. USA v. Byron Mitchell, 2003. No. 02-2859.

[61] United States Supreme Court. Daubert et. al. v Merrell Dow Pharmaceuticals, 1993. 509 U.S. 579.

[62] USCATC. United States Court of Appeals for the Third Circuit: USA v. Byron Mitchell, 2003. No. 02-2859.

[63] X. Wang, J. Li, and Y. Niu. Definition and extraction of stable points from fingerprint images. *Pattern Recogn.*, 40(6):1804–1815, 2007.

[64] C. Watson, M. Garris, E. Tabassi, C. Wilson, R. McCabe, and S. Janet. *User's Guide to NIST Fingerprint Image Software 2 (NFIS2).* NIST, 2004.

[65] C. Watson and C. Wilson. NIST special database 4: 8-bit gray images of fingerprint image groups. http://www.nist.gov/srd/nistsd4.htm, March 1992.

[66] K. Wertheim and A. Maceo. The critical stage of friction ridge and pattern formation. *Journal of Forensic Identification*, 52(1):35–85, 2002.

[67] Y. Zhu, S. C. Dass, and A. K. Jain. Statistical models for assessing the individuality of fingerprints. *IEEE Transactions on Information Forensics and Security*, 2(3-1):391–401, 2007.