On the Individuality of Fingerprints

Sharath Pankanti
IBM T. J. Watson Research Center
Yorktown Heights, NY 10598
sharat@watson.ibm.com

Salil Prabhakar
DigitalPersona, Inc.
Redwood City, CA 94063
salilp@digitalpersona.com

Anil K. Jain
Dept. of Comp. Sci. and Eng.
Michigan State University
East Lansing, MI 48824
jain@cse.msu.edu

Abstract

Fingerprint identification is based on two basic premises: (i) persistence: the basic characteristics of fingerprints do not change with time; and (ii) individuality: the fingerprint is unique to an individual. The validity of the first premise has been established by the anatomy and morphogenesis of friction ridge skin. While the second premise has been generally accepted to be true based on empirical results, the underlying scientific basis of fingerprint individuality has not been formally tested. As a result, fingerprint evidence is now being challenged in several court cases. We address the problem of fingerprint individuality by quantifying the amount of information available in minutiae points to establish a correspondence between two fingerprint images. We derive an expression which estimates the probability of falsely associating minutiae-based representations from two arbitrary fingerprints. For example, the probability that a fingerprint with 36 minutiae points will share 12 minutiae points with another arbitrarily chosen fingerprint with 36 minutiae points is 6.10×10^{-8}. *These probability estimates are compared with typical fingerprint matcher accuracy results. Our results show that (i) contrary to the popular belief fingerprint matching is not infallible and leads to some false associations, (ii) the performance of automatic fingerprint matcher does not even come close to the theoretical performance, and (iii) due to the limited information content of the minutiae-based representation, the automatic system designers should explore the use of non-minutiae-based information present in the fingerprints.*

1. Introduction

Fingerprint based personal identification is routinely used in forensic laboratories and identification units around the world and it has been accepted in the court of law for nearly a century. Until recently, the testimony of latent fingerprint examiners was admitted in courts without much scrutiny and challenges. However, in the 1993 case of Daubert vs. Merrell Dow Pharmaceuticals, Inc. [1], the Supreme Court ruled that the reliability of an expert scientific testimony must be established. Additionally, the court stated that when assessing reliability, the following five factors should be considered: (i) whether the particular technique or methodology in question has been subject to a statistical hypothesis testing, (ii) whether its error rate has been established, (iii) whether the standards controlling the technique's operations exist and have been maintained, (iv) whether it has been peer reviewed, and published, and (v) whether it has a general widespread acceptance. Subsequently, handwriting identification was challenged under *Daubert* (it was claimed that handwriting identification does not meet the scientific evidence criteria established in the Daubert case) in a number of cases between 1995 and 2001 and several courts have now ruled that handwriting identification does not meet the *Daubert* criteria. Fingerprint identification was first challenged by the defense lawyers under *Daubert* in the 1999 case of USA vs. Byron Mitchell [2] on the basis that the fundamental premises of fingerprint identification have not been objectively tested and its potential error rate is not known. The defense motion to exclude fingerprint evidence and testimony was, however, denied. The outcome of the USA vs. Byron Mitchell case is still pending. Fingerprint identification has been challenged under *Daubert* in more than 10 court cases till date since the USA vs. Byron Mitchell case in 1999.

The two fundamental premises on which fingerprint identification is based are: (i) fingerprint details are permanent, and (ii) fingerprints of an individual are unique. The validity of the first premise has been established based on the anatomy and morphogenesis of friction ridge skin. It is the second premise which is being challenged in recent court cases. The notion of fingerprint individuality has been widely accepted based on manual inspection (by experts) of millions of fingerprints. However, the underlying scientific basis [3] of fingerprint individuality has not been rigorously studied or tested.

What do we mean by fingerprint individuality? Finger-

print individuality problem can be formulated in many different ways. Two typical formulations are: (i) the individuality problem may be cast as determining the probability that any two individuals may have sufficiently similar fingerprints in a given target population; (ii) given a sample fingerprint, determine the probability of finding a sufficiently similar fingerprint in a target population. In this study, we define the individuality problem as the probability of a *false association*: given two fingerprints from two different fingers, determine the probability that they are "sufficiently" similar. If two fingerprints originating from two different fingers are examined at a very high level of detail (resolution), we may find that the fingerprints are indeed different. However, most human experts and automatic fingerprint identification systems (AFIS) declare that the fingerprints originate from the same source if they are "sufficiently" similar. How much similarity is enough depends on typical (intra-class) variations observed in the multiple impressions of a finger.

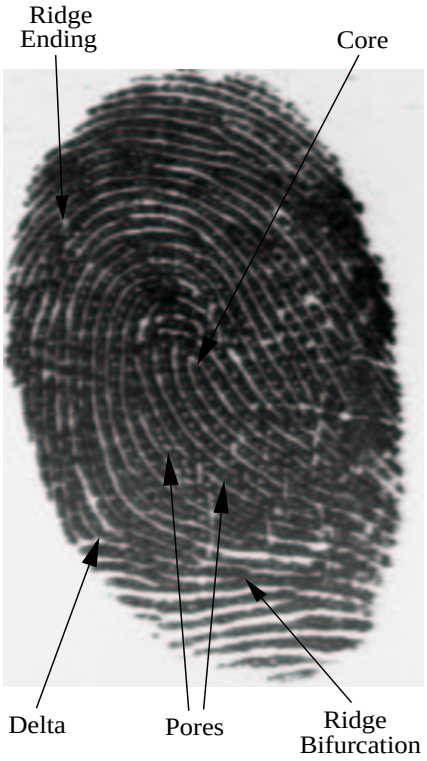


Figure 1. A fingerprint image of type "right loop". The overall ridge structure, singular points, and sweat pores are shown.

In order to solve the individuality problem, we need to first define *a priori* the representation of fingerprint (*pattern*) and the metric for the similarity. Fingerprints can be represented by a large number of features, including the overall ridge flow pattern, ridge frequency, location and position of singular points (core(s) and delta(s)), type, direction, and location of minutiae points, ridge counts between pairs of minutiae, and location of pores (see Figure 1). All these features contribute in establishing fingerprint individuality. In this study, we have chosen minutiae representation of the fingerprints because it is utilized by forensic experts, has been demonstrated to be relatively stable, and has been adopted by most of the automatic fingerprint matching systems. The similarity metric is the number of *corresponding* minutiae between the two minutiae sets (see Figure 2).

Given a representation scheme and a similarity metric, there are two approaches for determining the individuality of the fingerprints. In the empirical approach, *representative* samples of fingerprints are collected and using a *typical* fingerprint matcher, the accuracy of the matcher on the samples provides an indication of the uniqueness of the fingerprint with respect to the matcher. There are known problems (and costs) associated with collection of a large number of *representative* samples. In the theoretical approach to individuality estimation, one models all realistic phenomenon affecting inter-class and intra-class pattern variations. Given the similarity metric, one could then, estimate the probability of a false association. Theoretical approaches are often limited by the extent to which the assumed model conforms to the reality. In this work, we propose a theoretical formulation of the fingerprint individuality model based on a number of parameters derived from a database of fingerprint images. We also compare the probabilities obtained from this individuality model with the empirical matcher accuracy results.

The rest of the paper is organized as follows. Section 2 presents a brief summary of major fingerprint individuality studies. Section 3 presents the proposed fingerprint individuality model, and section 4 presents the results. Discussions are presented in section 5.

2. Background

Most of the early individuality studies examined the distinctiveness of a single fingerprint (without addressing the issues of intra-class pattern variation) under simplifying assumptions (e.g., implicit assumptions about statistical independence of events and that the corresponding event distributions are identical). We will refer to these total pattern variation-based fingerprint individuality estimates as the *probability of fingerprint configuration*. A summary of these studies is presented below.

The fingerprint individuality problem was first addressed by Galton in 1892 (cf. [8]), who considered a square region spanning six-ridges in a given fingerprint. He assumed that, on an average, a fingerprint can be covered by 24 such six-ridge wide independent square regions. Galton estimated that he could correctly reconstruct any of the regions with a probability of $\frac{1}{2}$, by looking at the surrounding ridges. Accordingly, the probability of a specific fingerprint con-

Author	P(Fingerprint Configuration)	N=36,R=24,M=72	N=12,R=8,M=72
Galton (1892)	$\frac{1}{16} \times \frac{1}{256} \times \left(\frac{1}{2}\right)^R$	1.45×10^{-11}	9.54×10^{-7}
Pearson (1930)	$\frac{1}{16} \times \frac{1}{256} \times \left(\frac{1}{36}\right)^R$	1.09×10^{-41}	8.65×10^{-17}
Henry (1900)	$\left(\frac{1}{4}\right)^{N+2}$	1.32×10^{-23}	3.72×10^{-9}
Balthazard (1911)	$\left(\frac{1}{4}\right)^N$	2.12×10^{-22}	5.96×10^{-8}
Bose (1917)	$\left(\frac{1}{4}\right)^N$	2.12×10^{-22}	5.96×10^{-8}
Wentworth & Wilder (1918)	$\left(\frac{1}{50}\right)^N$	6.87×10^{-62}	4.10×10^{-21}
Cummins & Midlo (1943)	$\frac{1}{31} \times \left(\frac{1}{50}\right)^N$	2.22×10^{-63}	1.32×10^{-22}
Gupta (1968)	$\frac{1}{10} \times \frac{1}{10} \times \left(\frac{1}{10}\right)^N$	1.00×10^{-38}	1.00×10^{-14}
Roxburgh (1933)	$\frac{1}{1000} \times \left(\frac{1.5}{10 \times 2.412}\right)^N$	3.75×10^{-47}	3.35×10^{-18}
Trauring (1963)	$(0.1944)^N$	2.47×10^{-26}	2.91×10^{-9}
Osterburg et al. (1980)	$(0.766)^{M-N}(0.234)^N$	1.33×10^{-27}	3.05×10^{-15}
Stoney (1985)	$\frac{N}{5} \times 0.6 \times (0.5 \times 10^{-3})^{N-1}$	1.2×10^{-80}	3.5×10^{-26}

Table 1. Comparison of probability of a particular fingerprint configuration using different models. We assume that an average size fingerprint has 24 regions ($R = 24$) as defined by Galton, 72 regions ($M = 72$) as defined by Osterburg et al., and has 36 minutiae on an average ($N = 36$). Note that all probabilities represent a full (N minutiae) match as opposed to a partial match (see Table 2).

figuration, given the surrounding ridges is $\left(\frac{1}{2}\right)^{24}$. He multiplied this conditional (on surrounding ridges) probability with the probability of finding the surrounding ridges to obtain the probability of occurrence of a fingerprint as $P(\text{Fingerprint Configuration}) = \frac{1}{16} \times \frac{1}{256} \times \left(\frac{1}{2}\right)^{24} = 1.45 \times 10^{-11}$, where $\frac{1}{16}$ is the probability of occurrence of a specific fingerprint type (such as arch, tented arch, left loop, right loop, double loop, whorl, etc.) and $\frac{1}{256}$ is the probability of occurrence of the correct number of ridges entering and exiting each of the 24 regions. Galton's formulation gives the probability that a particular fingerprint configuration in an average size fingerprint (containing 24 regions as defined by Galton) will be observed in nature. Pearson (cf. [8]) argued that there could be 36 (6×6) possible minutiae locations within one of Galton's six-ridge-square regions, and replaced Galton's probability of a six-ridge-square region of $\frac{1}{2}$ by $\frac{1}{36}$. A number of subsequent models by Henry (cf. [8]), Balthazard (cf. [8]), Bose (cf. [8]), Wentworth and Wilder (cf. [8]), Cummins and Midlo [5], and Gupta (cf. [8]) are interrelated and are based on a fixed probability, p, for the occurrence of a minutiae. They compute the probability of a particular N-minutiae fingerprint configuration as $P(\text{Fingerprint Configuration}) = p^N$. Roxburgh (cf. [8]), Amy (cf. [8]), and Kingston's (cf. [8]) models are more complex in that they compute the probability of a fingerprint configuration based on several additional fingerprint features.

Osterburg et al. [6] divided fingerprints into discrete cells of size $1\ mm \times 1\ mm$. They computed the frequencies of 13 types of minutiae events (including an empty cell) from 39 fingerprints (8,591 cells) and estimated the probability that 12 ridge endings will match between two fingerprints based on an average fingerprint area of $72\ mm^2$ as 1.25×10^{-20}. Sclove [7] modified Osterburg et al.'s model by incorporating the observed dependence of minutiae occurrence in cells and came up with an estimate of probability of fingerprint configuration that is slightly higher than that obtained by Osterburg et al.

Stoney and Thornton [8] critically reviewed earlier fingerprint individuality models and attempted to characterize pairwise minutiae dependence. They proposed a linear ordering of minutiae and recursively estimated the probability of a n-minutiae configuration from the probability of a $(n-1)$-minutiae configuration and the occurrence of a new minutiae of certain type/orientation at a particular distance/ridge counts from its nearest minutiae within the $(n-1)$-minutiae configuration. The probabilities of observing a particular minutiae configuration from different models are compared in Table 1.

The models discussed above measure the amount of detail in a single fingerprint, i.e., they estimate the probability of a fingerprint configuration. However, these models did not emphasize the intra-class variations in multiple impressions of a finger. We will refer to the quantifications of fingerprint individuality which explicitly consider the intra-class variations as *probability of correspondence*. Trauring [10] was the first to concentrate explicitly on measuring the amount of detail needed to establish correspondence between two prints from the same finger using an automatic fingerprint identification system. He observed that corresponding fingerprint features could be displaced from each other by as much as 1.5 times the inter-ridge distance. Using an AFIS, Meagher et al. (cf. Stiles [11]) empirically matched about 50,000 rolled fingerprints belonging to the same fingerprint class (left loop) with each other to compute the impostor distribution without considering realistic intra-class variations. Consequently, their assessment of the

probability of false association (10^{-97}) is a gross under estimate of the true probability.

3. Fingerprint Individuality Model

We have developed a model to obtain a realistic and accurate probability of correspondence between fingerprints. The probabilities obtained using this model will be compared against empirical values using an automatic fingerprint matching system [4]. To estimate the probability of correspondence, we make the following assumptions: (i) We consider only two types of minutiae features: ridge endings and ridge bifurcations. Additionally, we do not distinguish between the two types of minutiae because they can not be accurately discriminated. Since minutiae can reside only on ridges which follow a "flow" pattern, we implicitly model the statistical dependence between minutiae directions and locations. (ii) We assume a uniform distribution of minutiae in a fingerprint with the restriction that two minutiae cannot be very close to each other. This assumption approximates the slightly overdispersed uniform distribution of minutiae found by Stoney [9]. (iii) Correspondence of a minutiae pair is an independent event and each correspondence is equally important. (iv) We do not explicitly take into account fingerprint image quality in individuality determination since it is very difficult to reliably assign a quality index to a fingerprint.

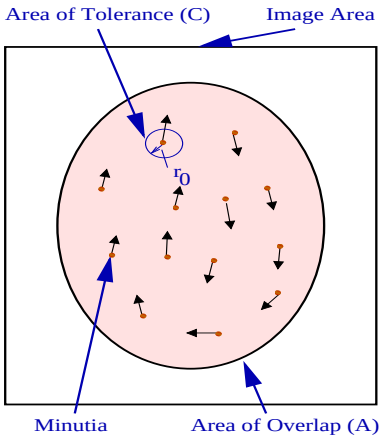


Figure 3. Fingerprint and minutiae.

The fingerprint correspondence problem involves matching a *template* with the *input*. We assume that a reasonable *alignment* has been established between the template and the input. The alignment of the input minutiae set with the template minutiae set is done so that the minutiae correspondences can be determined within a small tolerance. Given an input fingerprint containing n minutiae, our goal is to compute the probability that any arbitrary fingerprint (template in a database of fingerprints) containing m minutiae will have exactly q corresponding minutiae with the input. Since the fingerprint minutiae are defined by their location, (x, y), and by the angle of the ridge on which they reside, θ, the input and the template minutiae sets, I and T, respectively, are defined as:

$$I = \left\{\{x'_1, y'_1, \theta'_1\}, \{x'_2, y'_2, \theta'_2\}, ..., \{x'_n, y'_n, \theta'_n\}\right\}, \quad (1)$$

$$T = \{\{x_1, y_1, \theta_1\}, \{x_2, y_2, \theta_2\}, ..., \{x_m, y_m, \theta_m\}\}. \quad (2)$$

A minutiae j in the input fingerprint is considered as "corresponding" or "matching" to the minutiae i in the template, if and only if

$$\sqrt{(x'_i - x_j)^2 + (y'_i - y_j)^2} \leq r_0, \text{ and} \quad (3)$$

$$\min\left(\left|\theta'_i - \theta_j\right|, 360 - \left|\theta'_i - \theta_j\right|\right) \leq \theta_0, \quad (4)$$

where r_0 is the tolerance in distance and θ_0 is the tolerance in angle.

Let A be the total area of overlap between the input and the template fingerprints after a reasonable alignment has been achieved. The probabilities that any arbitrary minutiae in the input will match any arbitrary minutiae in the template, only in terms of location, and only in terms of direction, are given by Eqs. (5) and (6), respectively. Eq. (5) assumes that (x, y) and (x', y') are independent and Eq. (6) assumes that θ and θ' are independent. Let $\delta_x = x' - x$, $\delta_y = y' - y$, and $d = \sqrt{\delta_x^2 + \delta_y^2}$.

$$P(d \leq r_0) = \frac{\pi r_0^2}{A} = \frac{C}{A}, \quad (5)$$

$$P\left(\min\left(\left|\theta' - \theta\right|, 360 - \left|\theta' - \theta\right|\right) \leq \theta_0\right) = \frac{2\theta_0}{360}. \quad (6)$$

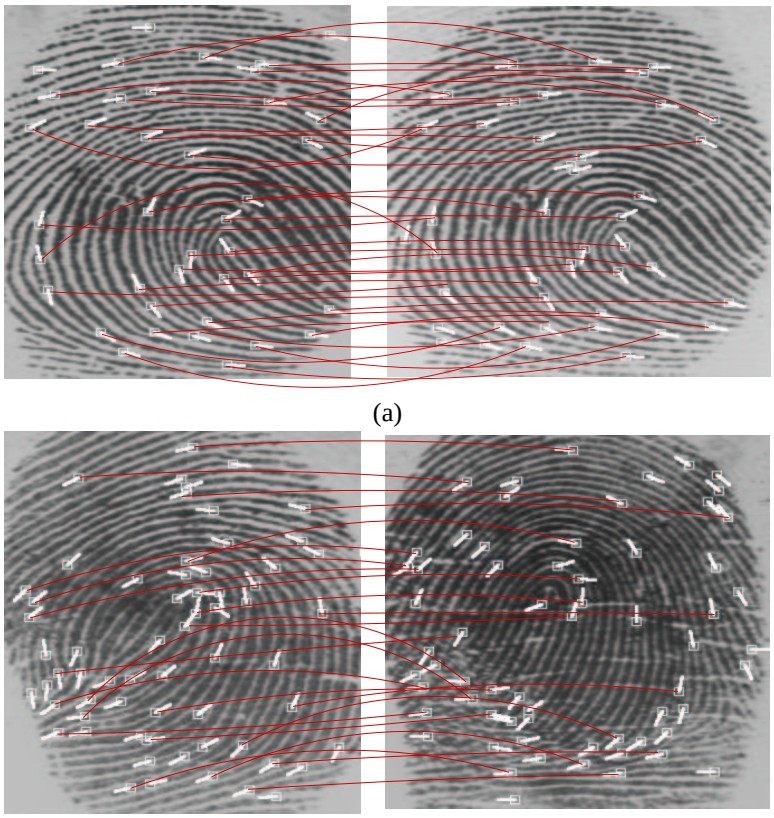
If the template contains m minutiae, the probability that only one minutia in the input will correspond to any of the m template minutiae is given by $\frac{mC}{A}$. Now, given two input minutiae, the probability that only the "first" one corresponds to any of the m template minutiae is the product of the probabilities that the first input minutiae has a correspondence ($\frac{mC}{A}$) and the second minutiae does not have a correspondence $\left(\frac{A-mC}{A-C}\right)$. Thus, the probability that exactly one of the two input minutiae matches any of the m template minutiae is $2 \times \frac{mC}{A} \times \frac{A-mC}{A-C}$, since either the first input minutiae alone may have a correspondence or the second input minutiae alone may have a correspondence. If the input fingerprint has n minutiae, the probability that exactly one input minutia matches one of the m template minutiae is

$$p(A, C, m, n, 1) = \binom{n}{1}\left(\frac{mC}{A}\right)\left(\frac{A - mC}{A - C}\right). \quad (7)$$

The probability that there are exactly ρ corresponding minutiae between the n input minutiae and m template minutiae is then given by:

$$p(A, C, m, n, \rho) = \binom{n}{\rho} \underbrace{\left(\frac{mC}{A}\right)\left(\frac{(m-1)C}{A-C}\right)\cdots\left(\frac{(m-\rho-1)C}{A-(\rho-1)C}\right)}_{\rho \text{ terms}} \times \underbrace{\left(\frac{A-mC}{A-\rho C}\right)\left(\frac{A-(m-1)C}{A-(\rho+1)C}\right)\cdots\left(\frac{(A-(m-(n-\rho+1))C}{A-(n-1)C}\right)}_{n-\rho \text{ terms}}. \quad (8)$$

The first ρ terms in Eq. (8) denote the probability of matching ρ minutiae between the template and the input; and remaining $n-\rho$ terms express the probability that $n-\rho$ minutiae in the input do not match any minutiae in the template.



(a)

(b)

Figure 2. Automatic minutiae matching. Two impressions of the same finger were matched in (a), 39 minutiae were detected in input (left), 42 in template (right), and 36 correspondences were found. Two different fingers are matched in (b), 64 minutiae were detected in input (left), 65 in template (right), and 25 "false correspondences" were found.

Dividing the numerator and denominator of each term in Eq. (8) by C, we obtain:

$$p(A,C,m,n,\rho) = \binom{n}{\rho} \left(\frac{m}{A/C}\right)\left(\frac{(m-1)}{A/C-1}\right)\cdots\left(\frac{(m-\rho-1)}{A/C-(\rho-1)}\right)\times \left(\frac{A/C-m}{A/C-\rho}\right)\left(\frac{A/C-(m-1)}{A/C-(\rho+1)}\right)\cdots\left(\frac{(A/C-(m-(n-\rho+1))}{A/C-(n-1)}\right). \quad (9)$$

Letting $M = A/C$, assuming M to be integer ($A \gg C$), and rearranging, we obtain

$$p(M,m,n,\rho) = \frac{\binom{m}{\rho}\binom{M-m}{n-\rho}}{\binom{M}{n}}, \quad (10)$$

which is a hyper-geometric distribution.

The above analysis considers a minutiae correspondence based solely on the minutiae location. Next we consider a minutiae correspondence that depends on minutiae directions as well as minutiae positions. For the sake of this analysis, let us assume that the minutiae directions are completely independent of the minutiae positions and matching minutiae position and minutiae direction are therefore independent events. Let l be such that $P\left(\min\left(|\theta_i' - \theta_j|, 360 - |\theta_i' - \theta_j|\right) \leq \theta_0\right) = l$ in Eq. (6). Given n input and m template minutiae, the probability of ρ minutiae falling into the *similar* positions can be estimated by Eq. (10). Once ρ minutiae positions are matched, the probability that $q \leq \rho$ minutiae among them have similar direction is given by

$$\binom{\rho}{q} (l)^q (1-l)^{\rho-q}, \quad (11)$$

where l is the probability of two position-matched minutiae having similar direction and $1-l$ is the probability of two position-matched minutiae taking different directions. Therefore, probability of matching q minutiae in both position as well as direction is given by

$$p(M,m,n,q) = \sum_{\rho=q}^{\min(m,n)} \left(\frac{\binom{m}{\rho}\binom{M-m}{n-\rho}}{\binom{M}{n}} \times \binom{\rho}{q} (l)^q (1-l)^{\rho-q} \right). \quad (12)$$

Until now, we have assumed that the minutiae locations are uniformly distributed within the *entire* fingerprint area. However, the number (or the area) of ridges across all fingerprint types is approximately the same. Since A is the

area of the overlap between the template and the input fingerprints, the ridges occupy approximately $\frac{A}{2}$ of the area, with the other half being occupied by the valleys. Since the minutiae can lie only on ridges, i.e., along a curve of length $\frac{A}{w}$, where w is the ridge period, the value of M in Eq. (12) should therefore be changed from $M = A/C$ to $M = \frac{A/w}{2r_0}$, where $2r_0$ is the length tolerance in minutiae location. This analysis assumes that the ridge direction information/uncertainty is completely captured by Eq. (6).

3.1. Parameter Estimation

Our individuality model has several parameters, namely, r_0, l, w, A, m, n, and q. The value of l further depends on θ_0. The values of r_0, l, and w are estimated in this section for a given sensor resolution. To compare the values obtained from the theoretical model with the empirical results, we will estimate the values of A, m, and n from two different databases in the next section.

The value of r_0 should be determined to account for the variation in the different impressions of the same finger. However, since the spatial tolerance is dependent upon the scale at which the fingerprint images are scanned, we need to calculate it for the specific sensor resolution. We used a database (GT) consisting of 450 mated pairs of fingerprints acquired using an optical scanner (from Identix Inc.) at a resolution of 500 dpi. The second print in the mated pair was acquired at least a week after the first print. The minutia were manually extracted from the prints by a fingerprint expert. The expert also determined the correspondence information for the detected minutiae. Using the ground truth correspondence information between duplex (two) pairs of corresponding minutiae, a rigid transformation between the mated pair was determined. The overall rigid transformation between the mated pair was determined using a least square approximation of the candidate rigid transformations estimated from each duplex pairs of the corresponding minutiae. After aligning a given mated pair of fingerprints using the overall transformation, (δ_x, δ_y) for each corresponding minutia pair was computed; distance offset $\left(d = \sqrt{\delta_x^2 + \delta_y^2}\right)$ estimates for all minutiae in all mated fingerprint pairs were pooled to obtain a distribution for distance between the corresponding minutiae. We are seeking that value of r_0 for which $P(d \leq r_0) \geq 0.975$, i.e., the value of r_0 which accounts for atleast 97.5% of variation in the minutiae positions of genuine fingerprint matchings. The value of r_0 is found to be 15 pixels for fingerprint images scanned at 500 dpi resolution.

To estimate the value of l, we first estimate the value of θ_0. The value of θ_0 can also be estimated using database GT. After aligning a given mated pair of fingerprints using the overall transformation, we seek that value of θ_0 which accounts for 97.5% variation in the minutia angles in the genuine fingerprint matchings, i.e., we seek that value of θ_0 for which $P(\min(|\theta' - \theta|, 360 - |\theta' - \theta|) \leq \theta_0) \geq 0.975$. The value for θ_0 is found to be $\theta_0 = 22.5^\circ$. In the second step, we determine the distribution $P(\min(|\theta' - \theta|, 360 - |\theta' - \theta|))$ for the impostor fingerprint matchings. Since we do not have correspondences marked by an expert between impostor fingerprint pairs, we depend on the automatic fingerprint matcher to establish correspondences between minutiae in impostor pairs. Again, we obtained the distribution from the GT database from which we determined that $P(\min(|\theta' - \theta|, 360 - |\theta' - \theta|) \leq 22.5^\circ) = 0.267$, i.e., $l = 0.267$. Note that under the assumption that minutiae directions are uniformly distributed and the directions for the minutiae that match in their location are independent, we obtain $l = \frac{2 \times 22.5}{360} = 0.125$. If minutiae orientations are considered instead of directions, the value for l determined from the experiments will be 0.417 as opposed to a value of $l = \frac{2 \times 22.5}{180} = 0.25$ determined under the assumption stated above.

The value of w was taken as reported by Stoney [9]. Stoney estimated the value of ridge period as 0.463 mm/ridge from a database of 412 fingerprints. For fingerprint sensors with a resolution of 500 dpi, the ridge period converts to ~ 9.1 pixels/ridge. Thus, $w \sim 9.1$.

The above analysis is based on the following additional assumptions: (i) Ridge widths are same across the population and spatially uniform in the same finger. This assumption is justified because the pressure variations could make non-uniform ridge variations uniform and vice versa. Further, there may be only a limited discriminatory information in the ridge frequency. (ii) The analysis of matchings of different impressions of the same finger binds the parameters of the probability of matching a minutiae in two fingerprints from different fingers. (iii) We assume that there exists one and only one alignment between the template and the input minutiae sets.

4. Experimental Results

Fingerprint images were collected in our laboratory from 167 subjects using an optical sensor manufactured by Digital Biometrics, Inc. (image size = 508×480, resolution = 500 dpi). Single impressions of the right index, right middle, left index, and left middle fingers for each subject were taken in that order. This process was then repeated to acquire a second impression. The fingerprint images were collected again from the same subjects after an interval of six weeks in a similar fashion. Thus, we have four impressions for each of the four fingers of a subject. This resulted in a total of $2,672$ ($167 \times 4 \times 4$) fingerprint images. We call this database DBI. Using the protocol described above, we also collected fingerprint images using a solid-state fingerprint sensor manufactured by Veridicom, Inc. (image size = 300×300, resolution = 500 dpi). We call this database

VERIDICOM. A large number of impostor matchings (over $4,000,000$) were generated using an automatic fingerprint matching system [4].

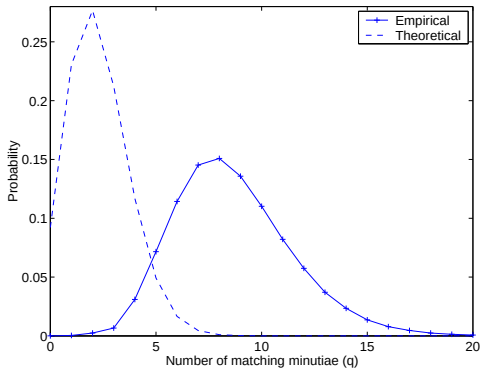


Figure 4. Comparison of experimental and theoretical probabilities for the number of matching minutiae in impostor fingerprint matches for the DBI database.

The mean values of m and n for the impostor matchings were estimated as 46 for the DBI database and as 26 for the VERIDICOM database. The average value of A for the DBI and the VERIDICOM databases was estimated to be $67,415$ pixels and $28,383$ pixels, respectively.

The probabilities of fingerprint correspondence obtained for different values of M, m, n, and q are given in Table 2. The probability values obtained from our model shown in Table 2 can be compared with values obtained from the previous models in Table 1 for $m = 36$, $n = 36$, and $q = 36, 12$.

Let us consider a typical latent print matching exercise. In a typical latent fingerprint examination, an expert matches a previously recorded entire fingerprint (template) with a partial (latent) print captured from the scene of crime. This examination consists of visually determining the overlapping area between the latent and the template and matching *all* the minutiae in the overlapping area with *all* the conflicting evidence that can be explained exogenously (e.g., dirt) [3]. Typically, a match consisting of 12-points (*the 12-point rule*) is considered as sufficient evidence in many courts of law. Assuming that an expert can correctly glean all the minutia in the latent, a 12 point match (see the last entry in Table 2) is an overwhelming amount of evidence, *provided* that there is no contradictory minutia evidence in the overlapping area.

Figures 4 shows the distribution of number of matching minutiae computed from the DBI database using the automatic fingerprint matching system [4]. This figure also shows the theoretical distributions obtained from our model described in Section 3 for the average values of M, m, and n computed from the DBI database. A similar behavior is shown by the distributions on the VERIDICOM database.

M, m, n, q	P(Fingerprint Correspondence)
104, 26, 26, 26	5.27×10^{-40}
104, 26, 26, 12	3.87×10^{-9}
176, 36, 36, 36	5.47×10^{-59}
176, 36, 36, 12	6.10×10^{-8}
248, 46, 46, 46	1.33×10^{-77}
248, 46, 46, 12	5.86×10^{-7}
70, 12, 12, 12	1.22×10^{-20}

Table 2. Fingerprint correspondence probabilities obtained from the proposed individuality model for different sizes of fingerprint images containing 26, 36 or 46 minutiae. M for the last entry was computed by estimating typical print area manifesting 12 minutia in a 500 dpi optical fingerprint scan.

The empirical distribution is shifted to the right of the theoretical distribution, which can be explained by the following factors: (i) some true minutiae are missed and some spurious minutiae are detected by the automatic system due to noise in the fingerprint images; (ii) the automatic algorithm cannot completely recover the non-linear deformation present in the fingerprint images; so the alignment between the input and template has some error. (iii) automatic feature extraction introduces error in minutiae location and orientations. (iv) the matcher seeks that alignment which maximizes the number of minutiae correspondences; consequently, the chance of false association increases.

The theoretical curve in Figure 4 provides an upper bound on the performance of an automatic fingerprint verification system; thus, it is possible to improve the performance of automatic fingerprint matching systems. At the same time, an automatic system can not perform better than the theoretical limit because of the limited information content in the minutiae-based representation.

Table 3 shows the empirical probability of matching 10 and 15 minutiae between two impostor fingerprints in VERIDICOM and DBI databases, respectively. The "typical" values of m and n, were estimated from the empirical distributions derived from our databases. The fingerprint correspondence probabilities (false acceptance rates) obtained on these databases are consistent with those obtained on similar databases by several other state-of-the-art automatic fingerprint verification systems reported in the FVC2000 Fingerprint Verification Competition [12]. On the other hand, the performance claims by several fingerprint verification system vendors vary over a large range (a false acceptance rate of 10^{-9} to 10^{-3}) due to the absence of standardized testing protocols and large standardized databases.

Database	m,n,q	P(Fingerprint Correspondence)
VERIDICOM	26, 26, 10	1.7×10^{-2}
DBI	46, 46, 15	1.4×10^{-2}

Table 3. Fingerprint correspondence probabilities obtained from matching impostor fingerprints for the VERIDICOM and DBI databases.

5. Conclusions

One of the most fundamental questions one would like to ask about any *practical* biometric authentication system is: what is the inherent discriminable information available in the input signal? Unfortunately, this question, if at all, has been answered in a very limited setting for most biometrics modalities, including fingerprints. The inherent signal capacity issue is of enormous complexity as it involves modeling both the composition of the population as well as the interaction between the behavioral and physiological attributes at different scales of time and space. Nevertheless, a first-order approximation to the answers to these questions will have a significant bearing on the acceptance of fingerprint- (biometrics-) based personal identification systems into our society as well as determining the upper bounds on scalability of deployed systems.

The model proposed here is relatively simple. It ignores most of the known (weak) dependencies among the features and does not directly include features such as ridge counts, fingerprint class, ridge frequencies, permanent scars, etc.[1] For these reasons, we suspect that the proposed model does not yet compete in predicting the performance of a human fingerprint expert matcher. By additionally considering a more detailed fingerprint representation (e.g., different minutiae types, sweat pore information), the confidence in genuine mates can be reinforced and the spurious associations among the impostors can be ruled out. In spite of the simplicity of our model, we believe that the individuality estimates predicted by this model is significantly closer to the performance of available automatic fingerprint matchers on realistic data samples than any other model. The extension of our proposed model to include additional features is a topic for our future research.

While the individuality of the minutiae based fingerprint representation based on our model is lower (i.e., the probability of false association is higher) than the previous estimates, our study indicates that the likelihood of an adversary guessing someone's fingerprint pattern (e.g., requiring matching 20 or more minutiae from a total of 36) is significantly lower than a hacker being able to guess a six character alpha-numerical case-sensitive (most probably weak) password by social engineering techniques or by brute force. Obviously, more stringent conditions on matching will provide a better cryptographic strength at the risk of increasing the false negative error rate.

The individuality problem in its present form is an ill-formulated problem in an information theoretic sense. A clear insight into this problem entails an understanding of the interactions among a number of confounding factors. It is only to be expected that the fingerprint-based personal identification, being one of the most mature, most well-understood, with strongest legitimate support from the biometrics community would be the first biometrics to be challenged for objective quantification of its distinctiveness. We believe that by objectively and quantitatively addressing individuality related issues, difficult as they may be, will force us to formalize the concepts of individuality. This eventually will lead us to establish the standards not only for other biometrics (e.g., face identification) but may also lay foundations for characterization/evaluation of high entropy complex pattern recognition systems.

[1] It also does not completely account for errors in the minutiae detection due to image acquisition problems/poor quality fingerprints which increase the error rates in both manual as well as automatic fingerprint matching systems.

References

[1] Daubert v. Merrell Dow Pharmaceuticals, 113 S. Ct. 2786 (1993).

[2] U.S. v. Byron Mitchell, Criminal Action No. 96-407, U.S. District Court for the Eastern District of Pennsylvania.

[3] S. A. Cole, "What Counts for Identity?" *Fingerprint Whorld*, Vol. 27, No. 103, pp. 7-35, 2001.

[4] A. K. Jain, L. Hong, S. Pankanti, and R. Bolle, "An Identity Authentication System Using Fingerprints," *Proc. IEEE*, Vol. 85, No. 9, pp. 1365-1388, 1997.

[5] H. Cummins and C. Midlo, Fingerprints, Palms and Soles: An Introduction to Dermatoglyphics, Dover Publications, Inc., New York, 1961.

[6] J. Osterburg, T. Parthasarathy, T. E. S. Raghavan, and S. L. Sclove, "Development of a Mathematical Formula for the Calculation of Fingerprint Probabilities Based on Individual Characteristics", Journal of the American Statistical Association, Vol 72, No. 360, pp. 772-778, 1977.

[7] S. L. Sclove, "The Occurrence of Fingerprint Characteristics as a Two Dimensional Process", *Journal of American Statistical Association*, Vol. 74, No. 367, pp. 588-595, 1979.

[8] D. A. Stoney and J. I. Thornton, "A Critical Analysis of Quantitative Fingerprint Individuality Models", *Journal of Forensic Sciences*, Vol. 31, No. 4, Oct 1986, pp. 1187-1216.

[9] D. A. Stoney, "Distribution of Epidermal Ridge Minutiae," *American Journal of Physical Anthropology*, Vol. 77, pp. 367-376, 1988.

[10] M Trauring, "Automatic Comparison of Finger-ridge Patterns", *Nature*, pp. 938-940, 1963.

[11] M. R. Stiles, "Government's post-Daubert Hearing Memorandum," United States District Court for the Eastern district of Pennsylvania, USA vs Mitchell, Criminal case No. 96-00407, http://www.usaoedpa.com/Invest/Mitchell/704postd.htm, 2000.

[12] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain, "*FVC2000:* Fingerprint Verification Competition", 15th IAPR International Conference on Pattern Recognition, Barcelona, Spain, Sep. 3-7, 2000. http://bias.csr.unibo.it/fvc2000/.