Not Such a Small Whorl After All

Brian Camley
Pascal Getreuer
Bradley Klingenberg
University of Colorado at Boulder
Boulder, CO

Advisor: Anne M. Dougherty

Summary

Fingerprint identification depends on the assumption that a person's fingerprints are unique. We assess the truth of this assumption by calculating the total number of distinct fingerprints.

We assume accurate fingerprints (ignoring procedural error) that are defined by 12 points of detail or *minutiae*. The number of distinct fingerprints depends also on the number of potential positions of these minutiae. Two historical methods and a geometric analysis estimate there to be 1,400 positions, a figure confirmed by our algorithm for counting ridges in a fingerprint.

We create two models to estimate the number of unique fingerprints:

- One model computes fingerprints as arrangements in minutiae;
- the other extrapolates the number of fingerprints from the Shannon entropy of the information that defines a fingerprint.

These two models agree to within an order of magnitude that there are 5×10^{33} unique fingerprints, a compelling validation of our general approach.

To handle the large number of fingerprints, we implement an approximation for the calculation of probabilities. Given a cumulative world population of 120 billion [Catton 2000], the probability of two people ever having the same fingerprint is 1.4×10^{-6}.

The probability of two humans living today sharing a fingerprint is 3.5×10^{-15}, which suggests that fingerprints are a theoretically more reliable method of identification than DNA analysis, which has a false positive probability of 10^{-9}. None of these calculations take into account procedural errors.

The UMAP Journal 25 (3) (2004) 245–258.

Introduction

The possibility of duplicate fingerprints, or fingerprints likely to be mistaken for each other, has led to recent criticism of fingerprints as a means of identification [Pankanti et al. 2001]. The key problem is:

How many distinct fingerprints are there?

We approach this problem with two general methods:

- "building" a fingerprint from the ground up, using different models; and
- using the information content of a fingerprint.

Assumptions

- **Fingerprint matching is done by comparing minutiae.** Comparing small features (minutiae) such as ridge endings and bifurcations within a fingerprint is a typical method (used by the FBI) for recognition of identity, and it provides very good accuracy [Andrew 2002; Hrechak and McHugh 1990; Jain et al. 2001].
- **Two fingerprints with the same minutiae configuration are identical.**
- **The distribution of minutiae within a fingerprint is uniform.** In fact, the minutiae are over-dispersed, but the uniform distribution is a good estimate for their locations [Stoney 1988].
- **Minutiae are statistically independent of one another.** This assumption is justified by studies of fingerprint individuality (Galton and Henry) that assume independence [Stoney 1988; Stoney and Thornton 1986].
- **Minutiae are either directed along the flow of the ridge of a fingerprint, or against it, that is, there are only two possible directions.** Attempting to measure more than two directions is very difficult [Stoney and Thornton 1986].
- **There is only one type of minutia, bifurcation.** We make this assumption to simplify the problem and to avoid dealing with minutiae (i.e., dots) that have no direction.
- **There are no errors in collection—we are dealing with "true" fingerprints.** The greatest source of error in fingerprint identification is not in recognition but in training of employees and the condition of equipment [Fickes 2003]. In addition, determining the minutiae of a fingerprint is nontrivial.

Model 1: Designing from the Ground Up

The Worst-Possible Case

There are no more than $10^{35,000}$ *possible fingerprints.*

The FBI standard for storing and comparing fingerprint data uses 500 dpi (250,000 pixels per square inch), with 8 bits per pixel and an average size of 1.5 square inches per fingerprint [Aboufadel and Schlicker 1999]. If the image is stored as a bitmap, there are $250,000 \times 1.5 \times 8 = 3 \times 10^6$ bits of information in a fingerprint. This implies an absolute maximum of $2^{3,000,000}$ possible fingerprint images without a pixel-for-pixel match.

However, the FBI does not store images in bitmapped form but instead uses the Cohen-Daubechies-Feauveau 9/7 or "Spline 9/7" wavelet for compression, by a factor of 26:1 with the thresholding used by the FBI. So there can be only $3 \times 10^6 \div 26 = 1.15 \times 10^5$ bits of information within a stored fingerprint. We compressed several typical fingerprints [Bio-Lab 2000] to about 26:1 using our implementation of the wavelet. Comparison of edges in the original and compressed images shows that information about the minutiae is lost at 52:1 and higher levels of compression (**Figure 1**).

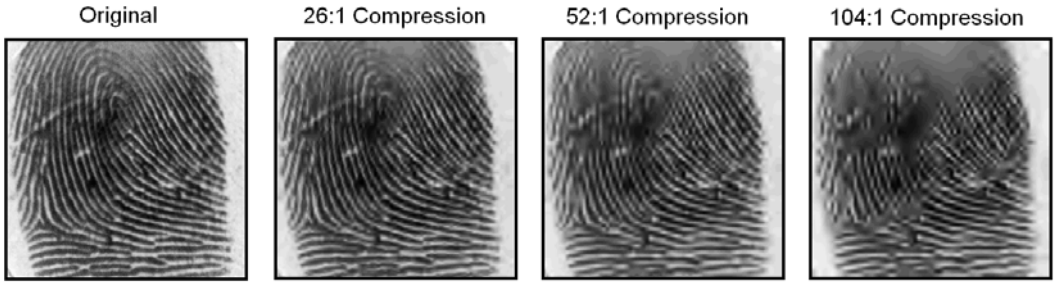

Figure 1. CDF 9/7 wavelet compression. Minutiae detail, such as bifurcations, is lost at compression greater than 26:1.

These results give the maximum number of fingerprint images as $2^{1.15 \times 10^5} \approx 10^{35,000}$.

Limited Space for Minutiae

In the previous subsection, we didn't make any assumptions about the image—it could have been a picture of a moose. If minutiae are limited to L physical locations, with at most one per location, the number of distinct fingerprints determined by n minutiae is $\binom{L}{n}$.

The Lower Bound for Total Fingerprints

Minutiae always occur on ridges, so it makes sense to represent the fingerprint as a set of ridges. We consider a typical fingerprint to be a set of 20 concentric ellipses. We assume that there are no minutiae closer than 5° away from each other—in other words, they are reasonably separated. Essentially, we are assuming that the minutiae are more or less uniformly distributed. These assumptions are equivalent to restricting minutiae to the intersections of 20 ellipses and 72 equally spaced radial lines (a simplified version of this is seen in **Figure 2**). This is similar to the empirical Roxburgh model [Stoney and Thornton 1986].

There are therefore $20 \times 72 = 1{,}440$ possible locations for minutiae.

Figure 2. Potential minutiae locations are on intersections of concentric ellipses and radial lines; 1,440 locations (far right) represent a fingerprint.

What number of minutiae should we choose? A typical fingerprint has 30 to 40 minutiae, but not all of these are significant. In fact, even as few as 6 minutiae may be important [Bhowmick et al. n.d.].

The number of distinct fingerprints that can be created by arrangements of 6 minutiae in 1,440 possible locations is :

$$\binom{1440}{6} = \frac{1440!}{(6!)(1440-6)!} \approx 1.23 \times 10^{16}.$$

This is a lower bound on the total number of distinguishable fingerprints.

Improving the Estimate

Though in some cases there are only 6 useful minutiae, typically there are about 30 to 40 minutiae in a fingerprint [Stoney and Thornton 1986]. If all of these were used for identification, and there were still only 1,440 possible locations for minutiae, then the value for the number of total fingerprints would be

$$\binom{1440}{35} \approx 2.23 \times 10^{70}.$$

However, the criteria for identity vary from about 10 to 16 matching features [Stoney and Thornton 1986], implying that using 35 features to define a fingerprint overestimates the number of fingerprints. We assume that 12 features are required to match a fingerprint (the FBI's "quality assurance" standard [Duffy 2002]). Then the number of distinguishable fingerprints is

$$\binom{1440}{12} \approx 1.59 \times 10^{29}.$$

Accounting for Direction of Minutiae

Allowing two orientations (with or against the flow of the ridge) for a minutia doubles the number of possible placements and increases the number of fingerprints to

$$\binom{2880}{12} \approx 6.64 \times 10^{32}.$$

Hence, the number of fingerprints is bounded between 1.23×10^{16} and 2.23×10^{70} but is most likely around 6.64×10^{32}.

We now narrow this range by using information theory.

Model 2: Information Theory

Clarification of Assumptions

We phrase our original assumptions in a new way:

- A single fingerprint contains n minutiae.
- A fingerprint can be effectively mapped as n minutiae falling into the squares of an $x \times x$ grid (at most one minutia per square). There are x^2 possible locations for the n minutiae, so $x = \sqrt{L}$, where L is the number of locations for minutiae.

Derivation of the Entropy of a Fingerprint

We can visualize a fingerprint as an x by x grid. If there is a minutia in a space, we mark it with an X; if not, we leave it blank (**Figure 3**).

We treat the two mutually exclusive states, minutia and non-minutia, as the elements of a two-letter alphabet, a. An x by x arrangement of this alphabet represents a fingerprint. Shannon's classic first-order equation gives the entropy of the alphabet:

$$H = -\sum_{i=1}^{m} P(a_i) \log_2 P(a_i), \qquad \textbf{(1)}$$

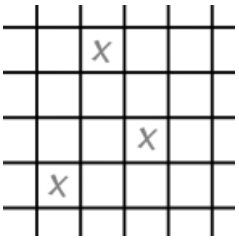

Figure 3. A section of a possible fingerprint configuration.

where $P(a_i)$ is the independent probability of the state i occurring in the fingerprint and m is the length of the alphabet [Shannon 1948].

For us, $m = 2$. The respective probabilities of each state (minutia or non-minutia) occurring in the alphabet are:

$$P(a_1) = \frac{\text{minutiae}}{\text{number of spaces}} = \frac{n}{x^2}, \qquad P(a_2) = \frac{\text{non-minutiae}}{\text{number of spaces}} = \frac{x^2 - n}{x^2}.$$

Substituting back into **(1)**, we are left with the following value for H:

$$H = -\left[\frac{n}{x^2}\log_2\left(\frac{n}{x^2}\right) + \frac{x^2-n}{x^2}\log_2\left(\frac{x^2-n}{x^2}\right)\right].$$

The Number of Fingerprints Based on Entropy

Because entropy is a measure of the minimum number of bits required to represent each element of the grid, it can be used to determine the total representative requirement of any fingerprint:

$$\text{bits required for fingerprint} = H \times \text{size of fingerprint} = Hx^2.$$

There are Hx^2 bits of information in a fingerprint, so there should be 2^{Hx^2} possible fingerprints. However, in our definition of the grid, we ignored the direction of minutiae. Each minutia has two possible directions, resulting in a total of 2^n possible directional configurations.

Combining these numbers, we find $2^{Hx^2} \times 2^n = 2^{Hx^2+n}$ possible fingerprints.

Using the values established earlier ($n = 12$, $L = x^2 = 1{,}440$), we get

$$H = -\left[\frac{12}{38^2}\log_2\left(\frac{12}{38^2}\right) + \frac{38^2-12}{38^2}\log_2\left(\frac{38^2-12}{38^2}\right)\right] \approx 0.07,$$

which leads to

$$H \times x^2 \approx 0.07 \times 38^2 \approx 100$$

and therefore $2^{Hx^2+n} = 2^{100+12} \approx 5.19 \times 10^{33}$ fingerprints.

Consistency of the Two Models

Experimentation with different values for L and n—not only for reasonable ranges of L (500–2,500) and n (6–18) but also for truly ridiculous numbers—indicates that the values from the combinatorial model and from the information theory agree to within an order of magnitude.

How Many Minutiae Locations Are There?

The physical dimensions of a minutia confirm the estimate of $L \approx 1{,}400$ in two different ways.

The Kingston Method

A visual inspection of a 300 $\times$ 300 pixel fingerprint image [Bio-Lab 2000] reveals that a minutia can be contained in a 9 $\times$ 9 pixel square (**Figure 4**). This would suggest that we can put a maximum of

$$\frac{300 \times 300}{9 \times 9} \approx 1{,}100$$

minutiae into one image. This method for estimating the number of possible minutiae locations in a fingerprint recalls the Kingston method, which calculates this value based on the area occupied by a minutia [Stoney and Thornton 1986]. This value for L confirms our initial geometric estimate.

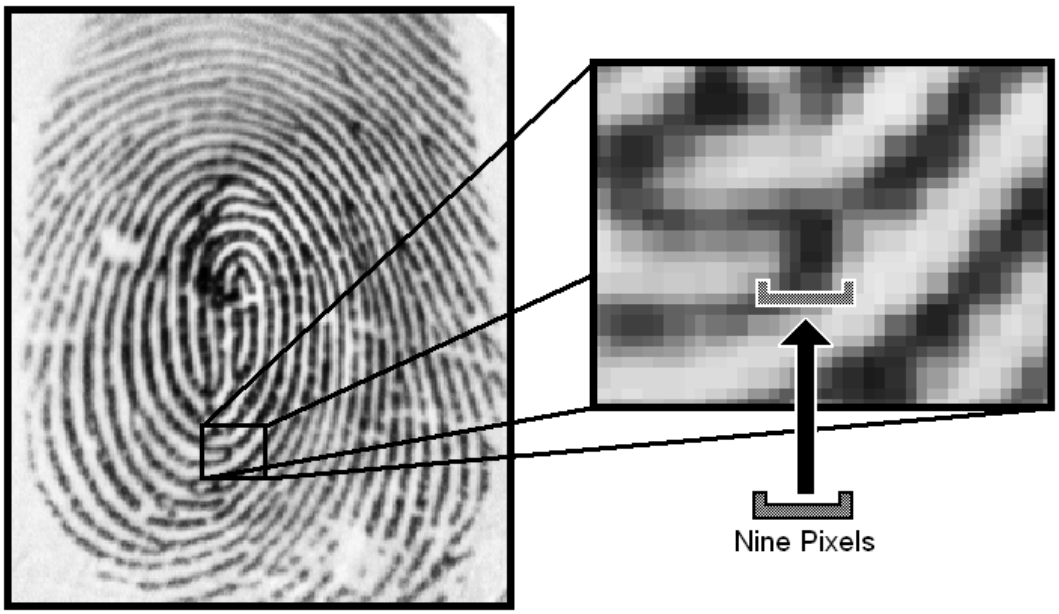

Figure 4. A minutia can typically be contained in a 9-by-9 pixel area.

The Amy Method

Amy's method [Stoney and Thornton 1986] calculates the number of potential minutiae positions L as

$$L = (\mathcal{I} - \iota + 1)^2,$$

where $\mathcal{I}$ is the number of ridge intervals on a side of a known fingerprint (see **Figure 5**). The studies of Roxburgh established the value of ι, the size of a minutia, to be $\sqrt{5/2}$ ridge intervals [Stoney and Thornton 1986].

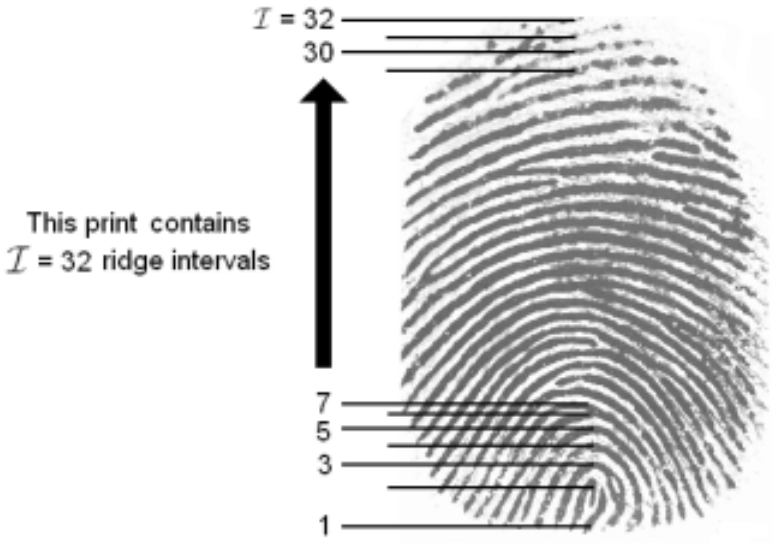

Figure 5. The number of ridge intervals on a fingerprint.

The average value of $\mathcal{I}$ is 38 for a typical fingerprint, from our ridge-counting algorithm (see the **Appendix**). We calculate the number of potential minutiae locations as

$$L = \left(38 - \sqrt{\frac{5}{2}} + 1\right)^2 \approx 1{,}400.$$

This value is almost exactly the number that we predicted using only the initial geometry of the fingerprint!

The geometry of fingerprint images and Amy's method confirm, in two unrelated ways, our prediction of $L = 1{,}440$.

Estimating the Odds of Duplicate Prints

General Method

Select a random person from a group of N. The probability that a second person selected at random has a different fingerprint from the first is $(f-1)/f$,

where f is the total number of fingerprints. Generalizing, as in the classic birthday coincidence problem, the probability $P(N)$ of picking N unique fingerprints (that is, no duplication) is

$$P(N) = \prod_{i=1}^{N-1} \left(\frac{f-i}{f} \right) = \frac{1}{f^{N-1}} (f-1)(f-2)(f-3)\cdots(f-N+1). \quad \textbf{(2)}$$

We calculated the probability of a duplication this by writing a C program, using arbitrary precision arithmetic to deal with the fact that f is very large. However, this calculation requires a lot of time for large N, so it is useful to have an easy-to-calculate approximation.

Approximation

We express **(2)** as

$$P(N) = \frac{1}{f^{N-1}} \left(f^{N-1} + c_1 f^{N-2} + c_2 f^{N-3} + \cdots + c_{N-1} \right),$$

where the c_i are integer coefficients of the powers of f. We can then write

$$P(N) = 1 + \frac{c_1}{f} + \frac{c_2}{f^2} + \cdots + \frac{c_{N-1}}{f^{N-1}}.$$

Since f is large, $f^{-2} \ll f^{-1}$ and we can discard everything except for the first term, as long as N is not of the same order as f, getting

$$P(N) \approx 1 + \frac{c_1}{f}.$$

Now, what is the coefficient c_1? The product $(f-1)(f-2)\cdots(f-N)$ is a sum of terms created by choosing either f or the number from each binomial. A term with f^{N-1} occurs when f is chosen for each binomial except one. There are N ways to do so, resulting in the coefficients $-1, -2, \ldots, -N$. Therefore,

$$c_1 = (-1) + (-2) + \cdots + (-N) = \frac{-(N^2+N)}{2}.$$

Hence

$$P(N) \approx 1 - \frac{N^2+N}{2f}.$$

The probability of a fingerprint duplication is

$$1 - P(N) = \frac{N^2+N}{2f}.$$

Table 1.
Probability of a fingerprint coincidence, for various numbers of fingerprints and population sizes.

N	Number of fingerprints		
	1.23×10^{16}	6.64×10^{32}	5.19×10^{33}
10^5	4×10^{-7}	10^{-22}	10^{-24}
10^6	4×10^{-5}	10^{-21}	10^{-22}
10^7	4×10^{-3}	10^{-19}	10^{-20}
10^8	0.334	10^{-22}	10^{-18}
10^9	0.999	10^{-22}	10^{-16}
6×10^9 (current world)	1	10^{-22}	3×10^{-15}
120×10^9 (cumulative world)	1	10^{-22}	1.4×10^{-6}

Odds of Misidentification

Table 1 gives the probability of a fingerprint coincidence for various population sizes, for each of our estimates of the number of fingerprints.

Based on either our best value (6.64×10^{32}) for the number of different prints, or the information theory estimate (5.19×10^{33}), there is essentially no chance of duplicating a fingerprint.

Conclusions

We use the basic geometry of a fingerprint and the known distribution of minutiae to determine that there are about 1,440 possible minutiae locations in a fingerprint. We confirm this by studying minutiae in digitized fingerprints.

Using this value, we calculate the number of possible distinct fingerprints given a certain number of minutiae to be used for identification. We choose the FBI "quality assurance" standard of 12 minutiae [Duffy 2002], which results in 6.64×10^{32} possible fingerprints, once minutiae direction is taken into account. This number was confirmed by using the amount of information in a model of a fingerprint, which estimated 5.19×10^{33} fingerprints using 12 minutiae.

If only six minutiae are used to determine a fingerprint, and thus there are only 1.23×10^{16} fingerprints, then the probability for a duplication in one billion people approaches unity. In fact, even in only 100 million people (the order of the FBI's fingerprint database), there would be a reasonable probability (.33) of a fingerprint duplication.

However, using 12 minutiae, the probability of a fingerprint duplication in six billion people is only 3×10^{-15}; the probability of a duplication among the 120 billion people who have ever lived is only 1.39×10^{-6}. Therefore, there is little risk of mistaken identity based on 12 minutiae, given perfect fingerprints. In the real world, fingerprints are not perfect, and the largest sources of error

are from mishandling and errors in the process [Fickes 2003].

DNA analysis has a theoretical probability of false positives on the order of 10^{-9}, though this figure is often disputed [Thompson et al. 2003]. Fingerprint identification is thus theoretically more reliable than DNA testing.

Strengths and Weaknesses of the Models

Strengths

- **Agreement of the models.** The same general number of possible minutiae locations is calculated in three completely separate ways (by assuming uniform distribution, by looking at the size of a minutia, and by counting ridges and using Amy's method). This consistency suggests that our result is reasonable. In addition, our two vastly different models (combinatorial and information theory) produce consistent numbers.

Weaknesses

- **Assumptions about the minutiae.** We assume that there is only one type of minutia, when in fact there are at least three (bifurcations, endpoints, and dots) [Stoney and Thornton 1986]. Some of these are orientable and others are not, but we assume that all minutiae have two directions. In addition, though our assumption of uniform distribution is borne out by study of real fingerprints, correlations between minutiae could strongly skew the number of possible locations.

- **The models work only in an ideal setting.** Many factors can create errors in a fingerprint, such as dirt, operator error, and mechanical breakdowns. Our models do not address this and only set a maximum theoretical accuracy for fingerprinting.

Appendix: Ridge-Counting Algorithm

Our ridge-counting algorithm, which estimates the number of concentric ridges in an image, supplies useful empirical results to support and validate our theoretical work. The steps of the algorithm are:

- edge detection and thresholding,
- selecting the ridge core location, and
- counting maximum number of ridges around the core.

Edge-Detection and Thresholding

Given an image $X(x, y)$, we use edge detection filters f_x and $f_y$to estimate the image gradient. These are directional derivatives of the general Gaussian function (**Figure 6**).

$$f_x(x, y) = -x \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right), \qquad f_y(x, y) = -y \exp\left(-\frac{y^2}{2\sigma_y^2} - \frac{x^2}{2\sigma_x^2}\right).$$

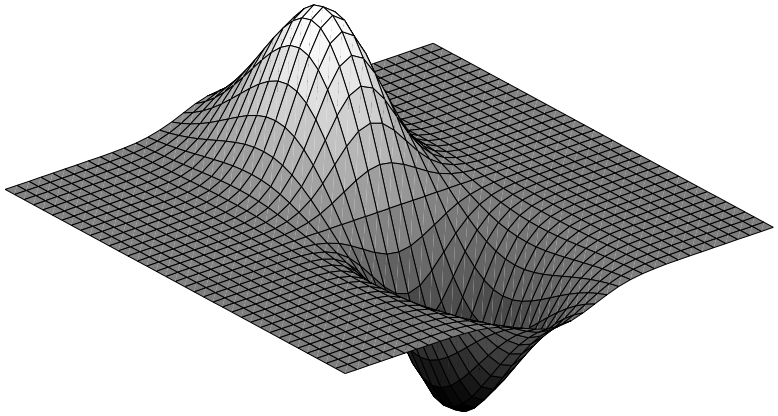

Figure 6. Edge detector filter.

With f_x operating horizontally and f_y vertically, the magnitude of the image gradient is

$$\nabla X(x, y) \approx \sqrt{[(X * f_x)(x, y)]^2 + [(X * f_y)(x, y)]^2}.$$

Thresholding the gradient yields a binarized image E of the edges,

$$E(x, y) = \begin{cases} 1, & \text{if } \nabla X(x, y) > \alpha; \\ 0, & \text{if } \nabla X(x, y) \le \alpha, \end{cases} \qquad (\min \nabla X < \alpha < \max \nabla X)$$

where α is the threshold level.

Selecting the Ridge Core Location

Most fingerprints are essentially concentric curves around a central part of the print, the ridge core. To find the number of ridges, one can begin at the core and move outwards, counting the ridges crossed in the path. Unfortunately, automatically determining the core location is difficult. One method to do this estimates the direction of the ridges through the directional field [Chan et al. 2004]. In this method, the image is segmented into $w \times w$-pixel blocks and a least-squares orientation method finds the smoothed orientation field $O(x, y)$. The value $\sin\big(O(x, y)\big)$ reflects the local ridge direction and indicates the core, which is where the direction curves fastest.

It is very easy to locate the core of a fingerprint by eye; and small deviations in the location of the core, like those expected from human error, do not make large differences in the measured number of ridges.

Maximum Number of Ridges around the Core

The next step is to use the $E(x, y)$ image to count the ridges, beginning at the core and counting the number of changes between 1 and 0. A typical edge is counted twice and a typical ridge has two edges; so dividing the count by 4 estimates the number of ridges. However, moving along only one path may miss the shorter ridges at the edge of the print or ridges with noisy edges. Instead, we count along many directions and use the highest count (**Figure 5**).

Results from Implementation

We applied the algorithm to 80 optically-scanned fingerprint images from Bio-Lab [2000]. The last nine images were too noisy for the edge detection and were discarded. The distribution of the remainder is roughly symmetrical about a mean of 38.0 ridges, with a standard deviation of 5.3.

References

Aboufadel, Edward, and Steven Schlicker. 1999. *Discovering Wavelets*. 1st ed. John Wiley and Sons.

Bolle, Ruud M., Andrew W. Senior, Nalini K. Ratha, and Sharath Pankanti. 2002. Fingerprint minutiae: A constructive definition. In *Proceedings ECCV Workshop on Biometrics*. `http://citeseer.ist.psu.edu/591141.html` .

Bhowmick, P., A. Bishnu, B.B. Bhattacharya, M.K. Kundu, C.A. Murthy, and T. Acharya. n.d. Determination of minutiae scores for fingerprint image applications. `http://citeseer.ist.psu.edu/554334.html` .

Bio-Lab, Univerity of Bologna. 2001. FVC 2000: Fingerprint Verification Competition. `http://bias.csr.unibo.it/fvc2000/` .

Brualdi, Richard A. 1999. *Introductory Combinatorics*. 3rd ed. Upper Saddle River, NJ: Prentice Hall.

Catton, William R., Jr. 2000. Worse than foreseen by Malthus (even if the living do not outnumber the dead). `http://desip.igc.org/malthus/Catton.html` .

Chan, K.C., Y.S. Moon, and P.S. Cheng. 2004. Fast fingerprint verification using subregions of fingerprint images. *IEEE Transactions on Circuits and Systems*

for Video Technology 14 (1) (January 2004) 95–101. http://ieeexplore.ieee.org/xpl/abs_free.jsp?arNumber=1262035 .

Cotton, R.W., and C.J. Word. 2003. Commentary on Thompson et al. [2003]. *Journal of Forensic Sciences* 48 (5): 1200.

Duffy, S.P. 2002. Experts may no longer testify that fingerprints "match." *The Legal Intelligencer* (9 January 2002). http://www.truthinjustice.org/print-match.com .

Fickes, Michael. 2003. Dirt: A fingerprint's weakest link. *Access Control and Security Systems* (1 February 2003). http://govtsecurity.securitysolutions.com/ar/security_dirt_fingerprints_weakest/ .

Hrechak, A.K., and J. McHugh. 1990. Automated fingerprint recognition using structural matching. *Pattern Recognition* 23: 893–904.

Jain, Anil, Arun Ross, and Salil Prabhakar. 2001. Fingerprint matching using minutiae and texture features. In *Proceedings of the International Conference on Image Processing (ICIP)* (Thessaloniki, Greece), 282–285. http://citeseer.ist.psu.edu/jain01fingerprint.html .

Pankanti, Sharath, Salil Prabhakar, and Anil K. Jain. 2001. On the individuality of fingerprints. *IEEE Transaction on Pattern Analysis and Machine Intelligence*: 1010–1025. http://citeseer.ist.psu.edu/472091.html .

Shannon, C.E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27: 379–423, 623–656.

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

_____, and J.I. Thornton. 1986. A critical analysis of quantitative fingerprint individuality models. *Journal of Forensic Sciences* 31: 1187–1216.

Thompson, W.C., F. Taroni, and C.G.G. Aitken. 2003. How the probability of a false positive affects the value of DNA evidence. *Journal of Forensic Sciences* 48 (1): 47–54.

See Cotton and Word [2003] for a commentary.