Can't Quite Put Our Finger On It

Séamus Ó Ceallaigh
Alva Sheeley
Aidan Crangle
University College Cork
Cork, Ireland

Advisor: James J. Grannell

Summary

There are two main paths to identification through fingerprints.

- Global analysis relies on the specific arrangement and characteristics of the ridges of a print.

- Local analysis assumes that the individuality of a print is based on the position and orientation of the two basic types of minutiae.

We subdivide a print into a grid of square cells, consider the distribution of ridge features, and calculate probabilities from combinatorial analysis. We make predictions by refining parameters of the model, such as the number of minutiae required for a positive match between two generic prints, and the size of a cell in the main grid. We compare our results to previous studies and discuss the relation to DNA profiling. The simplicity of our model is its key strength.

We conclude that it is extremely unlikely that any two randomly selected people have, or have ever had, the same set of fingerprints.

Despite the apparently simplistic nature of fingerprinting, it is vastly more reliable in identification than a DNA comparison test.

Introduction

In recent years, the scientific basis of fingerprint analysis has been questioned, in the U.S. Supreme Court ruling in the Daubert case that the reliability of expert scientific testimony must be established along the following five criteria:

更多数学建模资料请关注微店店铺 "数学建模学习交流"
https://k.weidian.com/RHO6PSpA

1. whether the particular technique or methodology in question has been subject to a statistical hypothesis testing;

2. whether its error rate has been established;

3. whether the standards controlling the technique's operation exist and have been maintained;

4. whether it has been peer-reviewed and published; and

5. whether it has a general widespread acceptance.

Our model tries to address the first two issues. We aim to produce a probabilistic method to measure the uniqueness of a particular print.

Assumptions

- A thumbprint is defined globally by ridge patterns and locally by a distribution of *minutiae,* which we refer to also as *features.*

- The area of interest typical thumbprint is a 20 mm × 20 mm square grid.

- There are two significant types of minutiae, the bifurcation and the ridge ending: all other minutiae are compositions of these [Osterburg et al. 1977].

- The probability of a minutia occurring in a grid box is .234 [Osterburg 1977].

- The orientation of the minutiae was not taken into account by Osterburg; we assign a minutiae one of eight angles, from 0° to 157.5°, in steps of 22.5°.

- When comparing two prints, we know one print arbitrarily well.

- The number of people who have ever lived is 1.064×10^{11} [Haub 1995].

The Model

Global Analysis

Ridge Patterns And Orientation Fields

Global analysis concerns ridge patterns, which distinguish prints into six main pattern groups: Arch, Tented Arch, Left Loop, Right Loop, Twin Loop and Whorl. Each pattern is determined by an orientation field, which may have specific stationary points, known as the the *delta* and the *core.* If a print contains 0 or 1 delta points and 0 or 1 core points, then it is classified as Lasso, and Wirbel otherwise.

The Lasso class consists of arch, tented arch, right loop, and left loop.

- If the fingerprint has 0 delta points or 0 core points, then it is an arch.

- Otherwise, if the core point and the delta point are aligned in the vertical direction, then the fingerprint is an arch if the length between the core point and the delta point is less than 2.5 mm and a tented arch otherwise.

- Otherwise, if the core point is to the right of the delta point, the fingerprint is a right loop.

- Otherwise, the fingerprint is a left loop.

    The Wirbel class consists of the whorl and the twin loop classes:

- If there are exactly two core points and exactly two delta points, then the fingerprint is a whorl if the two core points are aligned horizontally and a twin loop otherwise.

- Otherwise, the fingerprint is a whorl.

    The main aim of global analysis is a vector field or orientation field to the ridge lines of a fingerprint. We must find suitable parameters for such functions that give rise to the different classes of ridge pattern. The most basic pattern without stationary points is the arch **Figure 1**, modeled by the simple system

$$\frac{dx}{dt} = \mu y, \qquad \frac{dy}{dt} = -\nu,$$

with parameters μ and ν. The orientation fields for other ridge patterns are more complex, so the bulk of our model is directed at the print's local features.
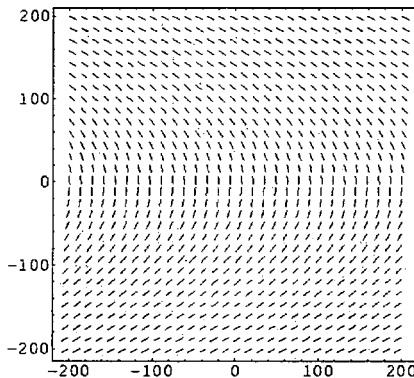

Figure 1. Arch orientation field.

Local Analysis

Estimates

    For an initial estimate of the probability of any two people having the same thumbprint, we must consider:

- the total number of people who have ever lived, estimated to be 1.064×10^{11} [Haub 1995] (that is, about 6% of all people are alive right now); and

- the total number of possible thumbprints that can be classified as "different."

To decide whether or not one thumbprint is the same as another, one must first decide on what exactly is a thumbprint. In our model, we take a typical area of the print as 20 mm × 20 mm, which is large enough to encompass the area of interest on any print. We divide this area into boxes 1 mm on a side, thus giving 400 boxes, each of area 1 mm^2. In principle, each box can be examined to determine whether or not it contains a minutia.

Minutiae are the features on a thumbprint that are used by almost all identification techniques to distinguish between prints. There are from 10 [Galton 1892] to 16 [Optel Ltd. 2003] different types of minutiae, but they are all composed of two fundamental types: ridge endings and ridge bifurcations. In our analyses, we consider only these two types. Also, if there are two minutiae in the same cell, it is impossible to resolve them separately.

Say that there are i resolvable features on the print. The number of ways that we can insert these i features into the 400 spaces in the grid is

$$\binom{400}{i}.$$

Remembering that there are two possibilities for each feature, the total number of combinations is then

$$\binom{400}{i} 2^i.$$

What is the value of i? The probability that any box contains a feature we take as .234 [Osterburg et al. 1977]. Then i has a binomial distribution:

$$P(i = x) = \binom{400}{x} (.234)^x (1 - .234)^{400-x},$$

with mean $\mu \approx 94$ and standard deviation $\sigma \approx 8$, so that the average number of cells containing features is 94.

Thus, the total number of possible thumbprints is

$$N = \sum_{i=0}^{400} \binom{400}{i} 2^i.$$

The binomial distribution for i, however, is concentrated mainly in the region $\mu - \sigma < i < \mu + \sigma$, or $94 - 8 < i < 94 + 8$. To be conservative, we consider only this range of numbers of minutiae; thus, there are approximately

$$N \approx \sum_{i=86}^{102} \binom{400}{i} 2^i \approx 1.19 \times 10^{128}$$

different thumbprints "available" for any actual thumb to hold. So very roughly,

$$P(\text{two people ever having the same thumbprint}) = \frac{1.064 \times 10^{11}}{1.19 \times 10^{128}} \sim 10^{-117}.$$

Comparison

This figure is the (approximate) probability that there have ever been two people who have had the same thumbprint. How might we take into account the chance that, when compared, two prints will be *judged* to be the same? To do this, we consider two hypothetical prints:

- A *control* print: an ideal known print, in which all i features are seen.

- A *sample* print: a print with more features i than the n available for comparison.

For two prints that are compared in a realistic circumstance, there will be at least $(i - n)$ features that are not included in the comparison. These features, in theory, could be in any combination of positions in the grid. The main question is: How many prints have $n < i$ features corresponding to a match? In other words, how many different ways can the remaining $(i - n)$ features be inserted into the grid and still produce a match with the control print? Knowing that, we can estimate how likely it is that two thumbprints not actually the same will match.

Incorrect Matching

We have $(i - n)$ features to distribute among $(400 - n)$ grid elements. The number of different ways to do this is, by previous reasoning,

$$\sum_{86}^{102} \binom{400 - n}{i - n} 2^{i-n}.$$

In criminal proceedings, a matching number of minutiae of anything from 8 [Collins 1992] to 15 [Vacca 2002] are accepted as conclusive proof of identification. Our model predicts that for $n = 12$, the total number of thumbprints that could have the same set of matching minutiae while not being the same print is $N = 1.3 \times 10^{117}$. But expressed as a fraction of the *total* number of possible prints, the probability of the print being one of these, if it selected from them, is

$$P(\text{false match}) = \frac{1.3 \times 10^{117}}{1.19 \times 10^{128}} \approx 1.09 \times 10^{-11}. \tag{1}$$

This is an extremely low probability.

Varying Parameters

The result **(1)** depends on the parameters, which can be varied according to circumstance and also as a way of refining the model:

- p: the probability of finding a feature in a grid cell. We take $p = .234$ [Osterburg et al. 1977]. Others [Thai 2003; Kingston 1964; Stoney and Thornton 1987; Dankmeijer et al. 1980] give values in the range $.19 < p < .25$.

- N: the number of cells in the grid. If there are more cells, on average, then more features will be observed, since p(feature) for a cell remains the same.

- n: the number of minutiae that one takes for comparison. We take $n = 12$.

- i: a variable, determined by p and N, that gives reasonable bounds for the summation.

- F: the number of different features that can appear in a grid cell. In our initial estimate, we take $F = 2$.

- L, A: the length of a side, and the area, of a grid cell. We take $L \approx 1$ mm, the average distance between features [Thai 2003].

If we wish to examine a thumbprint more closely, we should consider smaller and smaller areas of the print. It is not meaningful, however, to take L less than ~ 0.1 mm, since this is the typical ridge width.

The Dependence on L

We rework the model, taking the width of the generic grid cell to be 0.5 mm. Taking the overall area of the print to be the same, there are now 1,600 grid cells to consider, each with the same probability of having a feature. The binomial expression for i, the number of features observed on the whole print, is now

$$P(i = x) = \binom{1600}{x}(.234)^x(1 - .234)^{1600-x},$$

with mean $\mu = 374$ and standard deviation $\sigma = 17$. Thus, the region of relevance when summing is now $374 - 17 = 357 < i < 391 = 374 + 17$. This means that when the thumbprint is examined on a scale half that of the initial, about four times as many minutiae will be observed. Intuitively, the likelihood of a false match will decrease, since there are more possibilities for the number of prints:

$$N_{1600} = \sum_{i=357}^{391} \binom{1600}{i} 2^i \approx 3.5 \times 10^{502}.$$

The probability that any of the 100 billion people who have ever lived have had the same thumbprint is

$$P(\text{2 people ever having the same thumbprint}) = \frac{1.064 \times 10^{11}}{3.5 \times 10^{502}} \sim 10^{-492}.$$

We now determine the number of ways in which, when a certain number i of minutiae are selected for comparison, the remaining minutiae can be arranged. Following the same logic as before, this figure is

$$\sum_{357}^{391} \binom{1600 - n}{i - n} 2^{i-n},$$

which evaluates to 3.4×10^{491} for $n = 12$. The probability that any two compared thumbprints, judged to be identical by the standards of comparison, are actually different is therefore

$$P(\text{false match} \mid n = 12, N = 1600) = \frac{3.4 \times 10^{491}}{3.5 \times 10^{502}} \approx 10^{-11}.$$

This, interestingly, is not much greater than the probability for the previous estimate. The result is not therefore acutely dependent on the value of L, nor, by association, on the number N of grid cells. That said, it easy to examine details in any print to a scale of 0.5 mm.

The Dependence on p

The probability of a feature in a cell is, as of yet, a purely empirical figure. The formation of fingerprints, and their associated characteristics, is known: The foetus, at about 6.5 weeks, grows eleven "volar pads"—pouches on various locations of the hand [Anonymous 2001]. These shrink at about 11 weeks; and when they are gone, beneath where they lay are fingerprints. However, the mechanism of formation of the specific features is unknown. Genetic influences are present, but the environment is crucial also, evidenced by the fact that identical twins—who have the same DNA genotype—do *not* have the same fingerprints.

There is no way yet determined of predicting the frequency of occurrence of any type of minutia on the print of a particular person. Previous studies, cited in **Table 1**, show variation about ~ 0.2 minutiae/mm^2. It is not unreasonable to propose that the density depends on the print classification (i.e., whorl, loop, arch, etc.).

The range is $.204 < p < .246$. For 1,600 boxes, we have

$$P(\text{false match}) \approx 1.89 \times 10^{-12} \qquad \text{for } p = .204,$$
$$P(\text{false match}) \approx 1.78 \times 10^{-11} \qquad \text{for } p = .246.$$

The variation of p changes the final prediction by no more than an order of magnitude.

<div align="center">

Table 1.

The multiplication table of D_{10}.

</div>

Source	Number of prints	Mean density (minutiae/mm^2)
Osterburg et al. [1977]		.234
Dankmeijer et al. [1980]	1,000	.19
Stoney and Thornton [1987]	412	.223
Kingston [1964]	100	.246
Thai [2003]	30	.204

The Dependence on F

In out initial estimate, we take the number F of degrees of freedom of a feature in a print to be two: either a ridge ending or a ridge bifurcation. However, one can also consider the *orientation* of a feature. Each minutia lies on a ridge, which has a well-defined direction. We discretized this variable to one of eight possible directions, angles from 0° to 157.5°. Thus each feature, instead of having 2 degrees of freedom, now has 16.

The probability of a false match, taking 1,600 grid cells and a probability of occurrence $p = .234$, is now

$$P(\text{false match}) = \frac{\sum_{i=357}^{391} \binom{1600 - n}{i - n} 16^{i-n}}{\sum_{i=357}^{391} \binom{400}{i} 16^i}.$$

Taking $n = 12$ as before, we find that

$$P(\text{false match}) \approx 2.6 \times 10^{-22}.$$

This is an astonishingly smaller probability than the previous estimate of 10^{-11}. The orientation of a feature is no more difficult to determine in practice than its nature, so including it in the comparison process is a great improvement in efficacy with a modest increase in effort.

The Dependence on n

It is crucial to determine how many matching minutiae are necessary for a positive comparison. We have taken $n = 12$ in the preceding analyses; it is instructive to consider the variation of the probability of a false match with n. The graphs in **Figure 2** show that the probability falls off sharply, even as n increases beyond 1. A value of $n \approx 5$ is quite sufficient.
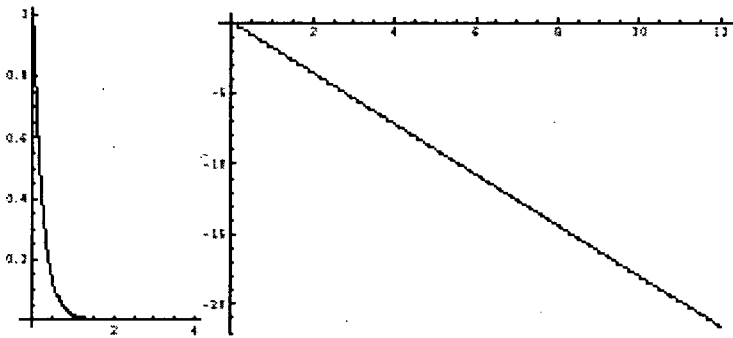
Figure 2. Probability of false match (linear and logarithmic scales) vs. n.

Conclusion

The Model

Our model, in the most general form, is

$$P(\text{false match}) = \frac{\sum_{i=\mu-\sigma}^{\mu+\sigma} \binom{N-n}{i-n} F^{i-n}}{\sum_{i=\mu-\sigma}^{\mu+\sigma} \binom{N}{i} F^{i}},$$

where

- F is the number of degrees of freedom and N is the number of grid cells,

- μ and σ are the mean and standard deviation of the binomial distribution determined by N and p,

- p is the probability of there being a feature in a cell, and

- n is the number of minutiae being used to make a comparison.

The preliminary result returned from our model, for $n = 12$ (a typical threshold for positive identification in many countries), is $P \approx 10^{-11}$. Further refinement of the parameters reduces this to $P \approx 7 \times 10^{-22}$. We conclude that 12 is a very reasonable comparison criterion, and that $n = 5$ or 6 is quite damning for any suspect so compared.

Thus, we conclude that to a very high degree of certainty, not only that no two people, now living or having ever lived in the past, have had the same thumbprint, but also that there is a vanishingly small chance that two prints are even close enough to be confused, given a small fraction of minutiae from their patterns to compare.

DNA Analysis

DNA identity testing is based on aspects of the DNA patterns called *loci*. For a 100 % match, the FBI [Thompson et al. n.d.] recommends that 13 loci be used. Using STR (Short Tandem Repeat) markers ensures that the inheritance profile at one location does not influence the inheritance at other locations. Each loci has two alleles, so 26 alleles must match. The FBI says that the possibility of a false match is 2.60×10^{-9} while other sources quote between 10^{-9} and 10^{-12}.

For two people chosen at random, the probability of a match based on the four most frequently analyzed alleles is between 1×10^{-5} and 1×10^{-8}. This is significantly higher than our estimated probability or a match for thumbprints. Hence, thumbprinting remains the most accurate form of biometric security known.

Strengths and Weaknesses

Strengths of the Model

- **Simplicity.** Our model is based on easily understood principles and simply expressed assumptions.

- **Realistic assumptions.**

- **Parameters.** The parameters in the model, such as the size of the grid-box, the total area, and the number of minutiae needed to match two thumbprints, can be easily varied.

- **Degrees of freedom.** In specifying two different kinds of possible minutiae and 8 orientation ranges for each one, the number of degrees of freedom is 16, greatly increasing the number of possible configurations of thumbprints and so minimising the probability of misidentification. Other studies [Osterburg et al. 1977; Galton 1892] do not take into account the orientation of the minutiae. By discretizing the directions of the features, we again keep the model simple.

- **Corroboration.** The probabilities returned by our model tie in with those given by previous studies by experts in the field (**Table 2**).

Table 2.

Comparison probabilities of studies.

Galton [1892]	1.45×10^{-11}
Osterburg et al. [1977]	1.33×10^{-27}
Stoney and Thornton [1987]	3.5×10^{-26}
our model	10^{-11} to 10^{-22}

Weaknesses of the Model

- **Multiple entries.** We assumed that in any given grid-box only one minutia can be present, which is sufficiently accurate for most types of minutiae. For example, both the bridge (consisting of two ridge bifurcations) and the spur (consisting of a ridge bifurcation and a ridge ending) have been defined [Osterburg et al. 1977] as being less than 2 mm in length. Thus if the bridge or spur is more than 0.707 mm in length, their constituent endings and bifurcations appear in different boxes and are counted as two separate minutiae.

  However, for minutiae consisting of ridge endings and ridge bifurcations in very close proximity, there is a chance that each will not be caught in a different box. An example is a dot. The distance between the two ridge endings that make up a dot is so small that it is unlikely that our model would catch these two occurrences of ridge endings in different boxes. A dot has been defined [Osterburg et al. 1977] as being large enough to encompass one pore, whose size ranges from 0.088 mm to 0.22 mm [Roddy and Stosz 1997]. Therefore, the two ridge endings will not appear in different boxes but will instead be misidentified as a single minutia.

- **Independence of minutia occurrence.** We assume that the placement of a minutia is completely unrelated to the placement of any others. This is not quite the case; there is a slight tendency for minutiae *not* to occur in direct proximity to each other.

- **Global analysis.** The overall ridge pattern of a thumbprint is entirely distinctive in its own right. We have not quantified this factor in our model.

Appendix: Classification

Minutiae

The ridges in a fingerprint or thumbprint form various patterns, those patterns being called *minutiae*. Ten different types are shown in **Figure A1**.

- **Ridge Ending.** A ridge ending occurs when a ridge ends abruptly. We define the orientation of a ridge ending as the direction the ridge came from.

- **Bifurcation.** A bifurcation is formed when two different ridges merge. We define the orientation as being the direction in which the merged ridge came from.

- **Island.** An island is a short ridge, comprised of two ridge endings whose orientations are in opposing directions. Two ridge endings occurring in neighbouring boxes with opposite configuration indicate the presence of an island.
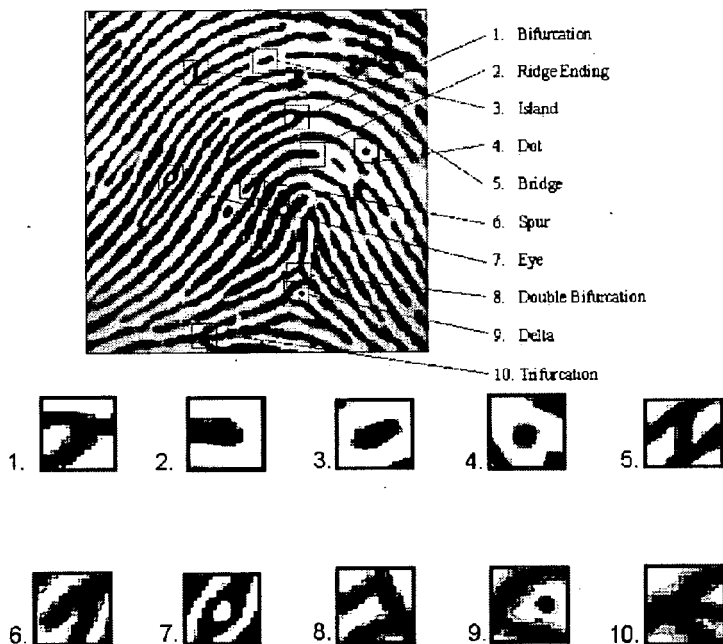
Figure A1. Ten types of minutia.

- **Dot.** A dot is an island but on a smaller scale.
- **Bridge.** A bridge is when a ridge branches out and merges with another ridge within a short region. It is composed of two bifurcations.
- **Spur.** A spur is when a ridge branches out and does not merge with another ridge. It is composed of one bifurcation and one ridge ending.
- **Eye.** An eye is formed by a ridge branching out into two ridges, and then recombining again a short distance later. It consists of two bifurcations.
- **Double Bifurcation.** As the name suggests, this type of minutia contains two bifurcations in succession.
- **Delta.** This type of minutia is composed of a dot and bifurcation, where the dot is between the merging ridges.
- **Trifurcation.** A trifurcation is a ridge that splits into 3 separate branches. It can be thought of as two bifurcations occurring in the same place.

Ridge Patterns

Figure A2 shows examples of the different classifications of a fingerprint, including the presence of cores and delta points.
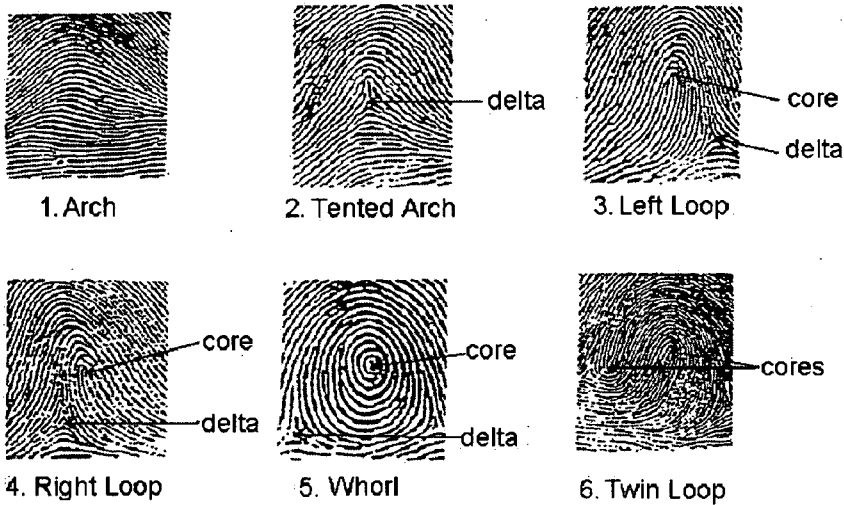
1. Arch   2. Tented Arch   3. Left Loop

4. Right Loop   5. Whorl   6. Twin Loop

Figure A2. Features in a fingerprint.

References

Anonymous. 2001. Friction skin growth. `http://www.ridgesandfurrows. homestead.com/Friction_Skin_Growth.html`.

Ballan, Meltem, F. Ayhan Sakarya, and Brian L. Evans. 1997. A fingerprint classification technique using directional images. In *Proceedings of the IEEE Asilomar Conference on Signals, Systems, and Computers* (Nov. 3–5, 1997, Pacific Grove, CA), vol. 1, 101–104. `http://www.ece.utexas.edu/~bevans/ papers/1997/fingerprints/`.

Beeton, Mary. 2002. Scientific methodology and the friction ridge identification process. *Identification Canada* 25 (3) (September 2002). Reprinted in *Georgia Forensic News* 32 (3) (November 2002): 1, 6–8. `http://www. ridgesandfurrows.homestead.com/files/Scientific_Methodology.pdf`.

Collins, Martin W. "Marty". 1992. Realizing the full value of latent prints. *California Identification Digest* (March 1992): 4–11. `http://www.latent-prints. com/realizing_the_full_value_of_late.htm`.

Dankmeijer, J., J.M. Waltman, and A.G. de Wilde. 1980. Biological foundations for forensic identification based on fingerprints. *Acta Morphologica Nederlando-scandinavica* 18 (1) (March 1980): 67–83. `http://www.ncbi.nlm. nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=7395553&dopt= Citation`.

Galton, F. 1892. *Finger Prints*. London: Macmillan.

Haub, C. 1995. How many people have ever lived on Earth? *Population Today* 23 (2) (February): 4–5. Reprinted. 30 (8) (November/December 2002): 4–5. `http://www.prb.org/Content/ContentGroups/PTarticle/Oct-Dec02/How_Many_People_Have_Ever_Lived_on_Earth_.htm`.

Kingston, C.R. 1964. Probabilistic analysis of partial fingerprint patterns. D.Crim. dissertation, University of California, Berkeley.

Lindsay, Don. 2002. Don Lindsay archive. `http://www.don-lindsay-archive.org/creation/definitions.htm`.

Nilsson, Kenneth, and Josef Bigun. 2003. Localization of corresponding points in fingerprints by complex filtering. *Pattern Recognition Letters* 24 (13) (September 2003): 2135–2144. `www.hh.se/staff/josef/public/publications/nilsson03prl.pdf`.

Optel Ltd. 2003. New numerical methods of fingerprints' recognition based on mathematical description of arrangement of dermatoglyphics and creation of minutiea [sic]. `http://www.optel.pl/software/english/method.htm`.

Osterburg, James W., T. Parthasarathy, T.E.S. Raghavan and Stanley L. Sclove. 1977. Development of a mathematical formula for the calculation of fingerprint probabilities based on individual characteristics. *Journal of the American Statistical Association* 72: 772–778.

Pankanti, Sharath, Salil Prabhakar, and Anil K. Jain. 2001. On the individuality of fingerprints. `http://citeseer.ist.psu.edu/472091.html`.

Roddy, A.R., and J.D. Stosz. 1997. Fingerprint features—Statistical analysis and system performances estimates. *Proceedings of the IEEE* 85 (9): 1390–1421. `http://citeseer.ist.psu.edu/roddy99fingerprint.html`.

Sclove, Stanley L. 1980. The occurrence of fingerprint characteristics as a two-dimensional process. *Communications in Statistics* (A) 9: 675–695. `http://tigger.uic.edu/~slsclove/papers.htm`.

Stoney, David A., and John I. Thornton. 1987. A systematic study of epidermal ridge minutiae. *Journal of Forensic Sciences*.

Thai, Raymond. 2003. Fingerprint image enhancement and minutiae extraction. Honours thesis, School of Computer Science and Software Engineering, University of Western Australia. `http://www.csse.uwa.edu.au/~pk/studentprojects/raymondthai/`.

Vacca, J. 2002. Biometric security solutions. Excerpt from *Identity Theft*. 2002. Indianapolis, IN: Prentice Hall PTR. `http://www.informit.com/articles/article.asp?p=29801&redir=1`.

Thompson, William C., Simon Ford, Travis Doom, Michael Raymer, and Dan Krane. n.d. Evaluating forensic DNA evidence: Essential elements of a competent defense review. `www.cs.wright.edu/itri/EVENTS/SUMMER-INST-2003/SIAC03-Krane2.PDF`.