Abstract

1. Introduction

2. History

3. Physiology

4. Automated Matching Principles

5. Uniqueness of Fingerprint Feature Configurations

6. Reliability - FRR Analysis

7. Performance

8. Dependence

9. Conclusion

Appendices

A.1 Glossary

A.2 Vital statistics of fingerprint features

A.3 Tables for Osterburg probabilities

A.4 Table of intra-ridge pore configurations

A.5 Tables and plots of ridge independent pore configurations

A.6 Tables and plots of pore sizes

A.7 Tables and images of reliability of pores

A.8 References

A.9 Suggested reading

ABSTRACT

As the need for personal authentication increases, many people are turning to biometric authentication as an alternative to traditional security devices. Concurrently, users and vendors of biometric authentication systems are searching for methods to establish system performance. This paper presents a model that defines the parameters necessary to estimate the performance of fingerprint authentication systems without going through the rigors of intensive system testing inherent in establishing error rates. The model presented here was developed to predict the performance of the pore-based automated fingerprint matching routine developed internally in the research and development division at NSA. This paper also discusses the statistics of fingerprint pores and the efficacy of using pores, in addition to the traditionally-used minutiae, to improve system performance. In addition, this paper links together the realms of automated matching and statistical evaluations of fingerprint features. The result of this link provides knowledge of practical performance limits of any automated matching routine which utilizes pores or minutia features.

1. Introduction

This paper provides a statistical analysis of fingerprint features, as well as a description and comparison between automated and manual (forensic) fingerprint matching techniques, which includes the identification of certain critical parameters involved in any automated matching technique. The authors use these critical parameters to determine the expected performance of any automated fingerprint matching routine, and apply their model to a specific system.

This paper establishes the performance estimates for a pore-based automated fingerprint matching routine that is under development in the research and development division at NSA. Many factors influence the performance of such a system; some have been explored previously in either studies of automated matching systems or in the forensic arena, but there has been little effort to tie together information from both areas to estimate performance. In order to assess performance, feature uniqueness and reliability and automated matching parameters must be understood. Literature from law enforcement and criminology tends to focus solely on the uniqueness of a configuration of fingerprint features, and literature concerning automated systems deals with processing and matching techniques without regard to feature uniqueness or variation of matching parameters. Neither source has addressed feature reliability, although knowledge of all these issues is intrinsic to the development of a sound fingerprint matching technique.

Currently, the performance of biometric systems is gauged mostly by error rates. Errors in a fingerprint recognition system can be one of two types. A false accept occurs when an unauthorized user is identified as an authorized user and is therefore accepted by the system. A false reject occurs when an authorized user is not recognized as such, and is rejected by the system. In order to describe the performance of a system, both the FAR (false accept rate) and FRR (false reject rate) must be determined. These FARs and FRRs are accepted currently as the metrics by which biometric system performance is judged today.

Although error rates serve as a good indicator of system performance, the most common method of determining FARs and FRRs requires extensive testing, which is very time consuming. If the

system parameters change, then the testing must be redone. One motivation of this paper is to provide an alternative, more immediate, method for projecting system performance. In this paper, a model for fingerprint matching is developed that encompasses both uniqueness (from forensic analyses) and parameters needed for automated matching techniques. This probabilistic model is generated using knowledge acquired from several sources: existing models for uniqueness of sets of fingerprint features developed in forensic studies, a physiological model based on morphogenesis of fingerprints, a mathematically derived matching scheme, and results from measurements of real fingerprint images. The resulting model is intended to simulate a real matching system and provide the ability to estimate error rates for a given set of system parameters.

Inherent in developing this theoretical model, and especially in gauging the performance of the internally developed system at NSA, is exploring the efficacy of using pores to match fingerprints. Automated matching techniques traditionally have used configurations consisting of medium resolution features, such as branch points (bifurcations) and end points of epidermal ridges, but it is also possible to implement pores, which are high resolution features. The NSA internal system is one of only a few systems to date which utilizes pores as features to match fingerprints.

2. History

Branch and end points of epidermal ridges were used by Sir Francis Galton in 1872 to develop a probabilistic model of fingerprint individuality, and they have been used since then in both forensic (Cummins and Midlo, 1943) and automated matching (Blue, Candela, Gruther, Chellapa, and Wilson, 1994; Hrechak and McHugh, 1990). These Galton features, or minutiae, contain unique information that enables their use in probabilistic analyses. Each Galton feature has a specific type, i.e., branch point or end point, a unique location on the fingerprint, and a specific orientation (Stoney and Thornton, 1986). The orientation can be defined for an end point, for example, as the approximate tangent angle to the ridge ending.

Most probabilistic models to date have utilized Galton features exclusively; two of these models will be presented in this paper. The first model, published in 1977 by James Osterburg, et al at the University of Illinois, determines the probability of occurrence of a certain configuration of Galton features in a fingerprint. Two years later, a member of Osterburg's team, Stanley Sclove, published a paper presenting the occurrence of Galton features as a two-dimensional Markov model. Both of these models can be adapted to use pores instead of Galton features.

Pores have been used historically to assist in forensic matching. Although most matching methods have emphasized minutia comparisons and used pores as ancillary comparison features, the ability to match prints based on pore information alone has been documented (Ashbaugh, 1983; Locard, 1912; Stosz and Alyea, 1994). The concept of using pores to match prints has been essentially dormant during the rise of automated fingerprint recognition systems.

3. Physiology

The uniqueness of a configuration of pores depends on several factors, such as the number of pores involved, their respective shapes and sizes, the locations of these pores with respect to each other, and so on. These factors are all a function of morphology; thus, it would be helpful to

discuss briefly the genesis and formation of fingerprints, as well as the implications imposed in the development of pores.

Pores are formed where sweat glands in the subcutaneous layer of the skin generate sweat ducts; these sweat ducts grow through the subcutaneous layer and dermis to the epidermis, where the open duct on the skin's surface presents itself as a pore (Webster, 1992). (See Figure 1) According to a 1973 study on skin ridge formation (Hirsch and Schweichel, 1973), sweat glands begin to form in the fifth month of gestation. The sweat gland ducts reach the surface of the epidermis in the sixth month, forming pores.
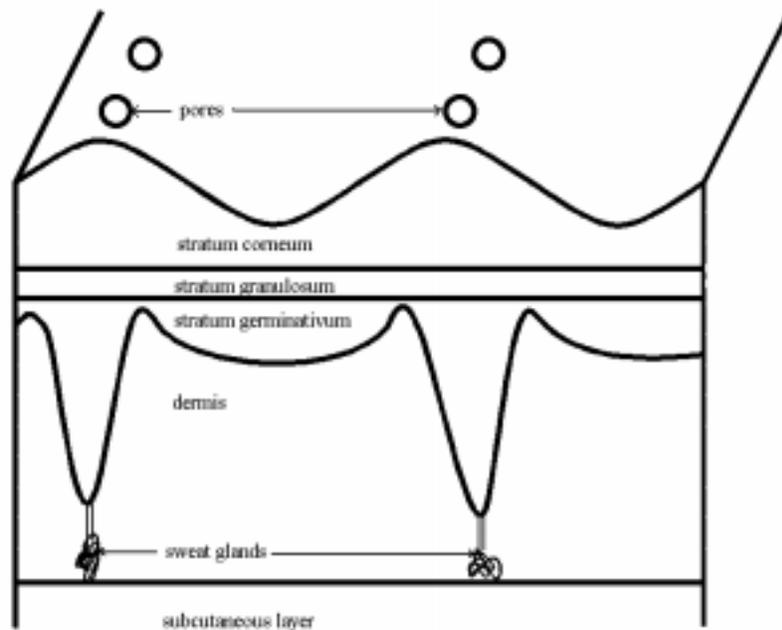

Figure 1. Physiology of the skin.

The epidermal ridges are not formed until after the sixth month; then, the pattern which has been forming in the glandular fold region is transferred to the epidermis. Hirsch and Schweichel concluded that several forces affect the epidermal pattern formation; one of these forces is the stabilization that occurs "when sweat gland secretion ducts open on to the surface, at regular intervals, in the papillary ridges." These openings of the ducts on the surface are the pores, and the regularity of their appearance plays a significant part in the uniqueness of pore configurations. Once these pores form on the ridge, they are fixed at that location. Considerable research has shown that pores do not disappear, move or spontaneously generate over time (Locard, 1917).

3.1 Model of Pore Spatial Distribution
In order to study the spatial distribution of pores, it is helpful to imagine a fingerprint based on an underlying lattice structure of sweat glands separated by distance *d*. Furthermore, pore distribution is assumed to be purely homogeneous and isotropic as a result of sweat glands being thus distributed (See Figures 2.a and 2.c). This first order approximation to actual pore formation is based on the assumption that the primary function of pores is heat transfer; therefore, pores should be evenly spaced. Matching a print with this kind of distribution would be trivial; if any

pore in the first print matches any pore in the second print, then all pores match (neglecting rotation effects).
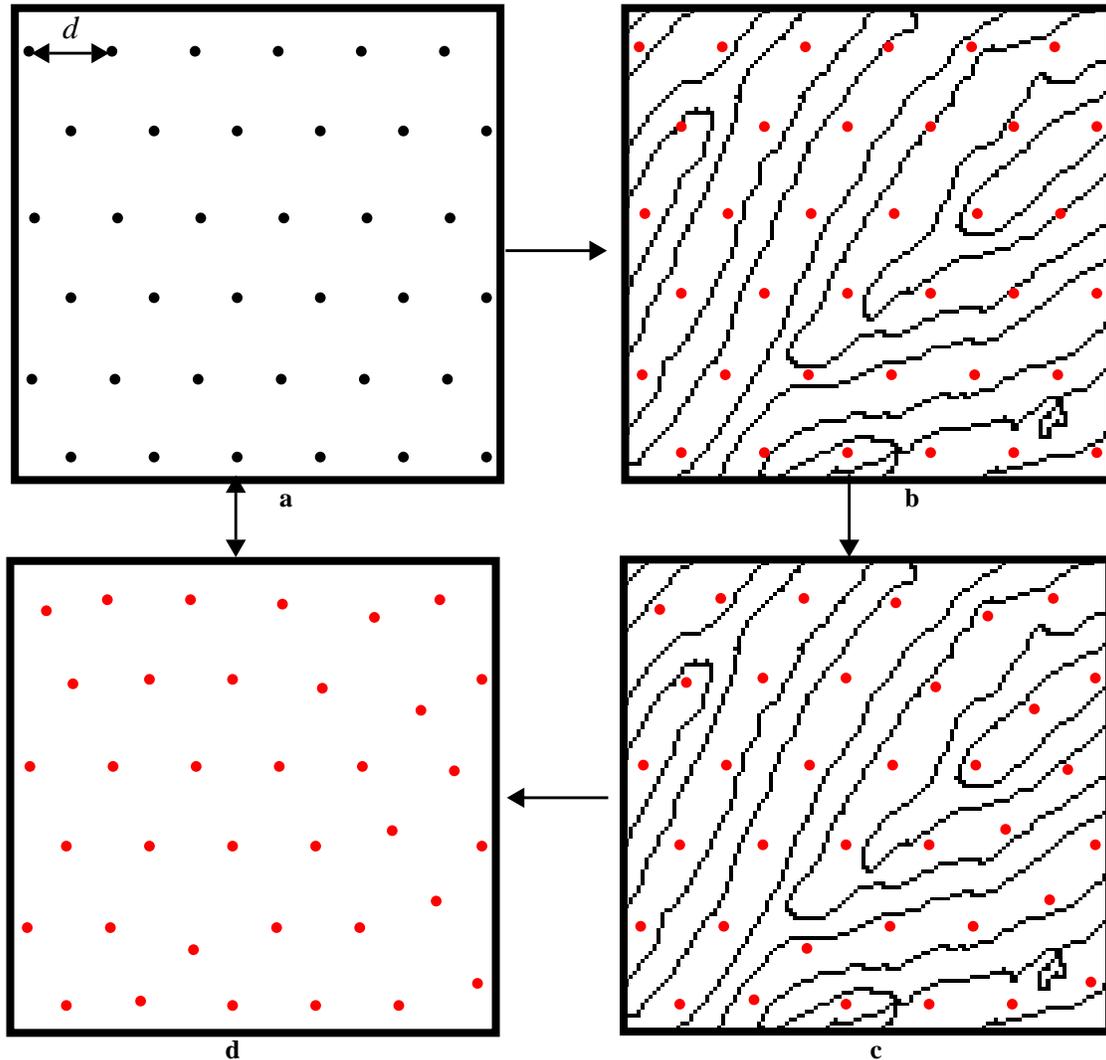

Figure 2. Lattice model for sweat gland/pore placement on the finger. In **a**, pores are represented in a lattice formation. In **b**, the lattice of pores is overlaid on a real fingerprint structure, in **c**, some pores must be moved to comply with the ridge structure; **d** is the resulting lattice with small random deviations added to some of the members. In a lattice formation, the positions are entirely deterministic. In **c** and **d**, the pores are distributed stochastically over the surface of the finger.

The lattice model must be adjusted since the formation of epidermal ridges constrains pores to appear on them. The lattice is modulated by the underlying ridge structure (See Figures 2.b and 2.d); however, a certain degree of regularity remains. Assume that each pore is located in a "sweat gland unit," where the units occur side by side on a ridge. Furthermore, assume that the position of the pore inside the sweat gland unit is a random variable which is uniformly distributed over the sweat gland unit dimensions.

A comparison of real and modeled pore spatial distribution is explored in Figure 3. Here, a fingerprint is shown in **a**, with pore location information extracted and displayed in **b**. A

rectangular lattice model is shown in **c**, and **d** is a modification of **c** which allows a small degree of random variation in the lattice point positions. The next image, **e**, depicts a lattice with an even greater degree of variation in the pore positions. Note that in **c**, **d**, and **e**, no ridge modulation has been performed and it is assumed that the underlying ridge structure is purely unidirectional (either vertical or horizontal rows of ridges). Unlike **d** and **e**, for **f**, the points were not constrained by a lattice structure in their positioning. The position of each point was considered to be a random variable with a uniform distribution (over the dimensions of the image). Therefore, an element is equally likely to occur at any location within the image. The real spatial pore distribution, in **b**, is more similar to the point distribution in **e** than **f**; thus, a model based on a uniform probability distribution for the pore positions is not accurate. In fact, the magnitude of small random perturbations added to each lattice location should be quite small based on inspection of the images in Figure 3.

Although regular spacing of sweat glands within a ridge is assumed to be the norm, one must allow for the possibility of an absent sweat gland. Along with the randomness of pore positions within the existing sweat gland units, physiological omissions of sweat glands contribute to the deviation from the expected distance between nearest neighbor pores (Okajima, 1975). These omissions are evident in Figures 3.a and 3.b.
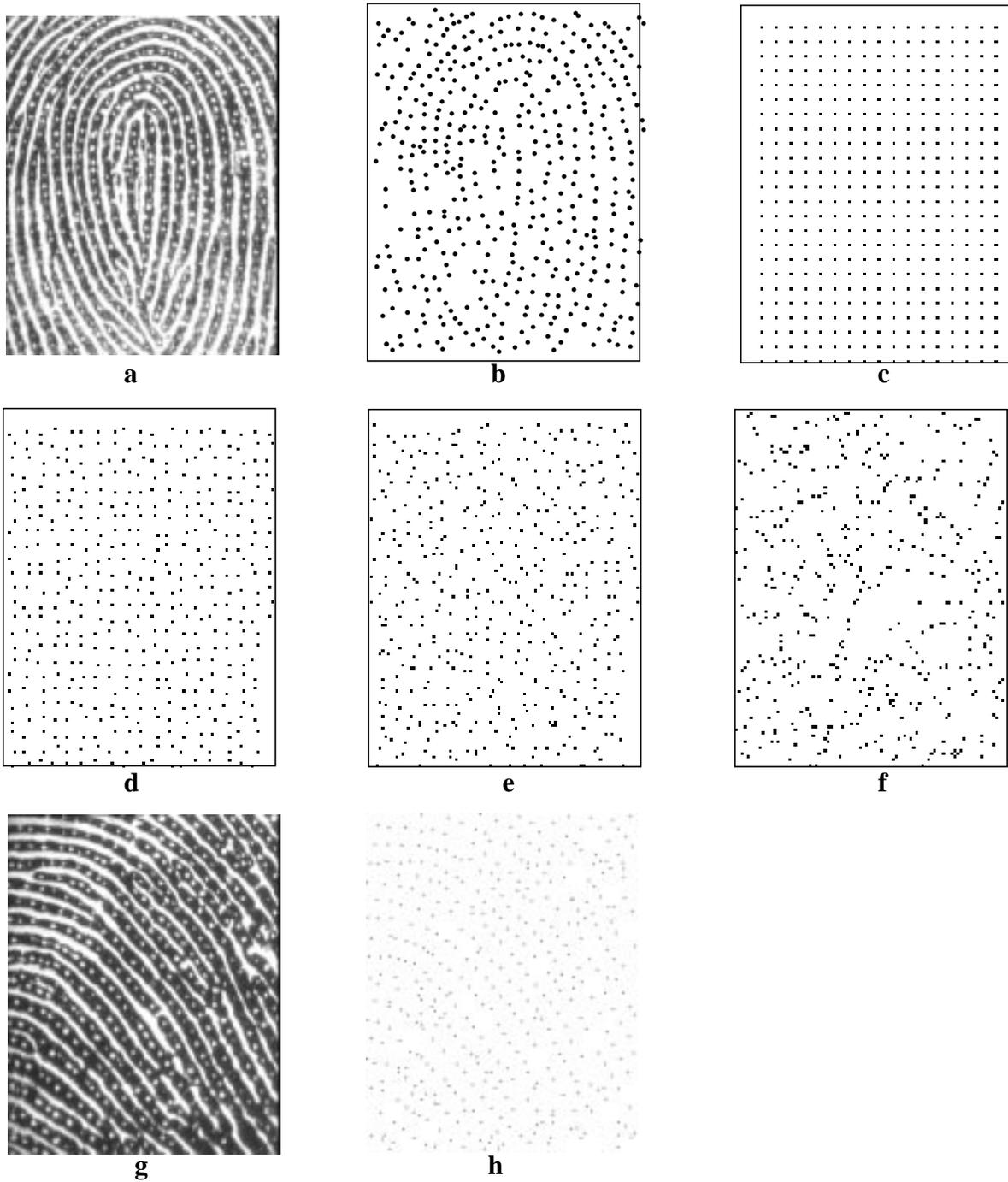
Figure 3. a. Original image. **b.** The position of more than 400 pores within a fingerprint image. **c.** Lattice of points. **d.** Lattice of points with random variation in position (within unit cell). **e.** Lattice of points with larger allowed position variation. **f.** Plot of a set of points with positions generated from uniformly distributed random variables (uniformly distributed set of points). **g** and **h** - fingerprint segment with more uniformly spaced pores than **a**.

Before progressing further into a discussion of system performance and model derivation, it is

important to consider how a matching routine works, especially since the matching routine is a critical component of any recognition system.

4. Automated Matching Principles

Consider the problem of comparing two different fingerprint sub-segments selected from complete prints. The first segment is obtained from a known user at the time of enrollment, or registration, into the system. The second segment results from a live-scan acquired for the purpose of verifying a user's identity. When the two segments are compared, there are two possible outcomes: The live-scan segment either matches or does not match the registered segment.
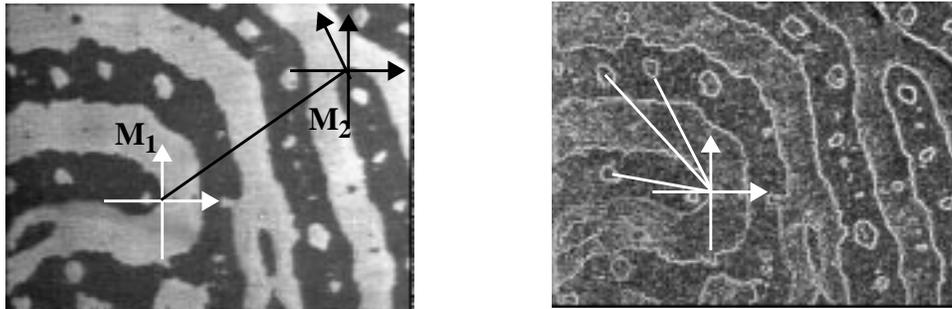

Figure 4. Fingerprint features. In **a**, features of minutiae points are demonstrated. In this case, both M_1 and M_2 are end points with orientation defined as the direction of ridge flow at the end point. Knowing the relative positions of M_1 and M_2 is sufficient to determine the degree of rotation between sets of images. The features of pores, their relative position (defined at the center of mass of the pore), size, and shape are seen in both **a** and **b**. Using only position information, a set of pores is unique.

An enrollment procedure is used to extract pertinent information from the fingerprint and store the information to a template (or feature vector) which then represents the user. When matching is based on minutiae or pores, the template consists of vital information about these features. In the case of minutia points, knowing the relative position, orientation, and type (branch or end) of each minutia in the set is sufficient:

minutia features: {(P_1,Θ_1,T_1), (P_2,Θ_2,T_2), ... (P_{ME},Θ_{ME},T_{ME})},

where there are M_E minutia enrolled, P is the position, Θ is the orientation, and T is the type. For pores, the position relative to a local reference point, size, and shape could be stored:

pore features: {(p_1,m_1,s_1), (p_2,m_2,s_2), ... (p_{PE},m_{PE},s_{PE})},

where P_E is the number of enrolled pores, p is the position (defined as the center of mass of the pore), m is the size, and s is the shape. See Figure 4, which is a high resolution fingerprint image, for examples of fingerprint features.

Whether pores or minutia are the basic features, it is possible to use a subset of the full feature set, (for instance, the location only) to represent the fingerprint. Authentication is then reduced to a

comparison of two sets of points in space and deciding if they match well enough:

enrolled features: {$f_{0,e}$, $f_{1,e}$,..., $f_{NC,e}$},
comparison features: {$f_{0,c}$, $f_{1,c}$,..., $f_{NE,c}$},

where N_E is the number of enrolled features (pores or minutiae) and N_C is the number of live-scan or comparison features (pores or minutiae). Figure 5 shows some basic matching concepts.
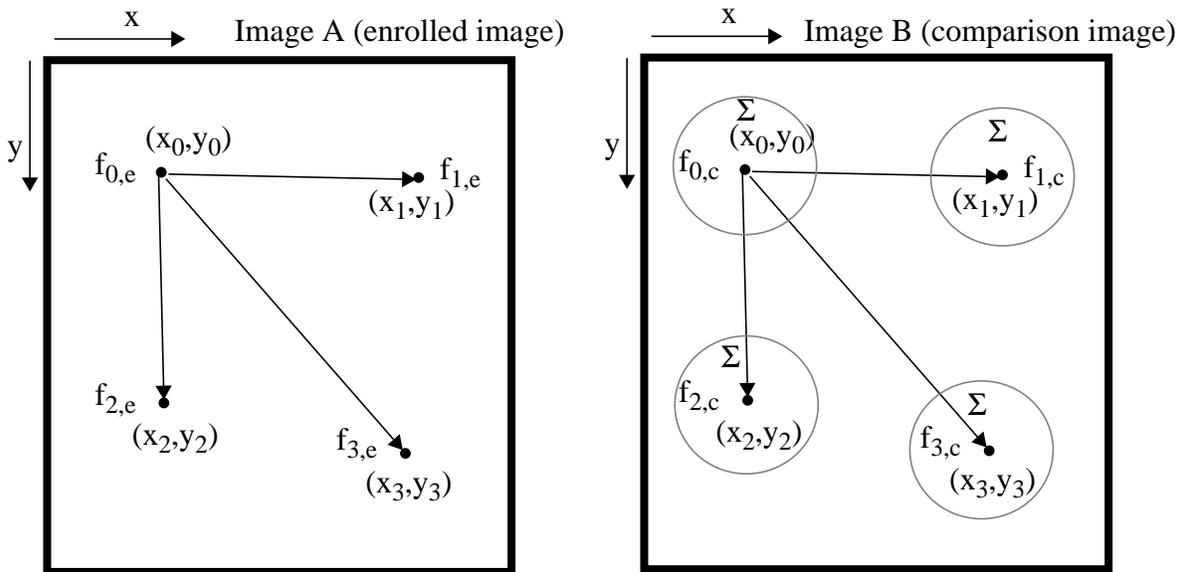

Figure 5. Basic matching principles. The concept of a search area represented by Σ, is shown. Also, measurement is based on relative positions as opposed to relying on an absolute coordinate system. Feature $f_{0,e}$ which corresponds to $f_{0,c}$ is a feature (local origin) which is used to establish relative positions of other features.

An important step in matching is determining a common reference point or origin in each of the print segments. For example, a particular minutia point may be used as a local origin from which to measure the position of a set of nearby pores. This minutia point must be properly identified in both images before the pores can be matched.

4.1 Pore Extraction Technique

The method used to extract the pores as fingerprint features is critical to the matching routine. The pore's position, size and shape are features which make it distinct from other objects in an image. Techniques used for the fingerprint data capture can be used to enhance the pore information. For example, high resolution scanning and manipulation of the gain and contrast camera controls can highlight the pores. The position of the pore is determined by processing the gray scale fingerprint image and transforming it to a skeleton representation. By applying models and processing routines to the skeleton to the skeleton of the fingerprint image, the pore locations can be extracted. Pores are transformed into isolated and connected short lines in the skeleton image. Given this information, the size of the pore can be determined by region growing routines operating on a binary version of the fingerprint image. More details can be found in references provided in the appendix (Stosz and Alyea, 1994; Stosz and Alyea, 1995).

4.2 Scanning Resolution

Some parameters become critical to the matching routine. For instance, the resolution at which the fingerprints are scanned determines the accuracy of feature location measurements. Inherently, there may be only one pore in a given 1 mm x 1 mm section of print, and at 1000 ppi (pixels per inch), this section is represented by approximately 40x40 (1600) pixels. In comparison, at a scanning resolution of 500 ppi, the same segment is represented by 20x20 (400) pixels. Therefore, the probability of another 1 mm^2 segment of print matching with respect to pore position is either 1/1600 or 1/400 depending on the scanning resolution. It can be seen that the false accept error rate will be reduced at a higher scanning resolution at the cost of an increased false reject error rate.

4.3 Feature or Search Area

The scanning resolution issue can be made invariant by defining an absolute area to be associated with each feature (a feature area or search area). For instance, the location of a pore in the enrolled print segment may be determined to be (x,y), but in the corresponding live-scan segment, its location may be shifted some distance Δ as a result of rotation, plasticity, or other distortion. For the purpose of matching these two segments, a search area Σ, ([x-Δx, x+Δx],[y-Δy, y+Δy]) as seen in Figure 5, can be defined such that if the feature is within Σ, the features match with respect to position.

The size of Σ is a parameter that influences the performance of the system (decreasing Σ produces a decreasing FAR and increasing FRR; increasing Σ produces increasing FAR, decreasing FRR). Practically, Σ should be large enough to account for effects such as plasticity of the finger and deviations in feature position due to variations in the data and effects of the processing algorithms, but not big enough for areas associated with distinct features to overlap. In a forensic comparison, plasticity and distortion of the finger are accounted for by human processing, but in an automated process, tolerances such as Σ must be incorporated to accommodate these inherent variations.

4.4 Finger Plasticity

The distance between two features can change significantly due to plasticity of the finger. This relative change of position is generally not significant for nearby features within small areas of print. Therefore, when measuring the position of small high density features such as pores, a local origin should be established. A minutia point can be used to establish a local origin.

4.5 Reliability

A critical factor when considering the performance of a fingerprint matching system is reliability. Within the scope of this paper, overall reliability is broken down into two components: inherent reliability and algorithm (processing) reliability. Inherent reliability refers to the physiological dependability of pores, which is the probability that a known pore will be visible in a particular live-scan print. Pores do not always appear on print images; factors such as temperature and skin condition can conspire to alter or suppress altogether the physical appearance of a given pore.

Algorithm, or processing, reliability must also be taken into account. Depending on the quality of the image, automated processing and detection algorithms make errors. There are two errors that the feature detection algorithm can make: a missed detect and an incorrect (false) detect. A

missed detect occurs when a feature (pore or minutia) is discernible in an image, yet is not picked up by the detection algorithm. A false detect occurs when the algorithm mistakenly marks a feature, when in fact, no feature is present. The degree of noise and degradation in the image influences the quantity of errors. The probability of incorrect detection, p_{fd}, and missed detection, p_{md}, are parameters on which the performance of the system depends. A high p_{fd} or p_{md} will tend to increase the FRR but have little effect on the FAR.

4.6 Match Score

A particular matching technique will produce a score representing the fraction of features matching between the enrolled and the live-scan prints. Figure 6 provides an example of matching based on either minutiae or pores for two segments from different fingers and also two segments from the same finger. Generally, the number of features detected in the two different prints, N_E and N_C, will be different. Therefore, the matching routine must compare two sets, or configurations, with a different number of elements. For example, a pore match score, S_P, can be defined as:

$$S_P = \frac{2n_m - n_n}{N_T}$$

Eq. 1

where

$N_T = (N_E + N_C) = $ *total number of pores in both segments*
$n_m = $ *number of pores that match*
$n_n = $ *number of pores that do not match*

and using

$$n_n = N_T - 2n_m,$$

The pore matching score, S_P, can be rewritten as:

$$S_P = \frac{4n_m - N_T}{N_T}$$

Eq. 2

A match occurs when a pore is detected in the comparison image at an enrolled pore's location. A mismatch occurs when a detected pore from either image does not correspond to one from the other image. Based on S_P, a decision is made to accept or reject the claimed identity of the user.

This score will be in the range [-1,+1], where +1 represents a perfect alignment of the pores in two different image segments. The local origin is defined to be a minutia point, and the relative rotation of the two image segments can be measured by determining the angles of corresponding minutia points.

20 pores
5 matches
39 mismatches

a

29 pores

b

34 pores
28 matches
7 mismatches

c

d

e
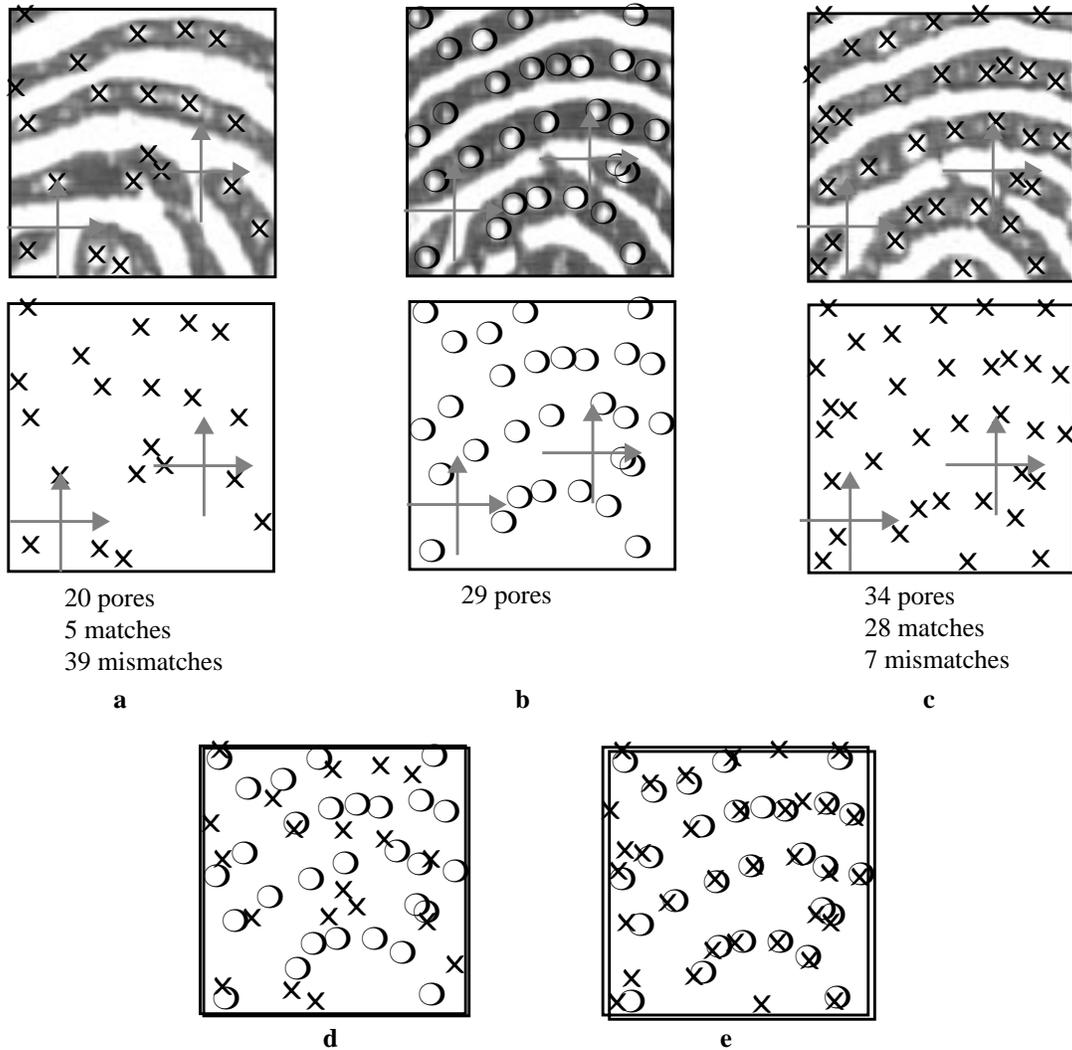
Figure 6. Pore based matching example. **b** and **c** are from the same finger and **a** is from a different finger. There are two very similar end point minutiae in both print segments. In fact, it is likely that the different print segments would match based on minutia comparison alone. If the minutiae are used to align the prints, the pore information matches for the center and right images but does not match for the center and left images.

5. Uniqueness of Fingerprint Feature Configurations

The goal of this paper is to establish the practical performance limits of fingerprint matching systems. Section 5 is devoted to determining uniqueness estimates which are related to the false accept error rate (FAR). The reliability of features, which provides false reject error rate limitations, will be presented later. In section 5.1, work done to establish the uniqueness of configurations of minutiae will be reviewed. Similar techniques using pores instead of minutiae will be presented in section 5.2 and 5.3. In addition, throughout section 5, practical matching algorithm issues will be introduced and their influence on uniqueness estimates addressed.

The uniqueness of a set of minutiae or pores is defined as the probability of occurrence of the set.

Therefore, the probability of an imposter match is directly related to the uniqueness. The more unique a set of features, the less likely an imposter is to match the set. Obviously, increasing the number of features used to represent the print will increase the uniqueness of the feature set, but the frequency of occurrence of particular arrangement of features will vary also.

A key issue in evaluating the uniqueness of a set of features is whether they are independent. Another issue is the feature distribution, for features may be regularly spaced, or the feature density (features per mm^2) may be non-uniform over the fingerprint. Addressing these issues is essential in determining the theoretical false accept rate of the system. For example, a particular combination of minutiae (such as the two end points of a ridge island) may occur frequently. In this case, an imposter is more likely to match both minutia points because of their close association, or dependence, than he would be to match two independent features. Therefore, features that exhibit dependence are less valuable than independent features.

In section 5.1, a review of research for a minutiae configuration study that assumes independent minutiae will be given. A method for studying minutia dependencies is summarized in Section 8.1. For pores, the dependence issue will be handled with a distinction between intra-ridge and inter-ridge pores.

5.1 Minutia Configurations
In this section, the methods used by Osterburg, <u>et al.</u> (Osterburg, Parthasarathy, Raghavan, and Sclove, 1977), to determine the uniqueness of a set or configuration of minutiae will be reviewed. The *P(configuration of features)* defines the uniqueness of a configuration of minutiae, in a given area of print, which is equivalent to the probability of two different print segments matching. Osterburg's results can be used to estimate the FAR of a minutia based matching technique, although his model does not include some very important parameters inherent to automated matching systems, such as feature reliability, detection errors, and search area variability. Therefore, the Osterburg model is just a starting point for determining the theoretical performance (which includes both FAR and FRR estimates) of practical fingerprint matching systems.

By examining 1 mm x 1 mm segments, or cells, of fingerprint (see Figure 7 for a perspective on scales used in fingerprint processing), Osterburg determined the frequency of occurrence of 13 possible outcomes based on Galton fingerprint features. The results are provided in Table 1 in Appendix A.3. The set of Galton features includes ridge endings, bifurcations, islands, dots, bridges, spurs, enclosures, double bifurcations, deltas, and trifurcations. Osterburg included three

other outcomes: a broken ridge, an empty cell, or some other multiple combination of features.
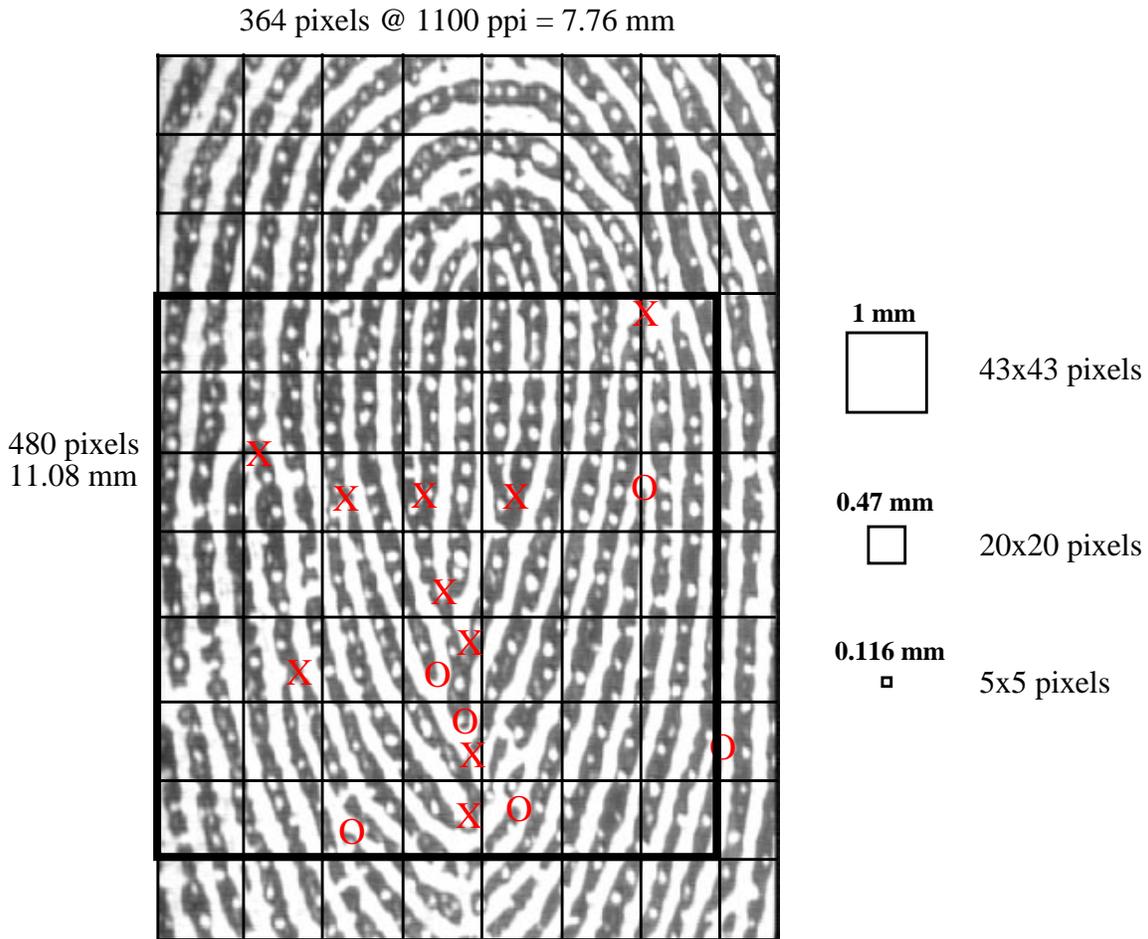

Figure 7. Scale and relative sizes used for fingerprint analysis. Inside the broad outline, the minutia end and branch points are marked with Os and Xs respectively.

The underlying assumption made by Osterburg is that the content of each cell is a random variable which is independent of all other cells. The implication is that any configuration of the same set of features has the same probability of occurrence meaning, for instance, that a tightly clustered pack of minutia is just as likely as the same set of minutiae being distributed uniformly over the print. Although the Osterburg study gives meaningful results, empirically the independence assumption is not valid because some configurations of Galton features are much less likely than others.

Based on the independence assumption, the individual feature probabilities are combined to yield the probability of a feature configuration:

$$P(\textit{configuration of Galton features}) = p_0{}^{k0}p_1{}^{k1}... \ p_{12}{}^{k12} \hspace{2cm} \text{Eq. 3}$$

where p_i (for $i = 0,...,12$) is the probability that a given type of Galton feature, i, will occur in a

cell, and k^i is the number of cells in which the feature occurs. The k^is sum to N, the number of cells.

As an example, the 7 mm by 7 mm block of data inside the broad box shown in Figure 7 has 4 cells containing one end point only, 8 cells containing one branch point only, 2 cells containing both a branch and an end point, and 35 empty cells (where only branch and end points have been considered for simplicity). Therefore, the probability of this configuration of 10 bifurcations and 6 end points (16 minutiae) in 49 mm^2 of print is:

$$P(configuration) = (0.0832)^4(0.0382)^8(0.0355)^2(0.8431)^{35} = 6.97 \times 10^{-22}.$$
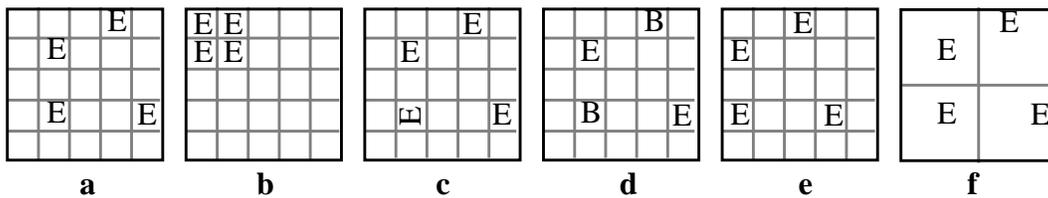
Figure 8. Osterburg minutia model and matching issues. In the Osterburg model, any configuration of the same set of features has the same probability of occurrence. The probability of configuration **a** or **b**, 4 end points and 12 blank grids in an area of 25 mm^2, is 1.78 x 10^{-7}. In a configuration, the relative position of the features is known as well as the type and orientation of the features. Note that the orientation of the minutia points is not the same in configurations **a** and **c** and although the configurations have the same probability of occurrence, they do not match. A common matching technique is to discard the type information; therefore, configuration **a** matches **d** but the two have different probabilities of occurrence. Depending on the defined origin or reference point of the print, configurations **a** and **e** may match. Finally, the resolution at which feature locations are measured (or the degree of allowed deviation in detected feature position) means that configurations **a** and **f** have a different probability of occurrence even though the minutiae are exactly the same (configuration uniqueness is a function of resolution).

Figure 8 demonstrates the need to expand the Osterburg model in order to determine practical values for the probability of matching. In Figure 8, **a** and **b** are two print segments that have the same set of features, but different configurations. These two prints would not match but would have the same probability of occurrence under the auspices of Osterburg model:

$$P(1st\ configuration) = P(2nd\ configuration) =$$
$$P(end\ point)^{\#\ cells\ with\ an\ end\ point}P(empty\ cell)^{\#\ empty\ cells} =$$
$$(0.0832)^4(0.766)^{21} = 1.78 \times 10^{-7}. \hspace{2cm} \text{Eq. 4}$$

For this example there are:

$$Combination(25\ 4) = 12650$$

different configurations possible for a set of 4 end points in 25 cells, each having the same probability of occurrence.

A comparison of a and c (in Figure 8) shows that although the feature types and locations match, the orientation of each feature is also crucial to the uniqueness of a configuration. The orientation is determined by the ridge flow direction, and the number of different orientations, N_O, is arbitrary (8 for instance). Whereas the probability of both configurations is the same under the Osterburg model, the different orientations actually distinguish the two configurations, thus:

$$P(\text{configuration with orientation}) =$$
$$P(\text{configuration without regard to orientation}) \, (1/N_O)^{NP} \qquad \text{Eq. 5}$$

where N_P is the number of points in the configuration for which the orientation is determined.

In figure 8, **d** demonstrates a situation where the location of all the features match **a**, yet the feature types do not match. These two configurations have different probabilities and do not match under the Osterburg model but will match for a routine which keys on geometry but is invariant with respect to type. The effect of translation is noted in **a** and **e**. Even though these configurations are different by definition, they may match depending on the choice of the reference point used to measure the minutia locations. Finally, the effect of the measurement accuracy (or resolution) can be seen comparing **a** with **f**. Even though the relative minutia positions and orientations are exactly the same, the cell size (feature area) is different resulting in a different probability of occurrence for the two configurations.

5.2 Pore Configurations

5.2.1 Pore Distribution
In order to describe the spatial regularity of pores, a deterministic model for pore distribution was proposed in section 4.1 in which neighboring pores are separated by a constant distance *d* and are arranged in a lattice formation (see Figure 5). This *d* represents the average distance between neighboring pores, where it is assumed that the pores are located in the center of a region containing only one sweat gland. The value of *d* is thus calculated from a live-scan image as:

$$d = (\text{area of ridges/number of pores})^{1/2} \qquad \text{Eq. 6}$$

For this extreme case, matching two images consists of simply lining up any pore on both images, since the rest of the pores would align themselves accordingly. The remaining pores in the image contribute no additional information; that is,

$$P(\text{all pores match})=P(\text{one pore matches}).$$

The model is made stochastic by assuming that each pore position can deviate by a small random amount. The methods used in the next several sections to model intra-ridge pores are closely tied to the lattice model. The intra-ridge models represent a single ridge, whereas the lattice model can be thought of as being composed of multiple independent ridges covering an area.

5.3 Intra-Ridge Pore Configurations

5.3.1 Ashbaugh Model

In this section, a model describing fingerprint pores presented by Roger Ashbaugh of the RCMP will be examined. Ashbaugh presented his work and reviews of prior work by Edmond Locard (circa 1912) in a series of articles on ridgeology, edgeology, and poroscopy (Ashbaugh, 1982).

According to Ashbaugh, the fingerprints begin forming on the fetus around the thirteenth week of development. Bumps form on the surface of the skin fusing together as they grow creating the ridges. The bumps, or pore pods, each contain one pore which originates from a sweat gland in the dermis. The pods are approximately equal in width and length (0.48 mm) resulting in a frequency of about 20.8 pores/cm of ridge. The pores are unique in shape, vary in size (from 88 to 220 μm), and change only in size due to growth of the skin.
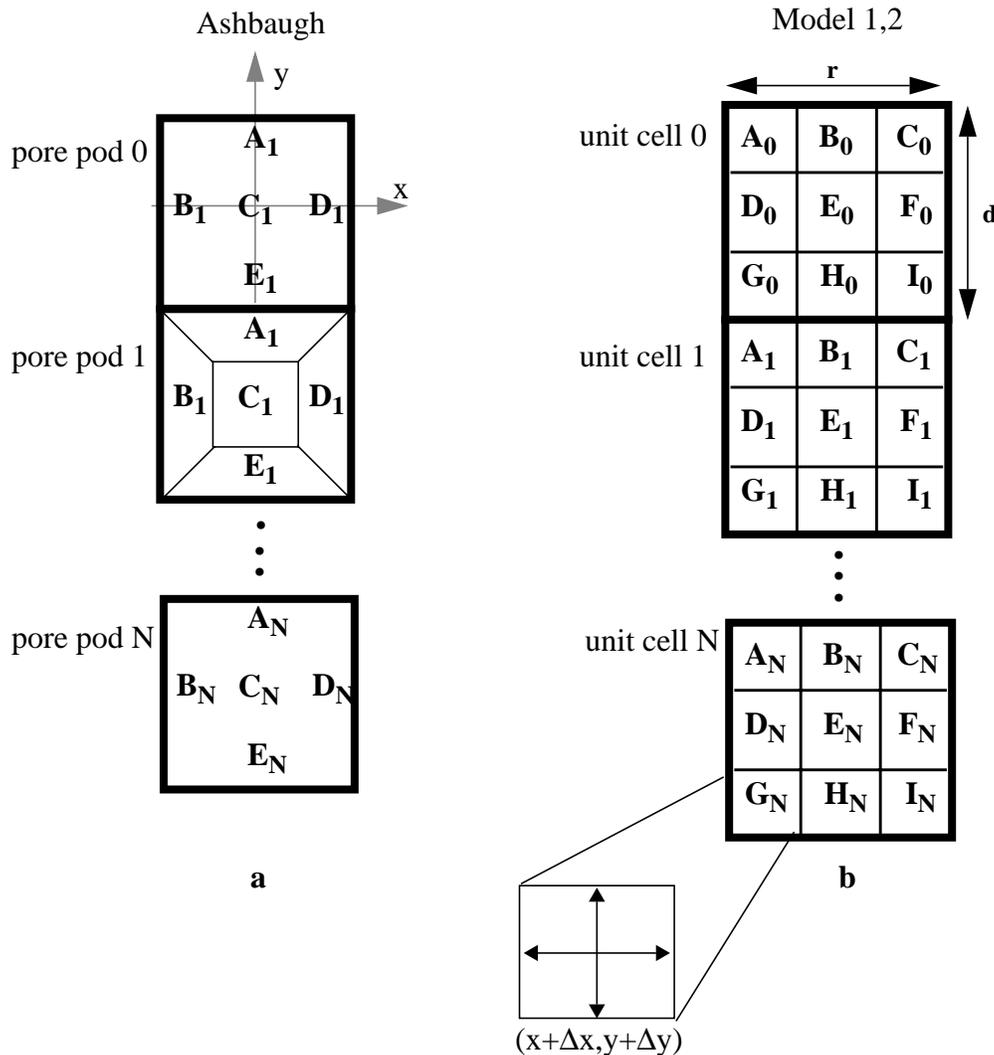
Figure 9. Models used to represent pore occurrence along a single ridge. These models can be used to predict the probability of occurrence of a sequence of N pores on the ridge, where **d** is the average distance between intra-ridge pores and **r** is the average ridge width.

Locard proposed independent poroscopy based identification and used this technique successfully to convict criminals in the early 1900s. In contrast, Ashbaugh claimed that pore and ridge

comparisons should be used in conjunction with appropriate weighting depending on the available data. Generally, the quality of the inked or latent print determines the usefulness of pore comparisons in law enforcement applications. According to Ashbaugh, the reliability of pores and their shape varies in inked and latent prints. It should be noted that the consistency of pore features acquired from live-scan devices will differ from that resulting from inked or latent developing techniques (see Table 5 in Appendix A.7). As a result of his concern over pore feature inconsistency, Ashbaugh promotes the comparison of pore locations only and states that shape and size should not be used in general.

Ashbaugh further contends that pore pods occur regularly, but the position of the pore within the pod is a random variable. In addition, he assumes independence between pore pods. In his model, each pod is divided into 5 general areas in which a pore may occur (as seen in Figure 9). The probability of a pore occurring in any of the 5 regions, A, B, C, D, or E of the pore pod, is:

$$P(pore\ in\ A) = P(pore\ in\ B) = P(pore\ in\ C) = P(pore\ in\ D) = P(pore\ in\ E) = P_P = 0.2$$

Under the assumption that the pods are independent,

$$P(a\ sequence\ of\ N\ intra\text{-}ridge\ pores) = (P_P)^N = (0.2)^N \qquad \text{Eq. 7}$$

given a sequence of N pores on the same ridge and assuming that each pore can occur in one of five equally likely states with probability, P_P.


a   **b**
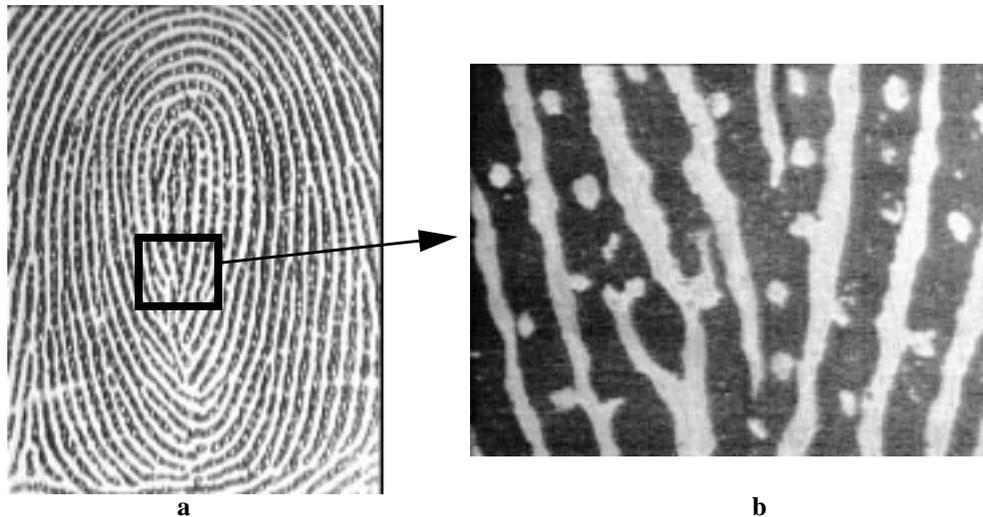
Figure 10. Sufficient information for identification. **a** was scanned at a resolution of 800 dpi. The small segment outlined with a dark box, and expanded in **b**, represents about 5 mm^2 of fingerprint area and contains more than 20 pores. According to Locard or Ashbaugh, the segment is sufficient to identify its owner based on the relative position of the pores alone (without even including shape or size descriptions). Note that sufficient information would exist for a minutia match by doubling the size of the rectangular region (about 12 minutia could be enclosed). Although any segment for which there are at least 20 pores may be used for identification, it is unlikely that sufficient minutia information would be present in such a small area.

Ashbaugh provides numerical examples for the uniqueness of a sequence of intra-ridge pores.

Based on his assumptions, the probability that two consecutive intra-ridge pores have the same relative position as two other pores is 0.04, and the probability of occurrence of a particular arrangement of 20 consecutive pores is:

$$P_{Ashbaugh}(a\ sequence\ of\ 20\ intra\text{-}ridge\ pores) = 0.2^{20} = 1.05\ x\ 10^{-14} \qquad \text{Eq. 8}$$

In addition, Ashbaugh supports the claim made by Locard that matching between 20 and 40 pores is sufficient to identify an individual. The implication of this statement is demonstrated in Figure 10 in which a small area of print is sufficient for identification. Since there are about 20 pores per cm of ridge, and a typical live-scan fingerprint usually contains over 50 cm of ridge, the amount of pore information required for identification is just a small fraction of the available data.

5.3.2 Distribution of Distances Between Sequential Intra-ridge Pores

In this section, a procedure for estimating the uniqueness of a sequence of pores based on measurements of real fingerprint data is summarized. To accomplish this task, pore locations along the ridges of live-scan prints were detected manually. Then, the distance between successive intra-ridge pores was calculated for individual prints. The plots of intra-ridge distance for individual fingerprints generally produced bimodal distributions with a dominant peak and an inferior peak resulting from missing or skipped pores. However, when all the data were combined, 3748 distance measurements resulted in a smoothed single mode distribution ($\mu_R = 0.377$ mm (16.955 pixels), $\sigma_R = 0.1820$ mm (8.1680 pixels)) with a significant upper tail, as seen in Figure 11.
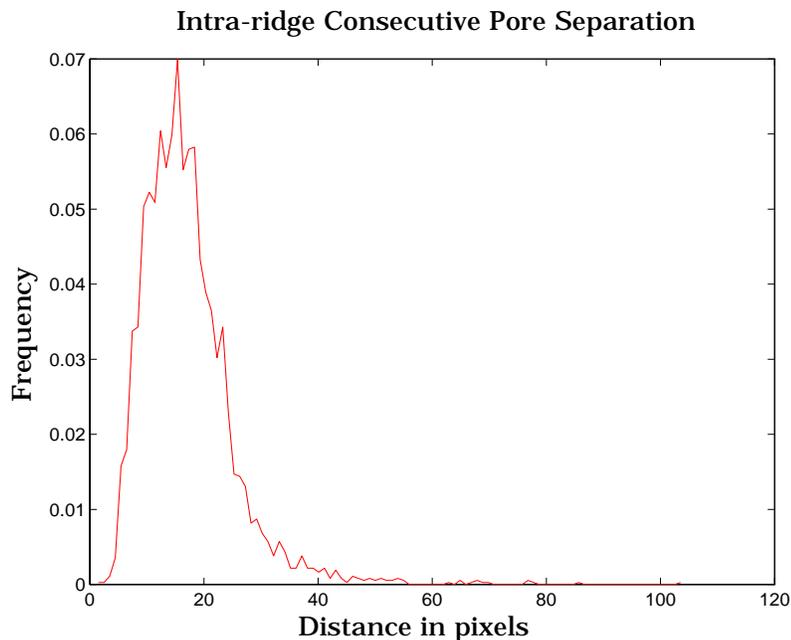
**Figure 11**. Plot of consecutive intra-ridge pore separation. The most frequent occurring separation is 13 pixels (0.3 mm).

The tail of the distribution begins at about 0.69 mm (30 pixels), and tapers off at 2.40 mm (104 pixels), the maximum observed distance between intra-ridge pores. The frequency value peaked at 0.30 mm (13 pixels), with a probability of occurrence of 0.0645 (at a measurement resolution of

1 pixel at 1100 ppi). This distance is defined as d_R, and its probability of occurrence is $P(d_R)$. In Figure 11, there are 104 bins represented, which is the maximum observed spacing between pores, meaning that each bin is the size of a pixel.

Given this distribution, the probability of occurrence of any sequence of intra-ridge pores can be calculated by assuming the pores are independent. For this case, the only parameter of interest is the distance between consecutive pores. In addition, a lower bound for the uniqueness can be calculated by assuming that all the pores in a sequence are spaced by d_R (the most likely separation). Then any sequence of the same number of pores is guaranteed to be at least as unique as this bound.

Table 2 in Appendix A.4 summarizes the uniqueness of sequences of intra-ridge pores with varying resolution. In this table, the results depend on the number of pores in the sequence and also on the measurement accuracy, or resolution, of a pore's position. As the measurement accuracy is decreased (fewer bins in the histogram), there are fewer distances possible between pores but the area under the distance probability density function remains constant, resulting in distances with higher probabilities. Therefore, d_R may not change but $P(d_R)$ will increase. The new values for $P(d_R)$ were determined by accumulating area of the normalized histogram around d_R symmetrically. From this table, with a resolution setting of r=3, the upper bound on the probability of occurrence is:

$$P_{Measured}(a\ sequence\ of\ 20\ intra\text{-}ridge\ pores) = 0.201^{20} = 1.16\ x\ 10^{-14} \qquad \text{Eq. 9}$$

for a sequence of 20 pores, which is in close agreement to Ashbaugh's results in Eq. 8.

**Probabilistic Value of Intra-Ridge Pore Configurations at Various Resolutions**
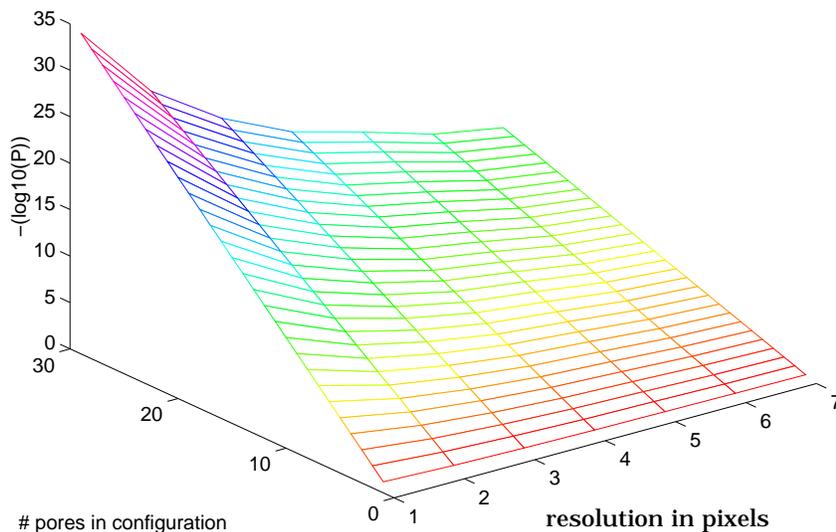
Figure 12. Plot of probability of a sequence of intra-ridge pores as a function of varying resolution. The y-axis shows the negative log probability. The highest resolution and greatest number of pores yields the lowest probability of occurrence.

A setting of 3 for the resolution parameter, r, is reasonable based on the fact that a typical pore has a diameter (assuming a circular shape) of about 5 pixels (115.5 µm) and the adjusted average spacing (compensating for skipped pores) between intra-ridge pores is 13 pixels (300.3 µm). Thus, setting r to 3 is equivalent to an allowed displacement of size 3 pixels (69.3 µm) in which to detect the pore. Therefore, the pore's position can vary slightly from its expected value, but too much variation will cause a mismatch. In addition, the "search area" is only big enough so that one pore is likely to be present and an adjacent pore is unlikely to overlap in this area. Figure 12 demonstrates how resolution and the number of pores affect the probability of occurrence.

5.3.3 Intra-ridge Pore Distribution from models

In Ashbaugh's treatment, pore pods are divided into five regions in which the occurrence of the pore is equally likely. In reality, the desired measurement accuracy determines the number of regions. In addition, dividing a pore pod into 5 distinct regions is not convenient for an automated system. Therefore, a similar but improved model for the pore placement will be discussed next which will be used to establish the distribution of intra-ridge pore spacing. The results that the model generates can be compared to the measured data to determine the accuracy of the model.

Assume that an individual pore occupies a unit area, or cell. Unit cells are bounded on the edge by the ridge border, and their length is defined as the average distance between intra-ridge pores. Unlike Ashbaugh's model, unit cells are not a physiological feature; they are simply defined as the average area of print associated with a single pore. The location of the pore within a unit cell is a random variable and therefore, a sequence of intra-ridge pores is represented by a random vector for which the elements are independent and identically distributed. In addition, there is a finite probability of a pore not occurring in a unit cell (an empty unit cell). This condition is used to explain the occurrence of relatively long segments of ridge which have no pores.

As shown in Figure 9.**b**, the unit cells are divided into 9 sub-regions, $A_i,...,I_i$, for a given unit cell i. For this analysis, two models will be examined. In the first model, Model 1, the position of the pore is uniformly randomly distributed over the entire unit cell. This assumption eliminates the need for the 9 sub-regions. In the second model, Model 2, the probability that a pore occurs in the center of the unit cell, or ridge, is greater than that for the edge of the ridge. For this case, the pore can be located anywhere within the sub-division of the unit cell with equal probability. Note that the intra-ridge pore models essentially represent one column (or row) of the stochastic modified lattice pore model described in Section 4.1.

The distribution of distances between consecutive intra-ridge pores in a sequence can be simulated using these two models and the results compared to the distribution obtained using the actual measured data in section 5.3.2.

5.3.3.1 Ashbaugh model

Ashbaugh's critical assumptions about the "pore pod" are that it is symmetric in shape (width = height), a series of pore pods form a ridge, and the pore can be located anywhere within the pod with equal probability. By generating uniformly distributed random vectors of length approximately equal to the number of pores measured in section 5.3.1, a string of pores was modeled and the distance between consecutive pores calculated. The resulting distribution is

roughly triangular in shape and a simulation is shown in Figure 13.
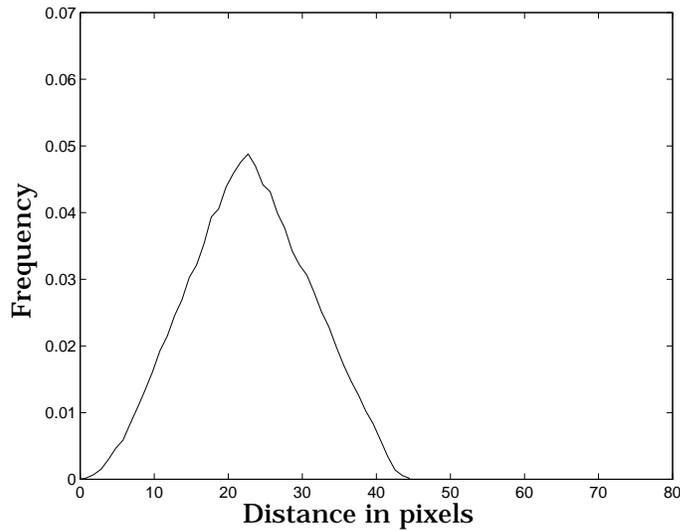
Figure 13. Distribution of intra-ridge pore spacing according to Ashbaugh's assumptions.

Comparing Figure 14.a, the distribution of the measured data, it is evident that these two distributions do not match; therefore, modifications to the Ashbaugh assumptions are required to model the distribution more accurately.

5.3.3.2 Model 1
Model 1 is an extension of the Ashbaugh model. The assumption of randomly located pores within a unit cell is similar to the assumption of randomly located pores within pods. The difference is that Model 1 incorporates unit cells with rectangular dimensions instead of squares with the width corresponding to the average width of a ridge and the length being the average adjusted spacing between intra-ridge pores (derived from measured data). In addition, the provision of a unit cell with no pore present (skipped pores) in Model 1 is used to account for the long tail of the distribution of the real data. The probability of a skipped pore is estimated using the distribution of distances derived from measured data and is approximated as 8.3%. Finally, Model 1 limits the minimum possible spacing between pores so that simulated pore positions are not unreasonably close together. The distribution of a set of simulated pore positions based on Model 1 is shown in Figure 14.c and is very similar to the distribution measured from real data.

5.3.3.3 Model 2
Model 2 is an extension of Model 1 and is meant to account for the fact that in real fingerprints the pores tend to be located on the center of the ridge. The assumptions used in Model 2 are the same as those used in Model 1 except that the probability of the pore being located in the center of the ridge is higher than the probability of a pore occurring on the edge of the ridge. It is seen from Figure 14.d that Model 2 simulates the real data distribution closely but there is not a significant difference between simulations of Models 1 and 2.

5.4 Ridge-independent pore configurations
Knowing the probability of occurrence of a sequence of intra-ridge pores is of value for proving the efficacy of using pores for identification. However, in practice, it may be unnecessary or too
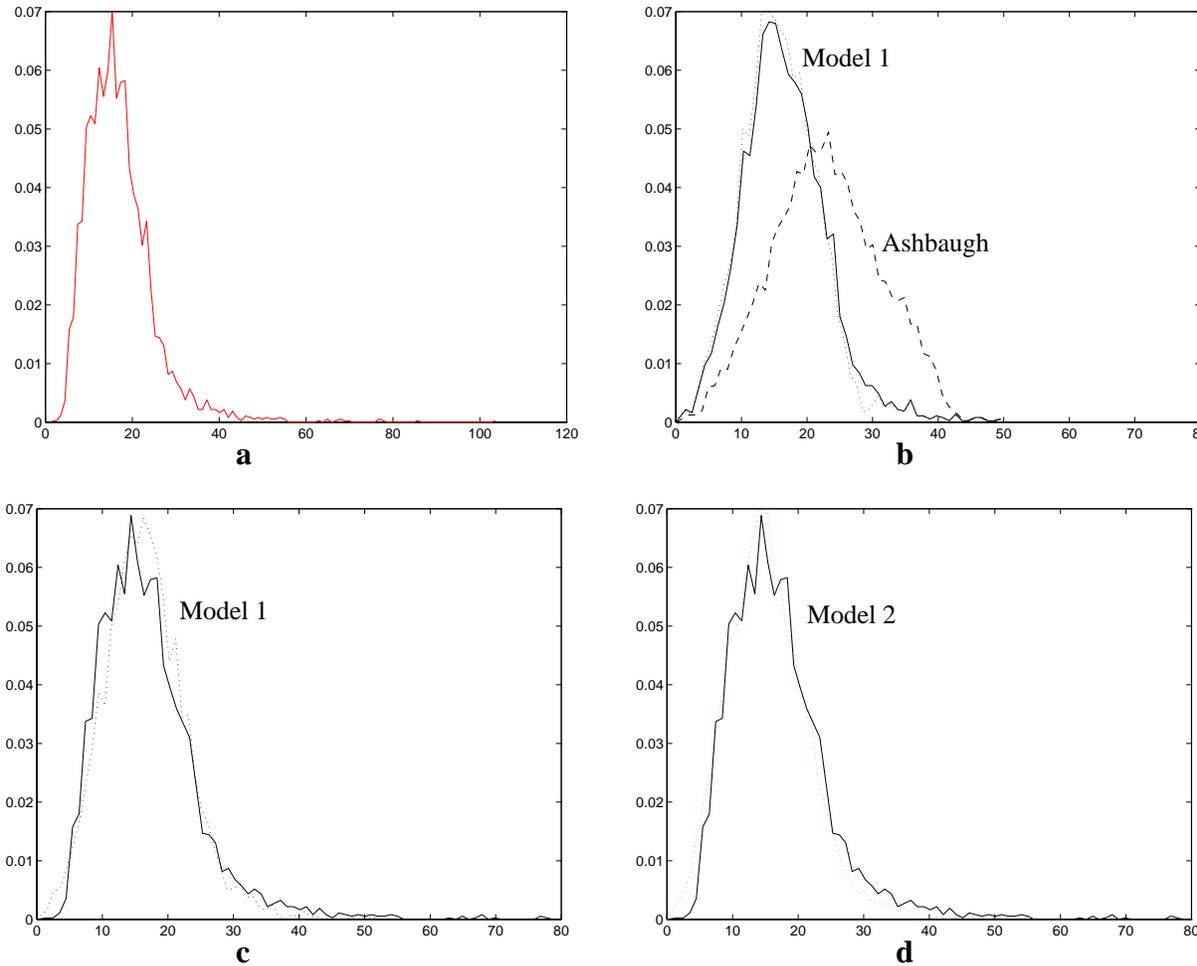
Figure 14. Distribution of intra-ridge pores. The measured distribution of spacing between consecutive intra-ridge pores is shown in **a** ($\mu = 16.96$, $\sigma = 8.16$) (where distance is measured in units of pixels). Ashbaugh's model for pore distribution generates the dashed curve in **b**. Model 1 is an extension of the Ashbaugh model which includes non-square unit cells and also a minimum allowed separation of pores and is represented by the dotted graph in **b**. Model 1 also has a provision for missing pores, which can be used to account for the tail of the real data distribution as shown in the solid curve in **b**. For Model 1, the unit cell size was 21 pixels wide by 13 pixels long, the fraction of skipped pores was 0.083, and the minimum allowed pore separation was 1 pixel (m = 16.7 and s = 6.5 for model 1 with 50000 samples). **c** shows the close fit between the real data and the Model 1 simulation. The second model, Model 2, includes the ability to force pores to be located on the center of the ridge with a higher probability than on the edges. With the probability of a pore occurring in the middle 1/3 of the ridge increased, the distribution is plotted in **d** along with the real data distribution for comparison.

difficult (for example with very noisy images) to associate pores with specific ridges. In these cases, extracting only the pore information while disregarding their ridge association is preferable. In this section, the value of ridge-independent configurations of pores will be examined. For this analysis, the configurations are made up of pores which exist in a local region of the fingerprint but may reside on several different ridges.

5.4.1 Binomial distribution of pores

The technique used by Osterburg, outlined in section 5.1, for determining the uniqueness of a configuration of minutiae can be extended for use with pores. For this purpose, 67 different prints of varying quality were analyzed on a 5x5 pixel scale (see Figure 7 to get a feel for this size segment). Each image was divided into 5x5 pixel segments and the number of pores per segment was counted. In such a small area, there is little chance of more than one pore occurring; therefore, a binomial distribution should result. From the data, 93.3% of search areas contained no pores and 6.7% contained one pore.

Defining

$$p_\alpha = P(one\ pore\ in\ a\ 5x5\ pixel\ cell)\ = 0.067 \qquad\qquad Eq.10$$

and

$$q = 1 - p_\alpha = P(no\ pores\ in\ the\ cell) = 0.933$$

and assuming that pores are independent, the probability of occurrence of a configuration of N_P pores in a region of N_C cells is given by:

$$P(N_P)\ =\ p_\alpha{}^{N_P}q^{(N_C - N_P)} \qquad\qquad Eq.\ 11$$

where there are

$$\binom{N_C}{N_P}\ =\ \frac{N_C!}{N_P!(N_C - N_P)!} \qquad\qquad Eq.\ 12$$

different configurations possible when N_P pores are present.

Assuming independence among the pores, the binomial distribution can be used to yield the probability of any particular configuration of pores occurring in any area of print. For example, a section of fingerprint which measures 20x20 pixels (0.462 mm x 0.462 mm) consists of an array of 16 grids of size 5x5 pixels each. Therefore, the probability of occurrence of a configuration with between 0 and 16 pores in that area can be calculated. Table 3 in Appendix A.5 summarizes the results for this analysis. From the table, the most likely configuration of pores in a 20x20 pixel region has a single pore. The probability of occurrence of a certain configuration with one pore is 0.0237 and since there are 16 configurations of one pore possible, the likelihood that one pore is present (regardless of the configuration) is 0.379.

In practice, a larger segment of fingerprint is desirable. The results for a 40x40 pixel (0.924 mm x 0.924 mm) area, 4 times the area of the 20x20 pixel segment, are discussed next. For this case, the segment consists of 64 grids of size 5x5 pixels (see Figure 15 for a typical example). A segment this size is large enough to contain significant ridge structure while not exhibiting distortion due to finger plasticity sometimes present in larger areas of print. As seen in Table 4, the most likely number of pores in this area is 4, with a probability of 20.0%. (Figure A.5.1 shows the relationship between the number of possible configurations for a given number of pores, the

probability of any particular configuration of a given number of pores, and the overall probability of a given number of pores occurring.) As an example, any configuration of 4 pores, like the one shown in Figure 15 has a probability of occurrence of 3.14×10^{-7}, and there will be 6.35×10^5 different configurations of 4 pores in a 64 grid area. For comparison, from the table, it is noted that the most likely configuration is the one with no pores, which occurs with a probability of 1.18%. For perspective, assuming the most likely sequence of pores occurs in a 40x40 pixel segment, the probability of a different fingerprint matching is 0.0118 based on comparison of pore location only with a resolution (or measurement accuracy) of 5x5 pixels.
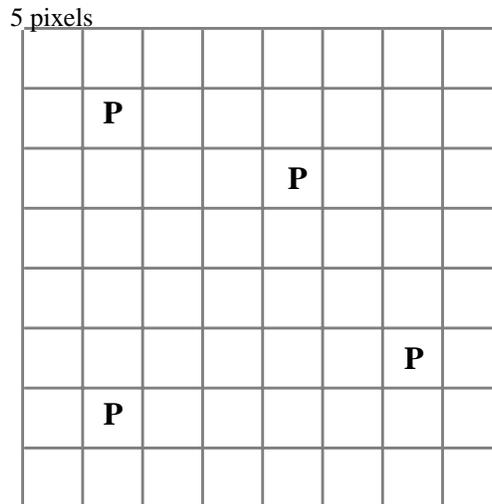

Figure 15. A possible pore configuration. Using a 5x5 pixel grid (equivalent to 0.115 mm x 0.115 mm area of print) a binomial pdf can be generated from the probability of a pore being present or absent in the 5x5 cell. With an analysis area of size 40 pixels x 40 pixels which is 0.92 mm x 0.92 mm, there are 8x8 grids of size 5x5 pixels. The probability of this configuration (or any configuration of 4 pores) is 3.14×10^{-7}.

In order to compare the results for intra-ridge sequences to ridge independent configurations, assume that 20 pores occur in an area of size 4 mm^2. This is a good approximation based on the density of pores being about 5 pores/mm^2. Again using cells of size 5x5 pixels, the area consists of 300 cells. Given this area, the probability of occurrence of a configuration of 20 ridge independent pores would be,

$$P_{RI}(a\ configuration\ of\ 20\ ridge\ independent\ pores) = 1.23 \times 10^{-32} \qquad \text{Eq. 13}$$

which is 1.06×10^{-18} times smaller than the probability of an intra-ridge sequence with the same number of pores (assuming all pores are spaced by the most likely separation).

5.4.2 Measuring configuration probabilities

In the previous discussions, the underlying assumption of independence makes uniqueness calculations possible. In reality, though, the independence assumption is not accurate. There appears to be a definite influence on a pore's position depending on the relative positions of the neighboring pores. If the independence assumption is not valid, then the assumption that all possible configurations of N pores are equally likely is also not valid. In this case, it is desirable

to determine the exact probability of occurrence of each possible configuration of N pores by finding the histogram or pdf of the configurations. This information could be used by the processing routine to ignore highly likely configurations and to search for very distinctive pore configurations providing inter-class separability. The expected outcome is a reduction in the number of false accept errors.

As alluded to before, a segment with an area of between 1 and 4 mm^2 should be optimal for processing pores. For this size area, the number of pixels (at 1100 ppi = 43.3 ppmm scanning resolution) is approximately 40x40 (1600) to 80x80 (6400). The resolution is effectively reduced by analyzing 5x5 pixel segments, leading to areas of 8x8 (64) or 16x16 (256) cells. In fact, to determine the histogram of all possible configurations is unreasonable even for the simplest case of low resolution and smaller area. There are 2^{64} possible configurations for the smaller area with low resolution and 2^{6400} configurations at high resolution and larger area. As an interesting note, about 4,000 km by 4,000 km of print area would be needed in order to fill each of the 2^{64} histogram bins with only one entry. Even with the enticement of free doughnuts, it is unlikely that the researchers could have gathered sufficient subjects to accumulate the data for this experiment.

These numbers are actually exaggerated, since only a small subset of the total number of theoretical configurations are really possible. Based on measurements, there were never more than 12 pores detected in a print area of size 40x40 pixels, and this event had the remote probability of 0.03%. The most probable number of pores to occur in this area was 4, with a probability of 22.3%. The complete distribution of measured data is shown in Figure 16. Even by applying such a realistic constraint on the pore density, the number of configurations is still enormous.
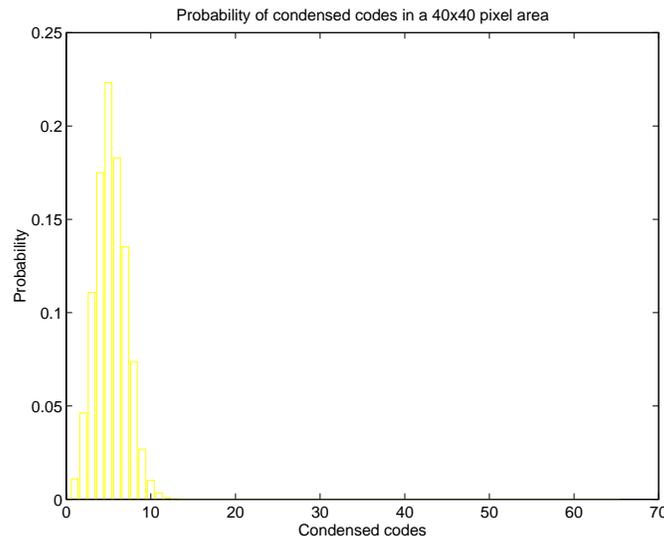
**Figure 16.** Measured density of pores (pores/mm^2) in a 40x40 pixel area of print.

5.4.3 Measuring configuration code probabilities
Even though the actual probability of a particular configuration cannot be determined, valuable information can still be gained by analyzing ridge independent configurations of pores within realistic sized regions. The next two sections will outline an analysis whereby a binary code

method was used to represent pore configurations. Implementing a binary code (1 = pore, 0 = no pore) in a 4 cell area of analysis to describe a configuration yields 16 different possibilities within those 4 cells (0000, 0001, ..., 1111). As already mentioned, executing a high area (many cells) or high resolution analysis is not practical. Therefore, using a pyramid paradigm, a low resolution and small area analysis was performed first and "tiers" of analysis areas were built around this basic unit. Once the accuracy of the information gathered from the small area, low resolution study was established, a higher resolution study was conducted. Sixty-seven different images of varying print quality were used for this analysis.

5.4.3.1 Low resolution configuration code study

The exact size of the basic unit of analysis was determined using the average pore density in a print, where

size of analysis unit = (total number in pixels/total number of pores)$^{1/2}$
*-> 19 x 19 pixels.*
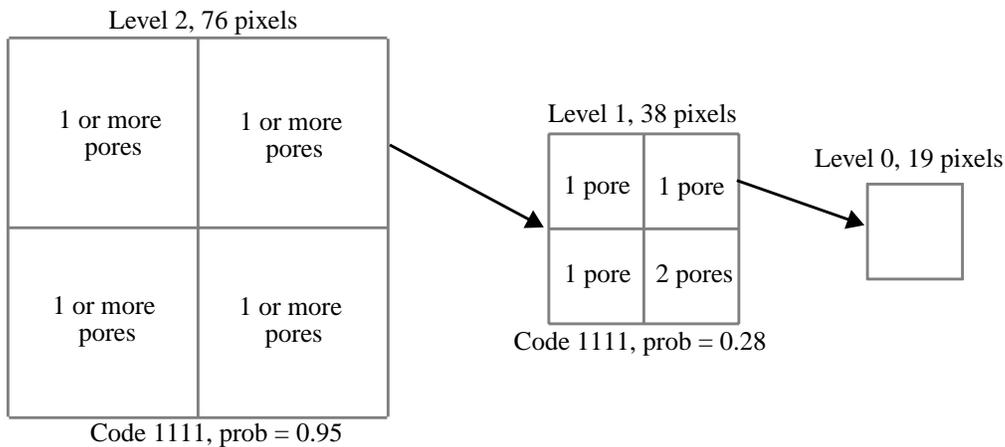
Figure 17. Ridge independent pore model. For example, in a 76x76 pixel area (level 2), the code is 1111, at higher resolution (level 1), each 38x38 pixel block has at least one pore present, giving a code of 1111. Finally, at the Level 0 analysis, the exact distribution of the number of pores present is known within a 19x19 pixel area. For this example, assume that each level 1 segment contains 5 pores (1 in each of three blocks and 2 in the remaining block), then the probability of this particular occurrence of 20 pores is 5.186 x 10^{-8}.

This square analysis area is the "Level 0" unit of a three-tiered construct (see Figure 17); within this level, the distribution of the number of pores present was determined. The most frequent occurrence was 1 pore in a Level 0 analysis area which happened in 48% of the units. The frequency of pores per analysis unit is shown in Figure 18.

Two higher levels are used to formulate the binary code discussed earlier. The next tier is composed of 4 units of the Level 0 area of analysis (38x38 pixels). Within each Level 0 unit, a one represents the existence of at least one pore, and a zero indicates that no pores are present in that area. The binary digits from these 4 blocks generate a 4-digit code that can be analyzed to determine which codes occur the most or least frequently. This model precludes establishing the exact pixel location of each pore.

The probability of one pore existing in a Level 0 unit is significantly different from the probability
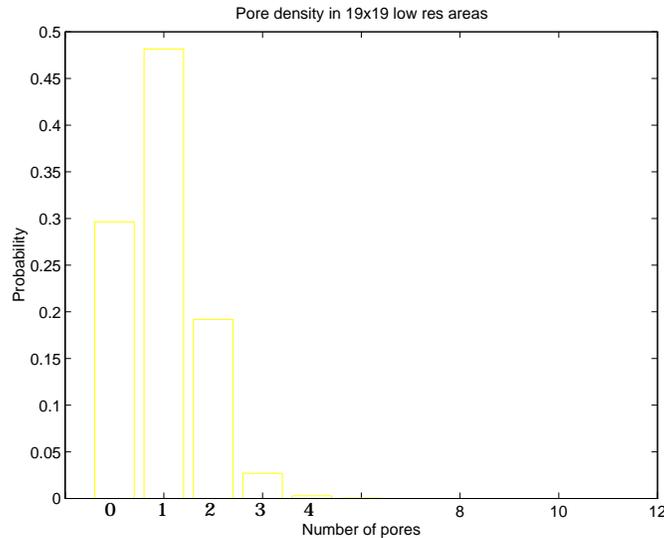
Figure 18. Measured density of pores in a 19x19 pixel unit cell.

of five pores residing in a unit, and thus these probabilities, in combination with the probabilities of the 4-digit codes, or Level 1 configurations, contribute to the overall probability of an actual pore configuration at Level 1. For Level 1, the most frequently occurring code was 1111, which occurred 28% of the time. The code frequencies are shown in Figure 19, where the next most frequently occurring codes were those containing three ones and a zero; each of these occurred about 10% of the time. The least frequent (and therefore the most valuable) code was 0000, which occurred with a frequency of 1.5% (this number actually matches closely the result established in section 5.4.1 in which the calculated probability of no pores is 1.18%).
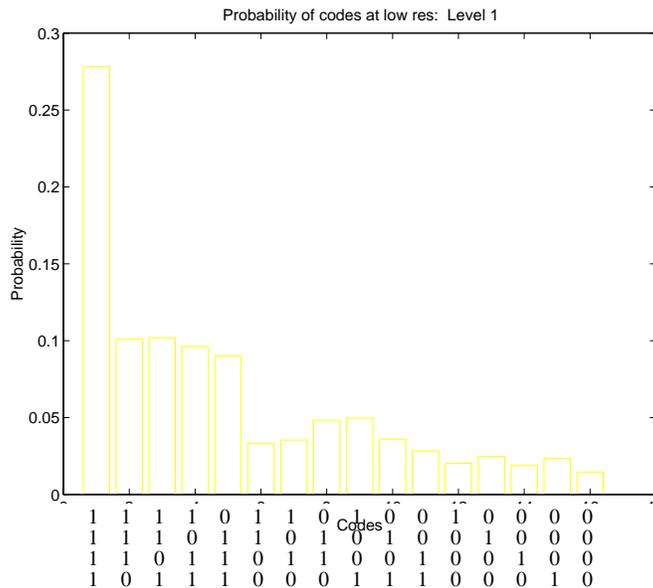
Figure 19. Measured frequencies of "codes" occurring in a 38x38 pixel area, where a binary "1" indicates the presence of at least one pore in a 19x19 pixel area, and a "0" indicates an absence of any pores. The codes are comprised of these 1s and 0s to represent configurations of pores.

The probability of a specific configuration occurring, for example, 1111 at Level 1 with

corresponding densities in the Level 0 units of 1, 2, 3, and 1 pores, is thus:

$$P(configuration) = P(code\ 1111\ at\ Level\ 1)x$$
$$P(density=1)P(density=2)P(density=3)P(density=1)$$
$$= (0.28)x(0.48)(0.19)(0.03)(0.48) = 3.68\ x\ 10^{-4}$$

which is the probability of a configuration of 7 pores in an area of 38x38 pixels.
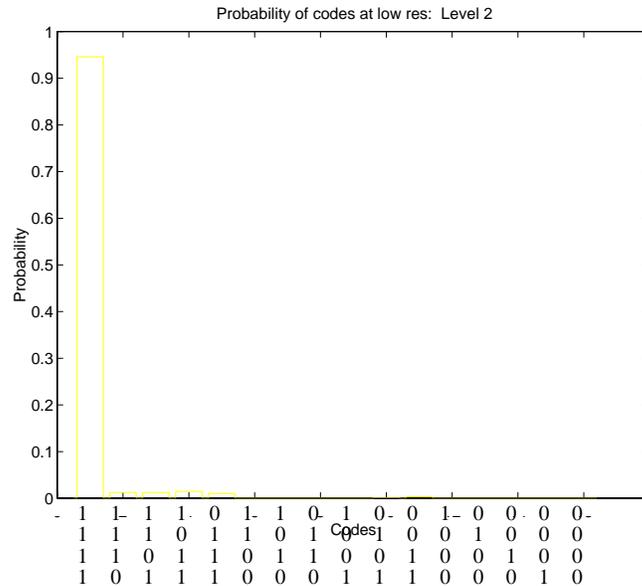
Figure 20. Measured frequencies of "codes" occurring in a 76x76 pixel area, where a binary "1" indicates the presence of at least one pore in a 38x38 pixel area, and a "0" indicates an absence of any pores. The codes are comprised of these 1s and 0s to represent configurations of pores.

Four Level 1 units can be combined to make up the third tier (Level 2), which is simply an extension of the Level 1 analysis. Each Level 1 unit contains a zero if none of its Level 0 units have any pores in it, and the Level 1 unit contains a one if any of its Level 0 units contain at least one pore. The 4 digits from these Level 1 units, four of which make up Level 2, generate another 4 digit binary code, the different possibilities of which are analyzed similarly to the codes generated by the Level 0 units. The Level 2 analysis generated codes of 1111 with a frequency of about 95%, which is evident in Figure 20. Because the Level 2 analysis is at such a low resolution (low measurement accuracy), many of the codes, especially those containing mostly zeros, never occurred.   Using an example with 20 pores (see Figure 17 for details), the probability of a particular configuration is determined to be,

$$P_{Levels}(a\ configuration\ of\ 20\ ridge\ independent\ pores) = 5.186\ x\ 10^{-8} \qquad \text{Eq. 14}$$

The codes generated at Levels 1 and 2 and the various local pore densities of Level 0 generate three different sets of probabilities, all of them combining to give the overall probability of pore configurations in an image.   This analysis, however, is performed at a very low effective resolution, and though it provides actual configuration information, a higher resolution analysis

will yield more valuable results.

5.4.3.2 High resolution configuration code study - Binomial Distribution

The previous analysis was repeated at a higher resolution. For this case, instead of using individual codes, such as 1101, the number of ones in the code were summed (i.e., 1101 = 3 ones). This limited analysis was necessary because the number of possible codes for a reasonable area of print is exceedingly large. The exact configuration information is lost when the codes are condensed in this way and the codes simply represent the number of pores in a given area. The result is a histogram of the number of pores in a given area of print which can be compared to the results achieved assuming independent cells with a binomial distribution as was done in section 5.4.1.

Instead of the 19x19 pixel area (Level 0 analysis) which was analyzed in section 5.4.3.1, a 5x5 pixel cell was used. As was stated before, a binomial distribution results with 93.3% of the search areas empty and 6.7% containing one pore. The actual probabilities of the condensed codes were found by analyzing an area 4 cells by 4 cells (20x20 pixels) and then an area 8 cells by 8 cells (40x40 pixels). These analyses yields the distribution of either 16 (small area) or 64 (large area) possible "condensed codes." Similar to Level 1 of the low resolution study, ones and zeros were assigned to blocks according to whether or not pores were present.

Theoretically, a similar distribution could be derived by extrapolating the binomial distribution (the probability of one or zero pores). This was done in Section 5.4.1 using Osterburg's method. However, this method assumes independence. Section 8 of this paper explores the concept of dependence among pores and demonstrates why extrapolating a distribution from the binomial is not as accurate as measuring the pore density directly.

6. Reliability - FRR analysis

Many factors that determine the FAR of a system were discussed in the previous sections; Section 6 is devoted to factors that can affect the FRR. Whereas the FAR analysis centers on the differences between an impostor and an authorized user, the FRR focuses on variability that occurs within an authorized user's fingerprints over time. These variations can be studied to determine the feature reliability, with regard to physiology and algorithm.

To ensure authentication of an individual with 100% accuracy, either the individual's live-scan fingerprint must be exactly the same as the enrolled print or, in the presence of noise and distortions, the features of the live-scan print corresponding to the enrolled print must be extracted without error. In the real world, noise and distortions are always present, and no automated process is perfect. In addition, the physiological reliability of pores falls short of 100%; therefore, the inherent and algorithm (or processing) reliability warrant further study.

6.1 Inherent reliability

The physiological reliability of pores (or inherent reliability, R_i) depends on environmental factors; temperature and skin condition can conspire to alter or suppress altogether the physical appearance of a given pore. Individual pores from 516 images of ten different fingerprints were

analyzed to determine R_i with respect to both pore visibility (or detection) and size (both absolute and relative to other pores in the image). Clarity of the pore, image quality, skin condition, and pore density were also recorded. The results can be seen in Table 6 of Appendix A.7. It should be noted that all of this data was collected manually (by eye) to prevent introducing algorithm errors.

The specific pores studied were visible on average in 91% of the images ($R_i = 0.91$). The least reliable pore was visible in only 75% of the fingerprint images. In this case, the reason for the low reliability is that during capture of these images, the individual altered his prints through a variety of means, i.e., gripping a cold soda can prior to image capture. Therefore, 75% can be estimated as a lower bound for R_i. Although pore size and shape are of significance, the most important aspect of R_i is whether or not the pore is actually present (detectable). It is also important to remember that, although the lower bound for detection is 75%, that lower bound is for one pore, not a configuration of pores. If twenty pores are used to match prints, the 75% refers to the reliability of only one pore out of the twenty.

Although the characteristic skin condition, image quality, size, and shape consistency varied somewhat among the individuals' prints, there were several correlations between categories. For instance, the more neutral prints (with regard to skin condition) had the highest image quality. Furthermore, prints of lower image quality tended to correspond to a dry skin condition, and their pores were less consistent in shape. Finally, circular pores proved to be the most reliable with regard to shape.

6.2 Algorithm reliability

The reliability of the algorithm, R_d, was also examined. Using high quality prints, the prevalence of missed detects and false detects was recorded and the causes for both types of errors were assessed. The detection algorithm missed 11% of the pores present ($R_d = 0.11$), and had a false detect occurrence of 1%. It was found that the predominant source of missed detections was the thresholding stage used in preprocessing to convert grey scale images to binary. These processing errors occurred in 4.8% of the detections. On the other side of the coin, most of the false detects were caused when the algorithm detected a pore in the middle of the valley. This physically impossible situation may have resulted when the curvature of the ridge implied that a pore was present on the edge of the ridge. The algorithm might interpret such a structure as a pore. A description of the events causing missed detects and false detects can be found in Table 7 in Appendix A.7.

These reliability statistics apply to a single pore in a configuration. The probability of the algorithm missing a given pore is 11%, but the probability of missing a configuration of many pores is orders of magnitude smaller.

In further discussions, R_i will be defined as the probability of a feature appearing in a fingerprint image, and R_d the probability that the feature is properly detected by the algorithm. Therefore, R (total reliability), is defined as the probability that a feature appears and is properly detected.

7. Performance

Section 5 examined the uniqueness of configurations of features and section 6 addressed feature reliability. In this section, uniqueness and reliability are conjoined to establish performance, which is defined in terms of the number of false reject and false accept errors the system produces. The number of false accept errors is related to the uniqueness of a configuration while the number of false reject errors depends on the reliability of the features. Some parameters which contribute to uniqueness have been discussed earlier: number of features, density of features, and feature area. Parameters critical to the reliability are the inherent feature reliability and the efficiency of the feature detection algorithm.
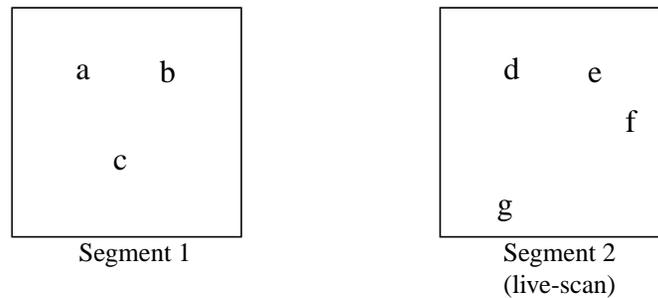

Figure 21. Matching the enrolled segment (segment 1) to the live-scan segment (segment 2). Features a and b in segment 1 match features d and e in segment 2. If both segments originate from the same finger, then features d and e are either reliable features which were correctly detected or they are false detections and a and b were unreliable. Features f and g are false detects, and feature c is an unreliable feature. If the segments originated from different fingers, then features d, e, f, and g are randomly positioned and can be either real features or false detects.

Consider comparing two fingerprint segments of equal size. Segment 1 is the enrolled segment and segment 2 is a subsequent live-scan which originated from either the same user or a different user. The comparison is based on feature location only and segment 1 contains n_1 features while segment 2 contains n_2 features. Assume that there were no errors in detecting features during enrollment (all real and no false features were detected). For the two possible sources of segment 2, it is necessary to determine the probability density function for the feature matching score.

Note that prior to this section all references to probability of matching or uniqueness of a set of features related to the entire set of features. Every feature was required to match for the entire set to match. In this section, a more realistic approach is taken in which the number of features in both segments as well as the number that actually match are taken into account. A matching score provides the degree of matching between two segments with a range of a complete non-match to a complete match.

Some relevant parameters needed for the matching problem are:

n_c - the number of cells or feature areas within the segment of print being analyzed
n_1 - the true number of features in segment 1 - the enrolled segment
n_2 - the number of features detected in segment 2 - the live-scan segment

n_m - the number of matching features

n_{fd} - the number of features falsely detected

p_{md} - probability that a valid feature is not detected

$p_{cd} = 1 - p_{md} = R_d$ - probability that a valid feature is correctly detected

p_{fd} - probability of detecting an invalid, or false, feature at any location

$p_{cm} = 1 - p_{fd}$ - probability that a feature is not detected in an invalid location.

If segment 2 originates from the same user as segment 1, then reliability must be addressed:

R - reliability

R_i - inherent reliability of the feature

R_d - algorithm detection reliability

$R = R_i \times R_d$

In addition, two sources contribute to the number of matching features:

$n_m = n_{m,R} + n_{m,F}$

$n_{m,R}$ - the number of correct (valid feature) matches

$n_{m,F}$ - the number of false detects (invalid features) that match real features

Figure 21 describes some of the relevant parameters required for performance analysis.

Reliability is defined as the probability of detecting a valid feature in the correct position. R can range from 0 (totally unreliable features) to 1.0 (no missed detects). When segment 2 is from an impostor print, R is considered to be 0 and p_{fd} is set equal to the measured value for the probability of a pore in a grid cell. This situation simulates a randomly located set of independent features in segment 2.

A feature match is defined as the detection of a feature in the live-scan segment at a valid (or enrolled) feature location. A feature mismatch is defined as any feature in the live-scan which does not match an enrolled feature. It is possible to make a false detection at the location of a valid feature which is unreliable. This situation results in an incorrect (false) match but is not a mismatch.

Define the feature matching score to be:

$$S_F = \frac{n_m}{n_1} - \frac{n_2 - n_m}{n_{2,\,max}}$$

Eq. 15

where $n_{2,max}$ is the maximum number of features allowed in segment 2.

The range of S_F is [-1,+1] where a score of -1 corresponds to the case where $n_m = 0$ and $n_2 =$

$n_{2,max}$. This is the worst possible match; no features match and segment 2 contains the maximum number of detected features allowed. A score of +1 results when $n_1 = n_2 = n_m$, a perfect match with no mismatching features.

Given n_c, n_1, n_2, $n_{m,R}$, $n_{m,F}$, R, and p_{fd}, the probability of n_m pores matching can be determined by calculating the probability of the score resulting from each variation of the input parameters and then accumulating the probability of like scores. The result is the pdf of the matching score, S_F, which can be used to calculate false accept and false reject error rates.

First assume that the two fingerprint segments originate from different fingers. Furthermore, for simplicity, assume that the features are independent. Given an enrolled feature set, the number of features in segment 2 and their positions are random. Therefore, the number of matching features is a random variable. Whether features detected in segment 2 are real or result from detection errors is transparent since the only concern is how many features match. Given n_c, n_1, and n_2 the probability of matching n_m features between segment 1 and segment 2 is:

$$p_{FAR}(n_m) = \left[\binom{n_1}{n_m} \frac{\left(\prod_{i=0}^{n_m-1} n_2 - i \right)\left(\prod_{i=0}^{n_1-n_m-1} n_c - n_2 - i \right)}{\left(\prod_{i=0}^{n_1-1} n_c - i \right)} \right]\left[\binom{n_c}{n_2} p_\alpha^{n_2} (1 - p_\alpha)^{n_c - n_2} \right] \quad \text{Eq. 16}$$

where the product terms are valid for $i > 0$. Or, by defining:

$$p_{FAR}(n_m) = p_{n_m} p_{n_2} \quad\quad\quad \text{Eq. 16.a}$$

The first term in Eq. 16 or Eq. 16.a is p_{nm}, the probability that n_m features from segment 1 match in segment 2. The second term is p_{n2}, the probability that there are n_2 features in segment 2, where p_α, which was determined in section 5.4.1, is the probability that there is a feature in a cell. In this situation, the second term incorporates p_{fd} into p_α and reliability is not an issue since the print segments originate from different fingers.

As an example, assume that for a given $n_c = 300$, $n_1 = 20$, and $n_2 = 25$, that there are $n_m = 10$ matching features. The probability of n_2 having 25 features is 0.0457 and the probability that 10 of them match those in segment 1 is 2.5175×10^{-7} giving an overall probability of 1.1505×10^{-8} of this situation occurring. The corresponding match score, S_F, from Eq. 15 is 0.45 assuming that $n_{2,max}$ is equal to n_c.

Next, assume that the two print segments originate from the same finger. Furthermore, the rotation and positioning of the segments are assumed to be known exactly. In this case, the feature reliability and the number of false detects are of critical importance. A feature may or may not be correctly detected in the live-scan segment depending on its reliability, leading to a reduction in the number of features matched. In addition, there are random false detection errors in the live-scan image segment which, depending on their position, will match (improve the score)

or mismatch (reduce the matching score). The probability of matching n_m features is:

$$p(n_m) = p_{n_{m,R}} p_{n_{m,F}} p_{n_2}$$

Eq. 17

where $n_m = n_{m,R} + n_{m,F}$

$$p_{n_{m,R}} = \left[\binom{n_1}{n_{m,R}} R^{n_{m,R}} (1-R)^{n_1 - n_{m,R}} \right]$$

Eq. 18

is the probability of $n_{m,R}$ matches corresponding to real features which were correctly detected.

$$p_{n_{m,F}} = \left[\binom{n_1 - n_{m,R}}{n_{m,F}} \frac{\left(\prod_{i=0}^{n_{m,F}-1} n_2 - n_{m,R} - i \right)\left(\prod_{i=0}^{n_1 - n_{m,R} - n_{m,F} - 1} n_c - n_2 - i \right)}{\left(\prod_{i=0}^{n_1 - n_{m,R} - 1} n_c - n_{m,R} - i \right)} \right]$$

Eq. 19

where $p_{nm,F}$ is the probability of $n_{m,F}$ matches which are matches of falsely detected features in the live-scan segment randomly occurring at valid feature locations.

And p_{n2}, the probability that there are n_2 features in segment 2, is given as:

$$p_{n_2} = \left[\binom{n_c - n_{m,R}}{n_{fd}} p_{fd}^{n_{fd}} (1 - p_{fd})^{n_c - n_{m,R} - n_{fd}} \right]$$

Eq. 20

In Eq. 20, p_{fd} is used instead of p_α, which was used in Eq. 16.

In Eq. 17, by setting $R = 0$, $p_{nm,R}$ is 1, $n_{m,R} = 0$ therefore, $n_m = n_{m,R}$, and $n_{fd} = n_2$. In addition, if p_{fd} is set equal to p_α which was determined in section 5.4.1 then Eq. 17 reduces to Eq. 16. This situation corresponds to totally unreliable features and a false detection rate which provides the same feature density as the measured feature density. The result is that segment 2 is equivalent to an imposter print segment.

If no false detects are allowed, then p_{fd} is 0 for Eq. 17, and $n_{m,F} = 0$ and $n_{fd} = 0$, therefore, $n_2 = n_m = n_{m,R}$. For this case, Eq. 19 and Eq. 20 reduce to 1 and the result is:

$$p(n_m) = p_{n_{m,R}} = \left[\binom{n_1}{n_m} R^{n_m} (1-R)^{n_1 - n_m} \right]$$

Eq. 21

which is just the probability that n_m features are reliable given n_1 features to start.

If the reliability is 100%, then $n_{m,R} = n_1$, Eq. 18 and Eq. 19 both reduce to 1, and Eq. 20 becomes:

$$p_{n_2} = \left[\binom{n_c - n_1}{n_{fd}} p_{fd}^{n_{fd}} (1 - p_{fd})^{n_c - n_1 - n_{fd}} \right]$$

Eq. 22

which represents the probability of n_{fd} falsely detected features in segment 2 given that there are n_1 correctly detected ($n_2 = n_1 + n_{fd}$).

Eq. 17, 18, 19, and 20 are used to determine the expected performance of a system. The parameters such as the number of features enrolled, the accuracy of measurement (feature area), the feature reliability, and the algorithm efficiency can all be evaluated. The plots in Figure 22 show how variations in the critical system parameters affect performance.
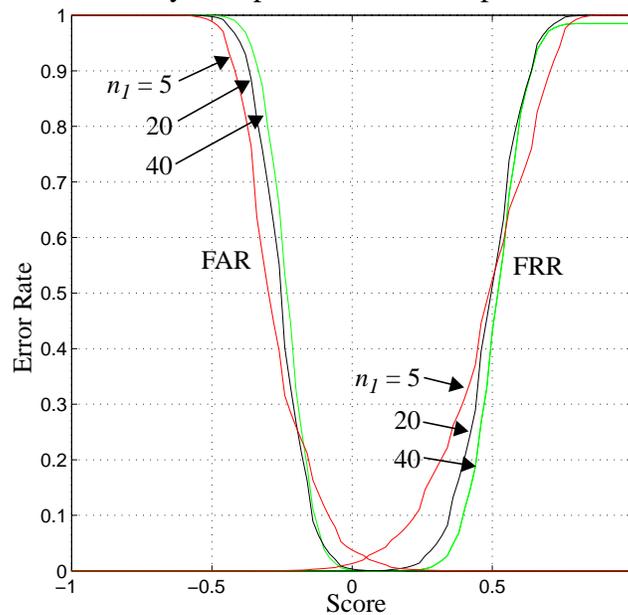
Figure 22.a. Error rate plots for variation in enrolled feature density. The parameters used in this simulation were: $n_c = 300$, $R = 0.8$ (for FRR and $R = 0$ for the FAR plots), $p_{fd} = 0.067$, n_2 ranges from [0,60], and n_1 is set to 5, 20, or 40. The Equal Error Rate (EER) is 2.43 x 10^{-2} for $n_1 = 5$, 6.2 x 10^{-4} for $n_1 = 20$, and 3.2 x 10^{-4} for $n_1 = 40$.
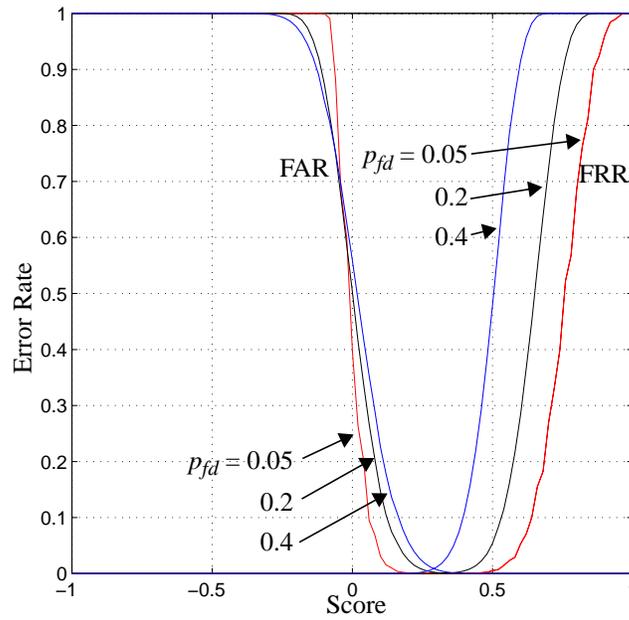
Figure 22.b. Variation in feature detection error rate (fingerprint noise level). The parameters used in this simulation were: $n_c = 300$, $R = 0.8$ (for FRR and $R = 0$ for the FAR plots), $n_1 = 20$, n_2 ranges from [0,300], and p_{fd} is set to 0.05, 0.2, or 0.4. For a lower detection error rate (cleaner fingerprint image), an enrolled user will gain access more often (the FRR error rate curve is steeper and more towards the right). For high detection error rates (noisy images), an imposter will have a higher feature density and will have a greater chance of gaining access. The Equal Error Rate (EER) is 7.2 x 10^{-4} for $p_{fd} = 0.2$, and 9 x 10^{-3} for $p_{fd} = 0.4$.
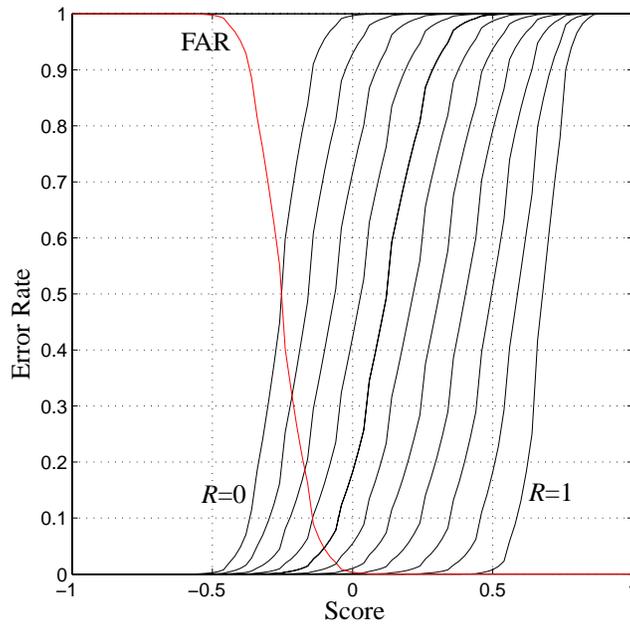
Figure 22.c. Error rate plots for variation in feature reliability. As R decreases, the performance of the system degrades. An FAR plot is shown for comparison of actual system performance as a function of the feature reliability. The parameters used in this simulation were: $n_c = 300$, $n_1 = 20$, $p_{fd} = 0.067$, n_2 ranges from $[0,60]$, and R ranges from 0 to 1.0 in 0.1 increments ($R = 0$ for the FAR plot).
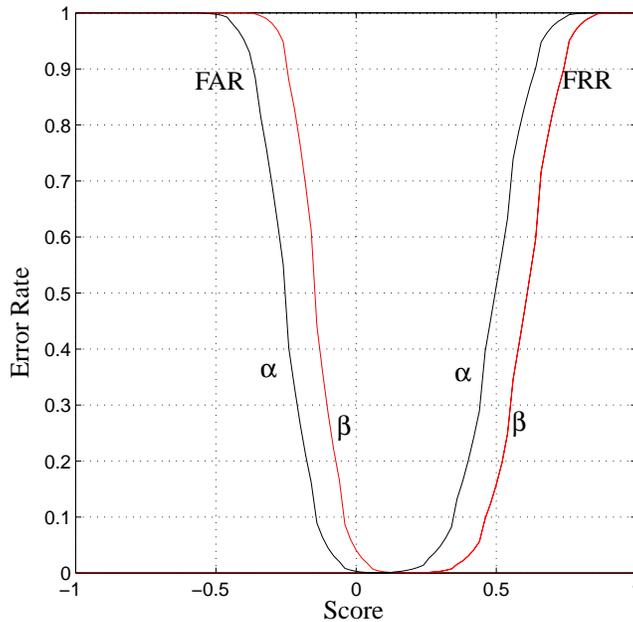
Figure 22.d. Resolution or measurement accuracy curves. Parameters for α curves: $n_c = 300$, $n_1 = 20$, $R = 0.8$ (for FRR), $p_{fd} = 0.067$, n_2 ranges from $[0,60]$. For β curves: $n_c = 200$, $n_1 = 20$, $R = 0.8$ (for FRR), $p_{fd} = 0.1$, n_2 ranges from $[0,60]$. Both sets of parameters simulate a comparison of two segments with the same area and detection error rates. The α curves simulate a system using a more precise feature position determination and corresponding smaller search area than the β curves. Higher resolution will tend to make it more difficult for both a valid user and an imposter to match (but the EER may be the same as for lower resolution settings) as seen from the curves.

8. Dependence

The models describing feature position discussed in earlier sections were based on an assumption of independence between the features (either minutiae or pores). Although this assumption is attractive because it simplifies calculations, it does not accurately describe actual feature placement. As alluded to in previous discussions of physiology and pore placement models, the placement of a feature depends in some way on the relative position of neighboring features.

8.1 Sclove's Model of Dependence Between Galton Features (Minutiae):

The idea of dependence of minutia placement on the location of other minutiae is not a new one. In 1979, Stanley Sclove, a member of the Osterburg team, published a paper proposing that minutia occurrence could be described by a Markov-type model (Sclove, 1979). He asserted that the presence of minutiae in a given cell is influenced by the presence of minutiae in the adjacent cells. This study supplanted the model assuming independent cells with one incorporating dependence between cells.

To summarize Sclove's theory: For a 3 cell by 3 cell area, the cell under scrutiny is the center one, or the fifth in a linear ordering. The number of cells in this ordering containing a Galton feature is represented by a, the number of adjacencies, which ranges between 0 and 4 (in a linear ordering, only the cells preceding the fifth cell will affect that cell). For a random vector $\underline{Y}(c)$ describing the outcome of the cth cell, Sclove's probability of a set of Galton features is represented as:

$P=P[\underline{Y}(c)=\underline{y}(c)]P(\underline{Y}(2)=\underline{y}(2)/\underline{Y}(1)=\underline{y}(1)]P[\underline{Y}(3)=\underline{y}(3)/\underline{Y}(2)=\underline{y}(2),$
$\underline{Y}(1)=\underline{y}(1)]...P[\underline{Y}(5)=\underline{y}(5)/\underline{Y}(4)=\underline{y}(4),\ \underline{Y}(3)=\underline{y}(3),\ \underline{Y}(2)=\underline{y}(2),\ \underline{Y}(1)=\underline{y}(1)].$ Eq. 23

For simplicity, the four preceding adjacent cells $(\underline{Y}(1)...\underline{Y}(4))$ are referred to as matrix $\underline{X}(c)$, where the conditional probability of Eq. 23 is reduced to

$P[\underline{Y}(c)=\underline{y}(c)/\underline{X}(c)].$ Eq. 24

Assuming that the cells in $\underline{X}(c)$ exert influence over cell c, the probability of feature occurrence involves the number of adjacencies, a, as well as the different types of features, f; if a cell is occupied, there are still 12 different event possibilities (using data from Table 1 in the Appendix). Sclove's probability of a minutia (or Galton feature) occurrence in a cell is then:

$$P \ = \ [k(0|a)\log P(0|a)]$$

$$+ \ \sum_{f\ =\ 1}^{12} [k(f)\log P(f)] - ((k(1) + k(2) + ... + k(12))\log P(E))$$

$$+ \ \sum_{a\ =\ 1}^{4} m(a)\log P(E|A=\ a)$$
                                                            Eq. 25

where $k(0|a)$ is the number of empty cells with a adjacencies, $k(f)$ is the number of cells containing the probability of minutia type f, $P(E)$ is the probability of an event occurring, and $m(a)$ is the number of occupied cells with a adjacencies. Sclove found that as a, the number of

adjacencies increases, the probability of the central cell being occupied increases monotonically. Therefore, clustering of minutiae happens more frequently than uniformly spaced minutia configurations. Thus is born the concept of dependence among minutiae. This dependence would be evident in a pdf as a non-uniform distribution, whereas the equally-likely configurations associated with independence give rise to a uniform pdf.

8.1.1 Dependence of pores: intra-ridge

Although Sclove dealt exclusively with Galton features, his idea of feature dependence can also be applied to pores, albeit pores are features with properties differing from minutiae. Minutiae, on one hand, are considered "accidental;" their spatial distribution serves no apparent purpose. In contrast, pores transfer heat, so they must have a spatial distribution that can support this function. Whereas minutiae are expected to cluster, pores are expected to be spread out over the finger.

Two events are independent if "the probability of either one is unaffected by the occurrence of the other," (Lapin, 1990). An event is the placement of a pore at a given position, thus:

$$f_{X,Y}(x,y) = f_X(x)f_Y(y)$$
<div align="right">Eq. 26</div>

where $f_X(x)$ and $f_Y(y)$ are the probability density functions of the independent random variables.

If pores were indeed placed independently of each other, what would the frequency distribution of intra-ridge distances look like? In order to plot such a distribution, a vector of zeroes was used to represent a one-dimensional "ridge." 3370 "pores" (ones) were substituted for zeroes at random intervals along the ridge, simulating a situation where the placement of one pore was entirely independent of the placement of other pores. The distances between pores was calculated and the frequency of these distances was plotted (Figure 23).
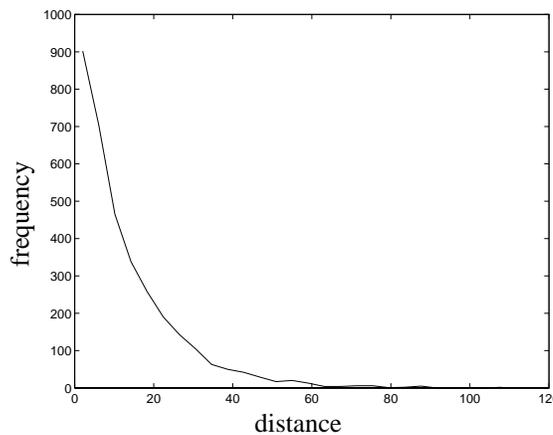
Figure 23. Distribution of distance measured between consecutive intra-ridge pores. This data was simulated using a model for which the pores are assumed to be independent and occurring at a random position on a one-dimensional axis. Therefore, there is no constraint on the placement of any particular pore and no influence on the position of a pore from its neighboring pores. The distribution decreases monotonically with distance, and there are numerous pores that are close together.

In order to prove that pore placement is a dependent phenomenon, this calculated distance distribution should be compared to a measurement of actual intra-ridge distances. However, intra-

ridge positions would have to be measured in the one-dimensional longitudinal direction only. Unfortunately, the intra-ridge distances obtained in this study were of a two-dimensional nature and cannot be used justifiably in a comparison.

Had the data been obtained in a manner suitable to comparison, independence could be proven or disproved by virtue of a chi square analysis. The chi square statistic is defined by Lapin as:

$$\chi^2 = \frac{1}{n}\sum_i\sum_j\frac{(fij - \hat{f}ij)^2}{\hat{f}ij}$$

Eq. 27

where f_{ij} are the actual intra-ridge distance frequencies obtained, and $\bar{f}_{ij}$ are the calculated distance frequencies.

The simulated distribution obtained in Figure 23 demonstrates a preponderance of small distances between pores. If a transverse variation were allowed to be added to the ridges in the analysis (simulating a two-dimensional process), it is surmised that many of the distance measurements between pores would increase, thereby changing the shape of the distribution.
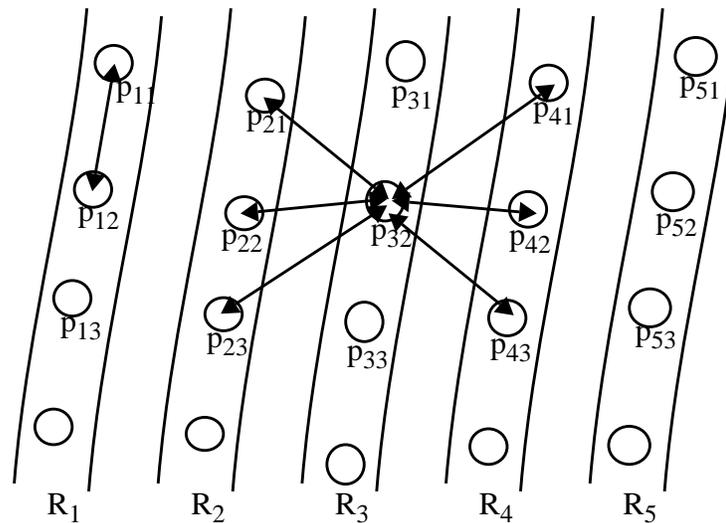
Figure 24. Spatial dependence. *Intra-ridge:* If the positions of intra-ridge pores were independent random variables, then their spacing would differ from the regularity observed in real fingerprints. *Inter-ridge:* The location of pore p_{32} would depend on the location of all the surrounding pores if pore positions were not independent random variables. p32 depends on the position of p31 and p33 but may also depend on p21, p22, p23 and p41,p42,p43.

Although most of the discussion thus far has concerned pore distribution along a given ridge, the discussion of dependence bids us to look at the spatial distribution of pores across ridges: In addition to an expected distance between nearest neighbor intra-ridge pores, there exists also an expected linear relationship between the inter-ridge pores. As shown in Figures 24 and 25, pores on adjacent ridges exhibit some degree of alignment. Inherent in this relationship is a linear deviation $\overline{\Delta}$ that measures the difference between the expected linear relationship and the actual pore pattern on the print. If pores were generated independently of each other, large $\overline{\Delta}$s would result, and no visible linear pattern would be evident.
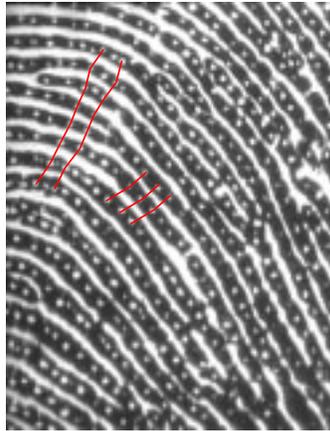
Figure 25. Dependence - pores on adjacent ridges demonstrate a high degree of alignment.

8.2 Degree of alignment of inter-ridge pores

Data from live-scanned fingerprints were examined to study alignment of adjacent pores across ridges. Frequency distributions of the alignment between nearest neighbor pores on three consecutive ridges and three alternating ridges were obtained (where ridges 2 and 4 are skipped). The degree of alignment may indicate a measure of dependence. The data are presented in Figure 26. Here, the plots for both three consecutive ridges and three alternating ridges are compared to a simulation in which pores are constrained to occur at random positions in sweat gland units. The model includes the possibility of empty sweat gland units.
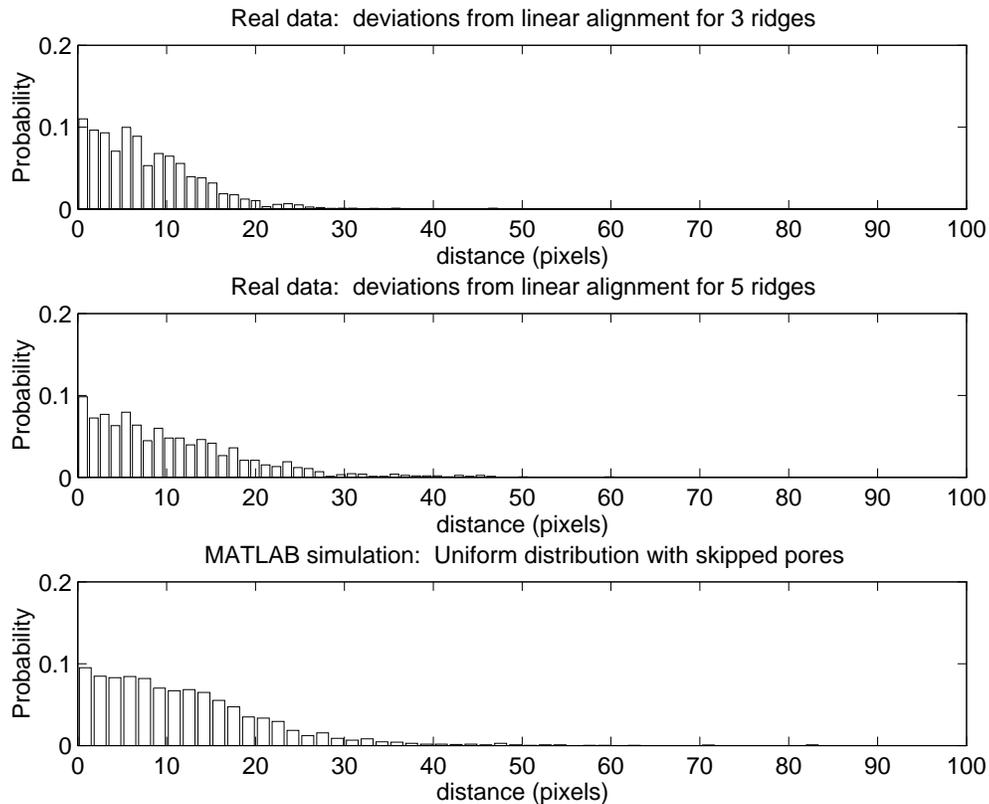
Figure 26. Frequency of degree of inter-ridge deviations from alignment. These plots demonstrate that most deviations from alignment are between 0 and 10 pixels (0.22 mm), evincing dependence.

All three plots show a maximum frequency of occurrence for deviations close to 0. As the degree of deviation increases, there is a marked decrease in probability of occurrence for all three plots. However, the plot of frequency of deviations for three alternating ridges follows the simulation more closely than the plot for three consecutive ridges. This can be attributed to a greater degree of alignment among pores on three consecutive ridges; the more proximal the ridges, the greater the degree of alignment between the pores across those ridges.

8.3 Dependence determination: Ridge-independent distributions

The effect of dependence among pores is evident also in the ridge-independent configuration studies discussed in section 5.4.3.1 and 5.4.3.2. Here, the probability of pore occurrence in a 5 pixel by 5 pixel area was found (the binomial distribution discussed in Section 5.4.1) as well as the pore density for 20 pixel by 20 pixel area (Figure 27).

Density of pores in a 20x20 pixel area – high resolution
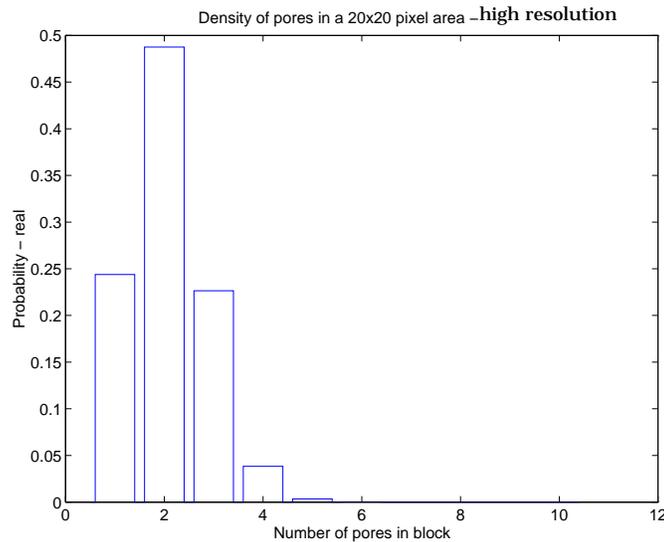
Figure 27. Measured density of pores in a 20 pixel by 20 pixel area. This distribution was obtained from actual data; the number of pores in a 20x20 pixel area was tallied and the frequencies of the individual densities was plotted.

If the occurrence of each pore were a phenomenon independent of other pores, the density for a larger area could be determined by using the data obtained from the binomial probability alone:

$$(P_{pore})^{\text{\# cells with pores}}(P_{no\ pore})^{\text{\# cells with no pores}} = (0.067)^{\text{\#cells with pores}}(0.933)^{\text{\#cells with no pores}}$$

However, when this extrapolation is attempted for a 20 by 20 pixel area, the resulting probability density function shows a higher incidence of empty cells than was detected in the actual density. This increase in empty cells is compensated by a reduction in the number of cells that contain one pore (Figure 28).

The independence assumption does not account for phenomena that determine the actual spatial distribution of pores; ridge flow is one such factor. Perhaps the ridge flow in a print constricts the spatial distribution so that there are fewer possible empty cells; this would account for the disparity between Figure 27 and Figure 28.

Interestingly enough, when the density of pores in a 40 by 40 pixel area is determined and compared to an extrapolation from the binomial distribution, the real density is similar to the extrapolation. It is proposed that the degree to which dependence affects pores is higher for smaller areas, much like it is greater in an inter-ridge capacity for more proximal ridges.
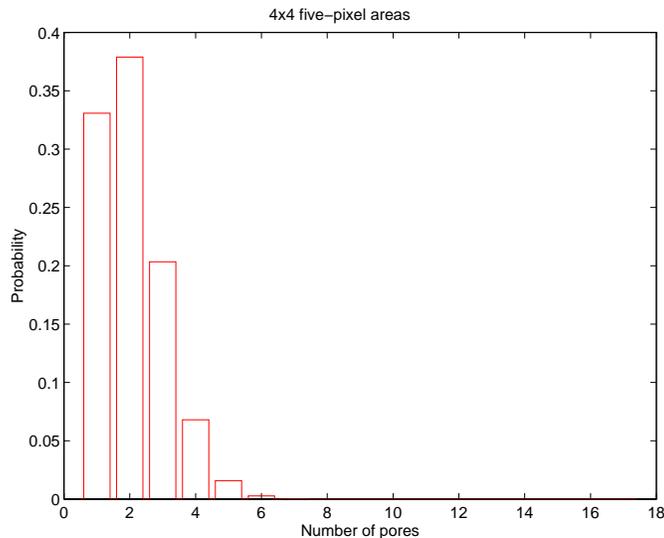
Figure 28. The binomial distribution discussed in Section 5.4.1 was extrapolated to produce this plot, which shows a theoretical frequency distribution of pore densities, assuming that pores occur independently of each other.

9. Conclusion

Recognition errors in automated fingerprint recognition systems, like other biometric systems, can be grouped into two classifications: false accepts and false rejects. When addressing the problem of a false accept, the question is: how does one differentiate an authorized user from an unauthorized user? The uniqueness analyses in this study provide answers to this question: Given a certain number of pores along a ridge or a number of pores in a constellation, the probability of someone else having an identical configuration is sufficiently low to preclude a false accept.

Still, false accepts persist. The value of the uniqueness of a configuration is reduced when using an automated system, for many parameters which were designed to decrease the number of false rejects actually increase the probability of a false accept. For example, automated matching requires accommodation of phenomena such as plasticity and distortion; therefore, parameters such as search areas are built in to allow a degree of flexibility in feature detection. Thus, the possibility exists that features in an impostor's print may be falsely detected as matching features in an enrolled user's print.

On the other side of the coin, however, is the question of false rejects. The problem that must be addressed in this case is: How does one recognize an authorized user as such, regardless of changes that have occurred since the enrollment procedure? These changes can include differences in location, orientation, shape or size of respective features due to distortion or

plasticity. Errors can occur as a result of the processing and feature detection stage or as a result of some features being physiologically unreliable.

Performance is defined by these error rates. In this paper, a model for an automated matching system was developed, which incorporates the parameters determining the error rates. From this model, performance of a generic automated fingerprint recognition system can be predicted.

A.1 Glossary

attribute - a sub-feature - the position, shape, and size are attributes of a pore.

authentication - confirmation of proper identity.

dependence - the ability of a feature to affect the attributes of another feature - cellular or biological dependence may produce a measured dependence of the pore positions.

dermis - the layer of skin directly beneath the epidermis; contains living elements such as sweat glands, nerves, and blood vessels.

distribution (probability) - the probability density function (pdf) of a random variable.

distribution (spatial) - the way that a set of points is positioned in space or in an image.

dpi - refers to scanning resolution measured in dots (pixels) per inch.

EER - Equal Error Rate: the value at which the FAR and FRR are equal.

epidermis - the outermost layer of skin; acts as a protective layer for the dermis.

FAR - False Accept (imposter) error Rate: fraction of attempts for which the system allows access to an imposter or invalid user.

FRR - False Reject (valid user) error Rate: fraction of attempts for which a fingerprint system denies access to a valid user.

feature: a characteristic; pores and minutia points are fingerprint features.

feature area - search area - The area assigned to an individual feature in which no other feature is assumed to exist. A small area of fingerprint surrounding the feature location in which the features exact position is not important. Related to resolution and search area.

feature characteristic (sub-feature) - attributes of features such as shape, size, and location for pores; type, orientation, and position for minutiae.

feature configuration - a feature set for which the specific arrangement of the features within the area occupied by the set is known.

feature density - the number of features per unit area [features/mm2].

feature position or location - defined as the center of mass of a pore or the center of the ridge at the point at which it ends (for end points) or branches (for bifurcations).

feature set: {a,b,c,...} a group of features associated with a specified area of fingerprint.

Galton feature - any of a set of 10 distinct fingerprint features which include minutiae (branches and ends) as well as special ridge structures such as ridge islands, dots, bridges, spurs, enclosures, double bifurcations, deltas, and trifurcations.

homogeneous: uniformly spatially distributed - the density of pores (or sweat glands) is constant over the entire area of print

identification: a scenario in which the identity of the user who presents a live-scan image to the system is unknown. The system must determine who the unknown user is from a database of valid users.

inked fingerprint - an image of the fingerprint resulting from applying ink to the surface of the finger and then rolling the finger on paper - results in a rolled fingerprint impression unlike a live-scan fingerprint.

inter-ridge pores - pores which are not on the same ridge.

intra-ridge pores - pores which are on the same ridge.

intra-ridge separation - measured value of separation between sequential pores on the same ridge.

intra-ridge separation (adjusted) - intra-ridge separation corrected for missing pores.

isotropic - having the same properties independent of direction or orientation.

latent - the fingerprint impression left on an object's surface resulting from contact with a finger.

live-scan - an image of the fingerprint acquired using an electronic scanner for the purpose of real-time fingerprint processing or matching. A live-scan represents a pressed finger as opposed to a rolled print.

match - of a feature: a feature represented in the enrolled template (or fingerprint) corresponds to a feature from the live-scan fingerprint.

mismatch - of a feature: a feature represented in the enrolled template (or fingerprint) which does not correspond to a feature from the live-scan fingerprint, or a feature from the live-scan print which does not correspond to a feature in the enrolled template.

minutia - a ridge structure which differs from the usual (normal) continuous and non-diverging flow, examples are ridge branches (bifurcations) and ridge ends.

MxN pixels - an area of fingerprint M pixels in width and N pixels in height.

measurement accuracy - the accuracy of determined locations of features; higher resolution

allows more precise estimation of an actual feature's position, shape, and size.

performance: measure of FRR and FAR for a given system; match time and cost should also be considered but are not addressed in this paper.

pixel - can be used as a unit of length or area; magnitude is established by the magnification of the input image and the dimensions of the image produced by the framegrabber.

pore - opening of a sweat gland which is visible on the surface of the finger ridges.

ppi - refers to scanning resolution measured in pixels per inch of fingerprint.

ppmm - refers to scanning resolution measured in pixels per millimeter of fingerprint.

regular distribution - uniform distribution of objects in space.

reliability - inherent - (key to FRR), the probability that a given feature (a pore for example) will be visible in different images of the same fingerprint.

reliability - algorithm: (key to FRR), the probability that the algorithm will correctly detect a visible (real) fingerprint feature.

resolution of database images - 1100 ppi where 1 pixel corresponds to 23.1 microns.

resolution - scanning - number of samples per unit length (or area), determined by the degree of magnification of the fingerprint image on the CCD sensor.

resolution - sensor - defined as the number of active pixels on the CCD imaging sensor in a video camera (sometimes quoted without regard to CCD dimensions).

scanned area - the area of fingerprint incident on the active area of the CCD sensor device and represented in the fingerprint image.

search area - a small area in which a feature is searched for; designed to account for detected feature position deviations due to noise, plasticity, distortion, or processing variations. Increasing the search area is equivalent to reducing the scanning resolution reducing the accuracy of detection of the feature position.

sub-feature - an attribute of a fingerprint feature.

subcutaneous layer - the layer of skin beneath the dermis; contains fat.

sweat gland - a structure within the dermis that produces sweat. Composed of a coil, which secretes the sweat, and a duct, which carries sweat to the surface. The duct opening on the skin surface comprises a pore.

template - a set of data which is extracted from a fingerprint and then used to represent that finger. Fingerprints are matched against templates or templates are matched against templates.

uniform distribution - defining the position of a feature as a random variable, a uniform pdf means that the feature has an equally likely probability of occurring anywhere in image (flat 2-D pdf in x,y). In one model of pore distribution, if there are N pores per fingerprint, then each pore's position is assumed to be a uniformly distributed random variable.

uniform distribution of objects - a homogeneous spatial distribution.

uniqueness - (key to FAR) probability of occurrence of a configuration of features.

verification: a scenario in which a user claims an identity (enters a PIN) and the system then authenticates the user's claim by matching his live-scan print against the template corresponding to his claimed identity.

A.2 Vital statistics of fingerprint features

Pore density
There are approximately 2700 to 3350 pores/in2 (4.19 to 5.19 pores/mm2)

Intra-ridge pore density
The average separation of pores on a ridge is 0.39 mm, ($\sigma = 0.207$mm).
There are 25.6 pores/cm of ridge on average.

Features
Placement:
    Pores are found on the ridge only.
    The pore's exact position on the ridge is random.
    The position is defined as the center of mass of the pore.

Size:
    Pores are generally less than 220 μm across.
    The average diameter of a pore is 109 μm assuming a circular pore shape.
    The fraction of print area occupied by pores is 3.9%.

Shape:
    The shape of each pore is unique.
    There is a lot of variation in the general shape of pores ranging from square to circular.

Pore reliability
Pore reliability is a function of the sub-feature type (position, shape, and size), capture method, and skin condition.

Ridge width
The average width of a ridge is 0.48 mm for males and shorter for females.

Galton feature density
The density of Galton features is approximately 0.234 features/mm^2.

Minutia (branch and end point) density
The density of minutia is approximately 0.241 minutiae/mm^2.

Ridge (Galton) feature placement
The placement of Galton features is random.

Ridge (Galton) feature reliability
The reliability of Galton features is relatively high and depends on the quality of the print, skin condition, and capture method. With respect to minutiae, sometimes it is difficult to distinguish between ridge bifurcations and ends.

A.3 Osterburg

Table 1: Osterburg probability of feature occurrence

Feature	Frequency	Probability of Occurrence
Empty cell	6584	0.766
End point	715	0.0832
Branch point	328	0.0382
Island	152	0.0177
Bridge	105	0.0122
Spur	64	0.00745
Dot	130	0.0151
Lake	55	0.00640
Trifurcation	5	0.000582
Double bifurcation	12	0.00140
Delta	17	0.00198
Broken ridge	119	0.0139
Other multiple occurrences	305	0.0355
Total	8591	1.00

A.4 Resolution and intra-ridge pore configurations

Table 2: Probabilistic values of intra-ridge pore configurations

# pore	r=1 pix	r=2 pix	r=3 pix	r=4 pix	r=5 pix	r=6 pix	r=7 pix
probability= (max probability at r)$^{\text{number of pores}}$							
2	0.0645	0.1246	0.1848	0.2444	0.2879	0.3450	0.3839
3	0.0042	0.0155	0.0341	0.0597	0.0829	0.1190	0.1474
4	2.6821e-04	0.0019	0.0063	0.0146	0.0239	0.0411	0.0566
5	1.7297e-05	2.4123e-04	0.0012	0.0036	0.0069	0.0142	0.0217
6	1.1155e-06	3.0064e-05	2.1531e-04	8.7115e-04	0.0020	0.0049	0.0083
7	7.1936e-08	3.7467e-06	3.9781e-05	2.1287e-04	5.6934e-04	0.0017	0.0032
8	4.6391e-09	4.6694e-07	7.3500e-06	5.2016e-05	1.6391e-04	5.8215e-04	0.0012
9	2.9918e-10	5.8192e-08	1.3580e-06	1.2710e-05	4.7187e-05	2.0086e-04	4.7223e-04
10	1.9294e-11	7.2523e-09	2.5091e-07	3.1058e-06	1.3585e-05	6.9304e-05	1.8131e-04
11	1.2443e-12	9.0382e-10	4.6358e-08	7.5891e-07	3.9110e-06	2.3912e-05	6.9614e-05
12	8.0242e-14	1.1264e-10	8.5652e-09	1.8544e-07	1.1259e-06	8.2505e-06	2.6728e-05
13	5.1748e-15	1.4038e-11	1.5825e-09	4.5314e-08	3.2415e-07	2.8467e-06	1.0262e-05
14	3.3372e-16	1.7495e-12	2.9239e-10	1.1073e-08	9.3319e-08	9.8221e-07	3.9401e-06
15	2.1522e-17	2.1803e-13	5.4022e-11	2.7056e-09	2.6866e-08	3.3890e-07	1.5128e-06
16	1.3879e-18	2.7172e-14	9.9813e-12	6.6113e-10	7.7344e-09	1.1693e-07	5.8082e-07
17	8.9507e-20	3.3863e-15	1.8442e-12	1.6155e-10	2.2267e-09	4.0345e-08	2.2301e-07
18	5.7723e-21	4.2203e-16	3.4073e-13	3.9475e-11	6.4103e-10	1.3920e-08	8.5622e-08
19	3.7225e-22	5.2595e-17	6.2954e-14	9.6459e-12	1.8455e-10	4.8030e-09	3.2874e-08
20	**2.4007e-23**	**6.5547e-18**	**1.1632e-14**	**2.3570e-12**	**5.3130e-11**	**1.6572e-09**	**1.2622e-08**
21	1.5482e-24	8.1689e-19	2.1491e-15	5.7594e-13	1.5296e-11	5.7179e-10	4.8461e-09
22	9.9841e-26	1.0181e-19	3.9706e-16	1.4073e-13	4.4035e-12	1.9729e-10	1.8606e-09
23	6.4388e-27	1.2688e-20	7.3362e-17	3.4389e-14	1.2677e-12	6.8071e-11	7.1439e-10
24	4.1523e-28	1.5812e-21	1.3555e-17	8.4030e-15	3.6497e-13	2.3487e-11	2.7429e-10
25	2.6778e-29	1.9706e-22	2.5044e-18	2.0533e-15	1.0507e-13	8.1037e-12	1.0531e-10
26	1.7269e-30	2.4559e-23	4.6271e-19	5.0174e-16	3.0249e-14	2.7961e-12	4.0434e-11
27	1.1137e-31	3.0606e-24	8.5492e-20	1.2260e-16	8.7084e-15	9.6474e-13	1.5524e-11
28	7.1822e-33	3.8143e-25	1.5796e-20	2.9958e-17	2.5071e-15	3.3287e-13	5.9605e-12
29	4.6318e-34	4.7537e-26	2.9184e-21	7.3203e-18	7.2176e-16	1.1485e-13	2.2885e-12
30	2.9870e-35	5.9243e-27	5.3921e-22	1.7888e-18	2.0779e-16	3.9627e-14	8.7866e-13

A.5 Ridge-independent pore configurations

By assuming that pores occur independently of each other, in a way similar to the treatment of minutiae, the probability of a configuration of pores can be calculated. For a given area of fingerprint, and a defined "cell size," there will be N cells or grid divisions in the print. In each cell, there can be either one pore or no pores present. For an area of print about 0.46 mm x 0.46 mm, the number of cells is 16. The number of configurations possible (using n pores), the probability of a given configuration of n pores, and the probability of n pores in the total analysis area are provided in Table 3.

Table 3: Calculated probability of n pores occurring in a 4x4 grid area

# Pores n	# Empty Cells m	# configurations combination (N n)	configuration probability $p^n q^m$	p(n pores in 20x20 pixel region) = $(N n) p^n q^m$
0	16	1.00000e+00	3.29690e-01	3.29690e-01
1	**15**	**1.60000e+01**	**2.36755e-02**	**3.78808e-01**
2	14	1.20000e+02	1.70017e-03	2.04021e-01
3	13	5.60000e+02	1.22092e-04	6.83713e-02
4	12	1.82000e+03	8.76756e-06	1.59570e-02
5	11	4.36800e+03	6.29611e-07	2.75014e-03
6	10	8.00800e+03	4.52132e-08	3.62067e-04
7	9	1.14400e+04	3.24682e-09	3.71436e-05
8	8	1.28700e+04	2.33159e-10	3.00075e-06
9	7	1.14400e+04	1.67434e-11	1.91545e-07
10	6	8.00800e+03	1.20237e-12	9.62857e-09
11	5	4.36800e+03	8.63438e-14	3.77150e-10
12	4	1.82000e+03	6.20046e-15	1.12848e-11
13	3	5.60000e+02	4.45264e-16	2.49348e-13
14	2	1.20000e+02	3.19750e-17	3.83700e-15
15	1	1.60000e+01	2.29617e-18	3.67387e-17
16	0	1.00000e+00	1.64891e-19	1.64891e-19

Each grid is of area 5x5 pixels. N = 16, total number of grids. Binomial distribution: P(pore in cell) = p = 0.067, P(cell empty) = q = 0.933.

For an area of print about 0.92 mm x 0.92 mm, the number of cells is 64. The number of configurations possible (using n pores), the probability of a given configuration of n pores, and the probability of n pores in the total analysis area are provided in Table 4. The entries in the tables are plotted in figure A.5.1 for a more insightful description of the underlying process of configuration probabilities.

Table 4: Calculated probability of n pores occurring in an 8x8 grid area

# Pores n	# Empty Cells m	# configurations combination (N n)	configuration probability $p^n q^m$	p(n pores in 64x64 pixel region) = (N n)$p^n q^m$
0	64	1.00000e+00	1.18148e-02	1.18148e-02
1	63	6.40000e+01	8.48435e-04	5.42998e-02
2	62	2.01600e+03	6.09272e-05	1.22829e-01
3	61	4.16640e+04	4.37527e-06	1.82291e-01
4	**60**	**6.35376e+05**	**3.14194e-07**	**1.99631e-01**
5	59	7.62451e+06	2.25627e-08	1.72030e-01
6	58	7.49744e+07	1.62026e-09	1.21478e-01
7	57	6.21216e+08	1.16353e-10	7.22803e-02
8	56	4.42617e+09	8.35546e-12	3.69827e-02
9	55	2.75406e+10	6.00017e-13	1.65248e-02
10	54	1.51473e+11	4.30880e-14	6.52668e-03
11	53	7.43596e+11	3.09421e-15	2.30084e-03
12	52	3.28421e+12	2.22200e-16	7.29751e-04
13	51	1.31369e+13	1.59526e-17	2.09618e-04
14	50	4.78557e+13	1.14585e-18	5.48357e-05
15	49	1.59519e+14	8.22854e-20	1.31261e-05
16	48	4.88527e+14	5.90902e-21	2.88672e-06
17	47	1.37937e+15	4.24335e-22	5.85315e-07
18	46	3.60169e+15	3.04721e-23	1.09751e-07
19	45	8.71988e+15	2.18824e-24	1.90812e-08
20	44	1.96197e+16	1.57141e-25	3.08306e-09
21	43	4.11080e+16	1.12845e-26	4.63882e-10
22	42	8.03474e+16	8.10354e-28	6.51099e-11
23	41	1.46721e+17	5.81926e-29	8.53810e-12
24	40	2.50649e+17	4.17889e-30	1.04744e-12
25	39	4.01039e+17	3.00092e-31	1.20348e-13
26	38	6.01558e+17	2.15500e-32	1.29636e-14
27	37	8.46637e+17	1.54754e-33	1.31020e-15
28	36	1.11877e+18	1.11131e-34	1.24330e-16
29	35	1.38882e+18	7.98044e-36	1.10834e-17
30	34	1.62029e+18	5.73086e-37	9.28565e-19
31	33	1.77709e+18	4.11541e-38	7.31345e-20
32	32	1.83262e+18	2.95533e-39	5.41601e-21
33	31	1.77709e+18	2.12226e-40	3.77145e-22
34	30	1.62029e+18	1.52403e-41	2.46936e-23
35	29	1.38882e+18	1.09442e-42	1.51996e-24
36	28	1.11877e+18	7.85921e-44	8.79265e-26
37	27	8.46637e+17	5.64381e-45	4.77825e-27
38	26	6.01558e+17	4.05289e-46	2.43805e-28
39	25	4.01039e+17	2.91044e-47	1.16720e-29
40	24	2.50649e+17	2.09003e-48	5.23863e-31
41	23	1.46721e+17	1.50088e-49	2.20211e-32
42	22	8.03474e+16	1.07780e-50	8.65984e-34
43	21	4.11080e+16	7.73982e-52	3.18169e-35
44	20	1.96197e+16	5.55807e-53	1.09048e-36
45	19	8.71988e+15	3.99133e-54	3.48039e-38
46	18	3.60169e+15	2.86623e-55	1.03233e-39
47	17	1.37937e+15	2.05828e-56	2.83912e-41
48	16	4.88527e+14	1.47808e-57	7.22080e-43
49	15	1.59519e+14	1.06143e-58	1.69318e-44
50	14	4.78557e+13	7.62225e-60	3.64768e-46
51	13	1.31369e+13	5.47364e-61	7.19064e-48
52	12	3.28421e+12	3.93070e-62	1.29092e-49
53	11	7.43596e+11	2.82269e-63	2.09894e-51
54	10	1.51473e+11	2.02701e-64	3.07038e-53
55	9	2.75406e+10	1.45562e-65	4.00887e-55
56	8	4.42617e+09	1.04530e-66	4.62668e-57
57	7	6.21216e+08	7.50646e-68	4.66314e-59
58	6	7.49744e+07	5.39049e-69	4.04149e-61
59	5	7.62451e+06	3.87099e-70	2.95144e-63
60	4	6.35376e+05	2.77981e-71	1.76622e-65
61	3	4.16640e+04	1.99622e-72	8.31704e-68
62	2	2.01600e+03	1.43351e-73	2.88996e-70
63	1	6.40000e+01	1.02942e-74	6.58831e-73
64	0	1.00000e+00	7.39243e-76	7.39243e-76

Each grid is of 5x5 pixels. N = 64, total number of grids. Binomial distribution: P(pore in cell) = p = 0.067, P(cell empty) = q = 0.933.
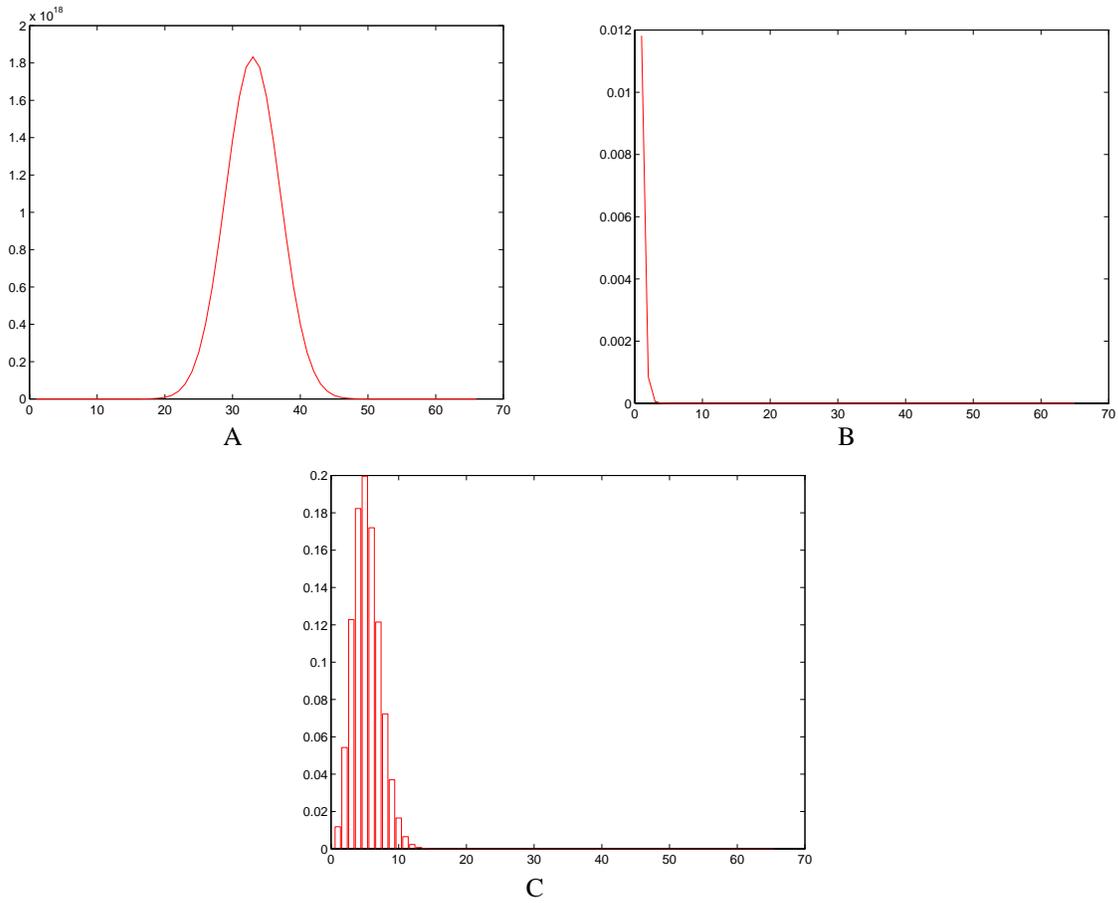
Figure A.5.1. Figures relate to a 40x40 pixel area divided into 64 grids of area 5x5 pixels each. For this situation, the probability that a grid contains one pore is 0.067 and the probability that the grid is empty is 0.933. Figure A is a plot of the number of possible configurations of pores as a function of the total number of pores. Figure B is a plot of the probability of a particular configuration of n pores. Figure C is a plot of A multiplied by B, which is the probability of a given number of pores, n, occurring inside the 40x40 pixel area (binomial distribution).

A.6 Size of Pores

The size of an individual pore may vary from one scan to the next leading to a relatively unreliable pore sub-feature. In addition, there is a large variation in the size of pores between individuals. In some cases, the range of pore sizes for different prints of the same finger will vary significantly. Figure A.6.1.A shows the general trend of a decreasing number of pores as pore size increases. The caption in Figure A.6.1 provides details of the nature of pore size.
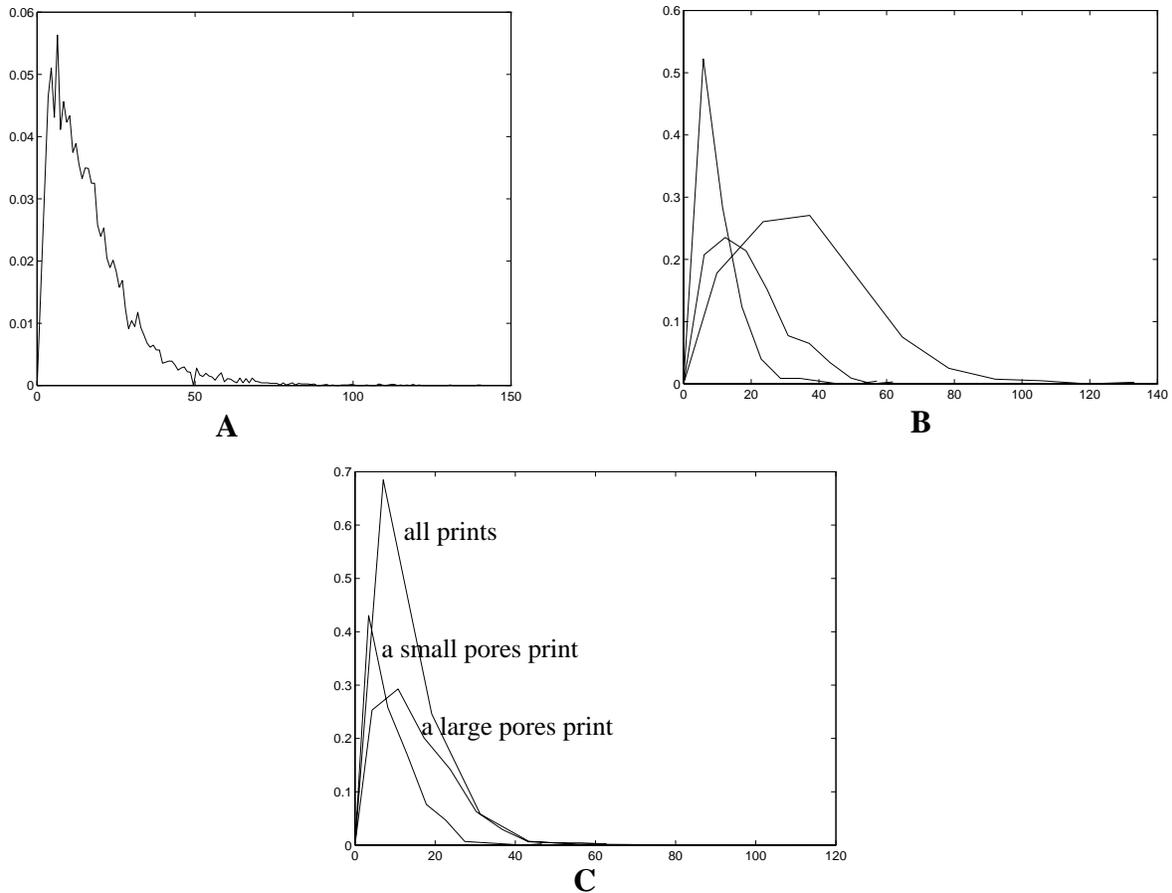

Figure A.6.1. Distribution of the area of pores (measured in units of pixels, where the resolution was 1100 pixels/inch and 1 pixel has dimensions 23.1μm x 23.1 μm). For plots A and B no pores smaller than 3 pixels in area were allowed in order to reduce noise effects. Pores were extracted and measured automatically. Figure A represents 40 images (13,197 pores with $\mu = 17.4$, $\sigma = 13.6$, max = 141). Figure B shows the variation in pore area between three different people's fingers: small pores ($\mu = 10.2$, $\sigma = 7.3$), medium pores ($\mu = 18.8$, $\sigma = 11.0$), and large pores ($\mu = 35.4$, $\sigma = 19.6$). Figure C demonstrates variation in the distribution of pore sizes for different live-scan images of the same finger (reliability of pore size): plot of size for 10 images together ($\mu = 11.0$, $\sigma = 9.2$), live-scan when pores are relatively large ($\mu = 14.8$, $\sigma = 10.2$), distribution when pores are relatively small ($\mu = 8.4$, $\sigma = 6.7$).

A.7 Tables and images of reliability of pores

Table 5: Pore reliability for various parameters

Print Type	Detection	Size	Shape
latent	Variable - depends on surface on which latent is found and method of latent extraction	Variable	Variable
inked	Good - pores can fill in if too much ink is used	Best - Inked impression should provide accurate representation of size of pore	Variable - bleeding of ink can cause problems
live-scan	Best - pores can fill in if the finger is oily or wet. Pores may not show up for very dry fingers.	Variable - what is seen in the live-scan image depends on the condition of the finger (oily, dry, neutral)	Variable - inherent variability of pore shape - again mostly due to amount of moisture on finger

Table 6: Observed inherent reliability of pores

Print	Vis. [0,1]	Size: Absolute Range [0,1]			Size: Relative [0,1]			Shape [0,1]	Clarity [0,1]			Quality [1,5] [0,1]		Density [1,5]
		Small	Med.	Large	Small	Med.	Large		Clear	Shad.	Fuzzy	Image	Print	
W52	.93	.48	.52	0	.82	.18	0	.85	.56	.33	.11	4.24	.0024	3.91
L253	.98	.70	.30	0	.62	.38	0	.94	.96	.02	.02	4.09	.0493	3.59
W254	.75	.62	.38	0	.62	.36	.2	.91	.71	.17	.12	3.48	.195	2.84
L92	.94	.13	.83	.04	.28	.71	0	-	.57	.37	.07	3.23	.186	3.29
L19	.95	.08	.60	.33	.17	.71	.12	-	.37	.46	.17	2.69	.128	2.83
A26	.79	.175	.65	.175	.175	.635	.19	-	.30	.40	.30	2.58	.245	2.39
L133	1.00	.14	.86	0	.16	.84	0	.95	.74	.21	.05	3.79	.035	3.74
L255	.85	.73	.27	0	.79	.18	0	.85	.79	.18	.03	3.82	.0518	3.44
L112	.91	.65	.35	0	.65	.35	0	1.00	.71	.23	.06	3.59	.0874	2.79
W259	1.00	.83	.17	0	.83	.17	0	1.00	1.00	0	0	3.83	.0372	3.72

Visibility is the fraction of instances in which a given pore was detectable by eye. Absolute size is the observed size of a pore. Relative size is the size of the pore in relation to all other pores in that image. Shape - fraction of occurrences for which the pore's shape is consistent. For clarity, clear means that the pore is in focus and high contrast; fuzzy denotes an out of focus pore. Shadowy describes pore's which were mostly clear but parts were fuzzy. Image quality (range [1,5] where 5

is high quality) is determined by the amount of noise and degradation in the image, whereas print quality (range [-1,1], wet-neutral-dry) is a measure of the condition of the skin. Density is the relative number of pores per image.

Table 7: Algorithm reliability

Quality	# pores	Missed Detect						False Detect				
		Proc.	Flow	NN	nonisol.	flow-nonisol	2 as 1	mid-valley	unknown	bridge	noise	washed out
Excel.	486	5.8	.62	2.3	1.6	1.0	.21	.41	.21	.21	.21	.21
Excel.	407	4.1	0	.25	.41	.41	.25	.98	0	0	.98	0
good	504	.4	4.2	.2	2.2	.79	0	0	.2	0	.4	.21
Good	427	.47	4.0	.47	2.8	.47	0	0	0	0	0	.47
OK	286	5.9	3.5	0	2.5	.35	0	0	0	0	.7	0
OK	489	5.5	3.1	.61	1.2	.2	0	.2	0	0	0	0
Poor	352	3.4	4.3	.85	1.4	0	0	1.1	0	0	1.1	0
Poor	411	14.1	4.4	0	4.1	2.4	0	0	0	0	.49	0

The percentage of algorithm detection errors attributable to various causes in eight prints of varying quality. The total error rate (missed detects plus false detects) is the sum of each row. Quality is the image quality. # pores is the number of true pores detected in the image accounting for missed and false detected pores (errors). Processing represents the percentage of pores erroneously discarded during the process of converting the image from gray level to binary. Flow represents the percentage of pores which are so large that they appear as a break in the normal flow of the ridge. NN is the percentage of close but separated pores for which only one pore is detected. Non-isolated is the percentage of pores on the edge of a ridge which are not detected. Flow non-isolated is the percentage of missed detected pores which occur on the center of a ridge but appear to touch the edge of the ridge because of noise. 2 as 1 represents the number of pairs of connected pores incorrectly detected as a single pore. Mid-valley is the percentage of instances in which the algorithm falsely detected a pore in the middle of a valley. Bridge is the percentage of falsely detected pores occurring at breaks in a ridge. Noise represents the percentage of instances in which noise was classified as a pore. Washed out represents a different noise process which causes low contrast areas in the image.

| a | b | c |
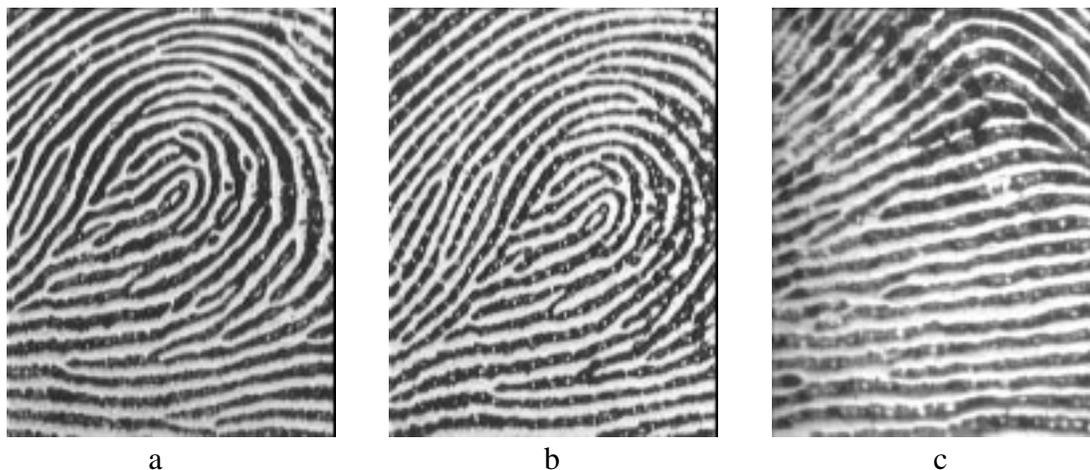
Figure A.7.1. Reliability of pores. The fingerprints in figure A.7.1, have a very low detectable pore density. For these prints, the use of poroscopy may not be reliable. **a** and **b** are images of the same finger showing a large disparity in pore density. Figure **c** is a print which also has a very low pore density, although the algorithm was able to find at least 84 pores.

A.8 References

Ashbaugh, Cpl. D.R., "Poroscopy," RCMP Gazette, Vol. 45, pp. 12-17, 1983.

Blue, J.L., G.T. Candela, P.J. Grother, R. Chellappa, and C. L. Wilson, "Evaluation of Pattern Classifiers for Fingerprint and OCR Applications," Pattern Recognition, Vol. 27, No. 4, pp. 485-501, 1994.

Cummins, H. and C. Midlo, Finger Prints, Palms and Soles, 2nd ed. New York: Dover Publications, Inc., 1943.

Hirsch, W. and J.U. Schweichel, "Morphological Evidence Concerning the Problem of Skin Ridge Formation," Journal of Mental Deficiency Research, Vol. 17, pp. 58-72, 1973.

Hrechak, A.K. and J.A. McHugh, "Automated Fingerprint Recognition Using Structural Matching," Pattern Recognition, Vol. 23, No. 8, pp. 893-904, 1990.

Lapin, Lawrence L, Probability and Statistics for Modern Engineering, 2nd ed. Boston: PWS-KENT Publishing Company, 1990, Chapters 5 and 15, pp. 220, 631-655.

Locard, E., "Les Pores et L'Identification des Criminels," Biologica, revue scientifique de medecin, Vol. 22, pp. 357-362, 1912.

Okajima, M., "Development of dermal ridges in the fetus," Journal of Medical Genetics, Vol. 12, pp. 243-250, 1975.

Osterburg, J.W., 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, December, 1977.

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

Sclove, S.L., "The Occurrence of Fingerprint Characteristics as a Two-Dimensional Poisson Process," Commun. Statist. - Theor. Meth., Vol. A9, No. 7, pp. 675-695.

Stoney, D.A. and J.I. Thornton, "A Method for the Description of Minutia Pairs in Epidermal Ridge Patterns," Journal of Forensic Sciences, JFSCA, Vol. 31, No. 4, pp. 1217-1234, October, 1986.

Stosz, J.D. and L.A. Alyea, "Automated system for fingerprint authentication using pores and ridge structure," Proceedings of the SPIE, Automatic Systems for the Identification and Inspection of Humans. Volume 2277, pp. 210-223, July, 1994.

Stosz, J.D. and L.A. Alyea, "Fingerprint Authentication," Proceedings of CardTech SecurTech

'95, pp. 201, 219. April 1995, Washington, DC.

Webster, J.G., ed., Medical Instrumentation, 2nd ed. Boston: Houghton Mifflin Company, 1992, Chapter 5, pp. 249-250.

A.9 Suggested Reading

Andersen, L.B. and K. Thestrup-Pedersen, "Sweat pore density on the fingertips of atopic patients," British Journal of Dermatology, Vol. 117, pp. 225-230, February, 1987.

Ashbaugh, Cpl. D.R., "Ridgeology," RCMP Gazette, Vol. 44, No. 5, 1982.

Bartlett, M.S., The Statistical Analysis of Spatial Pattern, 1st ed. London: Chapman and Hall, Ltd., 1975, Chapters 1, 3, pp. 9-12, 59-63.

Besag, J., "Spatial Interaction and the Statistical Analysis of Lattice Systems," Journal of the Royal Statistical Society, Ser.B, 36, pp. 192-236, 1974.

Bhattacharya, P.K., "Order of Dependence in a Stationary Normally Distributed Two-Way Series," The Annals of Mathematical Statistics, Vol. 43, No. 6, pp. 1792-1807, 1972.

Clegg, D.L.J., "Poroscopy in Practice," Journal of Forensic Identification, Vol. 44, No. 1, pp. 15-21, 1994.

Cullison, A.D., "Identification by Probabilities and Trial by Arithmetic (A Lesson for Beginners in How to be Wrong with Greater Precision)," Houston Law Review, Vol. 6, pp. 471-518, 1969.

Froelich, J.W. and E. Giles, "A Multivariate Approach to Fingerprint Variation in Papua New Guinea: Implications for Prehistory," American Journal of Physical Anthropology, Vol. 54, pp. 73-91, 1981.

Ibamoto, K., "Prediction Chart of Thermal Sensation Levels Derived from a Sweat Pore System Model," J. Human Ergol., Vol. 6, pp. 153-158, 1977.

Liu, J.H., C.H. Lin, J.W. Osterburg, J.D. Nichol, "Fingerprint Comparison II: On the Development of a Single Fingerprint Filing and Searching System," Journal of Forensic Sciences, JFSCA, Vol. 27, No. 2, pp. 305-317, April, 1982.

Moayer, B. and K.S. Fu, "An Application of Stochastic Languages to Fingerprint Pattern Recognition," Pattern Recognition, Vol. 8, pp. 173-179, 1976.

Rao, C.V.K., "On Fingerprint Pattern Recognition," Pattern Recognition, Vol. 10, pp. 15-18, 1978.

Rao, K., and K. Balck, "Type Classification of Fingerprints: A Syntactic Approach", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-2, No. 3, pp. 223-231, May, 1980.

Reed, B., "Automated Fingerprint Identification: From Will West to Minnesota Nine-Fingers and Beyond," Journal of Police Science Administration, Vol. 9, pp. 317-326, 1981.

Sampson, A.R. and R.L. Smith, "An Information Theory Model for the Evaluation of Circumstan-

tial Evidence," IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-15, No. 1, pp. 9-16, January-February, 1985.

Sparrow, M.K. and P.J. Sparrow, "A Topological Approach to the Matching of Single Fingerprints: Developing of Algorithms for Use on Latent Fingerprints," NBS Special Publication 500-126, pp. 1-67, October, 1985.

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

Stoney, D.A. and J.I. Thornton, "A Systematic Study of Epidermal Ridge Minutiae," Journal of Forensic Sciences, JFSCA, Vol. 32, No. 5, pp. 1182-1203, 1987.

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

Trauring, M., "Automatic Comparison of Finger-Ridge Patterns," Nature, Vol. 197, pp. 938-940, March, 1963.