A STATISTICAL STUDY ON THE FINGERPRINT MINUTIAE DISTRIBUTION

Jiansheng Chen, Yiu-Sang Moon

Department of Computer Science and Engineering
The Chinese University of Hong Kong
{jschen, ysmoon}@cse.cuhk.edu.hk

ABSTRACT

Fingerprint minutiae distribution is the key issue of fingerprint individuality study. A method for studying fingerprint minutiae distribution by analyzing their second order statistical properties is proposed in this paper. Experiments have been performed on 467 different fingerprints selected from three major fingerprint databases. Results show that fingerprint minutiae tend to overdisperse on a small scale; and cluster on a large scale. Our findings which have successfully explained and unified various previous research observations should enlighten the study of fingerprint minutiae pattern modeling, an important foundation for boosting improvement in the fingerprint authentication technology.

1. INTRODUCTION

Fingerprint verification is the most commonly used biometric authentication technology today. Its wide social acceptance comes from the common belief on the universality, stability and uniqueness of human fingerprints, among which uniqueness, or individuality, is the key to the discriminative power of fingerprints. While fingerprint universality and stability can be confirmed by empirical anatomic observations, the individuality of fingerprints requires more deliberate theoretical analysis.

The study of fingerprint individuality can be traced back to more than 100 years ago. F. Galton was among the first to use certain fingerprint ridge line patterns, or minutiae, to study the individuality of fingerprints. In [1], Galton claimed that a possibility of occurrence of a fingerprint configuration is 1.45×10^{-11}. From then on, most fingerprint individuality studies have focused on minutiae based representations [2, 3, 4]. Although different models have been proposed in these studies to describe fingerprint configuration and evaluate the fingerprint individuality quantitatively, the basic problem of minutiae based fingerprint individuality study for finding a good description of the spatial distribution of fingerprint minutiae still remains. Previous researches on this problem, mainly performed before 1990's, all revealed that fingerprint minutiae locations, when considered as two dimensional spatial point patterns, are NOT uniformly distributed [5, 6]. However, different opinions have been raised on how fingerprint minutiae patterns deviate from the uniform distribution. In [5], S. L. Sclove addressed this problem by considering fingerprints in terms of grids of 1-mm cells. The cells were categorized into different types according to the existence of minutia in a cell; or what kind of minutia the cell contains. Sclove showed that minutiae tend to cluster by analyzing the dependence among different cells of the grids. D. A. Stoney studied this problem in a different way [6]. For each of 412 thumbprints, Stoney chose one centrally located focal minutia. The ridge distances between these focal minutiae and their immediate neighbor minutiae were extracted. After studying the distribution of these distance values, Stoney found that fingerprint minutiae follow a slightly overdispersed uniform distribution; or, in fact, shows a slight tendency towards a regular distribution.

The purpose of our work here is to unify these two seemingly opposite findings by investigating the second-order statistical properties of fingerprint minutiae patterns. To ensure the generality of our work, the fingerprints for studying were selected from three different fingerprint databases collected in distinct regions and from various populations. This paper is organized as follows. Section 2 introduces the approach for analyzing spatial point patterns using their second order statistical properties. Experiments on statistical analysis of the minutiae patterns from selected fingerprints are reported and discussed in Section 3. The last section is a conclusion of our work.

2. SPATIAL POINT PATTERN ANALYSIS

For a fingerprint, its minutiae form a two dimensional point pattern after extraction. A partial fingerprint and its minutiae pattern are shown in Figure 1. Generally speaking, each minutia has three major properties: location, direction and type. In this work, we will only concentrate on the statistical analysis of fingerprint minutiae locations.

To analyze a spatial point pattern, a very natural starting point is to perform the test of Complete Spatial Randomness (CSR). The hypothesis of CSR asserts: (i) the number of points in any planar region A with area $|A|$ follows a Poisson distribution with mean $\lambda|A|$; (ii) given n points x_i in a region A, the x_i are independent random samples from the uniform distribution on A [7]. CSR hypothesis (i) and (ii) are actually self-consistent. Many statistical properties have been proposed for the CSR test: inter-event distances; nearest neighbor distances; first order properties such as intensity function and second order properties such as K function [7].

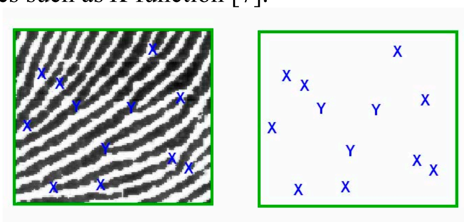

Figure 1. A partial fingerprint and its minutiae pattern (X-termination, Y-bifurcation)

In our work, we choose the K function for the CSR test of fingerprint minutiae patterns due to two main reasons. First, compared to the first order properties such as intensity function, the K function is more suitable for small samples as it is more related to the probability density function of the distances between pairs of points [7]. A fingerprint minutiae pattern is a pretty small sample for statistical analysis considering the number of minutiae for one fingerprint is usually smaller than 100. Second, the K function is invariant under random thinning. If each point of a given pattern is retained or not according to a series of mutually independent Bernoulli trials, the K function of the result thinned point pattern is identical to that of the original unthinned point pattern [7]. In our work, minutiae are marked manually by human beings. The case of missing any minutia can be considered an independent event with constant probability and thus will not affect the K function.

For a stationary isotropic spatial point process, its K function is defined as $K(t)=\lambda^{-1}E[N(t)]$, where λ is the expectation of the point density and $E[N(t)]$ is the expectation of the number of further points within distance t of an arbitrary point. Under the hypothesis of CSR, we can get $K(t)=\pi t^2$. For a given spatial point pattern containing n points on a planar region A with the area $|A|$, an unbiased estimator of $K(t)$ was given by Ripley in [8] as

$$\hat{K}(t)=\{n(n-1)\}^{-1}|A|\sum_{i=1}^{n}\sum_{j\neq i}\omega(x_i,u_{ij})^{-1}I(u_{ij}\leq t),$$

where u_{ij} is the distance between points x_i and x_j; $I(\bullet)$ denotes the indicator function; $\omega(x, u)$ was introduced by Ripley to eliminate the possible negative bias caused by boundary effects. It is defined as the proportion of the circumference of the circle with center x and radius u which lies within A. The explicit formula for $\omega(x, u)$ can be deduced if A is rectangular [7].

For a given spatial point pattern, $D(t)=\hat{K}(t)-\pi t^2$ can be used to evaluate its compatibility with the CSR assumption [7]. The sampling distribution of $\hat{K}(t)$ under the CSR assumption is analytically intractable. However, when A is a rectangle, the variance of $\hat{K}(t)$ can be explicitly expressed [9] as,

$$v_{LS}(t)=\{n(n-1)\}^{-1}|A|^2\{2b(t)-a_1(t)+(n-2)a_2(t)\}$$

and

$$b(t)=\pi t^2|A|^{-1}(1-\pi t^2/|A|)+|A|^{-2}(1.0716Pt^3+2.2375t^4)$$
$$a_1(t)=|A|^{-2}(0.21Pt^3+1.3t^4)$$
$$a_2(t)=|A|^{-3}(0.24Pt^5+2.62t^6),$$

where P denotes the perimeter of A. All the above four equations are exact when t is smaller than or equal to a quarter of the length of the shorter side of A [9]. As suggested in [7], $\pm 2\sqrt{v_{LS}(t)}$ can be used as the upper/lower limits for $D(t)$. If $D(t)$ lies within these limits for all the valid values of t, then the spatial point pattern under investigation can be regarded as compatible to the CSR assumption; otherwise, a deviation from CSR is suggested. Diggle suggested drawing a D-curve ($D(t)$ and $\pm 2\sqrt{v_{LS}(t)}$ against t) to visualize the CSR test result [7].

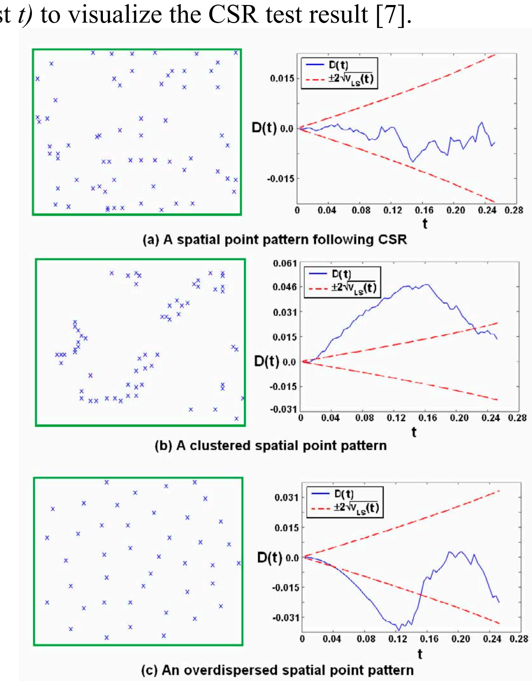

Figure 2. Typical point patterns and their D-curves

D-curves for three typical spatial point patterns (not fingerprint minutiae patterns) extracted from [10] are shown in Figure 2. Without losing generality, the units were chosen to make the patterns as unit squares. In Figure 2(a), the CSR assumption is supported. The D-curves in Figure 2(b) and 2(c) both suggest obvious deviation from the CSR assumption but in opposite directions. This can be explained by investigating the physical meaning of $\hat{K}(t)$. By definition, $\hat{K}(t)$ is essentially an average of point counts in circles of radius t. If the point pattern under investigation tends to cluster for certain values of t, the point counts in the circles will become much higher than the expectation under the CSR assumption because it is very probable that a large number of points aggregate 'into' the circles. However, if the point pattern has a tendency to overdisperse, the point counts in the circles will be essentially lower then expectation because t may not be big enough for the circles to 'reach' enough number of points. We can also observe that in both Figure 2(b) and 2(c), when t is big enough, the value of $D(t)$ drops back within the limits. This is because the tendency of clustering/overdispersing will be gradually 'averaged out' when t is getting bigger. Actually, t can be considered as the 'scale' of the CSR test. For a certain range of t, or a certain 'scale' of the test, the relation between $D(t)$ and the upper/lower limits can be used to describe the distribution tendency of the spatial point pattern on this 'scale'. If $D(t)$ is smaller than the lower bound, the pattern tends to overdisperse; or if $D(t)$ is bigger than the upper bound, the pattern tends to cluster; otherwise, the CSR assumption becomes applicable.

3. EXPERIMENTS AND DISCUSSIONS

We applied the second order property analysis method discussed in Section 3 to a set of fingerprint minutiae patterns. The fingerprint set contains 467 fingerprints selected from three different fingerprint databases: 134 fingerprints from NIST4 database (512×512; ~500dpi); 94 fingerprints from FVC2002 DB1 (388×374; 500dpi); and 239 fingerprints from FP383, a fingerprint database collected in our own laboratory (256×256; 450dpi) [11].

Three criteria were followed during the fingerprint selection process. First, all the fingerprints selected were from different finger tips. Second, only fingerprints with relatively high image quality were selected. Since the minutiae of the selected fingerprints had to be manually marked, high image quality would increase the reliability of the marking results. Third, only fingerprints with a big enough ROI (region of interest) were selected. This is to ensure that for each fingerprint minutiae pattern; there would be enough number of minutiae (sample size) for the statistical analysis. One remarkable characteristic of this fingerprint set is that it contains fingerprints collected in different regions around the world and from various populations. This ensures that our conclusion made based on experimental results on this fingerprint set is universal.

All the fingerprints were first normalized to 500 dpi. Then the minutiae of each fingerprint in the set were carefully marked and double checked. Our marking strategy for the 5 most common minutiae types [12] other than termination and bifurcation is shown in the following table.

Lake	Ind. Ridge	Island	Spur	Crossover
ignore	two endings	ignore	ending bifurcation	two bifurcations

A rectangular area was selected inside each fingerprint. The rectangles were selected so that all areas inside are ROI (because only effective fingerprint areas should be considered) and as many as possible minutiae are included. In this way, bounding rectangles for the 467 minutiae point patterns were obtained. For each rectangular minutiae point pattern, a D-curve was generated with the value of t ranging from 1 pixel to 1/4 of the shorter side length of the rectangle. A sample minutiae pattern (from NIST4) and its D-curve are shown in Figure 3. To make it more illustrative, 'pixel' and 'millimeter' are used as the unit of t respectively. The D-curve in Figure 3 suggests a tendency to overdisperse for small t values and a tendency to cluster for big t values. This can be directly observed from the minutiae pattern. The local patterns on the left/right bottom corner more resemble regular distributions. Global examination reveals that the minutiae are somewhat clustering near the edges while the central part of the minutiae pattern is relatively more sparse with fewer minutiae located.

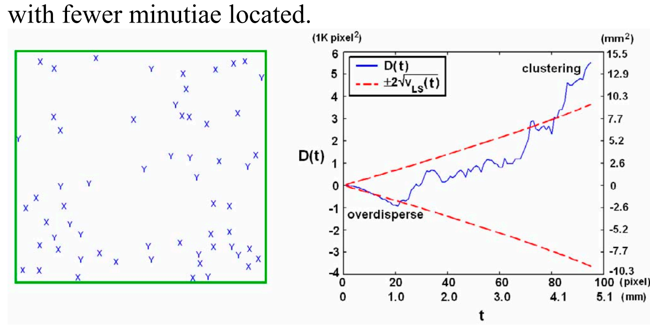

Figure 3. A sample minutiae pattern and its D-curve (X-termination, Y-bifurcation)

Table 1 lists the number of minutiae patterns whose D-curves strictly fall into the upper/lower limits for all the valid t values. Table 1 indicates that CSR is not an accurate model for describing fingerprint minutiae patterns. Identical assertion has also been made in [4, 5, 6].

Table 1. Number of CSR compatible minutiae patterns

Database	Total number of patterns	Number of CSR patterns	Ratio
NIST4	134	20	15%
FVC2002	94	43	46%
FP383	239	145	61%

To further study the distribution tendency of fingerprint minutiae patterns under different CSR test scales, we investigated the value of $D(t)$ for different ranges of t. We divided the values of t into sections with equal length of 10 pixels. For each section, the distribution tendency were evaluated as uniform, overdispersed or clustered according to the relation between $D(t)$ and the upper/lower limits. Table 2 shows the percentage of minutiae patterns which tend to overdisperse and cluster. In Table 2, 'O' stands for overdispersing and 'C' stands for clustering.

Similar trends can be observed from Table 2 for minutiae patterns from all the three databases. When t is relatively small (≤30pixels), an obvious tendency of overdispersing can be observed. This observation matches the finding made by Stoney in [6]. Since Stoney focused on studying the distances between focal minutiae and their immediate neighbors, suggesting that he made his tests in a relatively small scale. Stoney also explained this tendency using the growth stress model for minutia formation: 'minutia formation may alleviate local growth stress, thereby removing the impetus for formation of additional minutiae in the immediately surrounding region' [6]. More specifically, the tendency to overdisperse becomes prominent when the value of *t* is between 11 pixels and 20 pixels; converting to about 0.55mm to 1.0 mm for 500 dpi images. In [5], Sclove only studied the correlation among cells of 1mm; so it is not surprising that overdispersing was not demonstrated in his results since most overdispersing phenomena would simply happen inside the cells. It can also be noticed that the overdispersing in NIST4 is much more serious than the other two databases. This is because NIST4, unlike the other two databases, was created by scanning inked fingerprints. The ink technique requires the users to roll their fingers against the media with heavy pressure. The finger tip deformation thus caused would inevitably increase the inter distances between minutiae, or in other words, disperse the minutiae.

When *t* is relatively big (>50pixels), a tendency to clustering emerges as stated by Sclove in [5]. Actually, fingerprint minutiae tend to cluster around points where the ridge directions change abruptly, such as near the core point and delta point. A sudden change in the ridge direction may aggravate the local growth stress and force the change of ridge density, thereby increase the probability of the emergence of new minutiae.

Table 2. Minutiae distribution tendency on different scales

t (pixel)	NIST4		FVC2002		FP383	
	O (%)	**C (%)**	**O (%)**	**C (%)**	**O (%)**	**C (%)**
[1,10]	30	0	10	0	8	0
[11,20]	**74**	**0**	**31**	**2**	**20**	**1**
[21,30]	37	0	16	3	9	3
[31,40]	15	8	2	9	2	9
[41,50]	4	14	2	17	0	12
[51,60]	2	22	0	20	0	16
[61,70]	1	24	--	--	--	--
[71,80]	0	31	--	--	--	--
[81,90]	0	34	--	--	--	--
[91,100]	0	41	--	--	--	--

4. CONCLUSION

In this paper, the fingerprint minutiae distribution problem was studied by analyzing their second order statistical properties $K(t)$ (or $D(t)$). Experimental results show that uniform distribution (or CSR) is not an accurate model for describing spatial minutiae patterns. Further study on the minutiae distribution tendency on different test scales reveals that, when observed on a relatively small scale (small *t*), fingerprint minutiae tends to overdisperse; while clustering dominate when observed on a large scale (large *t*). The CSR assumption, as having been adopted in [4], seems to be approximately correct for describing fingerprint minutiae patterns only on the middle scale observation.

In conclusion, our work successfully explains and unifies different previous findings on fingerprint minutiae distribution study [5, 6]. We believe that a more accurate way for modeling fingerprint minutiae pattern is to employ a thinned process coupled with a simple inhibitory process; since such a thinned process displays small-scale regularity together with large-scale aggregation [7].

This work was partially supported by the Hong Kong Research Grants Council CERG Project 2150449, "Palmprint authentication using Time Series".

5. REFERENCES

[1] F. Galton, Finger Prints, London: McMillan, 1892

[2] M. Trauring, Automatic Comparison of Finger Ridge Patterns, Nature, pp. 938-940, 1963

[3] D. A. Stoney, J. I. Thornton, A Critical Analysis of Quantitative Fingerprint Individuality Models, J. Forensic Sciences, pp. 1187-1216, vol.31, no. 4, October 1986

[4] S. Pankanti, S. Prabhakar, A. K. Jain, On the Individuality of Fingerprints, IEEE Trans. on Pattern Analysis and Machine Intelligence, pp. 1010-1025, vol. 24, no. 8, August 2002

[5] S. L. Sclove, "The Occurrence of Fingerprint Characteristics as a Two Dimensional Process", J. of American Statistical Association, vol. 74, no. 367, pp. 588-595, 1979

[6] D. A. Stoney, "Distribution of Epidermal Ridge Minutiae", American J. of Physical Anthropology, vol. 77, pp. 367-376, 1988

[7] P. J. Diggle, Statistical Analysis of Spatial Point Patterns, Oxford University Press, 2003

[8] B. D. Ripley, "The Second-order Analysis of Stationary Point Processes", J. of Applied Probability, vol. 13, pp. 255-266, 1976

[9] H. W. Lotwick, B. W. Silverman, "Methods for Analysing Spatial Processes of Several Types of Points", J. of the Royal Statistical Society, vol. B 44, pp. 406-413, 1982

[10] http://www.maths.lancs.ac.uk/~diggle

[11] K. C. Chan, Y. S. Moon, P. S. Cheng, "Fast Fingerprint Verification Using Sub-regions of Fingerprint Images", IEEE Trans. on Circuits and Systems for Video Technology, vol. 14, no. 1, pp. 95-101, 2004

[12] D. Maltoni, D. Maio, A. K. Jain, S. Prabhakar, "Handbook of Fingerprint Recognition", Springer-Verlag, New York, Inc, 2003.