The Statistical Modelling of Fingerprint Minutiae Distribution with Implications for Fingerprint Individuality Studies

Jiansheng Chen   Yiu-Sang Moon
Department of Computer Science and Engineering,
The Chinese University of Hong Kong, Shatin N.T., Hong Kong
{jschen, ysmoon}@cse.cuhk.edu.hk

Abstract

The spatial distribution of fingerprint minutiae is a core problem in the fingerprint individuality study, the cornerstone of the fingerprint authentication technology. Previously, the assumption in most research that minutiae distribution is random has been proved to be inaccurate and may lead to significant overestimates of fingerprint uniqueness. In this paper, we propose a stochastic model for describing and simulating fingerprint minutiae patterns. Through coupling a pair potential Markov point process with a thinned process, this model successfully depicts the complex statistical behavior of fingerprint minutiae. Parameters of this model can be determined by nonlinear minimization. Furthermore, experiment results show that the statistical properties of our proposed model dovetails nicely with real minutiae data in terms of the false fingerprint correspondence probability. Such evidences indicate that the proposed model is a more accurate foundation for minutiae based fingerprint individuality studies as well as the artificial fingerprint synthesis when compared to the model of random distribution.

1. Introduction

Fingerprint authentication is believed to be the most commonly used biometric technology today [1]. Its wide social acceptance comes from the belief on the universality, stability and uniqueness of human fingerprints, among which the uniqueness, or the individuality, is the key to the discriminative power of fingerprints [2]. While fingerprint universality and stability can be confirmed by empirical anatomic observations, the individuality of fingerprints requires more deliberate theoretical analysis.

As a natural born feature of fingerprints, minutiae have been adopted for fingerprint representation and matching in most fingerprint authentication systems [2]. Hence, most previous fingerprint individuality research has focused on minutiae based representations [3, 4, 5, 6]. Although various models have been proposed to describe minutiae configuration to quantitatively evaluate the fingerprint individuality; the fundamental problem in the minutiae based fingerprint individuality study remains as finding a suitable model to depict the spatial distribution of minutiae. Previous research in this topic all reveal that fingerprint minutiae, when considered as two dimensional spatial point patterns, are NOT uniformly distributed [7, 8, 9]. In [7], Sclove showed that minutiae tend to cluster; while in [8] Stoney found that fingerprint minutiae demonstrate a slight tendency towards an overdispersed distribution. A unified view given in [9] claims that due to the growth stress during minutiae formation, fingerprint minutiae tend to be overdispersed when observed on a small scale; while clustering tendency dominate for large scales. Nevertheless, these qualitative conclusions do not provide enough help to the quantitative analysis of the fingerprint individuality.

Recently, a major advancement of this problem is the GMM (Gaussian Mixture Model) based minutiae model proposed in [10], in which the clustering tendency of minutiae is modeled. It is demonstrated that this mixture model gives rise to more realistic fingerprint individuality estimates than the uniform distribution model adopted in [3, 5, 6]. Nonetheless, this model assumes local independences among neighbor minutiae and ignores the minutiae overdispersing tendency [10]. Such an assumption could have limited the accuracy that the model may achieve.

To solve this problem, we propose a quantitative stochastic model for fingerprint minutiae distribution considering both the clustering and the overdispersing tendencies. Parameters of this model can be estimated through an *ad hoc* model fitting approach. Artificial minutiae patterns bearing similar statistical properties to real life fingerprints can be deterministically simulated from the model. Our model as well as the simulated artificial minutiae patterns can serve as a foundation for building more delicate fingerprint individuality models.

The remaining part of this paper is organized as follows. Section 2 introduces the spatial point analysis technologies. Section 3 describes the proposed model. Validating experiments performed on the simulated minutiae patterns will be presented in section 4. The last section is a conclusion of our work. The statistical calculations and simulations in this paper were implemented on the *R* platform and the following extension *R* packages have been used: *Spatstats*, *Splancs*, *Stats* and *MASS* [11].

2. Spatial Point Pattern Analysis

For a fingerprint, its minutiae form a two dimensional spatial point pattern. A fingerprint minutiae pattern from

978-1-4244-2243-2/08/$25.00 ©2008 IEEE

the NIST4 database is shown in Figure 1. For each minutia, its location, direction and type are the most widely used features [2]. In this work, we only concentrate on the statistical modelling of fingerprint minutiae spatial locations. All types of minutiae are treated equally because minutiae types cannot be automatically discriminated with a high level of accuracy [5]. For the minutiae direction, its statistical property and the way of incorporating it into fingerprint individuality models are fairly straightforward [6, 10], and will not be discussed in this paper.

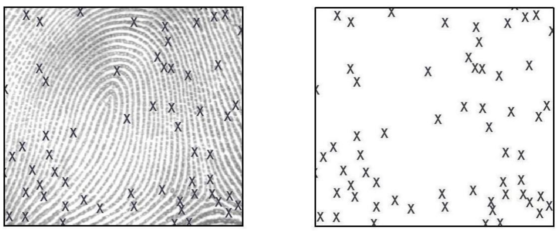


Figure 1: A NIST4 fingerprint and its minutiae pattern.

Statistical analysis for spatial point patterns has been long and widely used in the field of geography, astronomy, epidemiology and microanatomy [12, 13, 14, 15]. Given one or a group of spatial point patterns, there are two essential questions to be answered. First, what are the statistical properties of the given patterns? Second, how to formulate a parametric stochastic model (or process) that can be fitted to the given data appropriately?

For fingerprint minutiae patterns, the first question is partly answered in [9]. It is revealed that minutiae patterns are complex spatial point patterns demonstrating various distribution tendencies on different observation scales. As to the second question, there are generally two kinds of model fitting techniques: the likelihood-based method and the *ad hoc* method [13]. In the likelihood-based methods, the likelihood function of the model is calculated based on the input data; then the parameters are calculated by maximizing this function. As a formal statistical inference technique, the likelihood-based method has prevailed recently mainly due to the development of *Monte Carlo* methods [16] for calculating approximate likelihood functions for a wide range of stochastic models. However, this method is not suitable for minutiae pattern modelling due to two major reasons. Firstly, as is revealed in [9], fingerprint minutiae patterns need to be described by complex coupled models so that the resultant likelihood function will become notoriously intractable [13]. Secondly, the number of minutiae in a single fingerprint is relatively small so that the parameters cannot be stably estimated and may vary a lot among different fingerprints.

Therefore, we have adopted the *ad hoc* method which is based on comparing certain theoretical and empirical statistical properties, between the model and the original data. Compared to the likelihood based method, the *ad hoc* method is both computationally easier and is able to provide direct, visual methods for assessing the effectiveness of the fitting result. As suggested by Cressie in [12] and by Diggle in [13], we have chosen Ripley's *K function* as the statistical property for comparison. For a stationary isotropic spatial point process, its *K function* is defined as Equation (1), 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 [17].

$$K(t) = \lambda^{-1} E\left[N(t)\right] \quad (1)$$

Compared to first order statistical properties such as the intensity function, *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. A fingerprint minutiae pattern is a rather small sample for statistical analysis considering the number of minutiae in one fingerprint seldom exceeds one hundred. Also, the *K function* is invariant under a random thinning procedure in which each point of a given pattern is retained or not according to a series of mutually independent Bernoulli trials. Considering that the minutiae patterns we used as original data were marked by human experts, the case of missing any minutia can be approximated by a Bernoulli process and thus will not affect the *K function*.

$$\widehat{K}(t) = \{n(n-1)\}^{-1} |A| \sum_{i=1}^{n} \sum_{j \neq i} \omega(x_i, u_{ij})^{-1} \Lambda(u_{ij} \le t) \quad (2)$$

For a given spatial point pattern containing n points in a planar region A with the area $|A|$, an unbiased estimator of $K(t)$ was given by Ripley in [17] as Equation (2), in which u_{ij} is the distance between points x_i and x_j; $\Lambda(\bullet)$ denotes the indicator function; $\omega(x, u)$ was introduced by Ripley to eliminate the negative bias caused by boundary effects. It is defined as the proportion of the circumference of the circle with center x and radius u lying within A. The explicit formula for $\omega(x, u)$ can be deduced if A is rectangular [13].

$$D(t) = \widehat{K}(t) - \pi t^2 \quad (3)$$

If the target point pattern is a *CSR* (Complete Spatial Randomness) pattern, in which points are independently randomly distributed, its *K function* estimator should converge to πt^2 given n is big enough. Thus, $D(t)$ in Equation (3) can be used for indicating the point pattern's deviation from the *CSR*. Positive $D(t)$ values usually indicate a clustering tendency while negative $D(t)$ tells a overdispersing tendency [13]. This method is used in [9] to reveal the complex distribution tendency of minutiae.

For a bunch of replicated spatial point patterns generated by the same underlying process, their corresponding *K functions* are identically distributed. A reasonable overall estimate of the *K function* for the underlying process can be obtained by simply averaging the estimated *K functions* of all the replicated patterns using Equation (4), in which n_i is the number of points in the i^{th} point pattern, of which the *K function* estimator is $\widehat{K}_i(t)$.

$$\widehat{K}^*(t) = \sum_{i=1}^{r} n_i \widehat{K}_i(t) \Big/ \sum_{i=1}^{r} n_i \quad (4)$$

A critical step of the *ad hoc* method is to propose an appropriate stochastic model. Diggle recommends plotting the *K function* estimator for suggesting potential candidate models, and to provide initial parameter estimates [13]. The number of parameters in the model should be appropriate, since too few parameters will downgrade the model flexibility, while too many parameters will lead to convergence difficulties in the minimization step. Suppose the proposed model incorporates a parameter vector ζ. Let $K(t;\zeta)$ denotes the ground true *K function* of the model. A family of criteria for measuring the discrepancy between the model and the original data is suggested in [12] as Equation (5). The value of the parameter vector ζ can thus be found by applying nonlinear minimization or regression methods, such as the Gauss-Newton algorithm or the Golub-Pereyra algorithm [18], to minimize $Q(\zeta)$.

$$Q(\zeta) = \int_0^{t_0} \left\{ \left(\widehat{K}^*(t) \right)^C - \left(K(t;\zeta) \right)^C \right\}^2 dt \quad (5)$$

3. Fingerprint Minutiae Pattern Modelling

Fingerprints from three different databases were chosen for model designing and fitting. These three databases are: NIST4 (512×512, ~500dpi); FVC2002 DB1 (388×374, 500dpi) and FP383 [19] (256×256, 450dpi). These three databases were collected in different regions around the world and from various populations. Hopefully, they can ensure the universality of conclusions made based on them.

Not all the fingerprints in these databases were used. Three criteria were followed when selecting fingerprints. First, all the fingerprints selected were from different finger tips. Second, only fingerprints with sufficiently high image quality were selected. Since the minutiae of the selected fingerprints were manually marked, high image quality helps to improve the reliability of the marking results. Third, only fingerprints with a big enough (> 220×220 pixels) ROI (Region of Interest) were picked up to ensure that for each fingerprint minutiae pattern, there is enough number of minutiae for the statistical analysis. As a result, we selected 133 fingerprints from NIST4, 103 fingerprints from FP383 and 56 fingerprints from FVC2002 DB1. A stochastic model was designed and it was then fitted to the selected fingerprints from the three databases respectively. All the selected fingerprints were normalized to 500dpi; their minutiae were carefully marked and double checked by human experts. A square area of 220×220pixels was randomly selected inside the ROI of each fingerprint. The minutiae patterns inside these squares were used as the original data for the model fitting. Without loosing generality, we change the unit so that these squares become unity. Figure 2 shows three selected fingerprint samples.

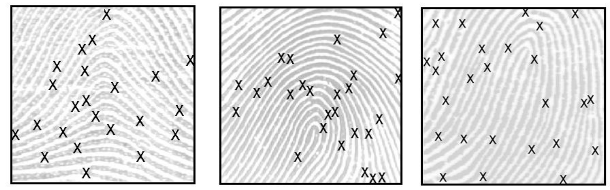


Figure 2: Sample selected fingerprint minutiae patterns

To get a universal *K function* estimator for minutiae patterns, we treat the fingerprint minutiae patterns from the same database as replicated spatial point patterns generated by one identical underlying process. Under such an assumption, Equation (4) can be applied. This seemingly bold assumption can actually be justified. Although there are numerous environmental and mental factors affecting the physical growth of an individual throughout her/his life, the formation of fingerprints is finished within a relatively short period and in a relatively stable environment. It is believed that fingerprints are fully formed at about seven months of fetus development and the formulated finger ridge configurations do not change throughout the whole life of an individual except due to accidents such as bruises and cuts on the fingertips [2]. A deterministic mathematical model has been proposed for modeling the mechanism of fingerprint formation [20]. Besides the formation mechanism, fingerprint acquisition is another major factor affecting the statistical properties of the minutiae patterns. The acquisition procedure can vary in many ways for different databases, such as acquisition technique, fingerprint sensors and collection settings. Nevertheless, fingerprints from the same database can be expected to have been collected through a relatively consistent procedure.

Figure 3 shows the estimated *K functions*, together with plus and minus two bootstrap standard deviations [13], for the three databases. From Figure 3, we can observe a clear ‘small scale overdispersing and large scale clustering’ [9] distribution tendency for all the three databases, leading to a ‘tick’ shape for all the three curves. We can also notice that that the overdispersing tendency is much more obvious in the NIST4 database then in the other two databases. This is because that NIST4, unlike the other two databases, was created by scanning inked fingerprints. The ink technique requires users to roll their fingers against the media with a heavy pressure. The finger tip deformation thus caused will inevitably increases the inter minutiae distances, in other words, further disperses the minutiae. Also, FVC2002 and FP383 have very similar *K functions*, indicating the rationality of our replicated pattern assumption.

Such a complex distribution tendency indicates that a composite stochastic model should be more suitable than any simple models for describing minutiae patterns. To model the small scale overdispersing tendency, we choose the *pair potential Markov point process*; to describe the large scale clustering tendency, a *thinned process* is used. Theoretically, a composite model consists of these two models is suitable for describing any point patterns whose *K functions* resemble the ‘tick’ shape curves in Figure 3.

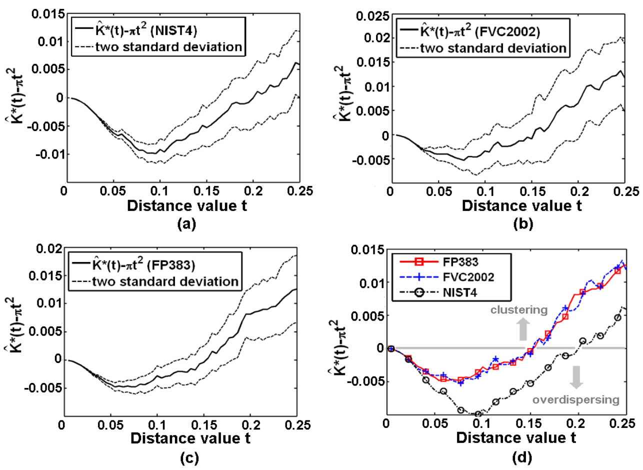


Figure 3: Estimated *K functions* for the three fingerprint databases.

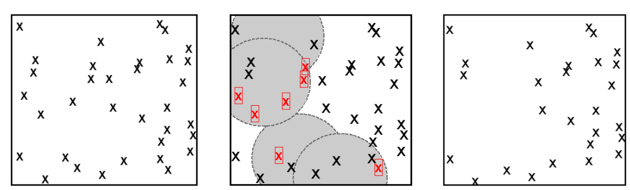
A *Markov point process* is a spatial point process in which the conditional intensity at any point s only depends on the points inside a closed ball of radius r_1 centered at s [12]. In other words, this kind of point process only involves local or *Markovian* dependences amongst points. A *Markov point process* is usually defined by its likelihood ratio function $f(\chi)$ with respect to a *Poisson process* of unit intensity. A *Poisson process* is defined to generate *CSR* point patterns. The *pair potential Markov point process* is a special kind of *Markov point process* whose $f(\chi)$ only depends on inter-point distances as is shown in Equation (6), where $h(\bullet)$ is a non-negative function of the inter-point distance and is usually called the interaction function. When $0 \le h(\bullet) < 1$, this model can stably generate overdispersed point patterns; while $h(\bullet)=1$ means no overdispersing and $h(\bullet)=0$ means strict inhibition; otherwise the smaller the value of $h(\bullet)$ is, the more obvious the overdispersing effect will be [13].

$$f(\chi) = \alpha\beta^n \prod_i \prod_{i \ne j} h\left(\left\| x_i - x_j \right\|\right) \tag{6}$$

To design a specific *pair potential Markov model*, it is crucial to propose a suitable form for $h(\bullet)$. Based on our observation, we propose an interaction function as Equation (7), which can be interpreted as an extension of the VSC (Very-Soft-Core) model proposed in [21], with one more degree of freedom. We let $r_1=0.2$ and r_0 be equal to the smallest t for the value of *K function* to be positive.

$$h(r) = \begin{cases} 0 & r < r_0 \\ 1 - \exp\left\{-\left(r - r_0\right) \times \left(ar + b\right)\right\} & r_0 \le r < r_1 \\ 1 & r_1 \le r \end{cases} \tag{7}$$

A *thinned process* $P(x)$ is defined by a basic point process $P_0(x)$ and a thinning field $Z(x)$. $Z(x)$ is also a stochastic process independent of $P_0(x)$. The points x_i in $P_0(x)$ are retained independently with probabilities $Z(x_i)$. Thus, $P(x)$ consists of the retained points of $P_0(x)$ [13]. It is a common practice to use the thinned process to model various large scale distribution tendencies. In our case, the basic point process $P_0(x)$ is the *pair potential Markov point process* defined above, and $Z(x)$ is designed to generate large scale clustering tendencies. We have observed that fingerprint minutiae tend to cluster around points where the ridge directions change abruptly, such as core points and delta points. In other words, fingerprint regions with smooth ridges usually have less probability of minutiae emergence. Thus we define $Z(x)$ as follows: discs of radius δ centered on the points generated by a *Poisson process* of intensity λ. As such, $Z(x)=p$ ($0<p\le1$) inside the union of all such discs, and $Z(x)=1$ otherwise. The resulting process displays small scale overdispersing due to the pair-wise interaction; and large scale clustering induced by higher minutia density in areas outside the discs defined by $Z(x)$. Figure 4 shows the generation of a sample point pattern through the proposed composite process.



(a) (b) (c)

Figure 4: Point pattern generation using the proposed model. (a) A *pair potential Markov point process* point pattern. Overdispersing can be observed. (b) The thinning field $Z(x)$ consists of four discs inside which $p=0.5$. Altogether 7 points are removed. (c) The resultant point pattern demonstrates large scale clustering caused by the thinning process. Points tend to concentrate in the region outside the four discs.

Summing up, the proposed model contains five parameters, or $\zeta = (a, b, \lambda, \delta, p)$ in Equation (5). Ideally, if the theoretical *K function* of this process can be explicitly expressed, the five parameters can be evaluated by minimizing $Q(\zeta)$. Unfortunately, although the *K function* of the thinned process can be explicitly formulated [13], the *K function* of the *pair potential Markov point process* defined by Equations (6) and (7) cannot be expressed in a closed form [12]. For such a situation in which the theoretical $K(t;\zeta)$ is unknown, Diggle suggested using $K_S(t;\zeta)$, the point-wise mean of the estimated *K functions* calculated from *S* simulated realizations of the model [13]. $K_S(t;\zeta)$ is calculated using a formula similar to Equation (4).

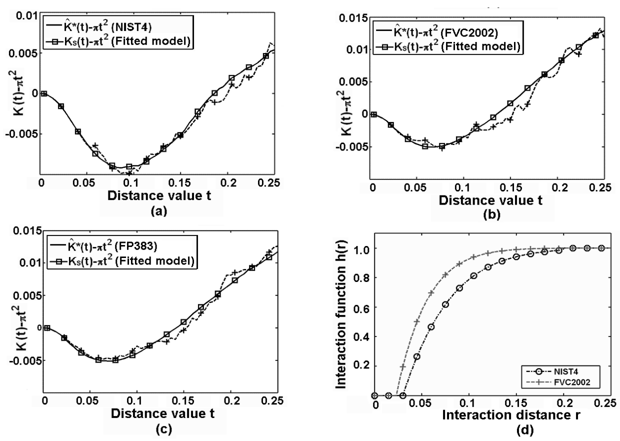


Figure 5: Results of the *K function* fitting for three databases.

The simulated realization of the proposed model consists of the simulations of the *pair potential Markov point process* and the *thinned process*. The definition of the *thinned process* is by itself the procedure for its simulation. For the *pair potential Markov point process*, Diggle proposed a depletion-replacement procedure conditioned on a given number of points, *n*, in the resultant simulated pattern [13]. In our work, we have used a simple mechanism to guarantee that the number of points in the simulated patterns has a distribution similar to that of the original fingerprint minutiae data. It is revealed in [9] that fingerprint minutiae number inside a given area approximately follows a *Poisson* distribution. Thus, we generated the values of *n* as random numbers following a *Poisson* distribution with mean m_p. The value of m_p can be calculated using Equation (8) from m_f, which is the mean number of minutiae in the original fingerprint data. The denominator of Equation (8) is the mean of the thinning field $Z(x)$. For NIST4, FVC2002 and FP383, m_f equals to 21.2, 29.0 and 27.7 respectively.

$$m_p = m_f \Big/ \left\{1-(1-p)\times\left(1-\exp\left(-\pi\lambda\delta^2\right)\right)\right\} \qquad (8)$$

We chose $c=0.5$ as suggested in [12] and $S=800$. The Newton-type algorithm proposed in [18] was adopted to perform the nonlinear minimization of Equation (5), in which $K(t;\zeta)$ should be replaced by $K_S(t;\zeta)$. The estimated model parameter values for the three databases are listed in Table 1. The values in the last column of Table 1 are the mean ratios of the fingerprint regions with high minutiae density (regions outside the thinning discs).

We can notice that FVC2002 and FP383 yield nearly the same values for all the five parameters. This is reasonable since the *K functions* of the two databases are very close to each other as shown in Figure 3d. Figure 5a-5c visualize the model fitting results in terms of the *K function* for the three databases. It can be seen that the proposed model is flexible enough to fulfill the estimated *K functions* for fingerprint minutiae patterns from different databases. Figure 5d compares the interaction functions $h(r)$ for NIST4 and FVC2002. NIST4 has smaller $h(r)$ than FVC2002 for most values of *r*, indicating a stronger overdispersing tendency caused by the ink technique.

Table 1: Estimated model parameters for three databases.

Database	a	b	λ	δ	p	$\exp(-\pi\lambda\delta^2)$
NIST4	31.003	18.990	3.478	0.199	0.121	0.649
FVC2002	40.947	29.703	2.712	0.270	0.256	0.538
FP383	40.952	29.709	2.712	0.269	0.261	0.539

Figure 6 shows several sample patterns generated by the proposed model using the parameters listed in Table 1. Visually, they resemble real fingerprint minutiae patterns demonstrated in Figure 1 and Figure 2 in terms of the distribution tendencies. This is expected since the *K function* used in the model fitting is an effective descriptor of the distribution tendency. In the next section, we will propose a novel statistical property, the *bipartite matching score distribution*, to further validate our proposed model. For fingerprint minutiae patterns, this statistical property is significant as it can be directly linked to the fingerprint matching score. We will demonstrate through experiments that in terms of this new statistical property, the point patterns generated from the proposed model are also analogous to the real fingerprint minutiae patterns.

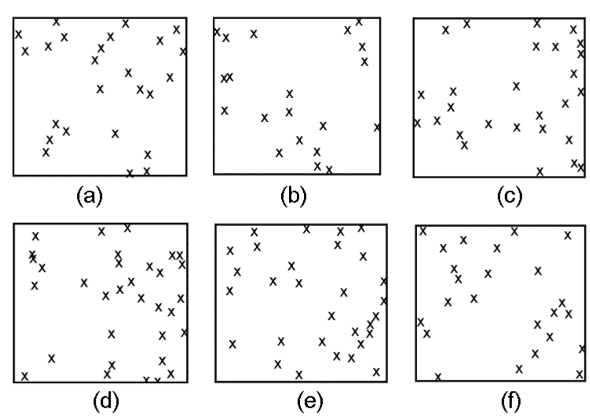


Figure 6: Sample simulated patterns of the proposed model. (a,b) NIST4; (c,d) FVC2002; (e,f) FP383.

4. Experiments

In this section, we will further validate our proposed statistical model by its application in investigating a fundamental problem in fingerprint individuality studies [3, 4, 5, 6], the imposter minutiae matching score distribution. In an imposter matching, the two fingerprints for matching are from different fingertips. Therefore, the probability distribution of the score for such a matching is by definition closely related to the fingerprint individuality, which refers to the differences among the fingerprints obtained from different fingers. As minutiae direction is not considered in this paper, we will only investigate fingerprint matching using the minutia location information.

The basic idea of the fingerprint matching is to count the number of minutia point correspondences between two fingerprints. Although many effective algorithms have been proposed for minutiae matching [2], most of them are heuristic and may be implemented in diverse ways. To avoid uncertainties, we propose a new deterministic and symmetric approach for point pattern matching. Suppose two point patterns are of the identical size, and are aligned through simple overlapping. Their *bipartite matching score* is defined as the maximum possible number of point correspondences between these two patterns. There are two criteria for two points, x and y, from the two patterns to be matched. First, the Euclidean distance between x and y is equal to or smaller than a threshold d_0. Second, neither x nor y is matched to any other points.

This problem can be formulated as a *bipartite graph* matching problem, for which the *Hungarian algorithm* is one of the most efficient solutions [22]. Suppose x_i and y_i are points from two point patterns. They are converted to the vertices of a *bipartite graph*, in which x_i and y_i are connected by an edge if and only if their Euclidean distance is equal to or smaller than d_0. Thus, the maximum number of point correspondences, or the *bipartite matching score*, is equal to the number of edges covered by the maximum matching of the graph. The *bipartite matching score* is invariant to the order of the input, and is thus symmetric. For a group of replicated point patterns generated by a stochastic process, the probability distribution $P_b(s;d_0)$ of the *bipartite matching scores* for all the possible point pattern pairs can be used as a statistical property for describing the underlying process. For fingerprints, $P_b(s;d_0)$ is merely the imposter matching score distribution of the one to one verification experiment.

For each database, $P_b(s;d_0)$ is computed for three groups of point patterns: a) the selected fingerprint minutiae patterns; b) 1000 simulated realizations of the proposed composite model; c) 1000 *CSR* point patterns. We use the *CSR* point pattern for comparison because randomness or quasi randomness is assumed for fingerprint minutiae distribution in most previous fingerprint individuality studies [3, 5, 6]. The results are shown in Figure 7. The score distributions for the proposed composite model nearly overlap those of real fingerprint minutiae patterns, indicating a high validity of our work.

Table 2: Probability values $P_b(s;d_0)$ (%) for relatively high scores.

d_0= 0.07	*Score*	*15*	*16*	*17*	*18*	*19*
	FP383	2.9	1.8	1.4	0.8	0.4
	Proposed Model	2.3	1.6	1.0	0.6	0.4
	CSR model	0.9	0.4	0.2	0.1	0.04
d_0= 0.114	*Score*	*22*	*23*	*24*	*25*	*26*
	FP383	2.5	2.1	1.9	1.3	1.2
	Proposed Model	2.5	1.9	1.5	1.1	0.8
	CSR model	1.4	0.9	0.5	0.3	0.2

For the *CSR* model, the distributions obviously deviate from that of the minutiae patterns. A considerable underestimate of the probability can be observed for relatively high matching scores scenarios. These scenarios are usually the focus of the fingerprint individuality study. For clarity, part of the data for plotting the curves in Figure 7(c) and 7(d) are listed in Table 2.

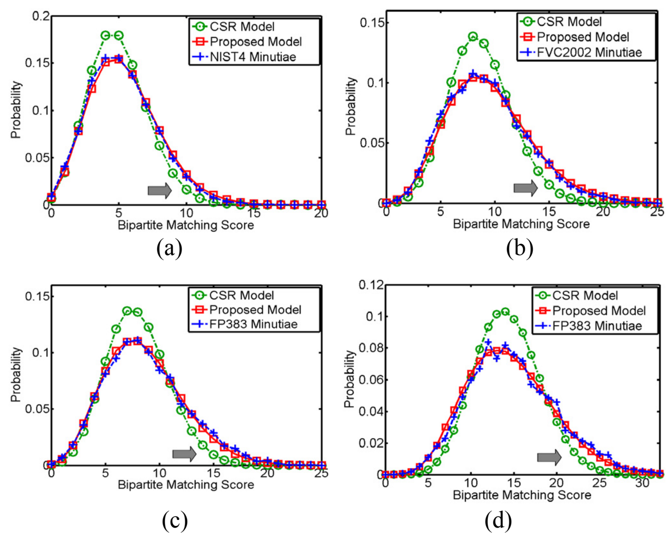


Figure 7: Probability distribution of the *bipartite matching scores* for three databases. In (a), (b) and (c), d_0=*0.07* (~15 pixels for 500 dpi images); in (d) d_0=*0.114* (~25 pixels).

Basically, this problem is caused by ignoring the clustering tendency of minutiae when assuming the random distribution. We have observed that the matching score between two minutiae patterns is greatly affected by the area of the overlapped clustering minutiae regions. Ignoring the clustering tendency causes an underestimate of the minutiae density inside these regions.

Underestimating the probability of false fingerprint correspondences is fatal for the fingerprint individuality study because it will probably impair suspects in a criminal investigation by preventing them from doubting the certainty level of the fingerprint matching [7]. Actually, ignoring the overdispersing tendency also bias the probability estimates. Hence, compared to the *CSR* model used in [3, 5, 6], our proposed model provides a more precise description of the fingerprint minutiae spatial distribution, and may serve as a better foundation for more accurate fingerprint individuality studies.

5. Conclusions

The spatial distribution of fingerprint minutiae is a core problem of the fingerprint individuality study. Due to the complexity of the formation process, the spatial distribution of fingerprint minutiae cannot be accurately described by simple statistical models. Based on the qualitative conclusions made in previous research, we have proposed a deterministic composite stochastic model for describing and simulating fingerprint minutiae patterns. This model consists of a *pair potential Markov point process* and a *thinned process*. The *Markov point process* is employed to simulate the overdispersing among minutiae. Its interaction function is designed to be exponentially increasing with respect to the inter minutiae distances. The *thinned process* simulates the large scale clustering of minutiae by creating low minutiae density regions where the probability of the emergence of minutiae is generally lower than the remaining parts of fingerprints. An *ad hoc* method is adopted to fit the model to real life fingerprint minutiae data through minimizing the discrepancy between the theoretical and estimated *K functions*.

Experimental results show that the proposed model over performs the commonly used *CSR* model by enabling more accurate estimation of false fingerprint correspondence probabilities, indicating that it can serve as a foundation for more accurate theoretical analysis of the fingerprint individuality. Moreover, the simulating realization of the model can be used in fingerprint synthesis systems for generating more realistic artificial fingerprints.

To further improve our model, there are mainly two challenges. Firstly, incorporating the minutiae direction into the proposed model is necessary. The Von-Mises distribution based random angle generation process proposed in [10] provides a possible solution. Secondly, the *K function* of the *pair potential Markov process* defined by Equation (7) cannot be expressed in a closed form [12]. This greatly jeopardizes the usability of the proposed model for any further theoretical analysis. Nonetheless, this difficulty can be overcome by approximating Equation (7) using a piecewise function, which enables the explicit expression of the *K function* [12].

Acknowledgements

The work described in this paper was substantially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. 415106).

References

[1] International Biometric Group, "Biometrics Market and Industry Report 2007-2012", www.biometricgroup.com.

[2] D. Maltoni, D. Maio, A.K. Jain, and S. Prabhakar, *Handbook of Fingerprint Recognition*, Springer 2003.

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

[4] 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.

[5] S. Pankanti, S. Prabhakar, A. K. Jain, "On the Individuality of Fingerprints", IEEE Trans. on PAMI, pp. 1010-1025, vol. 24, no. 8, August 2002.

[6] J.S. Chen, Y.S. Moon, "A Minutiae-based Fingerprint Individuality Model", Proceedings of CVPR'07, June 2007

[7] 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.

[8] D.A. Stoney, "Distribution of Epidermal Ridge Minutiae", Am. J. of Phy. Anthropology, vol. 77, pp. 367-376, 1988.

[9] J.S. Chen, Y.S. Moon, "A Statistical Study on the Fingerprint Minutiae Distribution", Proceedings of ICASSP'06, pp. II-169 – II-172, May 2006.

[10] Y. Zhu, S.C. Dass and A.K. Jain, "Statistical Models for Assessing the Individuality of Fingerprints", IEEE Trans. on IFS, vol. 2, issue 3, part 1, pp. 391-401, Sept. 2007.

[11] The *R* project: http://www.r-project.org.

[12] N.A.C. Cressie, *Statistics for Spatial Data*, John Wiley & Sons Inc., 1993.

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

[14] J. Moller et. al., *Statistical Inference and Simulation for Spatial Point Process*, Chapman & Hall/CRC, 2004.

[15] M. Jacobsen, *Point Process Theory and Applications*, Boston; Berlin: Birkhäuser, 2006.

[16] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, *Markov Chain Monte Carlo in Practice,* Chapman & Hall, 1996.

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

[18] J.E. Dennis and R.B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, NJ, 1983.

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

[20] M. Kuchen, A.C. Newell, "Fingerprint Formation", J. of Theoretical Biology, vol. 235, Issue 1, pp. 71 – 83, 2005.

[21] Y. Ogata and M. Tanemura, "Likelihood analysis of spatial point patterns", J. R. Statist. Soc. B 46, pp. 496-518, 1984.

[22] T.H. Cormen, C.E. Leiserson, R.L. Rivest and C. Stein, *Introduction to Algorithms*, the MIT press, 2001.