Compound Stochastic Models For Fingerprint Individuality

Yongfang Zhu, Sarat C. Dass
Department of Statistics & Probability
Michigan State University
East Lansing, MI 48824

Anil K. Jain
Department of Computer Science & Engineering
Michigan State University
East Lansing, MI 48824

Abstract

The question of fingerprint individuality can be posed as follows: Given a query fingerprint, what is the probability that the observed number of minutiae matches with a template fingerprint is purely due to chance? An assessment of this probability can be made by estimating the variability inherent in fingerprint minutiae. We develop a compound stochastic model that is able to capture three main sources of minutiae variability in actual fingerprint databases. The compound stochastic models are used to synthesize realizations of minutiae matches from which numerical estimates of fingerprint individuality can be derived. Experiments on the FVC2002 DB1 and IBM HURSLEY databases show that the probability of obtaining a 12 minutiae match purely due to chance is 1.6×10^{-5} when the number of minutiae in the query and template fingerprints are both 46.

1 Introduction

The problem of fingerprint individuality is to develop quantitative measures that characterize the extent of uniqueness of a fingerprint, i.e., given a query fingerprint, "What is the probability that the observed number of minutiae matches with a particular template is purely due to chance?". In this paper, we develop compound stochastic models that account for three sources of minutiae variability, namely, (i) the variability in the minutiae distributions in different fingers, (ii) variability due to local perturbations arising from non-linear distortion effects in multiple impressions of a finger, and (iii) variability due to the size of partial prints (or the area of finger region captured) in multiple acquisitions of a finger. The compound stochastic models are then used for synthesis and for obtaining estimates of fingerprint individuality; the main steps of our methodology are outlined in Figure 1.

There have been several previous studies that addressed the problem of fingerprint individuality using probabilistic models on minutiae [8]. However, the assumptions made
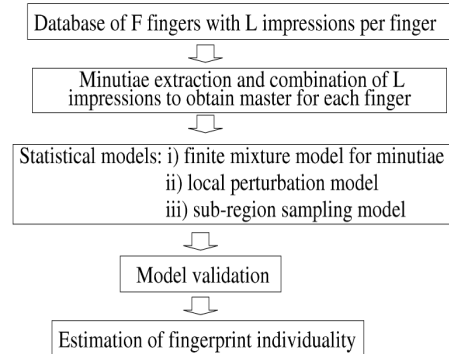
Figure 1. Flow chart for fingerprint individuality

in these studies did not satisfactorily represent the observed minutiae variations in multiple impressions of a finger in actual fingerprint databases; The "corrected" uniform model in [8] did not account for clustering tendencies of the minutiae. Dass et al [4] accounted for intra-class minutiae variability by fitting separate (and hence, uncorrelated) stochastic models for multiple acquisitions of a finger. The lack of satisfactory measures of fingerprint individuality has been the highlight of several recent court cases where the admissibility of fingerprint evidence is being challenged based on the premise that the uniqueness of fingerprints has not been objectively tested and matching error rates are unknown [5, 10].

2 Compound Stochastic Models

Suppose a fingerprint database consisting of F different fingers and L impressions per finger is available. Let $\mathcal{F}(f, l)$ denote the l-th impression of the f-th finger. The fingerprint features we consider are the minutiae locations and directions in each $\mathcal{F}(f, l)$. As a first step towards constructing the compound generative model for finger f, the
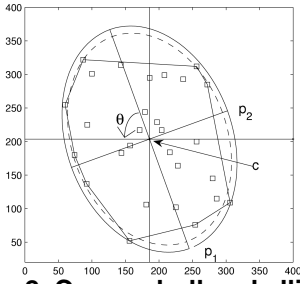
Figure 2. Convex hull and ellipses

minutiae in all the L impressions $\mathcal{F}(f,l), l = 1, 2, \ldots, L$ are combined to obtain a "master" set. A reference impression, without loss of generality, $\mathcal{F}(f,1)$, say, is first chosen. Next, for each $l = 2, 3, \ldots, L$, $\mathcal{F}(f,l)$ is aligned to $\mathcal{F}(f,1)$, and correspondences between minutiae in $\mathcal{F}(f,l)$ and $\mathcal{F}(f,1)$ are found. When a minutiae in $\mathcal{F}(f,l)$ does not have any corresponding minutiae in $\mathcal{F}(f,1)$, that minutiae is appended to the list of minutiae in $\mathcal{F}(f,1)$, thus increasing the total number of minutiae in $\mathcal{F}(f,1)$ by one each time such a minutiae is found. The consolidation of minutiae into the master set in this way results eventually in a total of n consolidated minutiae in $\mathcal{F}(f,1)$ with correspondence sets M_k, $k = 1, 2, \ldots, n$. We denote the elements in each M_k by $\{(X_{kj}, D_{kj}), j = 1, 2, \ldots, m_k\}$. For each set of correspondences, we define the mean, or center, of X_{kj} by $\bar{X}_k = \frac{1}{m_k} \sum_{j=1}^{m_k} X_{kj}$. The mean of D_{kj}, $\bar{D}_k$, is taken to be the phase angle of the complex number $\sum_{j=1}^{m_k} \cos(D_{kj}) + i \sin(D_{kj})$ (see also [7]). The deviations of locations and directions from their respective centers are given by

$$\mathcal{D}_k = \{(X_{kj} - \bar{X}_k, D_{kj} - \bar{D}_k), j = 1, 2, \ldots, m_k\}. \quad (1)$$

The first two stages of our compound generative model consist of developing statistical models on (i) the centers, and (ii) the deviations of the observed minutiae from their respective centers.

2.1 Finite Mixture Model On The Centers

We assume each minutiae center in the master, $(\bar{X}_k, \bar{D}_k)$, $k = 1, 2, \ldots, n$ is independently distributed according to the mixture density

$$f(s, \theta | \Theta_G) = \sum_{g=1}^{G} \tau_g f_g^X(s | \mu_g, \Sigma_g) \cdot f_g^D(\theta | \nu_g, \kappa_g, p_g), \quad (2)$$

where G is the total number of components; for the g-th component, $f_g^X(s \mid \mu_g, \Sigma_g)$ is a bivariate Gaussian density with mean μ_g and covariance matrix Σ_g, $f_g^D(\theta \mid \nu_g, \kappa_g, p_g)$ is the mixture of two Von-Mises densities (see [7]) with means ν_g and $\nu_g + \pi$, respectively, common precision parameter, κ_g, and mixing probabilities p_g and $1 - p_g$, respectively. The mixture model (2) enables flexible modeling of a

variety of observed minutiae patterns in fingerprint images. Similar mixture models have been used in [4] to model the distribution of minutiae locations and directions in different impressions of a finger.

2.2 Local Perturbation Models

For the local perturbation models, the domain of the master is first divided into a lattice of b_0 non-overlapping blocks, $\mathcal{B} = \{B_b, b = 1, 2, \ldots, b_0\}$. Next, the set $\mathcal{D}_k$ is assigned to block B if $\bar{X}_k \in B$. In block B, the location deviations, $\{X_{kj} - \bar{X}_k : \bar{X}_k \in B\}$ are modeled as a bivariate normal distribution with mean 0 and covariance matrix S_B, whereas the direction deviations, $\{D_{kj} - \bar{D}_k : \bar{X}_k \in B\}$ are modeled as a Von-Mises distribution with mean 0 and precision κ_B. The unknown parameters S_B and κ_B are estimated from the observed deviations in each block B.

2.3 Modeling the Variability of Partial Prints

The third and final component of our compound stochastic model accounts for minutiae variability due to partial prints in multiple acquisitions of a finger. For the impression $\mathcal{F}(f,l)$, a convex hull encompassing all minutiae locations is first determined (see Figure 2; minutiae are labeled as squares). Next, an ellipse (denoted by a dashed line in Figure 2) is obtained by the direct least square fitting method [6]. This method yields an ellipse that does not encompass all minutiae locations on the boundary of the convex hull. We increase the size of the ellipse to the minimal ellipse (denoted by a solid line in Figure 2) containing all minutiae locations on the boundary while maintaining the shape of the ellipse. We denote the parameters that uniquely determine this ellipse by the area $(A(f,l))$, length of major axis $(p_1(f,l))$, the orientation $(\theta(f,l))$ and the center $(c(f,l))$. In our experiments, we fixed the ratio of the lengths of the major to minor axes of each ellipse (say, r_0). Thus, the effective ellipse parameters reduce to the triplet $E(f,l) \equiv (A(f,l), \theta(f,l), c(f,l))$ only. We denote by $\mathcal{E}$ to be the collection of all ellipse parameters for all the fingerprint impressions in the fingerprint database. Also, let $\mathcal{T} \equiv \{T^{-1}(f,l)\}$, where $T(f,l)$ is the rigid transformation used to align $\mathcal{F}(f,l)$ to $\mathcal{F}(f,1)$.

For assessment of fingerprint individuality, we require to synthesize impressions from a finger consisting of a pre-specified number of minutiae, m_0, say. We describe a conditional simulation technique that performs this.

2.4 Fingerprint Synthesis

Figures 3 (a) and (b) give an instance of a finger f, and the constructed master set and the minimal ellipse,
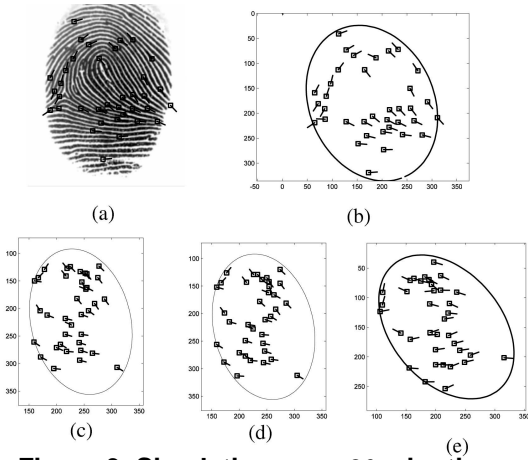
Figure 3. Simulating $m_0 = 36$ minutiae

Figure 4. FVC2002 DB1 (left), IBM (right)

respectively. First, (i) A random ellipse is generated where the area is a sample from a normal distribution with mean $h_0(m_0)$ and standard deviation σ_0 (h_0 is the least squares quadratic polynomial fitted to the scatter plot of $(m(f,l), A(f,l))$). The orientation and center are randomly selected from the second and third components of $\mathcal{E}$. (ii) A center is generated from the mixture model of master f. (iii) A deviation is generated according to the local perturbation model in Section 2.2, and compounded to the generated center from step (ii). This synthesized minutiae is retained if its location lies within the ellipse in step (i) and rejected otherwise. Steps (ii) and (iii) are repeated until m_0 synthesized minutiae features fall inside the ellipse.Figure 3 (c) shows the m_0 synthesized minutiae centers from the mixture model whereas Figure 3 (d) shows the synthesized minutiae after compounding with the local perturbation model. Finally, (iv) the m_0 minutiae are transformed by a random rigid transformation from $\mathcal{T}$ to form the synthesized impression (Figure 3 (e)).

3   Assessing Fingerprint Individuality

Given a query fingerprint, Q, with m_Q minutiae, and a template, T, with m_T minutiae, we wish to estimate the probability that Q and T share exactly w minutiae:

$$P\{M(Q,T) = k \,|\, \#Q = m_Q, \#T = m_T\}, \qquad (3)$$

where $M(Q,T)$ is the number of minutiae matches between Q and T as determined by a matcher M when (Q,T) are a pair of impostor fingerprints. We utilize the simulation technique in Section 2.4 to synthesize fingerprint impressions so that each Q (respectively, T) has m_Q (m_T) minutiae. In our experiments we took $m_Q = m_T$, so it was sufficient to generate only one synthetic fingerprint database with $m_Q(= m_T)$ minutiae features for each impression. We denote this synthetic database by $\mathcal{F}^*(f,h), h = 1, 2, \ldots, H$
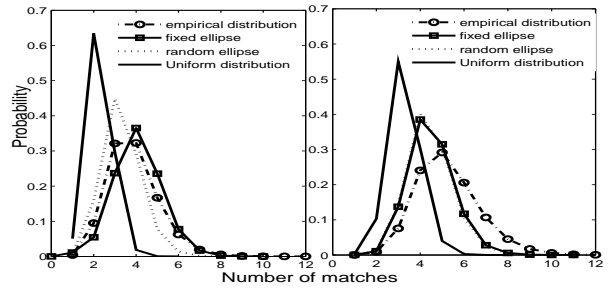
and $f = 1, 2, \ldots, F$ where $\mathcal{F}^*(f,h)$ is the h-th synthetic impression from finger f. The probability in (3) can be estimated by

$$p(w) = \frac{\displaystyle\sum_{h=1}^{H}\sum_{h'=1}^{H}\sum_{f=1}^{F}\sum_{\substack{f'=1 \\ f \neq f'}}^{F} I_w\{(f,h),(f',h')\}}{F(F-1)H^2} \qquad (4)$$

where $I_w\{(f,h),(f',h')\}$ is 1 if $M(\mathcal{F}^*(f,h), \mathcal{F}^*(f',h'))$ equals w, and 0, otherwise.

Note that (4) does not estimate (3) well when (3) is extremely small (this is the case when w is large). To obtain estimates of $p(w)$ for large w, we use the extrapolation formula

$$p(w) = p(w_0)\, 10^{\beta_1(w-w_0)}, \qquad (5)$$

where w_0 is the location of the modal value of $p(w)$, and β_1 is the slope of the least squares line fitted to the graph of $(w, \log_{10}(p(w)))$ which is obtained for moderate w values using (4).

4   Experimental Results

We used two databases in our experiments, namely, FVC2002 DB1 and IBM-HURSLEY. The FVC2002 DB1 database contains fingerprint images of $F = 100$ different fingers and $L = 8$ impressions per finger obtained using the optical sensor "TouchView II" by Identix. The IBM-HURSLEY database contains two impressions of 269 fingers with manually labeled minutiae and correspondences for 900 impressions. For IBM-HURSLEY Database, we manually checked the minutiae correspondence of a subset of the database with $F = 229$ fingers with $L = 2$ images per finger and use the subset for our experiment.

For validating the compound stochastic model, we generated a synthetic database consisting of F fingers with L impressions per finger in the following way: For finger f, a total of n minutiae were synthesized from the fitted mixture (for centers) and local perturbation (for deviations from centers) models. The parameterized ellipse and the

(m_Q, m_T, w)	Pooled Estimate
$(26, 26, 12)$	9.9×10^{-10}
$(26, 26, 26)$	1.4×10^{-23}
$(36, 36, 12)$	3.7×10^{-7}
$(36, 36, 36)$	1.5×10^{-26}
$(46, 46, 12)$	1.6×10^{-5}
$(46, 46, 46)$	8.1×10^{-28}

Table 1. Fingerprint-individuality estimates

(m_Q, m_T, w)	Our model	Model in [8]
$(26, 26, 12)$	6.8×10^{-10}	2.4×10^{-15}
$(36, 36, 12)$	6.5×10^{-7}	1.0×10^{-10}
$(46, 46, 12)$	2.0×10^{-5}	3.9×10^{-8}

Table 2. Fingerprint-individuality comparison

rigid transformation $T(f, l)$ were then, respectively, used to obtain a synthetic minutiae set and a synthetic impression. The distribution of the number of impostor minutiae matches for this synthetic database is obtained based on the matcher described in [9]. This distribution is represented by the solid line with squares ($\square$) in Figures 4. Another synthetic database of F fingers with L impressions per finger was constructed using the conditional simulation technique in Section 2.4 so that the number of minutiae for the l-th impression of finger f is $m(f, l)$, the observed number of minutiae. The ellipse and rigid transformation were randomly selected from $\mathcal{E}$ and $\mathcal{T}$, respectively, to generate a synthetic impression. The value of r_0 (ratio of lengths of major to minor axes of the best fitting ellipse) for the FVC and IBM databases were taken to be the mean values, namely 1.48 and 1.61, respectively. The corresponding distribution for the number of impostor matches is represented in Figures 4 (a) and (b) by dashed lines. We also obtained the empirical distributions based on the real fingerprint impressions, and based on the uniform distribution (no clustering tendencies) for the centers and deviations (denoted by the dot-dashed lines with circles and solid lines, respectively). Note the close agreement between the impostor distributions of the synthesized and empirical databases, thus, demonstrating the adequateness of the compound models in representing the distribution of minutiae in the two databases.

Fingerprint individuality estimates are given in Table 1 by pooling the impostor matches from the FVC2002 DB1 and IBM-HURSLEY databases ($H = 40$ synthetic impressions per finger were used). Table 2 gives the fingerprint individuality estimate from our model for the "12-point match criteria" based on FVC 2002 DB1 database. For comparison purposes, the fingerprint individuality estimates derived using the methodology of [8] is also given. One important point is that [8] computes the individuality estimate based on the number of minutiae in the query and template that occur in the *overlap* area. In this paper, m_Q and m_T represent the total number of minutiae in a query and template, respectively. In order to make valid comparisons, we found the mean number of minutiae occurring in the overlap area in our simulations; when $m_Q = m_T = 26$, this mean was approximately 17. Consequently, the estimate 2.4×10^{-15}

was calculated using the formula in [8] based on the combination $(17, 17, 12)$. Note that our estimates are orders of magnitude higher compared to [8]. We believe our model gives a more realistic estimate of fingerprint individuality compared to [8] as seen from the impostor matching distributions in Figure 4.

5 Summary and Conclusion

A compound stochastic model is developed for representing three sources of minutiae variability in fingerprint images. These models are used for synthesis and for obtaining estimates of fingerprint individuality. Our future work involves obtaining individuality estimates for larger databases and deriving a mathematical framework for calculating the probability associated with fingerprint individuality.

References

[1] S. Cole. Witnessing identification: Latent fingerprint evidence and expert knowledge. *Social Studies in Science*, 28(5–6):687–712, 1998.

[2] S. Cole. What counts for identity? The historical origins of the methodology of latent fingerprint identification. *Science in Context*, 27(1):139–172, 1999.

[3] S. Cole. *A History of Fingerprinting and Criminal Identification*. Harvard University Press, 2001.

[4] S. C. Dass, Y. Zhu, and A. K. Jain. Statistical models for assessing the individuality of fingerprints. *Fourth IEEE Workshop on Automatic Identification Advanced Technologies*, pages 1–7, 2005.

[5] Daubert v. Merrel Dow Pharmaceuticals Inc, 509 U.S. 579, 113 S. Ct. 2786, 125 L.Ed.2d 469 (1993).

[6] R. Halir and J. Flusser. Numerically stable direct least square fitting of ellipses. *Proc. of the 6th International Conference in Central Europe on Computer Graphics, Visualization and Interactive Digital Media*, pages 125–132, 1998.

[7] K. V. Mardia. *Statistics of Directional Data*. Academic Press, 1972.

[8] S. Pankanti, S. Prabhakar, and A. K. Jain. On the individuality of fingerprints. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1010–1025, 2002.

[9] A. Ross, S. Dass, and A. K. Jain. A deformable model for fingerprint matching. *Pattern Recognition*, 38(1):95–103, 2005.

[10] U. S. v. Byron Mitchell. Criminal Action No. 96-407, U. S. District Court for the Eastern District of Pennsylvania, 1999.