STATISTICAL MODELS FOR ASSESSING THE INDIVIDUALITY OF FINGERPRINTS

Sarat C. Dass

Joint work with Yongfang Zhu and Anil Jain

Michigan State University
Fingerprint Features

Global Features: (1) Ridge-valley structures

(2) Singular points—Core, Delta

Local Features: (1) Minutiae—Ridge endings and bifurcations

A minutiae is represented by its location $X = (x, y)$ and direction $D \in [0, 2\pi)$
Fingerprint Matching

Fingerprint matching finds the degree of similarity between two fingerprints in terms of their feature sets.

Let query Q be from person with claimed identity \( I_c \). Template T is retrieved from the database based on \( I_c \).

Test of hypotheses: \( H_0 : I_u \neq I_c \) vs. \( H_1 : I_u = I_c \)

where \( I_u \) is the true (and unknown) identity of the person.

Test statistic: \( W = \) the number of matched minutiae pairs between Q and T (in terms of location and direction).

Decision: For selected threshold \( w_0 \),

- If \( W \geq w_0 \), reject \( H_0 \).
- If \( W < w_0 \), accept \( H_0 \).

**What is the statistical confidence associated with the observed match?**
Fingerprint Individuality

Fundamental Question:

Given a query fingerprint, what is the probability of finding a sufficiently similar fingerprint in a target population?

Goals:

• Assess the likelihood that the query will share the observed no. of features with an arbitrary fingerprint from the target population.

• Need statistical models that adequately represent the distribution (or, variability) of fingerprint minutia in a population of individuals.

• Need an explicit numerical measure of confidence associated with the observed match
Motivation for fingerprint individuality

• Fingerprint individuality has been the highlight of several court cases in the U.S.
 ﬀ USA vs. Daubert (1993) : The court ruled that evidence based on biometric traits should be validated based on scientific principles.
 ﬀ USA vs. Byron Mitchell (1999) : Evidence based on fingerprints was first challenged in this case.
 ﬀ After the USA vs. Byron Mitchell case, fingerprint-based evidence has been challenged in more than 20 cases in the United States.

• Error rates associated with fingerprint matching are unknown.

The problem of individuality is also relevant to identification/verification systems based on biometric traits other than fingerprints.
Challenges in Assessing Fingerprint Individuality

1. Genuine fingerprints exhibit large intra-class variability

   ![Images of two genuine fingerprints](image1)

   Two impressions of the same finger that appear different from each other

2. Impostor fingerprints exhibit small inter-class variability

   ![Images of two impostor fingerprints](image2)

   Impressions from two different fingers that have similar ridge patterns

3. The quality of fingerprint impressions can vary significantly

   ![Images of different quality fingerprints](image3)

   Different quality of fingerprints. Assessing feature matches based on (c) can be difficult.
Literature on fingerprint individuality

(A) Galton (1892):

\[ P(\text{Fingerprint Configuration}) = \frac{1}{16} \times \frac{1}{256} \times \left( \frac{1}{2} \right)^{24} = 1.45 \times 10^{-11} \]

(B) Pearson (1933):

\[ P(\text{Fingerprint Configuration}) = \frac{1}{16} \times \frac{1}{256} \times \left( \frac{1}{36} \right)^{24} = 1.09 \times 10^{-41} \]

(C) Henry (1900), Balthazard (1911), Gupta (1968), and others:

\[ P(\text{Fingerprint Configuration}) = p^N \]

Drawback: These models were not validated on fingerprint databases.

(D) Pankanti et al. (2002): Proposed stochastic models which were validated on fingerprint databases based on a corrected uniform model for the minutiae features.
Mixture Models on \((X,D)\)

Let \(\{(X_i,D_i)\}_{i=1}^{n}\) be \(n\) minutiae locations and directions of a fingerprint. Each minutiae follows the density of a \(G\)-component mixture model

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

- \(f_g^X\) is a bivariate normal density;
- \(\tau = (\tau_1, \tau_2, ..., \tau_G)\) are weights summing to 1;
- \(f_g^D(\theta|\nu_g, \kappa_g, p_g)\) is the Von-Mises distribution and \(p_g\) is a number between 0 and 1.
Mixture Model Fitting And Parameter Estimation

- Convert the minutiae directions \(D_j\) to orientations \(O_j\) to obtain a standard mixture model and estimate all parameters except \(p_g\).

\[
\sum_{g=1}^{G} \tau_g f_g^X (X_j | \mu_g, \Sigma_g) \cdot f_g^O (\omega_j | \nu_g, \kappa_g)
\]

- In the \((k+1)\)-th iteration, \(p_g\) corresponding to component \(g\) is estimated as

\[
p_{g}^{k+1} = \frac{\sum_{j=1}^{k} I\{c_j^{k+1} = g, D_j \in [0,\pi)\}}{\sum_{j=1}^{k} I\{c_j^{k+1} = g\}}
\]

- BIC is used to estimate the number of components in the mixture model

\[
BIC(G) = 2 \sum_{j=1}^{n} \log f(X_j, D_j | \Theta_G) - |\Theta_G| \log(n)
\]

where \(|\Theta_G| = \# of model parameters\)
Examples Of Model Fitting

Fingerprint minutiae

Minutiae Clusters

Clusters in 3-D Space
Proposed model vs. uniform model
Extension To A Target Population: Identifying Clusters of Mixture Models

- Clustering Procedure: Hierarchical.

- Dissimilarity measure: Hellinger distance

\[
H(f, g) = \int_{x \in S} \int_{\theta \in [0,2\pi]} \left( \sqrt{f(x, \theta)} - \sqrt{g(x, \theta)} \right)^2 \, dx \, d\theta.
\]

- Elbow criteria used to determine the number of clusters:
  For j clusters \( C_1, C_2, \ldots, C_j \) given by the algorithm, define \( V_j \) as

\[
V_j = \frac{1}{j} \sum_{i=1}^{j} \frac{H(f, g)}{2|C_i|}
\]

- Let \( N^* \) be the number after which \( V_j \) does not change significantly. Choose \( B = N^* \) as the optimal number of mixture clusters in the population.
Computing the probability of a match

Assume \((X_Q, D_Q)\) and \((X_T, D_T)\), respectively, are minutiae from the query Q and template T with distributions \(H_Q\) and \(H_T\). Define the **probability of a match**, \(p(Q,T)\), as

\[
p(Q,T) = P\{(X_Q, D_Q) \text{ matches with } (X_T, D_T)\} = P\{|X_Q - X_T| < r_0, |D_Q - D_T|_d < d_0\}
\]

- \(r_0\) and \(d_0\) are matching thresholds.
- Define direction difference as

\[
|D_Q - D_T|_d = \min\left(|D_Q - D_T|, 2\pi - |D_Q - D_T|\right)
\]
Probability of a random correspondence (PRC)

- Analytical formula for the probability of exactly w matches is given by

\[ P(W = w \mid Q, T, m, n) = \text{Poisson}(w, \lambda(Q,T)) \]

where \( \lambda(Q,T) = m n p(Q,T) \). If \( Q \in C_b \) and \( T \in C_{b'} \), for \( b, b' = 1, 2, \ldots, B \), then \( p(Q,T) = p(C_b, C_{b'}) \).

- Overall probability of exactly w matches

\[ p^*(w) = \prod_{1 \leq b \neq b' \leq B} \pi_b \pi_{b'} P(W = w, C_b, C_{b'}). \]

- PRC = p-value = \( p^*(w_{\text{obs}}) \).
Some details on the derivation of the Poisson

• Q and T are minutiae sets of size m and n from clusters C and C’;
• The probability that T is matched with the first w minutiae of Q is

\[
\begin{align*}
&\frac{u_1}{1-u_1} \quad \frac{u_2}{1-u_1-u_2} \quad \frac{u_3}{1-u_1-u_2-u_3} \quad \cdots \quad \frac{u_w}{1-u_1-u_2-\cdots-u_w} \\
\frac{1}{w} \prod_{i=1}^{w} (1-u_i) &\quad \frac{1}{w} \prod_{i=1}^{w} (1-u_i-v_{w+1}) \quad \cdots \quad \frac{1}{w} \prod_{i=1}^{w} (1-u_i-v_{w+1}-\cdots-v_{j})
\end{align*}
\]

where

(1) \(u_i\) is the probability that a random minutiae from T, \((X_T, D_T)\), is matched with minutiae \((X_i, D_i)\) from Q, and

(2) \(v_j\) is the corresponding probability that the random minutiae, say \((X_T, D_T)\), from T is matched with minutiae \((X_j, D_j)\) from T.

• For large \(m\) and \(n\), this is approximately

\[
\prod_{i=1}^{w} u_i \exp\{ -\lambda(C, C') \}
\]
Recall: We want to calculate $P(W=w | C_i, C_j, m, n)$

- There are $\frac{n!}{(n-w)!}$ ways of $w$ minutiae from the template matching with the first $w$ minutiae of the query;

- There are $m \choose w$ ways to choose $w$ minutiae from query to match with minutiae in template;

The probability of exactly $w$ matches is

$$P(W(C_i, C_j) = w | m, n, H_0) = \frac{n!}{(n-w)!} m \choose w \sum_{i_1<i_2<...<i_w} u_{ik} e^{-\lambda(C_i, C_j)}$$

- When $m$ and $n$ are large, this is approximately

$$\text{Poisson}(w, \lambda(C_i, C_j)) = \frac{e^{-\lambda(C_i, C_j)} \lambda(C_i, C_j)^w}{w!}$$
Assessment of Fingerprint Individuality

The databases:

<table>
<thead>
<tr>
<th>Database</th>
<th>(m,n,w)</th>
<th>B</th>
<th>Mean fingerprint area</th>
<th>Mean λ</th>
<th>PRC</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIST (1,998)</td>
<td>(62,62,12)</td>
<td>33</td>
<td>2.5x10^5</td>
<td>3.45</td>
<td>4.1x10^-4</td>
</tr>
<tr>
<td>FVC2002 DB1 (100)</td>
<td>(63,63,12)</td>
<td>9</td>
<td>1.2x10^5</td>
<td>5.14</td>
<td>5.9x10^-3</td>
</tr>
<tr>
<td>FVC2002 DB2 (100)</td>
<td>(77,77,12)</td>
<td>12</td>
<td>1.8x10^5</td>
<td>5.20</td>
<td>8.4x10^-3</td>
</tr>
</tbody>
</table>

Comparison of PRCs:

<table>
<thead>
<tr>
<th>Database</th>
<th>(m,n,w)</th>
<th>Mixture Mean λ</th>
<th>Mixture PRC</th>
<th>Uniform Mean λ</th>
<th>Uniform PRC</th>
<th>Empirical Mean λ</th>
<th>Empirical PRC</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIST</td>
<td>(62,62,12)</td>
<td>3.46</td>
<td>4.1x10^-4</td>
<td>1.70</td>
<td>2.9x10^-7</td>
<td>7.1</td>
<td>3.4x10^-3</td>
</tr>
<tr>
<td>FVCDB1</td>
<td>(63,63,12)</td>
<td>5.14</td>
<td>5.9x10^-3</td>
<td>3.04</td>
<td>1.0x10^-4</td>
<td>8.0</td>
<td>1.4x10^-2</td>
</tr>
<tr>
<td>FVCDB2</td>
<td>(77,77,12)</td>
<td>5.20</td>
<td>8.4x10^-3</td>
<td>2.98</td>
<td>1.8x10^-5</td>
<td>8.6</td>
<td>1.9x10^-2</td>
</tr>
</tbody>
</table>

The PRCs from mixture model are higher than those derived from the uniform model, and closer to the empirical.
Assessment of Fingerprint Individuality

Comparison between the mixture and Pankanti’s models

Based on the overlapping area between Q and T

<table>
<thead>
<tr>
<th>Database</th>
<th>(m,n,w)</th>
<th>Mixture Mean λ</th>
<th>Mixture PRC</th>
<th>Pankanti Mean λ</th>
<th>Pankanti PRC</th>
<th>Empirical Mean λ</th>
<th>Empirical PRC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.3</td>
<td>4.4x10^{-3}</td>
<td></td>
<td>1.8</td>
<td>4.3x10^{-8}</td>
<td>7.1</td>
<td>3.9x10^{-3}</td>
</tr>
<tr>
<td>NIST</td>
<td>(52,52,12)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FVCDB1</td>
<td>(51,51,12)</td>
<td>4.9</td>
<td>1.1x10^{-2}</td>
<td>2.8</td>
<td>4.1x10^{-6}</td>
<td>8.0</td>
<td>2.9x10^{-2}</td>
</tr>
<tr>
<td>FVCDB2</td>
<td>(63,63,12)</td>
<td>4.8</td>
<td>1.1x10^{-2}</td>
<td>2.7</td>
<td>4.3x10^{-6}</td>
<td>8.6</td>
<td>6.5x10^{-2}</td>
</tr>
</tbody>
</table>

The PRCs from mixture model are, again, higher than those derived from the Pankanti’s model.

This is due to similar clustering tendencies of minutiae in different fingerprints.