Automatic Latent Fingerprint Value Prediction

Tarang Chugh †, Kai Cao †, Jiayu Zhou †, Elham Tabassi ‡ and Anil K. Jain †

† Michigan State University
‡ National Institute of Standards and Technology

Aug. 11, 2016
What are Latent Fingerprints?

Rolled  Plain  Latent
Challenges in Latent Matching

- Poor Ridge Clarity
- Partial Ridge Area
- Complex Background

• AFIS Performance (Rank-1 accuracy)
  - Plain: 98.5%
  - Latent: 67.2% (70.2% with image + markup)

Latent Matching: ACE-V Protocol

Value for Individualization (VID)
Value for Exclusion Only (VEO)

Value Determination

No Value (NV)
Retain/Discard

Feature Markup

Comparison

AFIS

Candidate List

C1 C2 C3 Ck...

Decision

Evaluation

Verification by second examiner

Individualization Exclusion Inconclusive

Fingerprint examiner
Latent Fingerprint Value

• Presumably based on two factors:
  – Image quality (e.g., clarity of friction ridges)
  – Information content (e.g., no. of minutiae)

VID
(Value for Identification)

VEO
(Value for Exclusion Only)

NV
(No Value)
Image Quality v. Information Content

- Low Image Quality, Low Information Content: Predicted Value: 2.65
- Low Image Quality, High Information Content: Predicted Value: 2.95
- High Image Quality, Low Information Content: Predicted Value: 2.65
- High Image Quality, High Information Content: Predicted Value: 2.95
Limitations of Examiner Value Determination

• Highly subjective
  • repeatability (intra-examiner variability): 84.6%
  • reproducibility (inter-examiner similarity): 75.2%

• Depends upon examiner’s skill and experience

• Time-consuming

Need for automatic value determination

Proposed Method

Expert Crowd

Crowdsourcing

Quality

Information content

Matrix Completion

Multidimensional Scaling

Minutiae-based features

Ridge-based features

Singularity-based features

Feature extractor

Underlying basis

Latent value prediction model

Lasso

Crowdsourced latent value

Predicted latent value

Off-line learning

On-line prediction

Training latents

Test latent
# Features for Value Assessment

<table>
<thead>
<tr>
<th>Feature No.</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>Number of minutiae</strong></td>
</tr>
<tr>
<td>2 - 8</td>
<td>Sum of <strong>minutiae reliability</strong> with reliability ( \geq t ), ( t = 0, 0.1, ..., 0.6 )</td>
</tr>
<tr>
<td>9</td>
<td><strong>Average area</strong> of minutiae Delaunay triangulation</td>
</tr>
<tr>
<td>10</td>
<td>Area of the convex hull of minutiae set</td>
</tr>
<tr>
<td>11 - 17</td>
<td>Sum of <strong>ridge quality</strong> blocks with quality value ( \geq t ), ( t = 0, 0.1, ..., 0.6 )</td>
</tr>
<tr>
<td>18</td>
<td><strong>Number of singular points</strong> (core and delta)</td>
</tr>
<tr>
<td>19</td>
<td><strong>Standard deviation of the ridge flow</strong> in the foreground</td>
</tr>
</tbody>
</table>
Feature Extraction

Input Latent with ROI → Ridge Flow Estimation → Normalization
Minutiae Extraction ← Ridge Quality Estimation ← Ridge Enhancement
Minutiae and Singular Points Extraction

Latent ID: G007
Latent ID: B106
Latent ID: U228

- Manually marked minutiae
- Automatic extracted minutiae
- Automatic detected core point
FingerprintMash: A Crowdsourcing Tool

Welcome user
Indicate (a) quality of each latent, and (b) which one (left or right) has more information for identification
Click on any image to zoom

Quality
High
Low

Information content
Left latent has much more
slightly more
similar
slightly more
Right latent has much more

Undo Previous
Submit
Skip to Next

24/100 Completed

Time elapsed: 215 second(s)

http://fingerprintmash.org
Crowdsourcing Details

• **Expert Crowd**
  • 31 experts (latent examiners and researchers)

• **Dataset**
  • 258 latents from NIST SD27
  • 258 latents from the Michigan State Police (MSP)

• **Protocol**
  • 5 levels for numerical quality rating; 1: low, 5: high
  • 5 relative levels for information content; 1: left is better
  • 100 randomly selected pairs presented to each expert
  • 1 pair is repeated at every 5th comparison for validity
Matrix Completion
Numerical Quality

Partially observed quality matrix $Q$

\[
\begin{pmatrix}
  l_1 & l_2 & l_3 & \ldots & l_m \\
  u_1 & 3 & - & 2 & \ldots & - \\
  u_2 & - & 3 & 2 & \ldots & - \\
  u_3 & 5 & - & - & \ldots & 2 \\
  \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
  u_n & - & 4 & - & \ldots & 2 \\
\end{pmatrix}
\]

Completed quality matrix $\hat{Q}$

\[
\begin{pmatrix}
  l_1 & l_2 & l_3 & \ldots & l_m \\
  u_1 & 2.9 & 3.1 & 1.9 & \ldots & 0.9 \\
  u_2 & 4.2 & 3.2 & 2.1 & \ldots & 1.1 \\
  u_3 & 5.0 & 3.5 & 2.1 & \ldots & 2.2 \\
  \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
  u_n & 5.0 & 3.9 & 2.0 & \ldots & 1.9 \\
\end{pmatrix}
\]

Matrix Completion
Pairwise Information Content

\[
\begin{bmatrix}
1 & - & \ldots & 1 \\
-1 & 1 & \ldots & - \\
-1 & -1 & \ldots & - \\
\vdots & \vdots & \ddots & \vdots \\
-1 & 1 & \ldots & 1 \\
\end{bmatrix}
\]

Partially observed pairwise information content \( C \)

\[
\begin{bmatrix}
2.7 & 2.1 & 3.6 & \ldots & 0.7 \\
3.2 & 3.3 & 3.5 & \ldots & 2.1 \\
4.0 & 2.5 & 3.4 & \ldots & 3.2 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
3.5 & 3.9 & 2.4 & \ldots & 1. \\
\end{bmatrix}
\]

Completed information content \( \hat{C} \)

Latent value \( V = (\hat{Q} + \hat{C})/2 \)

Yi et al., Inferring User’s Preferences from Crowdsourced Pairwise Comparisons: A Matrix Completion Approach, In HCOMP, 2013
Intra-Expert Variance

Numerical rating

Pairwise comparison
Intra-Expert Variance

**Numerical Rating**
- Mean = 0.35; Std. Dev. = 0.16

**Pairwise Comparison**
- Mean = 0.24; Std. Dev. = 0.15

**Low variance**
Values = [4,4,4,4]

**High variance**
Values = [1,2,2,4]

**High variance in comparison**
Value ranged from 1 to 4
Inter-Expert Variance

Examples of strong agreement among experts

Examples of strong disagreement among experts

Low Value (1)  High Value (5)  Value Range (2-5)  Value Range (1-5)
Inter-Expert Variance

Visualization of all 30 experts in a two-dimensional MDS space

numerial ratings  pairwise comparisons
Expert Crowd v. Value Determination

- NIST SD27: 210 VID, 41 VEO, and 7 NV from [4]

Rank-1 hit rates from AFIS; 250K reference prints

<table>
<thead>
<tr>
<th>Value Determination</th>
<th>VID</th>
<th>VEO</th>
<th>NV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>163/210</td>
<td>12/41</td>
<td>0/7</td>
</tr>
<tr>
<td>Expert Crowd</td>
<td>High Value</td>
<td>Medium Value</td>
<td>Low Value</td>
</tr>
<tr>
<td>(Median Value)</td>
<td>170/210</td>
<td>6/41</td>
<td>0/7</td>
</tr>
</tbody>
</table>

For a fair comparison, we identify, from crowdsourced value data, the top 210 valued latents as high value, next 41 as medium value and the remaining 7 as low value.

Multidimensional Scaling Analysis
## Underlying Bases

<table>
<thead>
<tr>
<th>Feature No.</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Number of minutiae</td>
</tr>
<tr>
<td>2 - 8</td>
<td>Sum of <em>minutiae reliability</em> with reliability $\geq t$, $t=0, 0.1, \ldots, 0.6$</td>
</tr>
<tr>
<td>9</td>
<td>Average area of minutiae Delaunay triangulation</td>
</tr>
<tr>
<td>10</td>
<td>Area of the convex hull of minutiae set</td>
</tr>
<tr>
<td>11 - 17</td>
<td>Sum of <em>ridge quality</em> blocks with quality value $\geq t$, $t=0, 0.1, \ldots, 0.6$</td>
</tr>
<tr>
<td>18</td>
<td>Number of <em>singular points</em> (core and delta)</td>
</tr>
<tr>
<td>19</td>
<td>Standard deviation of the ridge flow in the foreground</td>
</tr>
</tbody>
</table>
Value Prediction

• Learned a predictor for value assignment
• Average MSE for predicted value = 0.24

Examples of correctly predicted latent value

Examples of incorrectly predicted latent value

Crowd : 4.67
Predicted : 4.68

Crowd : 2.08
Predicted : 2.04

Crowd : 2.99
Predicted : 4.02

Crowd : 2.96
Predicted : 2.02
Value Prediction

Source: MSP
Rank Retrieval: 1
Predicted Value: 4.96

Source: MSP
Rank Retrieval: 1
Predicted Value: 4.62

Source: MSP
Rank Retrieval: 1
Predicted Value: 4.23

Source: NIST SD27
Rank Retrieval: 1
Predicted Value: 4.19

Source: NIST SD27
Rank Retrieval: 1
Predicted Value: 4.02

Source: NIST SD27
Rank Retrieval: 14,806
Predicted Value: 1.37

Source: MSP
Rank Retrieval: 11
Predicted Value: 1.25

Source: NIST SD27
Rank Retrieval: 22,442
Predicted Value: 1.19

Source: NIST SD27
Rank Retrieval: 21,662
Predicted Value: 1.08

Source: MSP
Rank Retrieval: 50,980
Predicted Value: 1.01
High Latent Fingerprint Value
Low AFIS performance

Predicted Value: 3.34
AFIS Retrieval Rank: 18,789

Predicted Value: 2.95
AFIS Retrieval Rank: 19,456
Summary

• Developed a crowdsourcing-based framework for understanding expert latent value assignment from the perspectives of latent quality and information content.

• Used MDS to identify the underlying bases for expert latent value assignment.

• Predicted quantitative latent value based on the underlying bases.

• **Wisdom of crowd** leads to better decision making over a single examiner.
Thank You

Any Questions?