Footwear Impression Research at NIST

Martin Herman
Information Technology Laboratory
NIST
November 7, 2018
MOTIVATION

• 2009 NAS; 2016 PCAST:
  – Footwear identifications are largely subjective
  – Questions about reliability
  – Questions about scientific validity
  – Need for quantitative assessments of footwear evidence
  – Need for increased empirically-tested objectivity of footwear analysis

• Need to improve quantitative analysis
• Need for algorithmic approaches for the forensic footwear community
GOALS

• Develop quantitative, objective methods for footwear impression comparisons
  – High degree of repeatability & reproducibility
  – Easier to measure accuracy with objective methods
• High performance – good discrimination power
• Provide prototype software tools to be evaluated for the following purposes
  – Use by practitioners in casework
  – Use by researchers to develop algorithms
SHOECALC

• A prototype system for footwear analysis that will allow
  – Researchers/developers to have a workbench for development of quantitative methods
  – Examiners to use these quantitative methods during casework
  – Development of this system is a long-term goal
Database consisting of:
- Real & staged crime scene impressions and metadata
- Catalogue of outsole designs and metadata
- Test impressions from shoes of arrestees or research volunteers
- Catalogue of acquired characteristics (RACs) along with shape, size, location, brand, outsole design, etc.
- Interfaces and formats for submitting and maintaining footwear data
Synthetic/augmented footwear impressions
- For research and testing, generates synthetic footwear impressions with user specified characteristics and with ground truth known
- Characteristics include outsole designs, wear amounts, sizes, and distributions of RACs; different matrix/substrate combinations
- Synthetic test & crime scene impressions
- Augmented data for research/tests
Comparison Measures
- A workbench for experimentation with different comparison scores. Some scores lead to better discrimination between mated and non-mated pairs of images than others.
- User inputs a function for computing a comparison score and applies it to any given pair of images; numerical score is reported.
- Also uses SHOEGULI to conduct experiments and produce ROC charts for comparing with a catalog of known, high performance comparison scores.
Quality Measures

- Measuring different characteristics that describe the degradation, distortion, completeness, number of features in the impression
- Input is any footwear image; output is a list of quality metrics
- May be used as a workbench for experimentation with different image quality metrics
• GUI for user interaction with the other modules of SHOECALC
• Allows user to upload images for calculation of comparison and quality scores
• Examine various choices of comparison metrics, scores and their ROC charts, and select choices for reporting the information in the evidence
• Exploratory analysis of data, charts, etc.
Today’s talks

1. Towards an end–to–end system for quantitative footwear impression comparisons – Martin Herman
   - End-to-end prototype system for use by examiners during casework

2. Image Alignment and Feature Extraction for Shoeprint Matching – Gautham Venkatasubramanian
   - As part of end-to-end system, alignment of questioned and known impressions, along with feature extraction to be used for image matching

3. Deep Learning based Feature Extractors for Shoeprint Matching – Sarala Padi
   - As part of end-to-end system, features learned in a DNN model are used for image matching

4. Matching Randomly Acquired Characteristics (RACs) in Footwear Impressions – Weiqing Chen
   - As part of the end-to-end system, RAC features are extracted and matched
Towards an End-to-End System for Quantitative Footwear Impression Comparisons

Presented by:

Martin Herman

Other Core Team Members:

Hari Iyer, Steve Lund, Gunay Dogan, Yooyoung Lee

Information Technology Laboratory, NIST

November 7, 2018
Use of SHOECALC: Quantitative Footwear Impression Comparisons

- For use by examiners in evidence evaluation
- FRStat for fingerprints (U.S. Defense Forensic Science Center) – currently in use

**Current Examiner Comparison Process**

1. Crime Scene Impressions
2. Test Impressions
3. Suspect Shoe

COMPARISON

Conclusion plus Report
Proposed Examiner Comparison Process

COMPARISON – Examiner Considers Additional Information: Comparison Scores, Context ( Relevant Reference Collection)

Crime Scene Impressions
Test Impressions
Suspect Shoe

Conclusion plus Report
Elements of the Comparison Score

• Features considered in total score
  – Shoe size
  – Outsole design features
  – Wear features
  – RACs

• Transparency for examiner
  – Examiner should be able to understand how the score is related to features above
  – Our goal is for the examiner to be able to relate the score to SWGTREAD “Range of Conclusions Standard”
Workflow for End-to-End Scoring System

Step 1: Image Alignment
Step 2: Feature Comparison
Step 3: RAC Comparison
Step 4: Final Score Computation

Test

Questioned
Step 1: Image Alignment

- Determine best alignment of the two impression images (Questioned and Test)
- If images do not align well, then (optional) SCORE < 0 & STOP

Described in talk later in session.
Step 2: Feature Comparison

• Compares features based mainly on combination of design, wear and size. RAC features play only very small part.
• Score considers combined features inside a Region of Interest

Described in talk later in session.
Step 3: RAC Comparison

- Compute score based on RACs
- RACs marked on test impression by examiner, then transformed to questioned impression after alignment.
- Then corresponding patches are compared.
- No marking of RACs in questioned impression.

Described in talk later in session.
Step 4: Computing Final Score

- Final score is combination of feature comparison & RAC comparison scores
  - Goal is to relate the individual feature and RAC scores, plus final score, to SWGTREAD conclusions scale
- The final score is computed using reference dataset of ground-truth-known mates and non-mates.
- Composite RAC score = combined Score-based Likelihood Ratio (SLR) of individual RAC SLRs
- Final score = SLR obtained from bivariate density of composite RAC score and feature score
End-to-End Score Computation: Examples
Example 1
Close Non-Match (left shoe flipped)
Everspry EverOS Scanner

Questioned 1

Test1
Example 2
Known Match

Questioned 2

Test 1
Example 3

Known Match
Dust Impression (Jacqueline Speir, WVU)
Comparison Scores

Feature Comparison Scores

<table>
<thead>
<tr>
<th></th>
<th>Q1 vs Test1</th>
<th>Q2 vs Test1</th>
<th>Q3 vs Test3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1 vs Test1</td>
<td>0.8771</td>
<td>0.8009</td>
<td>0.8760</td>
</tr>
</tbody>
</table>

RAC Comparison Scores

<table>
<thead>
<tr>
<th>Rac No.</th>
<th>Q1 vs Test1</th>
<th>Q2 vs Test1</th>
<th>Q3 vs Test3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0141</td>
<td>0.6002</td>
<td>0.2173</td>
</tr>
<tr>
<td>2</td>
<td>0.2042</td>
<td>0.5177</td>
<td>0.4651</td>
</tr>
<tr>
<td>3</td>
<td>0.0467</td>
<td>0.1392</td>
<td>0.3483</td>
</tr>
<tr>
<td>4</td>
<td>0.2992</td>
<td>0.5813</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.5409</td>
<td>0.8777</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.1849</td>
<td>0.2387</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.5997</td>
<td>0.6835</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.0938</td>
<td>0.7494</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.0272</td>
<td>0.7950</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.3495</td>
<td>0.7741</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0.3681</td>
<td>0.7558</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.1567</td>
<td>0.6828</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>0.4892</td>
<td>0.6302</td>
<td></td>
</tr>
</tbody>
</table>

Final Comparison Scores

<table>
<thead>
<tr>
<th></th>
<th>Q1 vs Test1</th>
<th>Q2 vs Test1</th>
<th>Q3 vs Test3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1 vs Test1</td>
<td>0.3831</td>
<td>0.6540</td>
<td>0.4501</td>
</tr>
</tbody>
</table>

Close non-match score is greater than match score. But scores are very close.

Q1 – close non-match; Q2 – known match; Q3 – known match
Context for Interpreting Comparison Scores - 1

• How do we determine what significance to give to any particular score?
• Answer: evaluate the score in the context of ground-truth-known mated and non-mated pairs that are representative of impressions obtained under conditions similar to the current crime scene.
  – E.g., same quality and quantity of information
• Provide context for
  1. Feature comparison score – size, design, wear
  2. RAC score
  3. Final score
- A score that lies mainly within mated pair scores indicates strong support for a match proposition.
- A score that lies mainly within mated pair scores indicates strong support for a match proposition.
- A score that lies mainly within non-mated pair scores indicates strong support for a non-match proposition.
- Scores that represent significant overlap of mated and non-mated pairs support neither proposition.
- A score that lies mainly within mated pair scores indicates strong support for a match proposition.
- A score that lies mainly within non-mated pair scores indicates strong support for a non-match proposition.
- A score that occurs nearly equally often among mated and non-mated pairs does not provide support for either proposition.
Thank You

martin.herman@nist.gov