A semi-automated algorithm to quantify the similarity between shoe outsole impressions

Soyoung Park & Alicia Carriquiry

CSAFE / Iowa State University

Center for Statistics and Applications in Forensic Evidence

National Institute of Standards and Technology
U.S. Department of Commerce

1. Introduction

- The comparison of shoe print images still largely relies on an expert's subjective assessment and experience.
- Crime scene (Q) vs. known out-sole test impression (K).

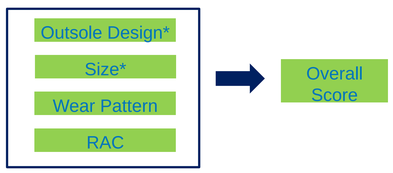
(A) Partial shoe print fount at crime scene (Q)

(B) Putative source shoe (K)

- Challenges : (1) Images lifted from crime scenes are typically partial prints, (2) For comparison, questioned image (Q) and known prints (K) need to be rotated and translated, (3) Images are subject to noise and background effects.

2. Goal

- Goal : Develop high performing scoring system(s) that quantify the degree of similarity between two out-sole 2D images.
- Contributors to the score :



3. Methods I

- Points of interest ("data") : Coordinates of all edges (Prewitt operator) and of 500 strong corners (the Harris-Stephens algorithm) in a 2D out-sole image.
- Maximum Clique (MC) : Largest subset of attributes in one image that is geometrically congruent to a corresponding collection of attributes in the other image. The advantage of using MC is its invariance to rotation and translation.

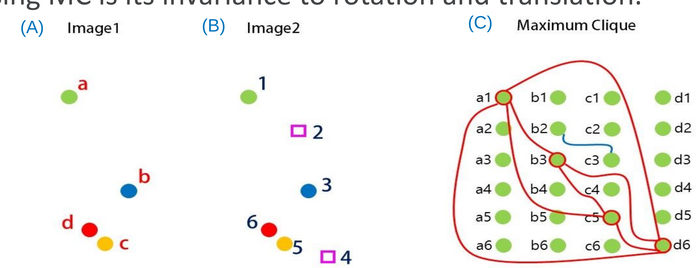
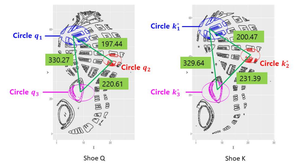


Figure (A) and (B) are example images. Image 1 has a set of points {a, b, c, d} and image 2 has a set of points {1, 2, 3, 4, 5, 6}. Figure (C) is application of the concept of maximum clique to find the correspondence between image 1 and image 2. The set of maximum clique for image 1 and image 2 is {(a,1), (b,3), (d,6), (c,5)}.

4. Methods II

- Signature of an out-sole image : We select three semi-arbitrary circular regions and the edge points within.



- Matching : After alignment of edges in Q and K, fix several circular areas in edges of Q and find the best matching area in K using MC.
- Similarity features : (1) Clique size, (2) SD of rot. angle, (3) % Overlap on Q, (4) % Overlap on K, (5) Med. Distance of OP, (6) Diff. in triangle Q-K.
- Score : Multiple similarity features are extracted and combined into single score using a random forest algorithm.

5. Matching Algorithm (MC-SCA)

- Step 1 : Extract edges from shoe Q and shoe K.
- Step 2 : Fix three circles ($\boldsymbol{q_1}$,$\boldsymbol{q_2}$,$\boldsymbol{q_3}$) in "interesting" areas of shoe Q.
- Step 3 : Search over several circular areas in shoe K, and find the circle with highest overlap with circle $\boldsymbol{q_1}$.
- Step 4 : From the circle found at step 3, find the corresponding circle $\boldsymbol{k_1^*}$ n shoe K that is best overlap with $\boldsymbol{q_1}$ in shoe Q.
- Step 5 : Repeat the entire process for at least two additional circles $\boldsymbol{q_2}$,$\boldsymbol{q_3}$ in shoe Q and find corresponding $\boldsymbol{k_2^*}$, $\boldsymbol{k_3^*}$.

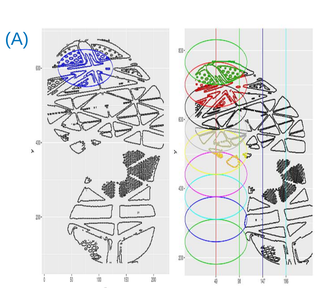


Figure (A) is comparison between shoe Q and shoe K in step 2 and 3. When fixing blue area in shoe Q, many candidate circles are searched over shoe K.

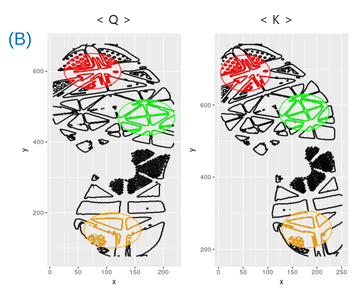


Figure (B) shows a final matching results in step 5. When fixing three sub areas in shoe Q, the corresponding circles in shoe K are found.

6. Data Analysis and results

- Data : CSAFE obtained replicate 2D images from 38 pairs of NIKE Winflow shoes size 8.5 and 22 pairs size 10.5. Shoes were used for about 6 months.
- Comparisons : There are 717 known mated (KM) comparisons and 600 known non-mated (KNM) comparisons.

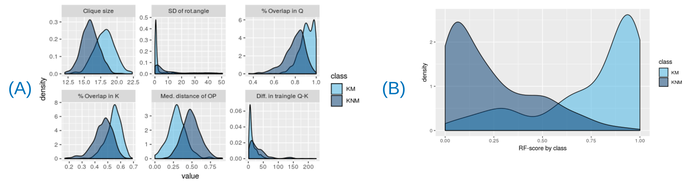
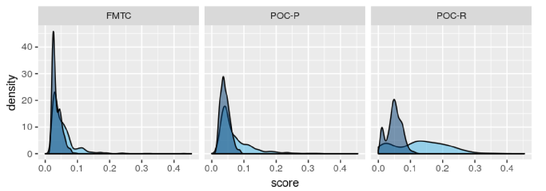


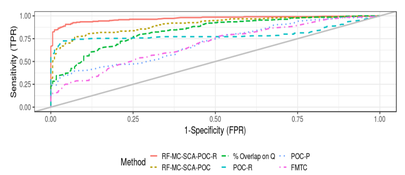
Figure (A) shows densities of class KM and KNM in six similarity features. Figure (B) is a final score predicted by RF on test comparisons.

- Methods : MC-SCA, POC with registration method, POC with principal axis and FMTC.



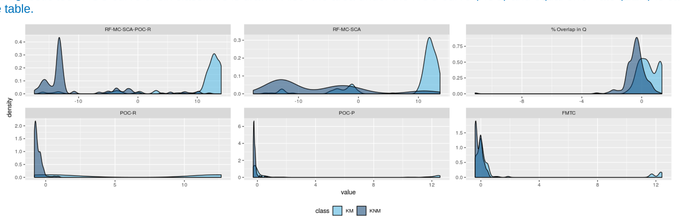
Density plot of KM and KNM from POC-R, C-P and FMTC.

- Performance on test set :



Classifiers	AUC	EER
RF-MC-SCA-POC-R	0.970	0.074
RF-MC-SCA	0.907	0.181
% Overlap on Q	0.835	0.246
POC - R	0.796	0.244
POC - P	0.705	0.387
FMTC	0.689	0.387

As a diagnostic tool, ROC curve of six classifiers is drawn. Also, the area under the ROC curve (AUC) and equal error rate (EER) is calculated in the table.



The density of score based likelihood ratio (SLR) in test comparisons is calculated by the training scores. The density of training score in KM and KNM is estimated by kernel density estimation.

7. References

- Richetelli, Nicole, et al. "Classification of footwear outsole patterns using fourier transform and local interest points." Forensic science international 275 (2017): 102-109.
- Wallis, W. D. "Graph Theory with Applications (JA Bondy and USR Murty)." (1979): 429.
- Kandel, Abraham, Horst Bunke, and Mark Last, eds. Applied graph theory in computer vision and pattern recognition. Vol. 52. Springer, 2007.
- Prewitt, Judith MS. "Object enhancement and extraction." Picture processing and Psychopictorics 10.1 (1970): 15-19.

This work was partially funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) through Cooperative Agreement #70NANB15H176 between NIST and Iowa State University, which includes activities carried out at Carnegie Mellon University, University of California Irvine, and University of Virginia.